

Introducing the Leiden Ranking Open Edition

Ludo Waltman , Nees Jan van Eck , Martijn Visser, Mark Neijssel , Lucy Montgomery, Cameron Neylon , Bianca Kramer, Kyle Demes and Jason Priem

Published January 30, 2024

Citation

Waltman, L., van Eck, N. J., Visser, M., Neijssel, M., Montgomery, L., Neylon, C., Kramer, B., Demes, K., & Priem, J. (n.d.). In *Leiden Madtrics*. Leiden Madtrics. <https://doi.org/10.59350/89wpz-hpz32>

Copyright

Copyright © Ludo Waltman et al. 2024. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Leiden Madtrics

When the Shanghai Ranking, also known as the Academic Ranking of World Universities (ARWU), was launched in 2003, Ton van Raan, director of CWTS at the time, sounded the alarm about the [problematic way in which the ranking uses bibliometric data](#), for instance in attributing publications to universities. In response to the Shanghai Ranking, CWTS decided to launch the [Leiden Ranking](#), aiming to demonstrate more appropriate ways to use bibliometric data for comparing universities. While the Leiden Ranking did not gain the same visibility as the Shanghai, THE, and QS rankings, it developed a strong reputation for offering a robust, high-quality approach for comparing universities in terms of bibliometric parameters.

Today, after almost two decades, the Leiden Ranking is going to take an ambitious next step in improving bibliometric approaches for comparing universities. The [Open Edition of the Leiden Ranking](#), launched today by CWTS, addresses one of the most challenging problems of bibliometric indicators: the lack of transparency of these indicators due to their dependence on proprietary data. Together with the [Curtin Open Knowledge Initiative \(COKI\)](#), [Sesame Open Science](#), and [OurResearch](#), CWTS has rebuilt the Leiden Ranking, making it fully transparent by working exclusively with open data and open source algorithms. Until now, the Leiden Ranking has always been based on data from [Web of Science](#), a proprietary data source owned by Clarivate. The Leiden Ranking Open Edition uses data from [OpenAlex](#), a fully open data source created by OurResearch. To the best of our knowledge, the Leiden Ranking Open Edition is the first fully transparent university ranking.

The importance of transparency of university rankings is widely recognized. CWTS emphasizes the need to be transparent in [its ten rules for ranking universities](#). Transparency also features prominently in the [criteria for fair and responsible university rankings](#) developed by the International Network of Research Management Societies (INORMS). These criteria were used in the [rethinking the rankings initiative](#). Likewise, transparency and openness are key elements in the [strategy for culture change regarding university rankings](#) that is currently being implemented in the Netherlands. More in general, the [UNESCO Recommendation on Open Science](#) highlights the need for "open bibliometrics and scientometrics systems". In a similar vein, the [Agreement on Reforming Research Assessment](#) stresses the importance of "independence and transparency of the data, infrastructure and criteria necessary for research assessment".

How did we create the Leiden Ranking Open Edition?

The Leiden Ranking Open Edition aims to reproduce the traditional Leiden Ranking as closely as possible, focusing on the [Leiden Ranking 2023](#) that was published in June last year. The Open Edition includes the same 1411 universities that are also included in the Leiden Ranking 2023. Instead of proprietary data from Web of Science, the Open Edition uses open data from OpenAlex. This data is harvested by OpenAlex from a [variety of sources](#), including [Crossref](#), [PubMed](#), and the websites of publishers. The Open Edition is based on the [OpenAlex snapshot](#) released on November 21, 2023.

Leiden Madtrics

Figure 1 summarizes the process of creating the Leiden Ranking Open Edition. We now discuss the different steps in this process.

