

# Sustaining the “frozen footprints” of scholarly communication through open citations

Zehra Taşkın<sup>1,2</sup> 

<sup>1</sup>Department of Information Management, Hacettepe University, Ankara, Turkey

<sup>2</sup>Scholarly Communication Research Group, Adam Mickiewicz University, Poznan, Poland

## Correspondence

Zehra Taşkın, Hacettepe University, Department of Information Management, Ankara, Turkey.  
Email: [ztaskin@hacettepe.edu.tr](mailto:ztaskin@hacettepe.edu.tr)

## Funding information

Türkiye Bilimsel ve Teknolojik Araştırma Kurumu, Grant/Award Number: 121C401; Scientific and Technological Research Council of Turkey's (TUBITAK) Domestic Postdoctoral Fellowship Program 2218

## Abstract

This review examines the role of open citations in fostering transparency, reproducibility, and accessibility in scholarly communication. Through a critical synthesis of diverse sources—articles, proceedings, presentations, datasets, and blog posts—it explores the motivations behind citing, the evolving meanings of citations, and key milestones in the open citation movement. Particular attention is given to initiatives like OpenCitations and the Initiative for Open Citations (I4OC), highlighting their contributions to advancing open scholarship. Key findings indicate that open citations democratize research by providing free access to citation data, improving discoverability, and facilitating the creation of public citation graphs. Technological advancements, such as advanced data models and reference mining tools, have significantly contributed to the management and utilization of citation data. Despite these benefits, challenges such as ensuring data quality and standardization, addressing structural inequalities in citation networks, and achieving universal publisher adoption persist. The study concludes with recommendations for future efforts, emphasizing policy advocacy, technological innovation, global collaboration, and educational initiatives to promote the widespread adoption and effective use of open citations. These strategies aim to make the “frozen footprints” of scholarly communication accessible to all, fostering a more equitable and transparent scientific landscape.

## 1 | INTRODUCTION

Cronin (1981) defined citations as frozen footprints in scientific achievement, stating that they offer a forensic tool that provides seductive power and versatility. Today, the meanings of citations and the motivations behind citing have evolved, and the number of citations in the literature has reached enormous proportions. Citation data is used for a variety of purposes, such as research evaluations, bibliometric research, and mapping scientific fields, making it a vital element of the knowledge organization domain for analytical study. However, existing systems pose challenges

in accessing citation data due to problems such as the lack of accurate and standardized citation data, the failure of publishers to support open citation movements and interoperability issues of information systems. To address these challenges and ensure accessibility to everyone interested, various “open citation” initiatives and organizations have been launched (Heibi et al., 2019b; Ortega, 2021; Peroni et al., 2015; Shotton, 2013, 2018). Funders require open citation data sharing, and countries support open citations in their national systems, recognizing their importance.

The primary objective of this review is not merely to summarize existing efforts but to critically examine how

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *Journal of the Association for Information Science and Technology* published by Wiley Periodicals LLC on behalf of Association for Information Science and Technology.

open citations can enhance transparency, reproducibility, and equitable access to scholarly data. To achieve this, a systematic review was conducted, encompassing meta-data and full texts from a variety of sources, including but not limited to articles, conference proceedings, blog posts, presentations, data models, initiatives, datasets, and other relevant materials.

This review is intended to serve as a comprehensive resource for researchers, practitioners, and policymakers, offering a holistic understanding of the open citation movement, its key milestones, and ongoing efforts. The main research question guiding this review is:

- How does the open citation initiative enhance research by fostering transparency, reproducibility, and accessibility in scholarly communication?

To address this overarching question, the following sub-questions are explored:

- What is the current landscape of open citations in scholarly communication, and how has it evolved over time?
- What are the key milestones in the development of open citation initiatives?
- How have open citation practices evolved within the field of information science?
- What are the primary challenges and opportunities highlighted in the literature regarding open citations?
- What fundamental challenges and motivations drive the development and adoption of open citation initiatives?

Before addressing these research questions, it is essential to understand why authors cite and the meanings of citations. Citations serve not only as a means to acknowledge prior work but also reflect the motivations and behaviors of researchers in scholarly communication. Understanding these diverse motivations is crucial because open citations aim to enhance transparency, fairness, and accessibility in citation practices. By recognizing how citations function and the biases that can influence them, we can better appreciate the need for open citation initiatives to ensure a more equitable and reproducible academic landscape.

## 1.1 | Why do authors cite, and what is the meaning of citations today?

According to the Cambridge Dictionary, a citation is “a word or piece of writing taken from a written work” (“Citation,” 2024). Through citations, authors establish

clear connections between their current work and previous studies in the vast archive of scholarly literature, symbolizing the concepts or ideas they discuss (Garfield, 1997). Citations are essential parts of research papers as they represent the relationship between the cited and citing documents (Smith, 1981), and as Ziman (1968) defines it, citation is the acknowledgment between documents.

Many researchers have explored the question, “Why do authors cite?” and have found similar answers, such as paying homage to pioneers in the field, explaining methods, providing a literature background, criticizing previous claims, confirming or falsifying ideas, or verifying data (e.g., Garfield, 1970; Weinstock, 1971). My classification scheme (Taşkın & Al, 2018) for citation content covers the meanings of citations (positive, negative, and neutral), their purposes (literature review, definition, methods, comparison, proof, support, etc.), their forms (mentioning authors' names, multiple citations, and quotation marks), and their arrangement (citation section, number of uses in the text, etc.). However, to count citations accurately, all citations must be considered equal, though this was not always the case from the beginning.

It is unrealistic to expect that all authors have meticulously and consistently cited their sources (Price, 1963), and citations have often been viewed as necessary trivialities by some authors (Kaplan, 1965). Authors' citing behaviors can be influenced by their expectations or environments. For instance, authors may cite many papers because scientific articles without reference lists are seen as “chatty” and can create suspicion (Tagliacozzo, 1967). Moreover, rewarding authors based on the number of citations can lead to practices such as gift citations to colleagues, and authors may avoid citing their competitors (Goudsmith, 1974). This issue still persists in the literature today. Authors may focus on some works while ignoring others (Sokolov, 2023). There are significant structural inequalities in scientific citation networks, particularly against female authors and minority researchers (Nettasinghe et al., 2021). These inequalities are exacerbated by the emerging trend of using artificial intelligence in writing literature reviews, with tools like Connected Papers<sup>1</sup> or scite<sup>2</sup> potentially deepening the Matthew Effect in citations (Merton, 1968), as they rely on current citation networks. Additionally, authors' tendency to cite papers without thoroughly reading them (Simkin & Roychowdhury, 2006) in the publish-or-perish world makes citations one of the gamed metrics (Ioannidis & Maniadis, 2024).

One key aspect is that “accessible citations” enable comprehensive analysis of the literature. A review from the early years of citation analysis revealed that citation-based studies measured the aging of scientific literature,

identified key research, discovered trends, and evaluated citation indexes (Tagliacozzo, 1967). However, when citations are not common and widely available, issues arise regarding the reproducibility and accuracy of scientific research (Peroni et al., 2015). Differences in citation numbers, indexed sources, and citation structures across databases, as well as restrictive access policies like fingerprint proof (Peroni et al., 2015; Piwowar & Vision, 2013), pose significant challenges. Today, with advanced techniques and network analysis, the importance and use of citations in information science and other fields have grown significantly. To ensure reproducible and accurate studies, there is a critical need for openly accessible structured citations. Thus, understanding the importance of open citations in scholarly communication is crucial.

## 1.2 | The politics and power of citations

As Latour argues (Latour, 1988), citations are more than mere acknowledgments; they are strategic tools that researchers use to lend strength to their arguments, bolster credibility, and navigate the politics of scientific discourse. The presence or absence of references can determine the perceived legitimacy of a scientific work. A paper rich with references creates an impression of authority, while one lacking citations is vulnerable, akin to “a child without an escort walking at night in a big city.”

Latour emphasizes that references do not simply support claims but also play a role in constructing scientific authority. By carefully selecting and arranging references, authors can neutralize opposition, strengthen alliances, and align themselves with influential paradigms. This practice highlights the rhetorical power of citations, which serve not only to reinforce the writer's position but also to shape the trajectory of future scholarship. For instance, a paper with meticulously curated references forces dissenters to engage with an extensive network of prior literature, adding weight to the author's argument and deterring superficial critique.

