



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

*Una manera de hacer Europa*

Bilbao, diciembre de 2021

# EFx-x.x. ACCESO A FUENTES EXTERNAS DE PUBLICACIONES

## Contents

<b>1. Introducción</b>	3
<b>2. Clasificación de los metadatos de una publicación</b>	4
<b>3. Análisis de fuentes externas de publicaciones</b>	7
3.1. Web of Science (WoS-Clarivate)	7
3.2. ORCID	9
3.3. PubMed	10
3.4. Semantic Scholar	11
3.5. arXiv	12
3.6. Crossref	12
3.7. Google Scholar	13
3.8. Open Citations	14
3.9. European PMC	15
3.10. Conclusiones sobre las fuentes analizadas	1
<b>4. Fuentes principales y complementarias</b>	7
4.1. PudMed	7
4.2. Semantic Scholar	8
4.3. CrossRef	8

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

1

Fecha

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

*Una manera de hacer Europa*

4.4. Open Citations	8
4.5. European PMC	9
<b>5. Compatibilidad de fuentes</b>	10
5.1 Análisis de la compatibilidad entre fuentes externas	10
5.2 Análisis de la compatibilidad de la fuente principal y complementaria	12
5.3 Compatibilidad de metadatos	12
5.4 Gestión de publicaciones incompletas	13
<b>6. Propuesta de proceso de reclamación de publicaciones</b>	15
<b>7. Otros tipos de publicaciones</b>	21
<b>Bibliografía</b>	22

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

***Una manera de hacer Europa***

## 1. Introducción

El objetivo de este entregable es realizar un análisis inicial de otras fuentes externas de ROs de tipo publicación que puedan ser una alternativa a Scopus-Elsevier, fuente externa para la que ya se ha desarrollado un microservicio en Hércules y que posee una serie de restricciones de cuotas o número de llamadas a la misma que obligan a realizar este estudio. Para ello, se analizarán las diversas fuentes externas, sus cuotas, sus restricciones de uso, los metadatos devueltos por cada publicación y otras características determinantes en la elección de la próxima fuente externa.

La principal finalidad de este entregable es encontrar las fuentes externas que más se adecuen a nuestras necesidades, y por tanto nos permita, con el mínimo número de *requests (peticiones)* posible, obtener los metadatos necesarios para describir todas las publicaciones que tiene un investigador. De este modo también se tendrá en cuenta si los identificadores, tanto de autores como de artículos obtenidos, son propios de dicha red (como el *scopus\_id* de Scopus), inexistentes (por ejemplo, nombres) o identificadores generales (ORCID y DOI).

Una vez elegidas las fuentes externas se implementarán los servicios necesarios que nos permitan acceder a las mismas y reclamar los datos de un investigador, tal y como ya se ha hecho para la fuente externa Scopus-Elsevier.

El documento también describe los pasos necesarios para llevar a cabo el proceso de reclamación de publicaciones de un determinado autor. El algoritmo se ha diseñado en base a la información que se requiere saber de cada artículo de un autor y al análisis de cada fuente externa seleccionada.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

*Una manera de hacer Europa*

Unión Europea

## 2. Clasificación de los metadatos de una publicación

A continuación, se listan los metadatos a obtener de cada publicación. Aquellos que están coloreados no podrán ser modificados por el usuario:

- Tipo de publicación<sup>1</sup>
- Título
- Abstract
- Idioma
- DOI
- Fecha
- Link
- Palabras clave
- Autor correspondiente
  - Nombre (lista)
  - ORCID
  - Identifier (ID en caso de no tener ORCID)
  - link
- Conjunto de autores (por cada uno):
  - Nombre (lista)
  - ORCID
  - Identifier (ID en caso de no tener ORCID)
  - link
- Página inicio
- Página fin
- Revista
  - Nombre
  - ISSN
  - ¿métrica?
- Métrica de publicación

---

<sup>1</sup> Este metadato se asociará a los tipos de publicaciones identificados en la ROH. En los primeros desarrollos únicamente se utilizarán las entidades AcademicArticle, ConferencePaper y, para el resto de tipos, Publication.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

*Una manera de hacer Europa*

- Nombre de red
- Número de citas.

Por otro lado, todos los artículos citan y referencian a otros. Dada una publicación principal como la de arriba tenemos que poder enlazarla con los artículos citados y referenciados, que serán considerados **ROs de tipo referencia o bibliografía**. Estos artículos citados y referenciados pueden ser caracterizados con los mismos metadatos que un RO de tipo publicación, pero la importancia de estos es diferente, ya que no se necesita obtener tanta información de estos. En el siguiente listado se indican con colores los metadatos mínimos necesarios para caracterizar un RO de tipo referencia o bibliografía.

- Tipo de publicación
- Título
- Abstract
- ID
- Idioma
- DOI
- Fecha
- Link
- Autor correspondiente:
  - Nombre (lista)
  - ORCID
  - Identifier (ID en caso de no tener ORCID)
  - link
- Conjunto de autores (por cada uno):
  - Nombre (lista)
  - ORCID
  - Identifier (ID en caso de no tener ORCID)
  - link
- Página inicio
- Página fin
- Revista
  - Nombre
    - ISSN
  - ¿Métrica?
- Métrica de publicación

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

*Una manera de hacer Europa*

- Nombre de la fuente externa
- Número de citas

El listado de colores significa lo siguiente:

- **Rojo** – Metadatos necesarios e IMPRESCINDIBLES. Son identificadores estándar.
- **Naranja** - Metadatos importantes, aunque no identificadores genéricos.
- **Negro** – Metadatos de tipo enlace o con información complementaria.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

*Una manera de hacer Europa*

### 3. Análisis de fuentes externas de publicaciones

En esta sección se expondrán diversas fuentes externas que contienen ROs de tipo publicación. En las siguientes subsecciones se detallarán cada una de ellas. Al final de la sección se expondrá una tabla en la que de manera resumida se expondrán los metadatos, en comparación con los deseados, que se obtienen de cada publicación, las APIs que permiten obtener estos metadatos, las ventajas y contras de estas APIs y, por último, una valoración respecto a la utilidad de dicha fuente externa en este proyecto. De esta manera, se podrá decidir qué nuevos microservicios de consumo de fuentes externas se deberán implementar en Hércules.

#### 3.1. Web of Science (WoS-Clarivate)

Esta fuente externa pertenece a Clarivate y está accesible a través de la [licencia nacional que gestiona la FECYT para las universidades españolas](#). En su web indican que para poder usarla hay que ponerse en [contacto con FECYT](#) a través de un formulario en el que se indica qué tipo de servicio desea habilitar.

Una vez hecha la solicitud, el procedimiento técnico para acceder a esta fuente externa exige estar registrado, dar de alta una aplicación desde la que se accederá a las APIs y suscribirse a cada una de las APIs que se van a usar para poder recibir posteriormente por correo electrónico una API-key para poder invocar a cada API. En la sección [How to get started](#) de su web se explican los pasos a seguir.