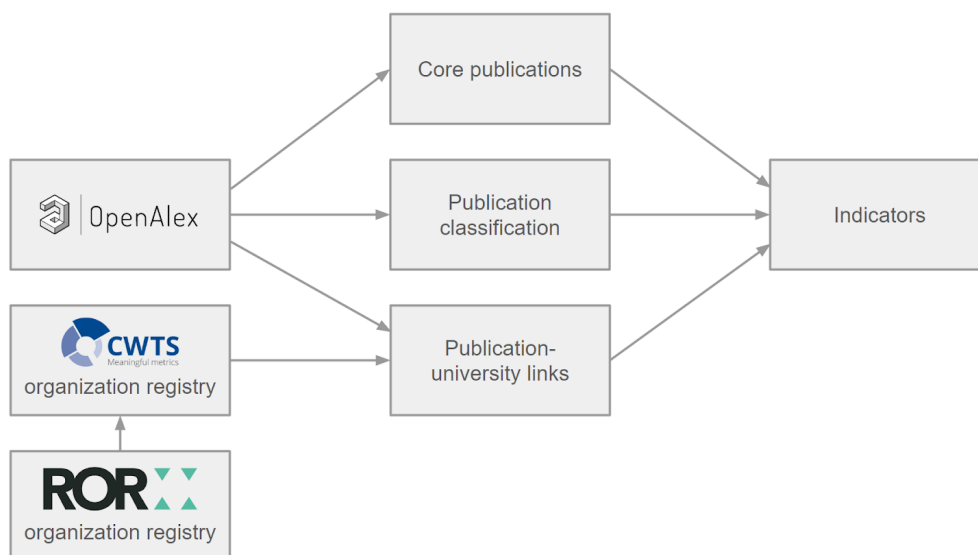


Figure 1. Summary of the process of creating the Leiden Ranking Open Edition.

Core publications

Web of Science and OpenAlex differ in their coverage of the scientific literature. Most publications covered by Web of Science are also covered by OpenAlex, but OpenAlex covers many publications that Web of Science does not cover. The traditional Leiden Ranking focuses on so-called core publications, which are identified based on a [number of criteria](#). The Open Edition also has a focus on core publications, identified based on [similar criteria](#). However, because of coverage differences between Web of Science and OpenAlex, the set of core publications considered in the Open Edition differs from the set of core publications considered in the traditional Leiden Ranking. This is discussed in more detail in [this blog post](#).

Publication classification

The Leiden Ranking relies on a detailed algorithmic classification of publications into research areas. This classification is used to assign publications to main fields and to calculate normalized citation impact indicators. For the Open Edition, CWTS created a new classification of publications based on OpenAlex data. [This blog post](#) provides more details.

CWTS organization registry

OpenAlex uses [ROR \(Research Organization Registry\) identifiers](#) for organizations, while the Leiden Ranking has traditionally used internal CWTS organization identifiers. For the Open Edition, CWTS created a new organization registry by matching CWTS organization identifiers to ROR identifiers. This was done for the 1411 universities included in the Open Edition and for the [affiliated organizations linked to these universities](#). Additional links between universities and affiliated organizations were obtained from the ROR organization registry. Links between

Leiden Madtrics

universities and their affiliated organizations are presented in a transparent way on the website of the Open Edition (see Figure 2).\



Figure 2. Screenshot of the website of the Leiden Ranking Open Edition showing the links between University of Amsterdam and its affiliated organisations.

Publication-university links

OpenAlex provides ROR identifiers for the affiliations of authors of publications. We linked publications to universities by connecting ROR identifiers obtained from OpenAlex to the ROR identifiers of universities and their affiliated organizations in the CWTS organization registry. In [this blog post](#) the CWTS team presents a comparison between the approach for linking publications to universities used in the traditional Leiden Ranking and the approach used in the Open Edition.

Indicators

The Open Edition includes the same bibliometric indicators as the traditional Leiden Ranking, except for [indicators of gender diversity](#). These indicators are somewhat more challenging to produce using OpenAlex data because OpenAlex does not make a distinction between first and last names of authors. This distinction is important for algorithmic gender inference. At the moment gender diversity indicators are therefore not included in the Open Edition. We may add them in the future.

How to access the data and software

The data used to create the Leiden Ranking Open Edition is [openly available in Zenodo](#) under a [CC0 public domain dedication](#). The data can also be [accessed through Google BigQuery](#) (to run queries you need to have a BigQuery account and your own project). Universities may for instance use the data to check whether publications have been correctly attributed to them. Performing such checks is not possible in the case of the traditional Leiden Ranking, since this ranking is based on proprietary data that cannot be shared openly.

The source code of the software used to create the Leiden Ranking Open Edition is [openly available in GitHub](#) under an MIT license. A Microsoft SQL Server database system was used to

Leiden Madtrics

perform the data processing and the calculations. The source code therefore consists mostly of SQL scripts.***

How to compare the Open Edition with the traditional Leiden Ranking

To what extent does the Open Edition provide results that are similar to those provided by the traditional Leiden Ranking? To address this question, we have created an [interactive dashboard](#) in which the values of the indicators in the two editions of the Leiden Ranking can be easily compared.