Moreover, Latour points out that citations contribute to the lifecycle of a scientific fact, as the validation of a claim depends on its adoption and reinforcement by subsequent research. Much like genes that require propagation, scientific statements rely on being cited by future works to gain stability and recognition within the scientific community. This perspective underscores the importance of open citations. When citation data is openly accessible, the dynamics of scientific discourse become more transparent, enabling a deeper understanding of how ideas evolve and how power structures within academia influence the flow of knowledge. Open citations

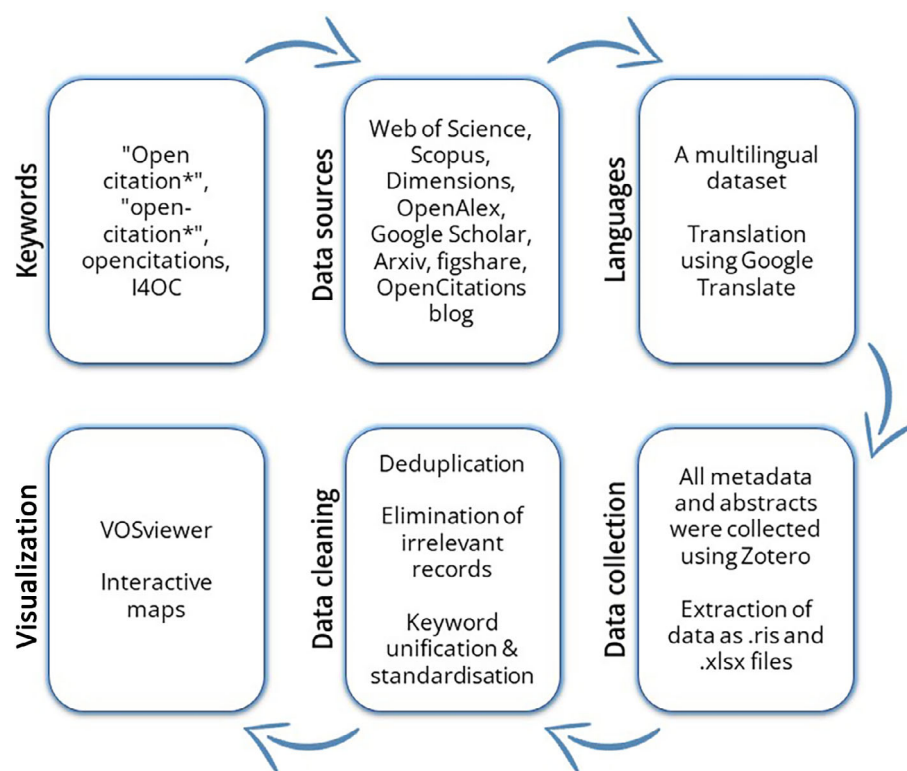
can thus democratize these processes, ensuring that the “frozen footprints” of scholarly communication are available for critical scrutiny and equitable use.

## 2 | METHODOLOGY

To provide a systematic review of open citations, a structured methodology was employed (see Figure 1). The details of the methodology are as follows:

- *Selection of Keywords:* The main keywords selected for this study are “Open citation\*,” “open-citation\*,” OpenCitations, and I4OC. Although terms like “citation data” or “bibliographic reference data” are used in papers on open citations, these keywords are used for many purposes in the literature, from defining methodology and sources of data to citation analysis studies, resulting in thousands of unrelated papers. Therefore, only keywords directly related to open citations and the main initiatives were considered.
- *Data Sources:* Most review studies cover only major commercial citation indexes such as Web of Science and Scopus. However, as of May 2024, there are only 53 papers indexed in Web of Science on this subject. To reach a wider range of outputs, the data sources for this review were extended (Web of Science, Scopus, Dimensions, OpenAlex, Google Scholar, Arxiv, and figshare), as suggested in the literature (Martin-Martin et al., 2021; Visser et al., 2021). Additionally, OpenCitations has a significant blog for current developments in the subject. The full texts of these blog posts were also included in the study.
- *Languages:* As multilingualism is a consideration for open citation studies (e.g., Malínek et al., 2024), making it problematic to focus solely on English-language publications. If searches retrieved outputs in different languages, they were included in the study by translating the titles and abstracts to English using Google Translate. However, it is important to note that only papers with English-language keywords in their metadata were included in the review, which is the main limitation of the study. Languages covered in the study are English, Spanish, German, Chinese, Japanese, and Finnish.
- *Data Collection and Cleaning Process:* All metadata and abstracts of 284 research items, including datasets, presentations, and images, were collected using a Zotero Library. Afterward, a deduplication and elimination process were conducted to remove duplicates and unrelated records from the dataset. In the end, 174 outputs were retained. Blogs were considered a separate dataset, with the full texts of 149 blog posts

FIGURE 1 Methodology of the review.



collected for a separate review. After creating the main datasets, all words used in titles, keywords, and abstracts were standardized. The standardization process included unifying plural/singular forms, abbreviations, and different spellings of words (such as “behavior” and “behavior”).

- **Visualization:** To create topic networks of the outputs, the Java-based software VOSviewer was used. The resolution parameter for all VOSviewer visualizations was set to 1.0, which is the default value in VOSviewer.

### 3 | FINDINGS

#### 3.1 | What is an open citation?

In 2013, Shotton (Shotton, 2013) highlighted the irony that in an era of open access, reference lists from journal articles—essential components of scholarly communication that allow for credit attribution and integration of independent research—are not freely accessible to all scholars, including the authors themselves. Citations are crucial in connecting scholarly knowledge and can be obtained directly from references, databases, or citation indexes. When citation data is freely accessible, downloadable, and reusable, it is termed “open citation” (Peroni & Shotton, 2018a).

It is important to note that being just accessible is not sufficient for being “open citation.” To qualify as “open

citation,” the citation must meet certain criteria (Initiative for Open Citations (I4OC), 2020; Peroni, 2018): It should be structured, separate, and open (see Figure 2 for examples).

Open citations provide numerous benefits that contribute to a more transparent, equitable, and efficient scientific research ecosystem. The key benefits of open citations can be listed as follows (Hutchins, 2021; Initiative for Open Citations (I4OC), 2020; Peroni & Shotton, 2020):

- **Transparency and reproducibility:** Open citations enhance the transparency and reproducibility of bibliometric and scientometric analyses by allowing the publication of source data.
- **Enhanced discoverability:** They increase the discoverability of scientific articles, enabling researchers to follow citation trails and find relevant literature easily. This is particularly beneficial for individuals outside academic institutions with subscriptions to commercial citation databases.
- **Commercial activity stimulation:** Open citation data stimulate downstream commercial activities by allowing businesses to develop new tools and services based on this data.
- **Improved science policy:** Access to open citation data aids in making informed decisions regarding science policy by providing comprehensive and accurate data.
- **Equitable access:** Open citations ensure that all scholars, regardless of institutional affiliation or financial resources, have access to crucial bibliographic data.



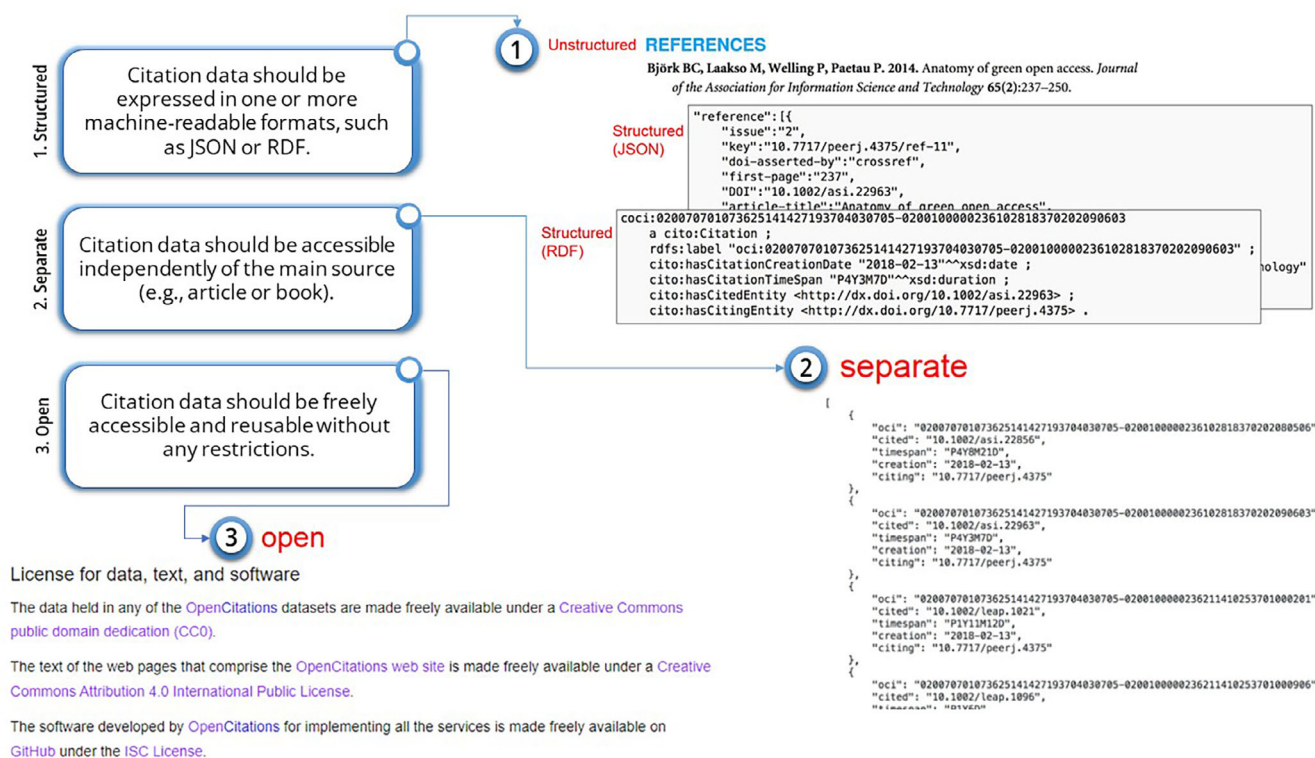


FIGURE 2 Open citation criteria (Initiative for Open Citations (I4OC), 2020; Peroni, 2018).