El [portal de Clarivate](#) proporciona varias APIs, además de las específicas de WoS. En concreto para obtener información de Web of Science (WoS) ofrecen las siguientes opciones:

- [InCites Document Level Metrics API](#) – Da información sobre las citas de un artículo.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

***Una manera de hacer Europa***

- [Web of Science API Expanded](#) - Support 'search' and 'data integration' using Web of Science data returned as JSON or XML. This API supports rich searching across the Web of Science to retrieve item-level metadata, including times cited counts and contributor addresses/affiliations. Additional operations support the ability to discover related records as well as cited references and citing items.
- [Web of Science API Lite](#) - This API supports simple searching across the Web of Science to retrieve core item-level metadata.
- [Web of Science SUSHI API](#) -
- [Web of Science Reviewer Locator API](#) - Combining the power of the Web of Science's author and reviewer databases, Web of Science Reviewer Locator is a full stack solution to find, screen, and connect with expert peer reviewers.
- [Web of Science Journals API](#) - This API provides journal-level metadata and metrics for all journals in the Journal Citation Reports™ covered in the Web of Science Core Collection, including the Journal Impact Factor™ and other new metrics. Integrate journal data into your internal systems or retrieve journal indicators for bibliometrics studies.

La [licencia nacional que gestiona la FECYT para las universidades españolas](#) permite el acceso a la “Web of Science API Expanded”, por lo que para realizar la primera pregunta y obtener información sobre los artículos de un autor se usará la consulta por identificativo de autor  $AI=(\{ORCID\_autor\})$  que ofrece esta API. La funcionalidad completa de esta API se encuentra descrita en este [enlace](#). El número de ítems devueltos en cada petición debe estar entre 1 y 100, por lo que para aquellos usuarios que tengan más de 100 publicaciones registradas deberá realizarse la misma petición tantas veces como sea necesario. Sin embargo, no se ha encontrado información respecto a cuotas o límite de *requests* con una determinada suscripción, aunque se seguirá indagando por si esta información está disponible en alguna sección de su web que aún no hemos encontrado.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

1

Fecha

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

### ***Una manera de hacer Europa***

En la tabla se encuentran diversos puntos a favor y en contra de esta fuente externa. En la tercera columna encontramos las APIS que se estiman necesarias para la obtención de los metadatos que deseamos. Cada una de estas APIs tiene un color con el que se ha coloreado en la última columna los metadatos que se obtienen con ella. Esta información está sujeta a cambios, ya que la información obtenida por las APIs no ha podido ser contrastada con una llamada.

### 3.2. ORCID

Esta [fuente externa](#) puede consultarse en el siguiente enlace. Para [acceder a la API](#) que nos proporciona esta fuente externa se debe solicitar un token, vale con que sea de tipo *read-publicly* no es necesario ser miembro para solicitarla. Es de tiempo prolongado, es decir, no caduca hasta dentro de 20 años y menciona que se pueden hacer varias llamadas con este mismo *token* pero no comentan si el número de llamadas es ilimitado o no.

Solo hay una API que permite, dado un autor, reclamar la información que se encuentra en los diferentes apartados (o todos) de su perfil. La información de cómo reclamar estos datos está en el segundo enlace de esta sección. No hay ningún ejemplo de qué metadatos se obtienen de cada artículo, pero dado que en la página web no hay demasiados metadatos por artículo, es posible que no obtengamos la información básica necesaria que deseamos para cada artículo. Este es el problema principal de esta fuente externa. Además, no está claro si esta es la fuente externa en la que los usuarios suelen tener la información actualizada de sus publicaciones. También se ha de comentar que los identificadores que posee y proporciona esta fuente externa son siempre genéricos, facilitando la combinación de diversas fuentes externas para obtener todos los metadatos que se requieren.

El resumen de la información obtenida de esta fuente externa sigue el mismo esquema descrito para la fuente externa WoS-Clarivate en la tabla de final de sección.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

**Una manera de hacer Europa**

### 3.3. PubMed

Las APIs de esta fuente externa son gratuitas y no se necesita una API-Key para acceder a las misma:

1. NCBI E-utilities --- <https://www.ncbi.nlm.nih.gov/books/NBK25501/> Po lo que tengo entendido con esta API puedes preguntar por todos los recursos de PubMed y de algún otro repositorio.
2. File validation tools, --- <https://www.ncbi.nlm.nih.gov/pmc/pub/validation/>
3. OA web service, --- <https://www.ncbi.nlm.nih.gov/pmc/tools/oa-service/>
4. OAI-PMH, --- <https://www.ncbi.nlm.nih.gov/pmc/tools/oai/> - Especifica de PubMed.
5. ID converter, --- <https://www.ncbi.nlm.nih.gov/pmc/tools/id-converter-api/>
6. FTP service, --- <https://www.ncbi.nlm.nih.gov/pmc/tools/ftp/>
7. Literature Citation Exporter. --- <https://api.ncbi.nlm.nih.gov/lit/ctxp>

Esta fuente externa es de dominio exclusivo biomédico, a diferencia de las otras fuentes externas examinadas que abarcaban varios dominios de la ciencia. El problema principal de estas APIs es el número de *requests* permitidos. En concreto de la fuente que permite obtener datos de publicaciones en PubMed se comenta lo siguiente:

*"If you are using a script that makes more than 100 requests of any kind, please run it outside of the PMC system's peak hours. Also, please make sure that your system does not make concurrent requests, even at off-peak times. Peak hours are Monday to Friday, 5:00 AM to 9:00 PM, U.S. Eastern time."*

En otras APIs que también permiten obtener información de PubMed (como es el caso de la primera), el número de *requests* permitido por segundo es 3, en comparación con otras fuentes externas que nos permiten desde 100-300 *requests* por segundo.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

1

Fecha

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

**Una manera de hacer Europa**

En la tabla de final de sección se encuentra un resumen de los pros y contras de esta fuente externa.

### 3.4. Semantic Scholar

La información sobre esta fuente externa puede consultarse en este [enlace](#). Esta fuente externa proporciona diversas APIs, aunque en muchas ocasiones se puede conseguir la misma información a través de una sola API aumentando los campos del parámetro *fields* en una misma petición o *request*. La información sobre las diversas APIs y los metadatos o campos de información que se pueden obtener con cada una de ellas se puede consultar en el siguiente [enlace](#). Para información más concreta sobre la APIs puede consultarse también el siguiente [enlace](#).

El mayor problema de estas APIs es que los IDs que identifican los artículos y autores devueltos son propios de *Semantics Scholar* y no aquellos que se pueden considerar como generales o estándar. Tiene una cuota de *requests*, pero dejan a disposición de cada organización o institución un formulario para rellenar en caso de que se quieran ampliar la cantidad de peticiones o *requests* por segundo que se van a realizar. Por lo demás, es gratuita y no necesita una API-key. Tampoco es posible determinar el ID de la revista en la que ha sido publicado un artículo, ya que únicamente devuelve una cadena de caracteres o *string* con el nombre de la revista. También es imposible realizar la búsqueda de los artículos de un cierto autor sin el id propio que se le otorga en esta fuente externa. Esta búsqueda tampoco puede hacerse con el ORCID de un investigador.

Al igual que con las anteriores fuentes externas, en la tabla adjunta a esta sección puede encontrarse información de las diversas APIs a utilizar, los metadatos obtenidos con ellas y los pros y contras de esta fuente externa.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

***Una manera de hacer Europa***

### 3.5. arXiv

Esta fuente externa puede encontrarse en el siguiente [enlace](#). Ofrece APIs de acceso público, pero hay un aviso sobre el uso comercial de las mismas que es conveniente revisar, ya que habría que informar a arXiv del uso que se pretende hacer y ello puede tener implicaciones en el modelo de negocio o de explotación de las herramientas desarrolladas en Hércules:

*“Commercial projects that utilize arXiv’s public APIs or other bulk data pipelines are requested to contact arXiv at nextgen@arxiv.org, review arXiv’s brand guidelines, sign a memorandum of understanding, and consider becoming an arXiv affiliate before embarking on the project. This includes any project created as a product for sale.”*

Esta fuente externa es diferente a las examinadas anteriormente, ya que permite declarar como publicación recursos que no han sido publicados en una revista científica. Además, no tiene perfil de los usuarios, por lo que los identifica únicamente con el nombre pudiendo dar lugar a problemas de identificación.

