UNIVERSIDAD DE GRANADA

Daniel Torres-Salinas
Wenceslao Arroyo-Machado
Nicolás Robinson-García

Principles of Evaluative Bibliometrics in a DORA/CoARA Context

Daniel Torres-Salinas
Wenceslao Arroyo-Machado
Nicolás Robinson-García

Principles of Evaluative Bibliometrics in a DORA/CoARA Context

Powered by

First Edicion, January 2025

Published by InfluScience Editions, a Temporary Autonomous Zone in Granada, in collaboration with U-CHASS, a Research Excellence Unit.

DTS & NRG are members of the Departament of Information and Communication at the University of Granada (Granada, Spain). WAM is member of INGENIO, CSIC-UPV (Valencia, Spain).

Corresponding author: torressalinas@ugr.es
Universidad de Granada, Campus Cartuja, Colegio Máximo
18071, Granada, Spain

DOI: 10.5281/zenodo.14672066

1. PREFACE

Bibliometrics, also referred to as scientometrics, is generally defined as the discipline that quantitatively analyzes scientific information. This field has been fundamental in the transformation of research and academic evaluation. Bibliometrics was consolidated from the 1960s onwards with the founding of the Institute for Scientific Information (ISI) by Eugene Garfield and the commercialization of the Science Citation Index (SCI), which allowed for the tracking and counting of the number of citations received by publications. This advance established the foundation for the development of all subsequent applications. The recognition of bibliometrics as a management tool has been global and undeniable, but the easy and massive availability of indicators has triggered new challenges and questioned a usage that has often been quite irresponsible.

Recent years have also witnessed significant technological transformations marked by social metrics and an overabundance of sources accompanied by the movement of responsible metrics. A large number of fairer and more equitable evaluation, promoting a more appropriate use of bibliometric indicators and fostering a more intensive adoption of qualitative methods supported by expert judgment. Therefore, new policies and evaluative criteria are being designed, which is leading the bibliometric community to reevaluate their roles and reflect on their practices.

This drive towards a reform of scientific evaluation should not be interpreted as a threat to those of us professionally dedicated to this field but rather offers a window of opportunity to foster more integrated collaboration. We must be self-critical and admit that many of the current problems are partly due to prevailing attitudes in our discipline. In recent years, we may have shown an evaluative condescension, a stance that has distanced us from maintaining rigorous standards and actively questioning evaluation systems and methods. Additionally, we have fallen into a certain self-absorption, generating a profusion of indicators that, although seemingly efficient, have shown limited holistic application. The incessant pursuit of more sources, data, and indicators has led to an overabundance, overshadowing fundamental debates such as the design of clear evaluation protocols and guidelines for their effective implementation. Adding to this the lack of detailed supervision by experts and managers and the absence of well-defined criteria, we find ourselves facing an "evaluative problem" that has been exacerbated.

The central thesis of this book argues that the theoretical-practical framework we seek has always been present; it just requires updating. We refer to evaluative bibliometrics, a proposal that, as will be detailed, took shape during the 1980s. This branch of bibliometrics offers us the necessary tools to integrate harmoniously with new evaluative systems without one invalidating the other. Evaluative bibliometrics has an integrative and collaborative essence that could broaden the conversation by incorporating different voices and highlighting the valuable contributions that experts can make to the field. To do this, it is vital to challenge the negationist trend towards bibliometrics, rescuing and recognizing the work of academics who are being unjustifiably sidelined. In fact, a careful examination of the works of pioneers in the field shows that many of the current principles had already been articulated and debated previously.

Our perception is that effective communication of advances in our field has been insufficient, and the conclusion is that we also need to improve the transfer of our proposals to the professional environment. It is crucial to rediscover and value the classical contributions in bibliometrics.

With this purpose in mind, this book aims to serve as a guide that links evaluative bibliometrics with other practices that support science policy-making and research managers. This book, due to its size and orientation, is not intended to be a manual but rather a guide, a starting point that introduces readers to the fundamentals and methodologies that can be developed in evaluation units. These guidelines, which have maintained their essence over time, require an organized compilation; they need an "aggiornamento". Our objective is precisely to address these issues from an updated perspective, seeking to transcend technical debates and, above all, providing clear guidance on how to optimally use bibliometric indicators and inviting multiple actors in evaluation systems to read these proposals.

