Leiden Ranking Open Edition

Nees Jan van Eck
Centre for Science and Technology Studies (CWTS), Leiden University

Towards Open Research Information, SciLake Plenary Meeting
Barcelona, Spain, November 9, 2023

Should open data sources such as OpenAlex replace traditional proprietary data sources such as Web of Science or Scopus?

Yes, definitely

Sometimes

No, not at all

CWTS

How can we make the use of open data sources such as OpenAlex the norm?

How can we convince the academic community that these data sources offer an improvement over traditional proprietary data sources?

CWTS

Last year CWTS, COKI and SIRIS also asked themselves these questions

Our answer:

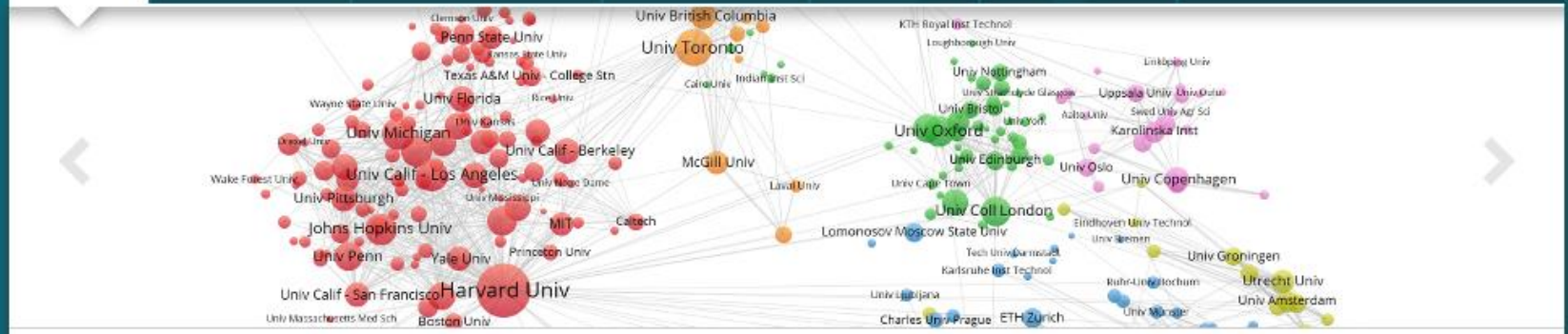
Let's show how open data sources can be used to introduce fully transparent and democratic approaches to university rankings

Any university that pays attention to rankings (for better or for worse) should start paying attention to open data sources

CWTS COKI CURTIN OPEN KNOWLEDGE INITIATIVE SIRIS A·C·A·D·E·M·I·C OpenAlex

CWTS Leiden Ranking

Meaningful metrics



CWTS Leiden Ranking 2023

The CWTS Leiden Ranking 2023 offers important insights into the scientific performance of over 1400 major universities worldwide. Select your preferred indicators, generate results, and explore the performance of universities.

List view

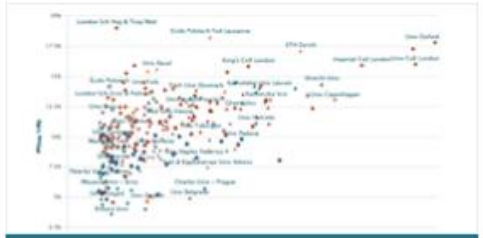


Chart view

Map view

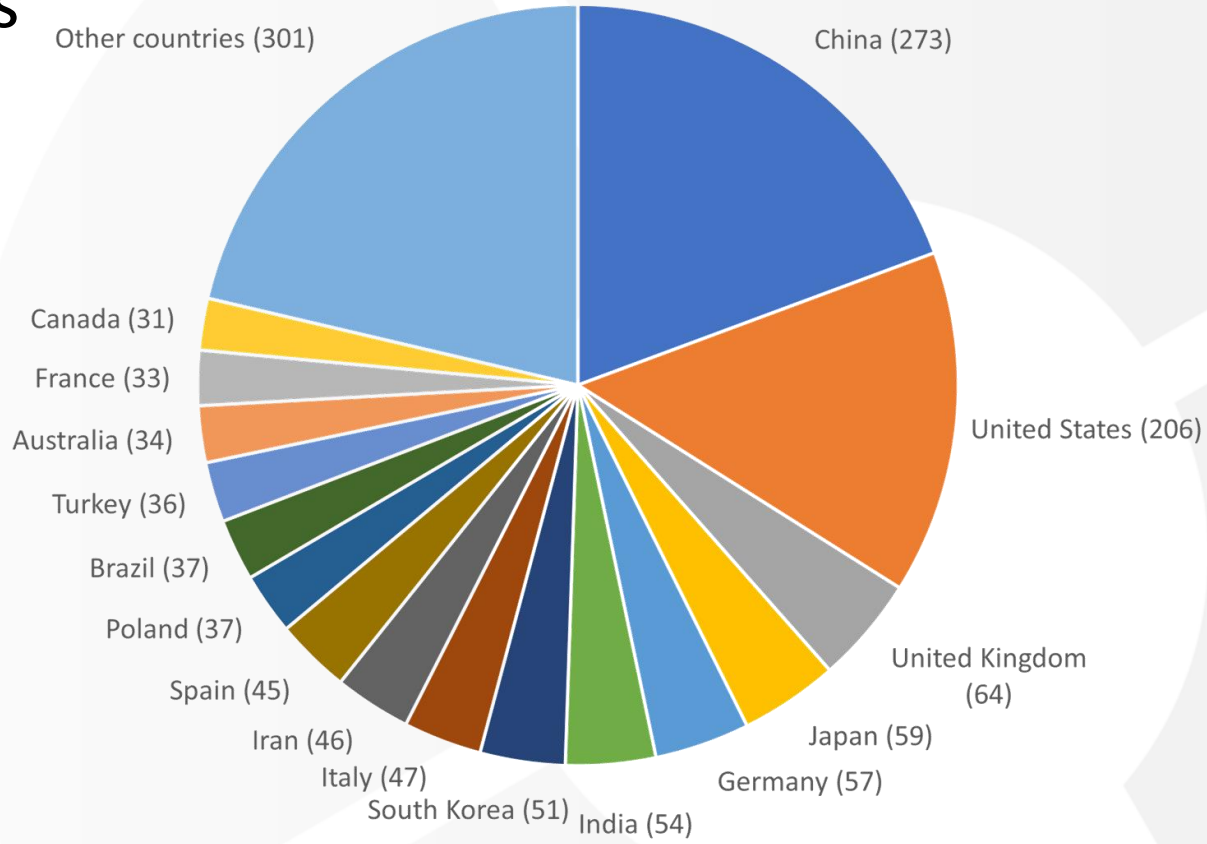
Select, generate and explore

The Leiden Ranking enables you to select sophisticated bibliometric indicators, to generate results based on these indicators and to explore the results from three different perspectives. Best known is the traditional *list view*, in which you can rank universities according to a selected

Subscribe to the Leiden Ranking newsletter

CWTS Leiden Ranking

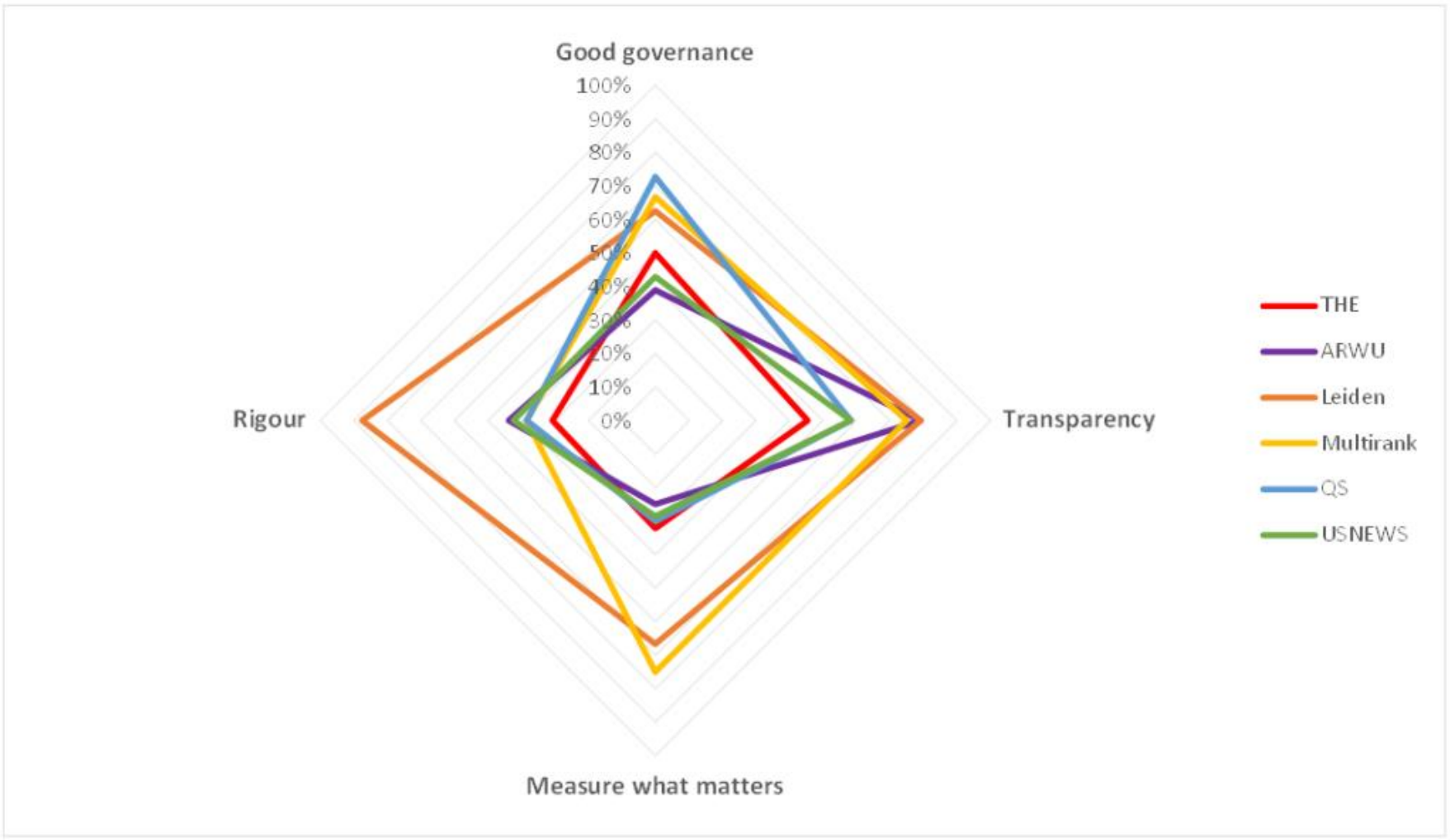
- 1411 universities from 72 countries
- Provides bibliometric indicators for four dimensions:
  - Scientific impact
  - Collaboration
  - Open access
  - Gender
- Based on Clarivate Web of Science data, with enrichments by CWTS