Esta fuente externa no está resumida en la tabla de final de sección debido a que se deja a los revisores de este informe la decisión de seguir considerándola como fuente externa candidata a incluir en Hércules, dado el tipo de publicaciones que se pueden encontrar en ella.

Si se decide abordar el consumo de esta fuente externa, podría considerarse la posibilidad de utilizar algún servicio desarrollado en ASIO que permitiera realizar la identificación de un usuario a partir de múltiples nombres que pueda tener en diferentes fuentes externas.

### 3.6. Crossref

Esta fuente externa puede consultarse en este [enlace](#). La documentación de las APIs proporcionadas y sus límites de cuotas, ya que tiene un número máximo de *requests* por

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

*Una manera de hacer Europa*

segundo, puede encontrarse en el siguiente [enlace](#). En las primeras pruebas realizadas para obtener todos los artículos de un autor a partir de su ORCID, las respuestas proporcionaban más información de la deseada, incluyendo artículos no publicados por dicho autor. Por tanto, parece recomendable usarla simplemente para complementar la información de un cierto artículo, es decir, como fuente externa secundaria o complementaria.

En cuanto a los metadatos obtenidos de un artículo, los relativos a los autores no incluyen información que los defina de forma unívoca como investigador, como por ejemplo el ORCID u otro ID propio de la red, sino que únicamente están definidos por su nombre, pudiendo producirse errores en su identificación. Además, los metadatos relativos a las referencias o bibliografía del artículo no son suficientes ni aclaratorios.

Cabe destacar que los servidores de esta fuente externa se cayeron cuando se realizaron diferentes pruebas de acceso, por lo que es posible que los resultados obtenidos y las conclusiones a las que hemos llegado estén sesgados por esta incidencia. También se encontraron algunos valores extraños en algunos metadatos, como por ejemplo un 0 en el número de artículos referenciados en la bibliografía, lo que refuerza el comentario anterior.

**[ACTUALIZACIÓN]** Al realizar nuevas pruebas de búsqueda, con el ORCID de un usuario únicamente se han obtenido publicaciones de dicho usuario, aunque su número era mucho menor que el de las obtenidas en fuentes externas como Scopus-Elsevier o WoS-Clarivate. El análisis de los metadatos obtenidos para las publicaciones está aún en curso.

### 3.7. Google Scholar

No ofrece una API para acceder a su base de datos porque Google solamente desea ofrecer esta información a través de una interfaz web para humanos. Sin embargo, hay un desarrollo **no oficial de Google Scholar** llamado [SerpApi](#) (Google Search API) que tiene una [API\(google-scholar-author-articles\)](#) para obtener las publicaciones de un autor.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

***Una manera de hacer Europa***

El problema es que está basado en *web-scraping* por lo que la información de cada artículo se limita a lo que ofrece la interfaz web de Google Scholar y sujeta a las posibles modificaciones que haga Google en dicha interfaz. Aunque Google asigna a cada investigador un ID interno, los datos de los autores que se obtienen tras llamar a la API son sus nombres, no un ID genérico ni el identificador único propio de Google Scholar. Esta API solo devuelve 100 resultados por búsqueda, por lo que habría que realizar varias peticiones para aquellos autores con más de 100 artículos.

Para poder realizar una consulta que permita obtener las publicaciones de un autor es necesario saber el *author\_id* del investigador en Google Scholar, que se puede obtener invocando a la [API \(google-scholar-profiles-api\)](#). Por lo tanto, surge la necesidad de filtrar los resultados con el nombre del investigador y posteriormente identificar al autor correcto y obtener el *author\_id* que le corresponde.

La [API \(google-scholar-author-citation\)](#) para saber más sobre el número de citas y la revista en la que está publicado un artículo necesita el *author\_id* de Google Scholar que devuelve la llamada inicial a la [API \(google-scholar-author-articles\)](#).

Este desarrollo no oficial exige obtener una API-key para consumir sus APIs, pero no está descrito el proceso para obtenerla, si bien hay un enlace para [registrarse en la web de SerpApi](#).

### 3.8. Open Citations

Dado el DOI de un artículo, esta fuente externa permite obtener los DOIs de los artículos que referencia en su bibliografía o que los que lo citan a él. No aporta mucha más información, es decir, no hay muchos más metadatos. Las APIs que permiten obtener esta información se encuentran en los siguientes enlaces:

- <https://opencitations.net/index/coci/api/v1#/references/{doi}>
- <https://opencitations.net/index/coci/api/v1#/citations/{doi}>

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

*Una manera de hacer Europa*

### 3.9. European PMC

Esta fuente externa puede encontrarse descrita en el siguiente [enlace](#) y permite obtener información sobre los artículos que citan y referencian a uno dado. El conjunto de APIs que se pueden consumir puede encontrarse en el siguiente [enlace](#), de las cuales son interesantes para Hércules las siguientes:

- [Articles RESTful API \(references\)](#)
- [Articles RESTful API \(citations\)](#)

Para poder realizar consultas es necesario obtener primero el ID interno del artículo en esta fuente realizando una petición extra a la siguiente URI utilizando el DOI del artículo, por ejemplo:

<https://www.ebi.ac.uk/europepmc/webservices/rest/search?query=DOI:10.1007/bf00197367>

Los artículos citados o referenciados aparecen en la respuesta con pocos metadatos, pero entre ellos se encuentran el DOI, razón por la cual esta fuente externa complementaria nos puede permitir enlazar un RO de tipo publicación con sus ROs de tipo bibliográfica y cita asociados.

En caso de que en la respuesta no haya suficiente información sobre un artículo citado, se deberá ejecutar una nueva petición extra a la URL anterior para completar los metadatos necesarios.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

1

Fecha

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

**Una manera de hacer Europa**

Fuente Externa	Pros y Contras	APIs interesantes/necesarias	Metadatos
Web of Science	[Contra] Hay que declarar específicamente qué APIs se van a usar porque cada una tiene una API-key diferente.	API para saber las publicaciones de un autor: Web of Science API Expanded	DOI Título Abstract Tipo Fecha Autores/autor principal -> Names Nombre Id Knowledge Area Etiquetas Número de artículos en la bibliografía Número de citas (refs) Palabras clave Links Página de inicio y fin Revista ID ID, (a confirmar) nombre Metadatos de revista (¿) Bibliografía -> Metadatos de cada artículo
	[PRO] Los metadatos obtenidos son muy extensos		
	[Contra] El número de ítems devuelto en la API es limitado.	API para obtener información de las métricas de una revista: Web of Science Journals API	
	[Contra] Son de pago. Hay que registrarse, pedir permiso y realizar una suscripción.		
	[PRO] WoS tiene muchas publicaciones y es una red muy enriquecida y muy usada.		
Web of Science: Como organización hay que mirar si es posible/aceptable, dadas las condiciones del proyecto, solicitar estas APIs. Los metadatos obtenidos son muy completos, obteniendo IDs generales, así como una gran gama de metadatos.			

