

Sci-Hub: ¿hackear la ciencia?

Sci-Hub: hack science?

Francisco García-Ull 

Instituto Mediterráneo de Estudios de Protocolo – Centro Adscrito UMH, Profesor, Alicante, España, fran.garcia@protocoloimep.com, <https://orcid.org/0000-0002-7999-4807>

Resumen

El sitio web Sci-Hub permite a los usuarios descargar versiones en PDF de artículos académicos, incluyendo muchos documentos de acceso restringido y de pago. Sci-Hub cuenta en la actualidad con un corpus estimado de 64 millones de artículos académicos. En otras palabras, permite el acceso a casi toda la literatura académica internacional.

Sci-Hub ha liberalizado los datos de acceso a artículos de sus registros de servidor en 2017. Tras el procesamiento de los datos, podemos observar que Sci-Hub proporciona acceso a un promedio de 400.000 solicitudes válidas por día.

En este artículo se utiliza la minería de datos para ofrecer un análisis minucioso del uso de la plataforma en España, con el objetivo de conocer con precisión su nivel de uso y penetración. Para ello se analizan, entre otros indicadores, el número de descargas y su procedencia regional, así como la identificación de editoriales académicas que más descargas acumulan y su clasificación en distintas áreas del conocimiento.

La plataforma Sci-Hub, queda enmarcado en la filosofía open data, que persigue que determinados tipos de datos estén disponibles de forma libre para todo el mundo, sin restricciones o mecanismos de control. Se pretenden ofrecer por tanto en este estudio herramientas para el debate entre la preservación de los derechos de autor y el libre acceso a la información científica en línea, que elimine limitaciones en el intercambio de conocimiento.

Palabras clave: open data, open access, piratería, sci-hub, hacktivismo

Abstract

Sci-Hub website allows users to download PDF versions of scholarly articles, including many paid and restricted access documents. Sci-Hub currently has an estimated corpus of 64 million scholarly articles. In other words, it allows access to almost all international academic literature. Sci-Hub has liberalized article access data from its server logs in 2017. After processing the data, we can see that Sci-Hub provides access to an average of 400,000 valid requests per day.

In this article, data mining is used to offer a detailed analysis of the use of the platform in Spain, with the aim of knowing precisely its level of use and penetration. To do this, among other indicators, the number of downloads and their regional origin are analysed, as well as the identification of academic publishers that accumulate the most downloads and their classification in different areas of knowledge.

The Sci-Hub platform is framed within the open data philosophy, which seeks to make certain types of data freely available to everyone, without restrictions or control mechanisms. Therefore, this study intends to offer tools for the debate between the preservation of copyright

and free access to scientific information online, which eliminates limitations in the exchange of knowledge.

Keywords: open data, open access, piracy, sci-hub, hacktivism

1. INTRODUCCIÓN

El sitio web Sci-Hub permite a los usuarios descargar versiones en PDF de artículos académicos, incluyendo muchos documentos de acceso restringido y de pago. Sci-Hub ha crecido rápidamente desde su creación en 2011, contando en la actualidad con un corpus estimado de 64 millones de artículos académicos. En la actualidad, Sci-Hub brinda acceso a casi toda la literatura académica internacional (Himmelstein et al, 2018).

La plataforma, no exenta de polémica, se ha visto envuelta en distintos procesos judiciales, acusada de vulnerar los derechos de propiedad intelectual por parte de algunas de las principales editoriales científicas, como Elsevier.

Sci-Hub ha liberalizado los datos de acceso a artículos de sus registros de servidor en 2017. Tras el procesamiento de los datos, podemos observar que Sci-Hub proporciona acceso a un promedio de 400.000 solicitudes válidas por día.

En este artículo se utiliza la minería de datos para ofrecer un análisis minucioso del uso de la plataforma en España, con el objetivo de conocer con precisión su nivel de uso y penetración. Para ello se analizan, entre otros indicadores, el número de descargas y su procedencia regional, así como la identificación de editoriales académicas que más descargas acumulan (mediante el análisis del Identificador de Objeto Digital, DOI) y publicaciones científicas más descargadas. Se trata del primer estudio que analiza la incidencia y el uso de esta plataforma en el ámbito español.

La plataforma Sci-Hub, queda enmarcado en la filosofía open data, que persigue que determinados tipos de datos estén disponibles de forma libre para todo el mundo, sin restricciones o mecanismos de control. Se pretenden ofrecer por tanto en este estudio herramientas para el debate entre la preservación de los derechos de autor y el libre acceso a la información científica en línea (Oppenheim, 2020), que elimine limitaciones en el intercambio de conocimiento. No hay mejor medio para ello que Internet, cuya arquitectura en red se constituye como estructura que posibilita este intercambio de información “de carácter abierto, descentralizado, distribuido y multidireccional” (Castells, 2001).

Cambio de modelo en el mercado editorial académico

En los últimos años el mercado editorial académico se ha transformado de manera radical. Hasta la segunda mitad del siglo XX, la mayoría de las comunicaciones occidentales eran publicadas por sociedades científicas o profesionales, editoriales universitarias y editoriales educativas. Se trataba de un mercado ocupado por imprentas especializadas, con un modelo de negocio con características únicas, no estrictamente sometido a las leyes de la oferta y la demanda.

