Information Processing and Management 60 (2023) 103385

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

New trends in bibliometric APIs: A comparative analysis

A. Velez-Estevez [a,1], I.J. Perez [b,1], P. García-Sánchez [c,1], J.A. Moral-Munoz [d,1], M.J. Cobo [b,*,1]

[a] *Department of Computer Science and Engineering, Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Cádiz, Escuela Superior de Ingeniería, Avda. de la Universidad, 10, Puerto Real 11519, Cádiz, Spain*
[b] *Department of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, 18071, Granada, Spain*
[c] *Department of Computer Architecture and Computer Technology, ETSIIT-CITIC, University of Granada, Spain*
[d] *Department of Nursing and Physiotherapy, University of Cadiz, Spain*

ARTICLE INFO

Keywords:
Scientific databases
API
Bibliographic information

ABSTRACT

The science of science practice requires the analysis of large and complex bibliometric data. Traditional data exporting from companies' websites is not sufficient, so APIs are used to access a larger corpus. Therefore, this study aims not only to establish a taxonomy but also to offer a comparative analysis of 44 bibliographic APIs from various non-profit and commercial organizations, analyzing their characteristics and metadata with descriptive analysis, their possible bibliometric analyses, and the interoperability of the APIs across four different data categories: general, content, search, and query modes. The study found that Clarivate Analytics and Elsevier offer highly versatile APIs, while non-profit organizations, such as OpenCitations and OurResearch promote the Open Science philosophy. Most organizations offer free access to APIs for non-commercial purposes, but some have limitations on metadata retrieval. However, CrossRef, OpenCitations, or OpenAlex have no restrictions on the metadata retrieval. Co-author analysis using author names and bibliometric evaluation using citations are the types of analyses that can be done with the data provided by most APIs. DOI, PubMedID, and PMCID are the most versatile identifiers for extending metadata in the APIs. Semantic Scholar, Dimensions, ORCID, and Embase are the APIs that offer the most extensibility. Considering the obtained results, there is no single API that gathers all the information needed to perform any bibliometric analysis. Combining two or more APIs may be the most appropriate option to cover as much information as possible and enrich reports and analyses. This study contributes to advancing the understanding and use of APIs in research practice.

1. Introduction

Currently, researchers, institutions, and different organizations linked to science produce many publications, such as journal articles, book chapters, and conference proceedings, among different document types. And, usually, this information is stored in bibliographical databases such as the well-known Scopus, Dimensions, or Web of Science. Also, notably, the number of databases has grown. In fact, sources such as lens.org (Penfold, 2020), OpenAlex (Singh Chawla, 2022), Semantic Scholar (Jones, 2015), or

* Corresponding author.
E-mail address: mjcobo@decsai.ugr.es (M.J. Cobo).
[1] All authors contribute equally to the whole manuscript.

https://doi.org/10.1016/j.ipm.2023.103385
Received 2 December 2022; Received in revised form 28 February 2023; Accepted 12 April 2023
Available online 13 May 2023
0306-4573/

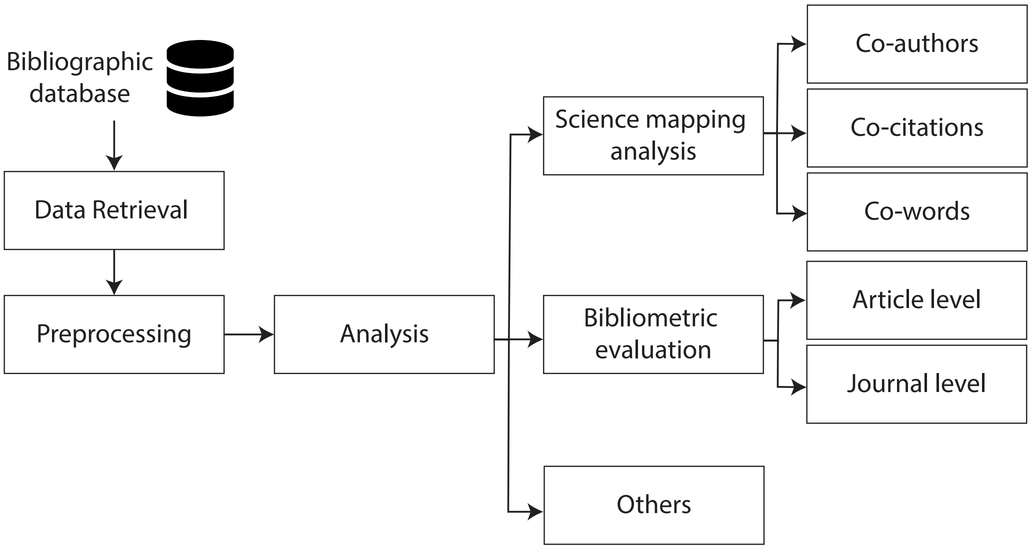

Fig. 1. Main process of bibliometric analysis.

Unpaywall (Else, 2018), among others, have recently appeared, thus complementing the well-established Web of Science and Scopus databases. In contrast, other databases have been discontinued, such as Microsoft Academic Graph (MAG) and Cobaltmetrics.

Moreover, there are multiple studies that have shed new light, helping analysts choose the best databases and tools to perform their analyses. In this sense, there are papers describing the features of bibliographical databases, such as Dimensions (Orduña-Malea & Delgado López-Cózar, 2018), Scopus (Baas, Schotten, Plume, Cote, & Karimi, 2020), CrossRef. (Hendricks, Tkaczyk, Lin, & Feeney, 2020), or OpenCitations (Peroni & Shotton, 2020), among others. Also, regarding the comparison of databases, there are also some studies comparing the coverage (Martín-Martín, Orduna-Malea, Thelwall, & Delgado López-Cózar, 2018; Singh, Singh, Karmakar, Leta, & Mayr, 2021), as well as their search features (Gusenbauer & Haddaway, 2020; Khurana, Ganesan, Kumar, & Sharma, 2022; Martín-Martín, Orduna-Malea, Thelwall, & Delgado-López-Cózar, 2019). Also, Moral-Muñoz, Herrera-Viedma, Santisteban-Espejo, Cobo et al. (2020) performed a deep review of the available tools to perform bibliometric analyses.

Because of the large number of scientific documents, the analysis of these, either for the analysis of scientific production and its impact, or for more complex analyses such as collaborative networks, topics, or economic analyses, requires specific techniques. In this sense, the so-called *Science of Science* (Fortunato et al., 2018) has taken advantage of the increasing availability of these data in a digital format, making possible the description of science as a network of different entities, such as researchers, projects, and papers, among others. Moreover, large analyses using a bigger corpus could be performed with the assistance of artificial intelligence, complex systems, mathematical models, and social science techniques. In fact, these analyses based on artificial intelligence (Liang, Mao, Lu, Ba, & Li, 2021) are gaining more attention in international collaboration, and therefore, greater citation impact (Velez-Estevez, García-Sánchez, Moral-Munoz, & Cobo, 2022).

Science of science is primarily based on techniques like bibliometric analysis and science mapping analysis, and is currently complemented with artificial intelligence methods and complex systems models to reveal new patterns and insights in scientific output. Studies based on bibliometric analysis generally follow a common process (Börner, Chen, & Boyack, 2003; Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011b), as shown in Fig. 1. The first step involves selecting a bibliographic database and defining a query to delimit the corpus to be analyzed. The data must be exported and added to existing bibliometric analysis tools (Moral-Muñoz et al., 2020), which is typically performed through the exporting capabilities of the databases' websites (e.g., Web of Science, Scopus). Before analysis, the data must be preprocessed to avoid errors and normalize the unit of analysis through a de-duplication process (Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011a). There are three main types of analysis: (i) science mapping analysis, which is based on bibliographic networks (Batagelj & Cerinšek, 2013); (ii) bibliometric evaluation, which can be done at the article or journal level and based on citations or altmetrics; and (iii) other analyses such as economic, leadership, and collaboration. Finally, the results can be visualized using different options and interpreted by the analyst.

As was mentioned, science of science analyses use to be more complex, facing and needing a large amount of data. For instance, if a researcher wants to analyze all the scientific output of Spain during the last decade, he/she will need to retrieve over 1 million documents (i.e., according to a query made on the Web of Science). Hence, manual access to that information through web forms and web pages is a tedious, complex, and low-performance process. In this sense, most bibliometric databases currently offer access to their information by using REST APIs (Torres-Salinas & Arroyo-Machado, 2022), which is a technique that allows us to automatically access and retrieve information between different machines. For instance, on Twitter, there are operations in their API that allow us to retrieve the tweets of a specific user or hashtag. Hence, as multiple APIs exist and they do not follow a common standard, there is a big variation in the features they offer to their users.

As previously said, the availability of access to large volumes of data through APIs is an important aspect, in order to make better and deeper analyses, that could extract the hidden knowledge in scholarly output, and find new patterns and behavior in science. Furthermore, as there are a great variety of bibliographical databases, with different characteristics, purposes, audiences, etc., the number of bibliometric APIs is also high. A previous description of bibliometric APIs was presented by Torres-Salinas and Arroyo-Machado (2022). In their paper, the authors shortly discuss the benefits of using APIs, as well as a list of 42 APIs and their URLs, along with a list of Python and R software packages for querying 13 of them. However, there is no comparison of the functionalities of each one, nor are they compared in detail. Therefore, it is necessary to perform a deeper comparative study to understand and uncover the differences, similarities, and possible applications of each API. Thus, this paper aims to develop a deeper comparative analysis of the main bibliometric and bibliographical APIs, comparing their core features and their possible usages, as well as the possible analyses that can be performed with them, and how to combine or jointly use multiple APIs to extend the information.

To do so, this paper is organized as follows. First, in Section 3, the inclusion criteria and methodology to analyze the different APIs are explained. Then, Section 4 shows the different categories of analyzed information as a global overview of the APIs. After that, we focus on the bibliometric analyses that can be performed using the APIs, and on the interoperability between them, in Sections 5 and 6, respectively. In Section 7, the results are discussed describing the usefulness and implications of this analysis. The lessons learned and future works are presented in Section 8. Finally, the paper ends with to Appendixes. The former is Appendix A, which summarizes the included APIs, and the latter is Appendix B, which details how the APIs can be used.

2. Objectives

The main objective of this paper is to develop a global comparative analysis of the principal bibliometric and bibliographical APIs. In order to do that, this objective is subdivided into several sub-objectives:

- Identify the APIs. To do that, we have collected the APIs from the main providers of bibliometric and bibliographical sources. For each identified API, we identify the general characteristics and build a taxonomy to classify their metadata for ulterior comparison.
- Comparison of the general characteristics and metadata provided by the APIs. To do that, we will analyze different aspects, such as global characteristics, search features, content features, and the ways to perform the query.
- Study of the bibliometric analyses that can be performed. In bibliometrics, there are different types of analyses that can be carried on. First, we will analyze the suitable APIs to conduct science mapping analysis. This analysis can be performed based on co-words (using authors' keywords or generated keywords), co-citation (using references' identifiers, references' authors, or references' journals), or co-authors (using the names' or authors' identifiers), each of them uncovering different aspects of science (Cobo et al., 2011b). Also, we will study the bibliometric evaluation considering metrics on the article and journal levels. At the article level, we considered traditional citations and altmetrics, whereas at the journal level, we studied metrics such as the JCR or SJR. Moreover, we will study the APIs that are suitable to conduct leadership, collaboration, and economic analyses.
- Analysis of the interoperability of the APIs from two points of view, the extension and the merging of metadata retrieved. That is, in the extension of metadata, we want to analyze how to extend a dataset retrieved from an API. To do that, we will analyze the APIs that allow us to search by identifiers present in the dataset to extend. Regarding the analysis of the merging of metadata, we will focus on the shared identifiers in the retrieved metadata of the APIs.

3. Methods

In this section, we present the criteria for selecting the different APIs that will be analyzed in the rest of the paper. Moreover, we explain the analyses we performed to obtain responses to the objectives of the paper.

3.1. Selection criteria

Currently, APIs are based on different architectural styles such as SOAP (Simple Object Access Protocol) (Snell, Tidwell, & Kulchenko, 2001), or REST (Representational state transfer) (Amundsen, Ruby, & Richardson, 2013). However, REST has currently become the most common way to build APIs, and several companies are employing this type of APIs.

As was previously mentioned, bibliographical databases are currently providing access to their data through REST APIs. Therefore, we analyzed the APIs provided by well-known organizations involved in bibliographical databases, which mainly cover the metadata of articles in journals, conferences, books, and/or book chapters, among others. Moreover, we focused on the APIs providing metadata and metrics about researchers (e.g., ORCID and Publons), also known as profiles. Furthermore, the traditional and alternative metric APIs were also selected for the study (e.g., InCites, SciVal, OpenCitations, and Altmetrics). Some unavailable APIs or discontinued were discarded (i.e., Microsoft Academic and PLOS Article Level Metrics).

3.2. Data collection and categories of information

Once the APIs were identified, the data describing their characteristics were collected. To do so, we gathered information on the individual APIs existing in March 2022, from their official websites, and used the APIs metadata. After the data were collected, we classified it into four categories, as it is described in Fig. 2, where each category has different dimensions, and each dimension has a set of features. Hence, the categories are:

1. **General category**. Three dimensions are present (i) the **requests** related features, such as the **time constraints** when doing the requests, the maximum results that can be retrieved with pagination, the page size allowed by the API (number of elements in each page), and the availability of the cursor pagination mode; (ii) the **subscription types** to the APIs, that cover the free, academic and commercial; (iii) and the **query options**, that comprise the sorting, wildcards and aggregation features.

2 and 3. **Search & Content categories**. They are about the search features and the metadata returned by the APIs, respectively. We have explained them together since they share some dimensions (the features in *cursive* mean that they are only available in the content category):

- **Bibliographical** dimension. It comprises the author's name, editors, title, abstract, date, journal, publisher, pages, volume, issue, ISBN, and ISSN features. Both categories contain these features.
- **Author**. The features collected were the ORCID, ResearcherID, corresponding author, Author ID (internal identifier of the API for authors), profile keywords (keywords that are related to the author e.g., ORCID profile keywords), the *name of the author*, *alternative names of the author*, the *EID* (Elsevier identifier of the author), the *ScopusID* (author identifier in the Scopus database), and the *education* of the author (i.e., Degree, PhD., among others).
- **Affiliation**. It is related to the organization where the authors belong. In this sense, the collected features in this dimension were the ROR ID, GRID ID, RINGGOLD ID, organization city, organization country, organization name, conference series name, *organization type*, and *department*.
- **Funding**. The collected characteristics were the funder countries, funder organization identifiers (i.e., GRID ID, FundRef ID and ROR ID), *GrantIDs*, *SourceID*, *funder name*, *alternative names*, *funder organization type*, and *funding amount*.
- **Metrics**. It is only present in the content category, covering the collection of the citations, h-index, impact factor, altmetrics, JCR, SJR, CIRC, and CiteScore features.
- **Identifiers**. It covers the identifiers of publications available in the content and in the search categories. The identifiers collected in this dimension were: database ID, DOI, EID, PMCID, PubMed ID, WOS ID, Dimensions ID, LUI, OCI, EMBASE ID, Altmetrics ID, arxivID, PII, PUI, SemanticScholar ID, and OpenAlex ID.
- The rest of the features are categorized in the **other** dimension. The document type, author keywords, referenced publications, clinical trials ID, CAS Registry Number, and MeSH terms are shared between the content and search categories. In the specific case of the content category, the additional collected features were the full text and the citing publications. In the case of the search category, the additional feature was the generated keywords.

4. **Query modes category**. In this category, two dimensions were analyzed:

- **Query-based**. The API provides a language that allows performing *boolean searches*, or even *grouping* and constructing complex conditions. The language to build the query can be free text or structured.
- **Resource-based**. The API provides several endpoints to filter the whole resource following single or several criteria. Multiple resources can be linked, so additional queries might be needed to retrieve additional information.

The former allows the user to be more concise and depending on the language used, allows also high expressiveness. In contrast, the latter is more technical and complex, and usually requires a high number of calls to retrieve additional information.

3.3. Analysis of the APIs features

As was mentioned above, bibliometric APIs have different characteristics and could be designed for different purposes. For that reason, a comparative framework should be developed in order to characterize each one, obtaining a global comparative that shows their common and specific characteristics. In that sense, three different comparative analyses are proposed: (i) descriptive, (ii) bibliometric, and (iii) interoperability. The descriptive analysis explains the metadata that can be retrieved and searched in each API. The information provided by this analysis is the main input for the types of bibliometric analyses that can be performed. Moreover, interoperability is also derived from the information provided by the descriptive analysis, as it is an analysis of what and how APIs can be extended or combined.