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

**Una manera de hacer Europa**

<b>ORCID</b>	[Contra] Pocos metadatos de las publicaciones.	Solo tienen <a href="#">una API</a> que es la que se usaría para saber los artículos de un autor (la primera a realizar).	Doi
	[Contra] No tiene diferentes APIs para poder completar la información obtenida.		Título
	[PRO] Conseguir el Token de acceso es fácil y además este dura un largo periodo de tiempo.		Abstract
	[--] Hay que considerar que no todos los usuarios declaran en esta fuente sus artículos.		Tipo
			Fecha
			Autores /autor principal -> Nombre, <b>Id</b>
			Knowledge Area
			Etiquetas
			Número de artículos en la bibliografía
			Número de citas
			Links
			Página de inicio. Página fin

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

**Una manera de hacer Europa**

<p>ORCID: Devuelve pocos metadatos y los usuarios han de tenerlo al día. En general la información devuelta es muy escasa y por tanto se necesitaría extraer información adicional de otras fuentes para completar la información. Sin embargo, la información devuelta está compuesta de identificadores genéricos, lo cual permite completar la información con otras fuentes.</p>	<p>Revista -&gt;  <a href="#">ID</a>,          nombre,          Metadatos de revista (¿)          Bibliografía -&gt;          Metadata de cada artículo</p>
--	---

<p><b>PubMed</b></p>	<p>[Contra] Es de un dominio de la ciencia en específico.</p>	<p>Dadas las restricciones y los contras tan grandes se ha decidido no verificar qué metadatos se pueden extraer de esta fuente externa.</p>	<p>Doi          Título          Abstract          Tipo          Fecha          Autores /autor principal -&gt;          Nombre          Id          Knowledge Area</p>
	<p>[Contra] Formato en el que devuelve la información es XML y no JSON.</p>		

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

**Una manera de hacer Europa**

	<p>[Contra] Cuota excesivamente restrictiva que no permite el volumen de consultas necesario en Hércules y que bloquea al usuario que sobrepasa los límites.</p>		<p>Etiquetas          Número de artículos en la bibliografía          Número de citas          Links          Página de inicio Página fin          Revista -&gt;              ID,              nombre              Metadatos de revista (i)          Bibliografía              Metadata de cada artículo</p>
<p>PubMed: No parece una fuente externa viable, dadas sus restricciones, aunque se podrían completar los datos de una publicación obtenida de la fuente Scopus-Elsevier, ya que esta devuelve el id de PubMed asociado a una publicación.</p>			
<p>Semantics Scholar  <a href="#">(link)</a></p>	<p>[--] Tiene cuota de cantidad requests por segundo, pero <b>se puede aumentar rellenando un formulario</b> (no se ha podido comprobar aún si es de pago).</p>	<p>La primera API es la que daría información sobre los papers de un autor dado su id.</p>	<p>Doi          Título          Abstract          Tipo          Fecha year</p>

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

**Una manera de hacer Europa**

	<p>[PRO] No se necesita ninguna API-Key ni token para reclamar la información.</p>	<p><a href="https://api.semanticscholar.org/graph/v1/author/{author_id}/papers">https://api.semanticscholar.org/graph/v1/author/{author_id}/papers</a></p>	<p>Autores /autor principal  <a href="#">Nombre</a>  <a href="#">Id -En su red</a>  <a href="#">Homepage</a>  <a href="#">Aliases</a>  <a href="#">externalIds</a></p> <p>Knowledge Area  Etiquetas  <a href="#">Número de artículos en la bibliografía</a>  <a href="#">Número de citas</a>  <a href="#">Links</a></p> <p>Página de inicio Página fin  Revista ID @venue  ID,  nombre  Metadatos de revista (;)</p> <p>Bibliografía  <a href="#">ID Semantic Scholar</a>  <a href="#">url</a>  <a href="#">title</a>  <a href="#">year</a>  <a href="#">authors</a>  <a href="#">número de citas</a></p>
<p>[Contra] No tiene el id identificativo de una revista. Por lo que esta información ha de ser completada con otras fuentes externas.</p>	<p>También tiene otras APIs para saber más información sobre un autor.</p>		
<p>[Contra] Los IDs devueltos son de su propia red. Al menos devuelve un identificador.</p>	<p><a href="https://api.semanticscholar.org/graph/v1/author/{author_id}?fields=...">https://api.semanticscholar.org/graph/v1/author/{author_id}?fields=...</a></p>		
<p>[PRO] Los metadatos de la bibliográfica son igual de extensos que los de la publicación principal.</p>	<p>Esta API es necesaria para obtener el id del autor del que queremos la información (la llamada de arriba).</p>		
<p>Opinión: el sistema de llamadas es muy fácil y cómodo. La información que devuelve no es muy extensa pero sí mínima y suficiente. La mejor opción es usarla para obtener los metadatos de una publicación de tipo referencia.</p>			

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

**Una manera de hacer Europa**

<p><b>Crossref</b> (se les ha caído el servidor)</p>	<p>[Contra] Tiene un límite de <i>requests</i> por segundo. <b>Creo que se puede aumentar, pero hay que declararlo.</b></p>	<p>La Api que obtiene las publicaciones de un autor. Esta petición proporciona demasiada información incluyendo artículos ajenos a dicho autor</p>	<p>Doi Título Abstract Tipo Fecha Autores /autor principal (orden)-&gt; Nombre Id Knowledge Area Etiquetas (un poco raras..) Número de artículos en la bibliografía -&gt; dudoso porque pone 0. Número de citas (is-referenced-by-count) Links Página de inicio Página fin Revista ID ID, Nombre Metadatos de revista (¿) Bibliografía Metadata de cada artículo</p>
	<p>[Contra] No sirve para obtener los artículos de un autor en concreto.</p>		
	<p>[contra] Hay casos en los que sí se devuelve el ORCID, pero solo en algunas personas.</p>	<p>API para la obtención de metadatos asociados a un autor <a href="#">/Works/{DOI}</a></p>	
	<p>[contra] Tampoco devuelve demasiada información de la bibliografía.</p> <p>[PRO] En algunos artículos da información sobre el doi de los artículos bibliográficos, incluyendo el doi.</p>		

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

**Una manera de hacer Europa**

Crossref. Devuelve una cantidad interesantes de metadatos, aunque no se le puede preguntar directamente por los artículos de un autor proporcionando el ORCID de este. Por tanto, es una fuente que se deberá usar para complementar la información de un artículo dado.

Google Scholar	[Contra] La llamada de <i>autor artides</i> da muy poca información sobre los artículos que ha publicado un investigador. Parámetro inicial: author_id	Si el autor no sabe su author_id hay que usar <i>Google Scholar Profiles API</i> , para lo cual se busca con el nombre y se debe filtrar el resultado hasta dar con el autor deseado.	Doi Título Abstract Tipo Fecha year Autores /autor principal (orden)-> Nombre Id Knowledge Area Etiquetas (un poco raras..) Número de artículos en la bibliografía Número de citas Links Página de inicio Página fin Revista ID
	[Contra] No se puede obtener información sobre la revista, su DOI, bibliografía o autores. Y lo obtenido no es genérico por lo que no permite búsquedas en otras fuentes.	<a href="#">Google Scholar Author Articles API</a> -- Información sobre los artículos que ha publicado un autor en concreto.	
	[Contra] Si el usuario no sabe su author_id entonces es difícil que con la API "profiles" se obtenga este valor.	La segunda llamada con el id de la primera obtendremos ( <a href="https://serpapi.com/google-scholar-author-citation">https://serpapi.com/google-scholar-author-citation</a> )	

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

***Una manera de hacer Europa***

Google Scholar: La obtención de metadatos es bastante limitada y hay que hacer uso de muchas APIs para tener la información deseada. Además, las publicaciones no son devueltas con IDs genéricos por lo que dificulta completar estos metadatos con otras fuentes.

