https://doi.org/10.51408/issi2025_154

# Toward Responsible Scientometrics: Normative Data Practices for Research Evaluation

Ying Huang[1M], Weishu Liu[2], Huizhen Fu[3], Jing Ma[4], Guijie Zhang[5], Yi Bu[6], Chao Min[7], Zhixiang Wu[8]

[1] *ying.huang@whu.edu.cn*
School of Information Management, Wuhan University, Wuhan 430072 (China)

[2] *Weishuliu.@zufe.edu.cn*
School of Information Management and Artificial Intelligence, Zhejiang University of Finance and Economics, Hangzhou 310018 (China)

[3] *fuhuizhen@zju.edu.cn*
School of Public Affairs, Zhejiang University, Hangzhou 310012 (China)

[4] *majing@xidian.edu.cn*
School of Economics and Management, Xidian University, Xi'an 710126 (China)

[5] *zgjzxmtx@163.com*
School of Management Science and Engineering, Shandong University of Finance and Economics, Jinan 250014 (China)

[6] *buyi@pku.edu.cn*
Department of Information Management, Peking University, Beijing 100084 (China)

[7] *mc@nju.edu.cn*
School of Information Management, Nanjing University, Nanjing 210023 (China)

[8] *cnwzx2012@njtech.edu.cn*
School of Economics and Management, Nanjing Tech University, Nanjing 211800 (China)

## Abstract

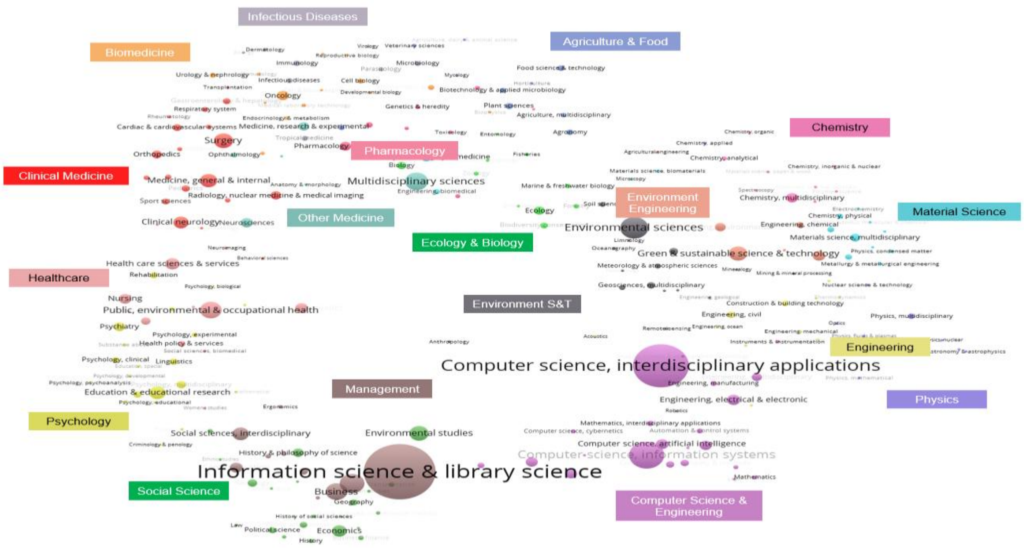
Conducting scientometric research in a rigorously normative manner requires careful attention to responsible metrics and data practices throughout the research life cycle. This study offers practical suggestions from a data perspective, addressing the key stages from source selection to data sharing while emphasizing transparency and policy implications for evaluation systems. First, the selection of data sources necessitates the alignment of data quality, inclusion criteria, and research objectives. Researchers must ensure that dataset characteristics align with their analytical goals and accurately document sources, including aspects such as coverage and granularity. Data retrieval and acquisition require a clear understanding of the database's scope and update mechanisms. A deliberate approach—utilizing tailored search terms, following platform-specific syntax, and transparently reporting retrieval strategies—ensures reproducibility. Data pre-processing involves a clear comprehension of the field and specific tasks (such as deduplication and normalization). Standardized protocols should connect raw data to analytical requirements while maintaining transparency in methodological choices. For analysis, selecting appropriate tools and differentiating standardization methods (e.g., citation versus collaboration network metrics) is critical to support responsible metrics in evaluation systems. Visualization should enhance interpretability without distorting evidence. Interpretation must strike a balance between robustness and restraint, avoiding over-claiming patterns or neglecting contextual nuances. Engaging domain experts helps mitigate disciplinary biases. Data

storage requires the systematic documentation of fields, derived variables, and backup protocols to ensure auditability—a key aspect of transparent and accountable research practices. Finally, obligations for sharing and citations include disseminating data through repositories and formally crediting open-access datasets to maintain scholarly standards. By integrating these guidelines, researchers can enhance scientometrics' methodological rigor, replicability, and ethical accountability, contributing to more responsible and policy-relevant evaluation systems. This framework highlights how normative advancements depend on meticulous data stewardship at each procedural stage, with broader policy implications for research assessment.

## Introduction

Scientometrics is an interdisciplinary discipline that uses quantitative methods to quantify all aspects of science and, as a whole, to reveal its development patterns. It is an important branch of the Science of Science domain and an active field in current scientific research. The term "science" here refers not only to science as a body of knowledge but also science as a social activity, establishment, and industry. Scientometrics, as used in this study, is a broad concept that emphasizes the use of quantitative methods to explore the laws of scientific development and, thus, it also includes informetrics and nascent altmetrics.

Scientometrics reveals the current state of development of a disciplinary field or research topic. As such, it can help researchers and research managers quickly understand a field's full picture. Simultaneously, increasingly sophisticated methodologies and software tools have further expanded the boundaries and groups of users of scientometrics, making it possible for scientometrics-related research to be applied beyond existing disciplines in a variety of fields, such as environmental science, psychology, and clinical medicine (as shown in Figure 1).