Descriptive analysis. We present and discuss the APIs following the four categories presented in the previous section, their dimensions and features, thus emphasizing their differences and similarities. The specific categories, dimensions, and features are summarized in Fig. 2.

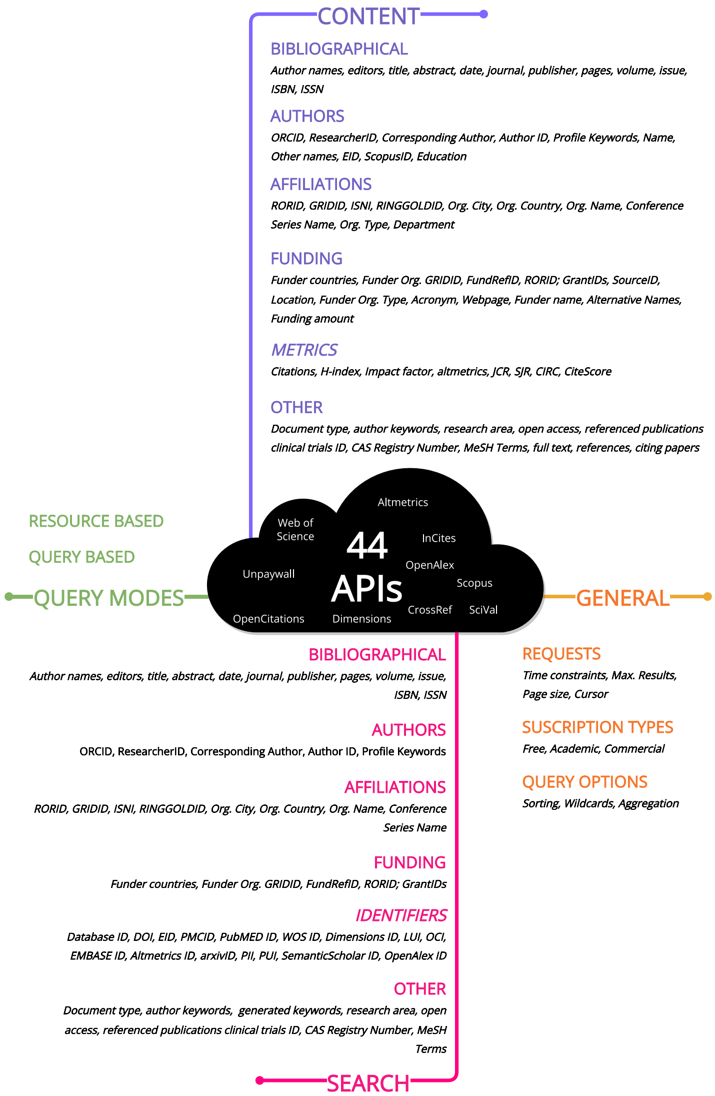

Fig. 2. Diagram of categories and dimensions of information about the APIs.

Bibliometric analyses. Each API gives a different vision of the data; therefore, some content could be shared among some of them, and other kinds of data could be specific to only one. There are a great variety of bibliometric analyses that need different kinds of data. Therefore, it is important to know what kind of bibliometric analysis could be done with each bibliometric API:

- *Science mapping analyses*. It checks the types of science mapping analyses that can be performed over the retrieved data. Concretely, we check if this analysis can be based on co-words (generated and author keywords), co-citation (references, authors, and journals), and co-author (name and identifier-based).
- *Bibliometric evaluation*. It checks two levels of evaluation, article (citations and altmetrics), and journal level (SJR, JCR, CIRC, CiteScore and Impact Factor).
- *Other*. We check if leadership (based on the corresponding author address), collaboration (based on the affiliation of the authors of the papers), and economic analyses (based on the funding) can be performed.

Interoperability. In this analysis, the focus is on two different views: (i) the ability to merge two or more datasets retrieved from the APIs and (ii) how to extend the metadata of a single dataset with metadata retrieved from other APIs. In the former, we analyze the fields that can be used to merge the collected data. In the latter, we study how, given a dataset, an analyst can extend the collected metadata with the metadata retrieved from other APIs.

Furthermore, the different proposed analyses are not mutually exclusive, but complementary. For instance, an analyst may want to perform a science mapping analysis based on co-words, but also a very specific analysis regarding other features. In this specific case, the analyst will check the conclusions of the bibliometric analyses and may then review the descriptive and interoperability analyses to find the best API, or a combination of APIs to perform the specific analysis.

4. Descriptive analysis

In this section, as explained in the methodology, the features of the selected APIs will be described, considering the classification in the four categories previously introduced. Hence, after searching and selecting the APIs, we identified the following ones (for a summary of these APIs, please refer to Appendix A):

- Semantic Scholar (Arum, 2016; Fricke, 2018; Jones, 2015).
- Lens (Penfold, 2020).
- InCites (Gray & Price, 2020; Panczyk, Woynarowska-Sołdan, Belowska, Zarzeka, & Gotlib, 2015).
- Journals.
- Publons (Ortega, 2017; Smith, 2016).
- Web of Science Expanded (Analytics, 2017).
- Web of Science Starter.
- BiorXiv (Abdill & Blekhman, 2019; Fraser, Momeni, Mayr, & Peters, 2020).
- ArXiv (Ginsparg, 2011).
- Altmetrics (Adie & Roe, 2013; Ortega, 2020).
- Dimensions (Hook, Porter, & Herzog, 2018; Thelwall, 2018).
- CIRC (Torres-Salinas, Bordons, Giménez-Toledo, Delgado López-Cózar, Jiménez-Contreras et al., 2010).
- Embase (Sampson et al., 2003).
- ScienceDirect Search (Hunter, 1998).
- ScienceDirect Article Metadata (Hunter, 1998).
- ScienceDirect Article (Full Text) Retrieval (Hunter, 1998).
- ScienceDirect Nonserial Title Metadata (Hunter, 1998).
- Scival (Dresbeck, 2015).
- Scopus Affiliations (Burnham, 2006) (aggregates search and retrieval APIs).
- Scopus Authors (Burnham, 2006) (aggregates search and retrieval APIs).
- Scopus Papers (Burnham, 2006) (aggregates Scopus Search and Abstract Retrieval APIs).
- Scopus Citations Count Metadata (Burnham, 2006).
- Scopus Citations Overview (Burnham, 2006).
- Scopus Serial Title Metadata (Burnham, 2006).
- Scopus PlumX Metrics (Champieux, 2015).
- Europe PMC (Ferguson et al., 2021; The Europe P.M.C. Consortium, 2014).
- IEEE Xplore Metadata (Griffin, 2002).
- IEEE Xplore Open Access (Griffin, 2002).
- CrossRef. (Pentz, 2001; van Eck & Waltman, 2022)
- iCite (Hutchins et al., 2019a; Hutchins, Davis, Meseroll, & Santangelo, 2019b; Hutchins, Hoppe, Meseroll, , erson et al., 2017; Hutchins, Yuan, Anderson, & Santangelo, 2016).
- NCBI PMC (Sayers et al., 2016).
- OpenCitations CCC.
- OpenCitations COCI (Heibi, Peroni, & Shotton, 2019b).
- OpenCitations CROCI (Heibi, Peroni, & Shotton, 2019a).
- OpenCitations Corpus (Peroni, Shotton, & Vitali, 2017).
- OpenCitations Indexes Unifying.
- ORCID (Haak, Fenner, Paglione, Pentz, & Ratner, 2012).
- OpenAlex (Singh Chawla, 2022).
- Unpaywall (Else, 2018).
- PLOS Search API (Federer et al., 2018; Savage & Vickers, 2009).
- PLOS Text Data Mining (Federer et al., 2018; Savage & Vickers, 2009).
- Springer Metadata API (Bekkari, 2012).
- Springer Open Access API (Bekkari, 2012).
- DBLP (Ley, 2002).

4.1. General category

The general accessing subscription and query option are shown in Table 1. The 44 APIs gathered in this paper belong to different organizations, being *Clarivate Analytics* and *Elsevier* the ones that own a high degree of multipurpose APIs. There are also APIs from

Table 1
API subscriptions, query options, pagination, and formats.

API	Free	Academic	Commercial	Sorting	Wildcards	Aggregation	Max. page size	Pagination	Cursor	JSON	XML
Semantic Scholar	✓		✓				100	✓		✓	
Lens	✓	✓	✓	✓			1K	✓	✓	✓	
Incites			✓				–			✓	✓
Journals			✓				50	✓		✓	✓
Publons			✓				10	✓		✓	
Web of Science Expanded			✓	✓	✓		100	✓		✓	✓
Web of Science Starter	✓		✓	✓	✓		50	✓		✓	✓
BiorXiv	✓					✓	100	✓	✓	✓	✓
ArXiv	✓			✓			2K	✓			✓
Altmetrics	✓		✓	✓			100	✓		✓	
Dimensions		✓	✓	✓	✓	✓	1K	✓		✓	
CIRC			✓				–	✓		✓	
Embase		✓	✓	✓	✓		25	✓		✓	✓
ScienceDirect Search		✓	✓	✓			200	✓		✓	✓
ScienceDirect Article Metadata		✓	✓				25	✓		✓	✓
ScienceDirect Article (Full Text) Retrieval	✓	✓	✓				–			✓	✓
ScienceDirect Nonserial Title Metadata	✓	✓	✓				200	✓		✓	✓
SciVal		✓	✓				500/100[a]	✓		✓	✓
Scopus Affiliations		✓	✓	✓	✓		200	✓		✓	✓
Scopus Authors		✓	✓	✓			200	✓		✓	✓
Scopus Papers	✓	✓	✓	✓	✓		25/200[b]	✓	✓	✓	✓
Scopus Citations Count Metadata		✓	✓				–			✓	✓
Scopus Citations Overview		✓	✓				25/200[b]	✓		✓	✓
Scopus Serial Title Metadata	✓	✓	✓				200	✓		✓	✓
Scopus PlumX Metrics		✓	✓				–			✓	✓
Europe PMC	✓						1K	✓	✓	✓	✓
IEEE Xplore Metadata	✓	✓		✓	✓		200	✓		✓	✓
IEEE Xplore Open Access	✓	✓					–	✓		✓	✓
CrossRef	✓			✓		✓	1K	✓	✓	✓	✓
iCite	✓						1K	✓		✓	
NCBI PMC	✓	✓	✓				10K	✓		✓	✓
OpenCitations CCC	✓			✓			–			✓	
OpenCitations COCI	✓			✓			–			✓	
OpenCitations CROCI	✓			✓			–			✓	
OpenCitations Corpus	✓			✓			–			✓	
OpenCitations Indexes unifying	✓			✓			–			✓	
OpenAlex	✓			✓		✓	200	✓	✓	✓	
Unpaywall	✓						50	✓		✓	
ORCID	✓		✓		✓		1K	✓		✓	✓
PLOS Search API	✓	✓	✓	✓			100	✓		✓	✓
PLOS Text Data Mining	✓	✓	✓	✓			100	✓		✓	✓
Springer Metadata API	✓		✓	✓	✓	✓	100	✓		✓	✓
Springer Open Access API	✓		✓	✓	✓	✓	100	✓		✓	✓
DBLP	✓				✓		1K	✓		✓	✓

[a] 500 for authors, 100 for other resources.
[b] 25 for non-commercial subscriptor, 200 for commercial. K stands for 1000 units.

non-profit organizations such as OpenCitations, and OurResearch, following the open science philosophy. Most organizations offer APIs that allow free non-commercial access to metadata, but some are limited and some are not. Also, there are APIs such as CrossRef, OpenCitations, or OpenAlex that offer their metadata for any purpose without limiting the maximum amount of metadata that can be retrieved. Moreover, most APIs return the information in the JSON format, with the single exception of ArXiv, which only does it in XML.

As a user, there are three main ways to access the APIs: free, academic, and commercial access. The APIs from *Clarivate*, *Elsevier*, and *Dimensions* cover various analyzed features, as we will discuss later. However, all the APIs of *Clarivate Analytics*, except for the *Web of Science Starter*, require a commercial subscription. In contrast, *Elsevier* APIs offer free, academic, and commercial access to their APIs, depending on the use case. Also, *Digital Science* offers academic (explaining the research project) and commercial access. There are several free APIs, such as *CrossRef*, *OpenAlex*, *Semantic Scholar*, *Lens*, *Unpaywall*, *Springer Metadata* and *Springer Open Access*, *OpenCitations Corpus* and *Indexes Unifying* that offer a wide range of bibliographical metadata.

From the point of view of a researcher, the retrieval of the records is fundamental to perform an analysis. Hence, in Table 2, we study request constraints, as well as the maximum results allowed. In this view, commercial rates are difficult to study because of the constraints being linked to private contracts between the clients and the provider. Therefore, focusing on the freely and academically

Table 2
APIs request constraints and maximum results. In the request constraints, *r* stands for *request*, and *s/m/d/w/M* stands for *second/minute/day/week/month*. Hence, for example, 2r/d stands for 2 requests a day. K stands for 1000 units.

API	Request constraints	Max. results (inc. pagination)
Semantic Scholar	100r/5 m (Free)	10K
Lens	20-100Kr/M (Commercial)	Unlimited
Incites	1Kr/d & 2r/s (Commercial)	100
Journals	5r/s (Commercial)	–
Publons	100 r/d (Free)	Unlimited
Web of Science Expanded	Basic (2r/s), Int (2r/s), Adv (3r/s), Prem (5r/s) (Commercial)	Basic (50K), Int (250K), Adv (1M), Prem (3M) (per year)
Web of Science Starter	Free (1r/s, max 50r/d), Institutional (5r/s, max 1Kr/d)	Free (2.5Kr/d), Institutional (50K/day)
BiorXiv	–	Unlimited
ArXiv	1r/3s (Free)	30K
Altmetrics	1r/s (Academic), Unlimited (Commercial)	Unlimited
Dimensions	30r/m (Reasonable use) (Academic)	50K
CIRC	–	–
Embase	6 r/s (Commercial)	–
ScienceDirect Search	2r/s (Commercial) (20Kr/w max.)	6K
ScienceDirect Article Metadata	6r/s (Commercial)	6K
ScienceDirect Article (Full Text) Retrieval	6r/s (Commercial)	1
ScienceDirect Nonserial Title Metadata	6 r/s (Commercial) (20Kr/w max.)	–
SciVal	6r/s (Commercial) (5Kr/w max.)	Unlimited
Scopus Affiliations	9r/s (Search, Academic); 6r/s (Retrieval, Academic) (5Kr/w max.)	5K
Scopus Authors	2r/s (Search, Academic); 3r/s (Retrieval, Academic) (5Kr/w max.)	5K
Scopus Papers	9r/s (Search, Retrieval, Academic); (max. 20Kr/w for search) (max. 10Kr/w for retrieval)	Unlimited with cursor (5K with offset)
Scopus Citations Count Metadata	10r/s (Academic) (max. 50Kr/w)	(depends on URL full length)
Scopus Citations Overview	4r/s (Academic) (max. 20Kr/w)	(depends on URL full length)
Scopus Serial Title Metadata	6r/s (Academic) (max. 20Kr/w)	–
Scopus PlumX Metrics	6r/s (Commercial) (1 request per metric type)	1
Europe PMC	–	Unlimited
IEEE Xplore Metadata	10r/s & 200r/d (Academic)	–
IEEE Xplore Open Access	10r/s & 200r/d (Academic)	1
CrossRef	50 r/s	Unlimited with cursor (80K with offset)
iCite	–	Unlimited
NCBI PMC	3r/s (without API Key); 10r/s (with API Key)	Unlimited
OpenCitations CCC	–	(depends on URL full length)
OpenCitations COCI	–	(depends on URL full length)
OpenCitations CROCI	–	(depends on URL full length)
OpenCitations Corpus	–	(depends on URL full length)
OpenCitations Indexes unifying	–	(depends on URL full length)
OpenAlex	< 100Kr/d	10K
Unpaywall	< 100Kr/d	–
ORCID	24r/s	10Kr (free), Unlimited (member & premium)
PLOS Search API	7.2Kr/d	–
PLOS Text Data Mining	7.2Kr/d	–
Springer Metadata API	300 r/m (max. 10Kr/d)	–
Springer Open Access API	300r/m (max. 10Kr/d)	–
DBLP	Unlimited	10K

Table 3
Bibliographical information. Ⓢ: means that the user can search for that feature. Ⓡ: means that the user can retrieve that feature.