Differences with other university rankings

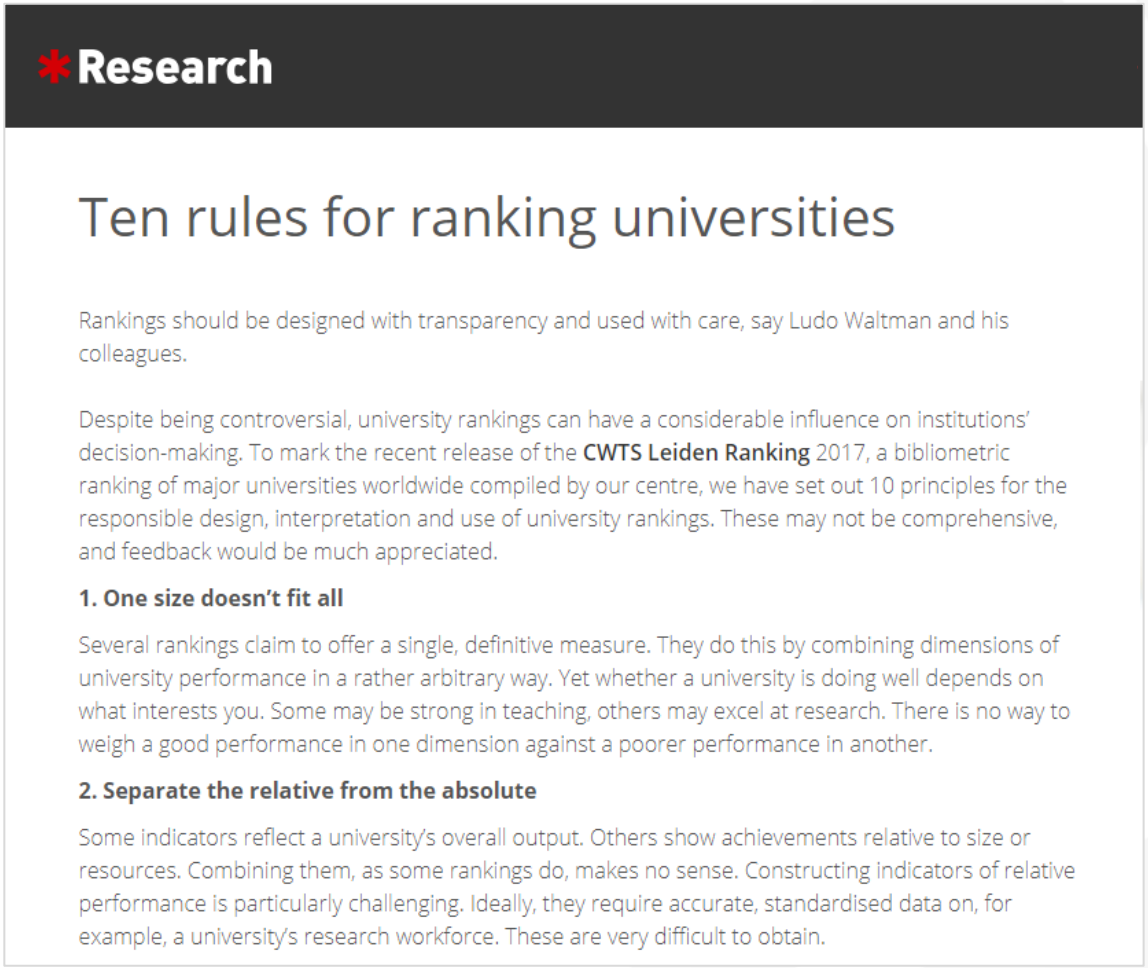
- No composite indicators
- Multi-dimensional
- Size-dependent and size-independent indicators
- Focused on research, not on teaching
- Based purely on bibliometric indicators; no survey data or data provided by universities
- High-quality bibliometric methodology
- Multiple views, not just a simple list

INORMS assessment of global university rankings



Source: https://doi.org/10.31235/osf.io/thx4j

Ten rules for ranking universities



Research

Ten rules for ranking universities

Rankings should be designed with transparency and used with care, say Ludo Waltman and his colleagues.

Despite being controversial, university rankings can have a considerable influence on institutions' decision-making. To mark the recent release of the **CWTS Leiden Ranking** 2017, a bibliometric ranking of major universities worldwide compiled by our centre, we have set out 10 principles for the responsible design, interpretation and use of university rankings. These may not be comprehensive, and feedback would be much appreciated.

1. One size doesn't fit all

Several rankings claim to offer a single, definitive measure. They do this by combining dimensions of university performance in a rather arbitrary way. Yet whether a university is doing well depends on what interests you. Some may be strong in teaching, others may excel at research. There is no way to weigh a good performance in one dimension against a poorer performance in another.

2. Separate the relative from the absolute

Some indicators reflect a university's overall output. Others show achievements relative to size or resources. Combining them, as some rankings do, makes no sense. Constructing indicators of relative performance is particularly challenging. Ideally, they require accurate, standardised data on, for example, a university's research workforce. These are very difficult to obtain.

Source: www.researchresearch.com/news/article/?articleId=1368350

CWTS

Ten rules for ranking universities

4. Be transparent

Users of rankings require at least a basic understanding of their design. Rankings therefore need to explain their methodology. Ideally, they should also make their underlying data available. Users could then see, for example, not only how many highly cited publications a university has produced but also what they are. Most rankings do not do this, because of the proprietary nature of data and the commercial interests of rankers.

CWTS

Leiden Ranking Open Edition

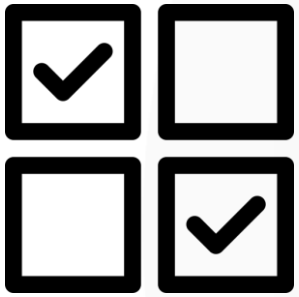
The idea of the Leiden Ranking Open Edition

- Based on OpenAlex
- Replicate methodology CWTS Leiden Ranking
- Explore what is possible
- Identify and solve issues
- Gain trust
- First edition planned for January 2024

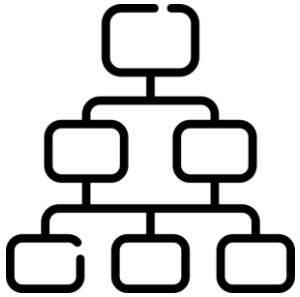
Main elements of the Leiden Ranking methodology

Linking publications to universities

Identification of core publications and sources



Creating a publication classification



Main elements of the Leiden Ranking methodology

Linking publications to universities

Identification of core publications and sources

Creating a publication classification

CWTS

Linking raw affiliation strings to universities

Leiden University
Universiteit Leiden
University of Leiden
Rijksuniversiteit Leiden
Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands
Huygens laboratory, Leiden University, P.O. Box 9504, 2300 RA Leiden, The Netherlands
Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands
Institute of Biology, Leiden University, Leiden, The Netherlands
Institute of Psychology, Leiden University, Leiden, The Netherlands
Instituut-Lorentz, Universiteit Leiden, P. O. Box 9506 2300 RA Leiden, The Netherlands
Mathematical Institute, Leiden University, Leiden, The Netherlands
Leiden Institute for Brain and Cognition, Leiden, The Netherlands
Leiden Observatory, Leiden, The Netherlands
Faculty of Archaeology, Leiden University, Leiden, the Netherlands
Kamerlingh-Onnes Laboratory, Leiden University, P.O.Box 9504, 2300 RA, Leiden, the Netherlands
Huygens Laboratorium der Rijksuniversiteit, Leiden, The Netherlands
LIACS, Leiden University, Leiden, The Netherlands
Centre for Science and Technology Studies, Leiden University, Leiden, The Netherlands
Universität Leiden, Leiden, Holland, Niederlande
Sterrewacht Leiden, PO Box 9513, 2300 RA Leiden, The Netherlands
Leiden Institute for Brain and Cognition, The Netherlands

CWTS

Linking publications to universities

Issues that have been addressed:

- Incomplete coverage of raw affiliation strings
- Incomplete institution identification
- Raw affiliation strings linked to no more than one institution