Como analiza Bodó (2016), se trata de un escenario donde la publicación de artículos científicos es bastante insensible a los cambios de precio. De un lado, los autores no publican con una finalidad económica y están dispuestos a proporcionar sus textos a la industria editorial de forma gratuita. Por otra parte, los artículos constituyen una aportación única a la investigación, por lo que la demanda de estos trabajos también actúa de manera independiente a su coste medido en valor monetario. De hecho, si bien existen publicaciones que generan ingresos con un modelo basado en suscripciones, el beneficio económico no es la prioridad y los cargos derivados son adoptados, tradicionalmente, por bibliotecas e instituciones, siendo pocos los lectores que realmente pagan directamente por las obras que consumen. La publicación académica, en esta primera etapa, sigue siendo principalmente un asunto interno de la comunidad científica.

En la segunda mitad del siglo XX, sin embargo, una serie de desarrollos llevaron a un aumento dramático de la influencia en el mercado de publicaciones académicas basadas en suscripciones, transformando un poco rentable mercado descentralizado, en una empresa global con gran rentabilidad y altamente concentrada, en su mayoría dirigida por editores comerciales.

Existen una serie de factores que propician este cambio de paradigma (Bodó, op. cit.):

- El auge posterior a la Segunda Guerra Mundial en la educación superior y la investigación occidentales, que aumentó el conjunto de autores, el número de artículos, el tamaño de las audiencias, y por lo tanto el tamaño del mercado editorial académico.
- El desarrollo de ISI (1960) y otros índices de citación académica, que identifican y acotan las principales revistas de cada disciplina, así como las publicaciones más relevantes en cada campo. Como resultado de estas métricas de citación e impacto aparentemente objetivas, las revistas se convierten en lugares de publicación altamente deseados tanto por los autores (cuya carrera académica depende de la publicación en revistas de gran prestigio), como por lectores (que confían en la selección de artículos y garantía de calidad de estas publicaciones). Esta demanda bilateral de determinadas publicaciones aumenta, a su vez, en gran medida su valor como potenciales activos comerciales.
- Digitalización, que reduce radicalmente los costos asociados a las tareas tradicionales del editor, sin suponer una disminución proporcional a los precios de las publicaciones.
- Una serie de fusiones y adquisiciones, que crean oligopolios en el mercado de las publicaciones académicas (como la fusión en 2015 de Springer Science y Business Media, Holtzbrinck Publishing Group's, Nature Publishing Group, Palgrave Macmillan, y Macmillan Education).
- La incapacidad de las bibliotecas para resistir los aumentos de precios en la era digital, viéndose obligadas a asumir los términos de acceso y licencia de las editoriales.
- La pasividad general de los científicos, que tolera las demandas de propiedad cada vez mayores de los editores. La publicación académica se convierte en herramienta para garantizar la primacía y la propiedad sobre las nuevas ideas y queda relegado a un segundo plano el papel de la publicación como medio de comunicación y difusión de las aportaciones académicas. El científico escribe para registrar su obra, no para difundir el conocimiento.

El modelo de acceso abierto

El movimiento de acceso abierto (open access, OA) a la literatura científica, supone empoderar a los usuarios para leer, descargar, copiar, distribuir, imprimir, buscar o enlazar los textos completos de los artículos científicos y usarlos con cualquier otro propósito legítimo. Este movimiento se enmarca en la filosofía open data, que persigue que determinados tipos de datos estén disponibles de forma libre para todo el mundo, sin restricciones o mecanismos de control.

Según Dialnet (2016) el acceso abierto se justifica “si tenemos en cuenta que la literatura científica se genera en la mayor parte de las ocasiones en el seno de entidades públicas, costeadas con recursos públicos. En contraste con ello, esta literatura científica ha venido estando monopolizada por los grandes grupos editoriales que cada vez han ido poniendo precios más elevados a la distribución de la información generada gracias a las aportaciones públicas”.

Como señala Suber (2012), el acceso abierto tiene tres puntos clave:

- Hacer que la literatura científica se encuentre disponible en línea.
- Eliminar las barreras económicas.
- Eliminar la mayoría de las barreras de permisos de reutilización.

El acceso abierto promueve eliminar las barreras económicas, legales y tecnológicas, tratando de obtener a cambio, como beneficios, una mayor accesibilidad para los documentos y una mayor visibilidad para los autores. De hecho, los documentos que están disponibles libremente son más consultados y tienen más posibilidades de ser citados (Eysenbach, 2006). Por otra parte, otro efecto deseado consiste en que los conocimientos y avances científico-técnicos se distribuyan de la manera más amplia posible, devolviéndole así a la sociedad los frutos de las investigaciones por ella subsidiada. En otros términos, devolver a la sociedad los beneficios de las inversiones realizadas en investigaciones científicas (Sala, 2011).

En la actualidad, como respuesta a este nuevo escenario de oligopolio editorial (Larivière et al., 2015) y debido, entre otros factores, al incremento en los precios de las suscripciones, el movimiento de acceso abierto ha comenzado a reclamar protagonismo en el proceso de publicación académica (Royster, 2016). Además del impacto académico y económico, el impacto social potencial de la publicación en acceso abierto está recibiendo una mayor atención (Tennant et al., 2016) y las grandes instituciones parecen estar de acuerdo con esta opinión, ya que cada vez más adoptan políticas de acceso abierto al material académico (Giles, 2004; Pérez et al, 2007; Melero, 2008).

