



There are now more than 60,000 manually confirmed external IDs linked with dblp author bibliographies. This is quite an improvement. Below you can see the number of external identifiers at the end of each year compared to the numbers in October 2018



As you can see, our definition of external identifier is a bit wider than usual. E.g., we include Twitter profiles here. ORCID, Google Scholar, Wikipedia and Twitter should be known to the readers. Wikidata is a LOD data store build on top of Wikipedia that provides query-able data. The others are

- GND: identifier from the German national Library [see here](#)
- ISNI: global authority file [see here](#)
- zbMath: biographies of Mathematicians [see here](#)
- Math Genealogy: the Mathematics Genealogy Project. Stores relations between researchers and the thesis advisers [see here](#)
- ResearcherID is a commercial authority service [see here](#)
- ACM DL is an entry in the ACM digital library [see here](#)

There are more types of external links. To see them hover the mouse over the small figure next to the name in the author bibliographies. E.g..



## Why are these identifiers important?

There are a number of reasons why we search for external identifiers. The most important ones are:

- External IDs provide valuable **information for curating bibliographies**. The bibliographic metadata dblp processes is very sparse, that is, we usually only know a publication title, author names, venue names, and a few extra information like



pagination. This is often insufficient to properly disambiguate authors (i.e., to make sure that all of an author's publications are listed on a single bibliography, and that the bibliography does not contain publications from other persons with the same name). Linked external data sources might provide author-curated publication lists, affiliation information, or details on an author's area of research which helps our editors. In particular, external identifiers are an important indicator when automatically scanning for bibliographies that need curation.

- Links to external resources allow our users to get more complete **information about an author**. An important role of a dblp bibliography is to provide an overview of an author's scholarly work. Maybe a user is studying a paper and is looking for more interesting work from its authors. External resources provide more information on the researchers. E.g., external databases may list works from outside of computer science, or a scientist's twitter channel gives more insight into her work.
- External resources give us **information about stuff that is missing**. Locating publications can be difficult. In particular smaller, community-run workshops can be well hidden. An linked external resources might provide a publication list that we can scan for such hidden venues.

## Why did the number of identifiers increase so fast?

Essentially, we spent a lot more of our time working on this topic.

But this is of course not the whole story. In particular, one important factor is that we were able to build upon existing open data collections. An important role in this process is played by Wikidata, which acts as a kind of open data hub for identifiers. Hence, a typical workflow is as follows:

1. We add a Wikidata identifier to a dblp bibliography.
2. We periodically harvest linked Wikidata entities for further identifiers, such as



## ORCIDs.

We still check all IDs we collect manually. But having a data hub like Wikidata ready makes our task much easier.

We also have become more effective at monitoring the quality of our bibliographies in dblp. Especially thanks to [the integration of ORCID data](#), we are now able to identify bibliographies that do either align with or show conflicts with linked ORCID profiles. These cases are easy to check and fix by our editors, and in the process we collect and link further ORCIDs with the curated bibliographies in dblp.

ORCIDs can themselves then be used to identify matching Wikidata entities, and vice versa. In doing so, we are able to aggregate and expand upon a network of linked identifiers for each bibliography. Of course, our findings are not only made available through dblp. We also periodically feed back the results of our work to Wikidata.