https://openalex.org/W1965914817

24 July 2014

SUAVE: a UV telescope for space weather and solar variability studies

L. Damé, M. Meftah, A. Irbah, A. Hauchecorne, P. Keckhut, E. Quémerais

Author Affiliations -

L. Damé,[1] M. Meftah,[1] A. Irbah,[1] A. Hauchecorne,[1] P. Keckhut,[1] E. Quémerais[1]

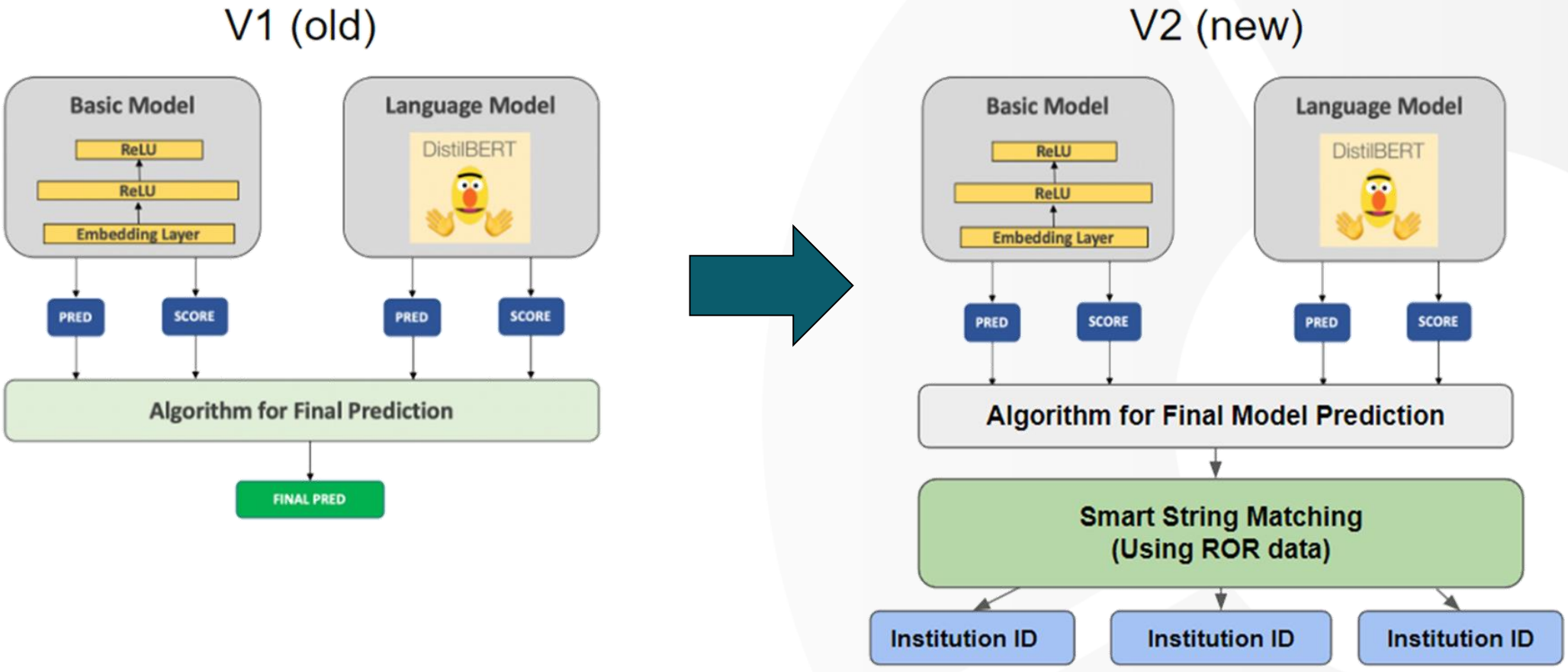
[1]LATMOS, Institut Pierre-Simon Laplace, CNRS, Univ. Versailles Saint-Quentin en Yvelines (France)

'LATMOS, Institut Pierre-Simon Laplace, CNRS, Univ. Versailles Saint-Quentin en Yvelines (France)"

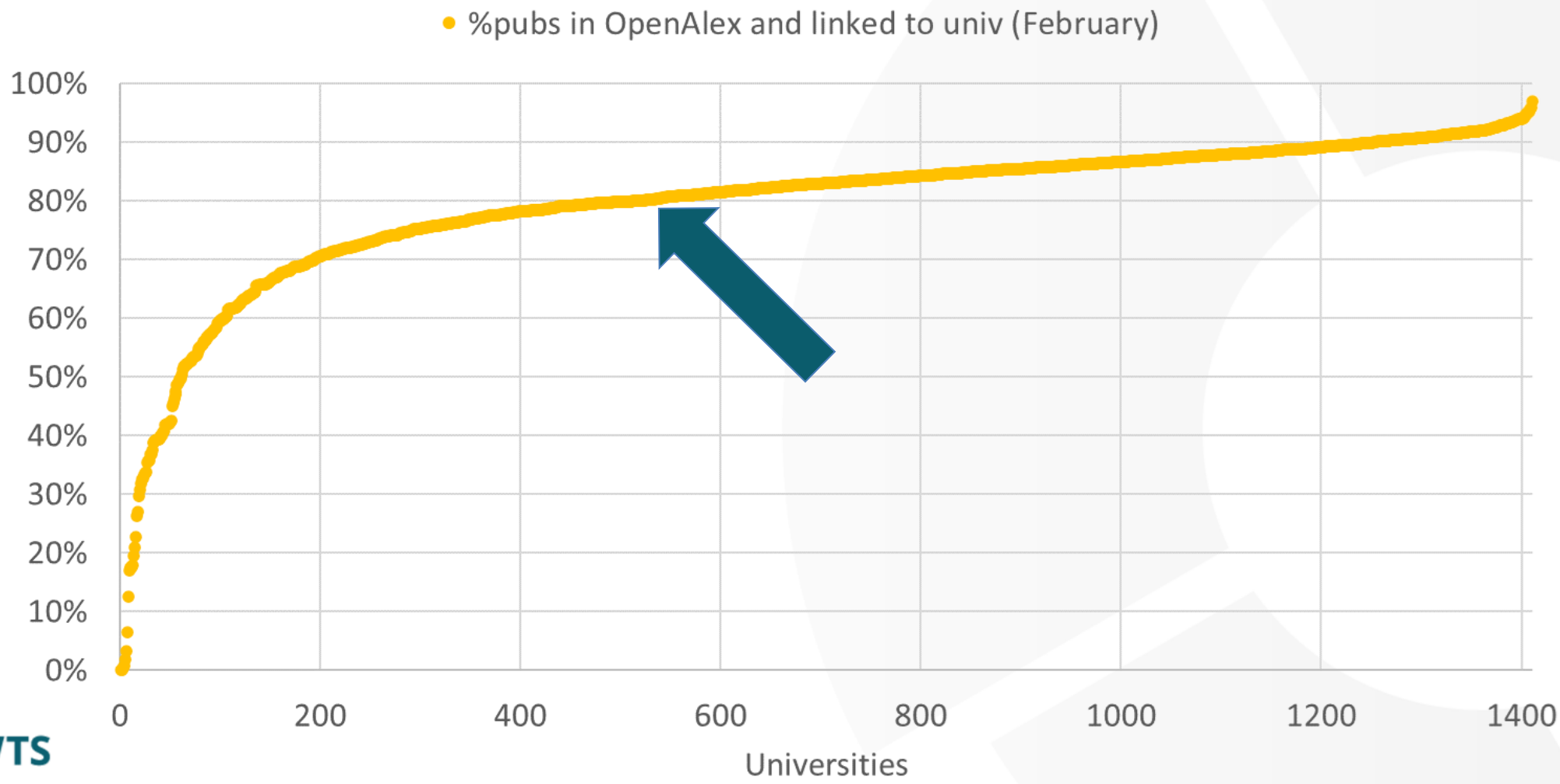
5ppf7q77 02haar591 02feahw73 03mkjjy25

Affiliations may refer to multiple non-hierarchical related institutions

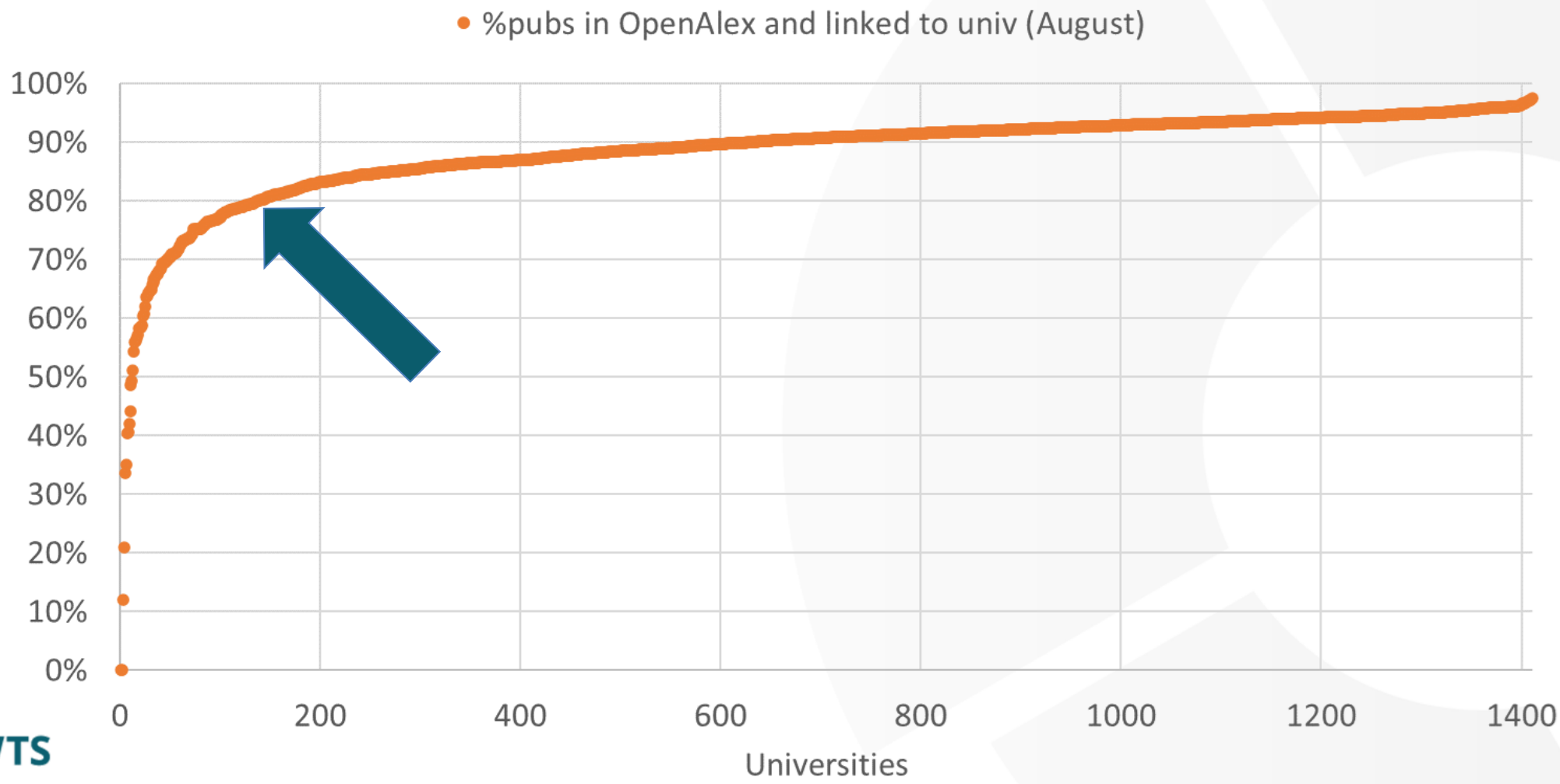
Improved OpenAlex institution parsing and linking system