Estos esfuerzos parecen tener un impacto, ya que un estudio de 2014 sobre publicaciones académicas en inglés descubrió que, si bien la adopción de acceso abierto varía entre las disciplinas académicas, un promedio de alrededor del 24% de los documentos académicos son accesibles de forma gratuita en la web (Khabsa, 2014).

Otra respuesta a estos cambios en el mundo editorial académico es el movimiento denominado Guerrilla Open Access (Balázs, 2016), Bibliogifts (Cabanac, 2015) o Black Open Access (Björk, 2017). Se trata, en resumen, del uso de "repositorios ilegales" (Azorín, 2018) para acceder a publicaciones académicas, como en el caso de Sci-Hub o LibGen (Gardner et al, 2017).

Especialmente Sci-Hub, que comenzó en 2011, se ha convertido en el centro de atención en los últimos años. Según la fundadora, Alexandra Elbakyan, el sitio web utiliza las credenciales de las bibliotecas donadas por contribuyentes para burlar los sistemas de acceso restringido (muros de pago o paywalls) de los editores y, por lo tanto, descarga gran parte de sus colecciones (Bohannon, 2016). Esta clara violación de los derechos de autor no solo condujo a una demanda por parte de Elsevier contra Elbakyan (Bernardo, 2017), sino también a que se la llamara "Robin Hood de la Ciencia" (Martín, 2017), episodios que han situado al repositorio Sci-Hub en el foco de la atención mediática.

A pesar de esto, ha habido poca investigación sobre cómo se utiliza Sci-Hub y a qué tipo de materiales se accede a través de él. Un estudio de 2014 analizó el contenido proporcionado a través de LibGen (Cabanac op. cit). En 2016, Sci-Hub publicó datos sobre 28 millones de descargas realizadas a través del servicio. Posteriormente, se analizaron estos datos para ver en qué países se está utilizando el sitio web, qué editores son más frecuentes y cómo la descarga de publicaciones a través de Sci-Hub se relaciona con factores socioeconómicos, como estar en un centro de investigación (Greshake, 2016) y cómo afecta a los préstamos interbibliotecarios (Gardner, op. cit.) y a los procesos de los repositorios bibliográficos (González-Solar, et al, 2019). En 2017, Sci-Hub publicó la lista de 62 millones de identificadores de objetos digitales (DOI) del contenido almacenado en su repositorio. El análisis de estos datos aporta cifras interesantes, como que aproximadamente un tercio de los artículos descargados corresponden a la editorial Elsevier o que, a nivel de publicaciones, las descargas muestran un sesgo hacia algunas disciplinas académicas, especialmente Química, lo que sugeriría mayores barreras de acceso para éstas (Greshake, op. cit.).

En enero de 2018, Sci-Hub publica el registro de descargas de la primera mitad de 2017 (en concreto, desde el 1 de enero hasta el 25 de junio: un total de 164 días). Este estudio es el primero en analizar los datos de descargas de las publicaciones a través de Sci-Hub, con el corpus completo publicado por el propio repositorio y que se centra específicamente en el análisis de su impacto en el ámbito español. Para ello se analiza mediante la minería de datos un corpus de más de 66 millones de descargas, con la intención de comprender cómo utilizan los usuarios españoles este repositorio, así como identificar cuáles son las editoriales y publicaciones académicas más descargadas.

2. METODOLOGÍA

El corpus de 66.352.252 descargas indexadas por Sci-Hub se obtuvo a partir del conjunto de datos publicado por el repositorio el 18 de enero de 2018. Para este análisis, se utilizó el software IBM SPSS Statistics 25 para identificar y clasificar las variables del conjunto de datos. Una vez

determinadas las mismas, se seleccionaron las siguientes variables: Fecha, País, Ciudad, DOI, Latitud y longitud.

Para identificar las editoriales que más descargas acumulan por parte de los usuarios españoles de Sci-Hub, se seleccionaron los DOI (Digital Object Identifier) de cada artículo y dividido en dos partes: la primera parte corresponde al editor, la segunda al objeto. Los DOI tienen un sistema jerárquico y son las empresas, normalmente las mismas editoriales, las que solicitan registrar sus objetos. Un DOI tiene dos partes separadas por una barra [/]. Una parte identifica a la entidad que registra y la otra parte identifica al objeto. Por ejemplo, el siguiente DOI: 10.1145/1067268.1067287, corresponde al documento “1067268.1067287” registrado por la ACM (código “10.1145”). Conociendo el DOI se puede llegar al artículo, utilizando el siguiente formato: <http://dx.doi.org/10.1145/1067268.1067287>. También la herramienta Google Scholar permite buscar artículos por DOI.

La geolocalización, mapa de calor y representaciones gráficas se ha elaborado introduciendo los datos correspondientes en la herramienta Google Fusion Tables. Para cada editor, el número de artículos descargados en España se comparó con el número esperado de descargas de la base de datos global de Sci-Hub.

3. RESULTADOS

3.1. Países con mayor número de descargas

China es el país que acumula un mayor número de descargas de artículos científicos desde el repositorio Sci-Hub durante la primera mitad de 2017, con un total de 10.977.138, que corresponde a un 16,5% del total de descargas registradas. España está en novena posición, con un total de 1.610.066 descargas (2,4% del total), siendo el segundo país europeo que acumula un mayor número de descargas, solo por detrás de Francia (1.916.217 descargas).