- *Support for open scholarship*: They support the broader open scholarship movement, promoting the free exchange of knowledge and reducing reliance on proprietary data sources.
- *Building new services*: The ability to build new services over the open citation data benefits publishers, researchers, funding agencies, academic institutions, and the general public, and enhances existing services.
- *Creation of a public citation graph*: Creating a global web of interconnected scholarly citation data improves the discoverability of both subscription-based and open-access content. This citation graph allows for the exploration of relationships between knowledge fields and tracks the development of ideas and academic disciplines.

By providing a robust framework for citation data, open citations foster an environment where scientific knowledge can be freely accessed, analyzed, and utilized for various scholarly and practical applications. However, the review reveals that while open citations significantly enhance discoverability and transparency, disparities in adoption persist among publishers, particularly those with substantial market control. These findings directly address the question of how open citations can democratize research by broadening access and promoting inclusive scholarly communication.

### 3.2 | Clarifying open citations: The movement, OpenCitations infrastructure, and I4OC

The terms open citations, OpenCitations, and the Initiative for Open Citations (I4OC) are often used interchangeably or misunderstood, even though they refer to distinct but related concepts within the scholarly communication landscape. For many, the boundaries between the broader open citations movement, the OpenCitations infrastructure, and the I4OC initiative remain unclear. To avoid confusion and provide clarity, it is essential to define each of these entities and explain their roles within the open scholarship ecosystem.

As mentioned earlier, the term open citations refers to citation data that is freely accessible, reusable, and downloadable without restrictions. Citations, which are essential to scholarly communication, serve as the connective tissue between research outputs by attributing credit and facilitating knowledge integration. Despite the rise of open access, Shotton (2013) pointed out that reference lists, which are crucial to scholarly work, have often remained inaccessible to many scholars, creating a paradox in the scholarly ecosystem. To address this gap, open citations emerged as a movement advocating for citation data that is structured, separable, and open.

Closely linked to this movement is the OpenCitations initiative, a key player in the open citations space. OpenCitations<sup>3</sup> is a not-for-profit infrastructure organization committed to publishing open bibliographic and citation data using semantic web technologies. OpenCitations fully embraces the principles of Open Science and the FAIR data principles—data should be findable, accessible, interoperable, and reusable. In addition to providing open citation data, OpenCitations plays an advocacy role as a founding member of I4OC. The organization has made significant strides in advancing open citation practices, including the creation of globally unique and persistent identifiers for bibliographic citations, known as Open Citation Identifiers (OCIs) (OpenCitations.net, n.d.).

The Initiative for Open Citations (I4OC),<sup>4</sup> launched as a collaborative effort among scholarly publishers, researchers, and other stakeholders, seeks to promote the unrestricted availability of scholarly citation data. I4OC emphasizes that for citation data to be truly open, it must be structured, separable, and open, facilitating both human and machine access. The initiative aims to establish a global public web of linked citation data that enhances the discoverability of scholarly content—particularly benefitting those without access to commercial citation databases. By making citation data openly available, I4OC not only fosters transparency and accessibility in research but also enables the creation of new services and insights, such as the ability to build a public citation graph to track the evolution of ideas and scholarly disciplines over time (Initiative for Open Citations (I4OC), 2020).

Together, these initiatives and movements form a collective effort to transform citation practices, ensuring that citation data becomes a public resource accessible to all. By distinguishing between the broad open citations movement, the OpenCitations infrastructure organization, and the I4OC initiative, this review clarifies their roles and contributions to the open scholarly landscape.

### 3.3 | Milestones of open citation movement

Significant milestones and developments from the open citation movement, highlighting the evolution and ongoing efforts to improve open access to citation data are as follows<sup>5</sup>:

- 2010: OpenCitations formally started as a one-year project funded by JISC with David Shotton as director. The project aimed to publish open bibliographic citation information in RDF and released the

OpenCitations Corpus (OCC), initially populated with citations from the Open Access Subset of PubMed Central.

- 2011: PubMed Central's Open Access full-text papers more than doubled since they were last harvested into the OCC.
- 2015: Silvio Peroni joined as co-director and technical manager. A new instantiation of the OCC was set up with a new metadata schema and technologies for automated citation metadata ingestion.
- 2016: OCC began ingesting, processing, and publishing reference lists of scholarly papers available in Europe PubMed Central, with additional metadata from Crossref and ORCID.
- 2017: OpenCitations received a grant from the Alfred P. Sloan Foundation for The OpenCitations Enhancement Project, improving infrastructure and developing citation indexes, including COCI.
- 2017: The Initiative for Open Citations (I4OC) launched, encouraging publishers to make citation data open.
- 2017: The scientometric community published a letter urging scholarly publishers to open their citations and respond to the I4OC call (Sugimoto et al., 2017). Additionally, the I4OC team published an open letter in the *eLife* journal for I4OC stakeholders (The I4OC Team, 2017).
- 2018: Springer Nature joined I4OC, promoting the open sharing of citation data.
- 2018: By September 2018, the percentage of publications with open citations increased from 1% to 52% out of 40.8 million articles with references stored in Crossref, resulting in more than 500 million citations being openly accessible.
- 2018: Shotton (2018) wrote an opinion piece in *Nature* urging funders to mandate open citations, and France added open citations in its National Plan for Open Science (Badolato, 2018).
- 2019: Another grant from the Wellcome Trust was awarded for the Open Biomedical Citations in Context Corpus project, aiming to create a new dataset distinguishing individual in-text references. OpenCitations was selected by the Global Sustainability Coalition for Open Science Services (SCOSS) for crowd-funding support, starting in January 2020, to transition into a global scholarly infrastructure organization.
- 2019: The editors of the *Journal of Informetrics* resigned due to growing dissatisfaction with Elsevier's actions and policies regarding open citations (Larivière, 2019).
- 2019: The NIH released its Open Citation Collection (Hutchins, Baker, et al., 2019).

- 2020: The Initiative for Open Abstracts (I4OA) was launched, pushing for open access to study abstracts. By 2021, 39% of articles with a Crossref DOI had open abstracts, almost doubling the proportion since 2018.
- 2020: Elsevier announced to open its citations (Waltman, 2021).
- 2021: Further development and integration of citation data from diverse sources such as PubMed Central, ArXiv, and CrossRef. Enhancements in data processing, including error correction and augmentation using bibliographic records and DOI identifiers.
- 2021: One of the important data sources for publication and citation data, Microsoft Academic Graph (MAG) was terminated (Chawla, 2021).
- 2022: After the termination of MAG, OpenAlex beta was launched (OpenAlex support, 2024).
- 2022: Continuous improvements in infrastructure and data handling capabilities. Expansion of citation data coverage and improved accessibility.
- 2023: Introduction of new user interfaces and visualization tools to enhance the usability and accessibility of citation data. Emphasis on personalized search capabilities and social media interactions.
- 2024: *The Barcelona Declaration on Open Research Information*<sup>6</sup> was published, marking a significant milestone in the open research movement. This declaration advocates for the unrestricted sharing of research information, including citations, to promote transparency, equity, and reproducibility in research. It emphasizes the importance of making all forms of research information, including citations, openly available for the global research community (Barcelona Declaration on Open Research Information, 2024).
- 2024: OpenCitations is managed by the Research Centre for Open Scholarly Metadata at the University of Bologna, ensuring its original aim of free provision of open bibliographic and citation data, services, and software is maintained. Ongoing efforts are focused on enhancing reference matching and incorporating user-generated annotations, with particular emphasis on CiTO-based citation typing. This ontology allows for the characterization of citations in terms of their factual and rhetorical nature, enabling these descriptions to be published online (Peroni & Shotton, 2012).

I want to highlight David Shotton's<sup>7</sup> tireless efforts, which are instrumental in propelling the open citation movement forward. His vision and leadership have guided OpenCitations from its inception in 2010 to its current status as a pivotal force in the open access landscape. As highlighted in one of his latest blog posts,

“From little acorns ... A retrospective on OpenCitations” (Shotton, 2021), Shotton's metaphor aptly captures the growth and impact of the initiative. Starting as a modest project, OpenCitations has blossomed into a robust platform that significantly advances the accessibility and transparency of scholarly citation data. Shotton's contributions, along with his commitment to the cause, have laid a solid foundation for future developments and have inspired a global push toward open science.

### 3.4 | Thematic classification of research subjects

The aim of this section is to systematically classify open citation efforts based on their subjects or areas of focus. This classification underscores the diverse approaches and impacts of open citations across various domains, synthesized from a collection of datasets. To achieve this objective, a co-occurrence network was created as an initial step to understand the basic concepts and main keywords of the subject. Figure 3 displays the co-occurrence network of keywords, illustrating the frequency and connections of keywords that appear together.

The co-occurrence map provides keywords from various dimensions, ranging from technical aspects to the open citation movement. However, examining only the keywords is insufficient to understand the main themes of the subject because not all papers include keywords, and keywords alone may not capture the full picture due to the limitation of five top words for all studies at most. To address this limitation, a term map, shown in Figure 4, was created using the titles and abstracts of the dataset, which includes scientific articles, proceedings, presentations, and more. The objective of the keyword analysis is to identify the main themes of the subject and highlight the most significant ones in this review from a broader perspective.