API	Author name	Editor	Title	Abstract	Date	Journal	Publisher	Pages	Volume	Issue	ISSN	ISBN
Semantic Scholar	ⓈⓇ		ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓢ					
Lens	ⓈⓇ		ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓡ	Ⓡ	Ⓡ	ⓈⓇ	Ⓡ
Incites												
Journals			Ⓡ		ⓈⓇ	Ⓢ	ⓈⓇ				Ⓡ	Ⓢ
Publons	Ⓡ		Ⓡ			Ⓡ	Ⓡ				Ⓡ	
Web of Science Expanded	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓡ	Ⓡ	Ⓡ	ⓈⓇ	ⓈⓇ
Web of Science Starter	ⓈⓇ	Ⓡ	ⓈⓇ	Ⓢ	ⓈⓇ	ⓈⓇ		ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ
BiorXiv	Ⓡ		Ⓡ	Ⓡ	ⓈⓇ		Ⓢ					
ArXiv	ⓈⓇ		ⓈⓇ	ⓈⓇ	Ⓡ	ⓈⓇ						
Altmetrics	Ⓡ	Ⓡ	Ⓡ		Ⓡ	Ⓡ					Ⓡ	Ⓢ
Dimensions	ⓈⓇ		ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓡ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	
CIRC						Ⓡ					ⓈⓇ	
Embase	ⓈⓇ	Ⓢ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓡ	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓢ
ScienceDirect Search	ⓈⓇ		ⓈⓇ		ⓈⓇ	ⓈⓇ		ⓈⓇ	ⓈⓇ	Ⓢ		
ScienceDirect Article Metadata	ⓈⓇ		ⓈⓇ	Ⓡ	ⓈⓇ	ⓈⓇ		Ⓡ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ
ScienceDirect Article (Full Text) Retrieval	Ⓡ		Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ
ScienceDirect Nonserial Title Metadata	Ⓡ	Ⓡ	ⓈⓇ				Ⓡ					ⓈⓇ
SciVal	Ⓡ		Ⓡ		ⓈⓇ	ⓈⓇ						
Scopus Affiliations	Ⓡ		Ⓡ		Ⓡ	Ⓡ		Ⓡ	Ⓡ	Ⓡ	Ⓡ	
Scopus Authors	ⓈⓇ		Ⓡ		Ⓡ	Ⓡ		Ⓡ	Ⓡ	Ⓡ	Ⓡ	
Scopus Papers	ⓈⓇ	Ⓢ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓢ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ
Scopus Citations Count Metadata												
Scopus Citations Overview	Ⓡ		Ⓡ	Ⓢ		Ⓡ		Ⓡ	Ⓡ	Ⓡ	Ⓡ	
Scopus Serial Title Metadata				Ⓢ		ⓈⓇ	ⓈⓇ				Ⓡ	Ⓢ
Scopus PlumX Metrics												Ⓢ
Europe PMC	ⓈⓇ		ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓢ
IEEE Xplore Metadata	ⓈⓇ		ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	ⓈⓇ	ⓈⓇ
IEEE Xplore Open Access												
CrossRef	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓡ	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓡ	Ⓡ	Ⓡ	ⓈⓇ	ⓈⓇ
iCite	Ⓡ		Ⓡ		ⓈⓇ	Ⓡ						
NCBI PMC	ⓈⓇ	Ⓢ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓢ
OpenCitations CCC	Ⓡ		Ⓡ		Ⓡ	Ⓡ		Ⓡ	Ⓡ	Ⓡ		
OpenCitations COCI	Ⓡ		Ⓡ		Ⓡ	Ⓡ		Ⓡ	Ⓡ	Ⓡ		
OpenCitations CROCI	Ⓡ		Ⓡ		Ⓡ	Ⓡ		Ⓡ	Ⓡ	Ⓡ		
OpenCitations Corpus	Ⓡ		Ⓡ		Ⓡ	Ⓡ		Ⓡ	Ⓡ	Ⓡ		
OpenCitations Indexes unifying	Ⓡ		Ⓡ		Ⓡ	Ⓡ		Ⓡ	Ⓡ	Ⓡ		
OpenAlex	ⓈⓇ		ⓈⓇ	Ⓡ	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓡ	Ⓡ	Ⓡ	ⓈⓇ	
Unpaywall	Ⓡ		ⓈⓇ		Ⓡ	Ⓡ	Ⓡ				Ⓡ	
ORCID	ⓈⓇ		ⓈⓇ		Ⓡ						ⓈⓇ	ⓈⓇ
PLOS Search API	ⓈⓇ	Ⓢ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓢ		Ⓢ	Ⓢ	ⓈⓇ	
PLOS Text Data Mining	Ⓡ		Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	
Springer Metadata API	ⓈⓇ		ⓈⓇ	Ⓡ	ⓈⓇ	ⓈⓇ	Ⓡ	Ⓡ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ
Springer Open Access API	ⓈⓇ		ⓈⓇ	Ⓡ	ⓈⓇ	ⓈⓇ	Ⓡ	Ⓡ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ
DBLP	ⓈⓇ		ⓈⓇ		Ⓡ	ⓈⓇ		Ⓡ				

accessible, the more relevant ones are: *Lens*, *Semantic Scholar*, *Dimensions*, the APIs of *Elsevier*, *OpenAlex* and *Unpaywall*. Additionally, we should highlight that some APIs allow us to retrieval an unlimited amount of records, such as *NCBI PMC*, *EuropePMC*, *CrossRef*, *Lens*, *BiorXiv* and *Scopus Papers* and the *OpenAlex* API.

4.2. Search & content categories

In this section, we introduce the information gathered for the categories of search and content, and their dimensions. First, it is important to introduce some notation to understand the following tables: if a specific feature can be searched, it will be marked with Ⓢ symbol, and if it can be retrieved with Ⓡ.

4.2.1. Bibliographical information

In Table 3, the main bibliographical information gathered by the APIs is analyzed. It is observed that the *author's name*, *title*, *abstract*, *date*, *journal*, *ISSN*, and *ISBN* are the most covered search features. In contrast, a user of these APIs may face some difficulties when trying to delimit the search by *editor*, *volume*, or *issue*. Focusing on the retrieved bibliographical features, most of them are covered in all APIs. In contrast, the editor of the books is a less common feature, as it is returned to specific APIs (i.e., *Web of Science Expanded* and *Starter*, *Altmetrics*, *ScienceDirect Nonserial Title Metadata* and *Crossref*). Other features such as the *abstract*, *publisher*, and the *ISBN* are missing in some APIs too. Also, *Incites*, *CIRC*, *Scopus Citations Count Metadata*, and *Scopus PlumX Metrics*, are more specific and focused on returning primarily metric information.

Table 4
Authors.Ⓢ: means that the user can search for that feature.Ⓡ: means that the user can retrieve that feature.

API	Name	Other names	Author ID	ResearcherID	Profile keywords	EID	ORCID	ScopusID	Education	Corresponding author
Semantic Scholar	Ⓡ	Ⓡ	Ⓢ							
Lens	Ⓡ						ⓈⓇ			
Incites										
Journals										
Publons	Ⓡ	Ⓡ	ⓈⓇ	ⓈⓇ			ⓈⓇ		Ⓡ	
Web of Science Expanded	Ⓡ		ⓈⓇ	ⓈⓇ			ⓈⓇ			Ⓡ
Web of Science Starter	Ⓡ		Ⓡ	Ⓡ						
BiorXiv	Ⓡ									Ⓡ
ArXiv	Ⓡ									
Altmetrics	Ⓡ									
Dimensions	Ⓡ		ⓈⓇ				ⓈⓇ			ⓈⓇ
CIRC										
Embase	Ⓡ						Ⓢ			
ScienceDirect Search	Ⓡ									
ScienceDirect Article Metadata	Ⓡ									
ScienceDirect Article (Full Text) Retrieval	Ⓡ									Ⓡ
ScienceDirect Nonserial Title Metadata	Ⓡ									
SciVal	Ⓡ		ⓈⓇ					Ⓡ		
Scopus Affiliations	Ⓡ		Ⓡ			Ⓡ	Ⓡ	Ⓡ		
Scopus Authors	Ⓡ	Ⓡ	ⓈⓇ			Ⓡ	ⓈⓇ	Ⓡ		
Scopus Papers	Ⓡ	Ⓡ	ⓈⓇ				Ⓡ	Ⓡ		Ⓡ
Scopus Citations Count Metadata										
Scopus Citations Overview	Ⓡ		Ⓡ					Ⓡ		
Scopus Serial Title Metadata										
Scopus PlumX Metrics										
Europe PMC	Ⓡ						Ⓢ			
IEEE Xplore Metadata	Ⓡ		Ⓡ							
IEEE Xplore Open Access										
CrossRef	Ⓡ						ⓈⓇ			
iCite	Ⓡ									
NCBI PMC	Ⓡ		Ⓢ							Ⓡ
OpenCitations CCC	Ⓡ									
OpenCitations COCI	Ⓡ									
OpenCitations CROCI	Ⓡ									
OpenCitations Corpus	Ⓡ									
OpenCitations Indexes unifying	Ⓡ									
OpenAlex	Ⓡ	Ⓡ	ⓈⓇ			Ⓡ	ⓈⓇ	Ⓡ		
Unpaywall	Ⓡ						Ⓡ			
ORCID	Ⓡ	Ⓡ	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓡ	ⓈⓇ	Ⓡ	Ⓡ	
PLOS Search API	Ⓡ									
PLOS Text Data Mining	Ⓡ						Ⓡ			
Springer Metadata API	Ⓡ									Ⓡ
Springer Open Access API	Ⓡ									Ⓡ
DBLP	Ⓡ		Ⓡ							

4.2.2. Authors

The author's dimension is also considered in the search and content categories. As shown in Table 4, the *ORCID* and the internal author identifiers are covered in most APIs. The *ResearchID* search feature only exists on the *Web of Science Expanded*, *Publons*, and *ORCID* APIs. Also, the corresponding author search is only possible in *Dimensions*, and the search by profile keywords is only possible in *ORCID*. Focusing on the content features of the table, the APIs differ enormously. The most common features are the name of the author, other names, and the ORCID. Some fields could be retrieved only in some APIs, such as education in Publons and ORCID, profile keywords in ORCID, and ResearcherID, which is covered in Publons, both versions of the Web of Science APIs and ORCID. Also, some APIs offer other names, which are important for some specific use cases (e.g., disambiguating authors). Moreover, others maintain their own internal identifiers, such as *Dimensions* and *OpenAlex*. Finally, it should be highlighted that *ORCID*, *Publons*, and *Scopus Authors* cover most of the author's features.

4.2.3. Affiliation

In this section, we show the search and retrieve categories for the dimension of the affiliation information (Table 5). On the one hand, in the search by affiliation information, the most common search features are the *organization name* and *organization country*.

Table 5
Affiliation.Ⓢ: means that the user can search for that feature.Ⓡ: means that the user can retrieve that feature.

API	ROR ID	GRID ID	ISNI	RINGOLD ID	Organization city	Organization country	Organization name	Organization type	Conference series name	Department
Semantic Scholar					Ⓢ	Ⓢ	Ⓢ			
Lens	Ⓡ	ⓈⓇ			Ⓢ	ⓈⓇ	ⓈⓇ		Ⓢ	
Incites										
Journals										
Publons						Ⓡ	ⓈⓇ			
Web of Science Expanded					Ⓢ	ⓈⓇ	ⓈⓇ		Ⓢ	
Web of Science Starter							Ⓢ		Ⓢ	
BiorXiv							Ⓡ			
ArXiv							Ⓡ			
Altmetrics										
Dimensions	ⓈⓇ	ⓈⓇ	ⓈⓇ		Ⓢ	ⓈⓇ	ⓈⓇ	Ⓡ		
CIRC										
Embase					Ⓢ	ⓈⓇ	ⓈⓇ		Ⓢ	Ⓡ
ScienceDirect Search										
ScienceDirect Article Metadata					Ⓢ	Ⓢ	Ⓢ			
ScienceDirect Article (Full Text) Retrieval										
ScienceDirect Nonserial Title Metadata										
SciVal						ⓈⓇ	ⓈⓇ			
Scopus Affiliations					Ⓢ	ⓈⓇ	ⓈⓇ			
Scopus Authors					Ⓢ	ⓈⓇ	ⓈⓇ			
Scopus Papers					Ⓢ	ⓈⓇ	ⓈⓇ		Ⓢ	
Scopus Citations Count Metadata										
Scopus Citations Overview										
Scopus Serial Title Metadata									Ⓢ	
Scopus PlumX Metrics										
Europe PMC		Ⓢ			Ⓢ	ⓈⓇ	ⓈⓇ			
IEEE Xplore Metadata					Ⓢ	Ⓢ	ⓈⓇ			
IEEE Xplore Open Access										
CrossRef	ⓈⓇ				Ⓢ	ⓈⓇ	ⓈⓇ			
iCite										
NCBI PMC					Ⓢ	Ⓢ	Ⓢ			
OpenCitations CCC										
OpenCitations COCI										
OpenCitations CROCI										
OpenCitations Corpus										
OpenCitations Indexes unifying										
OpenAlex	ⓈⓇ	ⓈⓇ				ⓈⓇ	ⓈⓇ	Ⓡ		
Unpaywall							Ⓡ			
ORCID	Ⓢ	ⓈⓇ		ⓈⓇ	Ⓢ	ⓈⓇ	Ⓢ			Ⓡ
PLOS Search API					Ⓢ	Ⓢ	Ⓢ			
PLOS Text Data Mining						Ⓡ	Ⓡ			
Springer Metadata API						Ⓢ			Ⓢ	
Springer Open Access API		Ⓡ				ⓈⓇ	Ⓡ		Ⓢ	
DBLP										

However, the search by identifiers related to the affiliation of the authors is less usual. In fact, the *GRID ID* search is only present at 5 APIs (*Lens*, *Dimensions*, *Europe PMC*, *OpenAlex* and *ORCID*). Similarly, the search by the new unique identifier for affiliations, the *ROR ID*, which substitutes the *GRID ID*, is only possible in 4 APIs (*Dimensions*, *CrossRef*, *OpenAlex* and *ORCID*). Regarding the *ISNI* and *RINGGOLD* identifiers, *Dimensions* allows searching by *ISNI* and *ORCID* allows searching by *RINGGOLD*. *Scopus Affiliations* provides its internal database affiliation ID, although it is not shown in the table.

On the other hand, in the content category, most APIs return the organization name and the organization's country. However, the organization type is returned only by *Dimensions* and *OpenAlex*, and the *department* is solely at *Embase* and *ORCID*. It is important to highlight that the *department* can also be found in other APIs, such as the *Web of Science* or *Scopus Papers*, but additional processing is required to extract it from the full affiliation. Regarding the identifiers, the *GRID ID* and *ROR ID* are present in *Lens*, *Dimensions*, *OpenAlex*, *ORCID*, and *Springer Metadata* and *Open Access*. Additionally, *GRID ID* is only available in *ORCID* and Springer APIs. Moreover, the *RINGGOLD ID* and *ISNI* identifiers are solely in *ORCID* and *Dimensions*, respectively. From another perspective, the APIs with higher coverage of these returned metadata about affiliations are *Dimensions*, *OpenAlex* and *ORCID* (although *ORCID* is about profiles, not documents).

Table 6
Funding.Ⓢ: means that the user can search for that feature.Ⓡ: means that the user can retrieve that feature.

API	Countries	GRID ID	FundRef ID	ROR ID	GrantIDs	*Source ID*	*Name*	*Alternative names*	*Type*	*Funding amount*
Semantic Scholar										
Lens	ⓈⓇ				Ⓡ		Ⓡ			
Incites										
Journals										
Publons										
Web of Science Expanded	ⓈⓇ				ⓈⓇ		Ⓡ			
Web of Science Starter										
BiorXiv										
ArXiv										
Altmetrics										
Dimensions	ⓈⓇ	ⓈⓇ		ⓈⓇ	ⓈⓇ	Ⓡ	Ⓡ		Ⓡ	Ⓡ
CIRC										
Embase										
ScienceDirect Search										
ScienceDirect Article Metadata										
ScienceDirect Article (Full Text) Retrieval										
ScienceDirect Nonserial Title Metadata										
SciVal										
Scopus Affiliations										
Scopus Authors										
Scopus Papers	ⓈⓇ				ⓈⓇ	Ⓡ	Ⓡ			
Scopus Citations Count Metadata										
Scopus Citations Overview										
Scopus Serial Title Metadata										
Scopus PlumX Metrics										
Europe PMC	ⓈⓇ		Ⓢ		ⓈⓇ	Ⓡ	Ⓡ			
IEEE Xplore Metadata										
IEEE Xplore Open Access										
CrossRef	ⓈⓇ		ⓈⓇ		Ⓡ	Ⓡ	Ⓡ	Ⓡ		
iCite										
NCBI PMC					ⓈⓇ		Ⓡ			
OpenCitations CCC										
OpenCitations COCI										
OpenCitations CROCI										
OpenCitations Corpus										
OpenCitations Indexes unifying										
OpenAlex										
Unpaywall										
ORCID	ⓈⓇ	Ⓡ	ⓈⓇ		ⓈⓇ	Ⓡ	Ⓡ			
PLOS Search API										
PLOS Text Data Mining					Ⓡ		Ⓡ			
Springer Metadata API										
Springer Open Access API										
DBLP										

4.2.4. Funding

Following the subsequent dimensions, we now analyze the funding in Table 6. In this way, regarding the funding-related search category, most of the APIs lack support for searching by funding. The best APIs in terms of search are *Web of Science Expanded*, *Dimensions*, *Europe PMC*, *CrossRef*, *ORCID* (related to authors' funding), and *Scopus Papers*, since they cover almost all the funding search features analyzed (i.e., funder countries, funder organization GRID ID, funder organization ROR ID, and GrantIDs). The more common search features are the *funder countries* and the *Grant identifiers*, whereas the identifiers of funder organizations: GRID, FundRef ID, and ROR ID are uncommon. In this way, to use these identifiers, the users must use *Dimensions*, *Europe PMC*, *CrossRef*, or *ORCID*. Regarding free APIs, the best ones to search by funding data are *Dimensions*, *Europe PMC*, *CrossRef*, and *ORCID*.