Note: TS=(Scientometrics OR Bibliometrics OR Informetrics) AND DT=("REVIEW" OR "ARTICLE") AND FPY=(1978-2024)

**Figure 1. Distribution of journals involved in research papers related to scientometrics.**

However, as scientometrics has expanded and developed, problems such as ambiguous data source selection, generalized data retrieval, lack of data pre-processing, misleading data analysis and visualization, arbitrary data interpretation, confusing data storage, and unclear representation of data sharing and citations have remained prevalent. To this end, this study aims to provide suggestions from a data perspective that help ensure that scientometrics research is stable, diverse, transparent (Bornmann et al., 2021), and reproducible (Velden et al., 2018; Waltman et al., 2018). Moreover, these practical recommendations are closely aligned with the broader normative debates on responsible metrics in research evaluation. By integrating insights from recent frameworks such as the Coalition for Advancing Research Assessment (CoARA), this study underscores the importance of adopting rigorous data practices to support evaluation reform efforts. Such alignment not only enhances the relevance of this work to the objectives of the Framework for Responsible Assessment Metrics in Research (FRAME) but also contributes to fostering a more equitable and sustainable research evaluation ecosystem.

## Addressing Normative Challenges in Scientometrics Through Data Practices

*Data source selection*

The selection of data sources has a fundamental effect on scientometric research. Improper data source selection significantly affects the reliability of the conclusions drawn. In recent years, both the new bibliographic databases (e.g., Dimensions, Semantic Scholar,  Crossref, OpenAlex, and OpenCitations) (Gusenbauer, 2022) and the new variables provided by traditional databases (such as "funding acknowledgment", "usage count", "enriched cited references" in the Web of Science) have enriched the user's range of data source selection. It is worth mentioning that open data sources help make scientometric research more transparent and reproducible, and they also help make the field of scientometrics more equitable, since they enable scientometric research to be carried out by researchers and institutions that cannot afford a subscription to commercial proprietary data sources. However, although rich data sources provide opportunities for scientometric research, they also challenge scientometric scholars in selecting appropriate data sources. Therefore, the following recommendations were made.

*Data quality must meet the research's needs*

The increasing diversification and greater availability of data sources have given scientometricans more choices, but the quality of these new data sources may not always meet our expectations. Even when the most widely used commercial databases are chosen as data sources, such as Scopus or the Web of Science Core Collection (Zhu & Liu, 2020), the studies have shown different types of errors. Some scholars have even described Scopus as a "museum of errors/horrors" (Franceschini et al., 2016). For example, Scopus has a substantial number of funding information errors, and users need to carefully check the data accuracy before using the Scopus database for funding analysis (Liu et al., 2020). With all data sources, such as Google Scholar, users first need to gauge the quality of the database in terms of accuracy, completeness, or whatever the appropriate metric and only reasonably use that data source on the premise that its quality meets the needs of the research.

*Inclusion criteria must meet the research's needs*

The content inclusion criteria of the selected bibliographic database(s) should also meet the research needs. Some bibliographic databases are selective (e.g., Web of Science), whereas others are not (e.g., Google Scholar). There are also differences in the selection criteria for including data sources. For example, the Emerging Sources Citation Index (ESCI), which has a large number of journals, is a collection of potential candidates for the other three classic journal citation indices of the Web of Science Core Collection, but its selection criteria are slightly lower than those of the other three authoritative indices (Huang et al., 2017). Hence, users should judge whether the criteria for including some literature in a specific data source meet the requirements of the research. Sometimes, a tradeoff between quality and quantity may need to be made.

*Data characteristics should match the research question*

A mismatch between the characteristics of the data source and the research question may lead to anomalous research conclusions that further mislead the practice. Several types of bibliographic databases are frequently used. These include multidisciplinary databases (e.g., Web of Science, Scopus, Dimensions, Crossref, and OpenAlex), disciplinary databases (e.g., Medline for medicine), and regional databases (e.g., the Chinese Science Citation Database, the Chinese Social Sciences Citation Index, the Russian Science Citation Index, and the SciELO Citation Index (South America). Different data sources also have language, regional, and discipline biases (Liu, 2017; Martín-Martín et al., 2021). Users need to be familiar with the characteristics and shortcomings of the data sources they consider, and choose the appropriate data sources according to an accurate analysis of the needs of the research problem. For example, if you plan to study the journal paper output of social science research in China, in addition to the Social Sciences Citation Index, the Chinese Social Sciences Citation Index is an important data source. Another example is that it is not appropriate to use acknowledgments from the Web of Science Core Collection to analyze funding in the social sciences prior to 2015. Web of Science began to include funding information for SCIE papers in 2008, but funding information for SSCI papers has only been systematically recorded since 2015 (Liu et al., 2020).

*Describe the data source precisely*

An accurate description of the data sources is the cornerstone of research reproducibility. However, some studies have found that even in the field of scientometrics, there are still serious problems with the descriptions of the data sources used in many studies (Liu, 2019). One prominent problem is the failure to accurately and unambiguously represent the Web of Science and its core collection. Web of Science is now a search platform provided by Clarivate, which contains databases such as the Web of Science Core Collection. However, the full Web of

Science Core Collection contains two chemical indices and eight citation indices (one of which is the Science Citation Index Expanded). Different institutions may subscribe to different subsets of core collections with different years of coverage. Therefore, when using the Web of Science Core Collection, the sub-datasets used and the corresponding coverage year information need to be clearly disclosed (Liu, 2019).

*Data retrieval and acquisition*

As the basis of subsequent data analyses, accurate and proper data retrieval and acquisition directly influence the credibility and effectiveness of the data analysis and the corresponding results. Improper data retrieval and acquisition strategies might not only lead to misleading and distorted information but also waste human, energy, and material resources. Appropriate data retrieval and acquisition strategies are built on adequate investigation and a solid understanding of both databases and research needs. This means being familiar with the advantages and limitations of databases and their search rules. We investigated the common problems with data retrieval and acquisition in existing scientometric studies and summarized the issues into the following four precautions and norms for researchers to pay attention to when engaging in scientometric studies.

*Be familiar with the scope and update rules of the database*

The different backgrounds and service objects of the different databases indicate that each database has its own preferences when it comes to which data to collect. Therefore, a good retrieval strategy should not exceed the scope of the database. Taking the Web of Science as an example, the SCI database began collecting abstracts, author keywords, and keyword-plus information in 1991 (Clarivate Analytics, 2020). Therefore, when the topic filter (topic includes title, abstract, author keywords, and keyword-plus) is used in SCI databases, it is recommended to avoid the watershed year of 1991 during the investigation period. Otherwise, there might have been a misleading jump in the number of publications in 1991, skewing growth trends (Liu, 2021).

