

# Summary Report of the OpenCitations Community Survey 2025

Chiara Di Giambattista; Silvio Peroni

## Introduction

OpenCitations is a not-for profit, community-driven open science infrastructure dedicated to the publication and dissemination of open bibliographic metadata and citation data. For OpenCitations, the research community plays a fundamental role: it contributes to data curation and provision, participates in shared governance processes, and supports the initiative financially.

To better understand the community's needs and identify barriers to adoption and community-informed priorities for future development, OpenCitations conducted an online survey between 20 October and 16 December 2025. The survey aimed to inform the OpenCitations' roadmap by collecting feedback on awareness, usage patterns, perceived challenges, and expectations within the scholarly communication ecosystem.

## Survey Design and Approach

The survey was conducted using an online questionnaire and distributed through academic and community channels. It remained open for responses for a total of 58 days in 2025 and received 55 completed responses.

The survey targeted a broad audience, including existing users of OpenCitations services, individuals who were aware of OpenCitations but had not previously engaged with it, and respondents encountering OpenCitations for the first time through the survey itself.

Responses were collected anonymously via Microsoft Forms through a questionnaire composed of a mix of multiple-choice and open-ended questions. The survey explored four main areas: (i) About You, (ii) Usage of OpenCitations, (iii) Engagement, and (iv) Open Feedback.

The analysis presented in this report is primarily based on a thematic interpretation of both structured responses and qualitative feedback. For open-ended questions, responses were examined to identify common themes, recurring concerns, and converging suggestions. These themes are synthesised narratively in the sections that follow, without reproducing individual responses or detailed numerical distributions. Any potentially identifying information provided by respondents has been excluded from this report.

The following questions were asked:

### Section 1: About You

The first section collected background information about the respondents.

1. What is your primary role?
2. What is your affiliation?

Both questions were required and designed as single-choice questions.

### Section 2: Use of OpenCitations

The second section focused on awareness, use, and perceptions of OpenCitations.

3. How familiar are you with OpenCitations?
4. What do you value the most about OpenCitations?

5. What OpenCitations services are you familiar with?
6. Have you ever used OpenCitations services in your research activities, projects, or professional work?

Questions 3 and 6 were single-choice questions. Questions 4 and 5 allowed multiple selections.

From this point, the questionnaire branched into two different paths depending on the response to Question 6.

#### Path A: Respondents Who Have Used OpenCitations (Users)

Participants who indicated that they had used OpenCitations services were presented with the following questions:

7. What do you use OpenCitations for?
8. What motivated you to use OpenCitations?
9. What are the biggest challenges or limitations you face when using OpenCitations?

These questions were open-text questions aimed at capturing qualitative descriptions of use cases, motivations, and challenges.

All respondents were then asked to complete the following ranking question:

10. Please rank the following possible future developments in OpenCitations from most urgent to least urgent:
  - *Including authors' affiliation information*
  - *Including funding information*
  - *Extending the coverage of bibliographic resources, including those not involved in citations*
  - *Extending the API operations by adding textual search functionality*
  - *Preparing tutorials to show how to use the APIs and other services*
  - *Developing a more robust web search system and browsing capabilities*
  - *Distributing the technical infrastructure on multiple nodes*

After the ranking question, respondents were invited to provide additional input:

11. Are there other types of data, integrations, and services you would like OpenCitations to provide in the future?

Question 11 was an open-text question.

#### Path B: Respondents Who Have Not Used OpenCitations (Non-Users)

Participants who indicated that they had not used OpenCitations services were presented with the following questions:

12. Why haven't you used OpenCitations so far?
13. Are there any other major challenges or limitations that prevent the use of OpenCitations?
14. What would make you more likely to try OpenCitations in the future?

Question 12 and Question 14 were single-choice questions, while Question 13 allowed an open-text response.

15. Please specify.

Question 15 was presented as a follow-up open-text question to allow respondents to elaborate on their previous answer concerning the alternative tools they use.

### Section 3: Engagement

The third section explored respondents' awareness of OpenCitations and their level of engagement with the project.

