

ARTÍCULO ORIGINAL



Buenas prácticas sobre publicaciones en acceso abierto de proyectos financiados por Horizonte 2020

1

Good Practices on Open Access Publications of Projects Funded by Horizon 2020

Elena Pastor-Ramon

Biblioteca Virtual de Ciències de la Salut de les Illes Balears (Bibliosalut)
elena.pastor@biblios Salut.com

Juan Antonio Casado

Responsable de tecnología y data en AlterBiblio
juanantonio@alterbiblio.com

Alejandra Campos

Universitat Autònoma de Barcelona
alejandra.campos@uab.cat

Received: 15-11-2024
Accepted: 13-01-2025

Resumen

Este estudio tuvo como objetivo analizar resultados de las políticas de publicación en acceso abierto establecidas para los proyectos financiados por el programa Horizonte 2020 de la Unión Europea. Se buscó específicamente: Identificar los requisitos específicos de publicación, cuantificar el número de artículos revisados por pares publicados en acceso abierto y analizar la cantidad de documentos depositados en repositorios institucionales. Como base de datos se utilizó el archivo de publicaciones de proyectos Horizonte 2020 proporcionado por la Oficina de Publicaciones de la UE. Se pudo determinar las publicaciones revisadas por pares informadas. A partir de aquellas que contaban con DOI, se cruzaron los datos con la plataforma *OpenAIRE* para analizar la información de acceso abierto y repositorios. A partir de los datos originales, se observa un total de 206.400 artículos revisados por pares, de los cuales 202.102 datos cuentan con DOI, 4.269 incluyen el ISSN de la publicación y en solo 29 no figuran datos para alcanzar el acceso. Cruzando los datos de aquellos que tienen DOI en *OpenAIRE* un total de 126.154 de estos artículos no cuentan con metadatos. Este análisis destaca la necesidad de mejorar el rastreo de la dirección de acceso a las publicaciones y fomentar el depósito en repositorios de libre acceso con licencias abiertas.

Palabras clave

Horizonte 2020; programas de ayudas; acceso abierto; difusión de la ciencia; financiación pública

Abstract

This study aims to analyze the results of open access publishing policies established for projects funded by the European Union's Horizon 2020 programme. It specifically seeks to identify specific

publication requirements, quantify the number of peer-reviewed articles published in open access, and analyze the number of documents deposited in institutional repositories. The archive of Horizon 2020 project publications provided by the EU Publications Office was used as a database. It was possible to determine the reported peer-reviewed publications. From those with DOIs, the data were cross-checked with the *OpenAIRE* platform to analyze open access and repository information. From the original data, a total of 206,400 peer-reviewed articles are observed, of which 202,102 have a DOI, 4,269 include the publication's ISSN, and only 29 have no data to reach access. Cross-referencing the data from those with DOIs in *OpenAIRE*, it is shown that a total of 126,154 of these articles have no metadata. This analysis highlights the need to improve the tracking of the access address of publications.

Keywords

Horizon 2020; grant programs; open access; science dissemination, public funding

Citación recomendada

Pastor-Ramon, Elena, Casado, Juan Antonio y Campos, Alejandra (2024). Buenas prácticas sobre publicaciones en acceso abierto de proyectos financiados por Horizonte 2020. *BiD*, 53. <https://doi.org/10.1344/bid2024.53.05>

1. Introducción

La Unión Europea (UE) invierte dinero público para implementar estrategias incluidas en sus agendas políticas, atendiendo retos sociales y promoviendo el desarrollo económico. La Comisión Europea (CE), como poder ejecutivo, diseña y ejecuta planes para alcanzar estos objetivos.

Desde 1984, los Programas Marco han sido el principal instrumento de inversión en investigación e innovación, evolucionando de 3,300 millones de euros en su primera edición a 80,000 millones en Horizonte 2020 (2014-2020). Esta inversión busca generar conocimientos, productos y servicios que respondan a los retos de la sociedad, bajo las directrices de las políticas públicas de la UE.

Tabla 1. Programas Marco de la Unión Europea. Fuente: elaboración propia a partir de datos de la CE.

Año	1984	1987	1990	1994	1998	2002	2007	2014	2021
Nombre del Programa Marco	Primer Programa Marco	Segundo Programa Marco	Tercer Programa Marco	Cuarto Programa Marco	Quinto Programa Marco	Sexto Programa Marco	Séptimo Programa Marco	Programa Marco Horizonte 2020	Programa Marco Horizonte Europa
Período	1984-1987	1987-1991	1990-1994	1994-1998	1998-2002	2002-2006	2007-2013	2014-2020	2021-2027
Financiación (en millones de euros)	3.300	5.400	6.600	13.200	14.900	19.300	55.900	80.000	99.500

Como se observa en la Tabla 1, la inversión en ciencia e innovación de la UE ha crecido de forma significativa. Estos programas fueron diseñados para cumplir las políticas estratégicas europeas, con el objetivo de generar un impacto positivo en la economía, la sociedad, el medio ambiente y la innovación (Chirlomez 2019).

