QiW

Qualität in der Wissenschaft

Zeitschrift für Qualitätsentwicklung in Forschung, Studium und Administration

Bibliometrie im Fokus

- SYoS & TOBI: Two projects to promote the implementation of a scientometric infrastructure for Switzerland
- Reframing scientometrics: How ontological understanding of science influences what we count and how we interpret it
- Towards transparent, democratic, and open research information: The evolving role of CWTS in the Netherlandsmetrics
- The Representation of Swiss Higher Education Institutions in Five Bibliometric Databases

4
2024

UVW UniversitätsVerlagWebler

18. Jahrgang
ISSN 1860-3041
www.universitaetsverlagwebler.de

Herausgeber*innenkreis

Hinweise für die Autor*innen

In dieser Zeitschrift werden i.d.R. nur Originalbeiträge publiziert. Sie werden doppelt begutachtet. Die Autor*innen versichern, den Beitrag nicht zu gleicher Zeit an anderer Stelle zur Publikation angeboten zu haben. Beiträge werden nur dann angenommen, wenn die Autor*innen den Gegenstand nicht in vergleichbarer Weise in einem anderen Medium behandeln. Senden Sie bitte das Manuskript als Word-Datei und Abbildungen als JPG-Dateien per E-Mail an die Redaktion (Adresse siehe Impressum).

Wichtige Vorgaben zu Textformatierungen und beigefügten Fotos, Zeichnungen sowie Abbildungen erhalten Sie in den Autor*innenhinweisen auf unserer Website: www.universitaetsverlagwebler.de.
Ausführliche Informationen zu den in diesem Heft aufgeführten Verlagsprodukten erhalten Sie ebenfalls auf der zuvor genannten Website.

Impressum

Verlag, Redaktion, Abonnementverwaltung:
UVW UniversitätsVerlagWebler
Der Fachverlag für Wissenschaft und Hochschule
Reepeweg 5, 33617 Bielefeld
Tel.: 0521-92 36 10-0

Satz: UVW, info@universitaetsverlagwebler.de

Anzeigen: Die Zeitschrift „Qualität in der Wissenschaft" veröffentlicht Verlagsanzeigen, Ausschreibungen und Stellenanzeigen. Aufträge sind an den Verlag zu richten. Die jeweils gültigen Anzeigenpreise erhalten Sie auf Anfrage beim Verlag.

Redaktionsschluss dieser Ausgabe: 11.12.2024

Umschlagsgestaltung: Wolff-Dietrich Webler, Bielefeld
Gesetzt in der Linotype Syntax Regular

Druck: Sievert Druck & Service GmbH,
Potsdamer Str. 220, 33719 Bielefeld

Abo-Bestellungen und die Bestellungen von Einzelheften sind an den Verlag zu richten. Eine Abo-Bestellvorlage finden Sie unter www.universitaetsverlagwebler.de.
Das Jahresabonnement verlängert sich automatisch um ein Jahr, wenn es nicht 6 Wochen vor Jahresende gekündigt wird.

Erscheinungsweise: 4mal jährlich

18. Jahrgang
ISSN 1860-3041

QiW

Qualität in der Wissenschaft

Zeitschrift für Qualitätsentwicklung in Forschung, Studium und Administration

Editorial

Qualitätsforschung

Qualitätsentwicklung /-politik

Calls for papers

Seitenblick auf die Schwesterzeitschriften

4 | 2024

Rüdiger Mutz

David Johann

An der evaluativen Bibliometrie entzündete sich in der Vergangenheit immer wieder Kritik. Für viele Fehlentwicklungen in der Wissenschaft, wie z.B. die Fokussierung auf Quantität statt Qualität, mangelnde Replikation von Ergebnissen, Beeinflussung von Metriken («Gaming») oder sogar wissenschaftliches Fehlverhalten, wird fast stereotyp die Bibliometrie und ihre Indikatoren (z.B. h-index, Journal Impact Factor) mitverantwortlich gemacht. Wichtige Deklarationen im Bereich des Research Assessment wie DORA (2012) oder COARA (2022) wurden letztlich auch in Abgrenzung zur traditionellen Bibliometrie abgefasst. Dabei wird jedoch übersehen, dass nach Goodhart`s Law jeder Indikator, der zum Kriterium einer Evaluation gemacht wird, in Wettbewerbssituationen selbst zum Ziel wird und damit seine Bedeutung als Indikator verliert, und dass es die Verantwortlichen in Hochschulen und Wissenschaft sind, die solche Indikatoren zu Zielkriterien erkoren haben. Die Quantitativen Wissenschaftsforschung (QSS) und die Bibliometrie weisen schon lange darauf hin, dass bestimmte Indikatoren wie beispielsweise der h-Index als Maß für die individuelle Forschungsleistung aufgrund methodischer Probleme ungeeignet sind, weshalb sie zumindest in diesen Disziplinen längst an Bedeutung verloren haben. Auch der Journal Impact Factor wird in der QSS und der Bibliometrie wieder in seiner ursprünglichen Bedeutung als bibliometrisches Maß für den Impact einer wissenschaftlichen Zeitschrift verwendet, und nicht als Maß für die Qualität der Arbeit von Forschenden.

Vor dem Hintergrund der häufig geäusserten Kritik an der Bibliometrie und ihrer Indikatoren ist es wichtig, darüber zu informieren, was Bibliometrie und ihre Indikatoren in verschiedenen Kontexten leisten können und wo ihre Grenzen liegen. Diesem Ziel ist die vorliegende Ausgabe der QiW gewidmet. Den Rahmen bildet die Veranstaltungsreihe „Swiss Year of Scientometrics" (SYoS), die von der ETH-Bibliothek der ETH Zürich in den Jahren 2023 bis 2024 durchgeführt und von swissuniversities, dem Pendant zur Hochschulrektorenkonferenz in Deutschland, als kompetitiv eingeworbenes Projekt finanziell gefördert wurde. Im Rahmen des Projekts konnten international renommierte Bibliometrikerinnen und Bibliometriker für Vorträge und Workshops gewonnen werden (https://yearofscientometrics.ethz.ch). Aus dieser Veranstaltungsreihe ist nun eine QiW-Ausgabe mit vier Beiträgen entstanden, die die Vielschichtigkeit der Bibliometrie und ihrer Anwendung exemplarisch widerspiegelt.

Diese Ausgabe beginnt mit einem Beitrag von *David Johann, Annette Guignard* und *Simon Willemin* mit dem Titel **SYoS & TOBI: Two projects to promote the implementation of a scientometric infrastructure for Switzerland.** Darin werden SYoS und ein weiteres Projekt zur Förderung einer szientometrischen Forschungsinfrastruktur in der Schweiz vorgestellt. **Seite 98**

Emanuel Kulcyzcki beleuchtet in seinem Beitrag **Reframing scientometrics: How ontological understanding of science influences what we count and how we interpret it** die in Vergessenheit geratenen russischen Wurzeln der Bibliometrie. Er vergleicht den eher atomistisch, westlichen Ansatz von de Solla Price mit dem (sowjet-)russischen, kollektivistischen Ansatz von Vasily Nalimov und untersucht ihre Bedeutung für die aktuelle Diskussion über den Einsatz von Bibliometrie. **Seite 102**

Simon Willemin und *Ludo Waltman* zeigen in ihrem Beitrag **Towards transparent, democratic, and open research Information: The evolving role of CWTS in the Netherlandsmetrics**, wie sich die Rolle und die Dienstleistungen des CWTS in Leiden, eines der weltweit führenden Zentren im Bereich QSS und Bibliometrie, mit den gesellschaftlichen Veränderungen in den Niederlanden gewandelt haben, etwa in Form einer Erweiterung des Aufgabenspektrums auf politische Entscheidungen und erweiterter Formen des Research Assessments. So sollen beispielsweise Metadaten zu Publikationen wie Zitationen, die bisher lizenzpflichtigen Datenbanken vorbehalten waren, öffentlich zugänglich gemacht werden. **Seite 109**

Den Abschluss bildet der Beitrag **The Representation of Swiss Higher Education Institutions in Five Bibliometric Databases** von *Julian Dederke, Michelle Koch* und *Simon Willemin*, der Überlegungen von Willemin und Waltman zu offenen Metadaten aufgreift und untersucht, inwiefern fünf bibliometrische Datenbanken, darunter zwei frei zugängliche (OpenAlex, OpenAIRE), für bibliometrische Analysen nutzbar sind und inwiefern die Schweizer Hochschulen darin repräsentiert sind. Das Vorgehen könnte auch für andere Länder und Institutionen von Interesse sein. **Seite 117**

Wir bedanken uns bei Professor Klaus Jonas für die finanzielle Unterstützung bei der Open-Access-Ausgabe.

Rüdiger Mutz und David Johann

David Johann, Annette Guignard & Simon Willemin

SYoS & TOBI

Two projects to promote the implementation of a scientometric infrastructure for Switzerland

David Johann

Annette Guignard

Simon Willemin

Unlike other countries, Switzerland lacks a network in which members of HEIs with scientometric expertise work together. In addition, it does not have a curated and comprehensive database of bibliometric data available to Swiss universities. ETH Zurich has launched two projects (SYoS and TOBI) to promote the implementation of a scientometric infrastructure for Switzerland. This article presents the objectives and approaches of both projects along with a discussion of some initial results.

1. Introduction

Similar to other countries, the Swiss science system has been influenced by the establishment of New Public Management in the late 1990s and early 2000s. As a result, researchers and universities today operate in a "quasi-market" in which they compete for scarce resources. Competition is desired by politics with the aim to increase the efficiency and effectiveness of researchers and universities internationally. Performance-based funding decisions are made using standardized criteria, which often include bibliometric indicators, such as the count of research articles in a selection of peer reviewed journals, or the proportion of publications amongst the most frequently cited ones (e.g. Nievergelt 2013; Enders et al. 2015; Johann et al. 2022).

In the two last decades, scientometric data and indicators have become more important. At first glance, this may seem plausible, because scientometric data and indicators can provide useful information and be a useful tool for managing universities and the science system, as long as they are used correctly, i.e. contextualized and interpreted appropriately. However, scientometric data and indicators require sensible management, analysis, and interpretation to be the basis of meaningful policy decisions.

It seems necessary to set up centers or networks that provide high-quality scientometric data, i.e., data that has been curated and cleansed of errors and inaccuracies, and comprehensive expertise in the analysis and interpretation of scientometric data – e.g. for data-based research evaluation. Examples from countries outside Switzerland include the Centre for Science and Technology Studies (CWTS) in the Netherlands and the Competence Network for Bibliometrics in Germany.

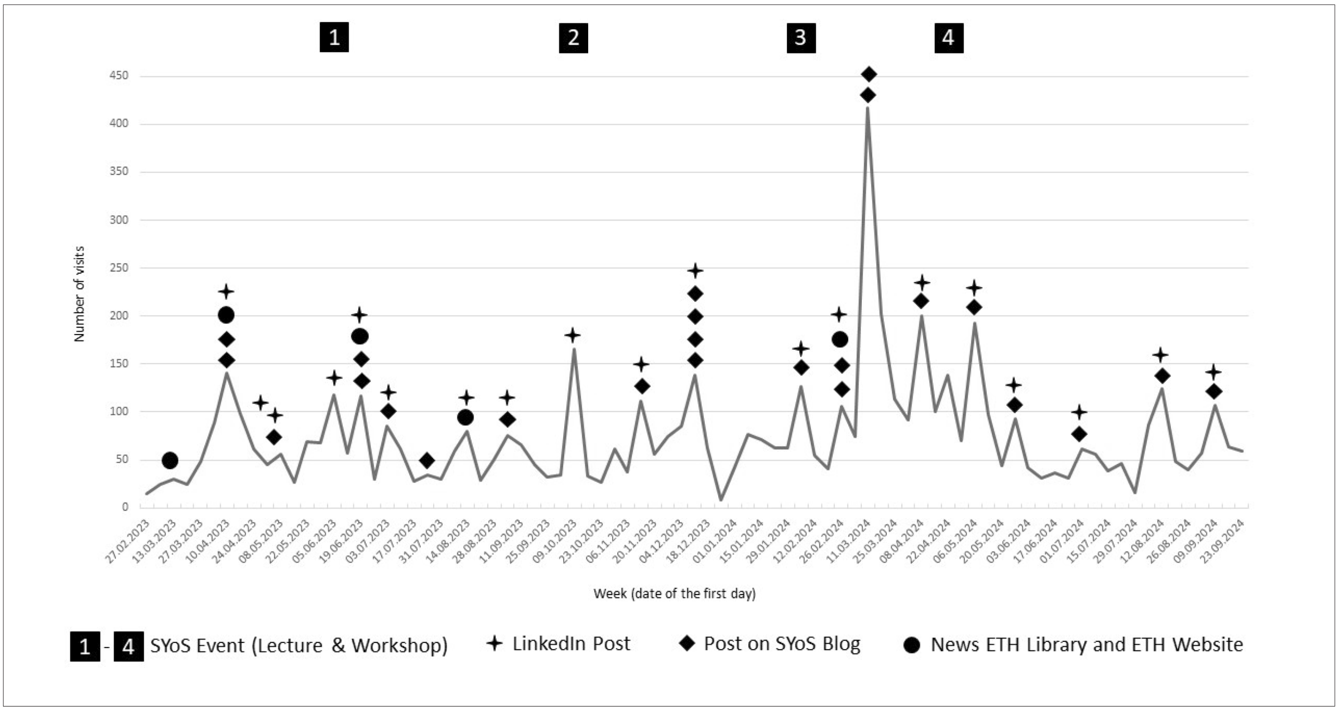
Given the paramount importance of science and universities for Switzerland, it may be surprising that Switzerland has neither established a competence network or center for scientometric analysis, nor set up a curated bibliometric database, yet. This has prompted the ETH Library to launch two projects working towards founding scientometric structures in Switzerland and exploring the possibility of using open and transparent bibliometric data sources for Switzerland: "Swiss Year of Scientometrics" (SYoS) and "Towards Open Bibliometric Indicators" (TOBI).

2. Swiss Year of Scientometrics (SYoS)

The Swiss Year of Scientometrics (SYoS) has several objectives:

- Establishing a network of relevant actors within the Swiss science system who deal directly or indirectly with scientometric data and analyses. Relevant actors include researchers in the fields of higher education and science studies, Swiss universities management and university libraries, as well as staff from other relevant Swiss institutions such as the

Fig. 1: Timeline of the SYoS project including information about communication activities



The figure also shows the weekly number of visits to the SYoS blog from February 2023 to September 2024. Please note: No distinction was made between internal (project staff) and external visitors when counting visits.

Swiss Science Council, the Swiss Library Service Platform (SLSP) or the Swiss Secretariat for Education, Research and Innovation (SERI).
- Raising awareness of the potential of scientometric services for Swiss research institutions.
- Investigating the possibilities, requirements, and restrictions for scientometric services in Switzerland – also in light of an increasing importance of open science.
- Promoting the transparent and responsible use of scientometric data and indicators in Switzerland – also considering the meaning of the San Francisco Declaration on Research Assessment (DORA) for Swiss universities.

To achieve the objectives, SYoS hosted a series of four 90-minute lectures with Q&A, dedicated to specific topics relating to the above goals, and featuring internationally renowned speakers, who shared their expertise on the following topics:

- Opportunities and challenges of scientometrics: Diversification of data sources and applications (Speaker: Prof Stefanie Haustein)
- Reframing scientometrics: How ontological understanding of science influences what we count and how we interpret it (Speaker: Prof Emanuel Kulczycki; for more information on the lecture, see Emanuel Kulczycki's contribution in this journal)
- The responsible assessment of open science (Speaker: Dr Elizabeth Gadd)
- Openness of research information – Democratizing the use of scientometrics (Speaker: Prof Ludo Waltman; for more information on the lecture, see the contribution by Ludo Waltman and Simon Willemin in this journal)

The lectures of Stefanie Haustein, Emanuel Kulczycki and Ludo Waltman, that took place at ETH Zurich, are available on the ETH Zurich video portal.[1]

Each of the lectures was accompanied by a one-day-workshop held the day after the lecture. While participation in the lecture was open to everyone interested, the number of participants in the workshops was limited to 20-30 contributors. The workshop allowed the participants to explore the respective topic in more depth through a chaired discussion. The participants came from different backgrounds and institutions, providing insights from different perspectives.

In addition, a blog[2] was set up and curated reporting on the core points from the lecture and workshop series (see Figure 1). It also covered thematic posts that addressed topics of key interest to the stakeholders (e.g., food for thought for decision-makers at universities). New blog posts were also announced on LinkedIn to increase visibility. The posts covered topics, such as "Shenanigans with Impact Factors"[3] or "Hierarchi-

1 ETH Zurich video portal: https://video.ethz.ch/
2 SYoS blog: https://yearofscientometrics.ethz.ch/.
3 Shenanigans with Impact Factors: https://yearofscientometrics.ethz.ch/?s=shenanigans

cal Classification System of Research Fields to Understand Research Interconnectedness"[4]. The blog also discussed bibliometric practices and activities in other countries (Austria: University of Vienna[5]).