After examining the distribution of keywords across different clusters and systematically evaluating the full texts of the research, five themes were identified.

#### 3.4.1 | Theme 1: Open citation movement and its impact

The open citation movement aims to enhance transparency, accessibility, and reproducibility in academic research by making citation data openly available. The studies within this theme provide comprehensive analyses of open citations, highlighting their impact on academic research and their potential to transform citation practices (e.g., Hutchins, 2021). Advocacy for







mandatory open citations by funders emphasizes the benefits of research transparency and accessibility (Shotton, 2018).

The study, “One Year of the OpenCitations Corpus” (Peroni et al., 2017) discusses the release of RDF-based OpenCitations Corpus, illustrating the project’s significant progress and its implications for the research community. Additionally, the Initiative for Open Citations (I4OC) is detailed in several presentations, describing collaborative efforts to release citation data from publishers and promote public ownership of the citation graph (Taraborelli, 2017, 2018).

Numerous studies in the literature, particularly those authored by the OpenCitations group (Shotton, Peroni, and Heibi), introduce new developments in this theme. These studies are cited in various parts of this review. Researchers seeking preliminary or foundational information about the subject should refer to the research in this theme.

### 3.4.2 | Theme 2: Data models and reference mining

Efforts in data models and reference extraction focus on improving the accuracy and efficiency of citation networks and managing bibliographic data. Articles in this group highlight various methodologies and tools developed to enhance the extraction, segmentation, and management of references. Some studies address domain-specific challenges and solutions in reference extraction (Arnold & Jäschke, 2022; Birkeneder et al., 2022). For example, “A Game with Complex Rules: Literature References in Literary Studies” emphasizes the unique challenges of reference extraction in literary studies compared with STEM fields and social sciences (Arnold & Jäschke, 2022). Similarly, “A knowledge graph embeddings based approach for author name disambiguation in open citation data” presents advanced methods using knowledge graph embeddings to disambiguate author names in open citation data, thereby improving network accuracy and efficiency (Santini et al., 2022). Additionally, tools like the EXCITE toolchain (Hosseini et al., 2019) and the Linked Open Citation Database (Alexiou et al., 2016; Lauscher et al., 2018) are designed to extract, match, and publish open literature references, significantly contributing to the availability and utility of citation data. This cluster also includes software reviews that present technical approaches to relevant parties (Heibi et al., 2019b). The research within this theme is intended for professionals in the field who are seeking technical solutions for open citations.

### 3.4.3 | Theme 3: Empirical research on research evaluations and open citations

Empirical research in open citations examines the practical implications and impacts of open citation data on academic publishing and research evaluation. Studies like “A construction and empirical research of the journal disruption” (Jiang & Liu, 2023) and “Do open citations give insights on the qualitative peer-review evaluation in research assessments? An analysis of the Italian National Scientific Qualification” (Bologna et al., 2023) investigate how open citations influence journal disruption and the qualitative aspects of research, respectively. The article “Coverage and correlations between open citations in Crossref and readership in Mendeley: Different fields of Brazilian science” (Maricato et al., 2023) provides an empirical analysis of the coverage and correlations between open citations, while “Open data to evaluate academic researchers: an experiment with the Italian Scientific Habilitation” (Di Iorio et al., 2019) examines the effectiveness of using open data for evaluating academic researchers in Italy. These studies collectively highlight the importance of open citation data in enhancing the transparency and reliability of research assessments.

### 3.4.4 | Theme 4: Comparative analysis

Comparative analysis of different bibliometric tools and data sources is crucial for understanding the strengths and limitations of various citation indexes. The paper “Can Crossref citations replace Web of Science for research evaluation?” (Chudlarsky & Dvorak, 2019) compares the coverage and feasibility of using Crossref’s Open Citation Index (COCI) against Web of Science for research evaluation purposes. Another study, “New trends in bibliometric APIs: A comparative analysis” (Velez-Estevez et al., 2023) evaluates the capabilities of new bibliometric APIs, emphasizing the role of open citation data in enhancing bibliometric research.

There are many studies in this cluster that compare data sources of citations (e.g., Web of Science, Microsoft Academic, Scopus, Dimensions, etc.). However, since most of these studies are discussed in the blue cluster subsection of the co-citation analysis part of this review, they are not mentioned here to avoid duplication.

### 3.4.5 | Theme 5: Infrastructure and organizational efforts

Infrastructure and organizational efforts in open citations focus on building and maintaining systems that support

open scholarship. The article “OpenCitations: an infrastructure organization for open scholarship” (Peroni & Shotton, 2020) details the goals and achievements of OpenCitations as an infrastructure organization dedicated to promoting open scholarship. Another pivotal comment in *Nature*, “Open citations” (Shotton, 2013) discusses the benefits and challenges of making bibliographic citation data freely available, highlighting the importance of open access and the role of the OpenCitations Corpus (OCC) in facilitating this process.

### 3.5 | Co-citation analysis

Co-citation refers to the frequency with which two documents are cited together (Small, 1973). The results of the co-citation analysis of the dataset are shown in Figure 5.

The co-citation network of the dataset outlines the path to open citations. The examination of the five defined clusters is as follows.

#### 3.5.1 | Red cluster: A path from citation indexing to semantic publishing and linked data

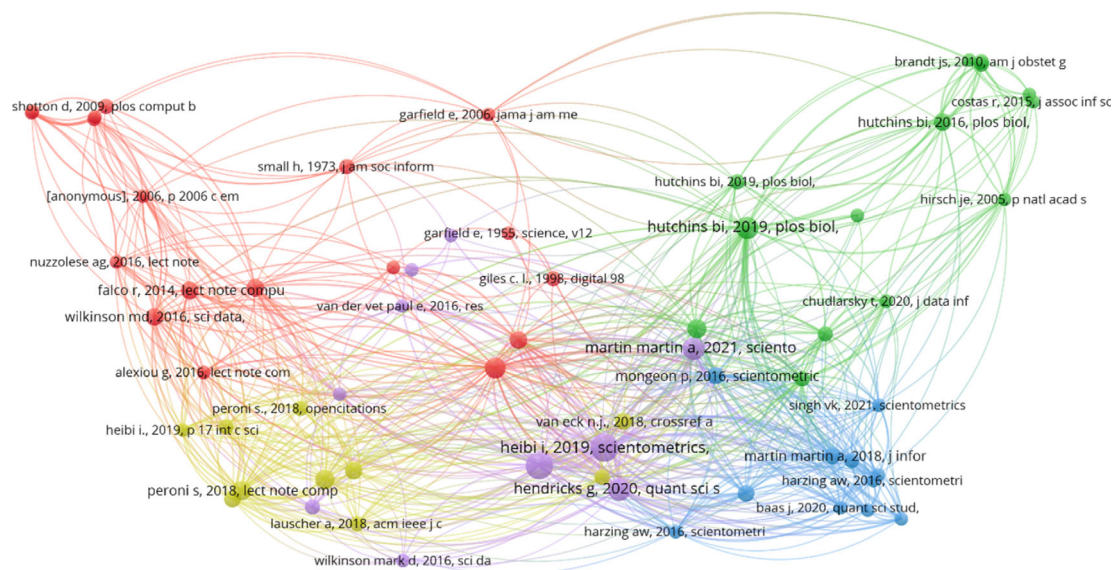
Co-citations in open citation papers trace their origins to the creation of citation indexes (Garfield, 1955), followed by Small’s paper on co-citations (1973). This indicates that citation data and the hidden connections within

them became significant for scientific studies over ~25 years. In 1998, the development of CiteSeer (Giles et al., 1998), an autonomous citation indexing system capable of parsing and identifying citations in different formats automatically, marked the beginning of automated methods for citation indexing. This was a pivotal advancement, as previous methods relied on manual indexing.

Post-2000, studies have focused on automatic classification of citations (Teufel et al., 2006), ontologies (Ciccarese et al., 2008; Peroni & Shotton, 2012; Shotton, 2010), and semantic publishing (Shotton et al., 2009). More recently, papers cited in this cluster pertain to linked data projects such as LOD (Alexiou et al., 2016) and The Semantic Web Dog Food (SWDF) (Nuzzolese et al., 2016). The importance of this cluster lies in illustrating the evolution of topics from the creation of citation indexes to the development of new techniques and data structures. The strongest node of this cluster is the pivotal *Nature* paper by David Shotton (2013) which covered all these issues above from citation indexes to open citation movement.