De este modo, la inversión en ciencia e innovación actúa como un motor clave de la economía basada en el conocimiento. Su propósito es incrementar el PIB mediante soluciones innovadoras de carácter interdisciplinar, intersectorial e internacional, beneficiando directamente a la sociedad que financia estas actividades con sus impuestos (Campos 2023).

Horizonte 2020 fue diseñado para fomentar un crecimiento económico inteligente, sostenible e inclusivo. Alineado con la estrategia Europa 2020, este programa buscó recuperar la economía tras la crisis de 2007. Su enfoque estaba en financiar avances, descubrimientos e innovaciones que transformaran ideas de laboratorios en productos y servicios de mercado (Comisión Europea 2014).

Según Carlos Moedas, comisario de investigación e innovación en ese momento, la inversión en H2020 se proyectaba con un impacto multiplicador de 8,5 veces sobre el PIB hacia 2030, además de la creación de miles de empleos. Este impacto trascendió a Europa, involucrando a participantes de prácticamente todo el mundo (Campos 2023).

El programa también se centró en financiar los mejores equipos, ya fuera en consorcios o de manera individual, a través de convocatorias abiertas y competitivas. Los resultados obtenidos se diseminaron ampliamente, siendo comunicados a audiencias diversas o explotados mediante patentes e innovaciones aplicadas.

En este artículo, hemos analizado los aspectos relacionados con la diseminación de los resultados de proyectos financiados por Horizonte 2020 (H2020), con especial énfasis en las publicaciones. Para ello, utilizamos documentación de referencia de la Oficina de Publicaciones de la Unión Europea disponible en su sitio web (Publications Office of the European Union, 26/6/2023).

Esta investigación reviste importancia porque H2020 ha destinado un total de 83.270 millones de euros para financiar 35,426 proyectos, con la participación de entidades de 178 países, incluidas universidades, centros de investigación, empresas, instituciones públicas, asociaciones y usuarios finales ("R&I Projects - Key Figures" 2024). Hasta su finalización, fue el programa público de investigación e innovación más grande del mundo (Reglamento del Parlamento Europeo y el Consejo 2018; Comisión Europea 2014), consolidándose como el programa europeo de referencia para todo el ecosistema de ciencia e innovación.

H2020 financió entidades de los 27 países de la UE, sus 24 territorios de ultramar, 16 países asociados y 124 terceros países con acuerdos de participación (EU Funding & Tenders Portal, 2024; Comisión Europea 2020).

Este programa marco incorporó el concepto de "investigación e innovación responsables" (RRI), que incluye aspectos como ética, ciencia abierta, balance de género, gobernanza colectiva, educación científica y participación ciudadana (RRI Tools, 2024). Estas herramientas fueron diseñadas para fomentar proyectos de ciencia e innovación como motores del crecimiento económico, maximizando las sinergias entre ciencia, tecnología, innovación y producción.

En este marco, H2020 promovía la ciencia abierta como la forma de hacer y concebir una investigación de vanguardia, aplicando el principio de la CE "open as possible, closed as necessary". Esto impulsó la publicación en acceso abierto, que constituye el enfoque principal de esta investigación.

2. Objetivos

El presente estudio tiene como objetivo analizar el cumplimiento de las obligaciones de diseminación en acceso abierto de los proyectos financiados por el programa Horizonte 2020 de la Unión Europea. En particular, se centra en el análisis de las publicaciones

reportadas por los beneficiarios, evaluando su publicación en acceso abierto o su depósito en un repositorio, con base en los datos proporcionados por la Oficina de Publicaciones de la Unión Europea.

El objetivo general de este estudio es observar y analizar las cifras de publicaciones en acceso abierto reportadas en el marco de los proyectos de Horizonte 2020, con los siguientes objetivos específicos:

- Analizar los requisitos específicos de publicación en acceso abierto de H2020 para los artículos revisados por pares.
- Detectar la cantidad de artículos revisados por pares publicados en acceso abierto y la información que se provee para acceder a ellos.
- Identificar la cantidad de artículos revisados por pares depositados en un repositorio.

4

A partir de estos objetivos específicos, se plantean las preguntas de investigación siguientes:

- P1: ¿Cuáles son los requisitos de H2020 respecto a la publicación en acceso abierto?
- P2: ¿Cuántos documentos que deben cumplir estos requisitos, reportados en el archivo facilitado por la Oficina de Publicaciones de la Unión Europea, han sido publicados en acceso abierto?
- P3: ¿Cuántos documentos reportados en el mismo archivo han sido depositados en un repositorio?