3. Towards Open Bibliometric Indicators (TOBI)

Towards Open Bibliometric Indicators (TOBI) is closely linked to SYoS. Until recently, bibliometric analyses were almost exclusively based on data from commercial databases, such as the Web of Science and Scopus. This has one core disadvantage: Access to these databases is expensive and restrictive, which is in direct contrast with the DORA declaration, demanding openness and transparency. The goal of TOBI is to identify alternative open and transparent data sources for bibliometric indicators and analyses for Swiss higher education institutions. As such, TOBI contributes to promoting a transparent, responsible, and DORA-compliant use of scientometric data, which is one of the core objectives of SYoS. Furthermore, TOBI embraces the idea of Open Science, which is becoming increasingly more important in Switzerland.

TOBI's approach is as follows: First, a large number of open data sources containing bibliometric data, altmetric data, and other relevant research information were identified. The most promising sources to compute bibliometric indicators were then extracted from these sources and, finally, analyzed in depth.

The project has analyzed the quality of the most important metadata for bibliometric analyses at the publication level. Guiding questions were:

- Have the metadata of publications retrieved from different sources been deduplicated?
- Is the type of publication (e.g. article, book, data) provided for each publication?
- Is information on the authors' affiliations available, and accurate, in the data sources?

The analysis of the quality of the metadata went hand in hand with the identification of possible bias in the data. For example, the language of publication (English vs. other languages), the location of the institutions (French-, Italian- and German-speaking parts of Switzerland), and the type of authors' affiliated institutions (universities, universities of applied sciences, universities of teacher education, other types of higher education institutions) were considered.

TOBI identified several hundreds of open databases, but only a handful were truly global in the sense that they contained more than 10 million bibliographic records and were neither discipline nor country specific. While some of these databases may be useful for in-depth literature research,[6] most of them did not contain affiliation data at all, or only useful data for a few publications. The lack of this information makes the databases largely unsuitable for bibliometric analyses. However, TOBI identified two promising open data sources: OpenAlex and OpenAIRE Graph. Both had a good scope and high quality, which makes them suitable for bibliometric analyses of Swiss higher education institutions. Other projects have already implemented open data: For example, Open-Alex has already become the central data source of the Swiss Open Access Monitor[7] and the Centre for Science and Technology Studies (CWTS) has published an open edition of its Leiden Ranking based on OpenAlex[8].

Overall, the lessons learned from TOBI could help decision-makers in the Swiss science sector to select a valid data source for their policy making (see also the contribution by Julian Dederke, Michelle Koch, and Simon Willemin in this journal).

4. Discussion and Outlook

Although the Swiss science system is highly competitive, and metrics play an important role in research assessment in Switzerland, the country has lacked a network and/or infrastructure for scientometrics that collects, provides, and promotes the responsible use of scientometric data and indicators.

SYoS and TOBI have taken major steps towards establishing such a network and a scientometric infrastructure: SYoS events were well attended and reached a broad audience from various Swiss institutions. They also helped establishing a stakeholder network on scientometric issues and raised willingness for further development of scientometrics in Switzerland, including the establishment of a scientometric infrastructure. The considerable numbers of visits to the SYoS blog indicates that the project is recognized (see Figure 1). The awareness should contribute to a more responsible use of scientometric data and indicators in Switzerland. The results of the TOBI project are also promising. They suggest that valid scientometric analyses of the Swiss research landscape is possible with open bibliometric data sources, such as OpenAlex. Expertise built on the basis of TOBI can be used to further develop an open data infrastructure for Swiss research institutions in the future.

4 Hierarchical Classification System of Research Fields to Understand Research Interconnectedness: https://yearofscientometrics.ethz.ch/a-hierarchical-classification-system-of-research-fields-to-understand-research-interconnectedness/

5 Austria: University of Vienna: https://yearofscientometrics.ethz.ch/change-is-the-only-constant-heraclitus-bibliometric-practices-and-activities-at-the-university-of-vienna-reloaded/

6 For instance, BASE has been used as a source for Open Knowledge Maps (see Kraker et al. 2024).

7 Swiss Open Access Monitor: https://oamonitor.ch/charts-data/jounral-monior/

8 CWTS Leiden Ranking Open Edition: https://open.leidenranking.com

Next steps are currently discussed with the network stakeholders. The successful SYoS events on scientometric topics will continue with an annual series of lectures and workshops that will take place at ETH Zurich and in other Swiss Higher Education Institutions. This will maintain the possibility for further exchange for the scientometric community in Switzerland in the future. In addition, planning for a suitable organizational structure for future collaboration between stakeholders has begun. A project to develop metadata standards that allows us to facilitate comparative bibliometric analyses has started and should help ensuring that Swiss research results are more visible internationally.

Acknowledgement

This article was written within the framework of the projects "Swiss Year of Scientometrics" and "Towards Open Bibliometric Indicators", which are co-funded by swissuniversities and the ETH Library. As such, the article includes information from the project proposal, project reporting, and the project websites of SYoS and TOBI. The authors acknowledge copyediting/proofreading and helpful comments on the manuscript by Kathrin Thomas. Finally, the authors of this article would like to thank Teresa Kubacka for making the TOBI project possible with her ideas and for making a significant contribution to its success.

References

Enders, J./Kehm, B. M./Schimank, U. (2015): 'Turning Universities into Actors on Quasi-Markets: How New Public Management Reforms Affect Academic Research'. In: Jansen, D./Pruisken, I. (eds.): The Changing Governance of Higher Education and Research. Cham, pp. 89–103.

Johann, D./Raabe, I. J./Rauhut, H. (2022): Under pressure: The extent and distribution of perceived pressure among scientists in Germany, Austria, and Switzerland. Research Evaluation, 31(3), pp. 385-409.

Kraker, P., Beardmore, L./Hemila, M./Johann, D./Kaczmirek, L./Schubert, C. (2024): Partizipative Modelle im Zusammenspiel von Bibliotheken und KI-Systemen – Drei Fallstudien zur Integration der visuellen Recherche-Plattform Open Knowledge Maps. b.i.t.online, 27(4), S. 327-335

Nievergelt, B. (2013): Professionalisierungsprozesse in der nationalen und internationalen Hochschullandschaft. Arbeitsplatz Hochschule im Wandel, S. 9-19 https://www.chess.uzh.ch/dam/jcr:ffffffff-ec7e-5bfe-0000-000036f53173/Zoom_ZUW_Nr3_Auszug_B_Nievergelt.pdf (17.04.2020)

■ **David Johann,** Dr. phil., Co-Head of Research Support Services and Head of Research Analytics Service,ETH Library, ETH Zurich,
E-Mail: david.johann@library.ethz.ch

■ **Annette Guignard,** Senior Information Specialist, ETH Library, ETH Zurich,
E-Mail: annette.guignard@library.ethz.ch

■ **Simon Willemin,** Data Scientist, ETH Library, ETH Zurich,
E-Mail: simon.willemin@library.ethz.ch

Emanuel Kulczycki

Emanuel Kulczycki

Reframing scientometrics

How ontological understanding of science influences what we count and how we interpret it

This paper critically examines the ontological assumptions underlying scientometrics, specifically contrasting the atomistic approach of Derek J. de Solla Price with the collective perspective of Vasily Nalimov. It argues that prevailing metrics, focused on individual outputs, fail to account for the importance of scientific communities in shaping knowledge. By applying political ontology, the paper suggests a paradigm shift in scientometrics, advocating for a collective-based perspective that can more effectively capture the complexity of scientific progress. Reframing scientometrics to recognize collectives as fundamental actors could better serve both descriptive and evaluative purposes, ultimately advancing the field.

Scientometrics is increasingly perceived as a practical or even evaluative activity, oriented towards counting publications and citations, often without engaging in theoretical reflection on the concepts it employs. These concepts, however, are crucial in shaping what we, as an academic community, consider real in science, and consequently, what we deem worth counting and assessing. Without such reflection, scientometrics – and science studies in general – risk being reduced to trivial descriptions, sets of tables and figures that merely illustrate what appears on the surface, without recognizing that our perception of research quality is shaped by the very concepts we use. Engaging with these concepts is crucial, particularly when we aim to change how we think about science and how we assess research. A shift in how we discuss science, and what we perceive as the fundamental elements of research communities, is one of the many necessary steps to improve not only scientometrics, but more importantly, science itself, which ultimately serves society. This paper tackles this issue by comparing two grand approaches to scientometrics introduced by Derek J. de Solla Price and Vasily Nalimov, demonstrating that a more collective approach to scientometrics is possible, and could better serve both its descriptive and evaluative purposes.

1. The forgotten alternative

The starting point for the reflection presented in this paper is the assertion that contemporary scientometrics, along with the effects of metrics-based research evaluation regimes – such as publication pressure and the misuse of metrics – largely stems from an atomistic and individualistic ontology pioneered by Price, particularly in his seminal work *Little Science, Big Science* (1963). While Price's approach remains largely unchallenged in Western scientometrics, this paper explores an alternative perspective that emerged in Eastern Europe, specifically Vasily Nalimov's marginalized ontology for scientometrics, presented in his influential *Naukometriya* coauthored with his doctoral student Zinaida Mulchenko (Nalimov/Mulchenko 1969).

This article presents key insights from my engagement in discussions within the fields of science of science and scientometrics. In 2023, I had the honor of delivering a keynote lecture at one of the quarterly events organized as part of the "Swiss Year of Scientometrics" at the ETH Library in Zurich, alongside esteemed scholars such as Stefanie Haustein (Co-director of ScholCommLab in Canada), Elizabeth Gadd (Vice-Chair of the CoARA Steering Board), and Ludo Waltman (Scientific director at the CWTS at Leiden University). The topic assigned to me was "reflexive bibliometrics," and I chose to address the tension between Price's individualistic vision of science and Nalimov's collective approach in my lecture, which was titled precisely the same as this article. While preparing the lecture, I realized that the perspective of political ontology, which I have developed largely thanks to the works of Krystian Szadkowski (Szadkowski 2023; Szadkowski/Krzeski 2019), is crucial for understanding a key point: the fundamental assumptions about what we recognize as "real" in the scientific realm shape our attitudes, actions, and decisions in scientometrics. These assumptions dictate how we measure science and what we consider its key outputs. In this paper, I seize the opportunity to compare the foundational works of Price and Nalimov, examining their perspectives through the lens of political economy.

2. Image of science in scientometrics

Most of the existing literature on scientometrics – whether descriptive or evaluative – focuses on creating new indicators to quantify aspects of scientific activity, such as gender disparities or interdisciplinary research. However, theoretical (meta)studies that critically examine the underlying assumptions guiding how we quantify science are relatively scarce. However, the way we think about and envision science profoundly influences what science becomes and the direction it takes. Maria and Stanisław Ossowscy, the creators of the first program for the science of science, articulated in their foundational 1935 paper, *The Science of Science* (*Nauka o nauce*), that "science, like all other fields of culture, is a part of that particular sphere of reality whose history depends on what we think of it" (Ossowska/Ossowski 1964, p. 82). Our conceptualization of science is not a natural given and does not derive from the objectivity of science itself, as it does not exist objectively in the sense of naive realism. As Ludwik Fleck (2008) would argue, our science is continuously constructed based on a tripartite relationship consisting of (1) individual scientists who conduct research, (2) research subjects investigated by these scientists, and (3) scientific communities (thought collectives in Fleck's terms) to which these scientists belong and which impose their thought styles. A thought style is the ability to perceive problems and articulate solutions based on the values and practices that define the "reference system" in which that though is created (cf. Condé/Jarnicki 2023).

Fleck demonstrated that thought styles within collectives evolve and that new facts do not alter old theories and thought styles. Only a new theory – a new way of perceiving research subjects – can effect such change. I aim to illustrate that the currently dominant thought style in scientometrics, which views science primarily as the product of individual scientists publishing papers, is not a dogma but a historically developed way of thinking about science. This prevailing thought style in contemporary scientometrics (both descriptive and evaluative) focuses on the achievements and successes of individual researchers, which poses significant challenges for studying and describing research groups or entire disciplines – essentially collectives in the Fleckian sense. While scientometrics does describe research groups or invisible colleges, it does so by aggregating the work and publications of individual researchers, rather than recognizing that scientists operate within collectives.

In other words, when contemporary scientometrics describes the productivity of an institution, analyzes ongoing discussions about a research topic, or identifies emerging fields in science, it often overlooks the true role of scientific communities. These communities are not merely aggregations of quantitative data about individual scientists. They constitute one of the three foundational elements of constructing knowledge, alongside scientists and research subjects. Genuine research takes place within the interactions of these three elements. More importantly, if we align with Fleck's perspective, we cannot consider individual scientists outside the prism of the thought collectives to which they belong. Therefore, when analyzing the scientific activity of individual researchers, we should first consider them as members of thought collectives. Reversing this relationship is of fundamental importance and seems inconceivable in current scientometrics. However, the history of science teaches us that what is inconceivable at a given moment is simply contrary to the prevailing thought style. We also know that thought styles, or paradigms in Kuhnian terms, change under the influence of new theories.

The situation with scientometrics is particularly intriguing, as an alternative vision of scientometrics and science, understood as the continuous work of collectives, emerged over half a century ago, conceived by none other than the originator of the term 'scientometrics,' Vasily Nalimov. However, the collectivist vision of Russian scientometrics was overshadowed by the individualistic vision of American scientometrics. It perhaps did not so much lose as it was never given a chance to present its program, as Nalimov's ideas, unjustly perceived as 'red scientist' notions, were largely ignored (Kulczycki 2023).

The current relationship between individuals and collectives in scientometrics has several causes. One significant factor is the highly individualistic and pragmatic concepts introduced by Price, often regarded as the founding father of scientometrics. Another factor is the redefinition of citations through the creation of the Science Citation Index by Eugene Garfield. Additionally, the sociology of science – particularly its transformation into the sociology of scientific knowledge – has largely moved away from focusing on scientific communities. As Jochen Gläser (2001) noted, this has led to a crucial disconnect between scientometrics and the sociology of science.

When I refer to 'current scientometrics,' I mean both the scientific discussions in so-called descriptive scientometrics, as seen in journals like *Scientometrics, Journal of Informetrics, Quantitative Science Studies,* and *Research Evaluation* as well as the entire commercial apparatus of university rankings, citation databases, and scientometric tools used in so-called evaluative scientometrics (Moed 2017). While these two areas intersect in many ways, the heuristic distinction between them is well established within the field itself (Bornmann/Leydesdorff 2014; Garfield 2009; Mingers/Leydesdorff 2015).

As researchers, we are always members of multiple thought collectives. We do not work in isolation. Even in the humanities, we engage in ongoing debate and discussion with colleagues and the works of classical authors. Yet, when using scientometric tools to analyze scientific activity, we invariably reduce everything to individual researchers. We collaborate, but then we count citations and publications separately for each scientist. When analyzing a given scientist, we often overlook their role within their collective or research group. For instance, we evaluate a dean by their publications, even though this is not their primary role or social function as a scholar.

Thus, we need a new approach to scientometrics that – in the current context – will flip this situation on its head. We should begin with the collective and then possibly move down to the scientists or up to the institution or country. In scientometrics, we start with the scientist and their publications because "that's the only way" with the tools we have. But this approach does not reflect how science functions. It reflects the design of the dominant version of scientometrics. It is high time to restore collectives and scientific communities as the foundation of the thought style in scientometrics.

The question arises: do we need to reinvent everything? Fortunately, no. In our journey toward a *collectivist scientometrics,* we can incorporate the effective aspects of the dominant perspective found in the works of Price and Garfield, while drawing inspiration from the somewhat forgotten collectivist vision of Nalimov. Therefore, this paper aims to present Nalimov's approach by contrasting it with Price's perspective through the lens of political ontology.

3. Perspective of political ontology

Fleck believed that knowledge is produced by scientists within thought collectives. Therefore, scientific facts exist and are considered true within these historical collectives and within the intersubjective social space when they become commonly accepted as commonsense knowledge. This perspective helps us understand what "real" scientific knowledge means. However, it raises further questions: How real are the scientists, the collectives, and the connections between them? What hierarchies exist among these elements? Prioritizing one of these elements can fundamentally change our perception of science. Addressing these questions requires moving beyond Fleck's framework, which primarily focuses on how scientists think and the knowledge they produce, emphasizing epistemology and methodology. But what about ontology? Ontology defines not only the elements of the tri-element relationship (research subjects, scientists, collectives), but also the relationships between them. It delimits the very nature of knowing, meaning that no analysis of scientific activity or reflection on science itself can be ontologically neutral. Ontological choices significantly shape the content of our theories.

If we choose to perceive science and the production of scientific knowledge through the lens of individual scientists, collectives, or institutions, our theories will be shaped accordingly, giving primacy to one of these elements. Consequently, as Hay emphasizes, any ontological choice will affect "our expectations about how the political drama will unfold" (Hay 2006, p. 79). Therefore, these choices have various epistemological, methodological, practical, and political consequences.

When comparing the visions of science within the science of science and scientometrics, focusing on how they perceive the roles of scientific communities and individual scientists, I enter the field of political ontology. Political ontology concerns the fundamental assumptions and concepts that shape our perception of scientific reality as a social reality, and in this way, as political. It addresses how we think about the foundations and 'essence' of science. What we recognize as 'real' in the scientific realm influences our attitudes, actions, and decisions in this field. Szadkowski highlights that political ontology is "a method of unveiling ontological assumptions underpinning historical manifestations of being and disclosing their political implications"(Szadkowski 2023, p. 62). Moreover, from the perspective of this paper, political ontology determines how we measure science, what we consider its key products, and the policies applied to it.