In addition, it usually takes time for a database to collect the newest data, and most databases have a fixed update frequency. For example, Web of Science generally updates on a bi-weekly basis. In addition, most databases have a time lag. Hence, the search date can also influence search results. For example, if we want to find documents published in 2020, a search on 1 January 1, 2021, will find fewer documents than a search on June 30, 2021, even if the same search strategy was used. Generally, a good buffer zone for accommodating the time lag is three to six months after the end of the study period.

*Deliberately select search terms*

Terms in a topic search are a common way to retrieve publications in a target field. Appropriate search terms can help ensure the accuracy of the data retrieval. A preliminary group of search terms should be selected only after the target field and the needs of a relevant search are fully understood. This preliminary search strategy can be further improved through discussions with relevant experts and multiple retrievals with spot checks and tests of the search results (Arora et al., 2013).

Evaluating a search strategy often requires an appropriate balance between recall and precision. Generally, when precision is high, recall is often low, and vice versa (Lambert, 1991). Researchers must also pay attention to the abbreviations used in the search terms. For example, the abbreviation of artificial intelligence is abbreviated as AI, but if AI is used as a search term, literature related to the biological index AI and the medical index AI would probably also be found (Wilson et al., 2022). In fact, it is very difficult to ensure that abbreviations are used only in the expected field. Generally, documents with the same abbreviations in nontarget fields are also found. Therefore, if precision is given priority, abbreviations need to be selected cautiously, and if abbreviations cannot be avoided, checking and cleaning the search results will be necessary to exclude irrelevant documents.

Attention should also be paid to the hypernyms and hyponyms of the search terms, and choosing a group of appositives is also recommended in one's search strategy. For example, (Fu & Waltman, 2022) used several appositives of climate change, including "climate change *", "climate change *", "climate variability *", "climate variability *", "global warming", "climate warming", "climate warming," in their data retrieval strategy to assemble a comprehensive body of literature on climate change. Alternatively, if the hyponym "greenhouse" of climate change research were to be included in the search terms, the search may result in a greater proportion of documents related to greenhouse effects, and this could overweight the characteristics of greenhouse across the whole field, skewing the research results.

*Understand the search rules of the database*

A data retrieval strategy and a detailed search query should be constructed based on a specific database. Before conducting a search, it is important to be familiar with the rules of the selected data platform. Common search rules include Boolean operators, proximity operators, quotation marks, wildcards, and truncation. The negligence and misuse of search rules have a significant impact on data retrieval and further analysis. Taking the use of quotation marks as an example, Topic=(solid waste *) and  Topic=("solid waste *") are two different search queries. The first search query found documents with solids and waste separately in different sentences. Thus, some documents irrelevant to solid waste were likely to be found. However, the second search query with quotation marks only found papers with solid waste as a phrase. This mistake was made in (Liu et al., 2014) and, as a result, almost twice the number of publications was found as there should have been. There were also significant differences in subsequent results, such as growth trends and country performance. In this case, the lack of quotation marks leads to inaccurate results and unreliable conclusions.

*Elaborate the strategy of data retrieval and acquisition*

The detailed strategies for data retrieval and acquisition need to be explained in the final publicly released publications to ensure the transparency of data retrieval and acquisition. A clearly stated search strategy provides the basis for judging the reliability of scientific activities. It also helps ensure the authenticity and certainty of scientific knowledge. It is suggested that the general description of the search strategy specifies the database, subdatabase, search query, investigation period, actual retrieval time, and any other salient information. The basic criterion for judging whether a search strategy is clearly explained is that if a reader were to follow the explanation, they would retrieve approximately the same dataset and reproduce similar search results. Even though data sources such as Web of Science, Scopus, and others are continuously being updated, new records are being added, and old records may be updated or may even be removed by documenting the search strategy, other researchers may be able to approximately reproduce the analysis.

### Data pre-processing

Data pre-processing is defined as a series of necessary processes that occur after data retrieval and before data analysis and modeling. It primarily includes tasks such as data integration, data cleaning, and data transformation. For instance, multi-source data are often used in scientometric research, but the structure of the data gathered from each source can be inconsistent. Simultaneously, with the increasing amount of scientific data, problems such as missing data, duplicate data, and anomalies in the data may seriously affect the validity and reliability of data analysis results. High-quality data pre-processing can effectively improve the quality of the analysis results and conclusions. It can also make the data fit the models and analysis tools better, which can help improve the efficiency of the analysis. Therefore, we have provided the following suggestions for ensuring efficient and high-quality data pre-processing.

#### *Adequately understand the data fields*

Owing to the inconsistency of field formats and statistical calibers, data fields from different data sources must be merged carefully. For example, the "Times Cited" fields of different databases are usually limited to the internal statistics of their respective databases, which means that the counts given by different databases are likely to be different. Therefore, the integration of such fields needs to be handled with caution (Pech & Delgado, 2020). Therefore, the characteristics of a field should be fully considered during pre-processing. For example, there are several fields regarding funding information in the Web of Science, among which the "FU" field records the funding institution and authorization number, while the "FX" field records funding information as unstructured text. Hence, the "FU" field is better structured and, therefore, less difficult to pre-process, whereas the "FX" field provides more complete information. In addition, keywords, as subject tags of research content, are also commonly used in scientometrics research. However, there are two keyword fields in WOS, "Author's Keywords" and "Keywords Plus". Unlike the Author's Keywords, the Keywords Plus field is based on the original records and their references, and is designed to make data retrieval more convenient. Therefore, in some research fields, Keyword Plus may not be sufficient to describe the uniqueness and comprehensiveness of the content of the literature and should be used with caution (Zhang et al., 2016).

*Clarify data pre-processing tasks*

In scientometric research, the original field content retrieved from a database cannot usually be directly used for subsequent analyses. Rather, the data needs to be pre-processed. Pre-processing tasks generally include data integration, deduplication, field extraction, and cleaning. However, in practice, the pre-processing methods to be applied should be selected according to the requirements of the specific analysis. Data deduplication and merging are basic pre-processing operations. Even when data have only been collected from a single database, there can still be problems with duplicate records. For example, the same paper may be included with different statuses, such as published, online, in-press, correction, and withdrawal. Duplicate records should be merged and deleted (Gagolewski, 2011).