En los casos en los que no se incluya la sección de marco teórico o referencial en el artículo, deben quedar claros los apartados mencionados aquí en la sección introducción. Es importante utilizar representaciones gráficas para aclarar ideas e ilustrar conceptos y conclusiones, como lo son:

Tabla 1. Frecuencia de descargas por países y porcentaje respecto al total

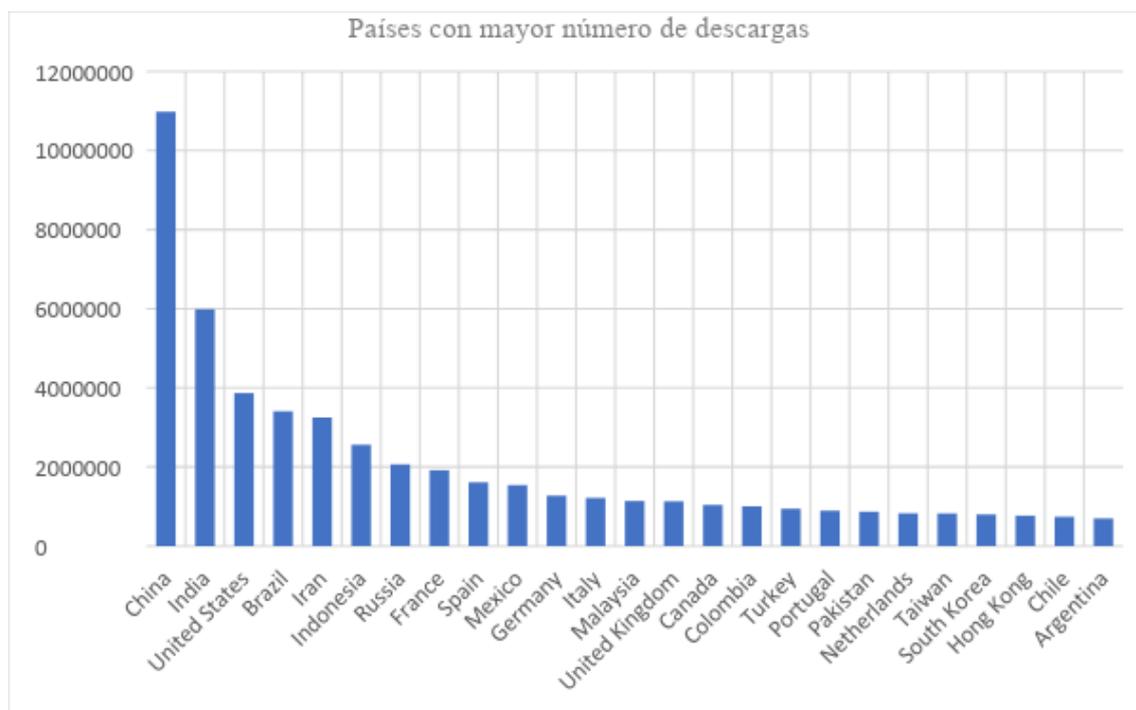
País	Frecuencia	Porcentaje respecto al total
China	10977138	16,5
India	5989065	9
United States	3866994	5,8
Brazil	3406326	5,1
Iran	3249740	4,9
Indonesia	2561122	3,9
Russia	2064172	3,1
France	1916217	2,9
Spain	1610066	2,4
Mexico	1539480	2,3
Germany	1273403	1,9
Italy	1216889	1,8

Malaysia	1138590	1,7
United Kingdom	1129746	1,7
Canada	1038091	1,6
Colombia	1003340	1,5
Turkey	942382	1,4
Portugal	892694	1,3
Pakistan	867101	1,3
Netherlands	826963	1,2
Taiwan	823548	1,2
South Korea	798215	1,2
Hong Kong	767697	1,2
Chile	739165	1,1
Argentina	697370	1,1

Fuente: elaboración propia

Como podemos observar en la Figura 1, China es con una importante diferencia la mayor consumidora de artículos científicos en Sci-Hub, seguida por India y Estados Unidos.

Figura 1. Los 25 países con mayor número de descargas en Sci-Hub



Fuente: elaboración propia

3.2. Descargas por ciudades en España

En la primera mitad de 2017 se producen descargas de artículos científicos desde un total de 1.648 localidades españolas. La ciudad con una mayor frecuencia de descargas es Madrid (con un total