16. How did you first hear of OpenCitations?
17. How would you describe your current level of engagement with OpenCitations?
18. Would you like to become more involved with OpenCitations? If yes, in what capacity?
19. What motivates you (or would motivate you) to engage more deeply with OpenCitations?
20. How would you prefer to engage in discussions with the OpenCitations team and community?
21. What kind of communications would you find most valuable?

Questions in this section were in single-choice and multiple-choice formats.

### Section 4: Open Feedback

The final section consisted of a single open-ended question:

22. Do you have any additional comments, suggestions, or thoughts about OpenCitations' current status and future direction?

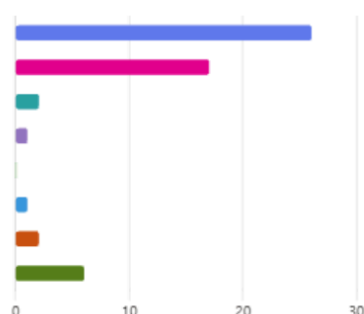
Microsoft Forms provided charts and word clouds for the responses, which are included as figures in the analysis.

## Section 1: About You

The survey respondents are primarily drawn from the **academic and research ecosystem**. Most participants identify as **researchers** (26) or **librarians/information specialists** (17), with smaller representations of scientometricians, software developers, students, and other roles; no policy makers participated, and only one respondent identified as a publisher. In terms of affiliation, the majority are based at **universities or academic institutions** (36), followed by infrastructure providers (5), libraries and research institutes (3 each), with limited representation from government bodies and commercial companies. Familiarity with OpenCitations varies: while 17 respondents have already used OpenCitations data or services, **a substantial proportion had heard of it but never used it** (26), and 12 encountered OpenCitations for the first time through the survey. Respondents most value the **openness** of OpenCitations services, along with transparency and reproducibility, absence of commercial influence, and its community-driven approach, while data coverage is less frequently cited. **Awareness of OpenCitations services is uneven:** bibliographic and citation databases are the most widely recognised, whereas more technical components such as SPARQL endpoints and the data model are familiar to a smaller subset of respondents, and 17 participants report no familiarity with any specific service.

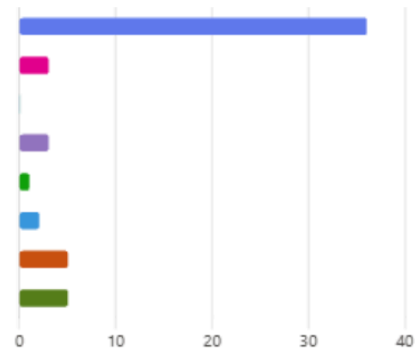
### 1. What is your primary role?

Researcher	26
Librarian/Information Specialist	17
Scientometrician	2
Software Developer	1
Policy Maker	0
Publisher	1
Student	2
Other	6



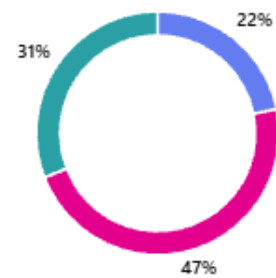
## 2. What is your affiliation?

University or Academic Institution	36
Library	3
Funding agency	0
Research institute	3
Government body	1
Commercial company	2
Infrastructure provider	5
Other	5



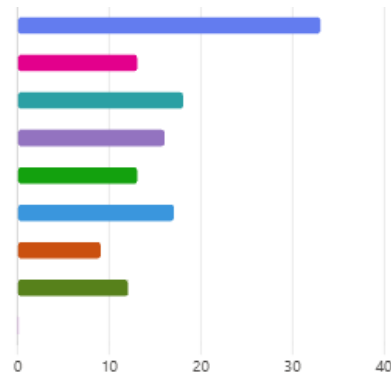
## 3. How familiar are you with OpenCitations?