Another common pre-processing task is to disambiguate author names, institutions, and references. If the research includes author analyses, author disambiguation should be considered first. This includes distinguishing author names and merging the different record forms of the same author. Different disambiguation strategies may impact the subsequent cooperative network construction or citation analysis (Kim & Diesner, 2015, 2016). Disambiguating Asian authors, especially Chinese and Korean authors, is generally considered more challenging because of regional and linguistic differences (Harzing, 2015; Strotmann & Zhao, 2012). Therefore, the author's other attributes, such as ORCID, institution, email address, research field, and ResearchGate records, can be considered to further improve disambiguated results (Abdulhayoglu & Thijs, 2017; Han et al., 2017; Porter, 2022; Youtie et al., 2017)

*Ensure the standardization and repeatability of data pre-processing*

Data pre-processing often relies on rules, stop word lists, and even manual processing to repeatedly revise and improve the results. Standardizing and making the data pre-processing procedure repeatable are key to ensuring the robustness and reliability of data analysis.

Taking text feature extraction as an example, common pre-processing steps include stemming, stop word list filtering, TF-IDF, fuzzy matching, and others. However, because text analysis usually requires consideration of the characteristics of disciplines or domains, different cleaning methods or processing steps may affect which text features are extracted (Newman et al., 2014; Zhang et al., 2014). For example, the word "nanotechnology" may be highly valuable in multidisciplinary studies. However, it is probably meaningless in a study of nanotechnology and may even be classified as a stop word. Therefore, data pre-processing should ensure that the procedural steps are rational in terms of stop word list design and threshold selection. Moreover, although in practice, it is usually not possible to document data-processing procedures in full detail in a research paper, making source codes or documentation openly available seems to be the best solution. It should also be emphasized that data pre-processing is an almost endless task of improvement, but it can be very time-consuming. Therefore, it is usually necessary to balance the quality of pre-processing with time, labor, and cost taken to do it. As the volume of pre-processing tasks increases, the results can be evaluated through sampling or other metrics to ascertain the quality and efficiency of the pre-processing regime.

*Bridge data pre-processing results to data analysis requirements*

The typical results of data pre-processing are cleaned data, extracted relationships, and preliminary statistics, which provide fundamental information for subsequent analysis tasks. For example, social network analysis tools, such as Gephi and Ucinet, have different requirements regarding the structure of the network data to be input, which can be either in the form of an edge or an adjacent matrix. Therefore, the structure and storage format of the pre-processing results should be considered to ensure that they fit the subsequent analysis tools and analysis methods. This will also help improve the efficiency of the data analysis.

## Data analysis

Data analysis is often carried out by means of some data analysis methods, tools, or software. Whichever means, the analysis should be rigorous and standardized, and the process should be as clear and transparent as possible. Anything less may lead to misjudgments in the subsequent data interpretation and could also raise the threshold for reproducibility of the research. Based on the common problems with data analysis in existing scientometrics research, we suggest that more attention should be paid to the selection of data analysis software, the use and parameter settings of any software used, data standardization, and visualization of data analysis.

*Select appropriate data analysis software*

Currently, many analytical tools can be used in scientometric research, so users must choose the appropriate one according to the characteristics of their data and the purpose of their analysis. For example, CiteSpace and VOSviewer can be used for co-citation, co-authorship, and keyword co-occurrence analysis. CitNetExplorer and HistCite were used for direct citation networks. Pajek, Ucinet, and Gephi can be used to visualize social network data and analyze network characteristics. Each tool has its advantages and limitations, and no single piece of software can accomplish all functions. Therefore, multiple tools must be used simultaneously. In addition, existing scientometric software has its limitations, and sometimes, researchers need to rely on general-purpose tools, such as Python and R. However, when using software, record which version of a software tool was used and which parameters were adjusted, which is very important for the optimization of the network, and explain the meanings of the parameters. For example, the layout area of VOSviewer provides "Attraction", "Repulsion", and "Advanced Parameters" to optimize the layout of a graph. Different parameters produce different rendering effects.

*Distinguish different data standardization*

In scientometric research, many similar issues related to data standardization require attention, and the counting method is one example of a broader set of issues. This refers to the calculation method of assigning ownership of scientific research papers according to certain rules. The counting method can be used to calculate the number

of papers published by authors, institutions, and countries and to analyze the frequency of citations. Most notably, this method can influence the allocation of scientific research resources and the content of science and technology policies (Sivertsen et al., 2019).
There are many ways to categorize these counting methods. The basic methods used are full counting and fractional counting. However, based on the correspondence between counting objects and counting units in various metrological problems, Gauffriau et al. (2007) divided the various counting methods into five categories: complete counting methods, complete-normalized counting methods, straight counting methods, whole counting methods, and whole-normalized counting methods. These can also be divided into different counting units. Here, we have full counting, country/organization/address/author level fractional counting, first author counting, and corresponding author counting (Waltman & van Eck, 2015). Full-counting methods reflect participation, whereas fractional-counting methods reflect contribution. Full counting is more commonly used at the individual level, as in the h-index proposed by Hirsch (Hirsch, 2005). Fractional counting is an aggregate metric in which the sum of the counts is the same as the number of papers. As such, this style of counting provides balance, consistency, and accuracy in standardized bibliometric measurements. In scientific research evaluations, it is usually necessary to compare the academic influence of papers in different fields. However, because different fields have different citation densities, these counts must be standardized before undertaking such a comparison. Moreover, the counting method used affects the results of standardization across different disciplines. Using different counting methods to calculate the number of citations in co-authored papers will result in different results, which will further affect any standardized citation impact indicators calculated when comparing the influence of papers across different fields (Lin et al., 2013).