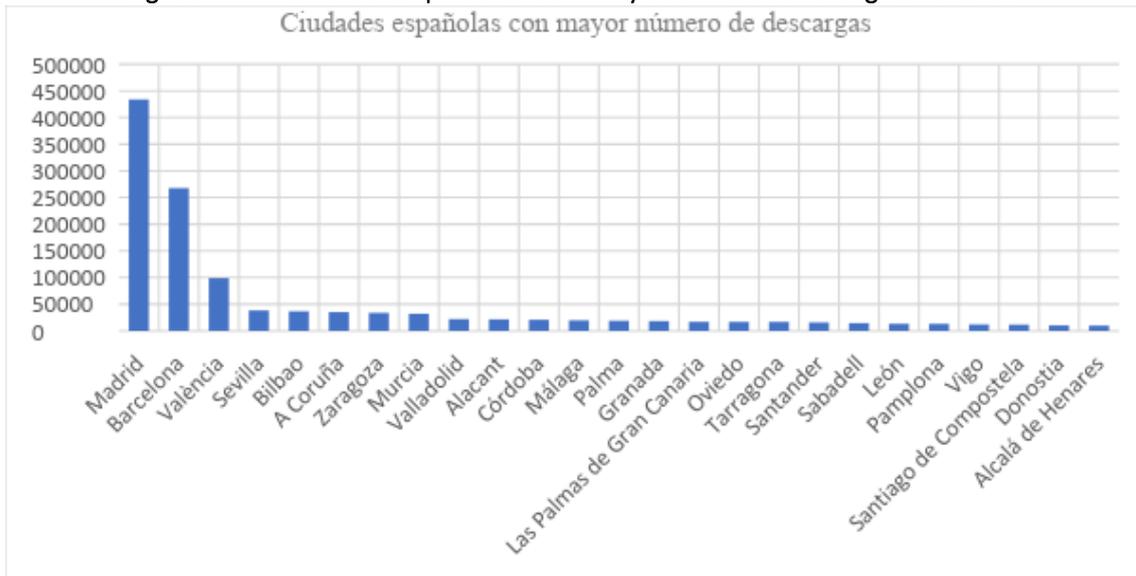
de 433.662 registros; el 26,9% del total en España), seguida por Barcelona (267.787; el 16,6% del total), Valencia (98.916), Sevilla (38.202) y Bilbao (36.646 descargas).

Tabla 2. Frecuencia de descargas por ciudades españolas y porcentaje respecto al total

Ciudad	Frecuencia	Porcentaje
Madrid	433662	26,9
Barcelona	267787	16,6
València	98916	6,1
Sevilla	38202	2,4
Bilbao	36646	2,3
A Coruña	35065	2,2
Zaragoza	33683	2,1
Murcia	32132	2
Valladolid	21997	1,4
Alacant	21572	1,3
Córdoba	20760	1,3
Málaga	19681	1,2
Palma	18849	1,2
Granada	18249	1,1
Las Palmas de Gran Canaria	16881	1
Oviedo	16802	1
Tarragona	16796	1
Santander	15702	1
Sabadell	14426	0,9
León	13149	0,8
Pamplona	13075	0,8
Vigo	11910	0,7
Santiago de Compostela	11571	0,7
Donostia	10134	0,6
Alcalá de Henares	9782	0,6

Fuente: elaboración propia

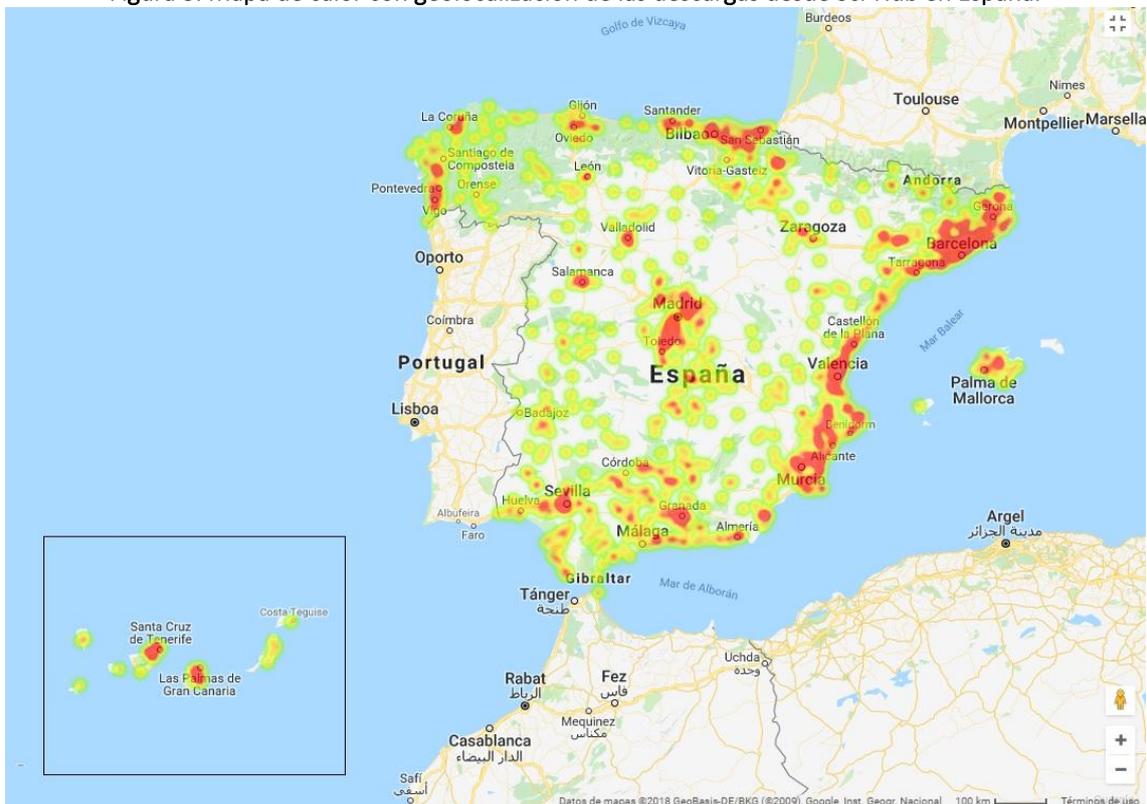
Figura 2. Las 25 ciudades españolas con un mayor número de descargas en Sci-Hub



Fuente: elaboración propia

Como se observa en el siguiente mapa de calor (Figura 3), la Comunidad de Madrid, el País Vasco, las principales capitales de Andalucía y Galicia, así como el litoral mediterráneo, son las zonas geográficas que presentan un mayor número de descargas desde Sci-Hub.