ID,  
Nombre  
Metadatos de revista (¿)  
Bibliografía  
Metadatos de cada artículo

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

*Una manera de hacer Europa*

### 3.10. Conclusiones sobre las fuentes analizadas

Por tanto, en este entregable se han analizado diversas fuentes externas con el objetivo de posteriormente realizar un microservicio en una fuente o no. Se han considerado las restricciones de las fuentes externas examinadas, tales como el número de *requests* permitidos por segundo, la facilidad de uso o los procesos necesarios para obtener las credenciales necesarias. También se ha adjuntado información de las APIs que estas fuentes externas ponen a disposición de los investigadores con el fin de que estos puedan reclamar su trabajo. Se han analizado dichas APIs con el propósito de definir cuáles de ellas son necesarias para obtener los metadatos deseados, así como se ha obtenido que combinación de APIs necesaria para definir los ROs de tipo publicación en la red de Hercules.

Este análisis se ha concluido con una tabla en la que cada fuente externa examinada ha sido valorada. En ella se exponen los puntos a favor y en contra de cada fuente, las APIs necesarias y los metadatos, en comparación con los deseados, obtenidos con cada una de ellas.

En este análisis se han obtenido los problemas más relevantes en relación con cada fuente externa:

- **WoS:** suscripción y pago por las APIs que no estén incluidas en la licencia nacional gestionada por la FECYT (WoS Expanded API, WoS Lite API y LAMR Service). Necesidad de completar metadatos con fuentes secundarias.
- **ORCID:** metadatos insuficientes y respuestas en algunos casos con identificativos de los autores no genéricos. Necesidad de completar metadatos con fuentes secundarias.
- **Semantic Scholar y PubMed:** dificultades para buscar las publicaciones por los identificativos genéricos de un autor. Estas fuentes externas serían interesantes para completar la información de una publicación.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

***Una manera de hacer Europa***

- **Semantic Scholar:** usa identificativos internos propios no genéricos para autores y publicaciones. No permite obtener el DOI de las publicaciones.
- **Crossref:** Pendiente de revisión, ya que debido a que la caída de sus servidores no ha proporcionado información fiable. Algunos usuarios son identificados con el ORCID y otros no. Del mismo modo, en ocasiones devuelve datos de la bibliografía y en otros casos no.

Por tanto, la fuente externa más recomendable como principal (que se uniría a la de Scopus-Elsevier) en cuestión de cantidad de metadatos proporcionados e identificación de publicaciones y autores de manera genérica es WoS-Clarivate. ORCID por su parte también es una buena propuesta como fuente principal ya que a pesar de devolver un número mínimo de metadatos estos son genéricos y son los más importantes para definir una publicación de forma unívoca. Las fuentes externas PubMed o Semantic Scholar podrían usarse para completar la información obtenida de las fuentes principales. Estas dos fuentes externas tienen inconvenientes para ser consideradas como principales. En concreto PubMed debido a la limitación en el número de peticiones y Semantic Scholar debido a la falta del metadato DOI identificativo de las publicaciones.

En los primeros desarrollos, los ROs de tipo publicación únicamente se asocian a las entidades AcademicArticle, ConferencePaper y Publication (para el resto de tipos) de la ROH. Sin embargo, está previsto realizar un mapeo con los diferentes valores asociados al metadato “Tipo de publicación” que se pueden recuperar desde las diferentes fuentes externas principales:

- Scopus-Elsevier<sup>2</sup>:
  - **Article**-ar
  - **Abstract Report**-ab
  - **Book**-bk
  - **Book Chapter**-ch

<sup>2</sup> [https://dev.elsevier.com/sc\\_search\\_tips.html](https://dev.elsevier.com/sc_search_tips.html)

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

*Una manera de hacer Europa*

- **Business Article**-bz
- **Conference Paper**-cp
- **Conference Review**-cr
- **Data Paper**-dp
- **Editorial**-ed
- **Erratum**-er
- **Letter**-le
- **Multimedia**-mm
- **Note**-no
- **Press Release**-pr
- **Report**-rp
- **Retracted**-tb
- **Review**-re
- **Short Survey**-sh
- **WoS-Clarivate**<sup>3</sup>:
  - **Article**: Reports of research on original works. Includes research papers, features, brief communications, case reports, technical notes, chronology, and full papers that were published in a journal and/or presented at a symposium or conference.
  - **Abstract** of Published Item: Bibliographic-only data on a published paper. Generally finds records dating back to 1974 or before.
  - **Art Exhibit Review**: Reviews of gallery or museum showings of artworks.
  - **Bibliography**: A list, often with descriptive or critical notes, of writings relating to a particular subject.
  - **Biographical-Item**: Obituaries, articles focusing on the life of an individual, and articles that are tributes to or commemorations of an individual.
  - **Book**: A monograph or publication written on a specific topic.

<sup>3</sup> [https://images.webofknowledge.com/images/help/WOS/hs\\_document\\_type.html](https://images.webofknowledge.com/images/help/WOS/hs_document_type.html)

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

### *Una manera de hacer Europa*

- **Book Chapter:** A monograph or publication written on a specific topic within a main division in a book.
- **Book Review:** A critical appraisal of a book (often reflecting a reviewer's personal opinion or recommendation) that evaluates such aspects as organization and writing style, possible market appeal, and cultural, political, or literary significance.
- **Chronology:** A review of events on a specific topic or subject in their order of occurrence in time.
- **Correction:** Correction of errors found in articles that were previously published and which have been made known after that article was published. Includes additions, errata, and retractions.
- **Correction, Addition:** Correction of errors found in articles that were previously published and which have been made known after that article was published. Includes additions, errata, and retractions.
- **Dance Performance Review:** Reviews of solo dance recitals, complete dance productions, dance programs consisting of several works, and other types of performed dances.
- **Data Paper:** A scholarly publication describing a particular dataset or collection of datasets and usually published in the form of a peer-reviewed article in a scholarly journal. The main purpose of a data paper is to provide facts about the data (metadata, such as data collection, access, features etc) rather than analysis and research in support of the data, as found in a conventional research article.
- **Database Review:** A critical appraisal of a database, often reflecting a reviewer's personal opinion or recommendation. Refers to a structured collection of records or data that is stored in a computer system.
- **Discussion:** An article or paper that discusses questions in an open and usually informal debate. Generally finds records dating back to 1996 or before.
- **Early Access:** An article that has been electronically published by a journal before it has been assigned to a specific volume and issue.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