Publication coverage of LR2023 universities: False negatives in OpenAlex February snapshot

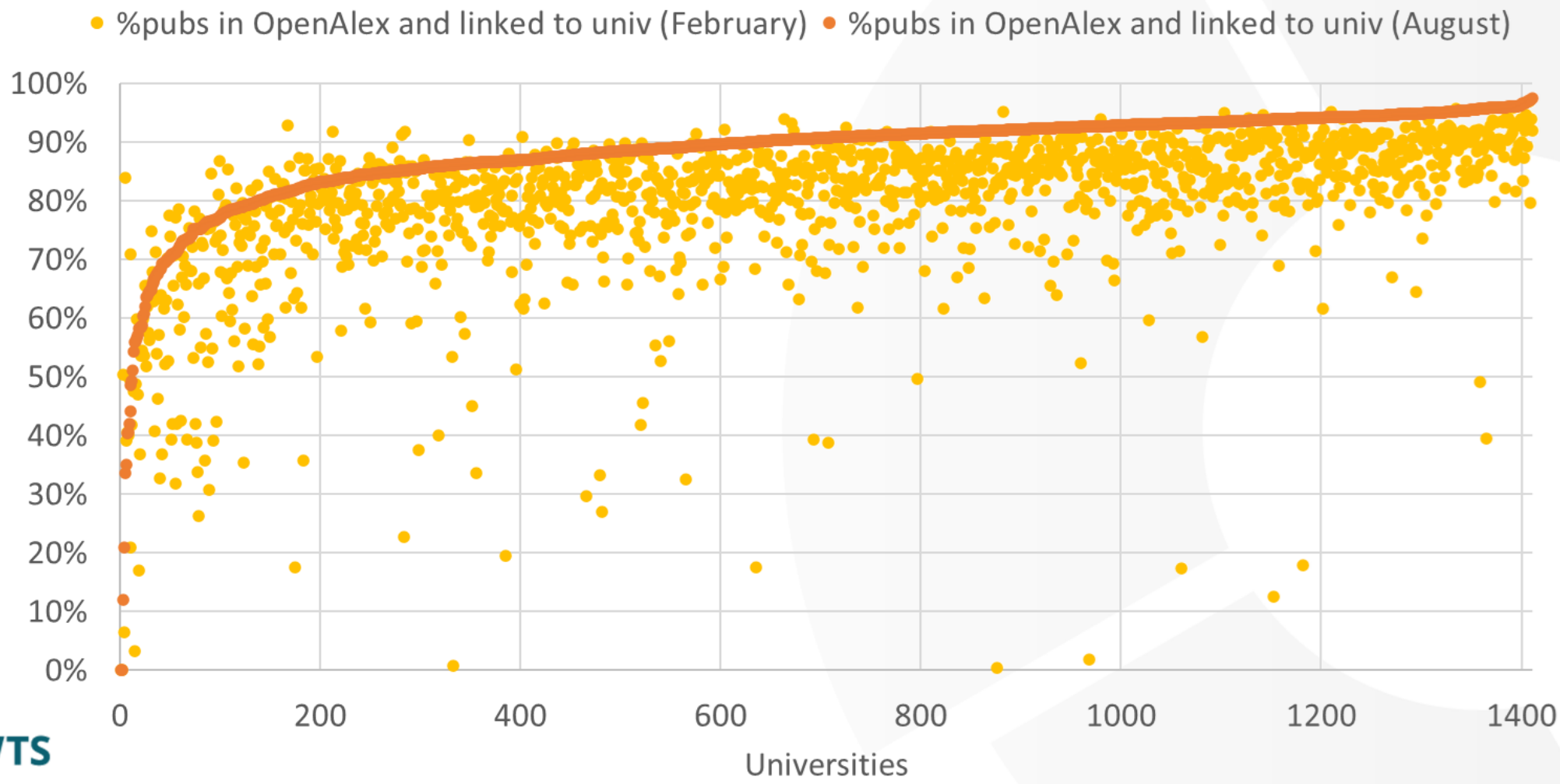


Publication coverage of LR2023 universities: False negatives in OpenAlex August snapshot



CWTS

Publication coverage of LR2023 universities: Improvement of false negatives in OpenAlex



CWTS

Main elements of the Leiden Ranking methodology

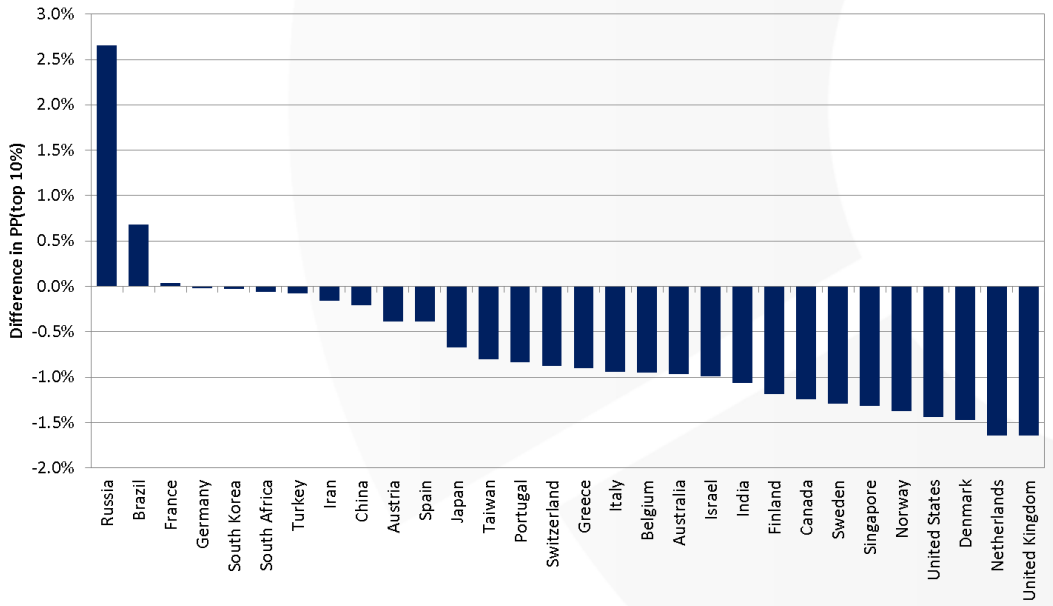
Linking publications to universities

Identification of core publications and sources

Creating a publication classification

Is more data always better?

- Universities from Brazil, China, Russia, France, Germany, etc. may not always benefit from having more data
- To perform meaningful comparisons, the Leiden Ranking bases citation impact indicators on a restricted database of publications
- Leiden Ranking excludes national scientific journals, trade journals, and popular magazines



Identification of core publications and sources

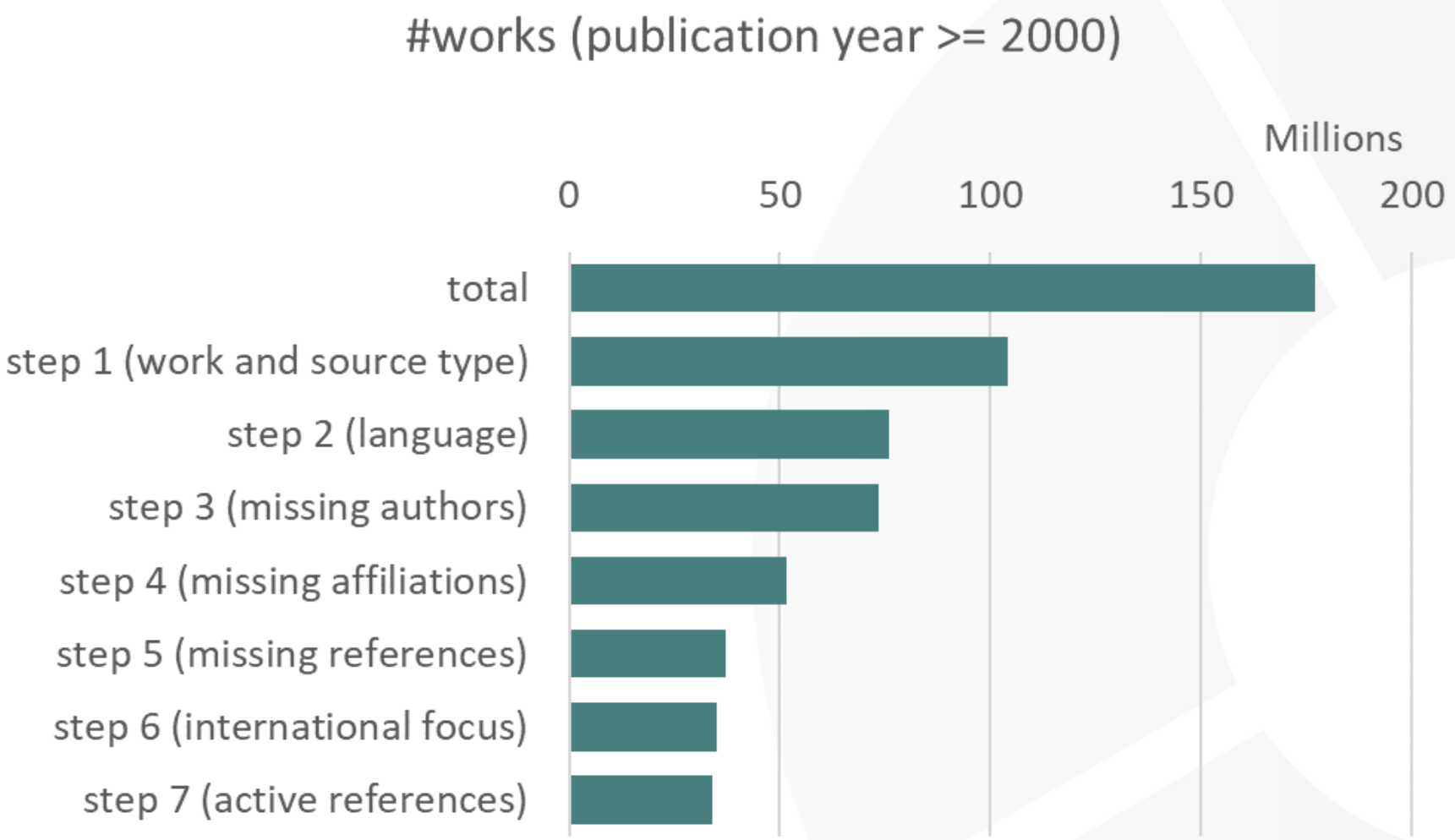
- Issues that have been addressed:
  - Works without a work type
  - Incorrectly classified sources
  - Works with a missing primary source assignment
  - Journal articles incorrectly assigned to a repository
- Other improvements:
  - Availability of language
  - Availability of countries
  - Modified work type
  - First attempt to identify errata/corrections (identification of letters and editorials will be implemented later on)
  - Improved identification of institutions