*Give full attention to data visualization*

To make good use of graph functions in analysis tools, appropriate dimensions and quantities must be selected. Only this way will one create a picture that is "worth a thousand words." On this basis, the selected visual map should be based on how best to present the data. For example, Citespace provides a network view, timeline view, and time-zone view. VOSviewer offers a network view and cluster density view. Pajek has 2D, 3D, and dynamic community maps, whereas SciMAT provides an item overlay map, an evolution map, and a cluster network.
In addition, there are other issues that need to be addressed when performing data visualization. Different authors often use different words for the same concept, so it may be necessary to combine synonyms such as e-health, eHealth, mobile learning, and m-learning; otherwise, the graph obtained will be neither refined nor accurate. There are often similar problems with the names of countries and institutions, such as England, the UK, the University of Tokyo, and Tokyo University. In the case of VOSviewer, the above problem can be normalized based on the vocabulary obtained before visualization.

## Data Interpretation

Reasonable and accurate interpretations are mandatory, regardless of the data analysis methods, tools, and/or software applications. Interpreting data requires a scientific nature and constructiveness. Here, a scientific nature means that one should strictly and standardly follow certain rules, regulations, and/or procedures of data interpretation and understand that inappropriate interpretations can greatly distort the results of data analyses. Constructiveness, on the other hand, is reflected by the fact that people from different domains may render various results because of factors such as professional background and subjective feelings. Hence, we suggest the following rules when interpreting data:

*Preventing over-interpretation*

There are many examples of overinterpreting data in scientometric research, among which people often confuse the difference between correlation and causation (Pearl & Mackenzie, 2018). At present, the majority of scientometric studies have been undertaken at the correlation level. This includes, for instance, (1) presenting the correlations found between two variables with some statistical charts (e.g., bivariate scatter plots, tri-variate bubble plots, etc.) and/or (2) depicting the results of statistical regression analyses between multiple variables. In studies indicating correlation-level results, one should avoid using words that imply causation when interpreting the results. For example, words such as "result in", "lead to", "influence", "impact", and "affect" all hint at causation. If causal inference has not been tested in the research, more careful and conservative expressions should be used. These include, but are not limited to, "relate to", "is proportional to", "as X increases, Y tends to...", etc.

In addition to confusing correlation and causation, scientists conducting descriptive research (e.g., mapping knowledge domains) may have made some inherent assumptions. These preconceived notions might lead them to focus on one part of the results that matches their psychological expectations when interpreting the results and ignore those that do not. Avoiding this pitfall requires vigilance when interpreting data.

*Avoiding under-interpretation*

At the other extreme, over-interpreting data is under-interpreting data. Generally, interpreting the results of scientometric research can be divided into four levels:

Level 1: Numerical level. Here is an example of an interpretation using the numerical dimension: "In biology, when the proportion of publications' references not belonging to the discipline is 0, the normalized value of their citations is about 0.5". This level is the most straightforward and serves as the basis for further interpretation. It should be noted that when interpreting at the numerical level, it is also necessary to focus on outliers, trends (temporally), multivariate relationships, and so on.

Level 2: Numerical comparison and inductive level. A typical example of Level 2 could be something like "as the value of the horizontal axis increases, the value of the vertical axis increases first and then decreases." Note that Level 2 still does not touch on any conceptual level of scientometric knowledge.

Level 3: Conceptual level. This layer links the numerical results (indicators) to scientometric concepts and/or domain knowledge. Suppose that the indicator on the horizontal axis is the proportion of a publication's references that do not belong to a discipline. The corresponding concept/construct of this indicator could be the degree of interdisciplinarity of a publication. If the indicator on the vertical axis is the relative number of citations of a publication in its discipline, the corresponding concept/construct of this indicator could be a publication's scientific impact. Considering the Level-2 interpretation we mentioned, we may further interpret it as, "As the interdisciplinary nature increases, the academic influence of scientific literature first increases and then decreases."

Level 4: Implications. Discussions at this level mainly focus on the significance and importance of the findings from Levels 1-3. Level 4 interpretations often vary and can be carried out from many distinct aspects, such as theories of scientometrics, methodology, scientific and technical policies, and pedagogy. For example, " with the increase of interdisciplinarity, the scientific impact of publications first increases and then decreases." may offer empirical evidence if it presents a causal relationship to help policymakers and funding providers form policies on interdisciplinarity and unidisciplinarity.

Understanding these four levels of data interpretation would help researchers paint a more vivid and comprehensive picture of the material under study. In addition, preventing underinterpretation generally requires a comparison with previous research results.

*Involving domain experts in the interpretation*

Scientometrics is a typical "meta-discipline". This means that empirical research in scientometrics often involves one or more disciplines. Therefore, domain experts are often required to interpret data. If one traces back the history of citation analysis and mapping knowledge in domain studies ( e.g., co-citation and bibliographic coupling analyses), for example, one impressive highlight worthy of recognition is that domain experts have strictly interpreted the results rather than simply discussing results without the perspective of any domain knowledge. Otherwise, normativeness, rigor, and persuasiveness can also be significantly reduced. More severely, the lack of interpretation by domain experts may hinder further reflection, optimization, and innovation of research methods (Bar-Ilan, 2008).

## Data storage

The organization and storage of data are essential steps in a bibliometric analysis. This kind of housekeeping ensures that one obtains reliable quantitative analysis results and that the results are repeatable (Ferrara & Salini, 2012). Data storage runs through the entire scientometric research process and plays an essential role in academic research and communication, manuscript submission and revision, and even after publication. Therefore, we make the following recommendations when storing data.

*Record all kinds of data fields and their derived variables*

There is a range of metadata fields for scientific and technological documents, each with rich meaning (Pranckutė, 2021). However, for convenience, these fields are often labeled in the form of abbreviations, making it difficult to determine their meanings from the name. For example, in the core collection field of Web of Science, there are at least four fields concerning citations and usage: TC indicates the number of citations in the core collection, Z9 indicates the citations in the core collection, U1 is the number of times used in the last 180 days, and U2 is the number of times used since 2013. Another example is the literature on patent type. The applicant and inventor have relationships with the subjects related to the patent right, but the former indicates the patent applicant (both the institutions and the individuals), and the latter indicates the patent's inventor (generally a specific person). These two are easily confused. The above are only original data fields derived from the scientific literature. In some applied follow-up research, new variables based on the original data fields will be generated, and some new variables will result in new versions owing to differences in algorithms and parameter settings. In this way, managing the data fields and derived variables may become more complicated, and errors may occur if one is not careful. Furthermore, as time goes by, it is easy to forget the specific meaning of fields and variables or become confused about them. Therefore, it is wise to provide each metadata field with a clear name and definition to avoid this problem. Moreover, if the variables for subsequent analysis have been generated based on metadata fields, it is even more necessary to name the fields and variables scientifically to ensure that they are scalable and that new variables can be generated and added.