3. Marco teórico

La Comisión Europea (CE) define la “diseminación” como los “medios para hacer públicos los resultados de un proyecto (por cualquier medio apropiado que no sea protegerlos o explotarlos, por ejemplo, mediante publicaciones científicas)” (DG EAC | Knowledge for Policy, 2023).

Esta actividad constituye una práctica esencial dentro de la comunidad científica, permitiendo la publicación de resultados o nuevos hallazgos de acciones de investigación e innovación para avanzar en el estado de la cuestión de un ámbito científico (The EU Guide to Science Communication - YouTube, 2024).

La diseminación de un proyecto, sin embargo, solo es posible si existen resultados tangibles derivados de éste (Comisión Europea 2023). No se limita a las publicaciones científicas (“DESCA: Horizon 2020 Model Consortium Agreement” 2016), sino que incluye todas aquellas acciones y actividades que facilitan dar a conocer los resultados a audiencias específicas, como científicos, grupos de pares, industrias e innovadores, quienes pueden utilizarlos para avanzar en sus propios ámbitos, proyectos o nuevos desarrollos (Comisión Europea 2023).

En este contexto, se fomenta que la diseminación se realice mediante canales de publicación en acceso abierto. Esta actividad es fundamental en los proyectos financiados por la UE, ya que representa una de las tres actividades clave que maximizan la inversión pública en ciencia e innovación, junto con la comunicación y la explotación de resultados (Campos 2023; Mea et al. 2016).

3.1. Publicación en acceso abierto

La Declaración de Budapest (Budapest Open Access Initiative, 2024) marca un hito al combinar una tradición ancestral con la tecnología moderna para crear un bien público único: el acceso abierto. Históricamente, científicos y académicos han estado dispuestos a compartir los resultados de sus investigaciones gratuitamente por el avance del conocimiento. Internet ha ampliado esta tradición al permitir la distribución electrónica mundial, ofreciendo acceso libre a la literatura revisada por pares, lo que acelera la investigación, mejora la educación y fomenta un intercambio global de ideas.

Aunque el acceso abierto ha impactado positivamente en pequeñas partes de la literatura académica, su éxito destaca la necesidad de eliminar barreras, especialmente económicas, para ampliar sus beneficios (Rico 2019). Este modelo asegura que artículos revisados por pares y preimpresiones no publicadas estén disponibles gratuitamente en Internet, permitiendo a los usuarios leer, descargar, copiar, distribuir y reutilizar estos textos sin restricciones financieras, legales o técnicas.

Existen diferentes vías para implementar el acceso abierto (Recolecta 2024): la vía verde, dorada, bronce y diamante o platino. Estas opciones varían según el tipo de depósito y las condiciones para compartir la información.

Vía verde (*green*): Consiste en el depósito de trabajos previamente publicados en repositorios de acceso abierto, que pueden ser institucionales (como los universitarios), temáticos (como *Docusalut* o *Recolecta*) o generalistas (como *Zenodo*). Un mismo trabajo puede almacenarse en múltiples repositorios.

Vía dorada (*gold*): En este modelo, las revistas publican artículos en acceso abierto inmediatamente después de su aceptación, manteniéndolos accesibles de manera permanente bajo licencias donde los autores conservan los derechos. Frecuentemente, estas revistas imponen tarifas conocidas como APC (*article processing charges*) para cubrir sus costos.

Vía bronce: Se refiere a artículos disponibles gratuitamente en los sitios web de las editoriales, pero sin una licencia abierta explícita que permita su distribución o reutilización.

Vía Platino o Diamante: Este modelo ofrece acceso sin costos tanto para los investigadores como para los lectores. Las revistas que lo adoptan suelen ser financiadas por instituciones académicas, gubernamentales u organizaciones científicas, como es el caso de muchas revistas universitarias.

Estas vías representan opciones flexibles para fomentar el acceso abierto, garantizando que los avances científicos sean ampliamente accesibles y reutilizables, promoviendo así una ciencia inclusiva y colaborativa.

3.2. Importancia de las buenas prácticas en acceso abierto

Según la Comisión Europea ("What Are 'Good Practices'? | European Website on Integration" 2021) la definición de *buenas prácticas* "implies strategies, approaches and/or activities that have been shown through research and evaluation to be effective, efficient, sustainable and/or transferable, and to reliably lead to a desired result."

