# Investigating the research output of institutions

Wolfgang G. Stock, Gerhard Reichmann, Christian Schlögl



Suggested Citation:
Stock, W. G., Reichmann, G., & Schlögl, C. (2025). Investigating the research output of institutions. Journal of Informetrics, 19(2), Article 101638. https://doi.org/10.1016/j.joi.2025.101638

Journal of Informetrics 19 (2025) 101638

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Research Paper

# Investigating the research output of institutions

Wolfgang G. Stock [a,*], Gerhard Reichmann [b], Christian Schlögl [b,1]

[a] Dept. of Information Science, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
[b] Dept. for Operations and Information Systems, University of Graz, Graz, Austria

ARTICLE INFO

*Keywords:*
Evaluative scientometrics
Institution
Publications
Affiliation
Address
Affiliated authors

ABSTRACT

Describing, analyzing, and evaluating research institutions are among the main tasks of scientometrics and research evaluation. But how can we optimally search for an institution's research output? Possible search arguments include institution names, affiliations, addresses, and affiliated authors' names. Prerequisites of these search tasks are complete lists (or at least good approximations) of the institutions' publications, and—in later steps—their citations, and topics. When searching for the publications of research institutions in an information service, there are two options, namely (1) searching directly for the name of the institution and (2) searching for all authors affiliated with the institution in a defined time interval. Which strategy is more effective? More specifically, do informetric indicators such as recall and precision, search recall and search precision, and relative visibility change depending on the search strategy? What are the reasons for differences? To illustrate our approach, we conducted an illustrative study on two information science institutions and identified all staff members. The search was performed using the Web of Science Core Collection (WoS CC). As a performance indicator, applying fractional counting and considering co-affiliations of authors, we used the institution's relative visibility in an information service. We also calculated two variants of recall and precision at the institution level, namely search recall and search precision as informetric measures of performance differences between different search strategies (here: author search versus institution search) on the same information service (here: WoS CC) and recall and precision in relation to the complete set of an institution's publications. For all our calculations, there is a clear result: Searches for affiliated authors outperform searches for institutions in WoS. However, especially for large institutions it is difficult to determine all the staff members in the time interval of research. Additionally, information services (including WoS) are incomplete and there are variants for the names of institutions in the services. Therefore, searching for institutions and the publication-based quantitative evaluation of institutions are very critical issues.

## 1. Introduction

The description, analysis, and evaluation of research institutions are important topics in informetrics, especially in scientometrics and research evaluation. Many scientometric studies are based upon collections of publications by the institutions under investigation (Cappalletti-Montano et al., 2022). Even world-wide university rankings bank on publication-based indicators including the Academic

* Corresponding author.
*E-mail addresses:* wolfgang.stock@hhu.de (W.G. Stock), gerhard.reichmann@uni-graz.at (G. Reichmann).
[1] Christian Schlögl died during the publishing process of this article.

https://doi.org/10.1016/j.joi.2025.101638
Received 21 July 2024; Received in revised form 22 December 2024; Accepted 15 January 2025
Available online 26 January 2025

Ranking of World Universities, the Quacquarelli Symonds World University Rankings, the Times Higher Education World University Ranking, the Leiden Ranking, and the U.S. News Best Global Universities, all of them applying data from multi-disciplinary bibliographic information services as Web of Science (WoS) or Scopus.

There are manifold research results in informetrics on research institutions. What is missing is research on the empirical basis of such investigations. As one of the main indicators for describing and analyzing institutions is the number of published documents by the institution within a defined time period (Bornmann et al., 2023), informetric studies on institutions have to ascertain the completeness or at least the representativity of the amount of found documents. It seems to be very easy to search for institutions: identify the institution's name, select an information service, search in the respective data field, download the search results, and analyze the hit set. Although this may have been done sometimes, the results of such an approach are certainly biased. Why? And how can this be done satisfactorily? The aim of our study is to close this research gap.

While performing evaluative scientometrics of institutions transparency of the entire scientometric processes is needed. In this article, we concentrate on correct searching and identifying institutions and their publication activities. Necessary conditions of empirical studies on institutions are complete lists (or at least good approximations) of the institutions' publications, and—in later steps—their citations, and topics. But how can we search for institutions, for instance universities, university institutes and departments, hospitals, or research units of private companies? Should we search for the name of the institution or for the names of the institution's researchers? Such informetric descriptions and evaluations of an institution's research production and its impact are necessary for the management of research units as well as for governmental decision making in research and technology policy (Huang et al., 2014). Scientometricians have been studying research institutions for decades; early examples include a university research group (Moed et al., 1985), a university (De Bruin et al., 1993), and a non-profit institution (Albrecht et al., 1994); however, not all methodological problems are actually successfully solved. In addition, it has been known for decades that the naming of institutions in publications as well as in information services is often inconsistent.

Research on departments or institutes is at the meso-level of scientometrics (Rousseau et al., 2018, p. 247). At this level a particular problem arises (Stock et al., 2023b): Who is a member of an institution in the considered time frame (Russell & Rousseau, 2009)? Is it possible to get such information due to data protection regulations? If we use professional online information services such as Web of Science (WoS), Scopus, Dimensions, or Google Scholar to obtain information at the meso-level, there are two options to search and count publications of an institution, namely (1) searching for all staff members of the institution in the observed time window or (2) searching directly at the institutional level (Stock et al., 2023a). Search results can be jeopardized if each approach provides different results.

For option 1 and the tasks of research evaluation, it is necessary to avoid double counting of publications written by two or more members of the same institution (Gauffriau et al., 2007; Korytkowski & Kulczycki, 2019). Therefore, one can only work with fractional author values ($1/n$ given $n$ co-authors per paper) and not with whole counting ($1$ for each co-author) (Aman & van den Besselaar, 2024). If we use option (1), i.e., searching for the institutions' authors and then aggregating the data at the institutional level, or option 2, i.e., searching directly for the institution, are there the same results for different information services or do the results differ? In an ideal scenario with accurate identification of institutions on the provider side and with correct search arguments on the searcher side, recall and precision should be optimal and the institution's visibility results should be the same. In this scenario, the relative visibility (Dorsch, 2017) of an institution with respect to an individual information service at the meso-level is identical to the aggregated value of the visibility of all individual researchers' articles at the micro-level. This is where we found our fundamental research question: Is this really the case? Using two university institutes as an illustration of the informetric problems and challenges, we are going to answer the research question using the example of the multi-disciplinary information service WoS. Our aim is not to theoretically discuss methods for disambiguating institution names, but rather the practice of searching for institutions in bibliographic online databases for the purposes of scientometrics and research evaluation. For institutions serving multiple scientific disciplines, additional problems arise (Abramo & D'Angelo, 2015), e.g., the different publishing cultures in those research fields. But also here the basis of performance studies is the data set of publications.

What is new in this study?

- To what extent do the retrieved hits differ, when searching for the institutions' names or when searching for the publications of their affiliated members?
- How to deal with authors who are co-affiliated, i.e., affiliated with more than one institution?
- Which informetric indicators should be applied for the description and evaluation of different search strategies for the same information service?
- How can we calculate an institution's relative visibility on an information service, also for different search strategies?

## 2. Institutions in online information services

In this paragraph, we will theoretically discuss the naming of institutions including their address data and their roles as affiliations of their researchers, the correctness of those data in journals and information services, the type of the institution (e.g., hospital, university institute, private company's research unit), the level of an institution (e.g., university, institute, department), and the indexing of institutions in information services including their identification through a unique number (as, for instance, the Research Organization Registry identifier). We will also describe the counting of publication shares for papers produced by more than one institution and by authors who are affiliated with more than one institution.

## 2.1. Naming of institutions

### 2.1.1. Address data, affiliations, and levels of an institution

What is the difference between an address and an affiliation? An address includes information about the location of an institution's building, e.g., the institution's name, street and house number, postal code and city name as well as the country. The affiliation is the institutional home of an academic or scholarly author of a paper in which the author conducted their research. It is "a part of academic culture" (Di Leo, 2003, p. 4).

Affiliations refer to different levels of research institutions (Fiala, 2013, 2014). Both address information and affiliation information may be structured in academic information services in a general scheme including defined abbreviations. A necessary precondition for any search for institutions is to indicate the correct affiliation–or if there are more than one–all affiliations of a paper's authors (Polonioli, 2024). Before searching for all members of an institution, it has to be clarified who is a member of the research institution, for instance, all staff members, all faculty, or all publishing people including students of the institution. However, it may be difficult to collect such personal data of all affiliated persons.

What is the type of the research institution, which should be evaluated? In fact, there are many research institutions that are universities (for instance, Massachusetts Institute of Technology), but we can also find manifold research institutions besides universities (Hardeman, 2013), for instance (and we stay in Massachusetts), institutes affiliated with a university (e.g., Center for Intelligent Information Retrieval at the University of Massachusetts Amherst), non-profit research institutes (e.g., Woods Hole Oceanographic Institute), agencies (e.g., Massachusetts' Executive Office of Education), private companies (Hardeman, 2013) (e.g., The Kraft Group, Foxborough, MA), and hospitals (Pylarinou & Kapidakis, 2017; Schmidt et al., 2016) (e.g., Massachusetts General Hospital). If private companies have subsidiaries, how should we count their research activities? For example, should we attribute the R&D output of Rand Whitney (which is a Kraft Group subsidiary) to The Kraft Group, Rand Whitney, or both?