CWTS

Adjusted methodology based on OpenAlex

- Exclude publications that:
  1. do not have an article work type and journal source type, or an article/book chapter work type and a book series source type
  2. are not in English
  3. do not have any authors
  4. do not have any affiliations
  5. do not have any references
- Exclude sources that:
  6. do not have a sufficiently international focus (to filter out national journals)
  7. do not have a sufficiently large proportion of publications with active references (to filter out trade magazines, business magazines, and popular scientific magazines)

Core publications

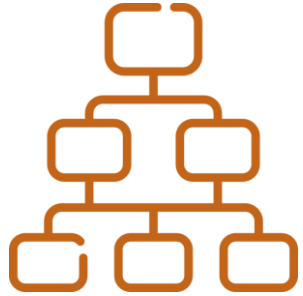


Main elements of the Leiden Ranking methodology

Linking publications to universities

Identification of core publications and sources

Creating a publication classification

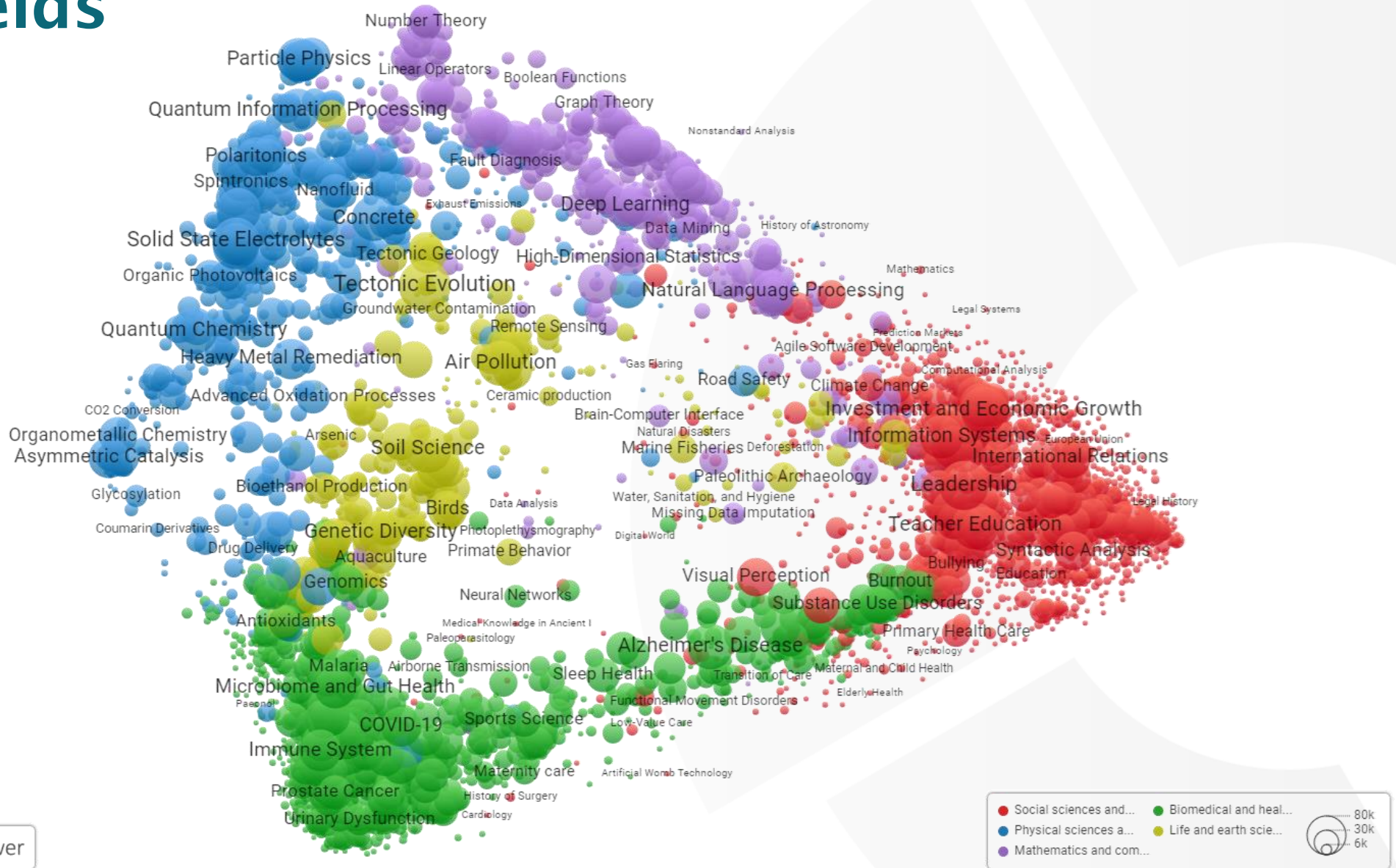


Creating a publication classification

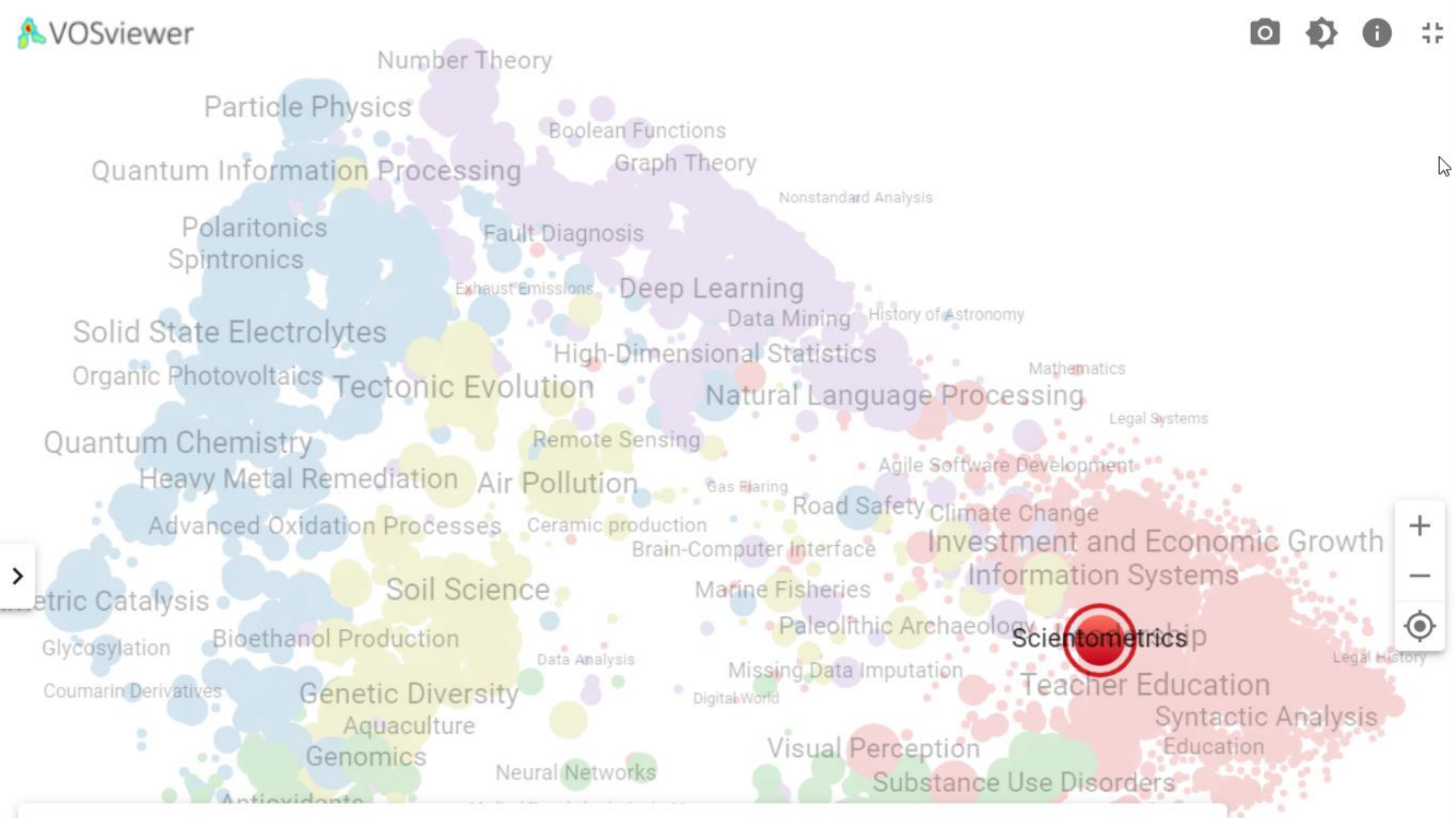
- Publications (not journals) are clustered into research areas based on citation relations
  - 216 million works
  - 1.7 billion citation relations
- Research areas are defined at different levels of granularity and are organized hierarchically
  - 5 main fields
  - 22 broad disciplines
  - 915 fields
  - 4,520 subfields
- Clustering is performed using the Leiden algorithm (Traag, Waltman & Van Eck, 2019)

CWTS

Breakdown of scientific literature into 4,520 subfields



VOSviewer



Find:

Publication	Cit
J. E. Hirsch (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569-16572. DOI, OpenAlex	7940
Nees Jan van Eck; et al. (2009). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523-538. DOI, OpenAlex	6979
Massimo Aria; et al. (2017). bibliometrix : An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. DOI, OpenAlex	3150
Chaomei Chen (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literat... *Journal of the Association for Information Science and Technology*, 57(3), 359-377. DOI, OpenAlex	2628
Matthew E. Falagas; et al. (2007). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *The FASEB Journal*, 22(2), 338-342. DOI, OpenAlex	2409
Stefan Wuchty; et al. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*, 316(5827), 1036-1039. DOI, OpenAlex	2376
Corinne A. Moss-Racusin; et al. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences of the United States of America*, 109(41), 16474-16479. DOI, OpenAlex	2079
Ivan Zupic; et al. (2014). Bibliometric Methods in Management and Organization. *Organizational Research Methods*, 18(3), 429-472. DOI, OpenAlex	2026
Eugene Garfield (2006). The History and Meaning of the Journal Impact Factor. *JAMA*, 295(1), 90. DOI, OpenAlex	1980
Philippe Mongeon; et al. (2015). The journal coverage of Web of Science and Scopus: a comparative analysis.	1692

Full name: Scientific Publishing and Research Evaluation
Main fields: Social sciences and humanities; Mathematics and computer science
Sources: Scientometrics; Nature; arXiv (Cornell University); Journal of Informetrics; Social Science Research Network
Keywords: scientific publishing; research evaluation; bibliometrics; citation analysis; peer review; altmetrics; open access; scholarly communication; academic performance; knowledge transfer
Summary: This cluster of papers focuses on various aspects of scientific publishing and research evaluation. It covers topics such as bibliometrics, citation analysis, peer review, altmetrics, open access, scholarly communication, academic performance, and knowledge transfer. The papers explore issues related to the conduct and correctness of mathematical publishing, the impact and scope of scientific research, the role of early-career university prestige stratification, the influence of peer effects in academic research, and more. Overall, this cluster examines the challenges and advancements in scientific publishing and research evaluation.

Micro field: Scientometrics | No. of pub. (2000-2022): 68682 | Main field: Social sciences and humanities | URL

Leiden Ranking Open Edition

Preliminary results

How do you expect the results of the open edition of the Leiden Ranking to compare to the results of the traditional closed edition?

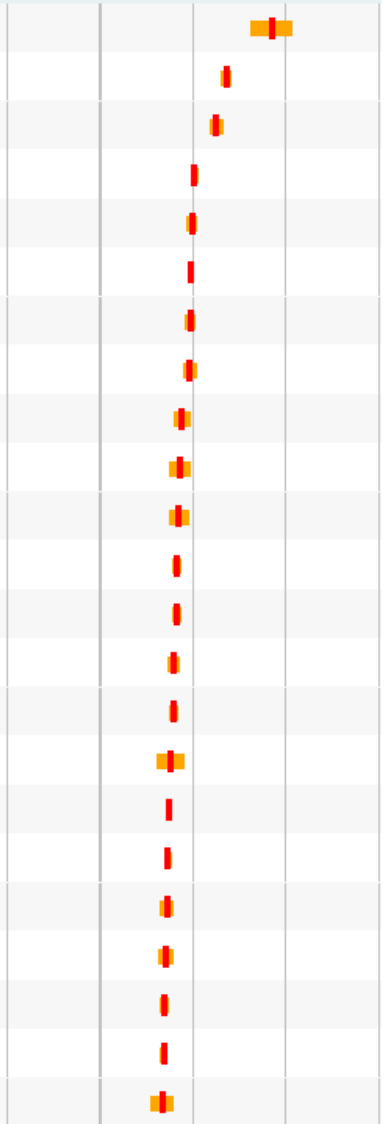
Almost identical results

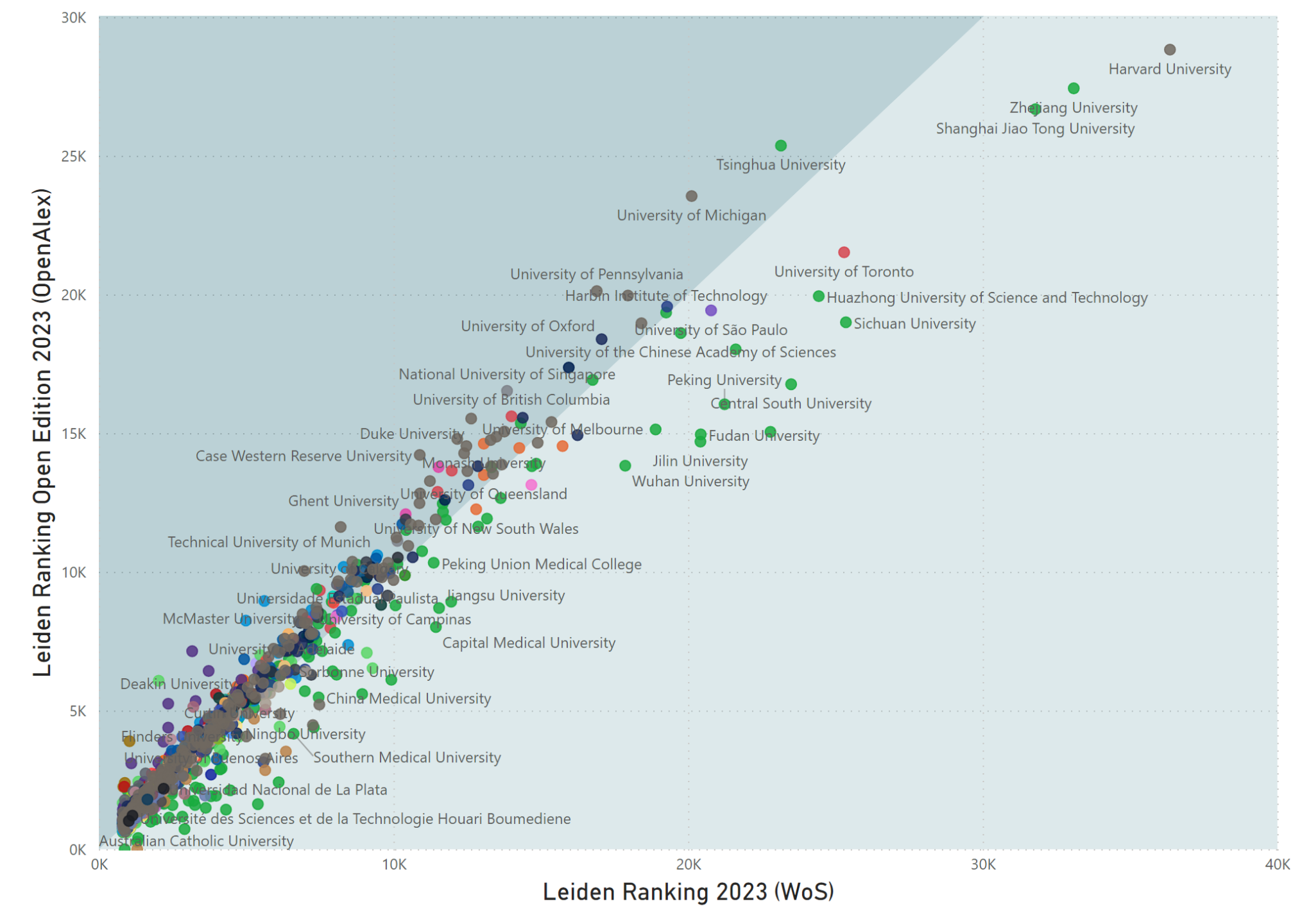
Some similarities and some differences

Completely different results

	University	P	P(top 10%)	PP(top 10%)
1	Harvard Univ	28827	5863	20.3%
2	Zhejiang Univ	27431	3813	13.9%
3	Shanghai Jiao Tong Univ	26681	3429	12.9%
4	Tsinghua Univ	25366	4027	15.9%
5	Univ Michigan	23549	3368	14.3%
6	Univ Toronto	21520	2854	13.3%
7	Univ Penn	20120	3118	15.5%
8	Stanford Univ	19963	3976	19.9%
9	Huazhong Univ Sci & Technol	19939	2971	14.9%
10	Seoul Natl Univ	19565	1751	8.9%
11	Univ São Paulo	19425	1323	6.8%
12	Harbin Inst Technol	19346	2521	13.0%
13	Sichuan Univ	18996	2225	11.7%
14	Johns Hopkins Univ	18961	2746	14.5%
15	Univ Chinese Acad Sci	18609	2533	13.6%
16	Univ Oxford	18388	3234	17.6%
17	Xi'an Jiaotong Univ	18022	2217	12.3%
18	Univ Tokyo	17368	1532	8.8%
19	Tianjin Univ	16921	2201	13.0%
20	Cent S Univ	16764	2323	13.9%
21	Natl Univ Singapore	16524	2818	17.1%
22	Peking Univ	16044	2284	14.2%
23	Univ British Columbia	15607	1993	12.8%

	University	P	P(top 10%)	PP(top 10%)
1	Rockefeller Univ	928	266	28.7%
2	MIT	10940	2599	23.8%
3	Princeton Univ	5487	1247	22.7%
4	Harvard Univ	28827	5863	20.3%
5	Univ California - Berkeley	10135	2035	20.1%
6	Stanford Univ	19963	3976	19.9%
7	Hunan Univ	9030	1795	19.9%
8	Caltech	5617	1112	19.8%
9	King Abdullah Univ Sci & Technol	3901	741	19.0%
10	Weizmann Inst Sci	2714	509	18.8%
11	Duy Tan Univ	1214	226	18.6%
12	Nanyang Technol Univ	11115	2047	18.4%
13	ETH Zurich	10516	1934	18.4%
14	Ecole Polytech Féd Lausanne	6400	1161	18.1%
15	Hong Kong Polytech Univ	8467	1528	18.1%
16	Australian Cath Univ	1002	178	17.7%
17	Univ Oxford	18388	3234	17.6%
18	Univ Cambridge	15556	2723	17.5%
19	Hong Kong Univ Sci & Technol	2915	506	17.4%
20	Univ California - Santa Barbara	4730	817	17.3%
21	Univ California - San Francisco	11713	2002	17.1%
22	Natl Univ Singapore	16524	2818	17.1%
23	Changsha Univ Sci & Technol	1929	326	16.9%