Figura 3. Mapa de calor con geolocalización de las descargas desde Sci-Hub en España.



Fuente: elaboración propia

3.3. Frecuencia de descargas por fecha

Durante la primera mitad del año 2017, los usuarios españoles descargan un promedio de 9.817 artículos científicos al día desde la web Sci-Hub. Mayo es el mes en el que se alcanzan un mayor número de descargas, con un total de 343.103, mientras que enero es el mes con menos registros, con un total de 213.494.

Tabla 3. Número de descargas en España por meses

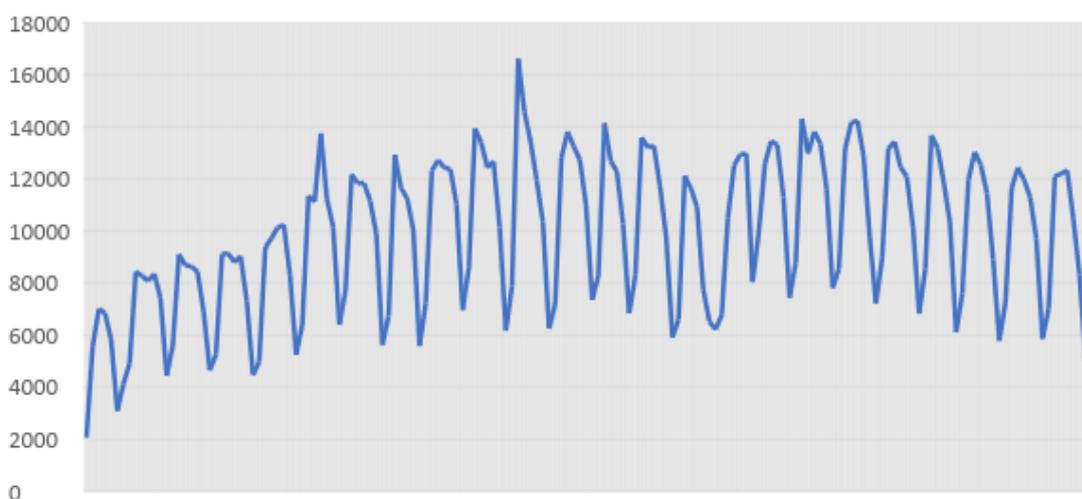
Mes	Descargas
Enero	213494
Febrero	264847
Marzo	336988
Abril	208101
Mayo	343103
Junio	243533
Total general	1610066

Fuente: elaboración propia

Como podemos comprobar, el gráfico en el que se representan las descargas de artículos científicos en España (Figura 4), expresa intervalos irregulares. Los picos se relacionan con los días laborales, mientras que los descensos en descargas se corresponden a fines de semana y días festivos. El lunes 13 de marzo de 2017, es el día que más descargas se registran, con un total de 16.616. De otro lado, el día 1 de enero es el día con un menor número de descargas, con un total de 2.058.

Figura 4. Descargas por día en España desde Sci-Hub

Descargas por día en España desde Sci-Hub en la primera mitad de 2017



Fuente: elaboración propia

3.4 Descargas por editorial

Se descargan en España en la primera mitad de 2017 desde Sci-Hub artículos académicos de 1.194 editores diferentes.

Tabla 4. Editoriales que acumulan más descargas en España

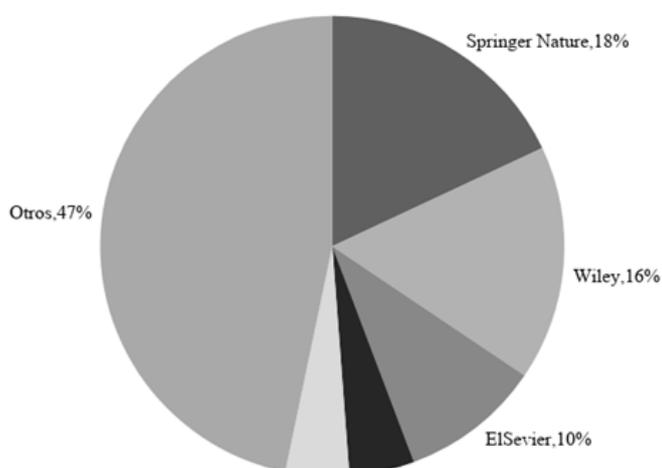
DOI	Descargas	Porcentaje respecto al total	Editorial
101007	289794	18,00	Springer Nature
101016	265182	16,47	Wiley
101002	157903	9,81	ElSevier
101111	73156	4,54	ACS
101038	71436	4,44	Taylor & Francis
101021	53767	3,34	NCBI
101080	44780	2,78	Sage Journals
101097	37781	2,35	RSC Royal Society Chemistry
101177	34725	2,16	IEEE
101039	29981	1,86	Oxford Academic Journals

Fuente: elaboración propia

Como se observa en la Tabla 4, la editorial con más publicaciones descargadas en España desde Sci-Hub es Springer Nature, con 289.794 descargas (192.655 Springer y 97.139 Nature), conformando un 18% del total de descargas. Wiley, que suma con dos DOI un total de 265.182 artículos descargados (un 16,47% del total) es la segunda editorial con más descargas, seguida por ElSevier (157.903 descargas y 9,81% del total), American Chemical Society (73.156 descargas) y Taylor & Francis (71.436 descargas).