### *Una manera de hacer Europa*

- **Editorial Material:** An article that gives the opinions of a person, group, or organization. Includes editorials, interviews, commentary, and discussions between individual, post-paper discussions, round table symposia, and clinical conferences.
- **Excerpt:** A selection from or a fragment of a literary or musical work, which cannot stand as a separate work in its own right.
- **Fiction, Creative Prose:** Includes short stories and other works of creative prose.
- **Film Review:** A review of a motion picture.
- **Hardware Review:** A critical appraisal of computer hardware, often reflecting a reviewer's personal opinion or recommendation. Refers to objects that you can actually touch, like disk drives, keyboards, printers.
- **Item About An Individual:** A review of the work(s) of a celebrated person in a particular field of study.
- **Letter:** Contributions or correspondence from the readers to the journal editor concerning previously published material.
- **Meeting Abstract:** A general summation of completed papers that were or will be presented at a symposium or conference.
- **Meeting Summary:** A paper that covers multiple meeting abstracts in a variety of subjects.
- **Music Performance Review:** Review of a live musical performance (recital, concert, and opera).
- **Music Score:** Transcript of the original and entire draft of a musical composition or an arrangement with the parts for the different instruments or voices written on staves one above another.
- **Music Score Review:** Review of a bound musical composition or bound collection of musical compositions.
- **News Item:** News, current events, and recent developments.
- **Note:** A paper that mentions or remarks on a published paper on a specific subject. Generally finds records dating back to 1996 or before.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

*Una manera de hacer Europa*

- **Poetry:** Compositions in verse; metrical writing.
- **Proceedings Paper:** Published literature of conferences, symposia, seminars, colloquia, workshops, and conventions in a wide range of disciplines. Generally published in a book of conference proceedings.  
Records covered in the two Conference Proceedings indexes (CPCI-S and CPCI-SSH) are identified as Proceedings Paper. However, the same records covered in the three indexes (SCI-E, SSCI, and A&HCI) are identified as Article when published in a journal.
- **Record Review:** Reviews of recorded music or speech.
- **Reprint:** An article that was previously published.
- **Retracted Publication:** An article that has been withdrawn by an author, institution, editor, or a publisher because of errors or unsubstantiated data.
- **Retraction:** A public notice that an article should be withdrawn because of errors or unsubstantiated data.
- **Review:** A renewed study of material previously studied. Includes review articles and surveys of previously published literature. Usually will not present any new information on a subject.
- **Script:** includes film scripts, plays, TV, and radio scripts.
- **Software Review:** A critical appraisal of computer software, often reflecting a reviewer's personal opinion or recommendation. Refers to programs, procedures, and rules, along with associated documentation pertaining to the operation of a computer system.
- **Theater Review:** Review of a performed play.
- **TV Review, Radio Review:** Reviews of television and radio broadcasts.
- **TV Review, Radio Review, Video Review:** Reviews of television, radio broadcasts, and video.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

*Una manera de hacer Europa*

## 4. Fuentes principales y complementarias

En este apartado se va a justificar por qué ciertas fuentes externas no pueden ser usadas como principales a la hora de realizar consultas sobre publicaciones de un autor, sino únicamente como fuentes complementarias para obtener información que permita enriquecer los metadatos de cada una de las publicaciones obtenidas a través de una consulta previa a una fuente principal.

Las fuentes externas que se consideran como principales serían **Scopus, WoSy ORCID**, ya que permiten realizar consultas de publicaciones de un autor a través de un identificativo estándar y genérico como ORCID ("Scopus Author ID" y "WoS ResearcherID" pueden estar asociados a ORCID) y, además, identifican las publicaciones recuperadas con su DOI, un identificador estándar y genérico con el que poder realizar nuevas consultas en otras fuentes externas para obtener más metadatos complementarios.

Las fuentes externas secundarias o complementarias serán aquellas a las que no se puede consultar la lista de publicaciones de un autor a través de alguno de sus identificativos estándar (ORCID, Scopus Author ID, WoS ResearcherID), pero que servirán para complementar la información de una publicación obtenida desde una fuente principal **cuando no se hayan obtenido los metadatos mínimos para describirla.**

En las siguientes secciones se exponen diversas fuentes externas y se comentará si se va a usar o no, así como qué tipo de fuente complementaria es cada una de ellas.

### 4.1. PubMed

Esta fuente externa no puede ser principal porque su API no permite la consulta de publicaciones de un autor a través de ningún tipo de identificador estándar del autor (p.ej. ORCID). No existe un identificador interno para cada investigador en esta fuente

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

***Una manera de hacer Europa***

externa por lo que no permite buscar por un autor a excepción de cuando se busca por su nombre. Además, los metadatos obtenidos de cada publicación son mínimos y no genéricos, por lo que no aportan información que pueda complementar a la que ya se dispone por una reclamación previa en una fuente principal que ya ha permitido obtener los metadatos mínimos de una publicación.

Además, la mayoría de los artículos que se encuentran en esta plataforma se pueden encontrar en otras fuentes externas como Scopus.

#### 4.2. Semantic Scholar

Esta fuente externa no puede ser una fuente principal porque no permite pedir información de publicaciones de un usuario a través de su ORCID, si bien permite realizar dicha consulta a través de un “author\_id” interno que cada investigador tiene asociado en esta fuente externa. Sin embargo, cuando devuelve información de las publicaciones de un autor, esta no incluye el DOI (identificativo estándar), lo que imposibilita complementar la información de la misma con otras fuentes externas. Por lo tanto, esta fuente externa solamente podría ser complementaria y deberá ser usada para obtener los metadatos necesarios de un artículo de tipo referencia.

#### 4.3. CrossRef

Esta fuente externa tampoco permite realizar peticiones con el ORCID de un autor. Además, tampoco se ha encontrado un identificador interno de cada usuario en esta fuente externa. Por lo tanto, tampoco se podría considerar como principal.

#### 4.4. Open Citations

Esta fuente de datos tiene la gran limitación de que solo ofrece la conexión con el DOI de los artículos que cita o referencia uno dado. Sin embargo, este metadato, como se verá

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

*Una manera de hacer Europa*

en posteriores análisis, es difícil de encontrar por lo que esta fuente es primordial para conseguir esta relación entre documentos. Además, al devolvemos el DOI de un artículo nos permite completar los metadatos de esta publicación con otra fuente externa.

#### 4.5. European PMC

Al igual que ocurre con Open Citations esta fuente está diseñada para obtener metadatos de los ROs de tipo referencia o cita. En este caso la búsqueda es un poco más enrevesada ya que en primer lugar se debe obtener el ID en la página de esta fuente externa. Posteriormente preguntar tanto por las citas como por las referencias de este ID.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

*Una manera de hacer Europa*

Unión Europea

## 5. Compatibilidad de fuentes

Las fuentes externas secundarias o complementarias serán aquellas a las que no se puede consultar la lista de publicaciones de un autor a través de alguno de sus identificativos estándar (ORCID, Scopus Author ID, WoS ResearcherID), pero que servirán para complementar la información de una publicación obtenida desde una fuente principal **cuando no se hayan obtenido los metadatos mínimos para describirla.**

Tal y como se ha analizado previamente, las fuentes complementarias que se utilizarán serán **Semantic Scholar, PubMed y Crossref. Scopus y WoS también se podrán usar como complementarias ya sea mediante APIs que ofrecen información de otro tipo sobre los artículos (como por ejemplo de la métrica de una revista p. ej. Web of Science Journals API) o cuando después de la llamada a las fuentes secundarias no se haya obtenido la información necesaria.**

### 5.1 Análisis de la compatibilidad entre fuentes externas

Esta invocación ha de hacerse de forma automática teniendo en cuenta diferentes compatibilidades para determinar qué fuentes complementarias se podrán usar en cada caso:

- **Compatibilidad de la fuente principal y la complementaria.** Esta compatibilidad se basa en que la fuente principal haya devuelto el metadato necesario para realizar una consulta en la fuente complementaria.
- **Compatibilidad de metadatos necesarios.** Dados los metadatos obtenidos en la fuente principal ha de elegirse la fuente complementaria que permitirá recuperar los metadatos no obtenidos previamente. Es decir, aquella que permita complementar mejor la información de la fuente principal.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

### *Una manera de hacer Europa*

Igualmente, se tendrá en cuenta el análisis general de las fuentes externas realizado en las secciones anteriores para determinar su capacidad (probabilidad de éxito) para devolver información sobre las publicaciones de un determinado autor. De este modo se podrá valorar qué fuente complementaria usar en caso de que haya varias opciones.