Regarding the content category of this dimension, it is important to highlight that almost all the APIs exclude funding-related metadata, and in case they do, the metadata returned has few details. With the previous considerations, the best APIs to retrieve data about funding are *Dimensions*, *CrossRef*, and *ORCID*. A special mention should be done on *Dimensions* since it is the only API that returns the amount of money (and the currency) that is linked to a grant.

4.2.5. Identifiers

In this section, we analyze the identifier dimensions in Table 7. Regarding the search category for this dimension, the most common identifiers of documents among the APIs are the DOI, the PMCID, and the PubMed Id. Moreover, the APIs that offer a

Table 7
Identifiers.Ⓢ: means that the user can search by that feature.Ⓡ: means that the user can retrieve that feature.

API	Paper database id	DOI	EID (Elsevier electronic identifier)	PMCID	PubMedID	WOS ID	Dimensions ID	Open Citation ID	LUI (ID embase internal)	EMBASE ID	Altmetric ID	ArXiv ID	Publication Item ID (PII)	PUI (Scopus)	Semantic Scholar ID	OpenAlexID
Semantic Scholar	Ⓡ	ⓈⓇ		Ⓡ	ⓈⓇ							ⓈⓇ			ⓈⓇ	
Lens	ⓈⓇ	ⓈⓇ		Ⓢ	ⓈⓇ										Ⓢ	
Incites	ⓈⓇ					ⓈⓇ										
Journals																
Publons	ⓈⓇ	Ⓡ														
Web of Science Expanded	ⓈⓇ	ⓈⓇ			ⓈⓇ	ⓈⓇ										
Web of Science Starter	ⓈⓇ	ⓈⓇ			Ⓢ	ⓈⓇ										
BiorXiv		ⓈⓇ														
ArXiv	ⓈⓇ	Ⓡ										ⓈⓇ				
Altmetrics	ⓈⓇ	ⓈⓇ			ⓈⓇ						ⓈⓇ	ⓈⓇ				
Dimensions	ⓈⓇ	ⓈⓇ		Ⓢ	ⓈⓇ		ⓈⓇ				ⓈⓇ	ⓈⓇ				
CIRC																
Embase	ⓈⓇ	ⓈⓇ		Ⓢ	ⓈⓇ			Ⓢ	Ⓡ	ⓈⓇ			ⓈⓇ	ⓈⓇ		
ScienceDirect Search	Ⓡ	Ⓡ											Ⓡ			
ScienceDirect Article Metadata	ⓈⓇ	ⓈⓇ	ⓈⓇ										Ⓡ			
ScienceDirect Article (Full Text) Retrieval	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓢ	ⓈⓇ								Ⓢ	Ⓢ		
ScienceDirect Nonserial Title Metadata																
SciVal	ⓈⓇ	Ⓡ	ⓈⓇ													
Scopus Affiliations	Ⓡ	Ⓡ	ⓈⓇ													
Scopus Authors	Ⓡ	Ⓡ	ⓈⓇ													
Scopus Papers	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓢ	ⓈⓇ								Ⓢ	Ⓢ		
Scopus Citations Count Metadata	ⓈⓇ	ⓈⓇ	Ⓡ		ⓈⓇ								ⓈⓇ	Ⓢ		
Scopus Citations Overview	ⓈⓇ	ⓈⓇ	Ⓡ		ⓈⓇ								ⓈⓇ			
Scopus Serial Title Metadata																
Scopus PlumX Metrics	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ							ⓈⓇ	ⓈⓇ			
Europe PMC	ⓈⓇ	ⓈⓇ		ⓈⓇ	ⓈⓇ											
IEEE Xplore Metadata	ⓈⓇ	ⓈⓇ														
IEEE Xplore Open Access																
CrossRef		ⓈⓇ														
iCite	Ⓢ	Ⓡ			ⓈⓇ											
NCBI PMC	ⓈⓇ			ⓈⓇ	ⓈⓇ											
OpenCitations CCC	ⓈⓇ	ⓈⓇ			Ⓢ			Ⓡ	Ⓢ							
OpenCitations COCI	ⓈⓇ	ⓈⓇ						Ⓡ	Ⓢ							
OpenCitations CROCI	ⓈⓇ	ⓈⓇ						Ⓡ	Ⓢ							
OpenCitations Corpus	Ⓡ	ⓈⓇ						Ⓡ								
OpenCitations Indexes unifying	ⓈⓇ	ⓈⓇ						Ⓡ	Ⓢ							
OpenAlex	ⓈⓇ	ⓈⓇ		ⓈⓇ	ⓈⓇ											ⓈⓇ
Unpaywall		ⓈⓇ														
ORCID		ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ						ⓈⓇ				
PLOS Search API		ⓈⓇ		Ⓢ												
PLOS Text Data Mining		ⓈⓇ														
Springer Metadata API		ⓈⓇ														
Springer Open Access API		ⓈⓇ														
DBLP		Ⓡ														

wider range of identifiers-related search fields are *ORCID*, *Scopus Papers*, *PlumX*, *Dimensions*, *Embase*, and *Altmetric*. A special mention should be done on the low coverage of a search feature to look for a document by the open citation identifier (OCI) in the rest of the commercial and free APIs.

Regarding the content, the coverage of identifiers is similar to the search category. The most common retrieved identifiers are the DOI, PMCID and PubMedID. *Scopus Papers*, *Dimensions*, *Embase*, and *ORCID* are the APIs that may return several identifiers for the same document.

4.2.6. Metrics

Another important category of content that can be retrieved is the metrics that we introduced in this section. These APIs are fundamental for understanding the dissemination and usage of the research work, as well as their social impact of them. It should be highlighted that the metrics dimension is available only for the content category.

Thus, as shown in Table 8, the *citations* are the most common metric in all APIs. Nonetheless, the coverage of additional indicators, such as the h-index, impact factor, altmetrics, JCR, SJR, CIRC, and CiteScore, is reduced in most of the APIs. In fact, focusing on the best APIs to retrieve metrics, we should highlight *Dimensions*, *InCites*, *Journals*, *Scopus Serial Title Metadata*, *CIRC*, and *SciVal*.

Table 8
Metrics.Ⓢ: means that the user can search for that feature.Ⓡ: means that the user can retrieve that feature.

API	Citations	H-Index	Impact factor	Altmetrics	JCR	SJR	CIRC	CiteScore
Semantic Scholar	Ⓡ							
Lens	Ⓡ							
Incites	Ⓡ		Ⓡ					
Journals	Ⓡ		Ⓡ		Ⓡ			
Publons								
Web of Science Expanded	Ⓡ							
Web of Science Starter	Ⓡ							
BiorXiv	Ⓡ							
ArXiv								
Altmetrics				Ⓡ				
Dimensions	Ⓡ			Ⓡ				
CIRC					Ⓡ	Ⓡ	Ⓡ	
Embase								
ScienceDirect Search								
ScienceDirect Article Metadata								
ScienceDirect Article (Full Text) Retrieval								
ScienceDirect Nonserial Title Metadata								
SciVal	Ⓡ	Ⓡ	Ⓡ					
Scopus Affiliations								
Scopus Authors	Ⓡ	Ⓡ						
Scopus Papers	Ⓡ							
Scopus Citations Count Metadata	Ⓡ							
Scopus Citations Overview	Ⓡ	Ⓡ						
Scopus Serial Title Metadata	Ⓡ		Ⓡ			Ⓡ		Ⓡ
Scopus PlumX Metrics	Ⓡ			Ⓡ				
Europe PMC	Ⓡ							
IEEE Xplore Metadata	Ⓡ							
IEEE Xplore Open Access								
CrossRef								
iCite	Ⓡ							
NCBI PMC	Ⓡ							
OpenCitations CCC	Ⓡ							
OpenCitations COCI	Ⓡ							
OpenCitations CROCI	Ⓡ							
OpenCitations Corpus	Ⓡ							
OpenCitations Indexes unifying	Ⓡ							
OpenAlex	Ⓡ							
Unpaywall								
ORCID								
PLOS Search API								
PLOS Text Data Mining								
Springer Metadata API								
Springer Open Access API								
DBLP								

Also, regarding alternative metrics, the best APIs are *Altmetric* and *Scopus PlumX*. A special mention should be made of the retrieval of the JCR and the SJR impact factors. The former is available at the *Journals* API of Clarivate, and the latter is found only in the *Scopus Serial Title Metadata*. We must also highlight that *CIRC* is an interesting API since it offers an aggregation of *JCR* and *SJR* and provides the *CIRC* indicator.

4.2.7. Other information

Apart from all search fields previously mentioned, in Table 9, we show a set of fields that we did not classify in any dimension, as its nature is very diverse. In that sense, most of the APIs allow searching by document type and research areas. The research area codes may vary from one API to another. For the sake of the example, the subject areas (Scopus), categories (Web of Science), and concepts (OpenAlex) are different, and they respond to a different classification. The document types and research area features can be used to delimit the search to some document types in a specific area (e.g., delimit the search to only return articles and not reviews belonging to the chemistry subject area).

Regarding the search by the author's keywords, we must highlight *Lens*, *Web of Science Expanded* and *Starter* APIs, *Embase*, *Scopus Papers*, *Springer Metadata, and Open Access*. Also, some APIs provide an open access search field (i.e., Lens, Dimensions, Science Direct Search and Science Direct Article Metadata, Scopus Papers, *Europe PMC*, *OpenAlex*, *Unpaywall*, *Springer Metadata* and *Springer Open Access*).

Table 9
Other.Ⓢ: means that the user can search for that feature.Ⓡ: means that the user can retrieve that feature.

API	Document type	Keywords of authors	Research field	Open Access	References	Clinical trials ID	CAS Registry number	Mesh terms	Full text	Citing papers
Semantic Scholar			Ⓡ	Ⓡ	Ⓡ					Ⓡ
Lens	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ		Ⓡ
Incites	Ⓡ			Ⓡ						
Journals			ⓈⓇ							
Publons										
Web of Science Expanded	ⓈⓇ	ⓈⓇ	ⓈⓇ		ⓈⓇ		Ⓡ	ⓈⓇ		Ⓡ
Web of Science Starter	ⓈⓇ	ⓈⓇ								
BiorXiv			Ⓡ	Ⓡ						
ArXiv			ⓈⓇ	Ⓡ						
Altmetrics	Ⓡ		Ⓡ	Ⓡ						
Dimensions	ⓈⓇ		ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ		ⓈⓇ		Ⓡ
CIRC										
Embase	ⓈⓇ	ⓈⓇ				Ⓢ	ⓈⓇ			
ScienceDirect Search				ⓈⓇ						
ScienceDirect Article Metadata	Ⓢ	Ⓡ		ⓈⓇ					Ⓡ	
ScienceDirect Article (Full Text) Retrieval	Ⓡ			Ⓡ					Ⓡ	
ScienceDirect Nonserial Title Metadata	Ⓡ		Ⓢ	Ⓡ	Ⓡ				Ⓡ	
SciVal	Ⓢ		Ⓢ							
Scopus Affiliations	Ⓡ									
Scopus Authors	Ⓡ		Ⓢ							
Scopus Papers	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ					Ⓡ
Scopus Citations Count Metadata										
Scopus Citations Overview	Ⓢ									
Scopus Serial Title Metadata	Ⓡ		ⓈⓇ	ⓈⓇ						
Scopus PlumX Metrics										
Europe PMC	ⓈⓇ		Ⓡ	ⓈⓇ	ⓈⓇ	Ⓢ	Ⓢ	ⓈⓇ	Ⓡ	
IEEE Xplore Metadata	ⓈⓇ		Ⓡ	Ⓡ						
IEEE Xplore Open Access									Ⓡ	
CrossRef	ⓈⓇ		Ⓢ		Ⓡ	ⓈⓇ				Ⓡ
iCite					Ⓡ					Ⓡ
NCBI PMC	ⓈⓇ		Ⓡ	Ⓡ	Ⓡ	Ⓢ	Ⓢ	ⓈⓇ	Ⓡ	
OpenCitations CCC					Ⓡ					Ⓡ
OpenCitations COCI					Ⓡ					Ⓡ
OpenCitations CROCI					Ⓡ					Ⓡ
OpenCitations Corpus					Ⓡ					Ⓡ
OpenCitations Indexes unifying					Ⓡ					Ⓡ
OpenAlex	ⓈⓇ		ⓈⓇ	ⓈⓇ	ⓈⓇ			Ⓡ		Ⓡ
Unpaywall	Ⓡ			ⓈⓇ						
ORCID										
PLOS Search API	ⓈⓇ		ⓈⓇ		Ⓢ					
PLOS Text Data Mining	Ⓡ		Ⓡ	Ⓡ	Ⓡ				Ⓡ	
Springer Metadata API	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ						
Springer Open Access API	ⓈⓇ	ⓈⓇ	ⓈⓇ	ⓈⓇ	Ⓡ				Ⓡ	
DBLP	Ⓡ									

In addition, regarding important fields to delimit the search in health-related disciplines such as Medicine or Nursery, the APIs analyzed present the clinical trials id, and the MeSH terms. The former search feature is only present in Lens, Dimensions, Embase, *Europe PMC*, *CrossRef*, and *NCBI PMC*; meanwhile, the latter search feature is solely found in *Lens*, *Web of Science Expanded*, *Dimensions*, *EuropePMC* and *NCBI PMC*. It is worth saying that EMBASE provides the search by Emtree, which gives additional features to the MeSH term thesaurus. Another important feature of the search is the possibility to look for documents by their references, allowing us to find the cited items of a concrete document. This is only possible with particular APIs, such as *Lens*, *Web of Science Expanded*, *Dimensions*, *Scopus Papers*, *Europe PMC*, *OpenAlex*, and *PLOS Search API*.

Regarding the features returned in this dimension, the most common are the *document type*, *open access*, *references*, and *citing papers*. However, we must highlight that access to references and citing papers require additional calls to the APIs, which reduces performance and access costs, since they sometimes require a commercial license, as in the *Web of Science Expanded*. In that sense, *Dimensions*, *OpenCitations*, *Semantic Scholar*, *Lens*, and *OpenAlex* offer this information without a cost. The research area is also returned by most of the APIs, although they might not be equivalent between APIs. Regarding specific and important content, we must mention the CAS Registry number (only available in Lens, Web of Science Expanded and Embase), the clinical trials (available in *Lens*, *Dimensions* and *CrossRef*), and the MeSH terms (available in *Lens*, *Web of Science Expanded*, *Dimensions*, *Europe*

Table 10
Query modes.

API	Resource based	Free text	Structured query
Semantic Scholar	✓	✓	
Lens			✓
Incites	✓		
Journals	✓	✓	
Publons	✓		
Web of Science Expanded	✓		✓
Web of Science Starter	✓		✓
BiorXiv	✓		
ArXiv			✓
Altmetrics	✓		
Dimensions	✓		✓
CIRC	✓		
Embase	✓		✓
ScienceDirect Search			✓
ScienceDirect Article Metadata			✓
ScienceDirect Article (Full Text) Retrieval	✓		
ScienceDirect nserial Title Metadata	✓		
SciVal	✓		✓
Scopus Affiliations	✓		✓
Scopus Authors	✓		✓
Scopus Papers	✓		✓
Scopus Citations Count Metadata	✓		
Scopus Citations Overview	✓		
Scopus Serial Title Metadata	✓		
Scopus PlumX Metrics	✓		
Europe PMC	✓		✓
IEEE Xplore Metadata	✓		✓
IEEE Xplore Open Access	✓		
CrossRef	✓		
iCite	✓		
NCBI PMC			✓
OpenCitations CCC	✓		
OpenCitations COCI	✓		
OpenCitations CROCI	✓		
OpenCitations Corpus	✓		
OpenCitations Indexes unifying	✓		
OpenAlex	✓	✓	✓
Unpaywall	✓		✓
ORCID	✓		✓
PLOS Search API			✓
PLOS Text Data Mining	✓		
Springer Metadata API			✓
Springer Open Access API			✓
DBLP		✓	

PMC, *NCBI PMC*and *OpenAlex*). Finally, the full-text content is returned by *ScienceDirect Article Metadata*, *Science Direct Article (Full Text) Retrieval*, *Science Direct Nonserial Title Metadata* (this is mainly about books), *Europe PMC*, *NCBI PMC*, *PLOS Text Data Mining* and *Springer Open Access*.