Political ontology assumes that our perception of what "exists" and the relationships between entities result from political conflicts. Szadkowski and Krzeski (2019), while emphasizing the usefulness of political ontology in studies of universities, note that political ontology allows us to analyze how different ontological assumptions limit the impact of what is considered "real" on reality. By combining the perspective of political ontology with Fleck's approach, we can explore how what we consider ontologically fundamental – whether individual scientists, thought collectives, or the relationships between them – shapes different views on what is real in the scientific sector, and consequently, leads to different actions and policies.

The perspective of political ontology helps us understand that the concept of the individual (specifically, the individual scientist) has been naturalized and given ontological primacy due to the success of classical political economy and the convergence of liberalism and capitalism. This naturalization of the individual as the foundational element is so strong that any attempt to revive collectives and establish their primacy – especially after the experiences of communist regimes in Europe – evokes images from the past. Consequently, in various fields, particularly in education and science, we encounter narratives that emphasize the individual who can gain personal benefits from education or science, benefits that can be exchanged in the market: education for employment, scientific achievement for rewards, and citations for academic promotion.

The myth of individual effort and flashes of genius persists, as evidenced by the Nobel Prizes awarded to individual scientists, even though many discoveries are the result of the collective work of hundreds of scholars. Elevating the individual in this ontological relationship contributes to the widespread acceptance of the privatization of knowledge through intellectual property and patents. Thus, the individual has taken center stage in the contemporary view of science, which is seen as a collection of individual – though collaborating – units. Modern approaches to science, including those in scientometrics and the science of science, do not consider collectives as the fundamental basis of this perspective. Instead, collectives are viewed merely as groups of scientists who make choices based on their individual will and thinking, rather than following a shared thought style, as Fleck suggested.

Szadkowski and Krzeski (2019) suggest that in the context of universities and higher education, we can identify a third approach: the political ontology of relationships connecting individuals. This approach emphasizes what is common, rather than what is private (as with individual primacy) or public (as with the primacy of collectives). The private/public dichotomy seems to play a larger role in higher education research than in scientometrics. Overcoming this dichotomy through the concept of the common could be crucial for shaping the future of higher education and universities. However, in thinking about science, we remain behind, as the individual/collective dichotomy (and the equality of these elements) is not fundamentally recognized, with the individual still being the primary unit shaping the vision of science. Therefore, in scientometrics, before we can move beyond this dichotomy and highlight the importance of relationships between individuals and collectives, we must first reestablish the significant role of scientific communities in our understanding of science.

4. Contrasting *Little Science, Big Science* and *Naukometriya*

Both Price and Nalimov wrote numerous works that could be considered important contributions to scientometrics. However, the foundational assumptions of their perspectives can be most effectively reconstructed by examining two key books: Price's *Little Science, Big Science* (1963) and Nalimov's *Naukometriya* (1969), which he co-authored with Mulchenko. Nevertheless, it is important to acknowledge that there are historical, methodological, and logical challenges in contrasting these two books, given that they were written in different times and contexts.

Little Science, Big Science is a short book based on four lectures delivered by Price at Harvard, at a time when the post-World War II United States was emerging as the stable heart of the global academic system. Published in 1961, the book quickly became a classic reference in scientometrics, cementing Price's reputation as the "father of scientometrics" and serving as the source of the ideas for the development of the Science Citation Index (SCI) by Eugene Garfield. Even if we agree with Andras Schubert (2019) that Friedrich Engels could be considered the great-grandfather of scientometrics, this does not diminish Price's contribution to stabilizing the law of "exponential growth" in science as a recognized social fact. However, this law was not a groundbreaking scientific discovery. It was rather a simple observation based on raw data. For this reason, it could be claimed by many other "parents." More interestingly, for Price, the law of exponential growth was merely a starting point for investigating the consequences of the shift toward logistic growth, or growth with a logistic curve shape. Yet, there are even more significant contributions in his small book. One such idea is the concept of the "invisible college" – a network of scientists working at the frontier of knowledge who must limit the scope of information to maintain active engagement. This insight provided a crude rationale for the development of the SCI by Garfield.

The SCI, launched in 1964, quickly became an indispensable tool not only in information sciences and libraries but also as the backbone of any scientometrics worth its name. Garfield, a close friend of Price, was undoubtedly inspired by him. However, Price was unable to use the index in his work, relying instead on crude data like the number of publications, entries from *American Men of Science*, and journal circulation statistics from public libraries. While Price referred to the importance of tracking citations and assessing the utility of scientists through more precise measures, he had to work with the limited data available to him, particularly when addressing geopolitical issues in science.

Nalimov and Mulchenko's *Naukometriya* followed in 1969. They openly acknowledged the significant influence of both Price and Garfield on their own work, to the point where one might feel they underplay their own contributions. Despite working in Soviet Russia, their book contains no mention of Marx, Marxism, or any other ideological references that could invite stereotypical accusations. To say that *Naukometriya* was written under the shadow of *Little Science, Big Science* is no exaggeration, but it would be an overstatement to suggest that this diminished the value of Nalimov's project.

Nalimov began his research on scientometrics in the 1950s while working as an abstractor at VINITI [All-Russian Institute for Scientific and Technical Information] (Hammarfelt/Dahlin 2024), where he encountered an Italian translation of Price's paper. Later, he developed his approach to scientometrics within the broader framework of the „science of science" program (Skalska-Zlat 2001; Wouters 1999), initiated in the 1920s and 1930s by Polish sociologists Maria Ossowska, and Stanisław Ossowski, whose article *Science of Science* [*Nauce o nauce*] (1935) was the very first program for a new metascience discipline. For Nalimov, science was a self-organizing system driven by information flows, and thus, the quantitative analysis of science should serve to improve science itself, not merely serve instrumental goals. *Naukometriya* was quickly translated into several languages in socialist countries, sparking significant discussions. It is notable that the only machine-aided English translation of the book, prepared by the Foreign Technology Division of the U.S. Air Force, used the term "measurement of science" instead of "scientometrics," reflecting a misunderstanding of Nalimov's intention to establish a new discipline.

Both Price and Nalimov saw science as a social process that could be analyzed and measured, but they had different views on the nature of its evolution. For Price, science was more orderly and predictable, while Nalimov emphasized its complexity and unpredictability. However, these are not the only divergences in their perspectives. The differences are most apparent when viewed through the lens of political ontology, across four key dimensions: the nature of science, the nature of scientometrics, the role of scholarly communication, and geopolitics. At this point, I would like to emphasize that the idea of examining these dimensions, summarized in Table 1, emerged during discussions within the Scholarly Communication Research Group at AMU, involving Krystian Szadkowski, Zehra Takın, and myself.

Price viewed science as an atomistic system, where individual scientists – acting as separate entities – drive progress, with "invisible colleges" forming networks of prominent researchers. In contrast, Nalimov's relational approach emphasized science as a self-organizing system shaped by information flows, with "invisible collectives" fostering faster scientific development. While Price focused on predicting science's growth through

Tab. 1: Price and Nalimov's Approaches through the Lenses of Political Ontology

Characteristics of Dimensions	*Little Science, Big Science*	*Naukometriya*
Dimension 1: Nature of Science		
Ontological understanding of science	Atomistic: A whole composed of individual scientists ("molecules") behaving in different ways.	Relational: A self-organizing system steered by information flows.
Group unit in science	Invisible colleges: social networks of eminent scientists.	Invisible collectives: groups promoting fast development of science and avoiding blockages in information flows.
Planning the development of science	Not possible.	Can be stimulated by invisible collectives.
Predicting development of science	Possible by extrapolating measurable trends.	Almost impossible due to the system's similarity to biospheres and unpredictable evolution.
Dimension 2: Nature of Scientometrics		
Task of scientometrics	To develop meaningful averages from individual behaviors.	To investigate information flows and their development.
Disciplines structuring scientometrics	Historical/statistical bibliographic methods, mathematics, sociology of science.	Mathematics, biology, cybernetics.
Primary unit of analysis	Individual scientist or paper.	Information flow, networks, relationships.
Area of science measurement	Eminence, productivity, talent distribution.	Relations, connectivity, idea circulation, networks.
Dimension 3: Role of Scholarly Publications		
Genealogy of the academic paper	Linked to intellectual property and journal articles.	A medium for faster information flows.
Primary function of the academic paper	To claim authorship and prestige.	To facilitate the flow of information.
Dimension 4: Geopolitics		
Ideological controversies	Built on the logic of eugenics (Galton).	None; avoids ideological references.
Modes of science-based modernization	Imitation of the centers ("cargo cult").	Unlocking peripheral blockages in information flow.
Understanding of geopolitics (of science)	Science race as economic competition between nations.	Epistemic competition between nations.
Role of the state	Securing funding.	Shaping science policy through funding choices.

observable trends and individual productivity, Nalimov argued that science's development was more unpredictable, resembling the biosphere's complex evolution. In scientometrics, Price prioritized measuring individual productivity and eminence using bibliographic methods, whereas Nalimov centered on analyzing relationships, information flows, and networks. Additionally, their views on scholarly communication diverged: for Price, academic papers were tools for claiming authorship and prestige, while Nalimov saw them as vehicles for accelerating information flow. These distinctions extend to geopolitics, where Price linked scientific progress to national economic competition, while Nalimov emphasized overcoming peripheral blockades in information flow, positioning science as a global, interconnected enterprise.

The comparison of *Little Science, Big Science* and *Naukometriya* through the lens of political ontology reveals how the fundamental assumptions about the nature of science, its measurement, and the role of scholarly communication profoundly shape our attitudes, actions, and decisions in scientometrics – particularly in terms of what we measure and what we regard as the key elements of the research process. Price's atomistic understanding of science, where individual scientists and their achievements are central to scientific progress, contrasts sharply with Nalimov's relational perspective, which emphasizes the self-organizing, collective nature of scientific knowledge production. While Price's focus on individual eminence and productivity aligns with a Western, liberal capitalist model of competition, Nalimov's emphasis on information flows and collectives reflects a more systemic, networked view of science, rooted in Soviet-era thinking. However, it is important not to oversimplify this comparison as merely a contrast between capitalist and communist visions of science. Both systems can support individualist or collectivist views of science, as seen in examples like Trofim Lysenko's prominence in the Soviet Union or the collective effort behind the Manhattan Project in the United States. Capitalism and communism operate on a broader ontological level, both using bibliometric indicators to describe or regulate scientific reality. De-individualizing or "socializing" scientometrics would not impart a communist character but would instead emphasize the fundamental role of thought collectives and scientific communities as key drivers of scientific progress.

These two perspectives highlight the critical role of ontological assumptions in shaping scientometrics. Choosing to focus on individuals or collectives, or prioritizing certain metrics over others, is not a neutral, technical decision. It carries political, epistemological, and methodological implications. The political economy of science, as seen through the work of Price and Nalimov, demonstrates that what we consider "real" in the scientific realm – whether it is the achievements of individual scientists or the collective advancement of knowledge – directly shapes how we measure and evaluate science. To improve scientometrics and research evaluation, we must critically reflect on the ontological assumptions embedded in scientometric practices and consider how these assumptions can be reimagined to better account for the complexities of scientific knowledge production.

5. Why is it worth to discuss reframing scientometrics?

In the prevailing view of scientometrics, science is fundamentally shaped by individual scientists who form groups, often referred to as 'invisible colleges' or 'research groups,' and work within various scientific or industrial institutions. This leads to a reductionist approach to the metricization of science, where the primary measurable aspect is scholarly communication, specifically scientific publications and their citations. As a result, the progress of science is predominantly monitored through the counting of scientists, publications, and institutions. This limited perspective often prioritizes the so-called hard sciences, such as chemistry and biology, while marginalizing fields like the social sciences and humanities. Such an exclusionary framework not only narrows the definition of scientific activity but also fails to recognize the full breadth of intellectual contributions across diverse disciplines.

While individualized scientometrics has been effective over the past six decades in describing, understanding, and evaluating the development of science, its focus on individual scientists as isolated actors has produced a range of negative consequences – chief among them, the intensification of publication pressure. A collectivist approach to scientometrics, however, does not disregard the importance of individual contributions. Even in disciplines like the humanities, where scholars may work individually, they do so within 'thought collectives' that shape their intellectual environment. Individual scientists are always embedded in wider collectives and research networks. This raises two important questions: Should we transition from an individualized to a collectivist approach in scientometrics? And if so, is it feasible?

My hope is that this paper will spark a broader discussion about transforming scientometrics to place greater value on thought collectives, and to ensure that scientists are described and evaluated within the context of their roles in research groups, institutions, disciplines, or thought collectives. Without rethinking its origins and underlying assumptions, scientometrics risks becoming even more radicalized, metricized, and commercialized. It is essential, therefore, to develop and use tools that genuinely serve both science and society, rather than distorting the practices of scientists and their collectives under the pressures of a purely metric-driven system.

References

Bornmann, L./Leydesdorff, L. (2014): Scientometrics in a changing research landscape. In: EMBO Reports, 15 (12), pp. 1228-1232. https://doi.org/10.15252/embr.201439608

Condé, M. L./Jarnicki, P. (2023): Ludwik Fleck: Thought Style and Thought Collective in the Historiography of Science. In: Condé, M. L./Salomon, M. (eds.): Handbook for the Historiography of Science. Springer International Publishing, pp. 1-17. https://doi.org/10.1007/978-3-030-99498-3_5-1

Fleck, L. (2008): Genesis and development of a scientific fact (R. K. Merton, Ed.; T. J. Trenn & F. Bradley, Trans.; Repr. 11. Aufl). Univ. of Chicago Press.
Garfield, E. (2009): From the science of science to Scientometrics visualizing the history of science with HistCite software. In: Journal of Informetrics, 3 (3), pp. 173-179.
https://doi.org/10.1016/j.joi.2009.03.009
Gläser, J. (2001): Scientific specialties as the (currently missing) link between scientometrics and the sociology of science, pp. 191-210.
Hammarfelt, B./Dahlin, J. (2024): Abstracting It All: The Soviet Institute of Scientific Information (VINITI) and the Promise of Centralisation, 1952–1977. Minerva.
https://doi.org/10.1007/s11024-024-09545-z
Hay, C. (2006): Political ontology. In The Oxford handbook of contextual political analysis. Oxford University Press, pp. 78-96.
Kulczycki, E. (2023): The Evaluation Game: How Publication Metrics Shape Scholarly Communication. Cambridge University Press.
Mingers, J/Leydesdorff, L. (2015): A Review of Theory and Practice in Scientometrics. In: European Journal of Operational Research.
https://doi.org/10.1016/j.ejor.2015.04.002
Moed, H. F. (2017): Applied Evaluative Informetrics. Springer International Publishing.
https://doi.org/10.1007/978-3-319-60522-7
Nalimov, V. V./Mulchenko, Z. M. (1969): Naukometrija: Izučenie razvitija nauki kak informacionnogo processa. Nauka.
Ossowska, M./Ossowski, S. (1935): Nauka o nauce. In: Nauka Polska, 20, pp. 1-12.
Ossowska, M./Ossowski, S. (1964): The science of science. In: Minerva, 3 (1), pp. 72-82.
Price, D. J. (1963): Little Science, Big Science. Columbia University Press.
Schubert, A. (2019): Friedrich Engels, the Great-grandfather of Scientometrics. In: ISSI Newsletter, 15 (4), pp. 64-67.
Skalska-Zlat, M. (2001): Nalimov and the Polish way towards science of science. In: Scientometrics, 52 (2), pp. 211–223.
https://doi.org/10.1023/A:1017911722525
Szadkowski, K. (2023): Capital in Higher Education: A Critique of the Political Economy of the Sector. Springer International Publishing.
https://doi.org/10.1007/978-3-031-38441-7
Szadkowski, K./Krzeski, J. (2019): Ontologia polityczna w badaniach szkolnictwa wyższego – indywidualna, publiczna, wspólna. In: Nauka i Szkolnictwo Wyższe, 1-2 (53-54).
https://doi.org/10.14746/nisw.2019.1-2.9
Wouters, P. (1999): The Citation Culture.
http://hdl.handle.net/11245/1.163066

■ **Emanuel Kulczycki,** PhD, Professor of Scholarly Communication, Head of Scholarly Communication Research Group, Faculty of Philosophy, Adam Mickiewicz University in Poznań,
E-Mail: emek@amu.edu.pl

Simon Willemin & Ludo Waltman

Simon Willemin

Ludo Waltman

Towards transparent, democratic, and open research information

The evolving role of CWTS in the Netherlands

Since its creation, the Centre for Science and Technology Studies (CWTS) at Leiden University in the Netherlands contributes to framing how Research Assessment is performed in the country. However, its role and the services it provides have evolved, in line with cultural changes, trends and opportunities. The present article briefly retraces the history of CWTS, showing that, from an initial focus on scientometrics, it has extended its work to policy-making and broader forms of research evaluation. Recent evaluation protocols, recommendation papers and declarations of national or international relevance allow us to understand the specificities of the Netherlands regarding the Research Assessment reform as well as the Open Science movement, and the evolving role played by CWTS in this context.

1. Introduction

This article draws from the content of the third Swiss Year of Scientometrics (SYoS) lecture "Openness of research information – Democratizing the use of scientometrics" held by Ludo Waltman at ETH Zurich on 24 April 2024. The starting point for this manuscript was a transcript of the lecture, which Simon Willemin enriched with additional background information, and which was then revised in an iterative process by Ludo Waltman and Simon Willemin.