First time hearing of it	12
Heard of it before this survey but never used	26
I knew it and I have used its data and services	17



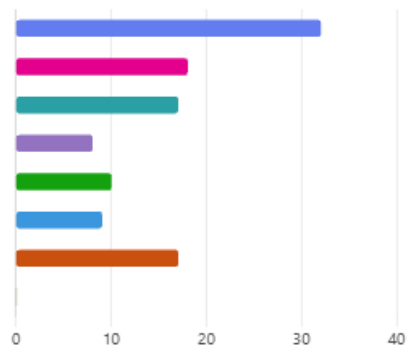
## 4. What do you value the most about OpenCitations? (select up to 3)

Openness of its services	33
Data quality and reliability	13
Transparency and reproducibility	18
Community-driven approach	16
Integration with Open Science services	13
No commercial influence	17
Data coverage	9
I don't know	12
Other	0



## 5. What OpenCitations services are you familiar with?

Bibliographic and citation databases (OpenCitations Meta and OpenCitations Index)	32
Data dumps	18
REST APIs	17
SPARQL endpoints	8
OpenCitations Data Model and SPAR Ontologies	10
Search interfaces	9
None	17
Other	0



## Section 2: Technical Usage

25% of participants reported that they have used OpenCitations in their research, while **the majority rely on alternative services or have not engaged due to technical barriers or limited coverage**. Among current users, OpenCitations is primarily valued for its reliability, openness, and integration within the open science ecosystem, particularly for citation tracking and metrics. Non-users cited the **need for easier access, improved documentation, and broader coverage as key factors that would encourage future adoption**.

Looking forward, respondents identified **expanding bibliographic coverage, including author affiliations and funding information, and enhancing APIs and web interfaces as the most urgent priorities**. Additional suggestions included making datasets more rapidly available and improving interoperability with other tools. What emerges is a strong appreciation for the project's openness and community-driven approach, while also pointing to clear opportunities to increase awareness, usability, and integration to broaden adoption.

### 6. Have you ever used OpenCitations services in your research activities/ projects/ job?



After this initial data collection section, the survey branched into two distinct paths, allowing respondents to indicate whether they were users or non-users of OpenCitations services, and presenting each group with a set of questions specific to their respective experiences.

### PATH A: RESPONSES FROM USERS (14)

#### 7. What do you use OpenCitations for?



OpenCitations is most commonly used **as a data source**. Many respondents described using it to feed citation data into other systems, such as repositories, dashboards, discovery services, and research infrastructures, including initiatives like OpenAIRE and Wikidata. In several cases, **this use is indirect**: users may not interact with OpenCitations services themselves but rely on its data as it is reused and redistributed through third-party platforms. Beyond this infrastructural role, OpenCitations is also used for **citation analysis and metrics**, supporting bibliometric studies, citation statistics, article-level indicators, and the development of analytical dashboards. In this context, it is widely perceived as an **open and transparent alternative to proprietary citation databases**. Finally, some respondents highlighted more research-oriented and exploratory uses, such as tracing citations to their own work or to influential publications, comparing scholarly outputs, building knowledge graphs, and supporting editorial workflows, for example, by identifying potential reviewers through citation patterns rather than co-authorship networks.

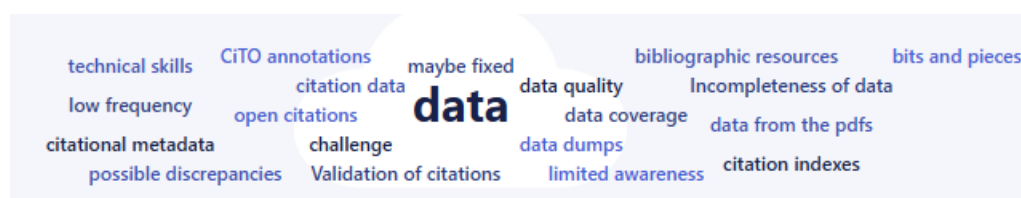
## 8. What motivated you to use OpenCitations?



Respondents value **openness**. **Reliability and trust** in the data and the team behind it are also highlighted.