A continuación, se detalla cómo se analizarán las compatibilidades mencionadas. Cabe mencionar que se trata de un **análisis preliminar** realizado a partir de la documentación oficial de las fuentes y de algunas interfaces web de prueba que estas ofrecen.

Es necesario realizar un análisis más exhaustivo basado en pruebas reales cuando se puedan realizar los registros en algunas APIs, además de contemplar si las restricciones de uso de algunas APIs son compatibles con el uso que se les quiere dar en Hércules. En concreto, está pendiente:

- El proceso de registrarse para acceder a la API de **ORCID** es complejo y hay varias opciones ([Registering a Member API Client](#) o [Registering a Public API client](#)). Estamos analizando los requisitos de cada una de las opciones y recabando la información que solicitan en su web.
  - **[ACTUALIZACIÓN]** Ya estamos registrados y haciendo pruebas.
- **WoS**aún no ha contestado a la solicitud de registro y no tenemos acceso programático vía API-key, por lo que no se ha podido probar realmente si los resultados que se obtienen coinciden con los ejemplos de su documentación y la interfaz web de la misma. Hay muchos metadatos mínimos necesarios para una publicación que debemos comprobar si son devueltos o no, como por ejemplo los relativos a la bibliografía del artículo.
  - **[ACTUALIZACIÓN]** Ya estamos registrados y haciendo pruebas.
- Determinar si las severas restricciones respecto al reducido número de consultas que permite **PubMed** la hacen candidata a fuente externa utilizable en Hércules.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

**Una manera de hacer Europa**

## 5.2 Análisis de la compatibilidad de la fuente principal y complementaria

En la siguiente tabla se muestran en las filas las fuentes externas principales y en las columnas las fuentes externas complementarias, indicando para cada una de ellas el nombre del metadato identificativo necesario para realizar búsquedas (en las principales ID de usuario y en las complementarias IDs de publicaciones).

Si la fuente principal proporciona el metadato identificativo necesario para obtener más información sobre una publicación en la fuente externa complementaria, entonces el fondo del cuadrante que une dichas fuentes será verde indicando compatibilidad. Si no, será rojo.

	Crossref (DOI)	PudMed (PudMedID)	Semantic Scholar (DOI)	Scopus (2ª llamada) (Scopus_ID) (DOI)	WoS revista (ISSN)	Open Citation (DOI)	European PMC (DOI)
Scopus (1ª llamada) (ORCID)				Scopus_ID			
WoS (ORCID)				DOI			
ORCID (ORCID)				DOI			

## 5.3 Compatibilidad de metadatos

En este análisis se han examinado los metadatos que se pueden obtener en las fuentes externas principales y complementarias, con el fin de ver que fuente complementaria es más adecuada para completar la información que devuelve cada fuente principal.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

*Una manera de hacer Europa*

Para ello se han creado tres tablas, una por cada fuente principal, cuya estructura es la siguiente:

- En la primera columna aparecen todos los metadatos a obtener para una publicación.
- La segunda columna pertenece a la fuente principal. En cada celda se indica con un 1 si es posible obtener el metadato de la fila correspondiente o un 0 en caso contrario.
- Las siguientes columnas van de dos en dos. Cada par de columnas pertenece a una fuente complementaria:
  - Las celdas de la primera columna indican con un 1 si es posible obtener el metadato de la fila correspondiente de la fuente complementaria o un 0 en caso contrario.
  - En las celdas de la segunda columna se realiza una suma, por fila, de los valores obtenidos entre la fuente principal y la fuente complementaria. Si el valor es mayor que 1 (celda con fondo verde), entonces ese metadato es recuperable, ya sea con la fuente principal o con la complementaria. Obviamente, un 0 (celda con fondo rojo) indica que ese metadato no es recuperable con esa combinación de fuentes.
  - En la última fila se ofrece un recuento de cuántos metadatos se obtienen en cada fuente o en la combinación de las fuentes (casillas moradas).

Dado que estas tablas ofrecen gran cantidad de información, se documentan en ficheros Excel aparte.

#### 5.4 Gestión de publicaciones incompletas

En caso de que con la combinación de fuente principal y fuente complementaria no se llegue a cubrir la información necesaria para registrar una publicación en el grafo de conocimiento de Hércules, se analizará la posibilidad de utilizar varias fuentes

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

**Unión Europea**

***Una manera de hacer Europa***

complementarias a la fuente principal. Por ejemplo, si la información de la bibliografía de una publicación no está completa, se podrían invocar a microservicios de fuentes externas complementarias a través de los DOIs de los artículos presentes en la bibliografía.

Posteriormente, habrá que determinar cómo gestionar las colisiones de metadatos que se puedan producir. La gestión se realizará de manera automática o manual, mostrando una interfaz al usuario para el tratamiento de dichos metadatos.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

*Una manera de hacer Europa*

## 6. Propuesta de proceso de reclamación de publicaciones

A continuación, se describen los pasos que se llevarán a cabo durante el proceso de reclamación de publicaciones de un determinado autor. El algoritmo se ha diseñado en base a la información que se requiere saber de cada artículo de un autor y al análisis de metadatos realizado en la siguiente sección.

**Nota: ORCID se retira debido a que nunca va a tener ni más información ni de mejor calidad que las dos fuentes externas principales identificadas en el presente análisis: Scopus (Elsevier) y WoS (Clarivate).**

### Función principal.

Primeramente, el investigador ofrece su ORCID y una fecha a partir de la cual quiere obtener sus ROs.

Se llamará a los servicios de WoS y Scopus para obtener la información de las publicaciones principales de este autor.

Se recorre cada una de las publicaciones obtenidas en WoS. Por cada una de ellas:

Se almacena el DOI en una lista para saber qué artículos ya hemos completado del investigador en cuestión.

Se llama al servicio de Semantic Scholar y se fusiona la información obtenida por este microservicio y la publicación que estamos examinando. (**Función de**

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

### *Una manera de hacer Europa*

**combinar dos publicaciones**). El resultado de esta unificación será la publicación que estamos observando.

A continuación, se llama a CrossRef y se fusiona la información obtenida por este microservicio y la publicación que estamos examinando. **(función de combinar dos publicaciones)**. El resultado de esta unificación será la publicación que estamos observando.

Se completa la bibliografía de este documento, ya que la fuente externa de CrossRef devuelve bibliografía, pero estas publicaciones tienen muy pocos metadatos **(función de completar bibliografía)**.

Se obtiene la bibliografía restante y las citas de la publicación que estamos observando **(función para obtener bibliografía y citas)**.

Se llama a la fuente externa Zenodo y en caso de encontrarse un fichero PDF con la publicación se añadirá como metadato.

Se llama al enriquecimiento de áreas temáticas y de palabras clave para completar la publicación.

Se añaden las métricas de las revistas.