Por lo tanto, entendemos que las buenas prácticas incluirán:

- **Transparencia y accesibilidad:** Importancia de hacer la investigación accesible para todos.
- **Reproducibilidad y validación:** Cómo el acceso abierto facilita la validación y reproducción de resultados científicos.
- **Impacto y visibilidad:** Beneficios del acceso abierto en términos de visibilidad e impacto de la investigación.
- **Selección de revistas y plataformas:** Criterios para elegir revistas y repositorios adecuados para la publicación en acceso abierto.
- **Derechos de autor y licencias:** Importancia de las licencias abiertas (p. ej., Creative Commons) y cómo elegir la adecuada.
- **Cumplimiento de políticas:** Asegurarse de cumplir con las políticas de acceso abierto.
- **Gestión de datos de investigación:** Estrategias para compartir datos de investigación en formatos accesibles y reutilizables.
- **Archivos en repositorios:** Uso de repositorios institucionales y temáticos para archivar *preprints* y versiones finales.
- **Preservación a largo plazo:** Métodos para asegurar la preservación a largo plazo de publicaciones y datos.

3.3. La diseminación en el programa marco H2020

Para cada proyecto financiado por H2020, se firmaba un contrato de subvención entre la CE y los beneficiarios, en el que se establecían todas las condiciones, requerimientos legales y procedimientos para gestionar la complejidad de las situaciones que pueden surgir durante la vida de los proyectos, que podían tener una duración de varios meses o hasta varios años.

En concreto, las obligaciones que exigía el contrato tipo de subvención de H2020 (Comisión Europea 2019) para la diseminación de resultados pueden sinterizarse en los puntos siguientes:

1) Diseminar los resultados:

- Los beneficiarios tienen la obligación de diseminar al público los resultados de los proyectos por aquellos medios que sean adecuados incluyendo publicaciones científicas, si fuera posible.
- Si un beneficiario tiene algún motivo justificable puede optar por no publicar los resultados, hasta que no sea conveniente, aunque debe formalmente notificar a la CE.

2) Asegurar el acceso abierto

- Cada beneficiario debe asegurar acceso abierto a todas las publicaciones científicas de revisión por pares (sin coste, accesible online y para cualquier usuario). De este tipo de publicaciones, una copia se deposita en un repositorio (la versión aceptada (*green*) o publicada (*gold*)).
- Se debe asegurar acceso abierto a la publicación depositada por medio del editor o realizar un depósito con un máximo de 6 meses de embargo, y en el caso de proyectos de ciencias sociales y humanidades con un embargo máximo de 12 meses.

3) Asegurar el acceso a metadatos

- Acceso a los metadatos estandarizados que identifiquen la publicación: un identificador permanente, nombre de la acción, acrónimo, identificador de subvención, fecha de publicación, longitud y tiempo de embargo.

4) Acceso a los datos de la investigación

- El Open Research Data Pilot pedía depositar los datos en un repositorio que cumplieran con ser accesibles, consultables, explotables, reproducibles y diseminables sin cargo. Los datos incluían los metadatos necesarios para validar los resultados de las publicaciones científicas. Era obligatorio dar acceso a los datos incluidos en el Data Management Plan (DMP) o plan de gestión de datos que se pedía presentar en el proyecto, como así también informar sobre las herramientas e instrumentos para validar resultados y proveerlos si estuvieran disponibles.
- En los primeros años del programa H2020 el proyecto piloto Open Research Data Pilot era de cumplimiento opcional. Se trataba de un programa piloto sobre ciencia abierta para la apertura de datos. Dos años después pasó a ser obligatorio, salvo razones justificables que se debían notificar a la CE.

5) Para los proyectos o acciones de tipo sanitario, se establecen condiciones específicas para publicar los datos en acceso abierto, que no enumeramos en detalle pero que se pueden encontrar en la cláusula 29.3 del Acuerdo de Subvención tipo (Comisión Europea 2019). Acordar la diseminación con el consorcio:

- En proyectos ejecutados por un consorcio de beneficiarios, cada uno de ellos debe avisar sobre aquello que diseminará al resto de beneficiarios del proyecto con un margen de tiempo preacordado. El acuerdo de consorcio o CA, firmado entre los beneficiarios de un proyecto que confirman un consorcio, acuerda reglas entre las partes para evitar que los resultados obtenidos queden desprotegidos antes de ser patentados o para proteger secretos comerciales (“DESCA: Horizon 2020 Model Consortium Agreement” 2016).

6) Obligaciones y visibilidad de los fondos europeos

- Las publicaciones debían indicar que la financiación es europea, incluyendo el emblema de la Unión Europea (UE) y un texto con las referencias del proyecto, siempre que sea posible.
- Las publicaciones debían dejar constancia que la publicación refleja la opinión de los beneficiarios y no necesariamente coincide con la de la UE.

Las consecuencias de no cumplir con las obligaciones del contrato de subvención relacionadas con la diseminación podían derivar en la reducción de la financiación. En este artículo nos centraremos en el punto número 2 de estos requisitos.