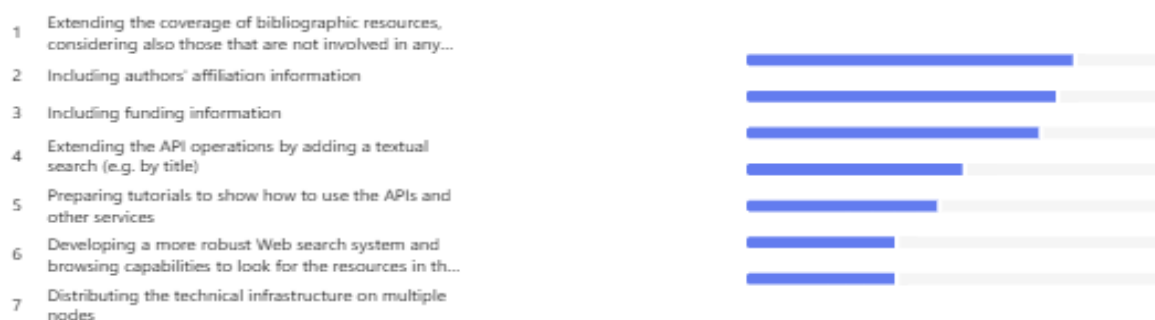
Some mention **technical accessibility**, such as APIs, SPARQL endpoints, and integration with external tools, while a few note practical use cases, such as finding relevant publications. Overall, OpenCitations is seen as a **trusted, researcher-driven, open infrastructure** aligned with open science principles.

## 9. What are the biggest challenges and limitations in using OpenCitations?



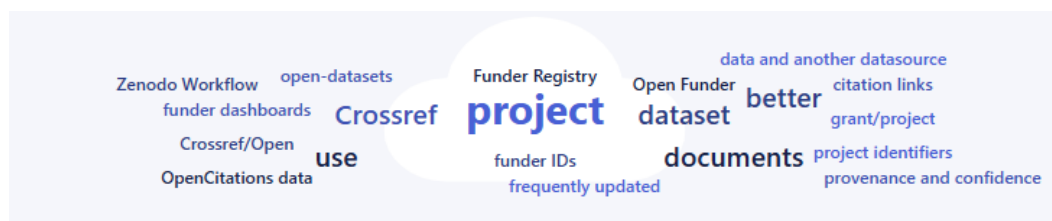
Respondents identified a set of recurring challenges that affect how OpenCitations is used and perceived. **Data coverage** emerged as a central concern: gaps in coverage, delays in data availability, and missing information limit the usefulness of the data for certain analytical and operational use cases. Closely related to this is **the issue of data quality and validation**, with some users pointing to discrepancies, duplicate records, and the difficulty of assessing the reliability of specific citation links. **Update frequency** was also mentioned, particularly in relation to data dumps, which some respondents felt are not refreshed often enough to support time-sensitive analyses. In addition, several participants highlighted **technical complexity as a barrier**, noting that non-technical users may struggle with the available interfaces and documentation. Finally, a more **cultural and institutional challenge** was reported: persuading colleagues, managers, or decision-makers to adopt open citation data instead of established commercial tools can be difficult, even when the principles of openness and transparency are valued.

## 10. Please rank the following possible future developments in OpenCitations from most urgent to least urgent:



Overall, the community emphasizes **better metadata coverage and improved access tools** as the main areas for OpenCitations' development.

11. Are there other types of data, integrations, and services you would like OpenCitations to provide in the future?



According to the responses, participants would like **richer semantic information around citations**, including support for CiTO intent annotations and clearer modelling of relationships between different research outputs (for example, from preprints to journal articles, or links to associated datasets). There is also interest in **extending coverage beyond the DOI ecosystem**, provided that such links are accompanied by transparent provenance and confidence information.

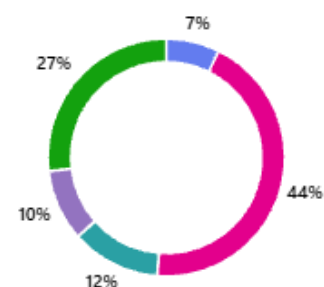
**Integration and interoperability** emerge as equally important. Respondents highlighted the value of incorporating persistent identifiers for funders and institutions (such as the Open Funder Registry and ROR) and of strengthening links with existing infrastructures like Crossref, OpenAIRE, and OpenAlex. These integrations should be supported by concise documentation, examples, and reusable scripts that make the combined use of multiple data sources easier.