Since its creation, the Centre for Science and Technology Studies (CWTS) at Leiden University in the Netherlands contributes to framing how Research Assessment is performed in the country. However, its role and the services it provides have evolved, in line with cultural changes, trends and opportunities. This article briefly retraces the history of CWTS and introduces recently published evaluation protocols, recommendation papers and declarations. Section 2 focuses primarily on the place of scientometrics to evaluate institutions or academics in the context of the Research Assessment reform. Section 3 concentrates primarily on the Open Science movement and on the promotion of transparent, democratic, and open data for scientometric analyses.

2. Cultural changes in the Netherlands

2.1 From quantitative studies to focal areas of interest

The Centre for Science and Technology Studies (CWTS) at Leiden University was created in 1989 around Anthony van Raan, Henk Moed and their research groups (Petersohn/Heinze 2017). It is possible to get an initial impression of the discussions of the time by consulting the *Handbook of Quantitative Studies of Science and Technology*, published by van Raan one year before the creation of the centre. The book contains theory and applications based on "mathematical, statistical, and data-analytical methods and techniques" (1988, p. 1). It appeared in the context of an increasing "demand from research policy and research management" for "new ways to monitor" (1988, p. 1). This demand is linked to a shift in the activites of bibliometric communities that was characterized as follows a few years later:

> Due to the rising costs of bibliographic data and the monopoly of vendors, many projects cannot be funded any more by usual grants. Today commissioned work for science policy and business is one of the preferred forms of bibliometric macro-level studies. Scientometric research has become indirectly dominated by the interest group "science policy and business". Its interest is clearly focused on "prompt" and "comprehensible" indicators, while the state of knowledge would allow the application of more sophisticated methods. Moreover, such research-reports tend to be only partially published and without the necessary methodological enhancements, which reduces its value for the bibliometric community. As a consequence a clear shift away from methodological research towards applied bibliometrics and "technology" can be observed. (Glänzel/Schoepflin 1994, p. 380)

The anchorage of CWTS in a university might have contributed to preventing such a shift away from methodological research. However, the development of the centre reflects the growing demand for applied bibliometrics. In 2002, for instance, in order to become more flexible regarding wages and staff, a small company called CWTS B.V. and working closely together with CWTS was created (Petersohn/Heinze 2017, p. 571). CWTS also expanded its activities and reached national and later international recognition as an expertise centre (Petersohn/Heinze 2017, p. 566). Currently, the centre is still active in scientometrics, and it continues to provide services in this area, but it has broadened its activities and has placed greater emphasis on policy-making in science and evaluation of research.

During the period 2017-2022, CWTS followed the research programme *Valuing Science and Scholarship* (CWTS, n.d.). On 1 January 2023, CWTS launched a strategic plan for the period 2023-2028, called the *CWTS knowledge agenda* (De Rijcke et al. 9 May 2023a). The name change from a *research program* to a *knowledge agenda* recognizes an embrace of a wider range of activities, not limited to fundamental research, but including interventions in policy and education or consultancy and contract work. It reminds us of the shift in the 1990s from methodological research to more applied practices. However, as in the 1990s, the centre will continue to have research activities. The knowledge agenda is based on three pillars, represented in Figure 1: *understanding*, *intervening* and *practicing*. These pillars allow us to understand the relationship between the more "basic" and the more "applied" parts of the activities of the centre. The first pillar covers basic research and scientific approaches. The second pillar has to do with the translation of findings into concrete actions to help improve the system. Among other things, this includes service work carried out via the company CWTS B.V. The third pillar is about CWTS using its knowledge of the research system to improve its own way of working.

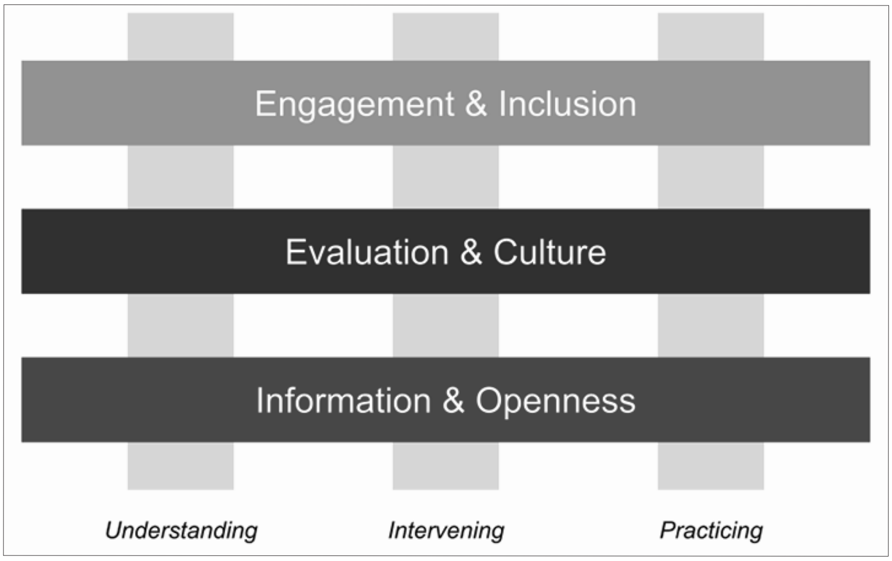
The three pillars are at the core of three focal areas also represented in Figure 1: *Engagement & Inclusion*, *Evaluation & Culture* and *Information & Openness*. Building on the rich history of CWTS in the field of scientometrics, all three focal areas rely partly on scientometric approaches. These approaches are especially important in the focal area *Information & Openness*, which deals with openness of the data needed to perform scientometric analyses. The current knowledge agenda is a sign of a shift where scientometrics is only a part of the activites of the centre.

2.2 Evolving forms of assessment in the Netherlands

This change from a centre focused on quantitative analysis to a centre with a broader spectrum of activities can better be understood if one considers the role played by CWTS in the Netherlands, especially regarding the Research Assessment reform. The cover of three recent documents produced in the Netherlands reflecting cultural changes around Research Assessment are represented in Figure 2 (page 111). They all handle the question of the use of indicators, be it to assess organizational units (Strategy Evaluation Protocol), to evaluate individual academics (Room for everyone's talent) or to rank universities (Ranking the university).

Research Assessment for institutional units in the Netherlands strongly relies on the Strategy Evaluation Protocol (SEP), referred to as the Standard Evaluation Protocol prior to 2021. The document describes how research institutes must be assessed, in a context where they need to be evaluated once every six years according to a protocol that also changes every six years. The evaluation is partly based on self-assessment and partly on external evaluation. In the middle of the 1990s, these protocols greatly contributed to the expansion of CWTS, since Dutch universities under evaluation commissioned the centre for scientometric analyses (Petersohn/Heinze 2017, pp. 570-571). The activities of CWTS mostly had to do with scientometrics. In the 2000s, new competitors emerged, and the database providers Web of Science and Scopus introduced ready-made indicators through InCites and SciVal (Petersohn/Heinze 2017, p. 573). Although the increased competition in the market of quantitative research assessment might have been seen as a threat to CWTS, this is not a major reason for the shift towards less scientometric-focused activities. Such a shift rather seems to come from critical reflection on the role of quantitative indicators in the SEP, which resulted in a declining demand for scientometric analyses in the Netherlands:

Fig. 1: Focal areas and pillars of CWTS



Three focal areas: "Engagement & Inclusion", "Evaluation & Culture", "Information & Openness", and three pillars: "Understanding", "Intervening", "Practicing" (De Rijcke et al., 9 May 2023b).

Fig. 2: Evaluation protocol for the period 2021-2027 (2020), position paper (2019), and recommendation paper (2023) from the Netherlands

The Strategy Evaluation Protocol (SEP) for the period 2021-2027 (VSNU, KNAW et al. 2020), the position paper from the Recognition & Rewards (R&R) program "Room for everyone's talent" (VSNU, NFU et al. 2019), and the recommendation paper "Ranking the university" (Universities of the Netherlands, 2023) are symptoms of cultural changes regarding Research Assessment in the Netherlands.

> The use of bibliometrics in quantitative research assessment received considerable criticism. This criticism resulted in an increased focus on societal relevance of research in the modified SEP (cycle 2015–21) […]. Hence, a noticeable decline in advanced bibliometric analyses commissioned by Dutch universities set in, while at the same time more bibliometric ad hoc analyses were conducted, thereby challenging the formerly dominant position of CWTS as an expert organization (Petersohn/Heinze,2017, p. 573)

Consequently, CWTS broadened its services, diversified its portfolio by including indicators for the societal impact of research and reached out more to international clients (Petersohn/Heinze 2017, pp. 573-574).

The current CWTS focal areas reflect not only the institutional needs as they appear in the SEP for the period 2021-2027, but also another programme named Recognition & Rewards (R&R), which started in 2018 and focuses on the needs regarding the assessment of individual academics. The latter has served as inspiration in other countries, for instance to design the Norwegian framework for assessment (Universities Norway 2021, p. 18). Both the protocol and the programme underline the fact that assessment should not solely rely on quantitative indicators. The protocol and the more recent road map for the programme include quality as an assessment criterion that should be "grounded in a narrative argument and supported by evidence" (VSNU, KNAW et al. 2020), or supported through the use of "evidence-based CVs and assessments portfolios", where achievements are described in "a coherent narrative" (Recognition & Rewards Programme 2023). In this context, quantitative methods and indicators are presented as possible tools to support evidence. More importance is given to the idea that each institutional unit is evaluated according to its own strategy, and that each individual has a wide variety of career paths, where the quantity and impact of research publications are not the most important promotion criteria.

The SEP and the ambitions of the R&R programme are hence not aligned with university rankings, where each institution is ranked according to the same set of criteria. Universities might hence decide to withdraw from some rankings, such as was the case for Utrecht University, which stopped supplying data to a ranking provider (Brent 2023). However, universities benefit from the so-called *ranking game*, since it provides visibility and prestige. It can also contribute to attracting students or researchers, and it can be used to increase funding or collaboration opportunities (Universities of the Netherlands 2023). To find a balance between the need to rethink assessment and the potential benefits of participation in university rankings, an expert group was formed by the Dutch universities and issued a recommendation paper, *Ranking the university*.

The paper focuses on *league tables*, that is, "one-dimensional university rankings that claim to reflect the overall performance of a university" (Universities of the Netherlands 2023, p. 7). Examples of well-known league tables are the Academic Ranking of World Universities (also known as Shanghai Ranking), the Times Higher Education World University Rankings, and the Quacquarelli Symonds World University Rankings. Multidimensional ranking tools such as U-Multirank offer an alternative to these one-dimensional rankings, and the paper encourages universities to contribute to such alternatives. As for league tables, the paper recommends that universities use them for marketing purposes only, are honest about their limitations, and do not to use them for evaluation. The paper also encourages institutions to make the data they supply to ranking providers openly available. The possibility to stop supplying data to ranking providers is mentioned, but the document recommends implementing such a change only as part of coordinated initiatives at the international level, and only for rankings that are not fully transparent and are one-dimensional (Universities of the Netherlands 2023, p. 22).

Despite the contentious nature of university rankings, they can be aligned with values of Open Science. In this context, CWTS constitutes a model by providing a university ranking based on open data, the CWTS Leiden Ranking Open Edition. This ranking presents an alternative to the traditional Leiden Ranking that CWTS has provided since 2007.

3. Open research information as default research information

3.1 The emergence of open bibliometric databases

In the middle of the 1990s, while there was a rising interest in scientometrics, Glänzel and Schoepflin identified the "symptoms of a crisis" (1994, p. 376; see also Gläser/Laudel 2007, p. 116), among which there is the "extension of a certain commercial way of thinking in bibliometrics" (Glänzel/Schoepflin 1994, p. 379). This increased commercialisation was characterized as follows:

> There are […] two classes of bibliometric research groups by now: the ones who can afford to buy expensive data sets, process complex data analyses and plan long-term bibliometric research programs; and the others who cannot do all this. At least in the domain of macro-research the latter have actually only the choice between buying data from the more fortunate colleagues or to restrict to research based on already published data. Here developing countries and Eastern Europe are the most concerned. (Glänzel/Schoepflin 1994, p. 379)

The allocation of financial resources to access the data was not the only difficulty. Already in the 1980s, data accuracy has been a big hurdle, since it could lead to disadvantaging specific groups. Van Raan hence set as a principle not only the systematic engaging with researchers under evaluation, but also cross-checking data accuracy (Petersohn/Heinze 2017, p. 570). The data used at CWTS could further be improved through the creation of an in-house database and through a long-term collaboration with Elsevier (1986-2019), which, according to van Raan, was of huge importance and also contributed to 25-30% of the budget of CWTS, together with other contracts (Petersohn/Heinze 2017, p. 571).

In the 1990s, these contracts and collaboration might have been considered as the best solution in a context where it seemed "unrealistic" to construct a database with sufficient accuracy:

> The publication policies and retrieval possibilities offered by the vendors tend to put limitations on bibliometric research. […] [T]he databases fall short of the expectations of bibliometricians. For the vast majority of the regular users, the databases may be sufficient in the actual form. We have to bear in mind that today, bibliometricians form only a small – although financially quite important – group in the total of all database users. Certain changes in the structure of the data or the standards would only make sense if they could be made effective back to at least 10 or 15 years. Changes to that extent seem to be definitely unrealistic in view of the costs for the customers. (Glänzel/Schoepflin 1994, p. 380)

More than ten years later, Gläser and Laudel asserted that the only available data adapted for citation studies are the indexes provided by Web of Science (2007, pp. 105-106). By then, Elsevier's Scopus had just been launched, and while it was identified as a product with promising potential (Gläser/Laudel 2007, p. 106), it was not yet perceived as a serious competitor to Web of Science:

> This is not only an absolute monopoly, which is very rare in the economy, it also creates the unusual situation whereby a whole scientific community […] depends on data that are not a public good but need to be bought. (Gläser/Laudel 2007, p. 106)

Together with the lack of a "competing public production of data" (Gläser/Laudel 2007, p. 106), the situation was seen as critical. This started to change with initiatives both on the side of Open Science, for instance with the development of an Open Citations Corpus (Shotton 2013) and the Initiative for Open Citations (I4OC) in 2017, and on the side of Research Assessment, notably with the DORA declaration in 2013, recommending that institutions base their assessment on open and transparent data.

While the aforementioned initiatives can be seen as bottom-up contributions, more recently, efforts towards Open Science have also been developed in a more top-down way. In the Netherlands, Open Science is encouraged through funding by the Ministry of Education, Culture and Science, which wanted to make "20 million euros per year available for Open Science until 2032" (NPOS 2022), before a reduction of the budget to 10 million euros (Waltman 2024, September 25). At the international level, the UNESCO Recommendation on Open Science (UNESCO 2021) has been adopted by many countries including the Netherlands.

These documents appear in a context where open databases are increasingly available. Several open citation databases, such as OpenCitations, Microsoft Academic Graph, OpenAIRE Graph, and OpenAlex, have emerged during the last decade. What seemed to be unrealistic in the twentieh century has started to become possible in a foreseeable future.

3.2 Transparency
CWTS is currently promoting a form of scientometrics that is fundamentally transparent, democratic and open. However, in the current context, it is not always possible to be as transparent as desired. For years, the CWTS Leiden Ranking has attracted positive reactions from people who were interested in getting further insights regarding some of its results and who asked CWTS to share the relevant data underlying the ranking. Because of restrictions imposed by Web of Science on the use of the underlying data, CWTS was not allowed to provide the requested information. CWTS and other organizations encounter similar issues when they use proprietary data in their anaylses. Restrictions on data-sharing not only prevent people who would be interested in getting more detailed analyses to explore the data further. It can also be critical when it comes to research evaluation, especially for institutions that have signed the DORA declaration (DORA 2013), in which it is recommended to be open and transparent regarding the data.

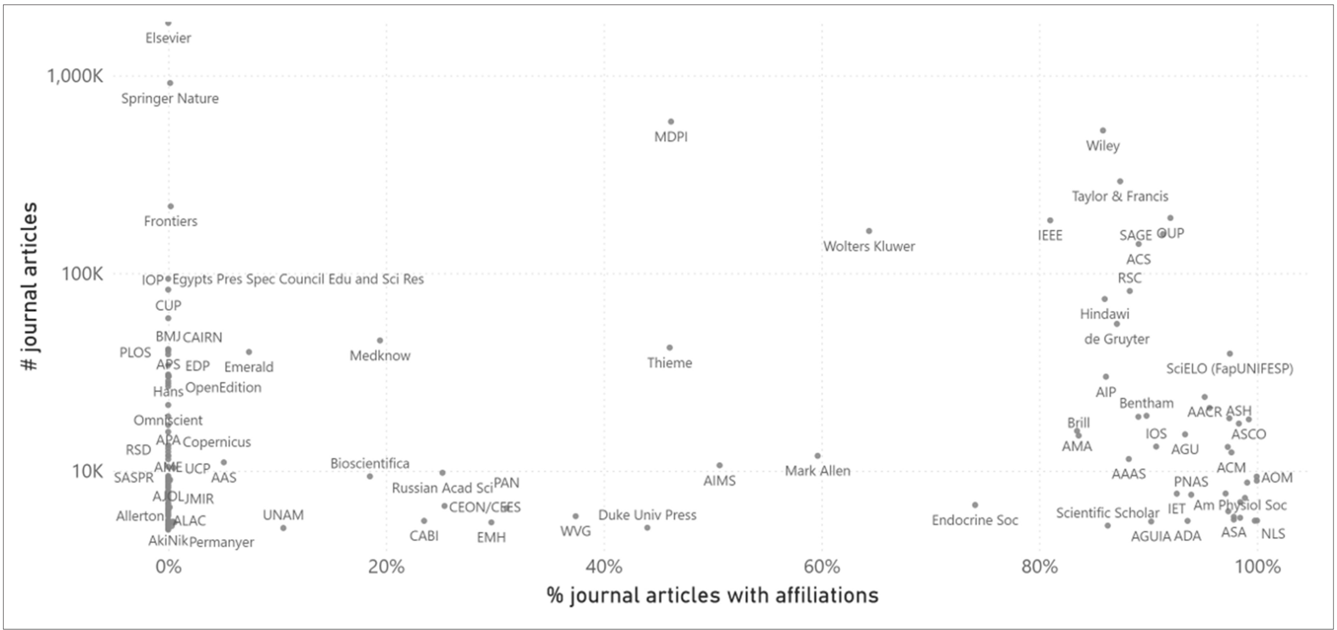
In the context of Research Assessment, the use of open data and the precise description of the methodology is a way to make clear to the evaluated on what information the assessment is based. It is also a way to enable discussion and criticism, both from the side of the assessor and the assessee. Criticism is not limited to the quality of the underlying data. It could also concern the promises of an indicator or the methodological background used to justify certain adjustments in a specific database. When research evaluation relies on ready-made indicators based on unavailable data, it is challenging to check or verify that an indicator provides what is expected. If the people involved in the evaluation process do not have access to the data, they may not have the possibility to provide evidence to challenge or to validate the use of certain indicators. In that sense, working towards making open data available is not only a movement towards transparency, but also a movement towards democratization.