#### 3.5.2 | Green cluster: Sources of citation data for bibliometric studies

This cluster began with the foundational paper on the h-index (Hirsch, 2005). Early papers in this cluster used Web of Science data to identify citation classics (Brandt



**FIGURE 5** Co-citation analysis of the dataset (not all research in my dataset includes reference data. Therefore, only articles that contain references were considered for the analysis. For records not indexed in WoS, the references of these articles were transformed into the WoS format. Only research that had been cited at least three times was included in the analysis. A total of 64 studies are shown in the network. The interactive map is available at that link).

et al., 2010, 2019) or to make comparisons with altmetrics scores and citations (Costas et al., 2015; Thelwall et al., 2013). However, following the publication of “Science of Science” (Fortunato et al., 2018), the limitations of citation data in Web of Science began to be discussed. Even Brandt et al. (2019) acknowledged this as a limitation of their study. Since then, new data sources such as the NIH Open Citation Collection (NIH-OCC) have been utilized (Hutchins, Baker, et al., 2019; Hutchins, Davis, et al., 2019; Mitra et al., 2021). Additionally, there have been discussions about using CrossRef as an alternative to Web of Science for research evaluation purposes (Chudlarsky & Dvorak, 2019), as well as comparisons of different data sources, including Scopus, Web of Science, Dimensions, CrossRef, and Microsoft Academic (Visser et al., 2021). The article entitled “Predicting translational progress in biomedical research” (Hutchins, Davis, et al., 2019) is the strongest node of this cluster with 46 links and 85 total link strengths.

### 3.5.3 | Blue cluster: Comparison of citation data sources

The blue cluster is one of the youngest among the cited document clusters. It began in 2016 with Harzing’s paper on Microsoft Academic (Harzing, 2016) and was followed by papers comparing citation data sources, including Crossref, Dimensions, Google Scholar, Microsoft Academic, Scopus, and Web of Science (Baas et al., 2020; Harzing, 2019; Harzing & Alakangas, 2016, 2017; Herzog et al., 2020; Hug & Brändle, 2017; Martín-Martín et al., 2018; Mongeon & Paul-Hus, 2016; Singh et al., 2021; Wang et al., 2020). These papers are all pivotal in their respective fields, as evidenced by their similar number of links (between 20 and 28) and comparable total link strengths (between 28 and 51).

### 3.5.4 | Yellow cluster: OpenCitations

The yellow cluster includes the influential papers authored by the managers of OpenCitations (Heibi et al., 2019a, 2019b; Peroni et al., 2017; Peroni & Shotton, 2018a; Peroni & Shotton, 2018b). Additionally, it encompasses significant developments in open citations, such as the letter to the scientometrics community on the need for open citations (Sugimoto et al., 2017), the establishment of the Linked Open Citation Database (LOC-DB) (Lauscher et al., 2018), and the inclusion of CrossRef as a data source in VOSviewer (Van Eck et al., 2018). This cluster also contains a comprehensive

study of book items and their citations using OpenCitations as a data source. Similar to the blue cluster, all nodes have a comparable number of links (from 19 to 30) and total link strengths (from 25 to 52).

### 3.5.5 | Purple cluster: A cluster connecting all others

Naming the purple cluster is challenging due to its wide range of topics, including bibliographic coupling (Kessler, 1963), paper recommendation systems (Beel et al., 2016), errors in citation networks (Van der Vet & Nijveen, 2016), FAIR principles for scientific data (Wilkinson et al., 2016), open citation data and data sources (Heibi, 2019; Hendricks et al., 2020; Martín-Martín et al., 2021), data models, infrastructures, and APIs (Daquino et al., 2020, 2022; Peroni & Shotton, 2020). Despite the diverse subjects, they all relate to the other clusters, making this cluster a connecting hub. This is confirmed by the fact that four of the papers with the highest total link strength (Heibi, 2019; Hendricks et al., 2020; Martín-Martín et al., 2021; Peroni & Shotton, 2020) are part of this cluster.

The results gathered from the co-citation network clearly demonstrate the evolution of the subject and its core sources. Additionally, two important findings have emerged. First, the diversity of document types, ranging from articles to letters and proceedings, underscores the need for diversified research evaluation systems that consider various document types. Second, the importance of citation data is highlighted as essential for properly investigating knowledge flows and the evolution of scientific fields.

## 3.6 | The importance of effective science communication: OpenCitations Blog

The OpenCitations organization has long engaged in significant science communication efforts to raise awareness about open citations. Through the OpenCitations Blog,<sup>8</sup> researchers and interested parties can easily stay informed on the topic without navigating through extensive scientific articles. Analyzing the blog, which featured 149 posts as of May 2024, revealed the thematic map presented in Figure 6.<sup>9</sup>

Before proceeding to the explanation of Figure 6, it is crucial to highlight that all developments, new models, publications, and data sets related to open citations are disseminated to the relevant audience through this blog. The term map, created using the full texts of the blog posts, clearly illustrates this point.



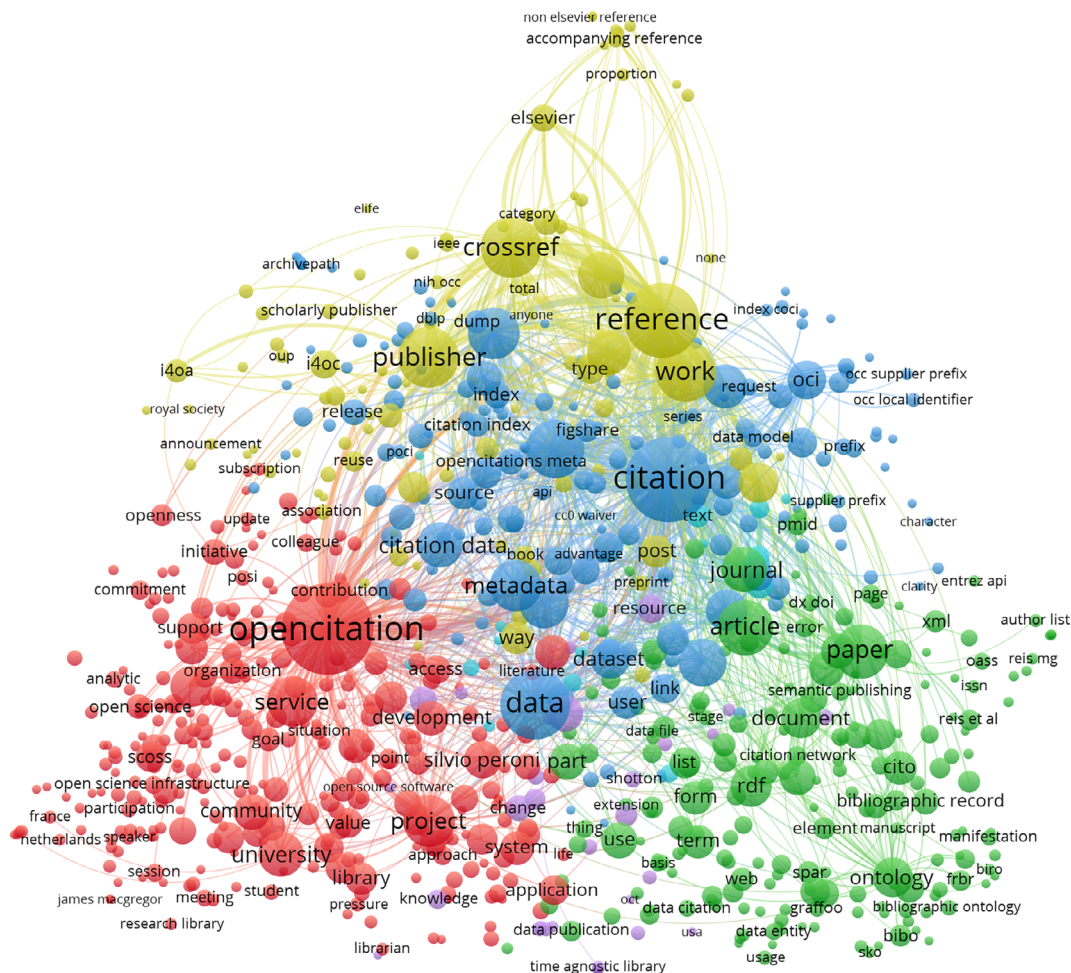


FIGURE 6 Term map of OpenCitation Blog posts (The interactive map is available at that link).

VOSviewer has identified six clusters. The largest cluster (294 keywords), shown in red, encompasses blogs discussing the services and infrastructures provided by OpenCitations. Additionally, it includes news about conferences, workshops, new collaborations, and project updates, all shared through the blog's influence. The average year for this cluster is 2019. The most common keywords in this cluster are *opencitations*, *project*, *service*, *university*, and *infrastructure*.

The second cluster (green, 202 keywords) highlights a critical aspect of open citations: semantic models, ontologies (such as SPAR, FaBiO, and CiTO), semantic infrastructures, Functional Requirements for Bibliographic Records (FRBR), semantic web standards, and Resource Description Framework (RDF). This cluster is significant because open citations have a technical dimension, and for those interested in the technical developments of open citations, the blog serves as an excellent resource for any technical needs or questions. The average year for blogs in this cluster is 2018. The most common keywords are *article*, *paper*, *ontology*, *RDF*, and *journal*.

Another significant cluster is the blue one, comprising 167 keywords. It encompasses APIs, indexes, corpuses, identifiers, data models, and datasets. This indicates that all new developments on the data side of OpenCitations are communicated to the public. The average year for this cluster is 2020, supporting this inference. The most frequent keywords are *citation*, *data*, *publication*, *COCI*, and *metadata*.

The yellow cluster, consisting of 97 keywords, highlights the ongoing tensions between publishers and open citation initiatives. It includes keywords related to major publishers (such as Elsevier, Taylor and Francis, American Chemical Society, IEEE, Wiley) as well as open citation corpuses like NIH OCC. The main keywords are *reference*, *CrossRef*, *work*, *publisher*, and *number*, with the average year for this cluster being 2019.