What is identity in academic culture, i.e., affiliation: Is it the university, the faculty, the institute, or a specific department or a working group, maybe assigned to more than one institute? We should take a look at an example (representing the former affiliation of one of this article's authors):

| | |
|---|---|
| Department | Dept. of Information Science |
| Institute | Institute for Language and Information |
| Faculty | Faculty of Arts and Humanities |
| University | Heinrich Heine University Düsseldorf. |

A search for an academic field, in this case information science, is only possible at the department or institute level; a search for the university's output is only possible at the university level. Some universities have introduced affiliation policies (e.g., Uni Düsseldorf, 2023) regulating the spelling of the name (in our example "Heinrich Heine University Düsseldorf" or "HHU Düsseldorf" as a valid abbreviation). Similar to the author identification number ORCID (Open Researcher and Contributor ID), ROR (Research Organization Registry) has been established to identify institutions (Lammey, 2020). ROR is an initiative of California Digital Library, Crossref, and DataCite. It contains data from the former institutional register GRID (Global Research Identifier Database). Our example university from Düsseldorf has the ROR ID *https://ror.org/024z2rq82*. In some cases, ROR supports parent-child organizational hierarchies as well as other types of relationships between institutional entities. However, affiliations that fall outside the scope of ROR include many university institutes and departments; therefore, detailed searches for these affiliations are not always possible–which is an important limitation for detailed institutional searches and scientometric studies.

### 2.1.2. Correctness of affiliations in information systems

Is the affiliation in the publishing journals or in the proceedings and in information services correct? Is it even stated? Dimzov et al. (2021) found for their case studies that there are dozens of name variants for the same institution. For papers from Canada, van Bellen (2023) stated that Dimensions was able to add a ROR identifier to approximately 62% of subscription-based articles and only 47% of open access papers. For WoS, we found some data on missing affiliation entries: Liu et al. (2017, p. 361) reported that 5% of the Science Citation Index Expanded (SCIE) records, 9% of the Social Science Citation Index, and 42% of the Arts & Humanities Citation Index records did not contain an institution name. There are two reasons for the absence of affiliation information: the articles themselves do not report it (about 60% of all cases examined by Liu et al., 2018) or they are not indexed in the WoS (40%). For articles from Spain and SCIE, García-Zorita et al. (2006) describe that 65% of the older literature (from 1985 to 1997) do not have a research address, while this is the case for 99.8% of papers from 1998 to 2004. Older studies identified problems with searching for institutions in bibliographic information services, including translation of foreign names (Stefaniak, 1987) and name changes when an institution merges with another or splits into new entities (Hoed & Wilson, 2003). Some addresses are incomplete, for instance, "U Berlin," because there is more than one university in Berlin (Winterhager et al., 2014). Another problem occurs if an author changes institutions and the article written at the old institution (*Inst)* and published months or even years later only names the new institution (*Inst')*. Problems may also arise with special characters in the institution's name, including German Umlauts (e.g., ü), accent marks in French (e.g., û), and other diacritic signs, not to mention names occurring in writing systems outside the Latin alphabet.

Multiple name variants of an institution are not always errors on the part of information services, but can also result from misleading data in the publications originated by the authors. For instance, Bador and Lafouge (2005) found in their case study of French authors that they frequently provided incorrect or even no data on their affiliations in their articles. Similar findings were

reported by Khalifa et al. (2023) based upon observations of the publishing behavior of Egyptian researchers.

#### 2.1.3. Indexing of institutions

Aside from the intellectual indexing of the institution names (Winterhager et al., 2014; Donner et al., 2020), there are automatic approaches to disambiguate institutions name in professional information services. The providers of WoS (Birkle et al., 2020), Scopus (Baas et al., 2020), and Dimensions (Herzog et al., 2020) claim that their databases are suitable for bibliometric analyses. However, Garfield, the founder of the Science Citation Index, which has been part of WoS for decades, emphasized that his database was built primarily for bibliographic searches and not for scientometric analyses (Garfield & Stock, 2002). Despite some limitations in terms of functionality, Google Scholar is also used for bibliometric studies (Moed et al., 2016; Delgado López-Cózar et al., 2019; Gusenbauer, 2019; Pereira & Mugnaini, 2023). Of course, discipline-specific databases would also be useful sources for searches for institutions, but only if they offer searchable institution information. Examples include the Derwent World Patents Index for the patent literature or Medline for the life sciences.

If one wants to find all publications of an institution, all of its papers have to be recorded in an information service. However, the leading bibliographic information services such as WoS, Scopus, Dimensions, and Google Scholar are incomplete compared to researchers' personal publication lists (Dorsch et al., 2018; Hilbert et al., 2015) and sometimes distorted in terms of language and discipline.

How do information providers handle affiliation indexing? All of the information services mentioned indicate that they use strategies for disambiguating institutions, resulting in preferred names referred to as "Organizations Enhanced" (OE) in WoS and "Scopus Affiliation ID" (AFID) in Scopus (Pranckutè, 2021, pp. 15 f.; Donner et al., 2020, p. 155). WoS focuses on higher education institutions in its affiliation search field and separates university hospitals from their parent institutions. The address field can be used to search for parts of universities, e.g., institutes or departments. Scopus works primarily at the top institutional level (including universities, hospitals, and private companies), but provides some (very incomplete) information on the affiliation hierarchy.

Dimensions applies (only in the paid version) a mapping of the address information to a preferred version via ROR entries (van Bellen, 2023), and, finally, Google Scholar works in its Citation Pages with links from author pages to their institutions; however, there is no search field for institutions in Google Scholar, so it does not seem to be very useful for systematically searching for institutions (Orduña-Malea et al., 2017).

Since the results of WoS are heavily dependent on the subscribed sub-databases of WoS (Stock, 2021), e.g., the inclusion or exclusion of the Emerging Sources Citation Index (De Filippo & Gorraiz, 2020), scientometricians have to exactly describe the sub-databases used and the subscribed time periods, e.g. WoS's Core Collection (WoS CC).

Purnell (2022, p. 117) found "that discrepancies in WoS were most frequently due to problems with unifying variants and in some cases, confusion clearly led to assigning records to the wrong institution. Discrepancies among Dimension's records were more likely found due to missing affiliations but there were also some issues with unification. In Scopus … there was no clear pattern and causes of discrepancy were mixed." Venet presented misleading affiliations, especially for Russian institutions, for both WoS (Venets, 2014) and Scopus (Venets, 2017).

#### 2.1.4. Counting multiple-institutions papers

When counting an author's publications, Egghe and Rousseau (1990, pp. 275 ff.) differentiated between whole counting (publication counts as *1* for all co-authors), and fractional counting (publication counts as *1/n* in the context of *n* co-authors). All in all, Gauffriau (2021) found 32 different counting methods for an author's share of a multi-authored research paper, with whole counting and fractional counting being the most popular. Since there are not only co-authors, but also collaborating institutions (i.e., two or more institutions are involved in the creation of an article), whole counting and fractional counting are mirrored at the institutional level–albeit a little more complicated. If all authors can be assigned to exactly one institution, the counting is carried out in a similar way to co-authorship. If we apply whole counting, each institution mentioned counts *1*. If we apply fractional affiliation counting (and we will work with this method in our illustrative study for the calculation of relative visibility), we count the number $m_{Inst}$ of authors of each participating institution in a given paper and divide this number by the number of all authors *n* of that paper. For instance, if there is a paper written by five authors and four of them are from institution *A* and one is from institution *B*, the paper counts *4/5* for *A* and *1/5* for *B*.

Nowadays there is no doubt that some authors may belong to several institutions, for example in a permanent position at institution *B* and as a visiting professor at institution *X*, which is called "co-affiliation" (Hottenrott & Lawson, 2022; Bachelet et al., 2019a,b). In the case of co-affiliations of authors, their share has to be divided by the number of different institutions *p*. Actually, we would have to differentiate more precisely here as soon as the work performance (work obligation) of an author differs at the individual institutions. For example, the University of Graz awards so-called "quarter professorships" ("*Viertelprofessuren*") worth 10 working hours per week. Assuming that these "quarter professors" are fully employed at their home universities, i.e., they are supposed to work for (at least) 40 hours per week, the shares would have to be divided in a ratio of 1:4. If this is neglected, it would be unfair if the university which employs such quarter professors tries to benefit from the publications of these researchers as much as possible, i.e., "ideally" in the form of full counting.