4.3. Query modes category

The last analysis of the general category is the query types. In that sense, Table 10, most APIs follow a hybrid approach between a language to search in the API, and then, once the results are obtained, several resource-based endpoints (operations) are used to retrieve additional metadata about the records obtained in the first place. In this sense, we will highlight the most important APIs in terms of query type expressiveness, conciseness, and possibilities. *Dimensions* has the most powerful query type, since its search language allows grouping, wildcards, and complex conditionals, as well as handling multi-value metadata. *Web of Science Expanded*, *Scopus*, and *ScienceDirect Search* enable boolean query search that allows high expressiveness and conciseness; however, when compared to *Dimensions*, the user cannot select the fields that he/her wants to retrieve or group by some field to perform a sum or a count. However, *Scopus* and*ScienceDirect Search* do not allow grouping or selection operators. We should also remark that some APIs have a protocol (OAI-PMH) for retrieving data such as Europe PMC and NCBI PMC (entrez utilities). The rest of the APIs with query languages also use boolean searches without additions such as Dimensions, so in most of them a user can make a structured query (i.e., to query for one or more different fields with different boolean operators). In contrast, other APIs, such as DBLP or Semantic Scholar, use a free text query. Moreover, regarding the resource-based APIs, they are mainly single-purpose ones, and they are metrics-related APIs.

Table 11
Bibliometric analyses.

APIs	Science mapping analysis							Bibliometric evaluation							Other		
	Co-words		Co-citation			Co-author		Article		Journal							
	Author keywords	Generated keywords	References	Authors	Journals	Names	Identifiers	Citations	Altmetrics	SJR	JCR	CIRC	CiteScore	Impact factor	Leadership	Collaboration	Economic
Semantic Scholar			✓	✓	✓	✓		✓									
Lens	✓		✓	✓	✓	✓		✓								✓	
Incites								✓						✓			
Journals								✓			✓			✓			
Publons						✓	✓									✓	
Web of Science Expanded	✓	✓	*	*	*	✓	✓	✓							✓	✓	
Web of Science Starter	✓					✓	✓	✓									
BiorXiv						✓		✓									
ArXiv						✓											
Altmetrics						✓			✓								
Dimensions		✓	✓	*	*	✓	✓	✓	✓						✓	✓	✓
CIRC										✓	✓	✓					
Embase	✓					✓										✓	
ScienceDirect Search						✓											
ScienceDirect Article Metadata	✓					✓											
ScienceDirect Article (Full Text) Retrieval	✓			✓	✓	✓									✓		
ScienceDirect Nonserial Title Metadata			✓			✓											
SciVal								✓									
Scopus Affiliations						✓	✓										
Scopus Authors						✓	✓	✓								✓	
Scopus Papers	✓	✓	†	†	†	✓	✓	✓							✓	✓	
Scopus Citations Count Metadata								✓									
Scopus Citations Overview						✓	✓	✓									
Scopus Serial Title Metadata								✓		✓			✓				
Scopus PlumX Metrics								✓	✓								
Europe PMC			✓	‡	✓	✓		✓								✓	
IEEE Xplore Metadata	✓	✓				✓	✓	✓									
IEEE Xplore Open Access																	
CrossRef			✓	*	✓	✓										✓	
iCite			✓	✓	✓	✓		✓									
NCBI PMC			✓			✓		✓							‡	✓	
OpenCitations CCC			*	* ‡	* ‡	✓		✓									
OpenCitations COCI			*	* ‡	* ‡	✓		✓									
OpenCitations CROCI			*	* ‡	* ‡	✓		✓									
OpenCitations Corpus			✓			✓		✓									
OpenCitations Indexes unifying			*	* ‡	* ‡	✓		✓									
OpenAlex		✓	✓	*	*	✓	✓	✓								✓	
Unpaywall						✓											
ORCID						✓	✓									✓	
PLOS Search API						✓											
PLOS Text Data Mining			✓	‡	✓	✓									✓	✓	
Springer Metadata API	✓					✓											
Springer Open Access API	✓		✓	✓	✓	✓									✓	✓	
DBLP						✓	✓										

∗: requires additional steps to download the references. †: requires FULL or REF view download (check Appendix B for more information). ‡: it returns raw data, and requires additional processing first.

In addition, these APIs usually offer a richer set of content types. Nonetheless, to retrieve the whole metadata about an object, a user might need to perform several API calls. This negatively affects the request constraints. In concrete, the most complete views are returned in *Dimensions*, *Scopus* APIs, *OpenAlex*, and *CrossRef*. The rest of the APIs may require additional calls to retrieve additional information.

For more information on how to query the APIs, the Appendix B gives a detailed view of the usage of the APIs.

5. Bibliometric analyses

Another perspective of the paper is to shed light on the bibliometric analyses that can be performed with the data retrieved from each API. To do this, we analyzed the content category and check if the analyses are workable. Concretely, we focus on science mapping analysis, bibliometric evaluation, and other types of analysis.

As shown in Table 11, the science mapping analysis can be performed using mainly three units of analysis (Batagelj & Cerinšek, 2013; Cobo et al., 2011b). First, an analysis based on co-words can be performed using the author keywords or generated keywords. However, the most common analysis that can be done is a co-words analysis based on the author's keywords. Then, the co-citation analysis can be based on the references, the authors of the references, and journals in which the references of an article are published. This type of science mapping analysis is workable in almost all the APIs. However, the co-citation based on authors and journals may require additional steps to retrieve the necessary data. Subsequently, the co-author science mapping analysis can be based on the names of the authors or their unique identifiers. The analysis based on author's names can be performed using almost all the APIs, meanwhile the based on identifiers is less workable and only possible with the well-known APIs (e.g., Dimensions, Scopus, Web of Science, among others). It is important to highlight that some APIs may require additional retrieval steps to get the necessary data (marked with a ∗), or additional processing to perform the analysis (marked with a ‡).

Regarding the bibliometric evaluation, we focus on two levels: article and journal level. Concerning the article level, an evaluation based on citations is available in almost all the APIs. Moreover, the altmetrics-based evaluation is only possible in Altmetrics,

Table 12
Merge documents retrieved from APIs using several identifiers.

Identifier	Combination options
DOI	All APIs, except NCBI PMC.
PubMed	Lens, Web of Science Expanded, Dimensions, Embase, Science Direct Article (Full-Text Retrieval), Scopus Papers, iCite, NCBI PMC, OpenAlex, and ORCID APIs.
PMCID	EuropePMC, NCBI PMC, OpenAlex, and ORCID.
OpenCitations ID	OpenCitations APIs (i.e., CCC, COCI, CROCI, Corpus and Indexes Unifying).
EID	Science Direct Article (Full Text), Science Direct Article Metadata, Scopus Papers, and ORCID.
WOS ID	Web of Science Expanded and Starter APIs, and ORCID.
Arxiv ID	ArXiv, Dimensions, ORCID.
PII	Embase, ScienceDirect Search, and ScienceDirect Article Metadata.

Dimensions, and Scopus PlumX. In addition, the journal-level evaluation considers the SJR, JCR, CIRC, CiteScore, and the impact factor of the journals. However, these indicators are solely in specific APIs: CIRC has the SJR, JCR, and CIRC indicators for the journals; JCR is also in the Journals API, and SJR and CiteScore are in the Scopus Serial Title Metadata.

Finally, in the other types of bibliometric analysis, we consider leadership, collaboration, and economic analyses. The leadership analysis is workable in the well-known APIs (e.g., Web of Science Expanded, Dimensions, Scopus Papers, among others). Also, collaboration analysis is possible in almost all the APIs. The economic analysis is only possible with Dimensions.

6. Interoperability of APIs

In this section, we analyze the interoperability of the APIs from two points of view: (i) the merge of documents retrieved using the APIs, (ii) and how the metadata of the documents retrieved from an API can be extended using other APIs.

6.1. Merging collections of documents

In this section, we study the merging of collections of documents using their retrieved identifiers. As shown in Table 12, most of the APIs allow merging data using the DOI identifier, except for NCBI PMC. Also, the PubMedID can be used to merge the data of the most popular APIs (e.g., Lens, Web of Science Expanded, Dimensions, Scopus Papers). Regarding the PMCID, it is assigned to the full-text articles in PubMed Central, and only EuropePMC, NCBI PMC, OpenAlex, and ORCID documents can be combined using this identifier. The Open Citations Identifier can be used to merge the documents retrieved from the different OpenCitations APIs. Also, the Elsevier identifier (EID) can be used with ScienceDirect, Scopus, and ORCID. The Web of Science identifier can only be useful to merge the documents from the APIs of the Web of Science and ORCID. The ArXiv identifier is useful to merge Dimensions, ORCID, and arXiv documents. Finally, the Publication Item Identifier (PII), can only be used to merge documents retrieved from some Elsevier APIs (Embase and some ScienceDirect APIs).

As it is described, the most common identifiers for merging the documents retrieved from APIs are the DOI and the PubMed identifiers.

6.2. Extension of metadata of documents

The dataset or corpus extension is important in several use cases, in which we can highlight extending the metadata with different metrics, such as citations from different databases, altmetrics, etc. Another possibility can be to complete some metadata that is in one database but not in another. Therefore, we gathered and studied the options for the extension of these metadata using the main identifiers in the APIs.

In Fig. 3, the available extension options are shown. Thus, on the left side, there are the source APIs, so that using the identifier of the central part, one can extend the metadata using the target APIs on the right side of the figure. For instance, if we have a document of Scopus Papers, we can use the DOI to get the altmetrics of the document. Hence, in view of Fig. 3, the DOI, PubMedID, and the PMCID are the identifiers that allow most extensions in the APIs. Other identifiers cover a good number of APIs but are internal to some companies (i.e., WOSID for Clarivate-related APIs, EID, PII, PUI and LUI for Elsevier-related APIs). Regarding the APIs that are more extensible, we must highlight Semantic Scholar, Dimensions, ORCID and Embase.

7. Discussion

Bibliometric and science mapping analysis have become important approaches for researchers and practitioners, being used beyond their traditional research area. In fact, according to González-Alcaide (2021), the number of bibliometric analysis papers has grown uncontrollably. This rapid growth and the great interest of the entire scientific community, and even in science policy management, has been possible, in part, thanks to bibliometric software tools (Cobo et al., 2011b; Moral-Muñoz et al., 2020) and access to bibliographic databases. From a traditional perspective, data for bibliometric analysis are gathered from web platforms, such as the Web of Science or Scopus, among others. Nevertheless, due to the restriction of the platform, it is hard or impossible to get a large corpus in that way (Baas et al., 2020; Gusenbauer & Haddaway, 2020; Hendricks et al., 2020; Martín-Martín et al., 2018;

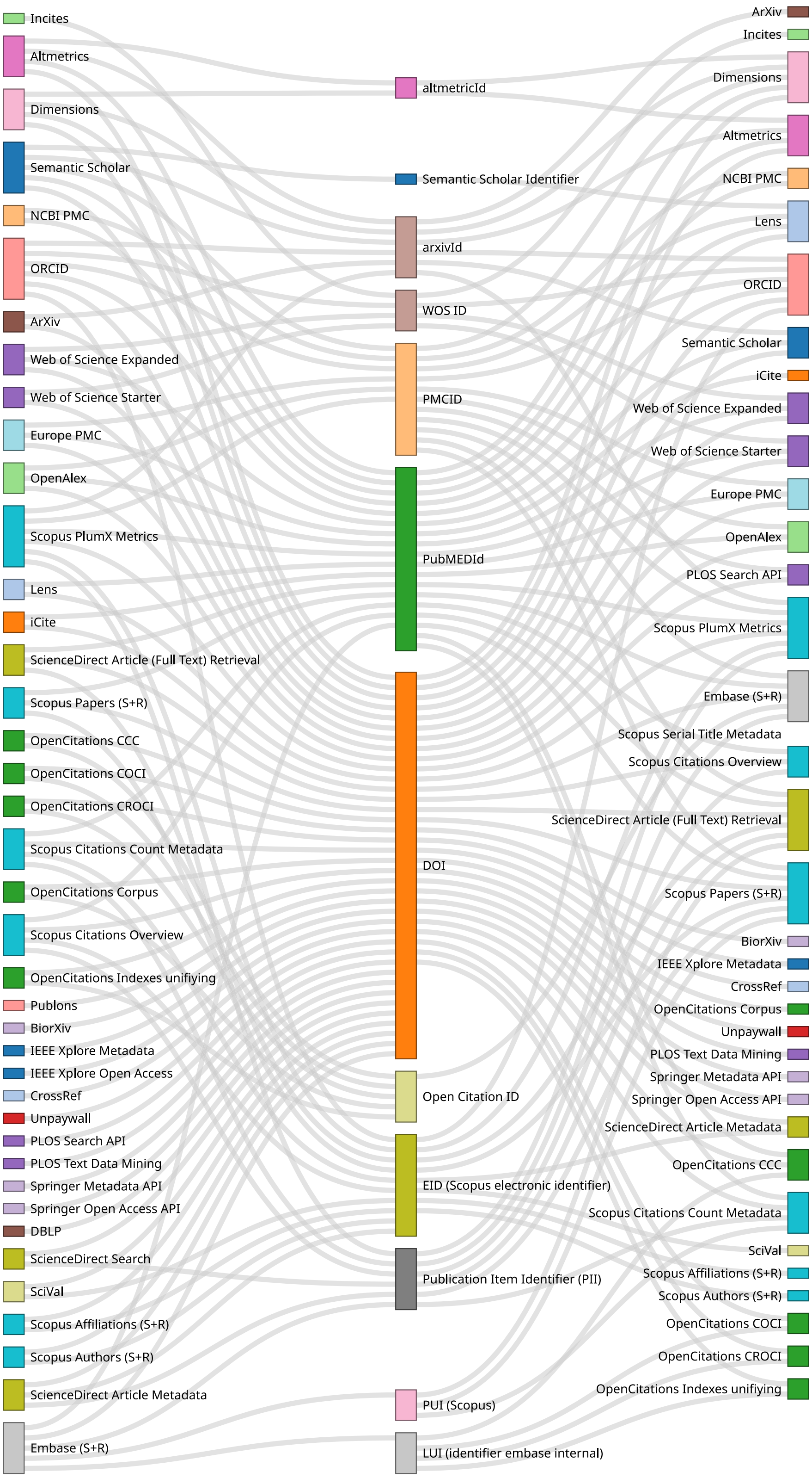

Fig. 3. Diagram of extensions through identifiers.

Orduña-Malea & Delgado López-Cózar, 2018; Peroni & Shotton, 2020; Singh et al., 2021). In that way, using APIs in bibliometric and science mapping analysis can help researchers gain deeper insights into research trends, improve their understanding of their field, and make more informed decisions about their research, since they allow us to access to large amounts of data (Lomborg & Bechmann, 2014), the automation, integration with other software tools (Perrotta, Gulson, Williamson, & Witzenberger, 2021), and standardization of data formats (Borgogno & Colangelo, 2019), making it easier to compare and integrate data from different sources. Therefore, in view of the results obtained in this study, some aspects related to the characteristics of the available APIs, the possibilities for using the information in the analyses and their interoperability will be discussed.

Regarding the results of the comparative analysis, first, we should highlight that most of the APIs offer the most common formats of metadata, JSON, and XML. Also, the most complete APIs in terms of features are the ones that require a commercial subscription (i.e. Web of Science, Scopus, etc.). In fact, commercial databases used to be those with better accuracy in their data, with less error and with the data (e.g. author's name, affiliations, etc.) more normalized or standardized (Sun, Kaur, Possamai, & Menczer, 2013), but in contrast, they used to have more limited access to their APIs (i.e. the number of petitions). Also, commercial databases are usually provided by an institutional account, which could cause researchers or universities with lower income cannot to access that. Moreover, there are a great variety of open or free databases that provide bulk access to their data, availability without a subscription, and sometimes a large coverage. However, it is important to note that open databases may contain more errors in their data, as cleaning and pre-processing tons of scientific data requires a great deal of effort that non-profit organizations sometimes cannot afford.