3.3 Democratization
The wider access to data and tools to perform complex quantitative analyses with bibliometric data – the so-called "desktop scientometrics" (Katz/Hicks 1997) – has been seen as an opportunity for its potential to compensate for the failing of peer review, or to increase transparency, accountability and opportunities for experimentation (Derrick/Pavone 2013; Rowlands 2018). Despite these positive views, the idea of *democratization* has not always been positively connoted in the context of scientometrics. In the 1980s, Moed and van Raan concluded from a project (the Leiden Science Indicator Project) that indicators "are not to be used by non-peers since background information is necessary to interpret the quantitative findings" (van Raan 1988, p. 3). Almost two decades later, Gläser and Laudel considered the production of "bibliometric analyses of an evaluative character by actors with little or no professional background in the field" (2007, p. 116). They regretted that the production of bibliometric analyses by people without a proper scientometric training "is more widespread than even the most pessimistic outsider would assume" (Gläser/Laudel 2007, p. 117). In this context, the popularity of bibliometrics is seen as an issue with undesirable consequences. Yves Gingras states, for instance, that the "democratization of bibliometrics is largely responsible for the dubious quality of many papers that are not peer-reviewed by experts in the field" of bibliometrics (2016, p. 39). Gingras also questions an assumption of Jorge E. Hirsch, according to which the *h*-index leads to a more democratic assessment (Rovner 19 May 2008): "On the contrary," states Gingras, "everything suggests that by ignoring the conditions of validity of an indicator, this supposed 'democracy' will quickly turn into evaluative populism" (Gingras 2016, pp. 43-44). Such a criticism depends on the questionable use of indicators like the *h*-index and the Journal Impact Factor as a "primary parameter" in research evaluation (DORA 2013). In that sense, the democraticization of bibliometrics could be considered to be at the roots of the Research Assessment reform.

While in the short term, the democraticization of scientometrics may indeed lead to inappropriate uses of quantitative indicators, in the longer term, it may be the only way to make sure that a broader community gets properly acquainted with scientometrics and becomes more experienced in its use. From this perspective, the main challenge is not to establish a small community of people widely recognized as scientometric experts, but to increase scientometric literacy more broadly, allowing anyone to benefit from the insights that scientometric indicators can bring if they are used in a meaningful way. In a context where the two most prominent data providers – Web of Science and Scopus – make their data accessible to subscribers only, ensuring that everyone has unrestricted and free access to scientometric data is increasingly seen as a crucial condition to make this form of democraticization a reality. Finally, democratization can entail other dimensions, such as ensuring that everyone is equitably represented in the openly accessible data (Babini et al. 2024, April 22).

3.4 Openness
The CWTS approach to open data is clearly visible in the orientation given by the centre to the CWTS Leiden Ranking, a ranking exclusively based on bibliometric parameters. The ranking does not aim to describe the overall performance of universities. It is limited to their performance in terms of publication counts and other purely bibliometric parameters, like citation impact, open access, gender and collaboration. Since its creation in 2007, the ranking was based on proprietary data provided by Web of Science. In January 2024, CWTS launched an open edition of the ranking based on open data from

Fig. 3: Percentage of journal articles with openly available affiliation metadata in Crossref



Wiley and Taylor & Francis, on the top right-hand side, have a large number of journal articles and affiliation metadata is openly available in Crossref for more than 80% of these articles. The big publishers Elsevier and Springer Nature, on the top left-hand side, also have a large number of journal articles, but do not supply affiliation metadata openly.

OpenAlex (Waltman et al. 2024, January 30). In the next few years, the latter ranking is expected to fully replace the ranking based on closed data.

The main difference between both rankings is the data source. The Leiden Ranking Open Edition is based on a subset of OpenAlex publications named core publications.[1] Due to this selection of core publications, the ranking appears to be less inclusive than one might expect. However, this limited sample enables CWTS to make the ranking resemble the well-established Leiden Ranking based on closed data.

Comparing the two rankings yields promising results, since the rankings show very small discrepancies for most of the universities. However, there are large differences for some universities, but these are exceptions (Van Eck et al. 30 January 2024). The main reasons for the discrepancies lie in differences regarding the data sources, the inclusion criteria and the data quality. Whereas errors are present in both data sources, overall the open data is still of lower quality, and this is an important parameter that explains the differences. However, the quality of the data in OpenAlex keeps improving, and the differences between the next open and closed versions of the ranking are therefore expected to be smaller. In any case, current results suggest that the ground is ready for a transition from proprietary data sources to open data.

Collaboration with providers of open data sources to improve their database is not the only way to contribute to facilitating the transition. One of the main sources used by open bibliometric data providers such as OpenAlex is Crossref. Figure 3 shows the percentage of journal articles for which affiliation metadata – one of the key metadata for bibliometric analyses at institutional level – are made openly available in Crossref. On the one hand, as of 2024, for some big publishers, such as Wiley/Taylor/Francis, affiliation metadata is openly available for more than 80% of their publications. On the other hand, other big publishers, such as Elsevier and Springer Nature, make no affiliation metadata (0%) available in Crossref (Van Eck/Waltman 2024).

If these big publishers made their metadata openly available, the accuracy of open bibliometric databases would improve considerably. However, they might not be interested in supplying such metadata openly, since the data may be perceived to have a significant commercial value. It can be argued that the negotiations on Open Access between university libraries and publishers should include clauses guaranteeing full access not only to publications, but also to the corresponding metadata. Putting pressure on publishers has shown to be successful. Since 2017, the Initiative for Open Citations (I4OC) has been joined by almost all of the larger publishers and has resulted in a major increase of openly available

1 "The Leiden Ranking Open Edition takes into account only a subset of the publications in OpenAlex. We refer to these publications as core publications" (https://open.leidenranking.com/information/indicators#publications, consulted on 7 October 2024).

citation data (Martín-Martín 27 October 2021; Van Eck/Waltman 2021; Van Eck/Waltman 2024). Allowing for better open scholarly metadata and developing strategies to negotiate contracts with big publishers belong to the goals that could be reached through the discussions and reflexions around the recent Barcelona Declaration on Open Research Information.

The Barcelona Declaration on Open Research Information was released on 16 April 2024, after a meeting in Barcelona between research-information experts, university librarians, research funders, and further stakeholders. Signatories of the Barcelona Declaration make four commitments related to openness of research information. The term *research information* refers to all information about research that is carried out, shared, communicated and published. This is broader than bibliographical data and for instance also includes data on research funding, grants, research organizations and research contributors. Signatories of the Barcelona Declaration agree (1) to foster the use and production of research information data that are open; (2) to work with services and systems that support and enable openness; (3) to support the sustainability of the required infrastructure; and (4) to support collective action to accelerate the transition (Barcelona Declaration 2024). On 16 April 2024, Leiden University, to which CWTS belongs, was among the initial signatories of the Barcelona Declaration, together with more than 40 other research organizations worldwide. After the publication of the Barcelona Declaration in April 2024, many more organizations decided to sign the Declaration. Six months later, in October 2024, there are almost 100 signatories.

The introduction of the Leiden Ranking Open Edition, discussed above, provides a concrete example of the actions that CWTS is taking to contribute to the transition to open research information. These actions and other activities that stem from the knowledge agenda of CWTS demonstrate how the center is actively engaged in the most recent trends regarding both the Research Assessment reform and the Open Science movement.

4. Conclusion

CWTS has evolved from a research centre initially focused almost entirely on scientometric analysis, with international recognition as an expertise centre in this area, to a centre where policy-making and research evaluation are considered strongly from the perspective of the Research Assessment reform and the Open Science movement, and more specifically from the perspective of changes in research assessment in the Netherlands. CWTS reflects on and contributes to changes in research evaluation practices. Members of the centre have also been active contributors to evaluation protocols, recommendation papers and declarations of national and international relevance. Such documents contribute to framing how research evaluation is performed in the country. They emerged from a science system in which research assessment was heavily dependent on quantitative indicators, while the country is now moving towards a system characterized by a more progressive position with regard to the Research Assessment reform and the Open Science movement. The experience gained in the Netherlands from these relatively recent developments can provide valuables insights for countries with similar concerns.

Acknowledgements

This article has been written within the framework of the project "Swiss Year of Scientometrics" (SYoS), which is co-funded by swissuniversities and the ETH Library. The authors acknowledge proofreading by Brian Cooper and helpful comments on the manuscript by David Johann.

Competing interests

Ludo Waltman is scientific director of the Centre for Science and Technology Studies (CWTS) at Leiden University.

References

Babini, D./Becerril-Garcia, A./Costas, R./Matas, L./Rafols, I./Rovelli, L. (2024, April 22): Not only Open, but also Diverse and Inclusive: Towards Decentralised and Federated Research Information Sources, Leiden Madtrics.
https://www.leidenmadtrics.nl/articles/not-only-open-but-also-diverse-and-inclusive-towards-decentralised-and-federated-research-information-sources

Barcelona Declaration on Open Research Information, Kramer, B./Cameron, N./Waltman, L. (2024, April 16): Barcelona Declaration on Open Research Information
https://doi.org/10.5281/zenodo.10958522

Brent, T. (2023, October 12): Utrecht University withdraws from global ranking as debate on quantitative metrics grows. ScienceBusiness.
https://sciencebusiness.net/news/universities/utrecht-university-withdraws-global-ranking-debate-quantitative-metrics-grows

CWTS (n.d.). CWTS Research Programme 2017-2022: Valuing science and scholarship: Meaningful relations between quality, impact and indicators.
https://www.cwts.nl/download/f-x2q2x2.pdf

De Rijcke, S./Waltman, L./Noyons, E. (2023a): Introducing the CWTS knowledge agenda 2023-2028. Leiden Madtrics.
https://www.leidenmadtrics.nl/articles/introducing-the-cwts-knowledge-agenda-2023-2028

De Rijcke, S./Waltman, L./Noyons, E. (2023b): The focal areas of CWTS. Leiden Madtrics.
https://www.leidenmadtrics.nl/articles/the-focal-areas-of-cwts

Derrick, G. E./Pavone, V. (2013): Democratising research evaluation: Achieving greater public engagement with bibliometrics-informed peer review. In: Science and Public Policy, 40 (5), pp. 563-575.
https://doi.org/10.1093/scipol/sct007

DORA (2013): San Francisco Declaration on Research Assessment.
https://sfdora.org/read/

Gingras, Y. (2016): Bibliometrics and Research Evaluation. The MIT Press.
https://doi.org/10.7551/mitpress/10719.001.0001

Glänzel, W./Schoepflin, U. (1994): Little scientometrics, big scientometrics… and beyond? In: Scientometrics, 30 (2-3), pp. 375-384.
https://doi.org/10.1007/BF02018107

Gläser, J./Laudel, G. (2007): The social construction of bibliometric evaluations. In: Whitley, R./Gläser, J. (eds.): The Changing Governance of the Sciences. Springer, pp. 101-123.
https://doi.org/10.1007/978-1-4020-6746-4_5

Katz, J. S./Hicks, D. (1997): Desktop scientometrics. In: Scientometrics, 38 (1), pp. 141-153.
https://doi.org/10.1007/bf02461128

Martín-Martín, A. (2021, October 27): Coverage of open citation data approaches parity with Web of Science and Scopus. OpenCitations blog.
https://opencitations.wordpress.com/2021/10/27/coverage-of-open-citation-data-approaches-parity-with-web-of-science-and-scopus/

NPOS (2022, December 7): Open Science 2030 in the Netherlands. NPOS2030 Ambition document and rolling agenda. Netherlands National Programme Open Science.
https://doi.org/10.5281/zenodo.7433767

Petersohn, S./Heinze, T. (2017): Professionalization of bibliometric research assessment. Insights from the history of the Leiden Centre for Science and Technology Studies (CWTS). In: Science and Public Policy, 45 (4), pp. 565-578.
https://doi.org/10.1093/scipol/scx084

Recognition & Rewards Programme (2023): Room for everyone's talent in practice: road map: how we are shaping a new system of recognition and rewards.

https://recognitionrewards.nl/wp-content/uploads/2023/04/Room-for-everyones-talent-in-practice-Road-map-Recognition-Rewards.pdf

Rovner, S. L. (2008, May 19): The import of impact: New types of journal metrics grow more influential in the scientific community. In: Chemical & Engineering News, 86 (20), pp. 39-42.
http://pubs.acs.org/cen/science/86/8621sci1.html

Rowlands, I. (2018): What are we measuring? Refocusing on some fundamentals in the age of desktop bibliometrics. In: FEMS Microbiology Letters, 365 (8), pp. 1-6.
https://doi.org/10.1093/femsle/fny059

Shotton, D. (2013): Publishing: Open citations. In: Nature, 502, pp. 295-297.
https://doi.org/10.1038/502295a

UNESCO (2021): UNESCO Recommendation on Open Science. United Nations Educational, Scientific and Cultural Organization.
https://doi.org/10.54677/MNMH8546

Universities of the Netherlands (2023): Ranking the University: On the effects of rankings on the academic community and how to overcome them.
https://doi.org/10.17026/SS/XGDX0H

Universities Norway (2021): NOR-CAM – A toolbox for recognition and rewards in academic careers.
https://www.uhr.no/en/_f/p3/i86e9ec84-3b3d-48ce-8167-bbae0f507ce8/nor-cam-a-tool-box-for-assessment-and-rewards.pdf

Van Eck, N. J./Visser, M./Waltman, L. (2024, January 30): Opening up the CWTS Leiden Ranking: Toward a decentralized and open model for data curation. Leiden Madtrics.
https://www.leidenmadtrics.nl/articles/opening-up-the-cwts-leiden-ranking-toward-a-decentralized-and-open-model-for-data-curation

Van Eck, N. J./Waltman, L. (2021): Crossref as a source of open bibliographic metadata. In: Proceedings ISSI 2021, pp. 1969-1174.
https://hdl.handle.net/1887/3264100

Van Eck, N. J./Waltman, L. (2024): Crossref as a source of open bibliographic metadata. MetaArXiv Preprints.
https://doi.org/10.31222/osf.io/smxe5

Van Raan, A. F. J. (1988): Introduction to the handbook. In: van Raan, A. F. J. (ed.): Handbook of Quantitative Studies of Science and Technology. Elsevier, pp. 1-8.
https://doi.org/10.1016/c2009-0-08476-1

VSNU, NFU, KNAW, NWO and ZonMw (2019): Room for everyone's talent: towards a new balance in the recognition and rewards of academics.
https://recognitionrewards.nl/wp-content/uploads/2020/12/position-paper-room-for-everyones-talent.pdf

VSNU/KNAW/NWO (2020): Strategy Evaluation Protocol 2021-2027.
https://storage.knaw.nl/2022-06/SEP_2021-2027.pdf

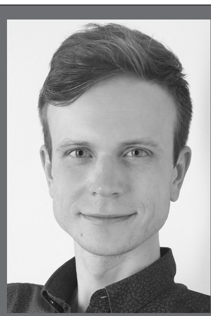
Waltman, L./Van Eck, N. J./Visser, M./Neijssel, M./Montgomery, L./Neylon, C./Kramer, B./Demes, K./Priem, J. (2024, January 30): Introducing the Leiden Ranking Open Edition. Leiden Madtrics.
https://www.leidenmadtrics.nl/articles/introducing-the-leiden-ranking-open-edition

Waltman, L. (2024, September 25): Bezuiniging op Open Science NL, superzonde! Leiden Madtrics.
https://www.leidenmadtrics.nl/articles/bezuiniging-op-open-science-nl-superzonde

■ **Simon Willemin,** Data Scientist, ETH Library, ETH Zurich,
E-Mail: simon.willemin@library.ethz.ch
■ **Ludo Waltman,** Prof. Dr., Centre for Science and Technology Studies (CWTS), Scientific Director, Leiden University,
E-Mail: waltmanlr@cwts.leidenuniv.nl

Julian Dederke, Michelle Koch & Simon Willemin

The Representation of Swiss Higher Education Institutions in Five Bibliometric Databases



Julian Dederke

Michelle Koch

Simon Willemin

This article examines the representation of Swiss higher education institutions in five bibliometric databases: Dimensions, Scopus, Web of Science (commercial), as well as OpenAIRE and OpenAlex (open). It emphasizes the importance of organisational identifiers, entity and language disambiguation, deduplication, definition of publication types, and affiliation-matching for findability and publication counts. The article presents similarities and differences across databases, noting that overall trends are consistent regardless of the database used. However, database choice can significantly affect comparisons between specific institutions. Non-university institutions in Switzerland appear disadvantaged in the commercial databases Scopus and WoS. This article suggests actions to improve representation of institutions, especially in open databases.