Finally, several comments focus on **usability and workflows**. Users asked for simpler data contribution processes and more timely access to updated datasets, including distribution through platforms such as BigQuery to enable faster integration with services like ORCID and Dimensions.

## PATH B: RESPONSES FROM NON-USERS (41)

### 12. Why haven't you used OpenCitations so far?

I don't need bibliographic data for my work	3
I use alternative services/tools	18
Technical barriers	5
Data coverage seems insufficient	4
Other	11



The other responses mostly refer to a lack of awareness of OpenCitations.

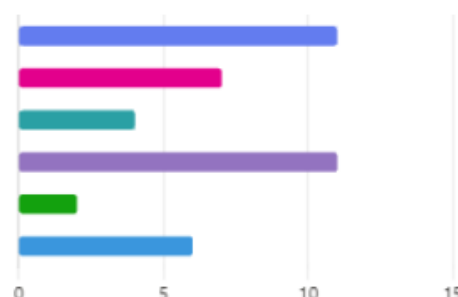
### 13. Are there any other major challenges or limitations that prevent the use of OpenCitations?



Based on the responses, several factors emerge as barriers to the use of OpenCitations. A recurring issue concerns missing or incomplete bibliographic metadata, particularly **the absence of book data and key fields such as author affiliations**. Limited familiarity with the service also plays a significant role: some respondents are **unfamiliar** with the data model, have never used OpenCitations, or feel that its advantages over other open bibliographic databases are not clearly communicated. **Usability and access issues** were mentioned as well, including the lack of an intuitive online interface for exploring citations and the effort required to clean data after download. **Technical barriers**, such as insufficient technological skills, further limit adoption. In some cases, **external constraints** apply, for example institutional requirements to use proprietary tools like Web of Science, or perceptions that OpenCitations does not adequately cover self-published but peer-reviewed content. Finally, a minority of respondents reported no particular issues or considered the question not applicable to their situation.

### 14. What would make you more likely to try OpenCitations in the future?

Easier access (e.g., web interface, tutorials)	11
Wider coverage	7
Integration with tools I already use	4
Training materials / documentation	11
Outreach through my institution	2
Other	6



4 respondents suggested additional tools to be integrated (OpenAlex, Python tools and Zotero) through **question 15 "Please specify"**, while the 6 "Other" respondents specified more comprehensive metadata; toolkits targeting multiple parties and easier access to training materials.

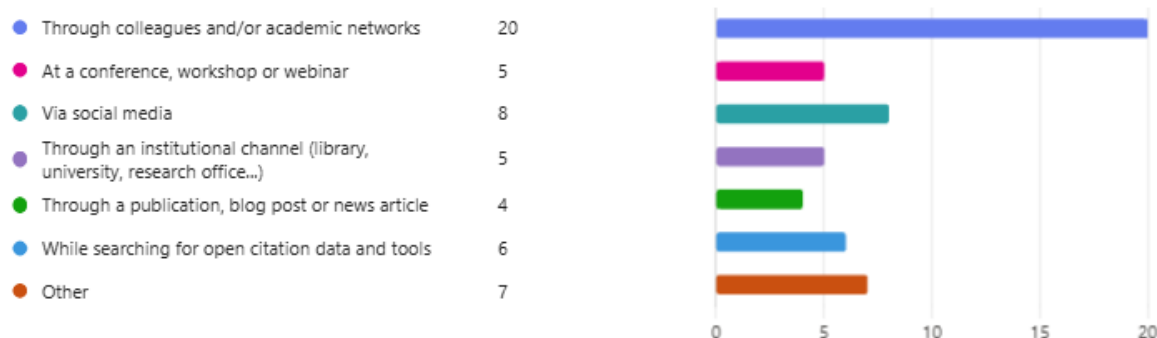


## Section 3: Community and Outreach

Most respondents first heard about OpenCitations through **colleagues or academic networks** (20 participants), followed by **social media** (8), while searching for citation tools (6), conferences or workshops (5), institutional channels (5), and publications or blogs (4). Seven respondents mentioned other sources. Regarding engagement, 21 participants know OpenCitations but have not yet engaged, while 11 follow updates passively. Active engagement is less common, with 6 involved in partner projects, 4 actively engaging with the community, and 1 part of the financial support network. When asked about potential future involvement, 27 participants prefer not to engage more deeply, while 13 would contribute data, 15 would act as testers, 6 as ambassadors/advocates, and 2 as partner institutions, with 6 other responses.