4. Metodología

A partir del análisis de los requerimientos de H2020 para la diseminación de publicaciones en abierto, hemos diseñado la siguiente metodología para responder a las preguntas de investigación que nos hemos planteado.

4.1. Materiales y métodos

Para responder a las preguntas de investigación planteadas, se utilizó como fuente principal de datos la hoja de cálculo facilitada por la Oficina de Publicaciones de la

Unión Europea. Este archivo contenía información detallada sobre las publicaciones de investigación financiadas por el programa Horizonte 2020 (H2020).

Los datos analizados se localizaron en la hoja 'data' del archivo denominado 'projectPublications.xlsx' (Publications Office of the European Union 2022). En este archivo, se seleccionó específicamente la columna 'doi', que incluye los identificadores DOI, fundamentales para el cruce de información.

Para el análisis se empleó *OpenAIRE* (*OpenAIRE*, 2024), una infraestructura tecnológica y de servicios creada en 2009 con el objetivo de apoyar, acelerar y medir la correcta implementación de las políticas europeas de acceso abierto a publicaciones científicas y datos de investigación. Esta herramienta recopila documentos depositados en repositorios de acceso público.

Según *Recolecta*, "*OpenAIRE* promueve la comunicación académica abierta y mejora la accesibilidad, reutilización, reproducibilidad y monitoreo de resultados de investigación, incluyendo datos, en todas las disciplinas científicas y áreas temáticas, tanto en Europa como a nivel global."

Para garantizar una recuperación eficiente de la información, se organizó el conjunto de datos seleccionando exclusivamente aquellos registros que contaban con DOI. Este identificador persistente, asignado a artículos, libros y otras publicaciones, permite una búsqueda única y fiable en cualquier base de datos.

Teniendo en mente esto quisimos comparar la información que teníamos en el archivo 'projectPublications.xlsx', con esta comparación queríamos responder tres preguntas:

- P1. Determinar cuántos artículos revisados por pares se han informado y cuántos de ellos tienen datos de acceso.
- P2. Saber cuántos de los documentos revisados por pares figuran indizados en el repositorio Open Access de *OpenAIRE* y,
- P3. Si existía la información, cuántos artículos revisados por pares podían haber sido depositados en un repositorio institucional.

Para la pregunta P1, se buscó determinar los documentos que cumplieran con los requerimientos de publicación en acceso abierto que exigía H2020 y se analizó el total de artículos informados que debían publicarse en abierto.

De un total de 206.400 artículos revisados por pares que se incluyen, un 98 % cuenta con DOI informado e información detallada de la revista. De los 4.298 restantes, un total de 4.269 contiene la información de la revista donde fue publicado con el ISSN identificatorio, y en los 29 restantes no figura el ISSN. Por tanto, en estos casos, cuando no figura el DOI, el acceso a los documentos es engorroso y debería facilitarse.

Para las preguntas P2 y P3, el proceso fue el siguiente:

- Se realizó un análisis exploratorio de datos (EDA) en el archivo original para identificar datos faltantes y posibles requisitos de limpieza.
- Se eliminaron las filas sin DOIs del conjunto de datos. Esto redujo el número de entradas que podíamos cruzar en *OpenAIRE* de 332.495 a 263.866 registros.
- Los datos restantes se categorizaron de la siguiente manera:
 - Artículos revisados por pares: 206.400 que incluyen 4.298 que no tenían información sobre DOI, pero sí había otra información de la revista donde se

habían publicado, por lo que tuvimos que descartarlos al no poder buscarlos de forma inequívoca en *OpenAIRE*. De forma que quedaron 202.102 artículos.

- Actas de congresos: 40.492
- Otros: 8.123
- Capítulos de libros: 7.974
- Artículos no revisados por pares: 3.487
- Libros monográficos: 1.198
- Tesis doctorales: 484

4.2. Limpieza inicial de datos

4.2.1. Recopilación de datos de *OpenAIRE*

- Se desarrolló un método de generación de URLs personalizado para acceder a la funcionalidad de búsqueda de *OpenAIRE* y recuperar datos en formato CSV.
- La estructura de la URL se diseñó para incluir múltiples DOIs en una sola solicitud, optimizando el proceso de recuperación de datos.
- Los DOIs se procesaron en lotes de 50 para gestionar el volumen de datos y las posibles limitaciones del sistema de búsqueda de *OpenAIRE*.
- Se creó un script en Python para:
 - Generar el URL apropiado para cada lote de DOIs
 - Enviar solicitudes HTTP a estos URLs
 - Descargar los archivos CSV resultantes
- Se descargaron un total de 5.278 archivos CSV, cada uno conteniendo datos de hasta 50 DOIs.
- Para garantizar una asociación precisa entre los DOIs y los datos de *OpenAIRE*, se mantuvo un archivo CSV separado (`registro_descargas.csv`), registrando cada DOI y su archivo de descarga correspondiente.