The final value *C* for the institution *Inst* with regard to one paper is the sum of all *1/p*1/n*-values of the *m* authors of the institution with respect to that paper (however, special problems such as quarter professorships are neglected, since data on such employment relationships are not always accessible):

$$C(\text{Inst}) = \sum_{k=1}^{m} \frac{1}{p(k)} \times \frac{1}{n(k)}, \quad (1)$$

where *C(Inst)* is the fractional counting value for the institution *Inst, m* counts the number of authors of *Inst, n* is the number of all co-authors, *p* is the number of different affiliations of each author at the time of research for the article, and *k* denotes the individual authors of *Inst*. Our example author from *B* and *X* counts ½ * 1/5 = 1/10 for *B* and also for *X*. If all four co-authors of *A* are only working for *A, C(A)* = 1/1*1/5 + 1/1*1/5 + 1/1*1/5 + 1/1*1/5 = 4/5; this value therefore remains the same as if co-affiliations of authors are not taken into account. If one of *A*'s authors is a visiting professor at *Y, C(Y)* = ½ * 1/5 = 1/10 and *C(A)* changes to 1/2*1/5 + 1/1*1/5 + 1/1*1/5 + 1/1*1/5 = 7/10.

## 3. Quality indicators for searches for research institutions

In this paragraph, we present informetric indicators for searching institutions including the classic indicators of recall and precision, the new indicators of search recall and search precision in order to describe differences of multiple research strategies for the same information service, and the relative visibility of institutions on information services. All of these indicators are clarifications or related terms of the broader concept of "coverage," a quality indicator defined as the ratio of the number of documents represented by an information service to the number of all available documents as reported by other sources (Hilbert et al., 2015). Recall is an aspect of the coverage of an information service, relative recall and relative precision describe the quality of the hits resulting from the search process, and, finally, relative visibility describes the coverage seen from the perspective of the institution. Using these indicators, we find our concrete research questions.

### 3.1. Recall and precision

Since the early days of information science researchers have been working with effectiveness measures of retrieval systems, especially with recall and precision (Stock & Stock, 2013, pp. 113-115). Recall (*R*) is the quotient of the number of relevant documents found and the sum of relevant documents found and not found in the information system; precision (*P*) is the quotient of the number of relevant documents found and the number of all documents found including ballast. Early retrieval systems were small; so the relevant documents not found were known by the researchers (e.g., Kent et al., 1955; Cleverdon, 1967). Due to the "big systems' syndrome" (Stock, 2000) this is not the case for huge information services such as WoS, Scopus, or Google. It is practically impossible to check the relevance of all documents in such large databases. If we know about all publications of an institution, we can use the number of these publications as the gold standard for all relevant documents. We refer to the effectiveness measures concerning this gold standard with the tradition terms "recall" and "precision." Additionally, we searched for the institution's publications using different search strategies on the same information system and received another standard value with the union of all hits from the different searches. We call this variant of effectiveness measures "search recall" and "search precision," as their values depend on the search strategies.

How did we calculate recall and precision? In our exemplary study, we applied two methods of searching for institutions, namely (a) a search strategy working with the name of the institution *s-f-i* ("search by institution names") and (b) a strategy working with the names of the authors who worked at the institution during the observation period *s-f-an* ("search by author names"). Let the number of all publications of an institution be *g* ("gold"), the number of correctly found documents *f* and *f'* ("found"), and the number of irrelevant documents in the hit list *ba* and *ba'* ("ballast"). For an information service *IS* (for instance, WoS, Scopus, or Dimensions), the recall *R* for a search for an institution *S-I* applying both strategies is calculated by

$$R_{(IS;S-I)(s-f-i)} = f/g \text{ and } R_{(IS;S-I)(s-f-an)} = f'/g, \quad (2)$$

the corresponding precision *P* is calculated by

$$P_{(IS;S-I)(s-f-i)} = f/(f+ba) \text{and } P_{(IS;S-I)(s-f-an)} = f'/(f'+ba') \quad (3)$$

At this point our first research question (**RQ1**) arises:

**RQ1a.** Is $R_{(IS;S-I)(s-f-i)} = R_{(IS;S-I)(s-f-an)}$, $R_{(IS;S-I)(s-f-i)} > R_{(IS;S-I)(s-f-an)}$, or $R_{(IS;S-I)(s-f-i)} < R_{(IS;S-I)(s-f-an)}$? Is the recall of a search for institutions applying a search for institution names equal, higher, or lower than that of a search for institutions applying a search for the institution's author names?

**RQ1b.** Is $P_{(IS;S-I)(s-f-i)} = P_{(IS;S-I)(s-f-an)}$, $P_{(IS;S-I)(s-f-i)} > P_{(IS;S-I)(s-f-an)}$, or $P_{(IS;S-I)(s-f-i)} < P_{(IS;S-I)(s-f-an)}$? Is the precision of a search for institutions applying a search for institution names equal, higher, or lower than that of a search for institutions applying a search for the institution's author names?

### 3.2. Search recall and search precision

As additional informetric indicators, we calculate "search recall" (*SR*) and "search precision" (*SP*). We searched for the institution's publications in an information service *IS* in multiple ways (for author and institution names), combined all results, and defined this combined set as the database-specific ideal search result *g(search,IS)* for an additional calculation of recall and precision in order to show the differences between author and institution searches. The amount of found hits is *f(search,IS),* and *ba(search,IS)* is the ballast, all values in relation to exactly one information service *IS*. Search recall *SR* for a search on an information service *IS* for an institution

applying our two strategies is calculated by

$$\mathrm{SR}_{(\mathrm{IS;S-I})(\mathrm{s-f-i})} = \mathrm{f}(\mathrm{search},\mathrm{IS})/\mathrm{g}(\mathrm{search},\mathrm{IS}) \text{and } \mathrm{SR}_{(\mathrm{IS;S-I})(\mathrm{s-f-an})} = \mathrm{f}'(\mathrm{search},\mathrm{IS})/\mathrm{g}(\mathrm{search},\mathrm{IS}), \tag{4}$$

the corresponding search precision *SP* is calculated by

$$\mathrm{SP}_{(\mathrm{IS;S-I})(\mathrm{s-f-i})} = \mathrm{f}(\mathrm{search},\mathrm{IS})/[\mathrm{f}(\mathrm{search}) + \mathrm{ba}(\mathrm{search})] \text{ and } \mathrm{SP}_{(\mathrm{IS;S-I})(\mathrm{s-f-an})} = \mathrm{f}'(\mathrm{search},\mathrm{IS})/[\mathrm{f}'(\mathrm{search},\mathrm{IS}) + \mathrm{ba}'(\mathrm{search},\mathrm{IS})]. \tag{5}$$

**RQ1c and RQ1d** are analogous to RQ1a and RQ1b, respectively; the most important difference is the value of the standard, i.e., *g (search,IS)* instead of *g*.

### 3.3. Relative visibility

Measurements of recall and precision do not cover all crucial quality aspects of information systems (Schumann & Stock, 2014). A useful addition to these criteria is measuring the relative visibility of an institution in an information service. Dorsch (2017) defined the relative visibility of an author $RV_{(A)}$ in an information service *IS* as the quotient of the number $d_{(A,IS)}$ of an author *A*'s publications found in the information service *IS* and the number of all their publications $r_{(A)}$ as found using the gold standard:

$$\mathrm{RV}_{(\mathrm{A,IS})} = \mathrm{d}_{(\mathrm{A,IS})}/\mathrm{r}_{(\mathrm{A})}. \tag{6}$$

Since we work with fractional counting according to formula (1), we do not count the number of publications as in the calculation of recall, but rather the author's share of the publications, expressed as publication points. So we have to add up all the publication points of an author *A*, which results in $\mathrm{d'}_{(\mathrm{A,IS})}$. If *A* has co-affiliations for some of their publications, we only consider the share of the relevant institution. The total number $r'_{(A)}$ of *A*'s "real" publication points results from the gold standard. An author's relative visibility using fractional counting, $RV_{(A,IS,FC)}$, is calculated similarly to Dorsch's formula:

$$\mathrm{RV}_{(\mathrm{A,IS,FC})} = \mathrm{d}'_{(\mathrm{A,IS})}/\mathrm{r}'_{(\mathrm{A})} \tag{7}$$

In the case of the search for author names *s-f-an*, an institution's *Inst* relative visibility *RV* in an information service *IS* is the sum of publication points $d'_{(A,\ IS)}$ of all its affiliated authors divided by the "real" publication points of the institution $r'_{(Inst)}$ according to the gold standard:

$$\mathrm{RV}_{(\mathrm{Inst,IS,FC})(\mathrm{s-f-an})} = \left(\sum_{k=1}^{m} \mathrm{d}'(\mathrm{A(k)},\mathrm{IS})\right) \Big/ r'(Inst) \tag{8}$$

In the case of the search for institution names *s-f-i*, we divide the publication points found on an information service *IS* $d'_{(Inst,IS)}$ by the "real" publication points of the institution $r'_{(Inst)}$:

$$\mathrm{RV}_{(\mathrm{Inst,IS,FC})(\mathrm{s-f-i})} = \mathrm{d}'_{(\mathrm{Inst,IS})}/\mathrm{r}'_{(\mathrm{Inst})} \tag{9}$$

At this point our second research question (**RQ2**) arises:

**RQ2.** Is $\mathrm{RV}_{(\mathrm{Inst,IS,FC})(\mathrm{s\text{-}f\text{-}an})} = \mathrm{RV}_{(\mathrm{Inst,IS,FC})(\mathrm{s\text{-}f\text{-}i})}$, $\mathrm{RV}_{(\mathrm{Inst,IS,FC})(\mathrm{s\text{-}f\text{-}an})} > \mathrm{RV}_{(\mathrm{Inst,IS,FC})(\mathrm{s\text{-}f\text{-}i})}$, or $\mathrm{RV}_{(\mathrm{Inst,IS,FC})(\mathrm{s\text{-}f\text{-}an})} < \mathrm{RV}_{(\mathrm{Inst,IS,FC})(\mathrm{s\text{-}f\text{-}i})}$? Is the institution's relative visibility applying a search for author names equal, higher, or lower than that applying a search for institution names?