De este modo, cinco editoriales (Springer Nature, Wiley, ElSevier, ACS y Taylor & Francis) suman más de la mitad de los artículos descargados en España desde Sci-Hub (en concreto el 53,26%).

Figura 5. Proporción de descargas por editorial con respecto a la totalidad de descargas en España.



Fuente: elaboración propia

3.5 Artículos más descargados

Los artículos más descargados en España durante la primera mitad de 2017 desde Sci-Hub pertenecen en su mayoría al campo de las ciencias tecnológicas. Podemos observar en este campo publicaciones correspondientes a disciplinas como la tecnología de materiales, tecnologías de construcción o química industrial. También son destacables las descargas de publicaciones del campo de las ciencias médicas. Entre los 10 artículos más descargados en España (Tabla 5), observamos que los tres con mayor número de accesos superan el millar de registros. Si analizamos la totalidad de la muestra, es destacable el hecho de que tan solo 17 publicaciones superen el centenar de descargas, dato que demuestra la gran disparidad de artículos que se han consultado desde el territorio español durante el periodo de tiempo analizado.

Tabla 5. Artículos científicos más descargados en España

DOI	Descargas	Editorial	Idioma	Campo	Año
10.1002/app.26296	1609	Wiley Online Library	Inglés	Ciencias tecnológicas	2007
10.1103/PhysRevB.73.153303	1609	APS	Inglés	Física	2006
10.12968/jowc.2017.26.sup2.s18	1227	Magonlinelibrary	Inglés	Ciencias médicas	2017
10.1002/9781444338386.wbeah03257	281	Wiley Online Library	Inglés	Historia	2013
10.1080/14786440808520347	246	Taylor & Francis	Inglés	Ciencias tecnológicas	1953
10.1002/14651858	232	Wiley Online Library	Inglés	Ciencias médicas	2004
10.1021/ie00106a015	188	ACS Publications	Inglés	Ciencias tecnológicas	1990
10.1021/ie010218+	175	ACS Publications	Inglés	Ciencias tecnológicas	2001
10.1016/s0257-8972(02)00116-0	164	Elsevier	Inglés	Ciencias tecnológicas	2002
10.1080/13606710802200829	156	Taylor & Francis	Inglés	Ciencias económicas	2008

Fuente: elaboración propia

4. CONCLUSIONES

Los análisis presentados aquí sugieren que Sci-Hub se utiliza por diversas razones y por distintos tipos de usuario. Si bien la mayoría del uso se centra en artículos publicados en los últimos veinte años, hay un subconjunto de usuarios interesados en obtener literatura académica histórica.

Tanto en el caso de España, como en su comparación con el conjunto de datos global, Sci-Hub parece ser una herramienta recurrente para profesionales del ámbito de las ciencias tecnológicas, especialmente para ingenieros y químicos industriales, como lo demuestra la extensa representación de artículos descargados que conforma la muestra.

Este hecho puede deberse al limitado acceso a la producción científica del que se dispone desde determinados sectores profesionales. Se trata de población que podría no tener herramientas para utilizar los repositorios académicos de las bibliotecas universitarias o centros de investigación y que se ve en la necesidad imperiosa de consultar un determinado texto científico. Los elevados precios de las suscripciones para lectores individuales y la accesibilidad ubicua e inmediata que proporciona Sci-Hub, hacen a estos usuarios recurrir a plataformas de Guerrilla Open Access, como la aquí analizada. El hecho de que la mayoría de las descargas se realicen durante los días laborales, refuerza esta idea.

En cuanto a la geolocalización de las descargas, no debería sorprendernos el hecho de que China sea, con mucha diferencia, el país desde donde se realizan el mayor número de accesos a artículos académicos. De hecho, de acuerdo a la Fundación Nacional para la Ciencia estadounidense (2018), China se convirtió en 2016 en el principal productor de literatura científica a nivel mundial, superando por primera vez a Estados Unidos.

Centrándonos en el caso español, es destacable que España se encuentre en la novena posición a nivel mundial en el número de descargas desde Sci-Hub. El análisis por ciudades nos ayuda a entender que los datos de acceso no están necesariamente relacionados con el número de habitantes. Encontramos, sin embargo, una mayor concordancia si relacionamos los datos de descargas de cada ciudad con la inversión en I+D de su Comunidad Autónoma.

Por último, cuando se trata de la representación de los editores, los datos de Sci-Hub muestran una concordancia en términos de uso real en comparación con la cantidad de literatura publicada, así como la evidencia de un oligopolio en el mercado editorial académico.

5. LIMITACIONES DE INVESTIGACIÓN Y FUTURAS LÍNEAS

Dada la idiosincrasia del portal analizado, así como sus constantes cambios de servidor (con el objetivo de burlar las restricciones de acceso vía IP), resulta complicado obtener información sobre el tráfico web o datos de usuario, más allá de los publicados eventualmente por la propia plataforma. Una comunicación directa con el personal responsable de Sci-Hub ayudaría, sin duda, a configurar un escenario más aproximado al delimitado en este estudio. En cuanto a futuras líneas de investigación, la autoría entiende que sería de gran interés un estudio longitudinal para evaluar la evolución de los perfiles de usuario y descargas. Un análisis de tipo cualitativo-cuantitativo dirigido a usuarios de Sci-Hub arrojaría también luz sobre los distintos perfiles de usuario y ayudaría a comprender sus necesidades y, sobre todo, su motivación a utilizar la herramienta.

