

You may not always notice this, but the dblp team is constantly working on the dblp website and its APIs in order to improve the quality of the services and the value for our users. Often these are just small details and fixes, but sometimes we introduce new features. Yet, in the past, we often rolled out those features silently with no major announcements. This has led to a number of improvements that many users of dblp may not be aware of. In order to make these features more widely known, we are starting this new “feature spotlight” series of blog posts. And the start will be a big one.

# Open citation data in computer science

Most of the time, we add new features because the community asks us to do so. And without any doubt, one of the most frequently requested features in the past years has been the addition of citations and references for the publications in dblp. While we fully understand the desire for this kind of data, our answer has always been: “Sorry, but this kind of data is unavailable to us, and there is no way that a small team like ours could collect all this data ourselves.”

The good news: This has changed during the course of the past few years. While our team is still too small to tackle such a task by ourselves, a huge chunk of the existing bibliographic citation data has been opened up to the public for reuse. And this is mainly true thanks to three actors out there. First, there is [Crossref](#), one of the major infrastructure providers behind the DOI system. For some years now, Crossref has been providing public access to publisher-deposited reference and citation data through their APIs. Second, there is [OpenCitations](#), an awesome infrastructure run by David Shotton (University of Oxford) and Silvio Peroni (University of Bologna). OpenCitations collects citation metadata from openly available sources (like Crossref) and provides access to the data via APIs and as OpenCitations Corpus (OCC). And finally, there is the [Initiative for Open Citations \(I4OC\)](#). I4OC is a joint initiative of researchers, infrastructure providers, and further stakeholders (including Crossref and OpenCitations) to promote the unrestricted availability of scholarly citation data. I4OC has already had a

remarkable impact by convincing more than 1200 bigger and smaller scholarly publishers, societies, and universities to deposit and open up their citation data with Crossref. As of September 2019, this corresponds to 59% of all Crossref-deposited articles with references.

#### How many citations are open today?



Thanks to the I4OC, the fraction of publications with open references has grown from 1% to 59% out of 47.6 million articles with references deposited with Crossref (as of September 2019).

Mid 2019, we conducted some experiments evaluating the coverage of citation data for computer science publications in the Crossref and OpenCitations data. To our surprise, the coverage was already quite significant. Out of 200,000 randomly sampled DOIs in dblp, we found that:

- 28.5% of the papers had references available at Crossref. On average, each of those papers listed 25.1 references (including many unstructured and non-DOI references).
- 52.2% of the papers had references listed at OpenCitations. On average, each of those papers listed 13.9 references (listing only DOI references).
- 59.6% of the papers had incoming citations listed at OpenCitations. On average, each of those papers listed 11.5 citing papers (listing only DOI sources).

## References and citations at dblp

Using the openly available reference and citation data of Crossref and OpenCitations, we were able to build a new “references & citations” details page for each publication in dblp that is assigned with a DOI. To find this details page, just click on the link in a paper’s drop-down menu:



■ [j11] Silvio Peroni, Alexander Dutton, Tanya Gray, David M. Shotton: **our bibliographic references free: towards open citation data.** Documentation 71(2): 253-277 (2015)

■ [10] **references & citations** authority control: Silvio Peroni, Fabio Tamburini, Fabio Vitali: **gs vs non-topical tags: Towards a bipartite classification? J.** (4): 486-505 (2015)

■ [c50] Angelo Di Iorio, Raffaele Giannella, Francesco Poggi, Silvio Peroni, Fabio Vitali: **Exploring Scholarly Papers Through Citations.** DocEng 2015: 107-116

As with all features that rely on external API calls from your browser (which might be a privacy consideration for you), you will need to opt-in once before any data can be retrieved. While we do not have any reason to believe that your connection data will be misused, we do not have any control over the remote servers either. So please proceed with care.

[ - ] **References**

---

load references from crossref.org and opencitations.net

**Privacy notice:**

By enabling the option above, your browser will contact the APIs of *crossref.org* and *opencitations.net* to load citation information. Although we do not have any reason to believe that your call will be tracked, we do not have any control over how the remote server uses your data. So please proceed with care and consider checking the Crossref privacy policy and the OpenCitations privacy policy.

Your decision whether to load or to not load citation data from *crossref.org* and *opencitations.net* will be stored by us in a cookie with your web browser and can be revoked at any time.

Once the citation data has been retrieved from Crossref and OpenCitations, it will be matched against the curated metadata records listed in dblp. The result will be displayed in one of the following ways:

- If the reference target is listed in dblp and successfully matches via its DOI, you will see a “pretty” list entry that states the full and curated metadata in the same way as usually seen in dblp bibliographies, together with its external links, export options, and so on.

Silvio Peroni , David M. Shotton **FaBiO and CiTO: Ontologies for describing bibliographic resources and citations.**  
J. Web Semant. 17: 33-43 (2012)

- If the reference target itself is not available in dblp (maybe because it is out of the scope of dblp, its addition to dblp is still pending, or the matching with its dblp record failed for some reason) you will see a textual bibliographic reference string describing the reference. We unfortunately cannot give curated metadata for items that are not indexed by dblp, so we use a plain description as given by the sources.

MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science*, 40(5), 342–349.  
doi:10.1002/(SICI)1097-4571(198909)40:5<342::AID-ASI7>3.0.CO;2-U

- In some rare cases, you might notice a list item labeled with “(missing metadata)”. This means that this is the first time that we have encountered that DOI in dblp, and we do not even have a textual reference string ready yet. In this case, a background process will make sure that a textual reference will be available soon, usually within minutes.

(missing metadata) [[doi:10.1093/bioinformatics](https://doi.org/10.1093/bioinformatics)]

- Of course, you might be unlucky and find that there is no open reference or citation data available for a publication. In that case, we unfortunately cannot list anything.

 Cited by 

---

load citations from opencitations.net

*no open data available, sorry*

Current statistics (as of November 2019) show that you can find at least a partial reference list for 51.6% of all publications in dblp, and that 45.7% of all publications list at least one citing paper. Please note that using open data sources, it is currently not

possible to retrieve citation data for publications without a DOI. So, in many ways, the reference and citation details are still a work in progress at dblp.

## The missing part

When using the reference and citation listings from dblp, you should always keep in mind that a number of important publishers in computer science (such as IEEE and Elsevier) are still not supporting open citation data. Hence, there is a systemic bias in the availability of such data. In particular, we would certainly not encourage anyone to conduct citation based studies based on such incomplete data. If you find that there is no open citation data available, this does not necessarily mean that an article has not been cited before.

It is our understanding that the reference lists given in scholarly publications are an integral part of a publication's metadata and need to be openly available to the public.

**If you find that citation data of your publications are not openly available yet, then please consider asking your publisher to release your reference lists to the public.** It is after all you, the researcher, who has been spending a lot of time and effort in order to compile those reference lists.

Alternatively, OpenCitations recently started crowdsourcing the collection of open citations with their new CROCI index. If you happen to have DOI-to-DOI citation information ready that you collected as part of your research, then you might consider donating your data to that CC0 corpus:

- <https://opencitations.wordpress.com/2019/02/07/crowdsourcing-open-citations-with-croci/>
- <https://github.com/opencitations/croci>

Another promising community initiative is [WikiCite](#). WikiCite is a project of the Wikimedia movement aiming at the creation of a comprehensive knowledge base with rich information about every scholarly reference in Wikipedia. This metadata can be curated openly by anyone within the Wikidata framework:

- <https://meta.wikimedia.org/wiki/WikiCite>
- [https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Source\\_MetaData](https://www.wikidata.org/wiki/Wikidata:WikiProject_Source_MetaData)