

A big change has just been made to the dblp website ... and, in case we did our job right, you may even haven't noticed yet: With the latest update, we introduced **major changes to the dblp URL scheme**. In particular, this applies to the URLs of all author bibliographies listed on dblp, which are now served under a new and persistent URL.

But don't worry, just like the first time we made such a change about eight years ago, we try to keep all previously existing URLs as a redirect for the foreseeable future.

In this post, we talk about the reasons that made us abandon our old URL scheme and why you will most likely want to update your hyperlinks and bookmarks anyway.

## So, what's the problem with the old URL scheme?

Author bibliographies have been an integral part of dblp since its inception in 1993. Since all author name strings are unique in dblp (even if we often have to enforce uniqueness by adding a magic [name disambiguation suffix number](#)) it was quite natural to just use an author's name in order to locate her bibliography. For example, [Jim Gray](#) is listed in dblp under the unique name "Jim Gray 0001". Hence, a simple rewriting rule derived the following URL for his bibliography webpage:

- [https://dblp.org/pers/g/Gray\\_0001:Jim.html](https://dblp.org/pers/g/Gray_0001:Jim.html)

This approach has served us well through the years. But recently, the need for persistent unique identifiers has become more and more evident. And those bibliography URLs tied to an author's name have two major problems: They are **neither unique, nor persistent**.

For instance, [Don Knuth](#) is listed in dblp as both "Donald E. Knuth" and "Donald Ervin Knuth". Hence, we used to end up with two URLs pointing to the same bibliography webpage, namely:

- [https://dblp.org/pers/k/Knuth:Donald\\_E=.html](https://dblp.org/pers/k/Knuth:Donald_E=.html)

- [https://dblp.org/pers/k/Knuth:Donald\\_Ervin.html](https://dblp.org/pers/k/Knuth:Donald_Ervin.html)

While this situation could be solved by serving HTTP redirects, the transient nature of those URLs poses a far more dire problem. As the URL is directly tied to the exact spelling of a name in dblp, any correction or edit that changed an author's name in the metadata did also change these URLs. E.g., at one point in time, we had Don Knuth's bibliography also listed without his middle name, that is, as:

- <https://dblp.org/pers/k/Knuth:Donald.html>

Since we corrected all of Don Knuth's metadata records to always contain his middle name, this URL has lost its purpose and is no longer showing any bibliography at all. This leaves external hyperlinks pointing to ugly "404: Not Found" error pages and breaks meaningful open data links in the semantic web.

## PIDs to the rescue!

As the need for stable links to dblp bibliographies had become evident, we already established a second set of persistent and unique identifiers some years ago: Every author has also been assigned an internal PID that is intended to never change. Attentive users may already have noticed these PIDs as part of the persistent short URLs given in the web UI.



However, for reasons deeply rooted in dblp's 25 years old technical layout (that are not

easy to explain in a simple blog post) PID-based URLs so far have only been able to play the role of a HTTP redirect and were not the prime web address displayed in your browsers URL bar. This has changed with the recent update.

Moving forward, the old name-based URLs will be retired and dblp bibliographies will be served exclusively based on their PIDs. We will, of course, try to keep all incoming links alive via HTTP 301 Redirects. However, **we strongly encourage you to update your bookmarks and hyperlinks** to the new, persistent scheme as the PID-based URLs will also become the basis of our updated data API URLs.

## The format of a dblp PID

So, what does this mean in practice? Take the bibliography of [Kristin Lauter](#) as an example. Her bibliography is, and will always remain, to be found using the unique dblp PID “08/1510” and the associated URL:

- <https://dblp.org/pid/08/1510.html>

Or, as a second example, have a look at the bibliography of [Ayanna Howard](#). In dblp, she is listed with PID “11/399” and, hence, under the URL

- <https://dblp.org/pid/11/399.html>

However, please be aware that by their look, dblp PIDs come in two flavors. Starting in 2009, newly created bibliographies have been assigned automatically minted, numerical PIDs like the two PIDs given above. The vast majority (more than 99%) of all PIDs you find in dblp today follow this numerical format.