The interactive dashboard can for instance be used to explore the correlation between the number of publications of a university in the traditional Leiden Ranking and in the Open Edition (see Figure 3), or the correlation between a university's proportion of highly cited publications in the two editions of the Leiden Ranking (see Figure 4).

\

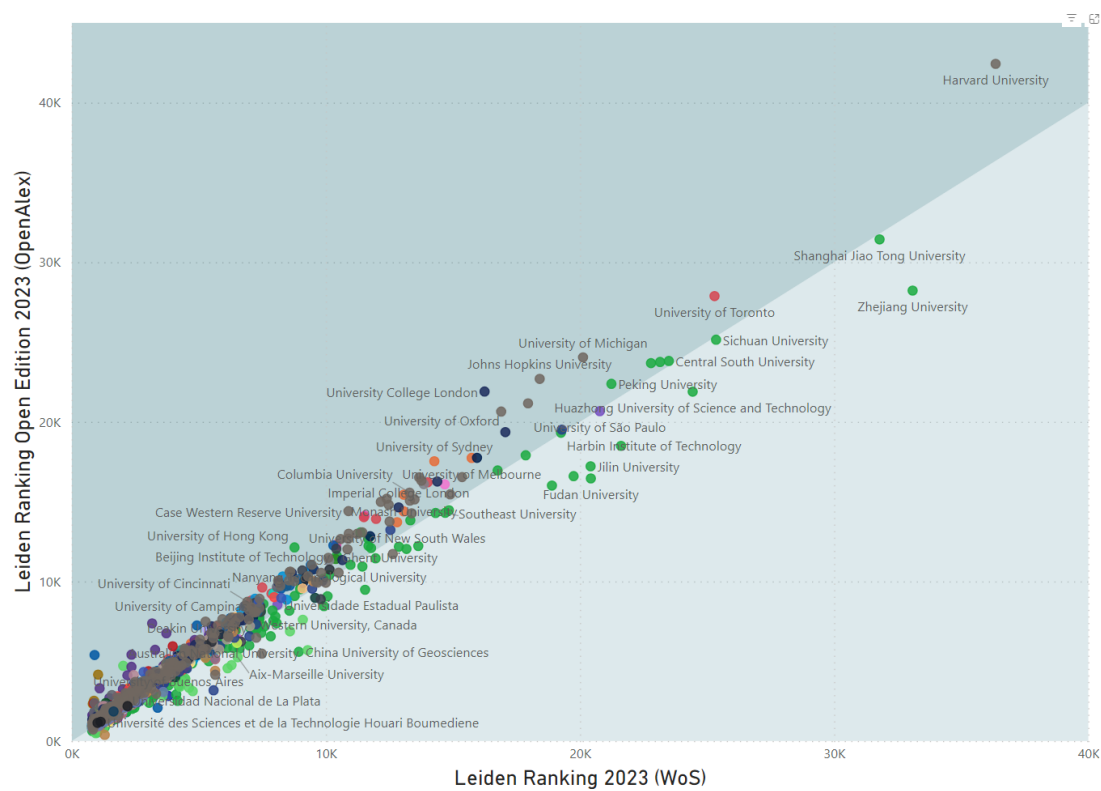


Figure 3. Correlation between the number of publications of a university in the traditional Leiden Ranking and in the Open Edition (fractional counting).