### 3.4. Differences between gold standard, author name search results, and institution name search results

If we are able to identify differences between our gold standard and the hits found as well as differences between the hit sets of the author and institution searches, an additional research question arises:

**RQ3.** What are the reasons for differences between the gold standard, the hit sets of an author search, and the hit sets of an institution search? How consistently is the institution's name used in information services?

## 4. Methods

In our illustrative example we analyze two information science institutions in German-speaking countries (Friedländer, 2014), the *Department of Information Science* at *Heinrich Heine University Düsseldorf* in Germany (Stock, 2023) and the *Institute for Information Science and Information Systems* at *Karl Franzens University Graz* in Austria (Reichmann et al., 2021; Reichmann & Schlögl, 2022), since 2020 part of the *Dept. of Operations and Information Systems*, for a period of ten years (2009 to 2018). The institution's researchers who are relevant to us are all faculty members including research assistants (*Wissenschaftliche Hilfskräfte*). If students declared their affiliation with the institution in their publications, such articles were also attributed to the institution if they were co-authored with a faculty member.

Since we were aware of the publications from individual publication lists (Düsseldorf) or from the university's existing research database (Graz), we defined this publication lists as our gold standard. Using formula (1), we calculated the authors' and institution's publication points for the gold standard. In the next step, we had to check whether all journal publications of all identified affiliated authors were included in the four WoS journal databases (SSCI, SCI, A&HCI, and ESSI) and carried out an *institution search* in a second

step. We also calculated the respective publication points as the results of fractional counting. Since 2008, the WoS has contained all affiliations of every single author of a publication (Clarivate, 2018, p. 35). The basis for the address data in the WoS is the affiliation data given in the journal (see Fig. 1, top).

As can be seen in Fig. 1 (middle), some transformations are performed by WoS. In particular, abbreviated institute and university names are used in this information service (e.g., "Inst Informat Sci & Informat Syst" instead of "Institute of Information Science and Information Systems" and "Graz Univ" instead of "University of Graz"). Since usually the German affiliation is specified in German-language journals, it follows that the institute name is usually different to the one given in English-language journals. As a consequence, both the German and English version of the institution name must be searched in WoS, if one wants to retrieve all documents of an institute.

In principle, there is an own field (SG) for searching for the sub-organization in WoS. The problem with this search is that it may retrieve also wrong results. For instance, the search statement

PY=(2009-2018) AND SG="Inst Informat Sci" AND OO="Karl Franzens Univ Graz" retrieves, among others, one record in which one author is from Karl Franzens University Graz (but from the Institute of Physics) and another author which is from an Institute of Information Science (& Education) in Japan (automatic internal truncation) (see Fig. 1, bottom). So, in order to ensure that the Institute of Information Science from Graz is found, we used the WoS address field (AD) in combination with the SAME operator, after which we specified the city:

PY=(2009-2018) AND AD=("Inst Informat Sci" SAME Graz).

This ensures that the information science institute is really from Graz.

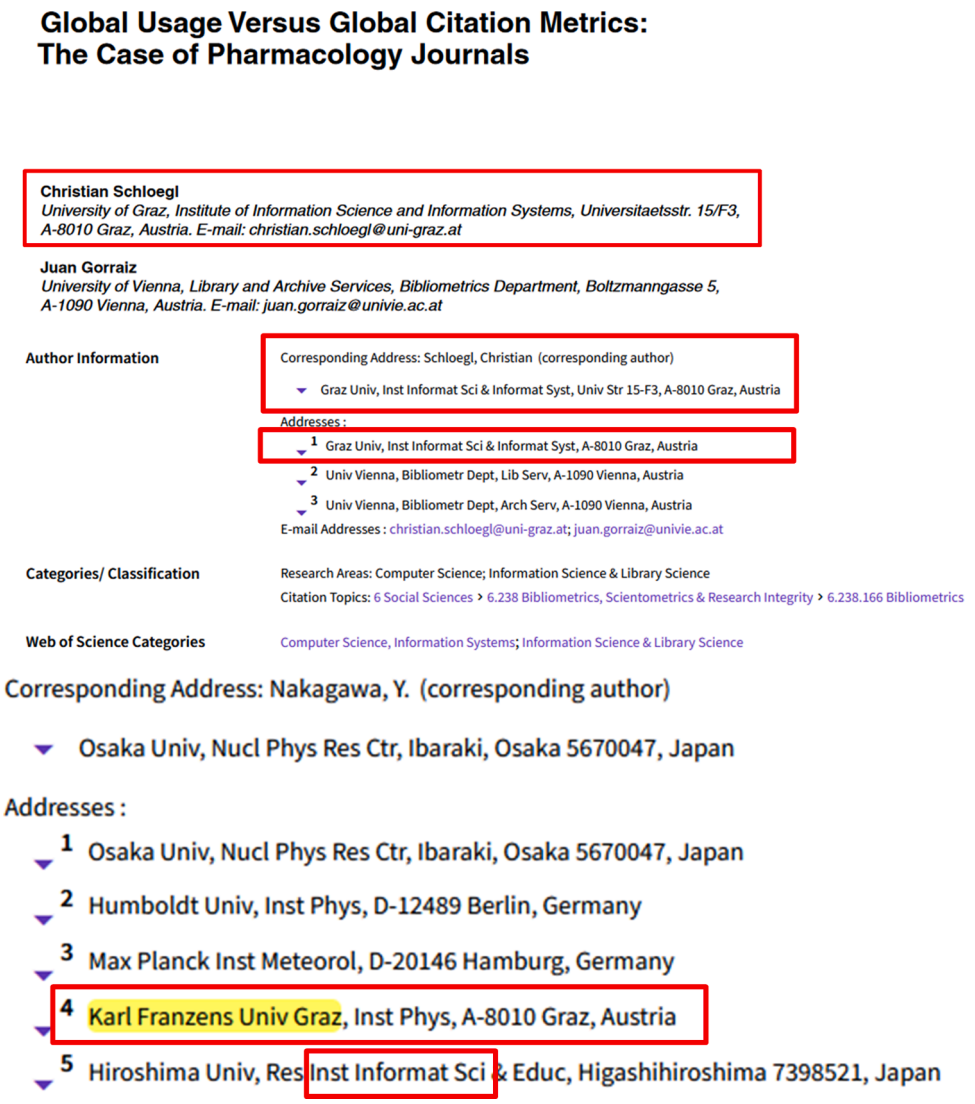


**Fig. 1.** Affiliation and address data given in a journal (top) and transformed address data in WoS (middle), and a misleading search in WoS (bottom). *Sources:* Journal of the American Society for Information Science and Technology, WoS.

In addition to the institution search, we also conducted an *author search*. In order to find all publications of an institution's research staff, we used search statements like the following: PY=(2009-2018) AND AU=Henkel M* AND CI=Dusseldorf. As can be seen, we only used the first initial of the first name followed by a truncation sign. Accordingly, there is the possibility not only to retrieve documents from "Henkel Maria" (being correct) but also from "Henkel Marius" (being false). To further narrow the author search we also included the city of the institution.

## 5. Results

All numerical results are included in Table 1. As is exhibited in Table 1.a, only a small part of the publications of the two institutes is included in Web of Science. Most of these publications were found in the author search (38.6 publications points or 44 publications for Düsseldorf and 16.7 publication points or 24 publications for Graz), a clearly lower part was retrieved in the institution search (33.8 publications points or 38 publications for Düsseldorf and 11.9 publication points or 16 publications for Graz). In the author search we only considered staff members and research assistants from the two institutes. Surprisingly, the institution search for Düsseldorf revealed 3 publications (2.75 publication points) which were only published by students (publications #5, #9 and #45 in Appendix A1) and which are therefore not included in the gold standard. For Graz it happened that one publication (0.33 publication points) was retrieved (#13 in Appendix A2) which was wrongly assigned to an author from Graz and therefore also not assigned to the gold standard. However, another publication which was originally not considered in the publication list but which was found in the institution search (#22 in Appendix A2) was added to the gold standard papers in WoS (and therefore also to the publication lists). The ideally searched hit set in WoS includes 46 publications from Düsseldorf (with our search strategy only found 44) and 25 from Graz (found 24).