The final two clusters are very small and not easily noticeable on the map. One cluster consists of keywords related to the time-agnostic library and browser developed by Arcangelo Massari (software) and Silvio Peroni (supervision) (Massari & Peroni, 2022). The other cluster



pertains to the creation of the Open Biomedical Citations in Context Corpus, a 1-year project funded by the Wellcome Trust (Daquino, 2020).

These findings on a blog also highlight the necessity of diversifying research evaluation systems to encompass not only scholarly articles but also science communication pieces. Given that OpenCitations has relatively few scholarly outputs due to its 14-year history, it is crucial to recognize the value of alternative dissemination methods. The blog, in particular, serves as an effective platform for keeping the audience informed about all developments in this movement.

## 4 | DISCUSSION

The landscape of scholarly communication has undergone a significant transformation with the advent of open citations. The systematic review of the literature reveals a rich tapestry of initiatives, technological advancements, and scholarly discussions centered around the theme of open citations. This section synthesizes the findings from various studies, highlighting the evolution, impact, and future prospects of open citations.

### 4.1 | Evolution of open citations

The journey of open citations began with early efforts to make citation data freely accessible, with pivotal contributions from projects like OpenCitations and initiatives such as the Initiative for Open Citations (I4OC). These efforts were driven by the need to enhance transparency, reproducibility, and accessibility in academic research. Key milestones include the establishment of the OpenCitations Corpus, the launch of the I4OC, and significant policy endorsements from various stakeholders. The collaborative push from the scientometric community and advocacy from influential figures like David Shotton and Silvio Peroni has been instrumental in advancing the open citations movement.

### 4.2 | Impact on scholarly communication

Open citations have fundamentally altered the way bibliometric and scientometric analyses are conducted. By providing free access to citation data, they have democratized research, enabling scholars from diverse backgrounds to engage with scientific literature without the barriers imposed by subscription-based databases. The increased transparency and reproducibility of

research facilitated by open citations have fostered greater trust in scientific findings.

The creation of public citation graphs and enhanced discoverability of research outputs are notable benefits. Researchers can now follow citation trails more efficiently, uncovering relevant studies that might have otherwise remained hidden. This increased visibility not only aids individual researchers but also contributes to the overall progress of scientific fields by making connections between disparate areas of research more apparent.

### 4.3 | Technological advancements

The development and implementation of advanced data models, reference mining tools, and semantic technologies have been crucial in managing and leveraging open citation data. Tools and various knowledge graph embedding techniques have improved the accuracy and efficiency of citation analysis. These technologies facilitate the extraction, segmentation, and management of references, enabling more sophisticated and comprehensive bibliometric studies.

The integration of open citation data with other scholarly infrastructures and platforms has also been significant. Projects like the Linked Open Citation Database and the NIH Open Citation Collection exemplify the successful amalgamation of open citations with broader data ecosystems, enhancing the utility and reach of citation data.

### 4.4 | Challenges and limitations

Despite substantial progress, several challenges persist in the realm of open citations. Ensuring the quality and standardization of citation data remains a critical issue, as variations in citation formats, the presence of errors, and inconsistencies across databases hinder the effective use of open citation data. Ongoing efforts to improve data cleaning, standardization, and error correction are essential to address these issues.

Structural inequalities in citation networks also pose a significant challenge. The tendency to cite certain works more frequently than others reinforces existing biases and disparities in scholarly recognition. Furthermore, the adoption of open citations by all publishers is not yet universal. Some publishers remain hesitant to release citation data openly, highlighting the need for continued advocacy and policy interventions.

Achieving the full potential of open citations requires targeted policy measures and the establishment of universal standards, directly supporting the research goal of

enhancing accessibility. By addressing these challenges through coordinated global efforts, the scholarly community can create a more equitable and transparent research ecosystem.

## 4.5 | Future prospects

The future of open citations looks promising, with several avenues for further development and enhancement. Policy and advocacy efforts must continue to push for mandatory open citations, encouraging more publishers and researchers to embrace open data practices. Technological advancements will play a crucial role, with ongoing improvements in data mining, natural language processing, and semantic web technologies expected to further enhance the management and utilization of open citation data.

Global collaboration and the inclusion of multilingual data sources will be vital in creating a truly comprehensive and inclusive citation database. Educational initiatives aimed at raising awareness and promoting best practices in open citations will also be crucial in ensuring widespread adoption and effective use of open citation data.

## 5 | CONCLUSION

Open citations represent a transformative advancement in scholarly communication, offering numerous benefits that enhance the transparency, accessibility, and efficiency of scientific research. The findings of this review highlight the critical role of open citations in fostering an open and inclusive research ecosystem. However, realizing the full potential of open citations requires addressing challenges related to data quality, structural inequalities, and publisher adoption. Future efforts should focus on policy advocacy, technological innovation, global collaboration, and educational initiatives to promote the widespread adoption and effective use of open citations. By embracing these strategies, the academic community can ensure that the “frozen footprints” of scholarly communication are accessible to all, paving the way for a more equitable and transparent scientific landscape.

## FUNDING INFORMATION

This study was conducted as part of the “Sustainable Open Citation Data for Scientific Literature in Turkish” project, supported by the Scientific and Technological Research Council of Turkey’s (TUBITAK) Domestic Postdoctoral Fellowship Program 2218 and hosted by TUBITAK ULAKBİM (Project No: 121C401).

## CONFLICT OF INTEREST STATEMENT

The author declares no conflict of interest.

## ORCID

Zehra Taşkın  <https://orcid.org/0000-0001-7102-493X>

## ENDNOTES

- <sup>1</sup> <https://www.connectedpapers.com/>
- <sup>2</sup> <https://scite.ai/>
- <sup>3</sup> <https://opencitations.net/>
- <sup>4</sup> <https://i4oc.org/>
- <sup>5</sup> The timeline of milestones are summarized using OpenCitations and I4OC websites, OpenCitations Blog and “Citation Data Are Now Open, but That’s Far from Enough,” 2022; Peroni, 2018.
- <sup>6</sup> <https://barcelona-declaration.org/>
- <sup>7</sup> The Founder and Co-Director of OpenCitations, Prof. David M. Shotton, passed away on Saturday, 18th May, 2024 (Giambattista, 2024).
- <sup>8</sup> <https://opencitations.hypotheses.org/>
- <sup>9</sup> The network displays 867 terms out of a total of 9274, with each term appearing at least five times. Stop words and nonsensical terms (such as month names) have been removed from the network. Full counting method has been deployed.

## REFERENCES

- Alexiou, G., Vahdati, S., Lange, C., Papastefanatos, G., & Lohmann, S. (2016). OpenAIRE LOD services: Scholarly communication data as linked data. In A. González-Beltrán, F. Osborne, & S. Peroni (Eds.), *Semantics, analytics, visualization. Enhancing scholarly data* (pp. 45–50). Springer International Publishing. [https://doi.org/10.1007/978-3-319-53637-8\\_6](https://doi.org/10.1007/978-3-319-53637-8_6)
- Arnold, F., & Jäschke, R. (2022). A game with complex rules: Literature references in literary studies. In T. Backes, A. Iurshina, & P. Mayr (Eds.), *Proceedings of the workshop on understanding literature references in academic full text* (pp. 7–15). CEUR.
- Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, 1(1), 377–386. [https://doi.org/10.1162/qss\\_a\\_00019](https://doi.org/10.1162/qss_a_00019)
- Badolato, A.-M. (2018). National plan for Open Science. Ouvrir La Science. <https://www.ouvirlascience.fr/national-plan-for-open-science-4th-july-2018>
- Barcelona Declaration on Open Research Information. (2024). Barcelona Declaration on Open Research Information. <https://barcelona-declaration.org/>
- Beel, J., Gipp, B., Langer, S., & Breiteringer, C. (2016). Research-paper recommender systems: A literature survey. *International Journal on Digital Libraries*, 17(4), 305–338. <https://doi.org/10.1007/s00799-015-0156-0>
- Birkeneder, B., Aufvenne, P., Haase, C., Mayr, P., & Steinbrink, M. (2022). Extracting literature references in German speaking geography – The GEOcite project. In T. Backes, A. Iurshina, & P. Mayr (Eds.), *Proceedings of the workshop on understanding literature references in academic full text* (pp. 34–41). CEUR.