4.2.2. Procesamiento y limpieza de datos

- Todos los archivos CSV descargados se combinaron en un único conjunto de datos completo utilizando *pandas* en *Python*.
- El conjunto de datos se refinó seleccionando las columnas relevantes: Título, Año de publicación, DOI, Tipo y Descarga desde.
- Se eliminaron las entradas duplicadas en el campo DOI, conservando solo la primera aparición de cada DOI único.

4.2.3. Análisis de datos

- Los datos limpios y procesados se guardaron como 'final_combined_results_subset.xlsx' en el directorio 'resultados'.
- Se realizó un análisis para determinar el porcentaje de artículos disponibles en *OpenAIRE* y en repositorios institucionales.

4.2.4. Identificación de los servidores

- Ya en *Excel*, se procesó el campo "Downloaded from" para extraer los URL base de cada archivo.

- Se eliminó la parte de la URL posterior al directorio principal, quedando con el formato "http(s)://directorio/".
- Este proceso permitió identificar los servidores o repositorios principales donde se alojaban los archivos, facilitando un análisis más preciso de las fuentes de los documentos.

Esta metodología permite la recopilación, procesamiento y análisis sistemáticos de datos de publicaciones de *OpenAIRE* utilizando técnicas de generación de URL personalizados y raspado web, con un enfoque en la identificación del uso de repositorios institucionales para los resultados de investigación financiados por H2020. La etapa final de identificación de servidores proporciona una visión más clara de las plataformas de alojamiento utilizadas para estos documentos de investigación.

5. Resultados

A partir de los datos originales, se ha podido determinar que un total de 206.400 artículos revisados por pares, un total de 202.102 artículos, es decir un 98 %, cuenta con DOI informado e información detallada de la revista donde se han publicado.

De todas maneras, de los 4.298 restantes, un total de 4.269 contiene la información de la revista donde fue publicado con el ISSN identificatorio, y en un total de 28 no se informa el ISSN. Por tanto, en estos casos, cuando no figura el DOI, el acceso a los documentos es engorroso y debería facilitarse.

A partir de allí, hemos podido analizar los documentos de los cuales se tenía el DOI, clave que nos permite la búsqueda de cualquier documento.

En la Tabla 2 podemos ver cuántos de los documentos del fichero inicial H2020 tenían DOI (263.866), después con esta información pudimos recuperar en *OpenAIRE* 240.757 documentos, de los que eran artículos revisados por pares 229.031. Además, hubo 68.629 que no estaban recogidos en esta base de datos. La respuesta a esto podría ser, por ejemplo, que:

- Se trataba de documentos publicados en papel, por lo que posiblemente no tenían asignado un DOI.
- Que se habían presentado en congresos en los que normalmente no se da un DOI a cada uno de los resúmenes, en todo caso al libro de comunicaciones, en global.

Por lo que asumimos que esta información no nos afectaba porque, como decimos en la introducción, el programa H2020 exige la publicación en acceso abierto o el depósito en repositorios de los artículos revisados por pares.

Tabla 2. Cruce de datos del fichero inicial H2020 con los datos en *OpenAIRE*

	Documentos	Diferencia	
Fichero inicial H2020	332.495		100,00 %
Con DOI	263.866	68.629	79,36 %
en <i>OpenAIRE</i>	240.757	23.109	72,41 %
Documentos revisados por pares	229.031	11.726	68,88 %
Artículos revisados por pares	202.102	26.929	60,78 %

En el archivo de datos en el que combinamos el fichero inicial con los datos extraídos de *OpenAIRE* fue incluido un campo denominado "con *download from*" (62.532), esta información ayudó a saber de dónde había recuperado *OpenAIRE* el documento a texto completo, si este existía. En muchos casos se trataba de repositorios, así que de esta forma podíamos saber cuántos se habían depositado además de haberlo publicado en acceso abierto.

En el fichero obtenido después de analizar cuántos documentos estaban indizados por *OpenAIRE*, encontramos la columna "download from" de estos documentos (62.532, un 30,94 %) se supone que están en acceso abierto al ser detectados por *OpenAIRE* y que informa que han sido descargados desde plataformas de revistas (*ScienceDirect*, *Springer*, *Wiley*, etc.), de plataformas de búsqueda (*Scielo*, *PubMed*, etc.), de webs de universidades (Universidad de Granada, Universidad de Córdoba, etc.), pero hay otros que no sabemos el motivo de que aparezca la URL de descarga porque no contienen textos completos que se puedan descargar directamente (*Scopus*, *Web of Science*, *DOAJ*, etc.). Además, dentro de la columna "download from", hemos detectado que 16.983 de estos documentos están depositados en un repositorio institucional.