Conversely, open databases APIs could improve reproducibility (Stodden, Seiler, & Ma, 2018), a key challenge, aiming to provide all the resources to the research community, allowing them to reproduce the experiment with exactly the same results. Some of them, such as OpenAlex, provide different snapshots of their data, facilitating access to the same portion of the data used in an analysis. In that sense, although bibliometric papers usually provide the query used to retrieve the data, without access to the provider, it is impossible to reproduce the study. Also, data are frequently updated in the databases, and, currently, it is impossible to obtain the same portion of the data.

Concerning the use of the obtained metadata to determine the social component of the research, ORCID, Scopus Authors, and Publons are the most complete APIs for authors-based analyses, while Dimensions, Scopus Affiliations, and OpenAlex are the most complete for institutions-based analyses. These metadata have been commonly used to uncover invisible colleagues through social network analysis (Barabási, Jeong, Néda, Ravasz, Schubert et al., 2002). In recent literature, these data have also been used to determine the international collaboration (Sugimoto, Robinson-Garcia, Murray, Yegros-Yegros, & Lariviere, 2017; Velez-Estevez et al., 2022), and how it affects the citation patterns. Moreover, access to funding data (Zhao, 2010) could help relate bibliometric analysis with economic data, analyzing the science from another perspective. In that sense, Dimensions is the API that covers most of the features related to funding metadata. Also, regarding the identifiers, the most common are DOI, PubMedID, and PMCID, and with them, APIs could be used cooperatively.

As was stated, the metadata provided by bibliographic databases are the main input for the bibliometric analysis. It should be noted that nearly all the APIs offer the bibliographical metadata of the documents, except those that delved only into metrics support, such as Altmetrics, InCites, SciVal, or PlumX. Metrics, whether citation-based or social media-based (altmetrics), are a crucial aspect of any bibliometric study, which could enrich them by providing more information and useful insights. Usually, citation count is provided by most APIs, but more advanced metrics, based on citations at the journal or article level, collaboration, leadership, or based on altmetrics are provided by specific APIs, such as Altmetrics, CIRC, Journals, etc.

The development of bibliometrics and science mapping analysis has led to the emergence of advanced analyses that require specific metadata. Very few APIs provide other important information for specific questions, such as the Mesh Terms, CAS Registry Number, full-text, references, or the citing papers of an article, limiting the analysis option. Therefore, APIs providers could make an effort to support these data, important in those advanced analyses, such as the detection of translational science (Padilla-Cabello, Santisteban-Espejo, Heradio, Cobo, Martin-Piedra et al., 2022), in which Mesh Terms are widely used, or detection of citation context which need the full text (Jebari, Herrera-Viedma, & Cobo, 2021).

8. Concluding remark

This manuscript presents a global comparative analysis of the most important bibliometric and bibliographic APIs, achieving its sub-objectives of identifying the APIs, comparing their characteristics and metadata, studying the possible bibliometric analyses, and analyzing the interoperability of the APIs. We collected a total of 44 APIs from various organizations, including non-profit and commercial entities. Our results show that *Clarivate Analytics* and *Elsevier* offer highly versatile APIs, while non-profit organizations such as *OpenCitations* and *OurResearch* promote the Open Science philosophy (Miedema, 2022). Additionally, almost all organizations offer free access to APIs, for non-commercial purposes, but some of them have limitations on the retrieval of the metadata. Nonetheless, there are APIs such as CrossRef, OpenCitations, or OpenAlex that have no restrictions on the purpose or over the metadata retrieval.

For bibliographic information, most APIs cover basic search features, such as author name, title, abstract, date, journal, ISSN, and ISBN. For authors, ORCID and internal author identifiers are widely supported. In the affiliation dimension, searching by organization name and country is most common, while searching by affiliation-related identifiers is less common. Searching by GRID ID is only available in Lens, Dimensions, Europe PMC, Open Alex and ORCID, and searching by the new unique affiliation identifier, the ROR ID, is only available in Dimensions, CrossRef, OpenAlex and ORCID. Most APIs return the organization name and country as content categories, while organization type is returned by *Dimensions* and *OpenAlex*. Regarding bibliometric analyses, co-author analysis using author names and bibliometric evaluation using citations are the types of analyses than can be done with the data provided by most APIs. It should also be noted that a large number of them allows to perform the other types of analyses. Furthermore, none of the identified API enables all types of analyses, but *Web of Science Expanded*, *Scopus Papers*, and *Dimensions* allow for most of them. For interoperability, DOI, PubMedID, and PMCID are the most versatile identifiers for extending metadata in the APIs. *Semantic Scholar, Dimensions*, *ORCID*, and *Embase* are the APIs that offer the most extensibility.

As a global claim, our comparison allows us to understand that APIs are usually dedicated to specific purposes, and there is no single one that gathers all the information to perform any kind of bibliometric analysis. Therefore, future bibliographic APIs should take into account not only to offer quality in a specific area (e.g., authors, addresses, or economic information), but also to focus on the quality of the relationships between the information serving different purposes. It is worth noting that, although there are APIs that can cover most of the analysis needs, the combination of two or more APIs may be the most appropriate option in order to cover as much information as possible and to enrich the reports and analyses that can be performed.

The results of our global comparative analysis of bibliometric and bibliographic APIs provide insights into the theoretical, practical, and methodological implications of using APIs in bibliometric studies. In terms of API theory, our findings highlight the strengths and weaknesses of existing bibliometric APIs, providing a global perspective of current API development. This knowledge can guide future API development efforts to either complement existing APIs or create new ones that incorporate desirable features. In terms of API practice, our study provides analysts with information about what metadata and metrics each API returns. This information can help researchers select the most appropriate API to retrieve the information they need for their studies. Finally, in terms of API methodology, the comparison of APIs presented in this paper, combined with an understanding of the typical workflow of bibliometric or scientometric analyses, can guide researchers in selecting the appropriate APIs to integrate into their methodologies and how they can affect the results. Overall, our study contributes to advancing the understanding and use of bibliometric APIs in research practice.

In view of the results obtained and the considerations made about the available APIs, some future work is proposed that would serve to enrich the use of these tools. As mentioned above, since there is no API with all the metadata, a novel API integrating all the characteristics would be desirable. However, the technical development of such an API is a challenging task, since it is almost impossible to make it perform well in all areas, since some information is protected by commercial regulations and cannot be part of a hypothetical single source of truth. Therefore, a good approach to tackle this problem might be to develop a Domain Specific Language (DSL) that allows to search, retrieve and combine bibliographic information from different bibliographic sources. Given the DSL and an engine that executes the definitions built with it, analysts would have a single front-end to science that facilitates the process of having an overview of a specific set of papers, authors, institutions, etc., by combining the information and analyzing multiple dimensions from the growing set of bibliographic sources. Other future works could focus on the accuracy of open versus commercial bibliographic databases, and how this affects the results of bibliometric analysis, or the development of techniques to improve the data quality and standardization of open databases, to make them more accessible and useful. On the other hand, researchers may also focus on the implications of the differences detected, such as the impact of commercial restrictions on access to bibliometric APIs and how this might affect the ability of researchers and institutions with lower incomes to carry out bibliometric analysis, or research on the interoperability of different bibliometric APIs, and how this might affect the ability to carry out a cross-database bibliometric analysis.

Data availability

No data was used for the research described in the article.

Acknowledgment

This paper has been supported by the Spanish State Research Agency through project PID2019-105381GA-I00/AEI/10.13039/501100011033 (iScience).

Appendix A. Summary of the included APIs

Table A.13, present a classification of different selected APIs. We can observe that several organizations, entities, or companies have developed up to 44 different APIs (composed of one or more endpoints) to retrieve the complete information from the databases automatically.

In the following, we will give a brief description of the APIs presented in Table A.13 to introduce and put them in context:

- *Semantic Scholar* (Arum, 2016; Fricke, 2018; Jones, 2015) was launched in 2015 for the Allen Institute for AI. It is a tool that uses artificial intelligence to search publications and their corresponding authors. Currently, it offers more than 196M documents from all the fields of science.
- *Lens* (Penfold, 2020), by Cambia, is another database that started back in 1999 with the name of Patent Lens. Now, Lens is an extended version that also serves over 200M of scholarly records, using a REST API, gathered from Microsoft Academic, Pubmed, and CrossRef and feeds that metadata with ORCID. As announced, MAG was discontinued at the end of 2021, so *Lens* took a dual strategy to replace MAG (Staff, 2021): the community engagement through Collective Action; and The Lens MetaRecord. The former supports the database through open content initiatives or collaboration. The latter is managing the complexities on the record variability by merging content sources and contextual metadata relevant to the original record, so they can flexibly ingest any open data (e.g., DataCite).

Table A.13
APIs classified by organization.

Allen institute for A.I.	
Semantic Scholar	Arum (2016), Fricke (2018), Jones (2015)
Cambia	
Lens	Penfold (2020)
Clarivate Analytics	
InCites	Gray and Price (2020), Panczyk et al. (2015)
Journals	
Publons	Ortega (2017), Smith (2016)
Web of Science Expanded	Analytics (2017)
Web of Science Starter	
Cold Spring Harbor Lab	
BiorXiv	Abdill and Blekhman (2019), Fraser et al. (2020)
Cornell University	
ArXiv	Ginsparg (2011)
Digital Science	
Altmetrics	Adie and Roe (2013), Ortega (2020)
Dimensions	Hook et al. (2018), Thelwall (2018)
EC3Metrics S.L.	
CIRC	Torres-Salinas et al. (2010)
Elsevier	
Embase	Sampson et al. (2003)
ScienceDirect Search	Hunter (1998)
ScienceDirect Article Metadata	Hunter (1998)
ScienceDirect Article (Full Text) Retrieval	Hunter (1998)
ScienceDirect Nonserial Title Metadata	Hunter (1998)
Scival	Dresbeck (2015)
Scopus Affiliations	Burnham (2006)
Scopus Authors	Burnham (2006)
Scopus Papers	Burnham (2006)
Scopus Citations Count Metadata	Burnham (2006)
Scopus Citations Overview	Burnham (2006)
Scopus Serial Title Metadata	Burnham (2006)
Scopus PlumX Metrics	Champieux (2015)
EMBL's European Bioinformatics Institute (EMBL-EBI)	
Europe PMC	Ferguson et al. (2021), The Europe P.M.C. Consortium (2014)
IEEE	
IEEE Xplore Metadata	Griffin (2002)
IEEE Xplore Open Access	Griffin (2002)
International DOI Fund.	
CrossRef	Pentz (2001), van Eck and Waltman (2022)
NIH	
iCite	Hutchins et al. (2019a, 2019b, 2017, 2016)
NCBI PMC	Sayers et al. (2016)
OpenCitations	
OpenCitations CCC	
OpenCitations COCI	Heibi et al. (2019b)
OpenCitations CROCI	Heibi et al. (2019a)
OpenCitations Corpus	Peroni et al. (2017)
OpenCitations Indexes Unifying	
ORCID Inc.	
ORCID	Haak et al. (2012)
OurResearch	
OpenAlex	Singh Chawla (2022)
Unpaywall	Else (2018)

(*continued on next page*)

Table A.13 (*continued*).

PLOS	
PLOS Search API	Federer et al. (2018), Savage and Vickers (2009)
PLOS Text Data Mining	Federer et al. (2018), Savage and Vickers (2009)
Springer	
Springer Metadata API	Bekkari (2012)
Springer Open Access API	Bekkari (2012)
University of Trier	
DBLP	Ley (2002)

- Also, we considered the well-known products of Clarivate Analytics (from now on will be referred as Clarivate), which has been serving science from 1864 with the creation of the Zoological Record. Nowadays, Clarivate offers several APIs dealing with bibliographical records such as publications, metrics, journals, and authors, as shown in Table A.13:
  - The *InCites* (Gray & Price, 2020; Panczyk et al., 2015) product is a tool used to benchmark and analyze science, and its API enables the retrieval of metrics of scholarly works.
  - The *Journals* API offers access to Journal Citation Reports (JCR) which provides metrics about journals.
  - *Publons* (Ortega, 2017; Smith, 2016) started in 2012 intending to verify and show peer reviews and contributions to academic journals with a unique identifier, the ResearcherID. Then, it was acquired by Clarivate in 2017.
  - *Web of Science Expanded* (Analytics, 2017) is an API that provides access to 171M records from different databases for multiple disciplines. The access to the WOS database is under a subscription, and the temporal coverage is from 1900 to the present.
  - *Web of Science Starter* is an API that provides access to the same database as the previous API, but the metadata returned is less complete. Also, there is free access for students and academics.
- *BiorXiv* (Abdill & Blekhman, 2019; Fraser et al., 2020), by Cold Spring Harbor Lab, is a preprint server for biology-related papers. It provides a free online archive and distribution site to upload unpublished preprints in life sciences.
- *ArXiv* (Ginsparg, 2011) by Cornell University serves for the same goal as *BiorXiv*, to provide a site where the unpublished works could be distributed and shared freely. It gives access to 1.9M scholarly articles in physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics.
- Digital Science was founded in 2010, since then they have launched several relevant products, that also have their corresponding API:
  - *Dimensions* (Hook et al., 2018), which is a scholarly database that integrates different open data sources such as Crossref, Altmetric, or GRID to offer a linked data structure between publications (121M), clinical trials (654K), grants (6M) and policy documents (715K).
  - *Altmetrics* (Adie & Roe, 2013; Ortega, 2020) is a database that tracks and analyze the online activity surroundings the research outputs, i.e., citing of articles in Twitter or Facebook.
- *CIRC* (Torres-Salinas et al., 2010) by EC3Metrics. *CIRC* is a classification of Humanities and Social Sciences scientific journals whose goal is to ease the building of bibliometric indicators to evaluate the research output.
- Another enterprise offering a considerable amount of bibliographic data in the form of APIs is Elsevier. Elsevier has several bibliographical content and metrics databases, (i) ScienceDirect contains full-text documents from journals and books, mainly published by Elsevier and some hosted societies. (ii) Scopus indexes metadata from papers and references of hundreds of sources, adding citation matching, and author and affiliation profiles. It indexes almost all the ScienceDirect database but without the full text of the documents. In this database, authors, affiliations, and documents are linked together. (iii) EMBASE is a medical literature database that provides an API exposing more than 32M of records indexed with data about drugs, diseases, and medical devices. They add Emtree, a taxonomy that aids in exploring biomedical concepts and terms. For each of these databases, there are one or more APIs delved into some specific kind of data or functionality:
  - *Embase* (Sampson et al., 2003) stands for Embase (Search + Retrieval). The former is used to perform a query against the database, while the latter is used to retrieve the full information of the records matching the query.
  - *ScienceDirect Search* (Hunter, 1998) is the API used to query against the ScienceDirect database with the same experience of the web-based search engine of ScienceDirect. It is recommended for federated search and free text search (Elsevier, 2018).
  - *ScienceDirect Article Metadata* (Hunter, 1998) is also used to query the ScienceDirect database but with advanced ways to search. It is recommended for information retrieval and text data mining applications (Elsevier, 2018).
  - *ScienceDirect Article (Full Text) Retrieval* (Hunter, 1998) enables the retrieval of the full text of the documents in a structured format.
  - *ScienceDirect Nonserial Title Metadata* (Hunter, 1998) allows to search among non-serial titles that Elsevier defines as a publication with an ISBN unless it is a report, part of a book series, proceeding (non-serial), or patent.

- *SciVal* (Dresbeck, 2015) allows retrieving metrics at author, country, institution, publication, Scopus sources, subject areas, topics, and at a world level (e.g., Hirsch index of an author or field weighted impact of an institution).
- *Scopus Affiliations* (Burnham, 2006), are the search and retrieval APIs of the affiliations in the Scopus database. With the former API, a search for an affiliation can be performed, and then, with the latter, they can be retrieved. In the retrieval API, one could retrieve the documents or authors linked with the affiliation searched.
- *Scopus Authors* (Burnham, 2006), are the search and retrieval APIs related to the authors in the Scopus database. Also, the documents linked with the searched author can be retrieved in the retrieval API.
- *Scopus Papers* (Burnham, 2006), are the search and retrieval APIs of the documents in the mentioned database.
- *Scopus Citations Count Metadata* (Burnham, 2006) API can be used to retrieve document citation counts of the Scopus database.
- *Scopus Citations Overview* (Burnham, 2006) API can be used to retrieve document citation counts broken down by year and also has the option of excluding self-citations.
- *Scopus Serial Title Metadata* (Burnham, 2006) is the way to retrieve information and metrics about Scopus indexed serial titles (journals, conference series, and book series).
- *Scopus PlumX Metrics* (Champieux, 2015) provides alternative metrics (altmetrics), of Scopus documents, books, and other types of content.

• The EMBL's European Bioinformatics Institute offers *Europe PMC* (Ferguson et al., 2021; The Europe P.M.C. Consortium, 2014) which offers access to abstracts, full texts through PubMed and PMC and other relevant sources (such as patent records) in a single access point.
• The Institute of Electrical and Electronics Engineers (IEEE) (Griffin, 2002) offers two APIs, giving access to literature in electrical engineering, computer science, electronics, and related disciplines.