Data availability: The data underlying this article are available in the ETH Research Collection (Koch et al. 2024).

1. Introduction

Rankings and bibliometric databases play a major role when benchmarking the performance and impact of research and teaching. Rankings such as the Times Higher Education Ranking (THE), Quacquarelli Symonds (QS) and the CWTS Leiden Ranking use bibliometric data (alongside other data about higher education institutions) to assess the performance of universities in a comparable manner. Some of these rankings also face criticism regarding their methodology (e.g., Fauzi et al. 2020). The bibliometric databases used by the leading ranking providers can also serve institutions, departments and researchers to analyse their scientific output. Besides the large commercial providers Dimensions, Scopus and Web of Science (WoS), in recent years more and more open, non-commercial bibliometric databases have emerged. These databases are less restrictive on the reuse of the data and allow for more transparency and reproducibility in times of the upturn of Open Science (UNESCO 2021), the reform efforts related to research assessment (DORA 2013), and considering the potential for universities to depend less strongly on proprietary databases – in Switzerland as elsewhere. Favouring their use can also be seen as a way to promote Open Science practices. However, favouring databases providing their data under open licences is just one aspect and decisions whether or not to license certain databases might ideally be made only after assessing and comparing the available alternatives.

With a variety of different types of higher education institutions (HEIs), Switzerland is set up with a very diverse landscape of research and teaching institutions – including federal and cantonal universities, universities of applied sciences, as well as universities of teacher education. All of them, with their respective profiles, add to Switzerland's strong track record (Benito et al. 2020) in research, science, teaching and education. To understand the scientific impact of Swiss HEIs, it is important to comprehend their representation in bibliometric databases. In the last two decades, a wide variety of such databases has emerged: CORE, Microsoft Academic Graph, OpenCitations, Semantic Scholar, The Lens and many others. In this publication, we compare the commercial bibliometric databases Dimensions, Scopus and WoS, and the open databases OpenAIRE and OpenAlex. This article addresses the following questions: Which Swiss HEIs are covered in the five compared bibliometric databases, and how do the publication counts of the different institutions vary across these databases?

Some previous research in Switzerland went in a similar direction with Machado et al. (2016) comparing the article numbers of ten Swiss HEIs based on Scopus data for 2001-2015. Hug and Brändle (2017) compare the coverage of Microsoft Academic, Scopus and WoS relative to the publications of the University of Zurich listed in the Zurich Open Archive and Repository (ZORA). They find that Scopus lists most of the journal articles found in ZORA, while Microsoft Academic surpasses the other two databases regarding coverage of book sections, conference items, edited volumes, and monographs. Drawing on a data sample from the University of Teacher Education Zurich, Stricker (2023) delivers a unique insight into database coverage for a small Swiss institution by examining whether its publications are adequately represented in, among others, Dimensions, Unpaywall and WoS. He detects a "non-adequate representation of publication output" (p. 388, translated from German), with the proportion of missing publications in these databases ranging between 52% and 84% (ibid.). Our paper goes beyond these approaches by looking at all Swiss HEIs, albeit without a comparison to internal repositories and publication lists (unlike Hug and Brändle 2017, or Sticker 2023). At the time of writing this article we are not aware of any publications that compare the coverage of a country's HEIs across databases in a similar way.

Following the introduction, we present the underlying data (Koch et al. 2024) and methods used for this study. We continue with a section on the findability of HEIs in the different databases and present remarkable differences. Thereafter, we stress the role of organizational identifiers, entity disambiguation and language disambiguation (which is particularly relevant in Switzerland, with its four official languages) for the findability of HEIs. A section on the publication counts compares the number of publications of Swiss HEIs across the investigated databases. The definition of publication types, deduplication processes and affiliation-matching procedures are identified as relevant factors for differences across databases. Finally, we conclude by summarising and highlighting key commonalities and differences across bibliometric databases and provide exemplary recommendations for actions that can improve open databases, such as requests in the Research Organization Registry (ROR) to add institutions or to improve their metadata, or curation efforts in OpenOrgs.

2. Data and Methods

For the analysis, we included all 52 institutions listed by swissuniversities[1] as "[a]ccredited Swiss Higher Education Institutions" (Swissuniversities 2024, status 21 February 2024). A focus was placed on the findability and publication counts of institutions in the databases Dimensions (Hook et al. 2018), Scopus (www.scopus.com), WoS (www.webofscience.com), OpenAIRE (Manghi et al. 2022) and OpenAlex (Priem et al. 2022). Such bibliometric databases can be used by downloading full or partial dumps of data, using application programming interfaces (APIs), or by using their web interfaces. For the commercial providers at least the first two options are only accessible to subscribing institutions, otherwise they have strong limitations. Moreover, using these available data(bases) beyond the easily accessible web interfaces usually requires data science competences that are not available to all interested institutions.

For all considered databases, publications for the years 2012 to 2022 were included that have the publication type (1) "article", (2) "book" or "monograph", (3) "(book) chapter", or (4) "conference paper" or "proceeding paper", respectively. Depending on the database, the available publication types differ (see Koch et al. 2024). For all included databases, the allocation of a publication to an institution can be based on different data, such as the metadata provided by publishers, or the authors' affiliation metadata that are directly available within the published version of an article. We have not performed any cleaning of the affiliation data, but instead took them as they are presented in the respective databases. Full counting was applied, i.e., a publication that has affiliations to several of the Swiss institutions is counted as one publication for each of these institutions.

2.1 Institutional Identifiers and Comparability across Databases

Since each database uses its own institutional identifiers, it is very difficult or even impossible to have an exact matching of institutional hierarchies (i.e., inclusion or exclusion of child or related institutions) across the databases. In essence, "ETH Zurich" or "University of Bern" can mean very different things in the different bibliometric databases. The methods we applied still try to mirror selected institutions as closely as possible and will be described in the following and are documented in more detail in Koch et al. (2024).

For the commercial databases Dimensions, Scopus and WoS, the web interfaces were used to manually select each of the 52 Swiss HEIs. The open databases have been explored using data dumps from March 2024. For the Dimensions data retrieval per institution, child institutions and related institutions were included, and identifiers for each of these institutions were documented. To approximate the role of child and related institutions as closely as possible in the case of Scopus, the count for "Documents, whole institution" was used, i.e., "the sum of those [documents] from the institution itself plus those from the affiliations contained within its hierarchy. Where an institution is not the top level of the hierarchy, only the affiliations below it are included in its total" (https://www.scopus.com, 20.11.2023). Similarly, for WoS, the full institution's identifier was used in the web interface (e.g., "OG = (ETH Zurich)"), which always includes a number of name disambiguations (e.g., Swiss Federal Institute of Technology Zurich) and (child) institutions within the hierarchy of the respective university.

[1] Swissuniversities is the rectors' conference of Swiss universities, i.e., the umbrella organisation of Swiss Higher Education Institutions (https://www.swissuniversities.ch/en/welcome).

While collecting data for the three commercial databases via their web interfaces, we also documented institutional identifiers for each of them: Dimensions' internal grid-identifier (e.g., 'grid.5801.c' for ETH Zurich) and the open Research Organization Registry (ROR) identifier (e.g. '05a28rw58'), Scopus' internal 8-digit identifier (e.g., '60025858' for ETH Zurich) and WoS' 'organization enhanced' identifier (e.g., 'OG=(ETH Zurich)') (see also Purnell 2022: Table 1, p. 100). We also extracted the respective institutions' identifiers in OpenAlex (e.g., 'I35440088' for ETH Zurich) and OpenAIRE (e.g., 'openorgs____::0000097748' for ETH Zurich).

Among the commercial databases, Dimensions with its grid-IDs allowed the most complete and most transparent way of selecting child institutions and related institutions (which have one grid-ID each). Therefore, we used the grid-ID list of each HEI's child institutions and related institutions and identified the ROR-identifiers that matched these grid-IDs. By sticking to the list of ROR-identifiers that correspond to Dimensions' grid-IDs, we made sure that data retrieved from OpenAIRE and OpenAlex would always include the same child institutions and related institutions via their ROR-IDs. Even though this selection does not necessarily correspond to what OpenAIRE or OpenAlex list as child and related institutions, it supports comparability across databases. We checked, for each institution, whether the ROR-ID is findable in OpenAIRE or OpenAlex, respectively. In OpenAlex, we inspected the authors' affiliations to determine the ROR-ID of each author's institution. By contrast, in OpenAIRE, the ROR-ID was used to search within the databases' 'PID' field. Whenever we could not find the ROR-ID with this approach, we conducted a manual search of the institution's name and its variants to ensure thorough coverage.

2.2 Name Variants as a Challenge for Identification

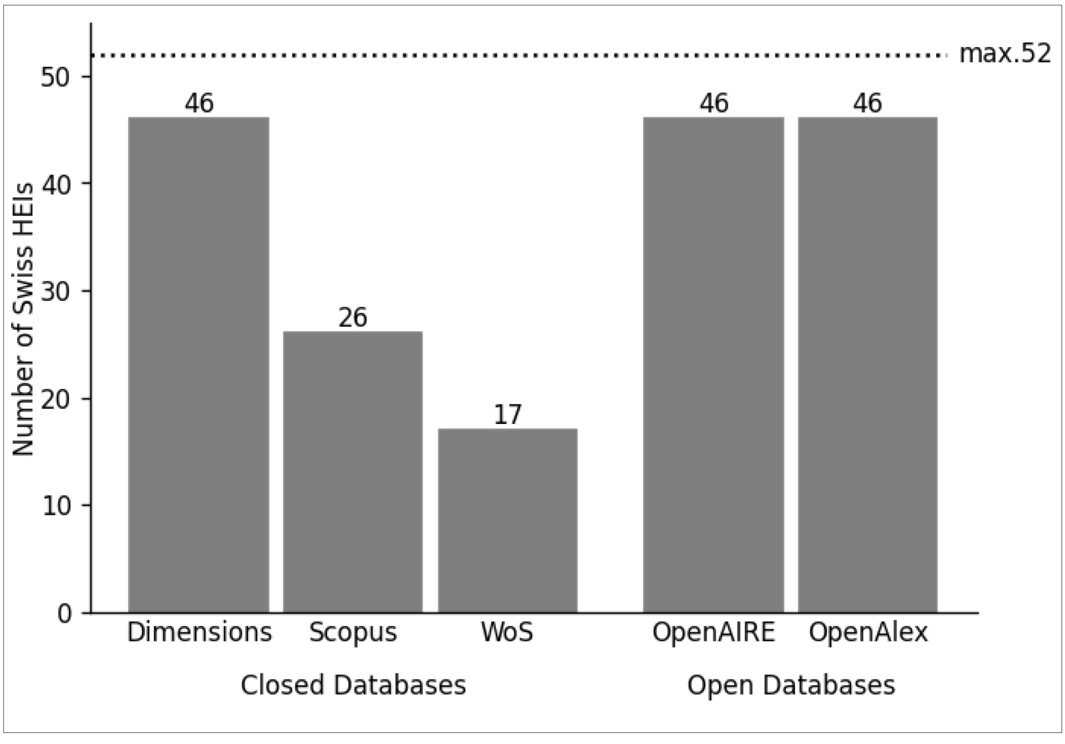
The findability and consistency of bibliometric data for research institutions is significantly hampered by the variability of institutions' names in the different databases. For example, using the organisation search of Dimensions, Scopus or the WoS web app to search for "Sankt Gallen" curiously does not yield the "University of St. Gallen", while "St Gallen" works. Institutional names in databases might also represent child institutions within the hierarchy of a HEI (e.g., Singapore-ETH Centre) or might just refer to one and the same institution in another language (e.g., "Università della Svizzera italiana" in Italian and "University of Italian Switzerland" in English). The latter point is particularly prevalent in Switzerland, with its four official languages (French, German, Italian, Romansh). While we did not check for Swiss language variants for every investigated institution, when searching institutions with their names in English and in one of the national languages, generally speaking, Dimensions' web app showed more complete results regarding findability of institutions than WoS. Similarly, language disambiguation of the WoS web app appeared as weaker than the one of Dimensions when searching institutions with their names in the respective national language. This finding is not surprising considering the scope of the respective database. WoS focuses mainly on English-language publications, while Dimensions is more inclusive, which is why a lack of language disambiguation for organisations might be more problematic for Dimensions.

3. Results

3.1 Findability of Swiss HEIs

Which Swiss HEIs are covered in the five investigated databases? To understand the extent to which the institutions' output is represented in the databases, we first need to know whether the institutions are identifiable in the respective databases at all. Figure 1 presents how many institutions have been identified per database. Overall, OpenAlex, OpenAIRE and Dimensions show the highest counts with 46 identified insti-

Fig. 1: Findability of 52 Swiss Higher Education Institutions in five databases



tutions[2], whereas Scopus and WoS appeared much more restrictive in this regard, with 17 and 26 identified institutions, respectively.[3]

These differences in the findability of institutions across databases are remarkable. Especially universities of applied sciences and universities of teacher education were at a disadvantage and often do not seem to be indexed within Scopus or WoS at all. It is important to note that the fact that a HEI was not found in one of the databases does not always imply that research output from that HEI is not at all represented in the respective database. Instead, it highlights that, with the Scopus and WoS web apps, it is either difficult or impossible to find and select those organisations and to analyse their research outputs. This can be problematic, as these web interfaces are frequent entry points to bibliometric data for research institutions. In particular for the commercial databases, it has been demonstrated that they might have different strengths and weaknesses when it comes to coverage of certain research fields, publication types, journals or certain research institutions either on a global scale or in specific countries (see, e.g., Asubiaro et al. 2024; Donner et al. 2020; Harzing 2019; Singh et al. 2021; Visser et al. 2021).
The pattern for Swiss HEIs is in accordance with previous studies on the coverage of commercial databases (e.g., Visser et al. 2021), among which WoS has been described as the most selective and Dimensions as the most comprehensive (see Singh et al. 2021 for journal coverage), and Dimensions as being abreast of Scopus, while WoS is viewed as less comprehensive (Harzing 2019). Given these extensive insights and the obvious differences in commercial databases, it is even more important to understand how the coverage of newer, open databases, as well as their strengths and weaknesses, relate to this.

3.2 Publication Counts of Swiss HEIs

The number of publications for each Swiss HEI depends on several factors such as institution type and publication type (e.g., Mongeon/Paul-Hus 2016; Singh et al. 2021). Figure 2 shows the counts of articles for institutions of each type[4] for each part of Switzerland: a university from the German-speaking part (ETH Zurich), a university of applied sciences from the Italian-speaking part (SUPSI), an institution from the French-speaking part (IHEID) and a university of teacher education from Eastern Switzerland (PHSG). This is merely an exemplary illustration to show the variety of institution types and article counts that the Swiss research and teaching landscape includes. Figures for each of the investigated Swiss institutions are provided in Koch et al. (2024).

In every database, universities are the institution type with the highest article counts during the period 2012 to 2022. The overall counts range from 1,063 articles (University of Lucerne, Scopus) to 86,586 articles (University of Zurich, OpenAIRE). Since some institutions from other types are missing in WoS and Scopus, the minimal number of articles they are affiliated to is zero. Universities of Applied Sciences reach their maximum at 5,059 articles (HES-SO, OpenAlex), Universities of Teacher Education at 614 articles (Pädagogische Hochschule Zürich, OpenAlex) and other institutions at 1,991 articles (IHEID, OpenAlex).

The article counts for universities across the five considered databases follow patterns observable in each database: The article counts from 2012 are smaller than those of 2021 for each institution in each database. In particular, for each subsequent year (except 2021-2022), ETH Zurich and the University of Bern increase their counts in each database; between 2019 and 2020, there is an increase in all the databases for all the universities, except for the University of Neuchâtel; between 2021 and 2022, there is a decrease in OpenAIRE and WoS for all universities, except for the University of Lucerne. Those counts reflect a global increase in publications (for trends until 2018, see Bornmann et al. 2021). The increase between 2019 and 2020 might be linked to the COVID pandemic (Rousseau 2023). Finally, the main factors for the decrease between 2021 and 2022 are unidentified. They might be related to subsequent effects of the increase observed during the COVID pandemic but also to changes in the data-processing in a context where, for instance, OpenAlex succeeded the discontinued Microsoft Academic Graph.
In most of the databases, publications other than articles are available for many institutions. These other types of publications are not presented in more detail here, but the data and visualisations for them are provided in Koch et al. (2024).