CWTS
Meaningful metrics

Period

2018-2021

Field

All sciences

Country

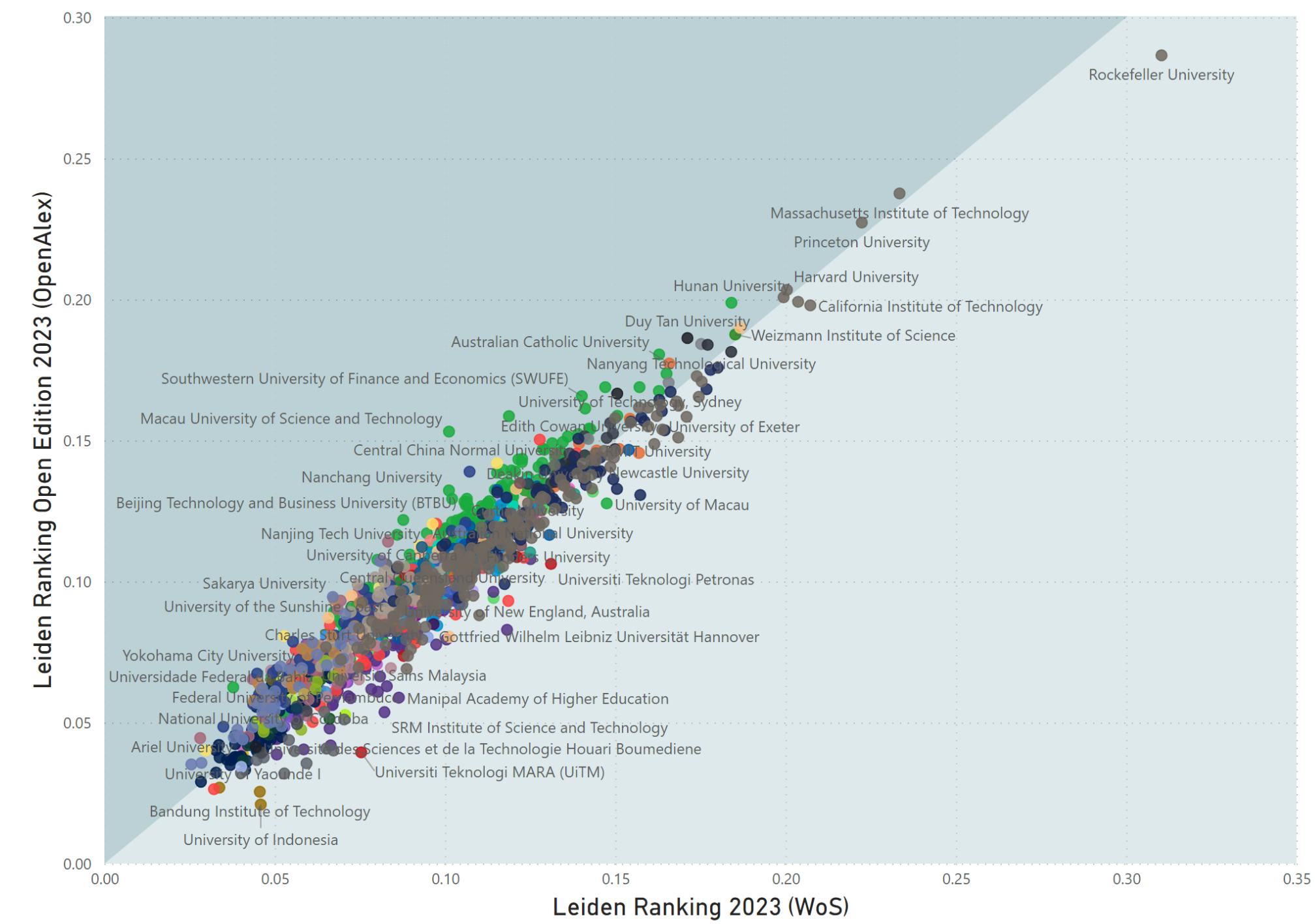
All

Counting method

Fractional counting

Indicator

P



CWTS
Meaningful metrics

Period

2018-2021

Field

All sciences

Country

All

Counting method

Fractional counting

Indicator

PP(top 10%)

CWTS
Meaningful metrics

Period

2018-2021

Field

All sciences

Country

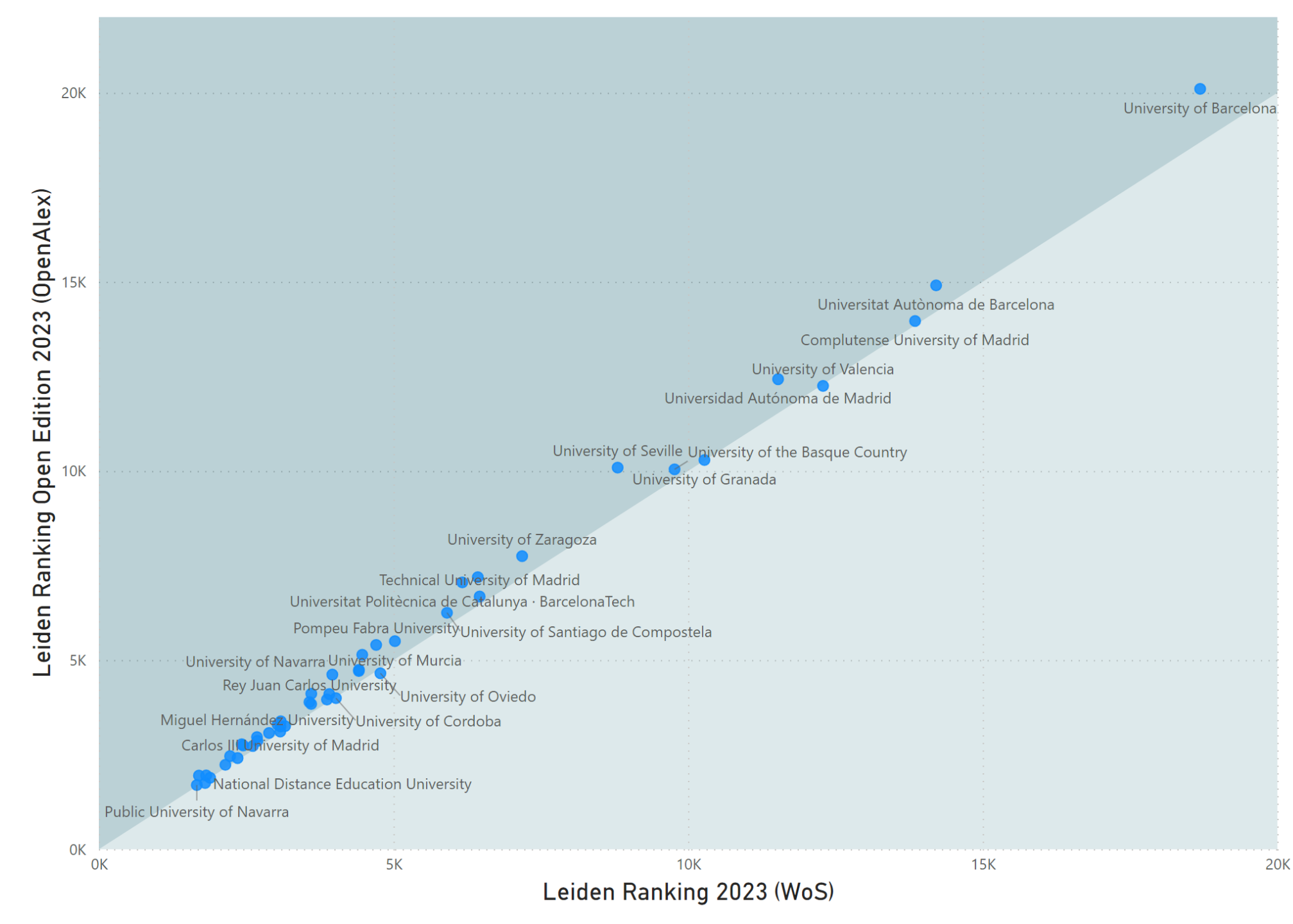
Spain

Counting method

Fractional counting

Indicator

P(top 10%)