A partir de lo anterior, ¿significa esto que ninguno de los otros documentos ha sido depositado en un repositorio? Creemos que no. Esto puede deberse a que se depositó en un repositorio no detectado por *OpenAIRE*, o que faltan datos en los metadatos de registro, o bien, que el depósito se realiza en archivos o repositorios no indexados con posibles demoras en la actualización de la información por parte de los repositorios.

Además, podrían existir errores en la coincidencia de los identificadores de los artículos que impedirían su correcta detección y registro en la propia plataforma.

Podemos ver que muchos de ellos se han depositado en revistas en acceso abierto. Según el archivo principal, vemos que los documentos se dividen en:

Tabla 3. Tipología documental del fichero principal.

Tipo	Documentos
<i>dataset</i>	189
publicación	166,649
otros	74

6. Conclusiones

Refiriéndonos al listado de buenas prácticas, podemos dividir las conclusiones y recomendaciones de la siguiente manera:

1. Transparencia y accesibilidad

Conclusión: De los 263.866 documentos del fichero inicial con DOI, se detectaron 240.757 con DOI coincidente en *OpenAIRE*. Sin embargo, 23.109 documentos no estaban disponibles en esta plataforma. Esto sugiere una limitación en la información analizada, posiblemente debido a la ausencia de DOI en artículos publicados en papel o en congresos, que a menudo no se indexan ni cuentan con identificadores únicos, pero la muestra no ha sido baja, ya que solo hay una diferencia de 68.629 que no tenían DOI.

Sin embargo, cabe destacar que la obligatoriedad de publicar en acceso abierto alcanza a los artículos revisados por pares, en donde las cifras de cumplimiento son más altas:

unos 206.400 (98 %) de los cuales solo 4.298 (2 %) no contienen DOI informado por lo cual no son detectados por *OpenAIRE*.

Recomendación: Mejorar la accesibilidad y trazabilidad de los documentos asegurando que todas las publicaciones tengan DOI y metadatos bien configurados, lo que aumentaría su visibilidad en bases de datos de acceso abierto.

2. Reproducibilidad y validación

Conclusión: Aunque una parte considerable de los documentos está depositada en repositorios, no todos están disponibles en texto completo, y hay casos en los que no se ofrece una URL de descarga directa. Esto dificulta la validación y reproducibilidad de los resultados.

Recomendación: Fomentar el depósito en repositorios que permitan la descarga de texto completo para facilitar la validación de resultados. Las políticas de acceso abierto deben garantizar plataformas accesibles para la verificación independiente.

3. Impacto y visibilidad

Conclusión: La clasificación de acceso abierto indica que 35,476 documentos están en "bronce", 40,183 en "dorado" y 47,455 en "híbrido". Sin embargo, más de 200,000 registros carecen de información sobre su estado de acceso, lo que limita el análisis.

Recomendación: Completar la información de los registros en el archivo de publicaciones financiadas para garantizar un análisis más preciso.

4. Revistas y plataformas seleccionadas

Conclusión: La selección de plataformas de publicación varía significativamente. Aunque muchos documentos están depositados en repositorios institucionales, otros solo están disponibles en revistas comerciales o bases de búsqueda, y algunos documentos híbridos no tienen un repositorio identificado.

Recomendación: Establecer criterios claros para la selección de revistas y repositorios que aseguren el libre acceso al público. Promover el depósito en repositorios abiertos para maximizar el alcance de los recursos financiados.

5. Derechos de autor y licencias

Conclusión: Más de 200.000 artículos revisados por pares carecen de información sobre licencias claras, lo que limita su uso y reutilización.

Recomendación: Adoptar y promover licencias abiertas como Creative Commons en todas las publicaciones financiadas por fondos públicos para facilitar la reutilización sin restricciones legales.

6. Cumplimiento de políticas de acceso abierto

Conclusión: Los documentos no encontrados en *OpenAIRE* o repositorios institucionales reflejan posibles carencias en el cumplimiento de las políticas de acceso abierto, problemas con metadatos o identificadores que pueden obstaculizar su visibilidad.

Recomendación: Realizar revisiones periódicas para garantizar que todas las publicaciones cumplan con las políticas de acceso abierto, actualizando metadatos y verificando identificadores.

7. Administración de datos y ficheros

Conclusión: Solo 189 datasets están registrados en repositorios, indicando una baja representación de datos en acceso abierto.

Recomendación: Fortalecer el depósito de datos en repositorios públicos, asegurando su accesibilidad.