### 5.1. Recall and precision and search recall and search precision (RQ 1)

Table 1.b indicates a low recall for the author search (Düsseldorf: 12.8 %, Graz: 10.5 %) which is even lower for the institution search (Düsseldorf: 11.0 %, Graz: 7.0 %). For the search recall (see Table 1.c) the values are clearly better than those based on the calculation of recall in relation to a gold standard; however, this is no wonder since we calibrated recall relatively to the merged hit sets on WoS (WoS-specific ideal hit set). The search recall of the author search is for both institutions more than 95 %, while it is only 82.6 % (Düsseldorf) and 64.0 % (Graz) for the institution search. This is in contrast to the precision of the institution search where we

**Table 1**
Numerical results.

Table 1.a Publication points (in brackets: publications) of our exemplary institutions.

| Institution | Total publication points (total publications) Gold standard, g | Publication points (publications) WOS-specific ideal hit set, $g_{(search,WoS)}$ | Publication points (publications) Author Search (WoS) | Publication points (publications) Institution Search (WoS) | Additional publication points (additional publications) (not gold standard) Institution Search (WoS)* |
|---|---|---|---|---|---|
| Düsseldorf | 276.9 (345) | 40.1 (46) | 38.6 (44) | 33.8 (38) | 2.75 (3) |
| Graz | 174.6 (229) | 17.2 (25) | 16.7 (24) | 11.9 (16) | 0.33 (1) |

**Table 1.b Recall and precision of author versus institution searches in WoS (based on publications).**

| Institution | Recall Author Search $R_{(WoS,S\text{-}I,s\text{-}f\text{-}an)}$ | Recall Institution Search $R_{(WoS,S\text{-}I,s\text{-}f\text{-}i)}$ | Precision Author Search $P_{(WoS,S\text{-}I,s\text{-}f\text{-}an)}$ | Precision Institution Search $P_{(WoS,S\text{-}I,s\text{-}f\text{-}i)}$ |
|---|---|---|---|---|
| Düsseldorf | 12.8 % | 11.0 % | 83.0 % | 100 % |
| Graz | 10.5 % | 7.0 % | 85.7 % | 100 % |

**Table 1.c Search recall and search precision of author versus institution searches in WoS (based on publications).**

| Institution | Search recall Author Search $SR_{(WoS,S\text{-}I,s\text{-}f\text{-}an)}$ | Search recall Institution Search $SR_{(WoS,S\text{-}I,s\text{-}f\text{-}i)}$ | Search precision Author Search $SP_{(WoS,S\text{-}I,s\text{-}f\text{-}an)}$ | Search precision Institution Search $SP_{(WoS,S\text{-}I,s\text{-}f\text{-}i)}$ |
|---|---|---|---|---|
| Düsseldorf | 95.7 % | 82.6 % | 83.0 % | 100 % |
| Graz | 96.0 % | 64.0 % | 85.7 % | 100 % |

**Table 1.d Relative visibility based on author and institution searches (based on publication points).**

| Institution | Relative visibility Author Search $RV_{(Inst,WoS,FC)\ (s\text{-}f\text{-}an)}$ | Relative visibility Institution Search $RV_{(Inst,WoS,FC)\ (s\text{-}f\text{-}i)}$ |
|---|---|---|
| Düsseldorf | 13.9 % | 12.2 % |
| Graz | 9.6 % | 6.8 % |

* Publications of non-faculty members, e.g., students, and publications which were assigned erroneously to an institute; excluded from further calculations

attained values of 100% for both institutes. These values are, of course, influenced by our search strategy. As mentioned in the method section, we used a broader approach for the author search. Accordingly, we found several homonymous author names (having the same last name and the same initial in the first name) (9 hits for Düsseldorf and 2 for Graz) where one of the co-authors worked at a university in Düsseldorf or Graz. In case of the institute from Graz, one staff member has left the institute within the publication period and was publishing for another research institution in Graz (which resulted in two more hits).

We learn from both calculations, i.e., gold-standard based recall and search recall, that the author search outperforms the institution search in WoS. However, this is more evident in the search recall, since the document base of an institute is smaller here.

### 5.2. Relative visibility (RQ 2)

As Table 1.d exhibits, only a relatively low part of the publications (in the form of publication points) of the two institutes is visible in WoS. This is due to two different reasons: (1) the sources of the institutions' articles are not covered by WoS CC and therefore not findable regardless of the search strategy, and, to a much lesser extent, (2) there were problems in the affiliation or address fields. The proportions for Düsseldorf are higher, which is in particular true for the institution search. This is also reflected in the recall which we analyzed before.

### 5.3. Reasons for differences between gold standard, author search results, and institution search results (RQ 3)

In this section, we want to explore in more detail the reasons for the in particular low recall and visibility values for the institution search. The main reason for weak recall values is the fact that the sources of the publications are not covered by WoS. As can be seen in Table 2.a (see also column "Name variants – inst." in Table A1 in the digital Appendix), there are eight different naming variants for the Department of Information Science in *Düsseldorf* which are due to three main reasons:

- inconsistent naming by WoS,
- inconsistent naming by authors of the papers,
- no institution names are given in the journal.

In Table 2, we exhibit in detail the found variants for the names of the two institutes und the universities separated by German-language and English-language publications.

## 6. Discussion

As our example has illustrated, the search for research institutions including university names is a critical issue. This is particularly true for institutions and universities outside Anglo-American countries also publishing in their mother tongue (and therefore giving their institution and university name in their mother tongue). In order to find their publications, one must also search for the institute and university names generated out of the native names (e.g., "Inst Informat Wissensch & Wirtschaftsinformat," "Karl Franzens Univ Graz").

The extent of inconsistent institute and university names was relatively high for the two institutes in our case study. The two main reasons are the inconsistent derivation of the institute or university name by WoS as well as the inconsistent use of the institute or university name by the authors or the publishing sources.

Concerning the first case, it happened that, for example, three different versions of the institution name ("Inst Informat Wissensch & Wirtschaftsinformat," "Inst Informations Wissensch Wirtschaftsinformat," and "Inst Informat Wissenschaft & Wirtschaftsinformat") were generated out of the affiliations given in the journal ("Institut für Informationswissenschaft und Wirtschaftsinformatik") by WoS. This was similar for the university name where, for instance, "Heinrich Heine University Düsseldorf" resulted in "Heinrich Heine Univ Dusseldorf" and "Univ Dusseldorf."

The second main sources for inconsistent naming are the authors themselves. Partly different name variants are given (e.g., "Inst Informat Sci & Informat Syst" vs. "Inst Informat Sci" vs. "Dept Informat Sci" or "Heinrich Heine Univ Dusseldorf" vs. "Heinrich Heine Univ"). However, it also happens that no institution name is mentioned in the publishing journal or that the name of the research center which is associated with the institute is given instead. In a few cases a journal can be the cause why the publications of an institute cannot be found. In our exemplary study, the affiliation was missing twice in the journal. In one case this was an editorial, in the other a book review. There may be additional reasons for *not* disclosing institutional affiliations (e.g., an author leaves an institution due to problems with colleagues or the management). It would be interesting to learn more about those reasons for concealing affiliations in a publication. This could be the subject of future research.

## 7. Conclusion

Since searching for institution names produces different results than searching for author names and there is also the problem of missing as well as corrupted name variants and missing publications in information services, we have to state issues and problems when investigating the research output of institutions.

The following conclusion can be drawn for searching for institutions. A search for the institution name definitely makes sense since also publications are retrieved which would not have been identified otherwise (e.g., which are even not included in the authors'

**Table 2**
Different naming variants for the institutions' and universities' names in WoS.

**Table 2.a Different naming variants for the institute name (Düsseldorf).**

| Name variant* | Institute name according to WoS | found pubs | % German/Engl. pubs |
|---|---|---|---|
| **German-language publications** | | | |
| **Inconsistent naming by WoS** | | | |
| I1 | Abt Informat Wissensch | 18 | 75% |
| I2 | Abt Informationswissenschaft | 2 | 8% |
| **Inconsistent naming by authors** | | | |
| I3 | Name of higher-level org. unit specified | 1 | 4% |
| I4 | Informat Wissensch & Sprachtechnol | 1 | 4% |
| **No naming in journal** | | | |
| I5 | no affiliation was given in journal (editorial and book review) | 2 | 8% |
| | | 24 | 100% |
| **English-language publications** | | | |
| **Inconsistent naming by authors** | | | |
| I6 | Dept Informat Sci | 23 | 92% |
| I7 | Informat Sci Dept | 1 | 4% |
| I8 | only university name was given | 1 | 4% |
| | | 25 | 100% |

**Table 2.b Different naming variants for the university name (Düsseldorf).**

| **Name variant*** | **University name according to WoS** | **found pubs** | **% German/Engl. pubs** |
|---|---|---|---|
| **German language publications** | | | |
| **Inconsistent naming by WoS** | | | |
| U1 | Heinrich Heine Univ Dusseldorf | 19 | 79% |
| U2 | Univ Dusseldorf (generated by WoS) | 2 | 8% |
| **Inconsistent naming by authors** | | | |
| U3 | Henrich Heine Univ Dusseldorf | 1 | 4% |
| **No naming in journal** | | | |
| U4 | no university name was given | 2 | 8% |
| | | 24 | 100% |
| **English-language publications** | | | |
| **Inconsistent naming by WoS** | | | |
| U5 | Heinrich Heine Univ Dusseldorf | 9 | 36% |
| U6 | Univ Dusseldorf (generated by WoS) | 13 | 52% |
| U7 | Heinrich Heine Univ | 1 | 4% |
| **Inconsistent naming by authors** | | | |
| U8 | Heinrich Heine Univ (without Dusseldorf) | 2 | 8% |
| | | 25 | 100% |