\

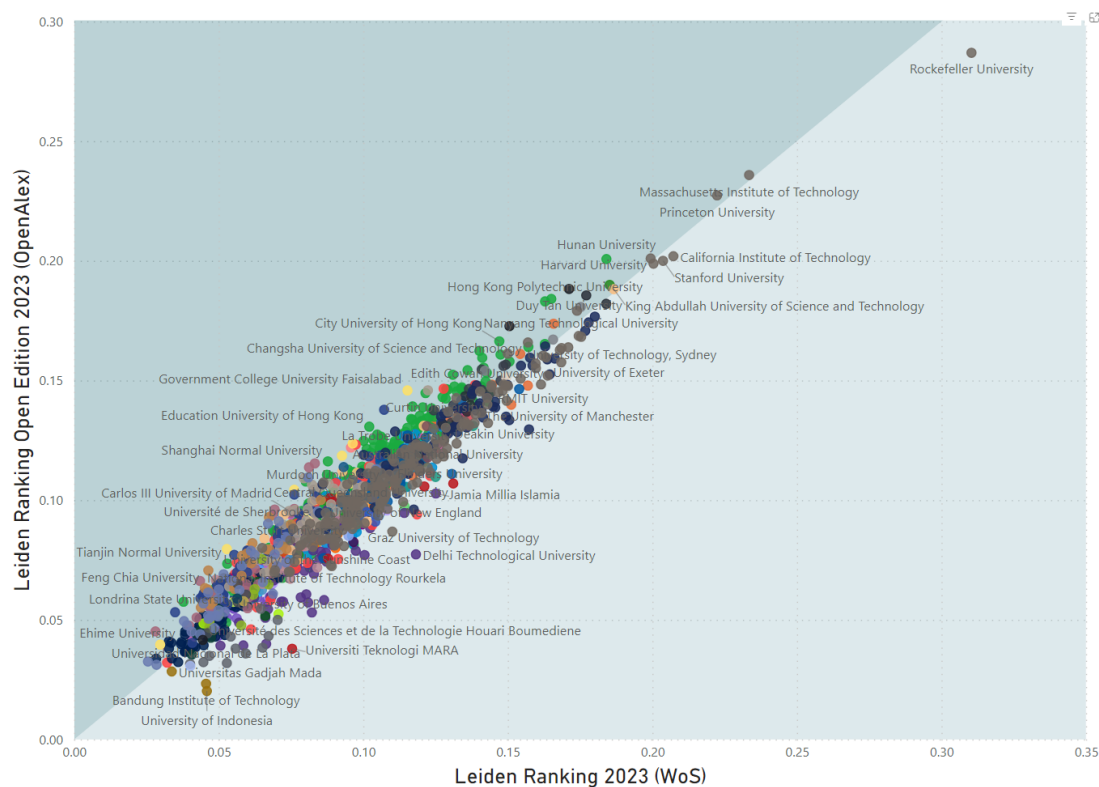


Figure 4. Correlation between the proportion of highly cited (top 10%) publications of a university in the traditional Leiden Ranking and in the Open Edition (fractional counting).

\

In general, we consider the correlations between indicator values in the two editions of the Leiden Ranking to be quite strong. However, for some universities there are large differences between the two editions of the Leiden Ranking. In many cases this is probably due to differences in the publications attributed to a university. This issue is analyzed in more detail in [this blog post](#).

What is next?

At the moment the Leiden Ranking Open Edition is still of an experimental nature. Building the ranking was an incredibly useful exercise for our organizations (CWTS, COKI, Sesame Open Science, and OurResearch). We also expect to learn a lot from the feedback we hope to receive from users of the Open Edition. Based on the lessons learned, we will make further improvements to the Open Edition. Within one or two years, we expect the Open Edition to be fully mature and to offer a full replacement for the traditional Leiden Ranking.

While our organisations are making further improvements to the Leiden Ranking Open Edition, CWTS will continue to release annual updates of the traditional Leiden Ranking based on Web of Science data. In the next one or two years, we expect the traditional Leiden Ranking and the Open Edition to co-exist. In the somewhat longer term, CWTS will make a [full transition to open research information](#). Within the next few years, all bibliometric indicators produced by CWTS, including those in the Leiden Ranking, will be based on open data.

Leiden Madtrics

The transition of the Leiden Ranking toward open data fits into broader discussions about openness of research information, responsible use of bibliometric indicators, and reform of research assessment. Critical reflection on the value of university rankings should be part of these important discussions.\

How to get in touch\

Would you like to know more about the Leiden Ranking Open Edition? CWTS is organizing two webinars about the Open Edition, providing an opportunity to discuss the ranking with members of the CWTS team. The [first webinar](#) takes place on February 6 at 10h CET, and the [second webinar](#) on February 9 at 15h CET. COKI, in collaboration with Future Campus, is offering a [webinar with an Australian focus](#) on January 31 at 11h AEDT. OpenAlex is organising a [webinar about institutional profiles and affiliation curation](#) on February 8 at 12h EST.

We also invite you to reach out to us using the [contact form](#) on the website of the Leiden Ranking Open Edition. We appreciate your comments and feedback on the Open Edition.

Do you want to start working yourself with open bibliometric data sources such as OpenAlex? You may then consider signing up for the [course *Scientometrics Using Open Data*](#) organised by CWTS, COKI, and Sesame Open Science. The next edition of this online course takes place from March 25 to March 28.\

Header image from <https://www.leidenranking.com>