- *IEEE Xplore Metadata* allows to query and retrieve metadata records including abstracts of IEEE Xplore.
- *IEEE Xplore Open Access* allows to query and retrieve full-text Open Access articles.

• *CrossRef* (Pentz, 2001), by the International DOI Fundation is a database that interlinks a large amount of records such as books, conference proceedings etc. *CrossRef* is also responsible of maintaining the FundRef Registry which is a way to provide unique and persistent identifiers for grant organizations.
• The National Institutes of Health (NIH), is a part of the U.S. Department of Health and Human Services. It offer two APIs:

- *iCite* is a tool to access to bibliometric measures and metadata. It has three modules: influence (Hutchins et al., 2017, 2016), translation (Hutchins et al., 2019b) and open citations (Hutchins et al., 2019a). The first provides relative citation ratio values; the second provides translation measures, which are related to the closeness of a paper to humans, molecular or animals; and the third provides citations from the NIH Open Citation Collection (NIH-OCC).
- *NCBI PMC* (Sayers et al., 2016) provides programmatic access to query the PubMed databases, which covers medical records such as abstracts, full-texts, books, and preprints.

• *OpenCitations* offers a set of APIs providing an infrastructure for open scholarship to publish open bibliographic and citation data by using the Semantic Web technologies. They offer the following APIs:

- *OpenCitations CCC* (Citations in Context Corpus) gives access to an open-access subset of XML articles hosted by Europe
- *OpenCitations COCI* (Heibi et al., 2019b) gives access to an RDF dataset containing all the citations that are specified by the open references to works present in Crossref. COCI does not index Crossref references that are not open, nor Crossref open references to entities that lack DOIs, as stated in their website.
- *OpenCitations CROCI* (Heibi et al., 2019a) is an index that contains citations deposited by individuals, identified by ORCID identifiers.
- *OpenCitations Corpus* (Peroni et al., 2017) is an API enabling access to an open repository of citation data, which provides accurate bibliographic references harvested from the scholarly literature.
- *OpenCitations Indexes Unifying* is an API that serves the content of all the indexes in OpenCitations (i.e., to the date of this paper, they are COCI & CROCI).

• *ORCID* (Haak et al., 2012) by ORCID Inc. is a database providing a unique identifier for scholars and a repository of profiles. It also provides an API giving access to metadata about the authors, like publications, funding, and employment. ORCID have two APIs, the public and the member ones. For the rest of the paper, we will use ORCID indistinguishably. These two APIs only differ in the capabilities of writing to ORCID registries and the synchronization mechanisms in the member API, but not in the search or metadata available.
• OurResearch was founded in 2011 and launched ImpactStory, that is an open-source tool to discover the online impact of the research of an author. Since then, they have created more tools to make research more open. The tools with API related to bibliographical metadata are:

- *Unpaywall* (Else, 2018) which finds Open Access content using data from CrossRef, DOAJ and monitoring more than 50K online content.

  - *OpenAlex* (Singh Chawla, 2022) which is an index of hundreds of millions of interconnected entities. With the discontinuation of MAG, it has gained significant popularity.

- *PLOS* (Federer et al., 2018; Savage & Vickers, 2009) is an open-access publisher aiming to accelerate process in science by leading a transformation in communication. They provide two APIs to access documents and full text together with the documents' metadata:
  - *PLOS Search API* enables the search of research articles over the whole corpus of PLOS.
  - *PLOS Text Data Mining* provides access to the metadata of the articles and also the fulltext of them.
- Springer (Bekkari, 2012), is a publisher of scientific content, mainly books and journals. The offer their data in two different APIs:
  - *Springer Metadata API*. It enables the access to the metadata of journal articles, book chapters, etc.
  - *Springer Open Access API* enables the access and retrieval of metadata and full-text content for documents from Springer Nature, including BioMed Central and SpringerOpen journals.
- The *Digital Bibliography & Library Project (DBLP)* (Ley, 2002) by the University of Trier, is aimed to provide open bibliographic information on computer science journals and proceedings. It counts with more than 5M publications.

Appendix B. How to query: a detailed view

In this appendix, we describe the ways to query the different APIs included in the paper. Although all the products/databases offer their APIs, they have to be queried differently. Therefore, we explain the details for querying the APIs below:

- **Semantic Scholar** allows a *query-based* search over the papers using a free text query, e.g.:
  `covid vaccination`
  Once the results are obtained, more endpoints to retrieve more data are provided. To search an author, the API is *resource-based*, as one provides an author name parameter and filters the resource by it.
- The **Lens** use a query-based boolean search. It is structured and can constraint the search to specific fields, combining them with boolean operators. Example:
  `title:Dimensions AND author.affiliation.name: (Harvard University)`
- **InCites** follows a *resource-based* search. It allows to search the specific metrics for a document filtering all the documents by the internal identifier.
- **Journals** use a free text query-based search to look for journals in the Web of Science (e.g., `nano*`; if an ISSN is provided, the ISSN is used to query). Also, several parameters can be specified to filter the results by journal impact factor, quartiles, year of JCR, among others. Once the identifier of the journal is obtained, other *resource-based* endpoints in the API provide more information about the journal, such as metrics.
- **Publons** follows a *resource-based* approach. It allows to filter the researchers resource by institution or by an unique identifier. Also, there exist an endpoint where the publications of a researcher can be obtained using this approach.
- **Web of Science Expanded** uses a *query-based* boolean search approach (e.g., `TS=(covid*)`). This first search outputs a query identifier, and it must be used with another endpoint to retrieve the results matched by the query. If the citing documents or the references are needed for a single document, two specific *resource-based* endpoints of the API must be used.
- **Web of Science Starter** uses a *query-based* boolean search approach. For instance:
  `OG=(Arizona State University)`
  Then the user can paginate over the same endpoint to retrieve the results. Moreover, the journals information endpoints use a *resource-based* approach, allowing to filter the journals by ISSN.
- **BiorXiv** allows to search using a *resource-based* approach filtering by a date interval and/or publisher.
- **ArXiv** follows a *query-based* structured search. It allows to define the search using constraints on specific fields (e.g., `au:del_maestro AND ti:checkerboard`). Then, the user can paginate over the same endpoint to retrieve the results.
- **Altmetrics** uses a *resource-based* search. It allows to filter the resources via some identifiers that have been already shown in Section 4.2.5.
- **Dimensions** has the most powerful *query-based* search engine of all the analyzed APIs. It allows to query for one resource, and return multiple resources counts of the database related to the first, for instance:
  `search publications for "malaria" return year return funders`
  Allows to retrieve the number of malaria publications supported by funding organizations broken down by year. However in some cases to achieve additional information, such as the citing documents of a document (in-going citations), more than one queries might be needed. Moreover, *Dimensions* allows grouping-clauses and where-clauses in the query syntax, giving the user a high level of conciseness and expressiveness.
- **CIRC** is a *resource-based* API allowing to search the CIRC classification and other metrics filtering all the journals (the resource) by ISSN.

- **Embase** follows a *query-based* boolean search approach. It also allows to specify several fields in the query. For instance: `collaboration:kw AND impact:kw`
  Would match all the documents that contain `collaboration` and `impact` as keywords. Then, the user can paginate over the results. Also, the API offers a set of endpoints that are *resource-based* to retrieve a single document through several identifiers (see Section 4.2.5).
- **ScienceDirect Search** is a *query-based* API. It allows boolean searches on ScienceDirect (e.g., `all(Articulation Disorders) AND pub-date aft 2017`). Once the search is performed, the user can use the returned identifiers to look for full texts in **ScienceDirect Article (Full Text) Retrieval**.
- **ScienceDirect Article Metadata** follows a *query-based* search model. It allows boolean searches to retrieve the articles metadata (e.g., `keywords(cold AND rhinovirus AND NOT influenza)`). This API has three views (i.e., a *view* is a subset of the whole available metadata): JISC; STANDARD and COMPLETE, which cover different metadata (the view with the highest coverage of metadata is the COMPLETE view). It also returns a link to an endpoint of **ScienceDirect Article (Full Text) Retrieval** that can be used to retrieve the full text.
- **ScienceDirect Article (Full Text) Retrieval** is a *resource-based* API. The full texts can be searched and retrieved filtering by identifier, as detailed in Section 4.2.5. There are different views that can be retrieved using this API:
  - *BASIC*: simple bibliographic information
  - *META*: additional information about pages, issue, document type, etc.
  - *META_ABS*: all the information in *META* plus the abstract of the document.
  - *META_ABS_REF*: all the information in the *META_ABS* view plus references.
  - *FULL*: all the available data, and full-text of non open access articles.

  The returned metadata of each view is a subset of the next view ($BASIC \subset META \subset META\_ABS \subset META\_ABS\_REF \subset FULL$).
- **ScienceDirect Nonserial Title Metadata** follows a *resource-based* search model. The API allows to search by filtering the Nonserial titles (books, monographs, among other documental types) by title, ISBN, etc. (see more in Section 4.2.1). Two views are available for this API: BASIC (i.e., ISBN, edition and publisher) and STANDARD (i.e., document type, authors, editors and, research area).
- **SciVal** follows a dual approach, *resource and query-based* search. It is worth to remember that SciVal allows to compute metrics for different types of resources: authors, countries, group of countries, institutions, group of institutions, publications, sources, subject areas, topics, topic clusters, and world metrics. Prior to retrieve the specific metrics of each with a *resource-based* model, one must search the identifiers using a *query-based* boolean search for countries, group of countries, institutions, institutions groups, sources, subject areas, topics, and topic clusters. In case of publications, authors, the search is not performed against **SciVal** but in **Scopus Papers** and **Scopus Authors** respectively. Regarding the world metrics, no search is needed, and a *resource-based* approach is followed, to get the world metrics for specific areas, or documental types, among other options.
- **Scopus Affiliations** also follows a dual approach. First, the user must search the affiliations with the *query-based* search (e.g., `AFFIL(university)`, search all the affiliations with university in their name). If the users want to feed his search with additional data, such as authors or documents of a single affiliation, a call to the retrieval part of the API is needed. This retrieval endpoint follows a *resource-based* approach, and has five different views for returning different metadata:
  - *BASIC*: only identifiers for the affiliation are returned.
  - *LIGHT*: basic information about the institution is returned, such as: city, country, address, or name variants.
  - *STANDARD*: it adds the institution profile to the *LIGHT* view.
  - *DOCUMENTS*: return a list of documents metadata associated with the affiliation.
  - *AUTHORS*: return a list of authors that are linked to the affiliation.
- **Scopus Authors**. In the same way, this API uses a dual search model. In a first stage, the user must submit a boolean query to retrieve a list of authors (e.g.,`SUBJAREA(CHEM)`, search the authors with documents classified under the Chemistry subject area). Then, if the user needs more data, the next step would be to call the retrieval endpoint, which is *resource-based*. This last endpoint, has the following views:
  - *BASIC*. It includes identifiers related information, such as the ORCID and the Elsevier ID of the author.
  - *METRICS*. It includes the h-Index of the author, the number of documents, the citation count, and the referenced by count, as well as the number of coauthors.
  - *LIGHT*. It includes information about the affiliation of the author, including all the fields in the *BASIC* and *METRICS* views.
  - *STANDARD*. It comprises all the previous views except for some metrics and adds the affiliation data related to the authors (also the historical affiliation).
  - *ENHANCED*. It retrieves all the metadata of the previous views.
  - *DOCUMENTS*. This view allows to retrieve the list of documents metadata associated with the author.
- **Scopus Papers**. This API also is *query and resource based*. To search the papers, the user must submit a query that may have boolean operators (e.g., `PUBYEAR = 1994 AND SUBJAREA(MATH)`, documents of the math subject area published in 1994). The search has two views:

- *STANDARD*. It includes all the bibliographic information, identifiers, and affiliation names.
- *COMPLETE*. It adds the affiliation id, authors, author keywords and funding information to the *STANDARD* view.

Once the search has been performed, the user might want to extend the metadata of the results (e.g., retrieve citing papers, notice that to search for the citing papers of a paper, the authors must contact Elsevier for them to provide the way to do that; referenced papers, etc.). To do that, additional calls to the retrieval endpoints of the API are required. The available views of the retrieval endpoints are:

- *BASIC*. It includes the identifier and information about open access.
- *META*. It includes basic bibliographical information except for the authors.
- *META_ABS*. It adds authors, affiliations identifiers, and abstracts of documents to the *META* view.
- *REF*. It includes basic information about the references: bibliographical information, authors, document identifiers, and citations.
- *FULL*. It includes all the information for the original paper and all the metadata of its references.

- **Scopus Citations Count Metadata** is a *resource-based* API. It allows to search by several DOIs, PIIs or PubMedIDs.
- **Scopus Citations Overview** uses a *resource-based* search. It allows to search by filtering by several document identifiers (see Section 4.2.5). The result of the search is a count citation matrix by year of the documents searched.
- **Scopus Serial Title Metadata** uses *resource-based* for searching too. It allows to search and retrieve information about serial titles (i.e., Book Series, Conference Series, Journals) using three views:
  - *BASIC*. It includes the ISSN and information related to open access.
  - *STANDARD*. It includes the metadata of the previous view and adds metrics such as the SJR and the CiteScore.
  - *ENHANCED*. This view includes all the previous metadata and adds some extra metrics such as the: number of documents by year, number of cited documents, percentage of not cited documents and percentage of review article documents of the serial title.
- **Scopus PlumX Metrics** follows a *resource-based* search model. It allows to filter the aggregated metrics by some identifiers such as DOI, ISBN, PMID, among others. It returns alternative metrics of mentions in social media, blog posts, number of clicks, downloads, and views of the paper, as well as citations broken down by source (i.e., CrossRef, Scopus).
- **Europe PMC** follows a *query-based* structured search (e.g., `auth:"Simon Hubbard" sort_date:y`) and then, to retrieve additional metadata a *resource-based* model for searching is used. Also, the endpoints for the grants of the API follow the same model. Moreover, there are specific endpoints to retrieve the citing publications of a document, and the full text of books and articles in XML format.
  In addition, this API provides an OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) version of the API. It is intended to retrieve big amount of data from document repositories. In this protocol, the model is *resource-based*, first the records are listed using a date interval, and then they are retrieved using their identifiers.
- **IEEE Xplore Metadata** uses a *query-based* structured model for searching, for instance:
  `publication_year:2018 AND auth:"Simon Hubbard"`
  Also, it has an endpoint which is *resource-based* that allows to search a document directly by DOI.
- **IEEE Xplore Open Access** follows a *resource-based* model. It allows to filter the documents by DOI to retrieve the full texts of the open access documents.
- **CrossRef** uses a *resource-based* approach. It has several endpoints for *works*, *funders*, *journals*, *types*.
- **iCite** uses a *resource-base* search. It has three endpoints, one to query all the publications, and the other two that allow to search by single or several PMIDs respectively.
- **NCBI PMC** follows a *query-based* model. The user can submit a boolean search to perform a query (e.g., `asthma[mesh] AND leukotrienes[mesh] AND 2009[pdat]`, documents with MeSH terms asthma and leukotrienes that were published in 2009). Then the user can use the fetch operations to retrieve the full record and full texts depending on the selected database.
- **OpenCitations CCC** is a *resource-based* API. It allows to search for the citations using the DOI but in the context of a full text document, enabling to retrieve the context of a citation. To this end, there are several endpoints providing all the in-text reference pointers for a given DOI. Moreover, there are more endpoints to detect citations in specific parts of the text (i.e., sentences, paragraphs and sections).
- **OpenCitations COCI** is also *resource-base*. This API allows to retrieve the citations and references of a document by filtering all the documents in the database (the resource) by DOI or OCI.
- **OpenCitations CROCI** has exactly the same structure as **OpenCitations COCI**. Their difference is the content of the database. CROCI contains citations deposited by individuals, identified by ORCID.
- **OpenCitations Corpus** follows a *resource-based* model. It has two endpoints: one to retrieve the document metadata by DOI; and the other for retrieving the co-authorship matrix of a set of DOIs given.
- **OpenCitations Indexes unifying** follows the same structure of **COCI** and **CROCI**, and gives access to the union of the data in all the OpenCitation Indexes (i.e., COCI and CROCI).
- **OpenAlex** uses a dual model. The entities of the database (*works*, *authors*, *venues*, *institutions* and *concepts*) can be searched using a structured query that does not allow boolean operators and gives exact matches. Also, the API offers a free text query

mode *works*, *authors*, *venues* and *institution*, that does not give exact matches. Once the records have been identified the user can retrieve the whole metadata for the objects following a *resource-based* model for each endpoint of the entity objects available.