**Table 2.c. Different naming variants for the institute name (Graz).**

| **Name variant*** | **Institute name according to WoS** | **found pubs** | **% German/Engl. pubs** |
|---|---|---|---|
| **German-language publications** | | | |
| **Inconsistent naming by WoS** | | | |
| I1 | Inst Informat Wissensch & Wirtschaftsinformat | 9 | 69% |
| I2 | Inst Informations Wissensch Wirtschaftsinformat | 1 | 8% |
| I3 | Inst Informat Wissenschaft & Wirtschaftsinformat | 1 | 8% |
| **Inconsistent naming by authors** | | | |
| I4 | Inst Informationswissenschaft | 1 | 8% |
| **No naming in journal** | | | |
| I5 | No affiliation in journal given (Editorial) | 1 | 8% |
| | | 13 | 100,0% |
| **English-language publications** | | | |
| **Inconsistent naming by authors** | | | |
| I6 | Inst Informat Sci & Informat Syst | 7 | 54% |
| I7 | Inst Informat Sci | 1 | 8% |
| I8 | Dept Informat Sci | 3 | 23% |
| I9 | Linked research center named Ctr Digital Commun (Graz Univ) | 1 | 8% |
| I10 | only university name was given by authors | 1 | 8% |
| | | 13 | 100% |

**Table 2.d Different naming variants for the university name (Graz).**

| **Name variant*** | **University name according to WoS** | **found pubs** | **% German/Engl. pubs** |
|---|---|---|---|
| **German-language publications** | | | |
| **Inconsistent naming by authors** | | | |
| U1 | Karl Franzens Univ Graz | 8 | 62% |
| U2 | Karl Franzen Univ Graz | 1 | 8% |
| U3 | Graz Univ ("Universität Graz") | 3 | 23% |

(*continued on next page*)

**Table 2** (*continued*)

| | | | |
|---|---|---|---|
| **No name in journal** | | | |
| U4 | no affiliation in journal (Editorial) | 1 | 8% |
| | | 13 | 100% |
| **English-language publications** | | | |
| **Inconsistent naming by authors** | | | |
| U5 | Karl Franzens Univ Graz | 3 | 23% |
| U6 | Graz Univ („University of Graz“) | 10 | 77% |
| | | 13 | 100% |

* In the digital Appendices A1 and A2 it is indicated which variant of the institute as well as university name can be found in which publication.

publication lists). In order to increase the recall, it is recommended to consider different naming variants of the institute name. In order to raise the precision when searching for author names, it is suggested to narrow the search statement. Instead of searching for the city of the institution, the university name should be included in the search. However, to ensure a high recall, different naming variants of the university name should be regarded (… AND OO=(Karl Franzens Univ Graz OR Graz Univ)).

If an institution maintains a complete data set of all of its publications in the sense of a “causative principle” (Fink, 1991; Stock, 1991) considered as authoritative and if it is openly accessible, it can be used beneficially for studies in evaluative scientometrics. However, it is a challenging to complete such publication lists with citation data from multidisciplinary information services such as WoS, Scopus, Semantic Scholar, or Dimensions.

We found another way to merge data in order to obtain at a representative set of an institution’s publications. Many personal publication lists can be found on the Web but they are not available in a standardized format. As such publication lists are open data this could be a task for future research on linked open data to merge the variations and create a centralized data set (Dorsch, 2017). However, we have to know who worked at the institution during the evaluation period. It would be very helpful to combine ORCID numbers (describing the individual researchers) with the ROR or other identifiers (for the institutions) with the assignment of the employment period, which is already realized for many ORCID entries at the university level (additionally including some DOIs of publications) (Haak et al., 2018). But ORCID does not cover sub-levels of universities such as, for example, faculties, institutes, and departments.

In order to see the quality differences between institution and author search more clearly, we have introduced *search recall* and *search precision* indicators. Since they consider the document base of the used information service only, they show the differences in the resulting recall and precision values more clearly. We used both indicators for the numerical description of the difference between searching for institution names and searching for names of their staff. Of course, search recall and search precision are valid indicators for all different search strategies on the same topic at the same information service.

For all our calculations, there is a clear result: A search for affiliated authors outperforms a search for institutions in WoS. However, both search strategies call for deep background knowledge on the institutions and their staff and also in-depth knowledge of advanced search strategies. It takes several research steps until you find the optimal search strategy.

Furthermore, it must be noted that despite a “perfect” search in the WoS, usually only a (often relatively small) proportion of an institution’s publications can be found, as numerous publications are not included in the WoS CC. There are many publications from Düsseldorf and Graz covered in the Proceedings Citation Index and in the Book Citation Index (Stock, 2021). However, these sub-databases are not subscribed by our university libraries. An additional reason for the low coverage in WoS is that the journals in which many articles were published are not included in the WoS. This is particularly true for non-English language journals. This problem was also evident in our example, which referred to two institutions in non-English-speaking countries. The relative visibility in the WoS for Düsseldorf was only 13.9% (12.2%) in the case of an author (institution) search. The corresponding values for Graz with significantly more German-language publications were only 9.6% and 6.8%, respectively. The best way to solve this problem is to use individual publication lists of all researchers, i.e., authors, working at an institution, which is, however, very resource-intensive. A good alternative might be the use of university research databases, provided they exist and are (reasonably) complete. In our example, this was the case for Graz.

With regard to our illustrative example, the following limitations apply: (1) The example only referred to two institutions, which can lead to biases. It should be noted in particular that both institutions are located in German-speaking countries which is why their relative visibility is below-average due to a correspondingly high number of non-English-language publications that are not included in WoS CC. However, it should be borne in mind that expanding the examples would be very resource-intensive and would probably provide little added value in terms of answering the research questions addressed here. In addition, the search and analysis carried out in our case study requires detailed knowledge of the institutions examined, which probably only staff members of the institutions can have. This was the case since the authors of this article are (former) staff members of the two institutions examined. (2) The gold standard we have defined, i.e., the recording of all publications of an institution, can probably never be completely met with certainty, even in the case of extensive retrieval and even when using individual publication lists. We also found–more or less by chance–publications outside of those in the individual publication lists of all authors we considered relevant. Reasons for this could be, for example, incomplete individual publication lists and incomplete or incorrect staff member information. (3) Concerning WoS data all possible name variants (including incorrect entries) with regard to authors and institutions can hardly be taken into account as they are unknown for the searcher. This also applies to our illustrative study. Therefore, realistically speaking, our calculated values for recall, precision, and relative visibility only represent approximate values.

Among other indicators, university rankings are based on data on institutions. The Leiden Ranking, for instance, uses data on publication output, citation impact, and scientific collaboration; all data are retrieved from WoS (Waltman et al., 2012). If one looks at such rankings in the light of our results, doubts about the reliability of such lists seem entirely justified. Using the example of the Shanghai ranking ("Academic Ranking of World Universities"), van Raan (2005, p. 140) calls such university rankings "quasi-evaluations" being "absolutely unacceptable". That is probably still true today.

What are the implications of our study for the informetric research practice? Since the data from only one information service are more or less incomplete and, additionally, personal or institutional publications lists are also not (or not always) complete, scientometricians could merge the publication data based on important information services (at least, from WoS, Scopus, Dimensions, Semantic Scholar, and Google Scholar) (Delgado-Quirós & Ortega, 2025) with the personal or institutional publication lists of the affiliated authors and their institutions into a single data sheet. We know from our study that the search for affiliated researchers outperforms the search for the institution name. Therefore, in a first step, we have to identify all affiliated researchers of an institution in the defined time interval before we search for the names found in a second step. However, this would mean a very high work load to identify all research staff in an institution with many researchers, e.g., in an entire university. Due to data protection laws and—especially in private companies—lists of staff that may be kept secret this task cannot always be successfully managed. Although we cannot guarantee completeness with our procedure, we can certainly guarantee a high level of quality with regard to the empirical basis for describing and evaluating a research institution. This means a very high level of effort in data collection and processing. If it is not possible to manage such a large amount of effort, one should refrain from describing the performance of research institutions in purely quantitative terms, otherwise the following applies: "Research performance values are not certain" (Abramo et al., 2015, p. 954). Numbers alone "cannot fully capture the value and importance of scientific research" (Torres-Salinas et al., 2024, p. 6) nor that of researchers and research institutions. Perhaps, "narrative bibliometrics" (Torres-Salinas et al., 2024) can help: Here, the combination of background knowledge about the institution, the analysis and contextualization of indicators, the completeness of applied information services, the search strategies on the information services, interviews with institution's members, other sources (such as newspaper articles or social media posts), and—of course—metrics-based research assessments of the research output generates acceptable and appropriate "stories and narratives" (Torres-Salinas et al., 2024, p. 6) about the research institution under investigation.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declaration of generative AI and AI-assisted technologies in the writing process

Not used.

## CRediT authorship contribution statement

**Wolfgang G. Stock:** Writing – review & editing, Writing – original draft, Project administration, Formal analysis, Conceptualization. **Gerhard Reichmann:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Christian Schlögl:** Writing – original draft, Methodology, Investigation, Data curation, Conceptualization.

