

WEBOMETRICS – RANKING WEB OF REPOSITORIES. TO COUNT AND NOT TO COUNT

Leif Longva

We, who work with repositories holding scientific documents, are keen to see how our repositories develop compared to other, comparable repositories. It is therefore with great interest we examine the ranking of such repositories, whenever a new such ranking is presented. Cybermetrics Lab, a research group belonging to the Consejo Superior de Investigaciones Científicas (CSIC) in Spain, calculate their ranking of the world scientific repositories based on four weighted criteria¹:

- The size, in terms of number of documents in the repositories
- Visibility, measured by the numbers of links in to the repository's documents
- The number of "rich files", meaning files in formats like pdf and word, to indicate that the repository holds more than metadata
- The degree of scholarly content, measured by documents identified by Google Scholar

We would like to express our gratitude to the Cybermetrics Lab for doing this. The rankings are very interesting to read, and we think the four mentioned criteria make sense.

Visibility

Visibility is an important criterion. The usefulness of our repositories is dependent of how visible the content is, and the number of links into the repository is a good indicator of this. And we therefore thought that the number of links would be interesting to count for the ranking purpose, no matter what kind of links were used. But this is not how Cybermetrics Lab sees it.

We prefer using persistent urls (purls), and our repository therefore allocates a handle, a widely used purl, to each item. And we always recommend these handle urls to be used while linking to documents in our repository, for the very reason why purls are wise to use: No risk of rotten links (as long as the handle

service is alive), regardless of possible future changes in our institution's or server's name. However, by examining the way Ranking Web of Repositories works, we realized that such purls do not count in the calculation of the visibility factor of the ranking. They only count links that carry the name of the institution's web domain. We found that rather strange. It should be possible to keep track of which purls point to which repositories. Surely, this must be technically solvable, we thought.

We therefore contacted Cybermetrics Lab on the matter. And we received a prompt reply, confirming that technical problems were not the issue. The answer said that since the handle system is owned by a private corporation unrelated to our university, we do not have any guarantee of its survival into the future. Moreover, the answer goes, links should include the institution's name in order to carry the information of which institution is behind the document.

Of course, Cybermetrics Lab is right that we have no guarantee for how the future looks like. And urls that include the institution's name do hold some information for the user. But what puzzles me is why this is an issue while trying to measure the visibility of repositories. The handle system is alive and kicking today, so the handle links do add visibility to our repository today. This is a fact, beyond discussion. Cybermetrics Lab has thus defined a way to count links that excludes many links, and therefore produces an inferior measurement. And this is done based on a moral view on how Cybermetrics Lab would like the repositories to build their links. I think Cybermetrics Lab rather should approach their task in a more scholarly way, and measure what is measurable, even those objects who act otherwise than what Cybermetrics Lab prefer. And by all means, Cybermetrics Lab is free to advocate their views on how to link.

So, while reading the Ranking Web of Repositories, beware what this metric is not measuring.

¹ <http://repositories.webometrics.info/en/Methodology>



Leif Longva University of Tromsø, Munin repository manager