CWTS
Meaningful metrics

Period

2018-2021

Field

All sciences

Country

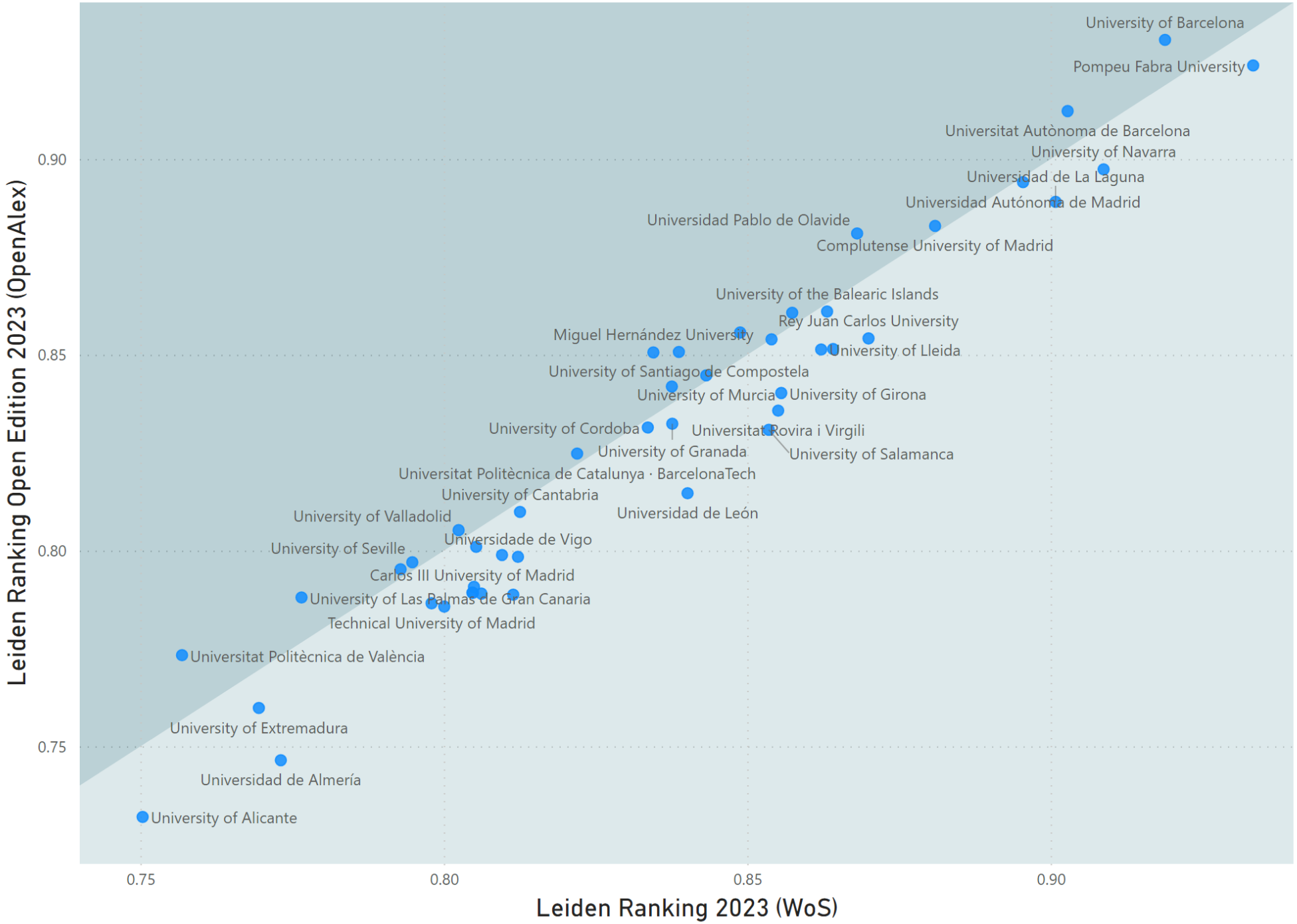
Spain

Counting method

Full counting

Indicator

P(collab)



CWTS
Meaningful metrics

Period

2018-2021

Field

All sciences

Country

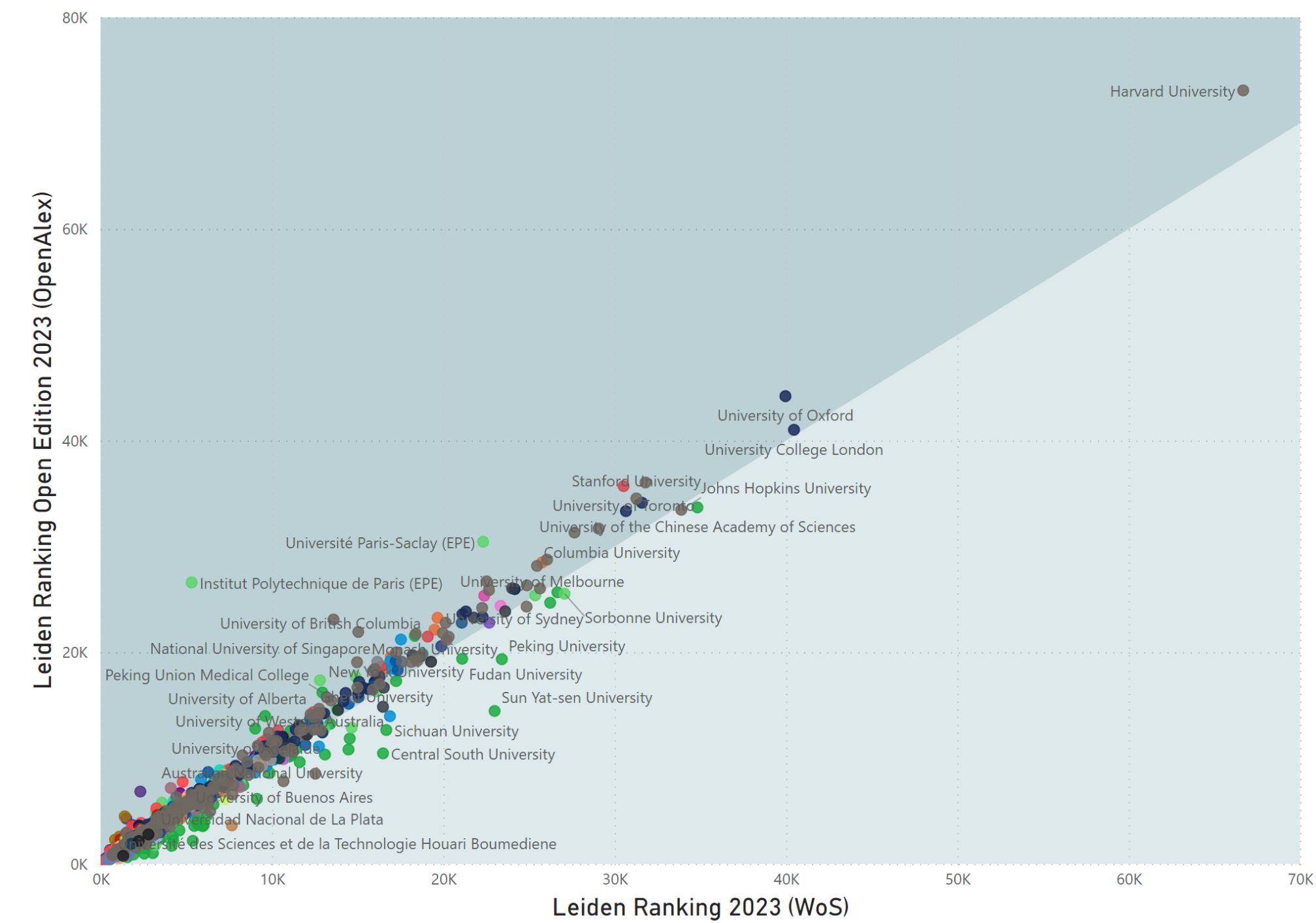
Spain

Counting method

Full counting

Indicator

PP(collab)



CWTS
Meaningful metrics

Period

2018-2021

Field

All sciences

Country

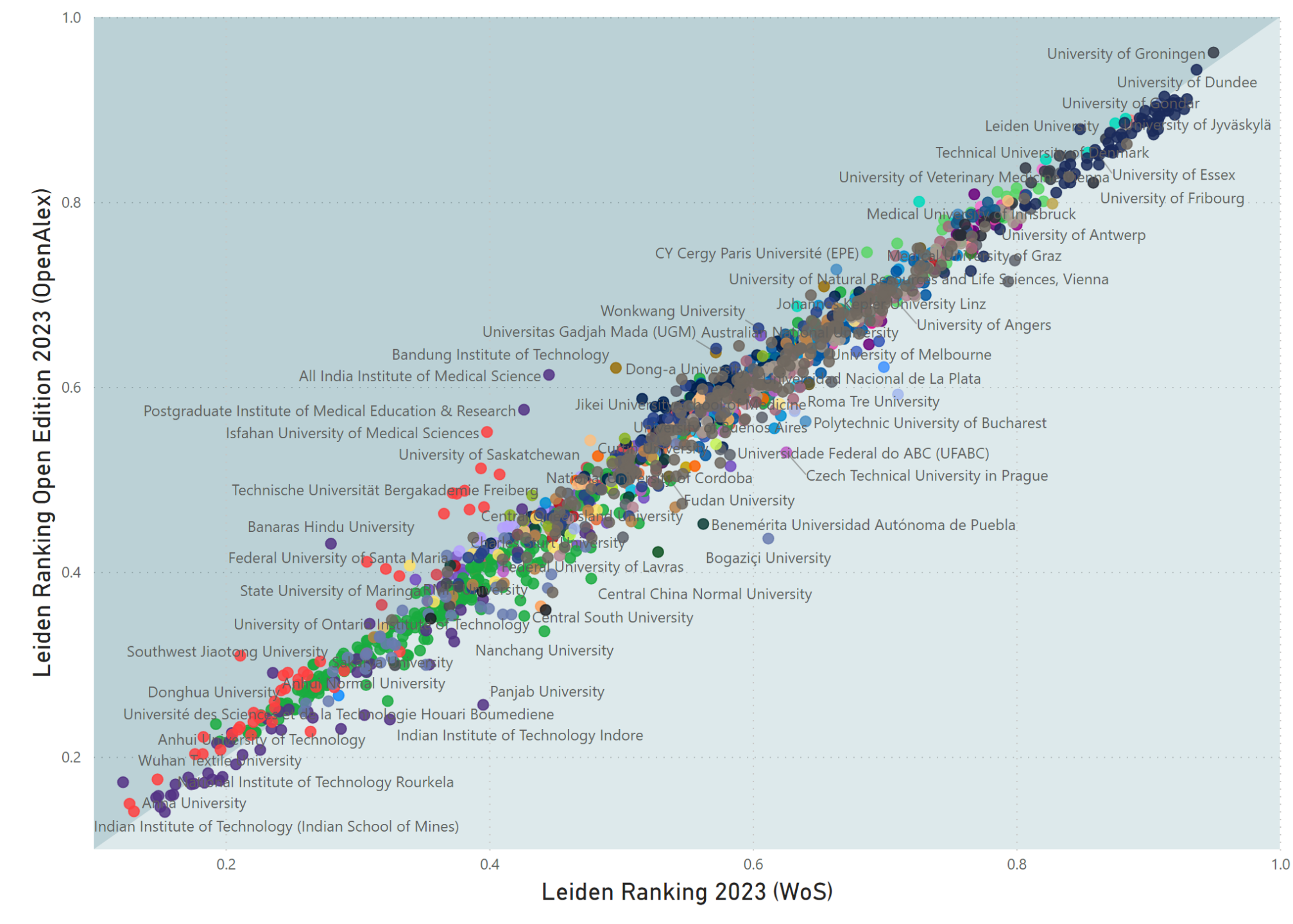
All

Counting method

Full counting

Indicator

P(OA)



CWTS
Meaningful metrics

Period

2018-2021

Field

All sciences

Country

All

Counting method

Full counting

Indicator

PP(OA)

Conclusions

- Open data sources democratize access to research information and provide the transparency that is needed for responsible decision making
- Inadequate coverage and quality of the data is sometimes given as a reason not to use open data sources yet
- Do not sit back and wait until everything is perfect
- By actively using open data sources, problems are made transparent and improvements can be implemented more quickly
- The Leiden Ranking Open Edition demonstrates how open data sources offer an improvement over traditional proprietary data sources!

CWTS

Stay tuned for the launch of the Leiden Ranking Open Edition in January 2024!