- Bologna, F., Di Iorio, A., Peroni, S., & Poggi, F. (2023). Do open citations give insights on the qualitative peer-review evaluation in research assessments? An analysis of the Italian National Scientific Qualification. *Scientometrics*, *128*(1), 19–53. <https://doi.org/10.1007/s11192-022-04581-6>
- Brandt, J. S., Downing, A. C., Howard, D. L., Kofinas, J. D., & Chasen, S. T. (2010). Citation classics in obstetrics and gynecology: The 100 most frequently cited journal articles in the last 50 years. *American Journal of Obstetrics and Gynecology*, *203*(4), 355.e1–355.e7. <https://doi.org/10.1016/j.ajog.2010.07.025>
- Brandt, J. S., Hadaya, O., Schuster, M., Rosen, T., Sauer, M. V., & Ananth, C. V. (2019). A bibliometric analysis of top-cited journal articles in obstetrics and gynecology. *JAMA Network Open*, *2*(12), e1918007. <https://doi.org/10.1001/jamanetworkopen.2019.18007>
- Chawla, D. S. (2021). Microsoft academic graph is being discontinued. What's next? *Nature News*. <https://www.nature.com/nature-index/news/microsoft-academic-graph-discontinued-whats-next>
- Chudlarsky, T., & Dvorak, J. (2019). Can crossref citations replace web of science for research evaluation? The share of open citations. In G. Catalano, C. Daraio, M. Gregori, H. F. Moed, & G. Ruocco (Eds.), *17th international conference on scientometrics & informetrics (ISSI2019)* (pp. 2551–2552). ISSI.
- Ciccarese, P., Wu, E., Wong, G., Ocana, M., Kinoshita, J., Ruttenberg, A., & Clark, T. (2008). The SWAN biomedical discourse ontology. *Journal of Biomedical Informatics*, *41*(5), 739–751. <https://doi.org/10.1016/j.jbi.2008.04.010>
- Citation. (2024). *Cambridge dictionary*. Cambridge University Press. <https://dictionary.cambridge.org/dictionary/english/citation>
- Costas, R., Zahedi, Z., & Wouters, P. (2015). Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, *66*(10), 2003–2019. <https://doi.org/10.1002/asi.23309>
- Cronin, B. (1981). The need for a theory of citing. *Journal of Documentation*, *37*(1), 16–24. <https://doi.org/10.1108/eb026703>
- Daquino, M. (2020). The open biomedical citations in context corpus: Progress report. OpenCitations Blog. <https://opencitations.wordpress.com/2020/01/27/the-open-biomedical-citations-in-context-corpus-progress-report/>
- Daquino, M., Heibi, I., Peroni, S., & Shotton, D. (2022). Creating RESTful APIs over SPARQL endpoints using RAMOSE. *Semantic Web*, *13*(2), 195–213. <https://doi.org/10.3233/SW-210439>
- Daquino, M., Peroni, S., Shotton, D., Colavizza, G., Ghavimi, B., Lauscher, A., Mayr, P., Romanello, M., & Zumstein, P. (2020). The OpenCitations data model. In G. Goos (Ed.), *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, 12507 LNCS (pp. 447–463). Springer International Publishing. [https://doi.org/10.1007/978-3-030-62466-8\\_28](https://doi.org/10.1007/978-3-030-62466-8_28)
- Di Iorio, A., Poggi, F., & Peroni, S. (2019). Open data to evaluate academic researchers: An experiment with the Italian Scientific Habilitation. In *International conference on scientometrics and informetrics* EdizioniEfes. (pp. 2133–2144).
- 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), ea00185. <https://doi.org/10.1126/science.a00185>
- Garfield, E. (1955). Citation indexes for science. *Science*, *122*(3159), 108–111. <https://doi.org/10.1126/science.122.3159.108>
- Garfield, E. (1970). Can citation indexing be automated? *Essays of an Information Scientist*, *1*, 84–90.
- Garfield, E. (1997). The concept of citation indexing: A unique and innovative tool for navigating the research literature. <https://garfield.library.upenn.edu/papers/vladivostok.html>
- Giambattista, C. D. (2024). Remembering David M. Shotton, Founder and Co-Director of OpenCitations [Billet]. OpenCitations Blog. <https://doi.org/10.58079/11seo>
- Giles, C. L., Bollacker, K. D., & Lawrence, S. (1998). CiteSeer: An automatic citation indexing system. In *Proceedings of the third ACM conference on digital libraries* (pp. 89–98). ACM Digital Library. <https://doi.org/10.1145/276675.276685>
- Goudsmith, S. A. (1974). Citation analysis. *Science*, *183*(4120), 28. <https://doi.org/10.1126/science.183.4120.28.b>
- Harzing, A.-W. (2016). Microsoft academic (search): A phoenix arisen from the ashes? *Scientometrics*, *108*(3), 1637–1647. <https://doi.org/10.1007/s11192-016-2026-y>
- Harzing, A.-W. (2019). Two new kids on the block: How do crossref and dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science? *Scientometrics*, *120*(1), 341–349. <https://doi.org/10.1007/s11192-019-03114-y>
- Harzing, A.-W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, *106*(2), 787–804. <https://doi.org/10.1007/s11192-015-1798-9>
- Harzing, A.-W., & Alakangas, S. (2017). Microsoft Academic: Is the phoenix getting wings? *Scientometrics*, *110*(1), 371–383. <https://doi.org/10.1007/s11192-016-2185-x>
- Heibi, I. (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, ISSI 2019—proceedings*. OSF. <https://cris.unibo.it/handle/11585/739946>
- Heibi, I., Peroni, S., & Shotton, D. (2019a). Enabling text search on SPARQL endpoints through OSCAR. *Data Science*, *2*(1-2), 205–227. <https://doi.org/10.3233/DS-190016>
- Heibi, I., Peroni, S., & Shotton, D. (2019b). Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. *Scientometrics*, *121*(2), 1213–1228. <https://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*(1), 414–427. [https://doi.org/10.1162/qss\\_a\\_00022](https://doi.org/10.1162/qss_a_00022)
- Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies*, *1*(1), 387–395. [https://doi.org/10.1162/qss\\_a\\_00020](https://doi.org/10.1162/qss_a_00020)
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, *102*(46), 16569–16572. <https://doi.org/10.1073/pnas.0507655102>
- Hosseini, A., Ghavimi, B., Boukhers, Z., & Mayr, P. (2019). EXCITE—A toolchain to extract, match and publish open literature references. In M. Bonn, D. Wu, S. J. Downie, & A.

- Martaus (Eds.), 2019 *ACM/IEEE joint conference on digital libraries (JCDL 2019)* IEEE (pp. 432–433). <https://doi.org/10.1109/JCDL.2019.00105>
- Hug, S. E., & Brändle, M. P. (2017). The coverage of Microsoft Academic: Analyzing the publication output of a university. *Scientometrics*, 113(3), 1551–1571. <https://doi.org/10.1007/s11192-017-2535-3>
- Hutchins, B. I. (2021). A tipping point for open citation data. *Quantitative Science Studies*, 2(2), 433–437. [https://doi.org/10.1162/qss\\_c\\_00138](https://doi.org/10.1162/qss_c_00138)
- 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), e3000385. <https://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), e3000416. <https://doi.org/10.1371/journal.pbio.3000416>
- Initiative for Open Citations (I4OC). (2020). <https://i4oc.org/>
- Ioannidis, J. P. A., & Maniatis, Z. (2024). Quantitative research assessment: Using metrics against gamed metrics. *Internal and Emergency Medicine*, 19(1), 39–47. <https://doi.org/10.1007/s11739-023-03447-w>
- Jiang, Y., & Liu, X. (2023). A construction and empirical research of the journal disruption index based on open citation data. *Scientometrics*, 128(7), 3935–3958. <https://doi.org/10.1007/s11192-023-04737-y>
- Kaplan, N. (1965). The norms of citation behavior: Prolegomena to the footnote. *American Documentation*, 16(3), 179–184. <https://doi.org/10.1002/asi.5090160305>
- Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10–25. <https://doi.org/10.1002/asi.5090140103>
- Larivière, V. (2019). *Resignation of the editorial board of the Journal of Informetrics*. International Society for Scientometrics and Informetrics. <https://www.issi-society.org/blog/posts/2019/january/resignation-of-the-editorial-board-of-the-journal-of-informetrics/>
- Latour, B. (1988). *Science in action: How to follow scientists and engineers through society (Revised edition)*. Harvard University Press.
- Lauscher, A., Eckert, K., Galke, L., Scherp, A., Rizvi, S. T. R., Ahmed, S., Dengel, A., Zumstein, P., & Klein, A. (2018). Linked open citation database: Enabling libraries to contribute to an open and interconnected citation graph. In *JCDL'18: Proceedings of the 18th ACM/IEEE joint conference on digital libraries* (pp. 109–118). Assoc Computing Machinery. <https://doi.org/10.1145/3197026.3197050>
- Malínek, V., Umerle, T., Gray, E., Heibi, I., Király, P., Klaes, C., Korytkowski, P., Lindemann, D., Moretti, A., Panušková, C., Péter, R., Tolonen, M., Tomczyńska, A., & Vimr, O. (2024). Open bibliographical data workflows and the multilinguality challenge. *Journal of Open Humanities Data*, 10, 1–14. <https://doi.org/10.5334/johd.190>
- Maricato, J. D. M., Regina Da Silva, M., & Ramos, T. B. C. (2023). Coverage and correlations between open citations in crossref and readership in Mendeley: Different fields of Brazilian science. *Journal of Scientometric Research*, 12(1), 176–186. <https://doi.org/10.5530/jscires.12.1.015>
- 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. <https://doi.org/10.1016/j.joi.2018.09.002>
- Martin-Martín, A., Thelwall, M., Orduna-Malea, E., & Delgado Lopez-Cozar, E. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations. *Scientometrics*, 126(1), 871–906. <https://doi.org/10.1007/s11192-020-03690-4>
- Massari, A., & Peroni, S. (2022). Time-agnostic-library [Computer software]. <https://cris.unibo.it/handle/11585/896040>
- Merton, R.K. (1968). The Matthew Effect in science. *Science*, 159, 56–63.
- Mitra, A. N., Aurora, N., Grover, S., Ananth, C., & Brandt, J. S. (2021). A bibliometric analysis of obstetrics and gynecology articles with highest relative citation ratios, 1980 to 2019. *American Journal of Obstetrics & Gynecology MFM*, 3(1), 100293. <https://doi.org/10.1016/j.ajogmf.2020.100293>
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- Citation data are now open, but that's far from enough. (2022). *Nature*, 609(7927), 441. <https://doi.org/10.1038/d41586-022-02915-1>
- Nettasinghe, B., Alipourfard, N., Krishnamurthy, V., & Lerman, K. (2021). Emergence of structural inequalities in scientific citation networks. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.2103.10944>
- Nuzzolese, A. G., Gentile, A. L., Presutti, V., & Gangemi, A. (2016). Conference linked data: The scholarly data project. In P. Groth, E. Simperl, A. Gray, M. Sabou, M. Krötzsch, F. Lecue, F. Flöck, & Y. Gil (Eds.), *The Semantic Web – ISWC 2016* (pp. 150–158). Springer International Publishing. [https://doi.org/10.1007/978-3-319-46547-0\\_16](https://doi.org/10.1007/978-3-319-46547-0_16)
- OpenAlex support. (2024). About us. OpenAlex. <https://help.openalex.org/hc/en-us/articles/24396686889751-About-us>
- OpenCitations.net. (n.d.). Welcome to the OpenCitations homepage! <https://opencitations.net/>
- Ortega, J. L. (2021). Open citations movement and its role in the transformation of research evaluation. *Arbor-Ciencia Pensamiento Y Cultura*, 197(799), a592. <https://doi.org/10.3989/arbor.2021.799007>
- Peroni, S. (2018). The open citations movement: The story so far [Presentation]. figshare. <https://doi.org/10.6084/m9.figshare.7110653.v3>
- Peroni, S., Dutton, A., Gray, T., & Shotton, D. (2015). Setting our bibliographic references free: Towards open citation data. *Journal of Documentation*, 71(2), 253–277. <https://doi.org/10.1108/JD-12-2013-0166>
- Peroni, S., & Shotton, D. (2012). FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. *Journal of Web Semantics*, 17, 33–43. <https://doi.org/10.1016/j.websem.2012.08.001>
- Peroni, S., & Shotton, D. (2018a). Open citation: Definition. 95436 bytes. <https://doi.org/10.6084/M9.FIGSHARE.6683855>