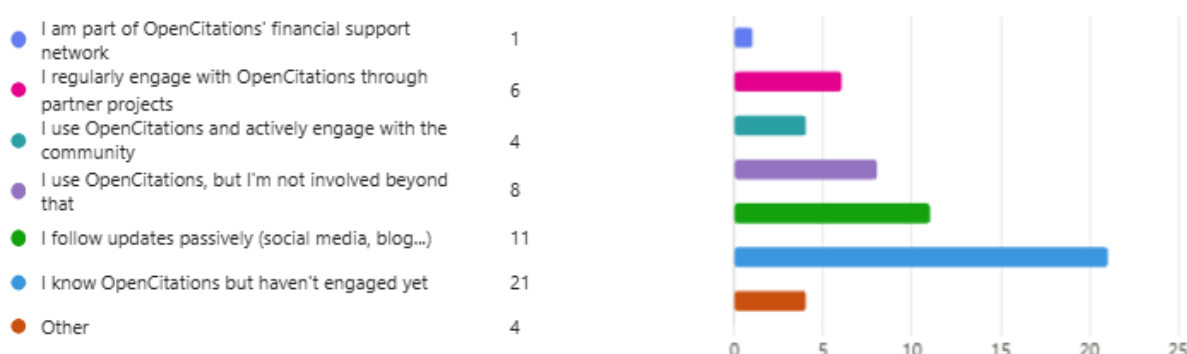
The main motivation for deeper involvement is contributing to open science and the public good (41 participants), followed by learning new skills (27), alignment with institutional priorities (18), access to additional resources or early releases (16), and professional networking opportunities (12). For communication, respondents **prefer mailing lists** (29 participants), **public GitHub issues** (24), and **regular webinars** (24), while in-person events (13) and real-time chat (7) are less popular. Most participants value updates on technical tutorials (32), release notes (31), roadmap updates (30), community use cases (30), as well as policy updates (11) and job/collaboration opportunities (13).

### 16. How did you first hear of OpenCitations?



The remaining responses fall into a residual “Other” category and mainly reflect less clearly defined discovery paths. These include respondents who do not remember how they first encountered OpenCitations, cases of indirect or highly technical discovery (for example, through documentation, SPARQL services, or platforms such as GraspOS), and a small number of answers that did not fit the predefined options, including one explicitly stating that ChatGPT was used.

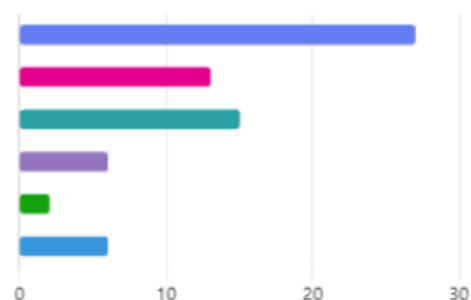
### 17. How would you describe your current level of engagement with OpenCitations?



In “other”, respondents stated they have no relationship, collaboration, or engagement with OpenCitations in any capacity.

### 18. Would you like to become more involved with OpenCitations? If yes, in what capacity? (check all that apply)

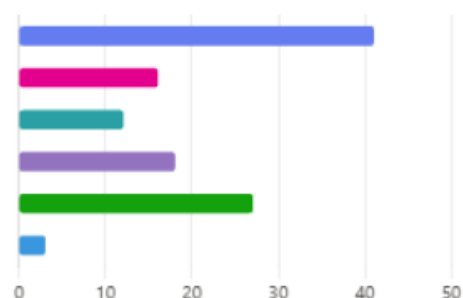
● I prefer not to engage more deeply	27
● As a data contributor	13
● As a tester	15
● As an ambassador/advocate	6
● As a partner institution	2
● Other	6



In “other”, some people want to be involved as active users, while four people are unsure about further engagement.