Yet, in the early days of dblp, internal keys have been created exclusively by hand. Hence, PIDs of those earliest authors ended up being, again, modeled after their name. As an example, have a look at [Barbara Liskov](#), who is listed with PID “l/BarbaraLiskov” and at URL:

- <https://dblp.org/pid/l/BarbaraLiskov.html>

The crucial difference to the name-based URLs we discussed earlier is that those name-

based PIDs have never been changed once they had been minted, even if the actual name had.

From a technical viewpoint, there is no difference between those flavors of PIDs. In general, a dblp PID should always be considered to be **just an arbitrary string of case-sensitive, alphanumeric ASCII letters** (plus occasional dashes and slashes) with no special meaning attached to them.

## A streamlined API URL scheme

Finding and browsing scholarly bibliography HTML websites is fine. But also having that same metadata easily accessible in a machine-readable format is even better. The dblp team is committed to making all of our metadata available as open data and to facilitate reuse. Hence, the URL scheme of our data APIs has also received attention, and its format has been streamlined in order to simplify its usage. From now on, the bibliography metadata API for all data formats will follow the same general scheme:

- <https://dblp.org/pid/<person-PID>.<data-format-file-extension>>

Again, take the bibliography of [Barbara Liskov](#) as a concrete example. Her bibliography is uniquely identified in dblp with the PID “l/BarbaraLiskov” and the resource URL:

- <https://dblp.org/pid/l/BarbaraLiskov>

By adding an appropriate file extension to that URL, you can then request the bibliography in your preferred data format, such as:

- <https://dblp.org/pid/l/BarbaraLiskov.html> (text/html, i.e.: the webpage)
- <https://dblp.org/pid/l/BarbaraLiskov.xml> (application/xml)
- <https://dblp.org/pid/l/BarbaraLiskov.rss> (application/rss+xml)
- <https://dblp.org/pid/l/BarbaraLiskov.bib> (application/x-bibtex)
- <https://dblp.org/pid/l/BarbaraLiskov.ris> (application/x-research-info-systems)

This is mirrored by our API serving metadata of a single publication. Here, given a

(persistent) dblp publication key (say, “journals/tocs/CastroL02”), the publication is uniquely identified by the resource URL

- <https://dblp.org/rec/journals/tocs/CastroL02>

and its web page and metadata can be retrieved using:

- <https://dblp.org/rec/journals/tocs/CastroL02.html> (text/html, i.e.: the webpage)
- <https://dblp.org/rec/journals/tocs/CastroL02.xml> (application/xml)
- <https://dblp.org/rec/journals/tocs/CastroL02.bib> (application/x-bibtex)
- and so on.

In addition to that, the API URLs (without file extension) also support content negotiation via the HTTP “Accept” header and MIME type. There is still more to tell about the dblp open data API, and we will most certainly dedicate a future blog post to that topic.

## Linking open data

If you are maintaining or building a service based on data from dblp, we strongly encourage you to update your URLs linking to dblp. This is particularly true in the context of linked open data in the semantic web, which relies on persistent URLs to identify entities.

Please note that while our PIDs are for the most part persistent, there may still be rare cases when they do change or expire. This usually only happens in scenarios where a personalized bibliography is so flawed beyond repair (e.g., a case of multiple homonymous authors mixed into one bibliography) that it is more reasonable to let the old bibliography expire and to recreate proper bibliographies from scratch, or when the bibliography has just been a disambiguation placeholder (say, a pseudo-bibliography of all unassigned publications of some “D. Wang”) that has outlived its purpose. We also have a redirection mechanism in place which lets a request to a deprecated PID respond with an HTTP status 301 Redirect in order to point to the most recent one. This is especially relevant in cases when two bibliographies have been merged.

Fortunately, we already had the new PID URLs in mind when we started adding [dblp PID links to WikiData](#) some time ago. Today, more than 40.000 person entity links already exist between [wikidata.org](https://wikidata.org) and dblp. Given that we finally adopted persistent URLs as first-class citizens, I am confident that many more will follow.