*Deal with data backups and file preservation*

Data from scientific literature have increasingly become a support for research results and the basis of scientific measurements. Therefore, long-term preservation and archiving of data is becoming a realistic demand for scientometric analysis. However, this raises many logical questions. Who will maintain a regular data backup? Where are data archives stored? If funding agencies are present, what specific requirements do they have regarding the format and scheme of data retention? All these questions must be considered. If only a small amount of data is available, a local storage solution may be sufficient. However, if there is a large amount of data, it may be better to choose a cloud platform as the storage medium.

Data security is also worth noting (Assunção et al., 2015). Scientific articles are often restricted by the copyright of the publisher, and security risks, such as leakage, need to be prevented. As time passes, physical aging of the storage medium may also lead to data loss and difficulties in recovery. The scientific community's demand for research reproducibility creates additional challenges for data openness and sharing (Fecher et al., 2015). Finally, creating a clear and concise data file will also make storing bibliographic data more convenient.

## Data sharing and data citations

Data sharing and data citations establish an interactive relationship between source and flow, where the source is the active distribution of data achievements to the scientific community by data creators, and the flow is the data citations. This relationship is the final stage of the scientometric life cycle and promotes the development of an open science culture. To this end, one of the main objectives of standardizing data storage is to support data sharing, while formal data citations honor data producers and, at the same time, advance data flows as creative elements of a scientific system. However, the significance of data sharing and citations has not been emphasized sufficiently. From the standpoint of responsible scientometric research, this section offers advice to the authors.

*Share data via a variety of channels*

Recently, many reputable journals have started requesting authors to disclose data, models, codes, and other pertinent attachment materials when publishing papers. To some extent, this ensures that research can be repeated and validated (Wilkinson et al., 2016). However, because most papers where the data are self-stored only provide access via an author's email address, the accessibility and usefulness of the data are limited. Consequently, we advise the authors to upload as much data as possible to a trustworthy platform. For instance, Mendeley Data, a platform for managing and sharing scientific research data created by Elsevier in 2015, offers a complete solution for data sharing, including data upload, data release, data storage, and data access (Bornmann & Haunschild, 2017); Zenodo, an open repository for all scholarships, enables researchers from all disciplines to share and preserve their research outputs, regardless of size or format. Free to upload and free to access, Zenodo makes scientific outputs of all kinds citable, shareable, and discoverable for the long term. Additionally, researchers can submit their data results for publication in peer-reviewed data journals, such as Scientific Data, which is part of the Nature Publishing Group. This is an important way to strike a balance between data sharing and performance acknowledgments. To encourage the exchange of high-quality data among peers, scientometric researchers should also try to share scientific research data more voluntarily, either through extensive collaboration or signing a data-sharing agreement. However, when someone shares the data, attention should be paid to the license under which data are shared (e.g., CC-BY or CC0), and they should not share data they are not allowed to share (e.g., raw data from Web of Science). To be effective in significantly increasing data availability, data-sharing policies should prescribe mechanisms for sharing that ensure reliable and long-term access to data (Federer et al., 2018), so including a data availability statement in research articles provides some potential solutions.

*Formally cite open-access research data*

Data citations generally refer to the practice of an author in identifying the source of the data used via a reference, footnote, or text note. The data cited include not only the sets of data that are crucial to the study but also the factual data that explain the background of the investigation and any additional data in the literature that attests to the strength of the study's findings. Some studies report that the average number of data citations in scientific literature is low, but the proportion of informal citations is high (Park et al., 2018). Informal citations typically describe data sources in the context of tables, acknowledgments, or other parts of a paper. However, a standardized data index such as the Data Citation Index (DCI) cannot fully index informal citation records (Park & Wolfram, 2017, 2019). Consequently, we advise the authors to properly cite their data, following the format of a traditional literature citation. DataCite, which is working to standardize data citation practices and provide DOIs to datasets, is a guide that authors can use. According to the DataCite standard, a data citation must, at the very least, list the data's title, creators, publisher, DOIs, and year of publication (Robinson-García et al., 2016). The URL of the original source website is required if the data does not have a DOI. At the same time, we advise scientometric researchers to actively and consciously retrieve, cite, and analyze data records in DCI, Zenodo (Andrea et al., 2021), or another data resource platform to promote the development of such services.

## Discussion and Conclusion

The scientometric framework proposed in this study systematically addresses key methodological challenges across the data lifecycle, from source selection to analytical interpretation. By deconstructing each operational phase, we revealed how technical decisions fundamentally shape research validity. While this work provides structured recommendations for standardization, it is critical to recognize that rigorous data practices alone cannot ensure responsible metrics in evaluation systems. The principles advocated in the San Francisco Declaration on Research Assessment (DORA) (DORA Group, 2013) and Leiden Manifesto (Hicks et al., 2015) remain essential complements, particularly regarding the contextualized use of metrics in research evaluation and their policy implications.

Moving forward, the scientometric field should treat methodological rigor and responsible metric use as interdependent requirements. Our phased framework provides scaffolding for reproducible research, but its implementation should actively engage with ongoing evaluation reform efforts, such as the Coalition for Advancing Research Assessment (CoARA). By explicitly addressing FRAME themes—such as enhancing transparency in evaluation processes, fostering a balanced use of bibliometric and qualitative insights, and promoting stakeholder engagement—this framework supports the development of more equitable and context-sensitive evaluation systems.

Additionally, the framework encourages the adoption of open science practices, such as sharing data and methodologies, to further enhance transparency and accountability. This aligns with the FRAME theme of promoting openness and trust in research evaluation. Only through this dual focus—optimizing technical processes while embedding ethical guidelines—can scientometrics fulfill its potential as a robust, socially accountable science of science. By aligning with initiatives like the Framework for Responsible Assessment Metrics in Research (FRAME), this study underscores the importance of transparency and accountability in shaping future evaluation systems. Such integration ensures that scientometric practices not only advance methodological rigor but also contribute to a fairer and more sustainable research ecosystem.