- Peroni, S., & Shotton, D. (2018b). The SPAR ontologies. In D. Vrandečić, K. Bontcheva, M. C. Suárez-Figueroa, V. Presutti, I. Celino, M. Sabou, L.-A. Kaffee, & E. Simperl (Eds.), *The Semantic Web – ISWC 2018* (pp. 119–136). Springer International Publishing. [https://doi.org/10.1007/978-3-030-00668-6\\_8](https://doi.org/10.1007/978-3-030-00668-6_8)
- Peroni, S., & Shotton, D. (2020). OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies*, 1(1), 428–444. [https://doi.org/10.1162/qss\\_a\\_00023](https://doi.org/10.1162/qss_a_00023)
- Peroni, S., Shotton, D., & Vitali, F. (2017). One year of the OpenCitations corpus releasing RDF-based scholarly citation data into the public domain. In C. Damato, M. Fernandez, V. Tamma, F. Lecue, P. CudreMauroux, J. Sequeda, C. Lange, & J. Heflin (Eds.), *Semantic Web—ISWC* (pp. 184–192). Springer International Publishing. [https://doi.org/10.1007/978-3-319-68204-4\\_19](https://doi.org/10.1007/978-3-319-68204-4_19)
- Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175. <https://doi.org/10.7717/peerj.175>
- Price, D. J. (1963). *Little science, big science*. Columbia University Press.
- Santini, C., Gesese, G. A., Peroni, S., Gangemi, A., Sack, H., & Alam, M. (2022). A knowledge graph embeddings based approach for author name disambiguation using literals. *Scientometrics*, 127(8), 4887–4912. <https://doi.org/10.1007/s11192-022-04426-2>
- Shotton, D. (2010). CiTO, the Citation typing ontology. *Journal of Biomedical Semantics*, 1(1), S6. <https://doi.org/10.1186/2041-1480-1-S1-S6>
- Shotton, D. (2013). Open citations. *Nature*, 502(7471), 295–297. <https://doi.org/10.1038/502295a>
- Shotton, D. (2018). Funders should mandate open citations. *Nature*, 553(7687), 129. <https://doi.org/10.1038/d41586-018-00104-7>
- Shotton, D. (2021). From little acorns. A retrospective on OpenCitations [Billet]. OpenCitations Blog. <https://doi.org/10.58079/sjek>
- Shotton, D., Portwin, K., Klyne, G., & Miles, A. (2009). Adventures in semantic publishing: Exemplar semantic enhancements of a research article. *PLoS Computational Biology*, 5(4), e1000361. <https://doi.org/10.1371/journal.pcbi.1000361>
- Simkin, M., & Roychowdhury, V. (2006). Do you sincerely want to be cited? Or: Read before you cite. *Significance*, 3(4), 179–181. <https://doi.org/10.1111/j.1740-9713.2006.00202.x>
- 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.
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265–269. <https://doi.org/10.1002/asi.4630240406>
- Smith, L. (1981). Citation analysis. *Library Trends*, 30, 83–106.
- Sokolov, M. (2023). The art of ignoring others' work among academics: A guessing game model of scholarly information search. *Social Studies of Science*, 53(2), 300–312. <https://doi.org/10.1177/03063127221119808>
- Sugimoto, C. R., Waltman, L., Larivière, V., van Eck, N. J., Boyack, K. W., Wouters, P., & de Rijcke, S. (2017). Open citations: A letter from the scientometric community to scholarly publishers. *International Society for Scientometrics and Informetrics*.
- Tagliacozzo, R. (1967). Citations and Citation indexes: A review. *Methods of Information in Medicine*, 6(3), 136–142. <https://doi.org/10.1055/s-0038-1636363>
- Taraborelli, D. (2017). Unlocking citations from tens of millions of scholarly papers. figshare. [https://figshare.com/articles/journal\\_contribution/Unlocking\\_citations\\_from\\_tens\\_of\\_millions\\_of\\_scholarly\\_papers/5674486/1](https://figshare.com/articles/journal_contribution/Unlocking_citations_from_tens_of_millions_of_scholarly_papers/5674486/1)
- Taraborelli, D. (2018). Remixing the graph [Presentation]. figshare. <https://doi.org/10.6084/m9.figshare.7038695.v3>
- Taşkın, Z., & Al, U. (2018). A content-based citation analysis study based on text categorization. *Scientometrics*, 114(1), 335–357. <https://doi.org/10.1007/s11192-017-2560-2>
- Teufel, S., Siddharthan, A., & Tidhar, D. (2006). Automatic classification of citation function. In *Proceedings of the 2006 conference on empirical methods in natural language processing* (pp. 103–110). ACL.
- The I4OC Team. (2017). An open letter to stakeholders of the initiative for open citations. *eLife*.
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLoS One*, 8(5), e64841. <https://doi.org/10.1371/journal.pone.0064841>
- Van Der Vet, P. E., & Nijveen, H. (2016). Propagation of errors in citation networks: A study involving the entire citation network of a widely cited paper published in, and later retracted from, the journal *Nature*. *Research Integrity and Peer Review*, 1(1), 3. <https://doi.org/10.1186/s41073-016-0008-5>
- Van Eck, N. J., Waltman, L., Larivière, V., & Sugimoto, C. R. (2018). Crossref as a new source of citation data: A comparison with Web of Science and Scopus. *CWTS*.
- Velez-Estevez, A., Perez, I. J., Garcia-Sanchez, P., Moral-Munoz, J. A., & Cobo, M. J. (2023). New trends in bibliometric APIs: A comparative analysis. *Information Processing & Management*, 60(4), 103385. <https://doi.org/10.1016/j.ipm.2023.103385>
- Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2(1), 20–41. [https://doi.org/10.1162/qss\\_a\\_00112](https://doi.org/10.1162/qss_a_00112)
- Waltman, L. (2021). Q&A about Elsevier's decision to open its citations. <https://leidenmadtrics.nl/articles/q-a-about-elseviers-decision-to-open-its-citation>
- Wang, K., Shen, Z., Huang, C., Wu, C.-H., Dong, Y., & Kanakia, A. (2020). Microsoft academic graph: When experts are not enough. *Quantitative Science Studies*, 1(1), 396–413. [https://doi.org/10.1162/qss\\_a\\_00021](https://doi.org/10.1162/qss_a_00021)
- Weinstock, M. (1971). Citation indexes. *Essays of an Information Scientist*, 1, 188–195.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Ziman, J. M. (1968). *Public knowledge: An essay concerning the social dimension of science*. Cambridge University Press.

**How to cite this article:** Taşkın, Z. (2025). Sustaining the “frozen footprints” of scholarly communication through open citations. *Journal of the Association for Information Science and Technology*, 1–17. <https://doi.org/10.1002/asi.24982>