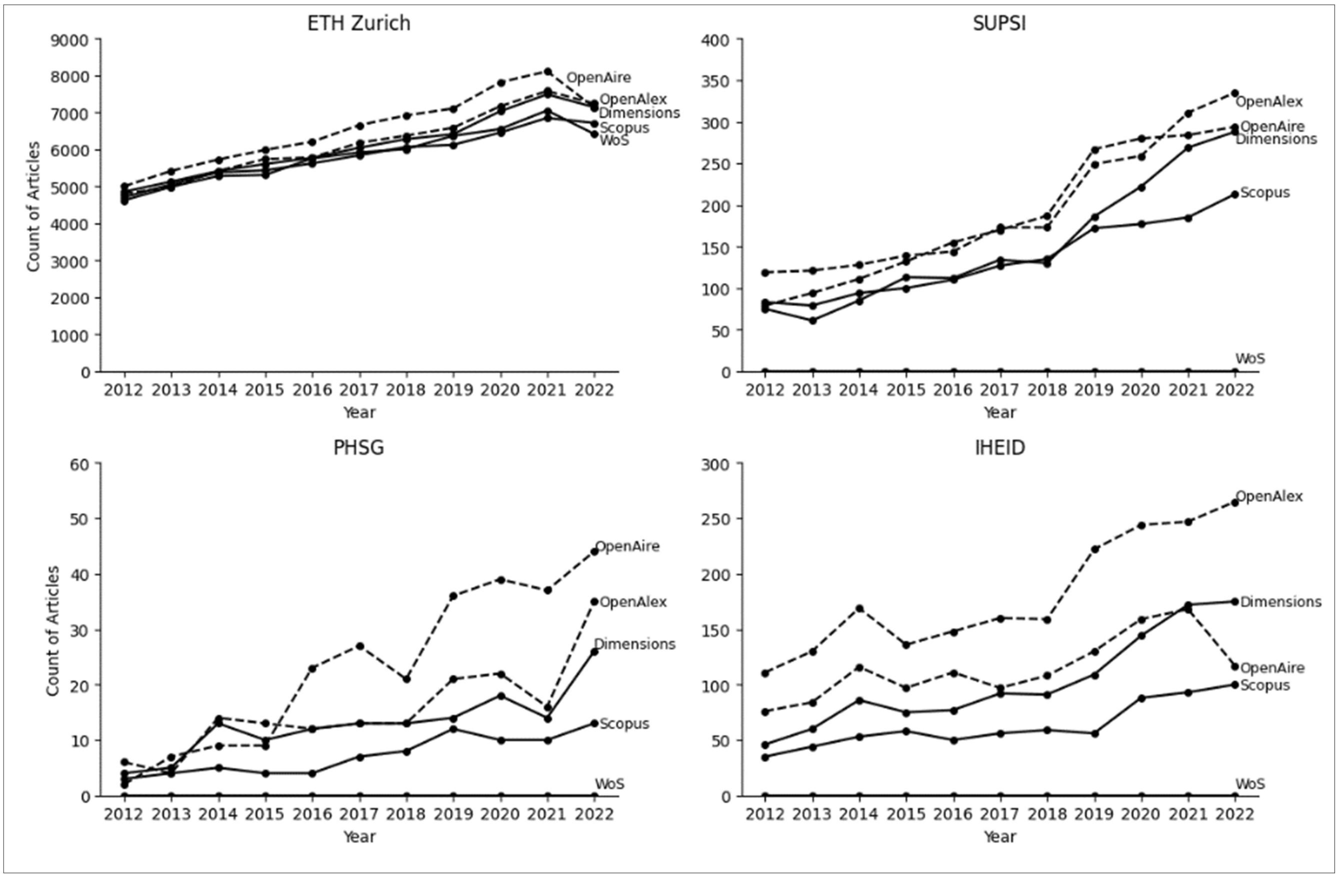
Although each database tends to show similar patterns across the type of institution and the years, the article counts for each institution of the same type can show high variation. As an example, Figure 3 illustrates the article counts of three universities across the five databases for 2021.

[2] Five institutions were not found in Dimensions, OpenAIRE and OpenAlex: Hochschulinstitut Schaffhausen (HSSH), Swiss Business School (SBS), Swiss UMEF (SUMEF), SUPSI – Dipartimento formazione e apprendimento (SUPSI-DFA) and Schweizerisches universitäres Institut für traditionelle chinesische Medizin (TCMUNI). The last missing institution was Pädagogische Hochschule Nordwestschweiz (PHNW/PH FHNW) in Dimensions and OpenAlex, and it was Stiftung Universitäre Fernstudien Schweiz, Brig (SUFS) in OpenAIRE. OpenAIRE classifies institutions into the categories 'approved' and 'pending'. In the 'approved' category, a total of 43 Swiss institutions were found, while 3 institutions appeared with a pending approval. For more details on all data-bases, see Koch et al. (2024).

[3] Despite a smaller coverage, Scopus and WoS include an institution that was not found in the three other databases, namely the Swiss Business School (SBS).

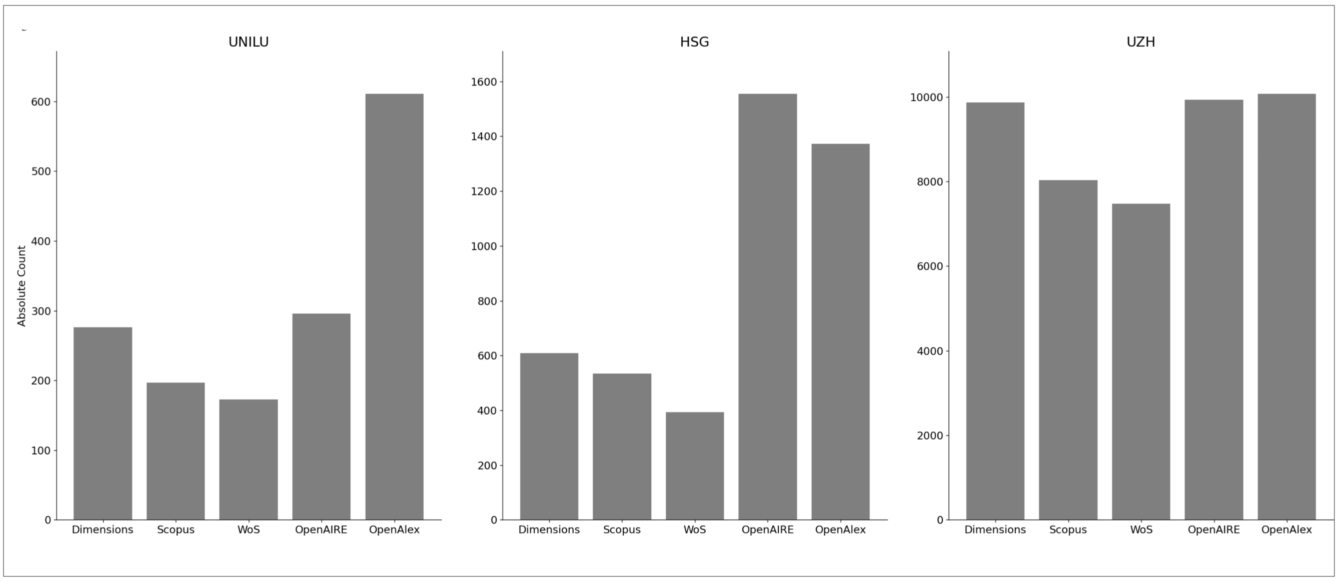
[4] Swissuniversities displays the 52 accredited institutions under four types: "Universities", "Universities of Applied Sciences", "Universities of Teacher Education", "Other institutions of the higher education sector" (Swissuniversities 2024). The list contains 12 universities: 2 federal institutes of technology (ETH Zurich, EPFL) and 10 cantonal universities (Basel, Bern, Fribourg, Geneva, Lausanne, Lucerne, Neuchâtel, St. Gallen, Svizzera Italiana, Zurich).

Fig. 2: Article counts of four Swiss HEIs across 5 databases



Note: articles 2012 to 2022; ETH Zurich = Eidgenössische Technische Hochschule Zürich, IHEID = Institut de Hautes Études Internationales et du Développement, SUPSI = Scuola universitaria professionale della Svizzera italiana, PHSG = Pädagogische Hochschule St. Gallen.

Fig. 3: Article counts of three Swiss universities in 2021 across 5 databases



Note: articles 2021; UNILU = University of Lucerne, HSG = University of St. Gallen, UZH = University of Zurich.

For the University of Zurich, two groups of counts were observed, one being Dimensions, OpenAlex and OpenAIRE, and the other one being WoS and Scopus. For the University of St. Gallen, also two groups can be identified but Dimensions would belong to the second one. For the University of Lucerne, OpenAlex shows much higher counts than the other databases.
This means that a comparison of articles produced by a Swiss institution in a single year, and without additional cleaning, might be significantly affected by the choice of the database, even though each database could possibly provide consistent results over the years.

When looking at article counts from the twelve universities for each year between 2012-2022 consistent differences can be observed. OpenAlex tends to indicate higher counts than any other database, while Scopus is the one indicating the smallest counts.[5] WoS and Scopus systematically show similar counts of journal articles, in the sense that their respective absolute relative differences |c2-c1|/c1 both have a mean smaller than 0.1, a standard deviation smaller than 0.075 and a median smaller than 0.06, while those values are higher for any other pair of databases. The highest absolute relative difference is 2.55 and arises for the University of St. Gallen in 2018: OpenAIRE identifies 785 articles, while 221 articles are to be found in Scopus.

4. Discussion

The lower counts in WoS and Scopus, in particular in comparison to Dimensions, can be explained by the fact that WoS and Scopus have stricter selection processes to include publications, e.g., regarding journal coverage (Singh et al. 2021) or certain subject categories (Martín-Martín et al. 2021, p. 900 for WoS). Singh et al. (2021) have also shown that WoS has a more restrictive journal coverage than Scopus. In light of this, the higher article counts in WoS compared to Scopus are unexpected but have not been investigated in this analysis. Moreover, the differences in counts might be due to differences in the definition of the publication type "article". The higher coverage of OpenAlex might be due to the inclusion of types such as "preprint" or "proceedings" that appear elsewhere in other databases.
The higher counts in OpenAlex might also be explained by a less restrictive selection process. However, a brief look into the publications attributed to the University of Lucerne allows us to identify at least two further reasons for the higher counts. First, one and the same article can appear under two different DOIs, one being linked to the publisher's website, the other to a version deposited on Zenodo (with a prefix 10.5281), a platform used by the University of Lucerne as a repository (https://zenodo.org/communities/lory_unilu). Some of these publications are not deduplicated and are hence counted as two in the counts provided in OpenAlex, although they refer to the same article.[6] Second, and this might affect the counts much more strongly, articles included in the counts have been incorrectly attributed to the University of Lucerne instead of the Cantonal Hospital of Lucerne. This issue affects most raw affiliations containing an English version of the institution name ("Cantonal Hospital Lucerne"), whereas raw affiliations containing a German version ("Kantonsspital Luzern") are correctly attributed. The latter issue also explains differences in counts for the University of St. Gallen, to which a significant number of articles by authors from the Cantonal Hospital St. Gallen are incorrectly attributed in OpenAlex.
When looking at ROR, the main source for institutions used by OpenAlex, one observes that the Cantonal Hospitals of Lucerne and St. Gallen only appear in their German version, "Luzerner Kantonsspital" and "Kantonsspital St. Gallen", without any alternative name in English. Hence, it seems plausible that the errors are due to an affiliation-matching algorithm based on a dataset that does not contain enough variant names. Both examples therefore point to the relevance of language disambiguation and name variants for findability, both of which have been discussed earlier in this article.
Open databases present the advantage of enabling, at least to a certain extent, the improvement of the data. Although users cannot directly modify OpenAlex, it is indirectly possible to update its content, for instance by contributing to the improvement of ROR. We hence tried to improve the affiliation-matching procedure in OpenAlex by including variant names for the Cantonal Hospital of Lucerne and the Cantonal Hospital of St. Gallen in ROR.[7] The integration of the alternative names in the institutional metadata of the concerned institutions takes only a few days or weeks. However, there are several steps in the complex updating process of OpenAlex (https://github.com/ourresearch/openalex-institution-parsing). The update of variant names in ROR might take several months to have an effect on the affiliation data in OpenAlex. However, in the longer term, we can expect the new variant names in ROR to have positive effects on the identification of affiliations.
The introduction of new variant names is not the only action that one can take to improve the data. We also requested the introduction of the six institutions that were not findable in ROR and in OpenAlex.[8] This action had concrete effects in the short term, since ac-

5 The article counts of OpenAlex are higher than those of OpenAIRE in 57% of the cases, higher than Dimensions in 89% of the cases, and higher than Scopus and WoS in 99% of the cases. The article counts in Scopus are smaller than WoS in 57% of the cases, than OpenAIRE in 84% of the cases, than Dimensions in 93% of the cases and than OpenAlex in 99% of the cases.

6 For instance, the two DOIs https://doi.org/10.5281/zenodo.5603246 and https://doi.org/10.1002/mus.27340 receive two different publication identifiers (W4287479005 and W3165812282).

7 The tickets for the respective requests can be found under https://github.com/ror-community/ror-updates/issues/12828 and https://github.com/ror-community/ror-updates/issues/12829.

8 In ROR, we opened requests for the six institutions on 24 May 2024. Except for one, they were validated in a week and integrated in OpenAlex in less than one or two months.

cepted institutions are findable both in ROR and in OpenAlex after only a few weeks. However, we observe that the update in the institutional data had no direct effects on the publications; i.e., in OpenAlex, each new institution is not associated with any publications. Due to the affiliation attribution procedure, it will probably take more time before publications are attributed to these newly introduced institutions.

OpenAIRE is also community-driven. It is hence possible to actively contribute to improving the database. Since ROR is one of the sources of OpenAIRE, the action described above might already contribute to its improvement. However, it is also possible to curate the data in a direct way through OpenOrgs. While in OpenAlex ROR identifiers play a central role and have a one-to-one relationship with the internal identifier, in OpenAIRE the internal identifiers depend on the curation happening in OpenOrgs. Each institution is associated with a main identifier, which has a recognizable prefix ('openorgs'), but further identifiers that have not been validated or merged are also available, recognizable by their prefix ('pending_org'). The presence of unvalidated identifiers reflects the differences in approaches taken by each data provider.

In OpenAlex, each institution is associated with one identifier (a ROR-ID). In OpenAIRE, each institution is also associated with a main identifier, which has a recognizable prefix ('openorgs') and which is usually linked to a ROR-ID but also provides further identifiers that have not been validated or merged (prefix 'pending_org'). The presence of unvalidated data reflects the differences in each data provider's approaches. Our method focussed on the identifiers that are linked to ROR and that could be found to use a clearly defined and efficient procedure. However, many unvalidated identifiers were ignored. The University of Basel, for instance, is associated with one approved identifier under a specific name (a), while many unvalidated identifiers with variant names are also available (b-g): a) 'University of Basel', b) 'Basel university', c) 'UNI: Basel Universität Basel CH', d) 'Universitätsspital Universität Basel', e) 'WWZ Uni Basel Universität Basel', f) 'Abteilung Wirtschaft und Politik Wirtschaftswissenschaftliche Fakultät (WWZ) Universität Basel', g) 'Abteilung Wirtschaftstheorie Wirtschaftswissenschaftliche Fakultät Universität Basel', h) …

The fragmentation due to variant names suggests that cleaning efforts, even at a basic level, might significantly increase the quality of the analysis. Analysts trying to perform a comparative analysis across the institutions of a country might hence potentially miss critical data or include duplicated records if they do not invest time to disambiguate each institution of interest. As of today, this could be discouraging. However, the possibility to curate a database ensures that such an effort is not in vain and reduces time for future curation processes and other users. In OpenOrgs, we were able to merge the listed unvalidated variant names (such as 'Basel university', pending_::00e97cf9dfb17d03db98d2db2cc583e7) with the main identifier ('University of Basel', openorgs____::a5124687d06ee9348a73a7dcfba96ec7). Although the unvalidated identifier seems to have disappeared in OpenOrgs, the variant name remains accessible and it is possible to track each change for each validated identifier. The fact that variant names are still accessible in OpenOrgs after the merging not only enables more traceable data but also raises awareness about the importance of providing metadata in the best format.

The curation of OpenOrgs for our institution – ETH Zurich – allowed us to deduplicate over 200 pending institutional identifiers. We can hence expect that the data dump following our curation will reflect a greater accuracy for this institution. Most of the institutions investigated are associated with less than five deduplicated entries. We hence recommend that representatives of institutions curate the open data that is available in ROR, OpenOrgs and any other database. In the longer term, we can hope that curation effort, in Switzerland and worldwide, will contribute to reducing such discrepancies. This study only looks at Switzerland. The situation in other countries is unknown to us, and data quality might be better in countries with more centralized research evaluation systems or procedures. However, the presented options for improvement remain applicable to any country, regardless of the data quality.

5. Conclusion

The findings presented in this article highlight the role of organizational identifiers, entity disambiguation, and language disambiguation for the findability and visibility of Swiss HEIs in bibliometric databases. In the Swiss context, many non-university institutions, or those with a stronger focus on applied sciences or teaching education, are at a disadvantage for monitoring their output since they are not findable in the commercial databases Scopus and WoS. Moreover, available HEIs in these two databases are usually associated with smaller article counts than in the databases Dimensions, OpenAIRE and OpenAlex. Explanations for the difference in counts are diverse and depend on inclusion criteria, definition of publication type, deduplication processes and affiliation-matching processes.