Agradecimientos

A mi viejo Lenovo 320-15ISK que con su procesador i3 y paciencia, ha analizado una base de datos de casi 70 millones de registros para realizar este artículo.

REFERENCIAS

- Andročec, D. (2017). Analysis of Sci-Hub downloads of computer science papers. *Acta Universitatis Sapientiae, Informatica*, 9(1), 83-96.
- Azorín, C. (2018). Intuiciones y certezas sobre los repositorios institucionales. En *InnovaTics. Congreso Internacional de Innovación Tecnológica* (No. 8º).
- Bernardo, A. (23 de junio de 2017) Condena millonaria a Sci-Hub, el 'Pirate Bay' de la ciencia. *Hipertextual*. Recuperado de: <https://hipertextual.com/2017/06/sci-hub-elsevier>
- Björk, B. C. (2017). Gold, green, and black open access. *Learned Publishing*, 30(2), 173-175
- Bodó, Balázs, (2016) Pirates in the Library – An Inquiry into the Guerilla Open Access Movement (July 6, 2016). Paper prepared for the 8th Annual Workshop of the International Society for the History and Theory of Intellectual Property, CREATE, University of Glasgow, UK, July 6-8, 2016., Available at SSRN: <https://ssrn.com/abstract=2816925> or <http://dx.doi.org/10.2139/ssrn.2816925>
- Bohannon J. (2016). Who's downloading pirated papers? Everyone. *Science Magazine*. Recuperado de: <http://www.sciencemag.org/news/2016/04/whos-downloading-pirated-papers-everyone>.
- Cabanac, G. (2016). Bibliogifts in Lib Gen? A study of a text-sharing platform driven by biblioleaks and crowdsourcing. *Journal of the Association for Information Science and Technology*, 67(4), 874-884.
- Castells, M. (2001). *La Galaxia Internet*. Madrid: Areté
- Davis, PM. (2017). Sci-Hub Downloads of APS Papers. *The Physiologist* 60: 3-9.
- Dialnet. (2016) El Acceso Abierto a la literatura científica. Recuperado de: <https://dialnet.unirioja.es/info/ayuda/oai>
- Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS biology*, 4(5), e157.

- Gardner, G. J., McLaughlin, S. R., & Asher, A. D. (2017). Shadow libraries and you: Sci-hub usage and the future of ill. In ACRL 2017, Baltimore, Maryland, March 22 - 25, 2017. [Conference paper]
- Giles, J. (2004). Trust gives warm welcome to open access. *Nature*.
- González-Solar, L., & Fernández-Marcial, V. (2019). Sci-Hub, a challenge for academic and research libraries. *Profesional De La información*, 28(1). <https://doi.org/10.3145/epi.2019.ene.12>
- Greshake, B. (2017). Looking into Pandora's Box: The content of Sci-Hub and its usage. *F1000Research*, 6.
- Himmelstein, D. S., Romero, A. R., Levernier, J. G., Munro, T. A., McLaughlin, S. R., Tzovaras, B. G., & Greene, C. S. (2018). Sci-Hub provides access to nearly all scholarly literature. *eLife*, 7, e32822.
- Khabsa, M., & Giles, C. L. (2014). The number of scholarly documents on the public web. *PloS one*, 9(5), e93949.
- Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PloS one*, 10(6), e0127502.
- Martín, B. (3 de diciembre de 2017) La ley alcanza, pero no derrota a la Robin Hood de la ciencia. *El País*. Recuperado de: https://elpais.com/elpais/2017/11/29/ciencia/1511971491_929151.html
- Melero, R. (2008). El paisaje de los repositorios institucionales open access en España. *BiD textos universitaris de biblioteconomia i documentació*. Universitat de Barcelona
- Murphy K. (2016) Should all research papers be free? New York Times. Recuperado de: <http://www.nytimes.com/2016/03/13/opinion/sunday/should-all-research-papers-be-free.html?smid=tw-share>.
- Oppenheim, C. (2020). A copyright overview. *Profesional De La información*, 29(1). <https://doi.org/10.3145/epi.2020.ene.06>
- Pérez, T. H., Mateos, D. R., & De la Fuente, G. B. (2007). Open Access: el papel de las bibliotecas en los repositorios institucionales de acceso abierto. En *Anales de documentación* (Vol. 10, pp. 185-204).
- Rosenwald, M. S. (2016). This student put 50 million stolen research articles online. And they're free. *Washington Post*, March, 30.
- Sala, H. E. (2011). Origen, consolidación, expansión e implicancias del Acceso Abierto (Open Access) en América Latina y el Caribe. *Revista Educación Superior y Sociedad*.
- Suber, P. (2012). Open access overview. *MIT Press*.
- Tennant, J. P., Waldner, F., Jacques, D. C., Masuzzo, P., Collister, L. B., & Hartgerink, C. H. (2016). The academic, economic and societal impacts of Open Access: an evidence-based review. *F1000Research*, 5.

Como citar:

García-Ull, F. (2022). Sci-Hub: ¿hackear la ciencia? *SIGNUM: Revista Internacional de Investigación en Eventos, Protocolo y Relaciones Institucionales*, 1(1): 11-23.