## Declaration of competing interest

The authors have no relevant financial or non-financial interests to disclose.

## Acknowledgement

The authors like to thank the peer reviewers for helpful comments and hints.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.joi.2025.101638.

## References

Abramo, G., & D'Angelo, C. A (2015). Evaluating university research: Same performance indicator, different rankings. *Journal of Informetrics, 9*(3), 514–525. https://doi.org/10.1016/j.joi.2015.04.002

Abramo, G., D'Angelo, C. A., & Grilli, L. (2015). Funnel plots for visualizing uncertainty in the research performance of institutions. *Journal of Informetrics, 9*(4), 954–961. https://doi.org/10.1016/j.joi.2015.08.006

Albrecht, K., Frost, M., & Handtke, U. (1994). Informetrische Vermessung eines Forschungsinstitut. In W. Rauch, F. Strohmeier, H. Hiller, & C. Schlögl (Eds.), *Mehrwert von Information – Professionalisierung der Informationsarbeit* (pp. 151–163). Universitätsverlag Konstanz. https://epub.uni-regensburg.de/10895/1/Mehrwert_von_Information_-_Professionalisierung_der_Informationsarbeit.pdf.

Aman, V., & van den Besselaar, P. (2024). Authorship regulations in performance-based funding systems and publication behaviour – A case study of German medical faculties. *Journal of Informetrics, 18*(2), Article 101500. https://doi.org/10.1016/j.joi.2024.101500

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies, 1*(1), 377–386. https://doi.org/10.1162/qss_a_00019

Bachelet, V. C., Uribe, F. A., Diaz, R. A., Vergara, A. F., Bravo-Cordova, F., Carrasco, V. A., Lizana, F. J., Meza-Ducaud, N., & Navarrete, M. S. (2019a). Author misrepresentation of institutional affiliations: Protocol for an exploratory case study. *BMJ Open, 9*(2), Article e023983. https://doi.org/10.1136/bmjopen-2018-023983

Bachelet, V. C., Uribe, F. A., Diaz, R. A., Vergara, A. F., Bravo-Cordova, F., Carrasco, V. A., Lizana, F. J., Meza-Ducaud, N., & Navarrete, M. S. (2019b). Misrepresentation of institutional affiliations: The results from an exploratory case study of Chilean authors. *Learned Publishing, 32*(4), 335–344. https://doi.org/10.1002/leap.1257

Bador, P., & Lafouge, T. (2005). Rédaction des adresses sur les publications: Un manque de rigueur défavorable aux universités françaises dans les classements internationaux. *La Presse Médicale, 34*(9), 633–636. https://doi.org/10.1016/s0755-4982(05)84000-x

Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies, 1*(1), 363–375. https://doi.org/10.1162/qss_a_00018

Bornmann, L., Gralka, S., de Moya Anegón, F., & Wohlrabe, K. (2023). Efficiency of universities and research-focused institutions worldwide: The introduction of a new input indicator reflecting institutional staff numbers. *Journal of Informetrics, 17*(2), Article 101400. https://doi.org/10.1016/j.joi.2023.101400

Cappelletti-Montano, B., Columbu, S., Montaldo, S., & Musio, M. (2022). Interpreting the outcomes of research assessments: A geometrical approach. *Journal of Informetrics, 16*(1), Article 101254. https://doi.org/10.1016/j.joi.2022.101254

Clarivate (2018). *Web of Science Core Collection. Descriptive document (version 2.0)*. http://library.khsu.ru/researchers/files/wos/Web_of_Science_Core_Collection_long_guide_public_OCT-18.pdf (Apr. 24, 2024).

Cleverdon, C. (1967). The Cranfield tests on index language devices. *Aslib Proceedings, 19*(6), 173–192. https://doi.org/10.1108/eb050097

De Bruine, R. E., Kint, A., Luwel, M., & Moed, H. F. (1993). A study of research evaluation and planning: The University of Ghent. *Research Evaluation, 3*(1), 25–41. https://doi.org/10.1093/rev/3.1.25

De Filippo, D., & Gorraiz, J. (2020). Is the Emerging Source Citation Index an aid to assess the citation impact in social science and humanities? *Journal of Informetrics, 14*(4), Article 101088. https://doi.org/10.1016/j.joi.2020.101088

Delgado López-Cózar, E., Orduña-Malea, E., & Martín-Martín, A. (2019). Google Scholar as a data source for research assessment. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 95–127). Springer. https://doi.org/10.1007/978-3-030-02511-3_4.

Delgado-Quirós, L., & Ortega, J. L. (2025). Citation counts and inclusion of references in seven free-access scholarly databases: A comparative analysis. *Journal of Informetrics, 19*(1), Article 101618. https://doi.org/10.1016/j.joi.2024.101618

Di Leo, J. R. (2003). Understanding affiliations. In J. R. Di Leo (Ed.), *Affiliations. Identity in academic culture*. University of Nebraska Press.

Dimzov, S., Matošić, M., & Urem, I. (2021). University rankings and institutional affiliations: Role of academic librarians. *The Journal of Academic Librarianship, 47*(5), Article 102387. https://doi.org/10.1016/j.acalib.2021.102387

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), 150–170. https://doi.org/10.1162/qss_a_00013

Dorsch, I. (2017). Relative visibility of authors' publications in different information services. *Scientometrics, 112*(2), 917–925. https://doi.org/10.1007/s11192-017-2416-9

Dorsch, I., Askeridis, J., & Stock, W. G. (2018). Truebounded, overbounded, or underbounded? Scientists' personal publication lists versus lists generated through bibliographic information services. *Publications, 6*(1), 1–9. https://doi.org/10.3390/publications6010007

Egghe, L., & Rousseau, R. (1990). *Introduction to informetrics. Quantitative methods in library, documentation and information science*. Elsevier. http://eprints.rclis.org/6011/.

Fiala, D. (2013). Suborganizations of institutions in Library and Information Science journals. *Information, 4*(4), 351–366. https://doi.org/10.3390/info4040351

Fiala, D. (2014). Sub-organizations of institutions in computer science journals at the turn of the century. *Malaysian Journal of Library & Information Science, 19*(2), 53–68. https://jice.um.edu.my/index.php/MJLIS/article/view/1789/2545.

Fink, S. (1991). Ein Jahr ifo Literaturdatenbank bei GENIOS. *Bibliotheksdienst, 25*(9), 1417–1418. https://doi.org/10.1515/bd.1991.25.9.1417

Friedländer, M. B. (2014). Informationswissenschaft an deutschsprachigen Universitäten: Eine komparative informetrische Analyse. *Information – Wissenschaft und Praxis, 65*(2), 109–119. https://doi.org/10.1515/iwp-2014-0018

García-Zorita, C., Martín-Moreno, C., Lascurain-Sánchez, M. L., & Sanz-Casado, E. (2006). Institutional addresses in the Web of Science: The effects on scientific evaluation. *Journal of Information Science, 32*(4), 378–383. https://doi.org/10.1177/0165551506065813

Garfield, E., & Stock, W. G. (2002). Citation consciousness. Interview with Eugene Garfield. *Password*, (6), 22–25. https://garfield.library.upenn.edu/papers/passwordinterview062002.pdf.

Gauffriau, M. (2021). Counting methods introduced into the bibliometric research literature 1970–2018: A review. *Quantitative Science Studies, 2*(3), 932–975. https://doi.org/10.1162/qss_a_00141

Gauffriau, M., Larsen, P. O., Maye, I., Roulin-Perriard, A., & von Ins, M. (2007). Publication, cooperation, and productivity measures in scientific research. *Scientometrics, 73*(2), 175–214. https://doi.org/10.1007/s11192-007-1800-2

Gusenbauer, M. (2019). Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics, 118*(1), 177–214. https://doi.org/10.1007/s11192-018-2958-5

Haak, L. L., Meadows, A., & Brown, J. (2018). Using ORCID, DOI, and other open identifiers in research evaluation. *Frontiers in Research Metrics and Analysis, 3*, 28. https://doi.org/10.3389/frma.2018.00028

Hardeman, S. (2013). Organization level research in scientometrics: A plea for an explicit pragmatic approach. *Scientometrics, 94*(3), 1175–1194. https://doi.org/10.1007/s11192-012-0806-6

Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies, 1*(1), 387–395. https://doi.org/10.1162/qss_a_00020

Hilbert, F., Barth, J., Gremm, J., Gros, D., Haiter, J., Henkel, M., Reinhardt, W., & Stock, W. G. (2015). Coverage of academic citation databases compared with coverage of social media: Personal publication lists as calibration parameters. *Online Information Review, 39*(2), 255–264. https://doi.org/10.1108/OIR-07-2014-0159

Hood, W. W., & Wilson, C. S. (2003). Informetric studies using databases: Opportunities and challenges. *Scientometrics, 58*(3), 587–608. https://doi.org/10.1023/B:SCIE.0000006882.47115.c6

Hottenrott, H., & Lawson, C. (2022). What is behind multiple institutional affiliations in academia? *Science and Public Policy, 49*(3), 382–402. https://doi.org/10.1093/scipol/scab086