- **Unpaywall** use a dual search model. The *query-based* search is a boolean search, e.g.:
  `cell OR thermometry`
  Also, to retrieve a single document, the API offers and endpoint that follows a *resource-based* search, that allows to filter by DOI.
- **ORCID** also follows a dual search model. First the user can query for a list of authors using the search endpoint. This search is a *query-based* boolean search (e.g., `doi-self:10.1087/20120404`, search for ORCIDs which have the DOI). Then, the user can use the ORCIDs matching the query, to call the *resource-based* endpoints for retrieving the *employments*, *education*, *works*, *profile keywords*, *external identifiers*, *affiliations*, etc.
- **PLOS Search API** follows a *query-based* model. The query is a boolean search that allows to retrieve the contents of the PLOS database (e.g., `title:"Drosophila" AND body:"RNA"`). Once the records are returned the user can retrieve the full text, if available, in the **PLOS Text Data Mining** API.
- **PLOS Text Data Mining** follows a *resource-based* search to retrieve the full text of documents. The filter is performed by the DOI identifier of a document.
- **Springer Metadata API** follows a *query-based* model in a single endpoint to retrieve the results (e.g., `subject:Chemistry`). It also has some options to return different formats.
- **Springer Open Access API** is identical to the Metadata API but offers the JATS format to return the full text of the articles and is solely focused on open access documents.
- **DBLP** is *query-based*. It supports a free text query for authors, publications and venues in separate endpoints.

References

Abdill, R. J., & Blekhman, R. (2019). Meta-research: Tracking the popularity and outcomes of all bioRxiv preprints. In E. Pewsey, P. Rodgers, & C. S. Greene (Eds.), *eLife*, *8*, Article e45133. http://dx.doi.org/10.7554/eLife.45133.

Adie, E., & Roe, W. (2013). Altmetric: enriching scholarly content with article-level discussion and metrics. *Learned Publishing*, *26*(1), 11–17. http://dx.doi.org/10.1087/20130103.

Amundsen, M., Ruby, S., & Richardson, L. (2013). *RESTful web APIs*.

Analytics, C. (2017). Web of science. *Trust the Difference. Web of Science Fact Book*.

Arum, N. S. (2016). *A look at semantic scholar and Google scholar*. Academia.

Baas, J., Schotten, M., Plume, A., Cote, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, *1*, 377–386. http://dx.doi.org/10.1162/qss_a_00019.

Barabási, A., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, *311*(3), 590–614. http://dx.doi.org/10.1016/S0378-4371(02)00736-7.

Batagelj, V., & Cerinšek, M. (2013). On bibliographic networks. *Scientometrics*, http://dx.doi.org/10.1007/s11192-012-0940-1.

Bekkari, S. (2012). Integration of MARC bibliographic records for electronic resources subscribed in the Springer collections: Towards a consolidated e-book service at the Orange Labs Infodoc Centre. *Information Services & Use*, *32*(1–2), 87–92. http://dx.doi.org/10.3233/ISU-2012-0632.

Borgogno, O., & Colangelo, G. (2019). Data sharing and interoperability: Fostering innovation and competition through APIs. *Computer Law & Security Review*, *35*(5), Article 105314.

Börner, K., Chen, C., & Boyack, K. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, *37*, 179–255.

Burnham, J. F. (2006). Scopus database: a review. *Biomedical Digital Libraries*, *3*(1), 1–8. http://dx.doi.org/10.1186/1742-5581-3-1.

Champieux, R. (2015). PlumX. *Journal of the Medical Library Association*, *103*, 63–64. http://dx.doi.org/10.3163/1536-5050.103.1.019.

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics*, *5*(1), 146–166.

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, *62*(7), 1382–1402. http://dx.doi.org/10.1002/asi.21525.

Dresbeck, R. (2015). SciVal. *Journal of the Medical Library Association*, *103*, 164–166. http://dx.doi.org/10.3163/1536-5050.103.3.018.

Else, H. (2018). How Unpaywall is transforming open science. *Nature*, *560*(7718), 290–291. http://dx.doi.org/10.1038/d41586-018-05968-3.

Elsevier (2018). ScienceDirect search API migration. https://dev.elsevier.com/tecdoc_sdsearch_migration.html. (Accessed 15 February 2022).

Federer, L. M., Belter, C. W., Joubert, D. J., Livinski, A., Lu, Y.-L., Snyders, L. N., & Thompson, H. (2018). Data sharing in PLOS ONE: An analysis of Data Availability Statements. *PLOS One*, *13*, Article e0194768. http://dx.doi.org/10.1371/journal.pone.0194768.

Ferguson, C., Araújo, D., Faulk, L., Gou, Y., Hamelers, A., Huang, Z., Ide-Smith, M., Levchenko, M., Marinos, N., Nambiar, R., Nassar, M., Parkin, M., Pi, X., Rahman, F., Rogers, F., Roochun, Y., Saha, S., Selim, M., Shafique, Z., ... McEntyre, J. (2021). Europe PMC in 2020. *Nucleic Acids Research*, *49*, D1507–D1514. http://dx.doi.org/10.1093/nar/gkaa994.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., & Barabási, A.-L. (2018). Science of science. *Science*, *359*(6379), eaao0185. http://dx.doi.org/10.1126/science.aao0185.

Fraser, N., Momeni, F., Mayr, P., & Peters, I. (2020). The relationship between bioRxiv preprints, citations and altmetrics. *Quantitative Science Studies*, 1–21. http://dx.doi.org/10.1162/qss_a_00043.

Fricke, S. (2018). Semantic scholar. *Journal of the Medical Library Association*, *106*, http://dx.doi.org/10.5195/jmla.2018.280.

Ginsparg, P. (2011). ArXiv at 20. *Nature*, *476*, 145–147. http://dx.doi.org/10.1038/476145a.

González-Alcaide, G. (2021). Bibliometric studies outside the information science and library science field: uncontainable or uncontrollable? *Scientometrics*, *126*, 6837–6870. http://dx.doi.org/10.1007/s11192-021-04061-3.

Gray, A., & Price, R. (2020). Using InCites responsibly: A guide to interpretation and good practice. http://dx.doi.org/10.25561/75946.

Griffin, L. (2002). IEEE Xplore. Version 1.3. *Online Information Review*, *26*, 285. http://dx.doi.org/10.1108/oir.2002.26.4.285.12.

Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other sources. *Research Synthesis Methods*, *11*, 181–217. http://dx.doi.org/10.1002/jrsm.1378.

Haak, L. L., Fenner, M., Paglione, L., Pentz, E., & Ratner, H. (2012). ORCID: a system to uniquely identify researchers. *Learned Publishing*, *25*(4), 259–264. http://dx.doi.org/10.1087/20120404.

Heibi, I., Peroni, S., & Shotton, D. (2019). Crowdsourcing open citations with CROCI - An analysis of the current status of open citations, and a proposal. In *17th international conference on scientometrics and informetrics* (pp. 1448–1454). http://dx.doi.org/10.48550/arXiv.1902.02534.

Heibi, I., Peroni, S., & Shotton, D. (2019). Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. *Scientometrics*, *121*, 1213–1228. http://dx.doi.org/10.1007/s11192-019-03217-6.

Hendricks, G., Tkaczyk, D., Lin, J., & Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies*, *1*, 414–427. http://dx.doi.org/10.1162/qss_a_00022.

Hook, D. W., Porter, S. J., & Herzog, C. (2018). Dimensions: Building context for search and evaluation. *Frontiers in Research Metrics and Analytics*, *3*, 23. http://dx.doi.org/10.3389/frma.2018.00023.

Hunter, K. (1998). ScienceDirect™. *The Serials Librarian*, *33*, 287–297. http://dx.doi.org/10.1300/J123v33n03_07.

Hutchins, B. I., Baker, K. L., Davis, M. T., Diwersy, M. A., Haque, E., Harriman, R. M., Hoppe, T. A., Leicht, S. A., Meyer, P., & Santangelo, G. M. (2019). The NIH Open Citation Collection: A public access, broad coverage resource. *PLoS Biology*, *17*(10), 1–6. http://dx.doi.org/10.1371/journal.pbio.3000385.

Hutchins, B. I., Davis, M. T., Meseroll, R. A., & Santangelo, G. M. (2019). Predicting translational progress in biomedical research. *PLoS Biology*, *17*(10), 1–25. http://dx.doi.org/10.1371/journal.pbio.3000416.

Hutchins, B. I., Hoppe, T. A., Meseroll, R. A., erson, J. M., & Santangelo, G. M. (2017). Additional support for RCR: A validated article-level measure of scientific influence. *PLoS Biology*, *15*(10), 1–3. http://dx.doi.org/10.1371/journal.pbio.2003552.

Hutchins, B. I., Yuan, X., Anderson, J. M., & Santangelo, G. M. (2016). Relative citation ratio (RCR): A new metric that uses citation rates to measure influence at the article level. *PLoS Biology*, *14*(9), 1–25. http://dx.doi.org/10.1371/journal.pbio.1002541.

Jebari, C., Herrera-Viedma, E., & Cobo, M. J. (2021). The use of citation context to detect the evolution of research topics: a large-scale analysis. *Scientometrics*, *126*(4), 2971–2989. http://dx.doi.org/10.1007/s11192-020-03858-y.

Jones, N. (2015). Artificial-intelligence institute launches free science search engine. *Nature*, http://dx.doi.org/10.1038/nature.2015.18703.

Khurana, P., Ganesan, G., Kumar, G., & Sharma, K. (2022). A comparative analysis of unified informetrics with scopus and web of science. *Journal of Scientometric Research*, *11*(2), 146–154.

Ley, M. (2002). The DBLP computer science bibliography: Evolution, research issues, perspectives. In *International symposium on string processing and information retrieval* (pp. 1–10). http://dx.doi.org/10.1007/3-540-45735-6_1.

Liang, Z., Mao, J., Lu, K., Ba, Z., & Li, G. (2021). Combining deep neural network and bibliometric indicator for emerging research topic prediction. *Information Processing & Management*, *58*(5), Article 102611. http://dx.doi.org/10.1016/j.ipm.2021.102611.

Lomborg, S., & Bechmann, A. (2014). Using APIs for data collection on social media. *The Information Society*, *30*(4), 256–265.

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, *12*(4), 1160–1177. http://dx.doi.org/10.1016/j.joi.2018.09.002.

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado-López-Cózar, E. (2019). Google scholar, web of science, and scopus: Which is best for me? *Impact of Social Sciences Blog*.

Miedema, F. (2022). *Open science: The very idea*. Dordrecht: Springer Netherlands, http://dx.doi.org/10.1007/978-94-024-2115-6.

Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). *Software tools for conducting bibliometric analysis in science: An up-to-date review*. EPI SCP, http://dx.doi.org/10.3145/epi.2020.ene.03.

Orduña-Malea, E., & Delgado López-Cózar, E. (2018). Dimensions: redescubriendo el ecosistema de la información científica. *El Profesional de la Información*, *27*, 420. http://dx.doi.org/10.3145/epi.2018.mar.21.

Ortega, J. L. (2017). Are peer-review activities related to reviewer bibliometric performance? A scientometric analysis of publons. *Scientometrics*, *112*, 947–962. http://dx.doi.org/10.1007/s11192-017-2399-6.

Ortega, J.-L. (2020). Altmetrics data providers: A meta-analysis review of the coverage of metrics and publication. *El Profesional de la Información*, *29*, http://dx.doi.org/10.3145/epi.2020.ene.07.

Padilla-Cabello, J., Santisteban-Espejo, A., Heradio, R., Cobo, M. J., Martin-Piedra, M. A., & Moral-Munoz, J. A. (2022). Methods for identifying biomedical translation: a systematic review. *Amercian Journal of Translational Research*, *14*(4), 2697–2708.

Panczyk, M., Woynarowska-Sołdan, M., Belowska, J., Zarzeka, A., & Gotlib, J. (2015). Bibliometric evaluation of scientific literature in the area of research in education using incites™ database of thomson reuters. In *Proceedings of INTED 2015 conference* (pp. 487–496).

Penfold, R. (2020). Using the Lens database for staff publications. *Journal of the Medical Library Association*, *108*, http://dx.doi.org/10.5195/jmla.2020.918.

Pentz, E. (2001). CrossRef: a collaborative linking network. *Issues in Science and Technology Librarianship*, *10*, http://dx.doi.org/10.5062/F4CR5RBK.

Peroni, S., & Shotton, D. (2020). OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies*, *1*, 428–444. http://dx.doi.org/10.1162/qss_a_00023.

Peroni, S., Shotton, D., & Vitali, F. (2017). One year of the OpenCitations corpus. In *The semantic web – ISWC 2017* (pp. 184–192). Cham: Springer International Publishing.

Perrotta, C., Gulson, K. N., Williamson, B., & Witzenberger, K. (2021). Automation, APIs and the distributed labour of platform pedagogies in Google Classroom. *Critical Studies in Education*, *62*(1), 97–113.

Sampson, M., Barrowman, N. J., Moher, D., Klassen, T. P., Platt, R., John, P. D. S., Viola, R., & Raina, P. (2003). Should meta-analysts search Embase in addition to Medline? *Journal of Clinical Epidemiology*, *56*, 943–955. http://dx.doi.org/10.1016/S0895-4356(03)00110-0.

Savage, C. J., & Vickers, A. J. (2009). Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. *PLoS One*, *4*, Article e7078. http://dx.doi.org/10.1371/journal.pone.0007078.

Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K., Connor, R., Fiorini, N., Funk, K., & Hefferon, T. (2016). Database resources of the national center for Biotechnology information. *Nucleic Acids Research*, *44*, D7–D19. http://dx.doi.org/10.1093/nar/gkv1290.

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, *126*(6), 5113–5142. http://dx.doi.org/10.1007/s11192-021-03948-5.

Singh Chawla, D. (2022). Massive open index of scholarly papers launches. *Nature*, http://dx.doi.org/10.1038/d41586-022-00138-y.

Smith, D. R. (2016). Will publons popularize the scientific peer-review process? *BioScience*, *66*, 265–266. http://dx.doi.org/10.1093/biosci/biw010.

Snell, J., Tidwell, D., & Kulchenko, P. (2001). *Programming web services with SOAP*.

Staff, C. (2021). The lens scholarly MetaRecord strategy: Beyond microsoft academic graph. https://about.lens.org/the-lens-scholarly-metarecord-strategy-beyond-microsoft-academic-graph/. (Accessed 15 February 2022).

Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences*, http://dx.doi.org/10.1073/pnas.1708290115.

Sugimoto, C., Robinson-Garcia, N., Murray, D., Yegros-Yegros, R., & Lariviere, V. (2017). Scientists have most impact when they're free to move. *Nature*, *550*, 29–31. http://dx.doi.org/10.1038/550029a.

Sun, X., Kaur, J., Possamai, L., & Menczer, F. (2013). Ambiguous author query detection using crowdsourced digital library annotations. *Information Processing & Management*, *49*(2), 454–464. http://dx.doi.org/10.1016/j.ipm.2012.09.001.

The Europe P. M. C. Consortium (2014). Europe PMC: a full-text literature database for the life sciences and platform for innovation. *Nucleic Acids Research*, *43*(D1), D1042–D1048. http://dx.doi.org/10.1093/nar/gku1061.

Thelwall, M. (2018). Dimensions: A competitor to Scopus and the Web of Science? *Journal of Informetrics*, *12*, 430–435. http://dx.doi.org/10.1016/j.joi.2018.03.006.

Torres-Salinas, D., & Arroyo-Machado, W. (2022). APIs en contextos bibliométricos: introducción básica y corpus exhaustivo. *Anuario ThinkEPI*, http://dx.doi.org/10.3145/thinkepi.2022.e16a09.

Torres-Salinas, D., Bordons, M., Giménez-Toledo, E., Delgado López-Cózar, E., Jiménez-Contreras, E., & Sanz-Casado, E. (2010). Clasificación integrada de revistas científicas (CIRC): propuesta de categorización de las revistas en ciencias sociales y humanas. *Profesional de la Información*, *19*(6), 675–684. http://dx.doi.org/10.3145/epi.2010.nov.15.

van Eck, N. J., & Waltman, L. (2022). Crossref as a source of open bibliographic metadata. http://dx.doi.org/10.31222/osf.io/smxe5.

Velez-Estevez, A., García-Sánchez, P., Moral-Munoz, J., & Cobo, M. (2022). Why do papers from international collaborations get more citations? A bibliometric analysis of library and information science papers. *Scientometrics*, http://dx.doi.org/10.1007/s11192-022-04486-4.

Zhao, D. (2010). Characteristics and impact of grant-funded research: a case study of the library and information science field. *Scientometrics*, *84*, 293–306. http://dx.doi.org/10.1007/s11192-010-0191-y.