Finally, the framework highlights the need for continuous dialogue between researchers, policymakers, and evaluators to ensure that bibliometric indicators are used in a way that aligns with societal values and research priorities. This participatory approach, rooted in the FRAME principle of inclusivity, further strengthens the framework's relevance to responsible research assessment.

## Acknowledgments

We thank Alan L. Porter (Georgia Institute of Technology), Jiang Li (Nanjing University), Liming Liang (Henan Normal University), Lin Zhang (Wuhan University), Ludo Waltman (Leiden University), Ronald Rousseau (Antwerp University), Xianwen Wang (Dalian University of Technology), and Yishan Wu (Chinese Academy of Science and Technology for Development), for their valuable comments on this study.

## References

Abdulhayoglu, M. A., & Thijs, B. (2017). Use of ResearchGate and Google CSE for author name disambiguation. *Scientometrics*, *111*(3), 1965-1985. https://doi.org/10.1007/s11192-017-2341-y

Andrea, S.-C., Nicolas, R.-G., van Thed, L., & Rodrigo, C. (2021). Exploring the relevance of ORCID as a source of study of data sharing activities at the individual-level: a methodological discussion. *Scientometrics*, *126*(8), 7149-7165. https://doi.org/10.1007/s11192-021-04043-5

Arora, S. K., Porter, A. L., Youtie, J., & Shapira, P. (2013). Capturing new developments in an emerging technology: an updated search strategy for identifying nanotechnology research outputs. *Scientometrics*, *95*(1), 351-370. https://doi.org/10.1007/s11192-012-0903-6

Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A. S., & Buyya, R. (2015). Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, *79-80*, 3-15. https://doi.org/https://doi.org/10.1016/j.jpdc.2014.08.003

Bar-Ilan, J. (2008). Informetrics at the beginning of the 21st century—A review. *Journal of Informetrics*, *2*(1), 1-52. https://doi.org/10.1016/j.joi.2007.11.001

Bornmann, L., Guns, R., Thelwall, M., & Wolfram, D. (2021). Which aspects of the Open Science agenda are most relevant to scientometric research and publishing? An opinion paper. *Quantitative Science Studies*, *2*(2), 438-453. https://doi.org/10.1162/qss_e_00121

Bornmann, L., & Haunschild, R. (2017). Measuring field-normalized impact of papers on specific societal groups: An altmetrics study based on Mendeley Data. *Research Evaluation*, *26*(3), 230-241. https://doi.org/10.1093/reseval/rvx005

Clarivate Analytics. (2020). *Web of Science Core Collection Help*. Retrieved 9 Oct 2022 from https://images.webofknowledge.com/images/help/WOS/hp_full_record.html

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

Egghe, L. (2015). Message from the retiring Editor-in-Chief. *Journal of Informetrics*, *9*(1), A1-A2. https://doi.org/10.1016/j.joi.2015.01.007

Fecher, B., Friesike, S., & Hebing, M. (2015). What Drives Academic Data Sharing? *Plos One*, *10*(2), e0118053. https://doi.org/10.1371/journal.pone.0118053

Federer, L. M., Belter, C. W., Joubert, D. J., Livinski, A., Lu, Y.-L., Snyders, L. N., & Thompson, H. (2018). Data sharing in PLOS ONE: An analysis of Data Availability Statements. *Plos One*, *13*(5), e0194768. https://doi.org/10.1371/journal.pone.0194768

Ferrara, A., & Salini, S. (2012). Ten challenges in modeling bibliographic data for bibliometric analysis. *Scientometrics*, *93*(3),
765-785. https://doi.org/10.1007/s11192-012-0810-x

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016). The museum of errors/horrors in Scopus.
*Journal of Informetrics*, *10*(1), 174-182. https://doi.org/10.1016/j.joi.2015.11.006

Fu, H.-Z., & Waltman, L. (2022). A large-scale bibliometric analysis of global climate change research between 2001 and 2018. *Climatic Change*, *170*(3), 36. https://doi.org/10.1007/s10584-022-03324-z

Gagolewski, M. (2011). Bibliometric impact assessment with R and the CITAN package. *Journal of Informetrics*, *5*(4), 678-692. https://doi.org/10.1016/j.joi.2011.06.006

Gusenbauer, M. (2022). Search where you will find most: Comparing the disciplinary coverage of 56 bibliographic databases. *Scientometrics*, *127*(5), 2683-2745. https://doi.org/10.1007/s11192-022-04289-7

Han, H., Yao, C., Fu, Y., Yu, Y., Zhang, Y., & Xu, S. (2017). Semantic fingerprints-based author name disambiguation in Chinese documents. *Scientometrics*, *111*(3), 1879-1896. https://doi.org/10.1007/s11192-017-2338-6

Harzing, A.-W. (2015). Health warning: might contain multiple personalities—the problem of homonyms in Thomson Reuters Essential Science Indicators. *Scientometrics*, *105*(3), 2259-2270. https://doi.org/10.1007/s11192-015-1699-y

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429-431.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569-16572. https://doi.org/10.1073/pnas.0507655102

Huang, Y., Zhu, D., Lv, Q., Porter, A. L., Robinson, D. K. R., & Wang, X. (2017). Early insights on the Emerging Sources Citation Index (ESCI): an overlay map-based bibliometric study. *Scientometrics*, *111*(3),
2041-2057. https://doi.org/10.1007/s11192-017-2349-3

Kim, J., & Diesner, J. (2015). The effect of data pre-processing on understanding the evolution of collaboration networks. *Journal of Informetrics*, *9*(1), 226-236. https://doi.org/10.1016/j.joi.2015.01.002

Kim, J., & Diesner, J. (2016). Distortive effects of initial-based name disambiguation on measurements of large-scale co-authorship networks.
*Journal of the Association for Information Science and Technology*, *67*(6), 1446-1461. https://doi.org/10.1002/asi.23489

Lambert, N. (1991). Online searching of polymer patents: precision and recall.
*Journal of Chemical Information and Computer Sciences*, *31*(4), 443-446. https://doi.org/10.1021/ci00004a002