Considering the differences across databases, we have argued that data-cleaning efforts within scientometric studies are important for ensuring the robustness of scientometric analyses and that curation efforts of databases and organisational identifiers are instrumental in improving the overall data quality of scholarly metadata in the long term. Since scholarly metadata and publication counts continue to be relevant for education and university benchmarking, these comparative insights into databases are also relevant for informed decision-making in HEIs – in Switzerland and elsewhere.

Acknowledgements

This article was written within the framework of the project "Towards Open Bibliometric Indicators" (TOBI), which is co-funded by swissuniversities and the ETH Library. Some of the insights of the analysis underlying this article have been previously presented in blogposts by Dederke (2024) and Koch (2024). The data underlying this article are published and presented by Koch et al. (2024). We are thankful to the OpenAIRE Graph team for providing us with a data dump containing the output associated with Swiss institutions.

References

Asubiaro, T./Onaolapo, S./Mills, D. (2024): Regional disparities in Web of Science and Scopus journal coverage. In: Scientometrics, 129 (3), pp. 1469-1491. https://doi.org/10.1007/s11192-024-04948-x

Benito, M./Gil, P./Romera, R. (2020): Evaluating the influence of country characteristics on the Higher Education System Rankings' progress. In: Journal of Informetrics, 14 (3), 101051. https://doi.org/10.1016/j.joi.2020.101051

Bornmann, L./Haunschild, R./Mutz, R. (2021): Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases. In: Humanit Soc Sci Commun, 8, p. 224. https://doi.org/10.1057/s41599-021-00903-w

Dederke, J. (2024, April 9): Exploring Swiss higher education institutions in commercial bibliometric databases. In: Swiss Year of Scientometrics SYoS. https://yearofscientometrics.ethz.ch/swissuniversities-institutions-in-commercial-databases/

DORA (2013): San Francisco Declaration on Research Assessment, https://sfdora.org/read/

Donner, P./Rimmert, C./van Eck, N. J. (2020): Comparing institutional-level bibliometric research performance indicator values based on different affiliation disambiguation systems. Quantitative Science Studies, 1 (1), pp. 150-170. https://doi.org/10.1162/qss_a_00013

Fauzi, M. A./Tan, C. N.-L./Daud, M./Awalludin, M. M. N. (2020): University rankings: A review of methodological flaws. Issues in Educational Research, 30 (1), pp. 79–96.koch_michelle@outlook.com https://search.informit.org/doi/10.3316/informit.085786218299205

Harzing, A. W. (2019): Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science? In: Scientometrics, 120 (1), pp. 341-349. https://doi.org/10.1007/s11192-021-03948-5.

Hook, D. W./Porter, S. J./Herzog, C. (2018): Dimensions: building context for search and evaluation. In: Frontiers in Research Metrics and Analytics, 3, Artikel 23. https://doi.org/10.3389/frma.2018.00023

Hug, S. E./Brändle, M. P. (2017): The coverage of Microsoft Academic: Analyzing the publication output of a university. In: Scientometrics, 113, pp. 1551-1571. https://doi.org/10.1007/s11192-017-2535-3

Koch, M. (2024, July 4): Exploring Swiss Higher Education Institutions in Commercial and Open Bibliometric Databases. https://yearofscientometrics.ethz.ch/exploring-swiss-higher-education-institutions-in-*commercial-and-open-bibliometric-databases/*

Koch, M./Dederke, J./Willemin, S. (2024): Publication counts for Swiss Higher Education Institutions in five bibliometric databases (2012-2022). ETH Zurich: Research Collection. https://doi.org/10.3929/ethz-b-000694041

Machado, A./Hoppmann, L./Knaus, J./Palzenberger, M. (2016): Bibliometric study of the Swiss Publication System. Zenodo. https://doi.org/10.5281/zenodo.167381

Manghi, P./Atzori, C./Bardi, A./Baglioni, M./Schirrwagen, J./Dimitropoulos, H./La Bruzzo, S./Foufoulas, I./Mannocci, A./Horst, M./Czerniak, A./Iatropoulou, K./Kokogiannaki, A./De Bonis, M./Artini, M./Lempesis, A./Ioannidis, A./Manola, N./Principe, P./Vergoulis, T./Chatzopoulos, S./Pierrakos, D. (2022): OpenAIRE Research Graph Dataset, Dataset, Zenodo. https://doi.org/10.5281/zenodo.3516917

Martín-Martín, A./Thelwall, M./Orduna-Malea, E./Delgado López-Cózar, E. (2021): Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. In: Scientometrics, 126 (1), pp. 871-906. https://doi.org/10.1007/s11192-020-03690-4

Mongeon, P./Paul-Hus, A. (2016): The journal coverage of Web of Science and Scopus: a comparative analysis. In: Scientometrics, 106, pp. 213-228. https://doi.org/10.1007/s11192-015-1765-5

Priem, J./Piwowar, H./Orr, R. (2022): OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. ArXiv. https://arxiv.org/abs/2205.01833

Purnell, P. J. (2022): The prevalence and impact of university affiliation discrepancies between four bibliographic databases – Scopus, Web of Science, Dimensions, and Microsoft Academic. In: Quantitative Science Studies, 3 (1), pp. 99-121. https://doi.org/10.1162/qss_a_00175.

Rousseau, D./Garcia-Zorita, C./Sanz-Casado, E. (2023): Publications during COVID-19 times: An unexpected overall increase. In: Journal of Informetrics, 17 (4), 101461. https://doi.org/10.1016/j.joi.2023.101461

Singh, V. K./Singh, P./Karmakar, M./Leta, J./Mayr, P. (2021): The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. In: Scientometrics, 126, pp. 5113-5142. https://doi.org/10.1007/s11192-021-03948-5

Stricker, M. (2023): Open Access Monitoring: Verzerrende Datenquellen und unbeabsichtigte Leerstellen – eine explorative Studie. in: Bibliothek Forschung und Praxis, 47 (2), S. 382-392. https://doi.org/10.3929/ethz-b-000622885

Swissuniversities (2024): Accredited Swiss Higher Education Institutions in accordance with HEdA (Swiss Higher Education Funding and Coordination Act, German: Schweizer Hochschulförderungs- und Koordinationsgesetz). https://www.swissuniversities.ch/en/topics/studying/accredited-swiss-higher-education-institutions-1. (21.02.2024)

UNESCO (2021): UNESCO recommendations on Open Science. https://doi.org/10.54677/MNMH8546

Visser, M./Van Eck, N. J./Waltman, L. (2021): Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. In: Quantitative science studies, 2 (1), pp. 20-41. https://doi.org/10.1162/qss_a_00112

■ **Julian Dederke,** Dr. sc., Consultant Research Data Management and Scientometrics, ETH Library, ETH Zurich, E-Mail: julian.dederke@library.ethz.ch

■ **Michelle Koch,** Bachelor of Arts in Economics and Management, Data Engineer Student Assistant, ETH Library, ETH Zurich, E-Mail: koch_michelle@outlook.com

■ **Simon Willemin,** Data Scientist, ETH Library, ETH Zurich, E-Mail: simon.willemin@library.ethz.ch

Jahresverzeichnis 2024

Das Inhaltsverzeichnis des Jahrgangs 2024 der Zeitschrift Qualität in der Wissenschaft finden Sie auf unserer Website als PDF-Datei:

https://www.universitaetsverlagwebler.de/jvz

Hauptbeiträge der aktuellen Hefte Fo, HSW, P-OE, QiW und ZBS

Auf unserer Website www.universitaetsverlagwebler.de erhalten Sie Einblick in das Editorial und Inhaltsverzeichnis aller bisher erschienenen Ausgaben. Nach zwei Jahren sind alle Ausgaben eines Jahrgangs frei zugänglich.

Fo

Forschung

Politik - Strategie - Management

Fo 1+2/2024
Transdisziplinarität

Sabine Maasen
Transdisziplinierung der Wissenschaft – zwischen Glanz und Elend

Christian Dayé, Raphaela Maier & Peter Obersteiner
Zukunftsbilder in transdisziplinären Settings erarbeiten
Erfahrungen mit Zukunftswerkstatt und Participatory Systems Mapping

Jessica Nuske
Disziplinäre und transdisziplinäre Sozialwissenschaften
Distinktionen und Synergiepotenziale wissenschaftlicher Forschungs- und Transferleistungen

Lena Theiler, Alexandra Lux, Michael Kreß-Ludwig & Flurina Schneider
Schlüsselkonzepte und Umsetzung transdisziplinärer Forschung

Sabine Hennig
Digitale Geomedien und ihr Einsatz in Bildung für Nachhaltige Entwicklung
Transdisziplinäre Entwicklung als Grundlage für zielgruppenzentrierte Lehr- und Trainingsmaterialien, Empfehlungen und Konzepte

HSW

Das Hochschulwesen

Forum für Hochschulforschung, -praxis und -politik

HSW 4+5/2024

Hannes Müller
Das nachlassende Interesse an technischen Universitäten in Österreich
Status quo und Gegenmaßnahmen

Philipp Schultes, Patrick Gollup & Jörg Holle
Studienplatzoffensiven als Instrument für mehr Lehrkräftenachwuchs? Evaluation eines Ansatzes aus Nordrhein-Westfalen und Gedanken zur Übertragbarkeit für das Lehramt an beruflichen Schulen

HSW-Gespräch zwischen Annette Greiner, Steffen Lang, Heike Thiele, Leo Lüngen & Wolff-Dietrich Webler
Zum Zustand des Schulunterrichts, zu Problemen der Schülerschaft und des Reformbedarfs der Lehrer(aus)bildung

Ewald Scherm
Die Wahl zum „Rektor des Jahres“
Kein Ranking (der Eignung) deutscher Universitätsrektoren

Roland Bloch & Sylvi Mauermeister
Von der Entropie zur Strategie? Organisationsmuster der Digitalisierung von Hochschulbildung nach dem Ende der Pandemie

Anouâr Gadermann & Chuansen Zhao
Sprungbrett oder Holzweg? Zur Gewichtung des Studiums in Deutschland bei der Wahl chinesisch-deutscher MINT-Studiengänge (am Beispiel der CDTF)

P-OE

Personal- und Organisationsentwicklung
in Einrichtungen der Lehre und Forschung

Forum für Führung, Entwicklung und Transformation

POE 2/2024
Die Hochschule als attraktiver Arbeitsplatz – People Development / FutureSkills

Natalie Hotz, Anke Diez & Heike Schäfer-Lammert
Etablierung eines strategischen Human Resources-Ansatzes in einer wissenschaftlichen Organisation am Beispiel des KIT

Katrin Heß, Bettina Krings & Katrin Klink
„Zukunftsthemen Personal“ – Die Entwicklung eines Reflexionsrahmens für die zukünftige Ausrichtung der Personalarbeit in einer Wissenschaftseinrichtung

Ines Köhler & Julia Schuller
Zielgruppenspezifische Personalentwicklung von High Potentials am Beispiel des Postdoc-Karriereentwicklungsprogramms „Young Investigator Group Preparation Program“

Anne Mazunga
Kreativität im Team & Kreativität als Team

P-OE-Gespräch mit Dr. h.c. Thomas Sattelberger
Zur Bedeutung von strategischer Personal- und Organisationsentwicklung an deutschen Hochschulen und Forschungseinrichtungen

HM

Hochschulmanagement

Zeitschrift für die Leitung, Entwicklung und Selbstverwaltung von Hochschulen und Wissenschaftseinrichtungen

HM 3+4/2024

Sascha Armutat, Jacqueline Albers & Maren Grün
Zielgruppenspezifische Rekrutierung von Professor*innen an Hochschulen für angewandte Wissenschaften (HAW)

Valerie Hug
Entscheidungsverhalten in der Ausgestaltung einer Grauzone – Hochschulpraktische Implikationen der Beforschung professoraler Personalauswahl

Kalle Hauss & Alexander Raev
Politiken der Personalgewinnung in der Wissenschaft – eine Analyse des wettbewerblichen Umfelds auf der Basis staatlicher Förderprogramme

Ayşegül Engin & Michaela Schaffhauser-Linzatti
Was tun mit Gewinnen? Eine Analyse zum strategischen Gewinnausweis österreichischer Universitäten

Karina Sopp & Josef Baumüller
Die Bedeutung der Nachhaltigkeitsberichterstattung gemäß CSRD für Hochschulen

Cedric Bardowieck & Joachim Prinz
Nicht ohne die Uni! Die Hochschule als vielversprechende Multiplikatorin in der nationalen Finanzbildungsstrategie

Michael Hölscher & Dieter Kaufmann
Wissenschaftsmanagement in Landeshochschulgesetzen – Bestandsaufnahme und kritische Reflexion

ZBS

Zeitschrift für Beratung und Studium

Handlungsfelder, Praxisbeispiele und Lösungskonzepte

ZBS 3+4/2024
Mediatisierung der Studienberatung. Von OSA's über Blended Counseling bis hin zur künstlichen Intelligenz

Dennis Mocigemba
Medienensembles und Medienrepertoires in der Studienberatung

Minnie Silfverberg & Martina Hörmann
Kompetenzentwicklung für Beratung im digitalen Raum

Malte Hübner
„Wie tickt die Generation Alpha?"

Franz Oberlehner, Nicole Undeutsch & Laura Waschulin
Änderungen der Psychischen Belastung und der Beratungsanliegen von Klientinnen und Klienten in der Psychologischen Studierendenberatung.

Julia Schaller
Studienabbruch als Systemische Heldinnenreise

Stephanie Kessens & Wiebke Lückert
Gutes besser machen. Qualitätsentwicklung in der Beratung an Hochschulen

ChatGPT4o, Niklas Kempter, Dennis Mocigemba & x mal Anonymous
Künstliche Intelligenz und Large Language Models in der Studienberatung? Oder: „Wer ist hier noch echt?"

Zwei neue E-Book-Publikationen im UVW

In der Reihe ***Hochschulwesen: Wissenschaft und Praxis*** sind zwei Sammelbände erschienen, die sich mit ganz unterschiedlichen Themenfeldern rund um Hochschule und Wissenschaft befassen.

Yvette E. Hofmann (Hg.)

Die psycho-soziale Situation von Studierenden in der (post-)pandemischen Zeit Stand der Forschung und Impulse aus der Praxis

Die Corona-Pandemie hat die Hochschulen als Orte der Begegnung, des Austauschs und der Vernetzung stark verändert. Dieser Sammelband beleuchtet aus inter-disziplinärer Perspektive die Herausforderungen, denen Studierende vor, während und nach der Pandemie begegneten. Expertinnen und Experten aus Gesundheitswissenschaften, Hochschulforschung, Pädagogik, Psychologie und Soziologie untersuchen nicht nur die psychischen und sozialen Belastungen, sondern präsentieren auch Lösungsansätze und Strategien zur Unterstützung der Studierenden. Dabei werden sowohl individuelle als auch strukturelle Faktoren berücksichtigt, um ein umfassendes Bild der Wechsel-wirkungen zu zeichnen.

Damit greift der Sammelband ein zentrales Thema auf, das in den letzten Jahren in der Hochschulforschung stark an Bedeutung gewonnen hat: die Rolle des sozialen Faktors für die psychische Gesundheit von Studierenden. Durch die Untersuchung der psycho-sozialen Situation in der (post-) pandemischen Zeit erhält diese Diskussion eine wichtige neue Dimension.

Der Sammelband bietet wertvolle Einblicke für Wissen-schaftler*innen, Praktiker*innen, Hochschulpolitiker*innen und Studierende. Er soll dazu anregen, die Bedürfnisse der Studierenden besser zu verstehen und Wege zu finden, ihre Studienzeit künftig noch gesünder und positiver zu gestalten.

Wolff-Dietrich Webler (Hg.)

Überzogener und überhitzter Wettbewerb in der Wissenschaft (Band II)

Wettbewerb ist etwas Normales in vielen Situationen des Lebens – so auch in der Wissenschaft. Auf seine Varianten wird in diesem Band eingegangen. Trotzdem ist festzustellen, dass es zahlreiche Anzeichen dafür gibt, dass in der Ausgestaltung von Wettbewerb von den Trägern des Wissenschaftssystems (meistens dem Staat) so gravierende Fehler begangen wurden und noch immer begangen werden, dass die Situation als „überzogen" bezeichnet werden muss. Dies geht vor allem zu Lasten von Forschung und Lehre und der Wissenschaftler*innen selbst. Worin liegen die Ursachen? Welche Kriterien müssen für das Verdikt „überzogen" erfüllt sein?

Wettbewerbe in der Wissenschaft von Seiten des jeweiligen Trägers zu inszenieren, verfolgt nicht immer wissenschaftsimmanente und der Wissenschaft förderliche Ziele. Diese Vorgänge in theoretischen und analytischen Perspektiven aufzuschlüsseln, hat sich dieser Band in seinen 5 Beiträgen vorgenommen. Dabei werden nicht nur wichtige Zusammenhänge deutlich, sondern den wissenschaftspolitischen Akteurinnen und Akteuren auch gravierende Fehler nachgewiesen, die kaum glaublich erscheinen.

Die Bände I und II zu dieser Thematik können in ihrem gesammelten Ergebnis als fundierte Politikberatung aufgefasst werden.

Die freie Verfügbarkeit der E-Book-Ausgabe dieser Publikation wurde ermöglicht durch die Deutsche Forschungsgemeinschaft, den Fachinformationsdienst Erziehungswissenschaft und Bildungsforschung und ein Netzwerk wissenschaftlicher Einrichtungen und Bibliotheken zur Förderung von Open Access in den Erziehungs- und Bildungswissenschaften.