8. Preservación a largo plazo

Conclusión: Falta información en algunos campos, lo que limita la posibilidad de preservar publicaciones.

Recomendación: Establecer políticas de conservación a largo plazo que garanticen la accesibilidad futura, promoviendo el depósito en repositorios establecidos.

6.1. Futuras perspectivas

H2020 ha marcado un hito en la historia de la diseminación de proyectos financiados por fondos públicos europeos, al iniciar la obligatoriedad de abrir tanto los resultados de los proyectos financiados por estos fondos, como también los datos, siempre que fuera viable con ciertas excepciones que se pueden aplicar y tener en cuenta.

Dada la documentación que hemos podido analizar, se evidencia un avance en la tendencia de publicación en abierto de los artículos revisados por pares, con un 98 % de ellos que informan un DOI de acceso. Por tanto, es evidente que esta dinámica está en marcha con un alto porcentaje de cumplimiento.

Referencias

"Budapest Open Access Initiative." 2024. <https://budapestopenaccessinitiative.org/read/>

Campos, Alejandra. 2023. "Comunicación efectiva de la Ciencia a audiencias amplias: ¿una obligación, un reto o una oportunidad?" TDX (Tesis Doctorals en Xarxa). <http://hdl.handle.net/10803/688845>

Chirlomez, Alina Danielescu. 2019. "European Commission to invest €11 billion in new solutions for societal challenges and drive innovation-led sustainable growth". European Cluster Collaboration Platform. 2019. <https://clustercollaboration.eu/news/european-commission-invest-eu11-billion-new-solutions-societal-challenges-and>

Comisión Europea. 2019. "AGA: Annotated Model Grant Agreement."

Comisión Europea. 2014. "HORIZON 2020 in Brief. Directorate-General for Research and Innovation". <https://doi.org/10.2777/3719>

Comisión Europea. 2020. "The Framework Programme for Research and Innovation."

Comisión Europea. 2023. "HE Social Media Guide: Using Social Media in EU Funded R&I Projects." Vol. 7.

"DESCA: Horizon 2020 Model Consortium Agreement." 2016

"DG EAC." 2023. https://knowledge4policy.ec.europa.eu/organisation/dg-eac-dg-education-culture_en

"EU Funding & Tenders Portal." 2024. <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/programmes/h2020>

European IPR HelpDesk. 2016. "Fact sheet plan for the exploitation and dissemination of results in Horizon 2020." *European IPR Helpdesk*, 1–11

Recolecta. "La doble vía al acceso abierto", 2024. <https://recolecta.fecyt.es/acceso-abierto/vias>

Mea, Marianna, Alice Newton, Maria C. Uyarra, Carolina Alonso, and Angel Borja. 2016. "From Science to Policy and Society: Enhancing the Effectiveness of Communication." *Frontiers in Marine Science* 3 (SEP): 215048. <https://doi.org/10.3389/FMARS.2016.00168/BIBTEX>

"OpenAIRE." 2024. Accessed July 18, 2024. https://catalogue.OpenAIRE.eu/service/OpenAIRE.discovery_portal/overview

Pastor-Ramon, Elena, Juan Antonio Casado, and Alejandra Campos Salvá. 2024. "H2020 Project Publications Compared to OpenAIRE." *Datasets*, November. <https://hdl.handle.net/20.500.13003/21209>

Publications Office of the European Union. 2022. "CORDIS: EU Research Projects under Horizon 2020 (2014–2020)." Publications Office of the European Union. January 21, 2022. <http://data.europa.eu/88u/dataset/cordisH2020projects>

"R&I Projects: Key Figures." 2024. https://dashboard.tech.ec.europa.eu/qs_digit_dashboard_mt/public/sense/app/d58f3864-d519-4f9f-855e-c34f9860acdd/sheet/KVdtQ/state/analysis

"Reglamento del Parlamento Europeo y del Consejo por el que se crea el Programa Marco de Investigación e Innovación «Horizonte Europa» y se establecen sus normas de participación y difusión por el que se crea el programa marco de investigación e innovación «Horizonte Europa» y se establecen sus normas de participación y difusión" 2018. Comisión Europea. 2018. <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52018PC0435>

Rico-Castro, Pilar. 2019. "¿Amigos o enemigos? Cómo la open science pone a las políticas de open access frente al espejo." *RUIDERAE: Revista de Unidades de Información*, 15.

"RRI Tools." 2024. <https://rri-tools.eu>

"The EU Guide to Science Communication". 2024. <https://youtube.com/playlist?list=PLvpwjZTs-LheOwu6uy8gr7JFfmv8EZuH>

"What Are 'Good Practices'? | European Website on Integration." 2021. https://migrant-integration.ec.europa.eu/page/what-are-good-practices_en