Lin, C.-S., Huang, M.-H., & Chen, D.-Z. (2013). The influences of counting methods on university rankings based on paper count and citation count. *Journal of Informetrics*, *7*(3), 611-621. https://doi.org/10.1016/j.joi.2013.03.007

Liu, A.-Y., Fu, H.-Z., Li, S.-Y., & Guo, Y.-Q. (2014). Comments on "Global trends of solid waste research from 1997 to 2011 by using bibliometric analysis". *Scientometrics*, *98*(1), 767-774. https://doi.org/10.1007/s11192-013-1086-5

Liu, W. (2017). The changing role of non-English papers in scholarly communication: Evidence from Web of Science's three journal citation indexes. *Learned Publishing*, *30*(2), 115-123. https://doi.org/10.1002/leap.1089

Liu, W. (2019). The data source of this study is Web of Science Core Collection? Not enough. *Scientometrics*, *121*(3), 1815-1824. https://doi.org/10.1007/s11192-019-03238-1

Liu, W. (2021). Caveats for the use of Web of Science Core Collection in old literature retrieval and historical bibliometric analysis. *Technological Forecasting and Social Change*, *172*, 121023. https://doi.org/10.1016/j.techfore.2021.121023

Liu, W., Tang, L., & Hu, G. (2020). Funding information in Web of Science: an updated overview. *Scientometrics*, *122*(3), 1509-1524. https://doi.org/10.1007/s11192-020-03362-3

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. *Scientometrics*, *126*(1), 871-906. https://doi.org/10.1007/s11192-020-03690-4

Newman, N. C., Porter, A. L., Newman, D., Trumbach, C. C., & Bolan, S. D. (2014). Comparing methods to extract technical content for technological intelligence. *Journal of Engineering and Technology Management*, *32*, 97-109. https://doi.org/10.1016/j.jengtecman.2013.09.001

Park, H., & Wolfram, D. (2017). An examination of research data sharing and re-use: implications for data citation practice. *Scientometrics*, *111*(1), 443-461. https://doi.org/10.1007/s11192-017-2240-2

Park, H., & Wolfram, D. (2019). Research software citation in the Data Citation Index: Current practices and implications for research software sharing and reuse. *Journal of Informetrics*, *13*(2), 574-582. https://doi.org/10.1016/j.joi.2019.03.005

Park, H., You, S., & Wolfram, D. (2018). Informal data citation for data sharing and reuse is more common than formal data citation in biomedical fields. *Journal of the Association for Information Science and Technology*, *69*(11), 1346-1354. https://doi.org/10.1002/asi.24049

Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. Basic books.

Pech, G., & Delgado, C. (2020). Assessing the publication impact using citation data from both Scopus and WoS databases: an approach validated in 15 research fields. *Scientometrics*, *125*(2), 909-924. https://doi.org/10.1007/s11192-020-03660-w

Porter, S. J. (2022). Measuring research information citizenship across ORCID practice. *Frontiers in Research Metrics and Analytics*, *7*, 779097. https://doi.org/10.3389/frma.2022.779097

Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The Titans of Bibliographic Information in Today's Academic World. *Publications*, *9*(1), 12. https://www.mdpi.com/2304-6775/9/1/12

Robinson-García, N., Jiménez-Contreras, E., & Torres-Salinas, D. (2016). Analyzing data citation practices using the data citation index. *Journal of the Association for Information Science and Technology*, *67*(12), 2964-2975. https://doi.org/10.1002/asi.23529

Sivertsen, G., Rousseau, R., & Zhang, L. (2019). Measuring scientific contributions with modified fractional counting. *Journal of Informetrics*, *13*(2), 679-694. https://doi.org/10.1016/j.joi.2019.03.010

Strotmann, A., & Zhao, D. (2012). Author name disambiguation: What difference does it make in author-based citation analysis? *Journal of the American Society for Information Science and Technology*, *63*(9), 1820-1833. https://doi.org/10.1002/asi.22695

Velden, T., Hinze, S., Scharnhorst, A., Schneider, J. W., & Waltman, L. (2018). Exploration of reproducibility issues in scientometric research Part 2: Conceptual reproducibility. *arXiv preprint arXiv:1804.05026*.

Waltman, L., Hinze, S., Scharnhorst, A., Schneider, J. W., & Velden, T. (2018). Exploration of reproducibility issues in scientometric research Part 1: Direct reproducibility. *arXiv preprint arXiv:1804.05024*.

Waltman, L., Larivière, V., Milojević, S., & Sugimoto, C. R. (2020). Opening science: The rebirth of a scholarly journal. *Quantitative Science Studies*, *1*(1), 1-3. https://doi.org/10.1162/qss_e_00025

Waltman, L., & van Eck, N. J. (2015). Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics*, *9*(4), 872-894. https://doi.org/10.1016/j.joi.2015.08.001

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship.
*Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Wilson, D. L., Tolson, J., Churchward, T. J., Melehan, K., O'Donoghue, F. J., & Ruehland, W. R. (2022). Exclusion of EEG-based arousals in wake epochs of polysomnography leads to underestimation of the arousal index. *Journal of Clinical Sleep Medicine*, *18*(5), 1385-1393. https://doi.org/10.5664/jcsm.9878

Youtie, J., Carley, S., Porter, A. L., & Shapira, P. (2017). Tracking researchers and their outputs: new insights from ORCIDs. *Scientometrics*, *113*(1), 437-453. https://doi.org/10.1007/s11192-017-2473-0

Zhang, J., Yu, Q., Zheng, F., Long, C., Lu, Z., & Duan, Z. (2016). Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. *Journal of the Association for Information Science and Technology*, *67*(4), 967-972. https://doi.org/10.1002/asi.23437

Zhang, Y., Porter, A. L., Hu, Z., Guo, Y., & Newman, N. C. (2014). "Term clumping" for technical intelligence: A case study on dye-sensitized solar cells.
*Technological Forecasting and Social Change*, *85*,
26-39. https://doi.org/10.1016/j.techfore.2013.12.019

Zhu, J., & Liu, W. (2020). A tale of two databases: the use of Web of Science and Scopus in academic papers.
*Scientometrics*, *123*(1), 321-335. https://doi.org/10.1007/s11192-020-03387-8