Huang, S., Yang, B., Yan, S., & Rousseau, R. (2014). Institution name disambiguation for research assessment. *Scientometrics, 99*(3), 823–838. https://doi.org/10.1007/s11192-013-1214-2

Kent, A., Berry, M., Luehrs, F. U., Jr., & Perry, J. W (1955). Machine literature searching. VIII: Operational criteria for designing information retrieval systems. *American Documentation, 6*(2), 93–101. https://doi.org/10.1002/asi.5090060209

Khalifa, A. A., Hussien, S. M., Ansary, E. M., & El-Gharably, A. A. (2023). Different reporting patterns of author affiliations: A cross-sectional evaluation of publications from an Egyptian medical academic Institute. *Turkish Medical Student Journal, 10*(1), 13–18. https://doi.org/10.4274/tmsj.galenos.2023.2022-5-3

Korytkowski, P., & Kulczycki, E. (2019). Publication counting methods for a national research evaluation exercise. *Journal of Informetrics, 13*(3), 804–816. https://doi.org/10.1016/j.joi.2019.07.001

Lammey, R. (2020). Solutions for identification problems: A look at the Research Organization Registry. *Science Editing, 7*(1), 65–69. https://doi.org/10.6087/kcse.192

Liu, W., Ding, Y., & Gu, M. (2017). Book reviews in academic journals: Patterns and dynamics. *Scientometrics, 110*(1), 355–364. https://doi.org/10.1007/s11192-016-2172-2

Liu, W., Hu, G., & Tang, L. (2018). Missing author address information in Web of Science. An explorative study. *Journal of Informetrics, 23*(3), 985–997. https://doi.org/10.1016/j.joi.2018.07.008

Moed, H. F., Bar-Ilan, J., & Halevi, G. (2016). A new methodology for comparing Google Scholar and Scopus. *Journal of Informetrics, 10*(2), 533–551. https://doi.org/10.1016/j.joi.2016.04.017

Moed, H. F., Burger, W. J. M., Frankfort, J. G., & van Raan, A. F. J. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy, 14*, 131–149. https://doi.org/10.1016/0048-7333(85)90012-5

Orduña-Malea, E., Ayllón, J. M., Martín-Martín, A., & Delgado López-Cózar, E. (2017). The lost academic home: Institutional affiliation links in Google Scholar Citations. *Online Information Review, 41*(6), 762–781. https://doi.org/10.1108/OIR-10-2016-0302

Pereira, F. A., & Mugnaini, R. (2023). Mapping the use of Google Scholar in evaluative bibliometric or scientometric studies: A bibliometric review. *Quantitative Science Studies, 4*(2), 233–245. https://doi.org/10.1162/qss_a_00231

Polonioli, A. (2024). The case for affiliation contribution statements. *Accountability in Research. Ethics, Integrity and Policy, 31*(4), 377–383. https://doi.org/10.1080/08989621.2022.2130775

Pranckuté, R. (2021). Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications, 9*(1), 1–59. https://doi.org/10.3390/publications9010012, 12.

Purnell, P. J. (2022). The prevalence and impact of university affiliation discrepancies between four bibliographic databases. Scopus, Web of Science, Dimensions, and Microsoft Academic. *Quantitative Science Studies, 13*(1), 99–121. https://doi.org/10.1162/qss_a_00175

Pylarinou, S., & Kapidakis, S. (2017). Tracking scholarly publishing of hospitals using MEDLINE, Scopus, WoS and Google Scholar. *Journal of Hospital Librarianship, 17*(3), 209–216. https://doi.org/10.1080/15323269.2017.1332934

Reichmann, G., & Schlögl, C. (2022). On the possibilities of presenting the research performance of an institute over a long period of time: The case of the Institute of Information Science at the University of Graz in Austria. *Scientometrics, 127*(6), 3193–3223. https://doi.org/10.1007/s11192-022-04377-8

Reichmann, G., Schlögl, C., & Thalmann, S. (2021). Das Institut für Informationswissenschaft an der Universität Graz: 1987 –2020. *Information – Wissenschaft und Praxis, 72*(1), 1–9. https://doi.org/10.1515/iwp-2020-2132

Rousseau, R., Egghe, L., & Guns, R. (2018). *Becoming metric-wise: A bibliometric guide for researchers*. Chandos. https://doi.org/10.1016/C2017-0-01828-1

Russell, J. M., & Rousseau, R. (2009). Bibliometrics and institutional evaluation. In R. Arvanities (Ed.), *Science and technology policy, II* (pp. 42–65). Eolss. https://www.eolss.net/sample-chapters/c15/e1-30-04-04.pdf.

Schmidt, C. M., Cox, R., Fial, A. V., Hartman, T. L., & Magee, M. L. (2016). Gaps in affiliation indexing in Scopus and PubMed. *Journal of the Medical Library Association, 104*(2), 138–142. https://doi.org/10.3163/1536-5050.104.2.008

Schumann, L., & Stock, W. G. (2014). The Information Service Evaluation (ISE) model. *Webology, 11*(1), 1–20, 115 http://www.webology.org/2014/v11n1/a115.pdf.

Stefaniak, B. (1987). Use of bibliographic data bases for scientometric studies. *Scientometrics, 12*(3-4), 149–161. https://doi.org/10.1007/BF02016289

Stock, W. G. (1991). Die Ifo-Literaturdatenbank: eine volkswirtschaftliche Online-Datenbank nach dem „Verursacherprinzip". *ABI-Technik, 11*, 311–316.

Stock, W. G. (2000). Automatische und intellektuelle Indexierung. Das „Big-Systems'-Syndrom". Eine Antwort auf Robert Fugmann. *Password*, (1), 18–20.

Stock, W. G. (2021). Die Teildatenbanken von Web of Science und deren Subskriptionen von Wissenschaftseinrichtungen: Auswirkungen auf Forschungsevaluation und Sichtbarkeit von Forschenden. *Zeitschrift für Bibliothekswesen und Bibliographie, 68*(3), 152–157. https://doi.org/10.3196/186429502068331

Stock, W. G. (2023). Informationswissenschaftliche Forschung in Düsseldorf (1968 –2023). In A. Imeri, K. Scheibe, & F. Zimmer (Eds.), *Informationswissenschaft im Wandel. Wissenschaftliche Tagung 2022 (IWWT22)* (pp. 18–73). Hülsbusch. https://doi.org/10.5281/zenodo.7456998.

Stock, W. G., Dorsch, I., Reichmann, G., & Schlögl, C. (2023a). Labor productivity, labor impact, and co-authorship of research institutions: Publications and citations per full-time equivalents. *Scientometrics, 128*(1), 363–377. https://doi.org/10.1007/s11192-022-04582-5

Stock, W. G., Dorsch, I., Reichmann, G., & Schlögl, C. (2023b). Counting research publications, citations, and topics: A critical assessment of the empirical basis of scientometrics and research evaluation. *Journal of Information Science Theory and Practice, 11*(2), 37–66. https://doi.org/10.1633/JISTaP.2023.11.2.4

Stock, W. G., & Stock, M. (2013). *Handbook of information science*. De Gruyter Saur. https://doi.org/10.1515/9783110235005

Torres-Salinas, D., Orduña-Malea, E., Degado-Vázquez, Á., Gorrais, J., & Arroyo-Machado, W. (2024). Foundations of narrative bibliometrics. *Journal of Informetrics, 18*(3), Article 101546. https://doi.org/10.1016/j.joi.2024.101546

Uni Düsseldorf. (2023). Publikationsrichtlinie der Heinrich-Heine-Universität Düsseldorf vom 07.11.2023. *Amtliche Bekanntmachungen der Heinrich-Heine-Universität Düsseldorf*, (34), 2–6. https://www.hhu.de/fileadmin/redaktion/ZUV/Justitiariat/Amtliche_Bekanntmachungen/2023/2023_11_07_AB_34.pdf.

van Bellen, S. (2023). 'O author, where are thou?' An analysis of affiliation indexing in Canadian journals and bibliometric research potential. In *51st Annual Conference of the Canadian Association for Information Science /L'Association Canadiénne des Sciences de l'Information* (p. 6). https://cais2023.ca/talk/04.bellen/04.Bellen.pdf.

van Raan, A. F. J. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics, 62*(1), 133–143. https://doi.org/10.1007/s11192-005-0008-6

Venets, V. I. (2014). Some problems associated with affiliation of the authors in the Web of Science. *Journal of Communications Technology and Electronics, 59*(6), 681–687. https://doi.org/10.1134/S1064226914060205

Venets, V. I. (2017). Some problems connected with affiliation search in Scopus and a method for their solution. I. Correction of the institution profile. *Journal of Communications Technology and Electronics, 62*(6), 713–717. https://doi.org/10.1134/S1064226917060249

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., van Eck, N. J., van Leeuwen, T. N., van Raan, A. F. J., Visser, M. S., & Wouters, P. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology, 63*(12), 2419–2432. https://doi.org/10.1002/asi.22708

Winterhager, M., Schwechheimer, H., & Rimmert, C. (2014). Institutionenkodierung als Grundlage für bibliometrische Indikatoren. *Bibliometrie – Praxis und Forschung, 3*, 1–22. https://doi.org/10.5283/bpf.209