### 19. What motivates you (or would motivate you) to engage more deeply with OpenCitations?

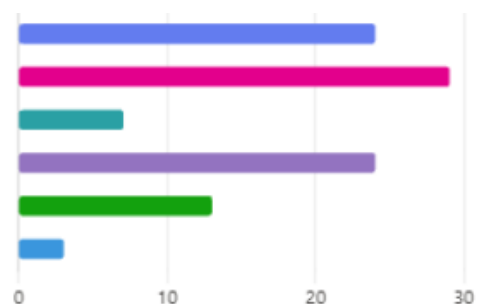
● Contributing to open science and the public good	41
● Access to additional resources, training, or early releases	16
● Professional networking opportunities	12
● Alignment with institutional/organizational priorities	18
● Learning new skills (technical, bibliometrics, data science)	27
● Other	3



The responses in “other” are those that do not express a motivation, but rather uncertainty or demotivation,

### 20. How would you prefer to engage in discussions with the OpenCitations team and community?

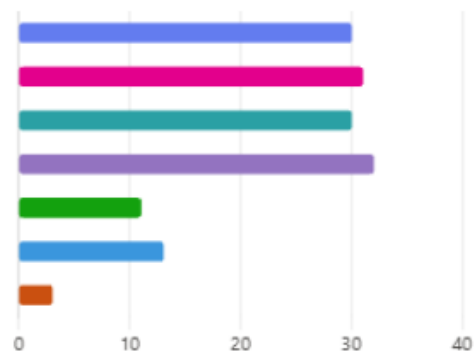
● Public GitHub issues	24
● Mailing list or Google Group	29
● Real-time chat (Discord)	7
● Regular webinars	24
● In-presence events	13
● Other	3



The “other” respondents refer to a lack of time.

### 21. What kind of communications would you find most valuable?

● Roadmap updates (progress and future plans)	30
● Release notes (new datasets, tools, features)	31
● Use cases from the community	30
● Technical tutorials and documentation	32
● Policy and advocacy updates	11
● Job / collaboration opportunities	13
● Other	3



Among the “other” responses, one participant mentions “disruptive use cases from the community”.

## Section 4: Open-ended feedback (optional)

The majority of the 15 respondents express **strong appreciation and support** for OpenCitations, repeatedly describing it as a *valuable, unique, and essential public infrastructure* for open science. Several comments explicitly thank the team and encourage them to continue their work.

Alongside this positive feedback, some **constructive suggestions and concerns** emerge:

- **Usability and accessibility:** multiple respondents highlight the need for a more **user-friendly web interface** and a **lower barrier to entry**, noting that OpenCitations can be difficult to start using.
- **Data expansion and integration:** there are calls to **expand data coverage** and improve **linkages with funder and institutional identifiers**, as well as interest in integrating or transferring citation data from platforms such as Wikidata.
- **Engagement and awareness:** a few respondents note that they are still **new to OpenCitations** or not yet active users, but are interested in becoming more engaged and learning more about its activities.

A small number of responses indicate **no additional comments** or provide external references without further explanation.

## Next steps

The OpenCitations Community Survey represents a starting point for a more systematic dialogue with the community. Based on the insights collected, the next phase will focus on **translating feedback into concrete actions**, while maintaining a realistic alignment with available resources and ongoing development priorities.

In the short term, the survey results are being analysed and discussed internally to identify areas where immediate improvements are feasible, particularly concerning **usability, documentation, and access to services**. These aspects were repeatedly highlighted as barriers to adoption and will be considered alongside existing technical and strategic constraints.

In the medium term, the results will model the OpenCitations' [roadmap](#). In particular, feedback related to **data coverage, metadata enrichment (e.g. affiliations and funding information), interoperability with other open infrastructures, and communication practices** will be used to reassess priorities and sequencing of planned activities.

Finally, this survey will serve as a **baseline for future community engagement initiatives**. Lessons learned from this first experiment (both in terms of participation and stakeholder mapping) will inform the design of surveys and consultation mechanisms, with the aim of broadening participation and strengthening long-term engagement. Continued feedback loops with the community will be essential to ensure that OpenCitations evolves in line with real needs and expectations.