Se recorren todos los documentos obtenidos por Scopus y para cada uno de ellos:

Si el DOI de esta publicación coincide con la publicación que estamos examinando entonces se combina la información **(función de combinar dos publicaciones)**.

En caso contrario no se hace nada.

Llegados a este punto la publicación central está completa, así como todas las bibliográficas y citas que la componen. Se guarda para devolverse.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

***Una manera de hacer Europa***

Recorremos la lista de publicaciones de Scopus con el fin de completar aquellas que no se han obtenido de WoS. Por tanto, por cada una de las publicaciones:

Si ya ha sido completada y almacenada antes, no hace nada con ella.

En caso contrario:

Se llama al servicio de Semantic Scholar y se fusiona la información obtenida por este microservicio y la publicación que estamos examinando (**función de combinar dos publicaciones**). El resultado de esta unificación será la publicación que estamos observando.

Se llama a CrossRef y se fusiona la información obtenida por este microservicio y la publicación que estamos examinando (**función de combinar dos publicaciones**). El resultado de esta unificación será la publicación que estamos observando.

Se completará la bibliografía de este documento, ya que la fuente externa de CrossRef devuelve bibliografía, pero estas publicaciones tienen muy pocos metadatos (**función completar bibliografía**).

Se obtiene la bibliografía restante y las citas de la publicación que estamos observando (**función para obtener bibliografía y citas**).

Se llama a la fuente externa Zenodo y en caso de encontrarse un fichero PDF con la publicación se añadirá como metadato.

Se llama al enriquecimiento de áreas temáticas y de palabras clave para completar la publicación.

Se añaden las métricas de las revistas.

Llegados a este punto ya tenemos completas todas las publicaciones de este autor.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

*Una manera de hacer Europa*

Unión Europea

## Función de combinar dos publicaciones

Con esta función se combinan todos los metadatos de las publicaciones recibidas. Cada metadato se combina de forma independiente. En el caso de los autores se hace de modo que no permita duplicidad de usuarios en el mismo conjunto de colaboradores de la publicación.

## Función completar bibliografía

Por cada artículo de la bibliografía que hay en una publicación se realiza lo siguiente:

Se llama al servicio de Semantic Scholar y se fusiona la información obtenida por este microservicio y la publicación que estamos examinando (**función de combinar dos publicaciones**). El resultado de esta unificación será la publicación que estamos observando.

Se llama a CrossRef y se fusiona la información obtenida por este microservicio y la publicación que estamos examinando (**función de combinar dos publicaciones**). El resultado de esta unificación será la publicación que estamos observando. En este caso no nos quedamos con la bibliografía devuelta por CrossRef.

Se llama a la fuente externa Zenodo y en caso de encontrarse un fichero PDF con la publicación se añadirá como metadato.

Se llama al enriquecimiento de áreas temáticas y de palabras clave para completar la publicación.

Se añaden las métricas de las revistas.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Unión Europea

*Una manera de hacer Europa*

Llegado a este punto la información de las publicaciones de bibliografía ya estaría completa.

### Función para obtener bibliografía y citas

Se llama al servicio externo de Open Citations con el que obtendremos una publicación con un listado de artículos bibliográficos (que solo tendrán como metadato el DOI) y una lista de artículos citados que también solo obtendrán el DOI.

Para cada elemento de la lista de bibliografía:

Se verifica que este elemento no está en la bibliografía que la publicación central que estamos analizando. En caso de que no esté almacenada se realizará lo siguiente:

Se llama al servicio de Semantic Scholar y se fusiona la información obtenida por este microservicio y la publicación que estamos examinando (**función de combinar dos publicaciones**). El resultado de esta unificación será la publicación que estamos observando.

Se llama a CrossRef y se fusiona la información obtenida por este microservicio y la publicación que estamos examinando (**función de combinar dos publicaciones**). El resultado de esta unificación será la publicación que estamos observando. En este caso no nos quedamos con la bibliografía devuelta por CrossRef.

Se llama a la fuente externa Zenodo y en caso de encontrarse un fichero PDF con la publicación se añadirá como metadato.

Se llama al enriquecimiento de áreas temáticas y de palabras clave para completar la publicación.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

*Una manera de hacer Europa*

Unión Europea

Se añaden las métricas de las revistas.

Llegado a este punto la información de la bibliografía ya estaría completa.

Para cada elemento de la lista de citas:

Se llama al servicio de Semantic Scholar y se fusiona la información obtenida por este microservicio y la publicación que estamos examinando (**función de combinar dos publicaciones**). El resultado de esta unificación será la publicación que estamos observando.

Se llama CrossRef y se fusiona la información obtenida por este microservicio y la publicación que estamos examinando (**función de combinar dos publicaciones**). El resultado de esta unificación será la publicación que estamos observando. En este caso no nos quedamos con la bibliografía devuelta por CrossRef.

Se llama a la fuente externa Zenodo y en caso de encontrarse un fichero PDF con la publicación se añadirá como metadato.

Se llama al enriquecimiento de áreas temáticas y de palabras clave para completar la publicación.

Se añaden las métricas de las revistas.

Llegado a este punto la información de las publicaciones de bibliografía ya estaría completa.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

*Una manera de hacer Europa*

## 7. Otros tipos de publicaciones

Existen a su vez otros ROs (Dataset, Presentación, Gráficos, Docs, Enlaces, Video, Poster, Lesson) cuyos elementos son en definitiva objetos identificados con un DOI, que a su vez serán referenciados o citados por otros ROs.

A priori se podrá utilizar una lógica similar a la del algoritmo diseñado para las fuentes externas de ROs de tipo publicación, aunque habrá que analizar otras fuentes externas para estos tipos de ROs:

- Zenodo. Permite realizar consultas mediante el ORCID de un autor.
- Figshare (Dataset, Presentación, Gráficos, Docs). ID propio.
- OpenAire. Permite realizar consultas mediante el ORCID de un autor.
- Slideshare (Presentación) (Ejemplo de presentación). ID propio.

Sin embargo, para estos ROs en principio no será necesario consumir fuentes externas secundarias o complementarias, ya que la información obtenida en la primera consulta de ROs para un determinado autor presumiblemente ya los describe suficientemente.

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

*Una manera de hacer Europa*

## Bibliografía

- [1] «Research Object Crate». [En línea]. Disponible en: <https://www.researchobject.org/>. [Accedido: 3-marzo-2021].
- [2] «Figshare». [En línea]. Disponible en: <https://figshare.com>. [Accedido: 3-marzo-2021].
- [3] «Share». [En línea]. Disponible en: <https://www.share-research.org>. [Accedido: 3-marzo-2021].
- [4] «GitHub». [En línea]. Disponible en: <https://github.com>. [Accedido: 3-marzo-2021].
- [5] «Research Object Crate». [En línea]. Disponible en: <https://www.researchobject.org/>. [Accedido: 3-marzo-2021].
- [6] «CiTO, the Citation Typing Ontology». [En línea]. Disponible en: <https://sparontologies.github.io/cito/current/cito.html>. [Accedido: 3-marzo-2021].
- [7] «Web Annotation Data Model». [En línea]. Disponible en: <https://www.w3.org/TR/annotation-model>. [Accedido: 3-marzo-2021].
- [8] «Dublin Core». [En línea]. Disponible en: <https://www.dublincore.org/specifications/dublin-core/dcmi-type-vocabulary/>. [Accedido: 3-marzo-2021].

Documento Acceso a fuentes externas de publicaciones (EFx-x.x)

Autor



Versión

Fecha

1

20/10/2021