It should also be noted that there is no specific work that tells us what the principles of evaluative bibliometrics are; there is also no precise and concrete consensus on these principles. That is the specific and ambitious task of this work. Therefore, we present general and integrated bibliometric principles, but each of them should be conceived as an ontological cluster that brings together multiple sub-principles and recommendations. We do not seek to provide a fixed list of rules but rather to suggest sets that can be adapted. Unlike other manifestos or decalogues, we do not promote a closed, fixed, and immutable structure but offer an umbrella that provides an open and flexible framework. We also want to vindicate our protocols and services that, as we will demonstrate, have always been committed to developing rigorous and responsible methodologies; this is not new.

To achieve this, we will rely on the intellectual contributions and scientific literature generated over more than forty years (1976-2023) by the founders and members of the Centre for Science and Technology Studies (CWTS), who have been responsible in Europe for describing the basic procedures of evaluative bibliometrics. Therefore, continuous references are made throughout the text to Van Raan or Moed. An exhaustive bibliographic review of the contributions from the mentioned period and contemporary authors who follow their direction has been conducted. This work also seeks to incorporate the ideas exchanged with Moed during his countless stays and visits to the University of Granada. Therefore, we venture to recover, reinterpret, and systematize these ideas, many of them discussed in informal settings. This book, in a way, acts as a tribute to his figure and seeks to continue his legacy. We have also tried to integrate our own vision from multiple roles in bibliometrics, such as promoters and managers of bibliometric units, as well as researchers, consultants, and trainers in bibliometrics, thereby offering a much richer and broader perspective. This text is above all a synthesis of those conversations and the experience gathered in recent years.

We have structured the work into two parts. First, an introduction tracing the origin and systematization of evaluative bibliometrics. Second, in the main body, we break down the five principles we have established, dedicating a specific section to each of them. Regarding the use of technical nomenclature, indicator names, or database names, we have followed a simple guideline to facilitate reading and understanding: they have been minimized, opting for generic terminology that allows for smooth reading. Therefore, we do not talk about CPP/FCS but about normalized impact indicators; we reduce mentions of commercial products like Web of Science and Scopus and refer to citation indexes or simply databases. This avoids the tongue-twister that works in our specialty often become. It also has a dual intention: to demonstrate that the principles are independent of the scientific and technological context and to ensure that this text has some timelessness and is not subject to constant updates due to the numerous changes occurring in the world of scientific information. Finally, to facilitate the assimilation of the principles, they are accompanied by small practical cases that minimally illustrate what is explained. The book you, the reader, have in your hands is a plain and sometimes very personal explanation of the world of evaluative bibliometrics.

2. INTRODUCTION

Definition and origins

There is a consensus that bibliometrics emerged in the first third of the 20th century, with significant contributions from Coles and Eales in 1917, Gros and Gros in 1927, and Bradford in 1934. However, some studies suggest evidence of a proto-bibliometrics, tracing it back to the 18th century (Hammarfelt and Hallonsten, 2023). The field of bibliometrics started to take shape after World War II. Key figures in this development are the contributions of Ranganathan in 1948, who was one of the first to mention Librametry, and later the contributions of Derek J. de Solla Price in 1965 and Alan Pritchard in 1969. It was Pritchard who, in his work "Statistical Bibliography or Bibliometrics?", coined the term "bibliometrics," defining it as "the application of mathematical and statistical methods to books and other media of communication." While this definition remains relevant, it is worth mentioning that evaluative bibliometrics did not appear until the 1970s, focused on the analysis of publications and citations to evaluate research activity. This development in bibliometrics occurred alongside the emergence of quantitative science studies in 1966 in Eastern Europe (Nalimov, 1966), a broader field that encompasses the study of the science system itself, under which bibliometrics is considered a subfield aimed at creating research performance indicators based on publications.

When referring to evaluative bibliometrics, we highlight two essential characteristics. First, it is applied to the analysis of scientific activity. Second, its results are explicitly intended to assist on the assessment of scientific performance. Therefore, evaluative bibliometrics is primarily concerned with supporting decision-making in the realm of science policy. Conversely, a report, memorandum, or other informative document not aiming to inform decision-making, will fall under the scope of descriptive bibliometrics. One of the main distinctions between evaluative and descriptive bibliometrics lies on data collection (van Leeuwen, 2004). Decision-oriented indicators involve a verification process in which the evaluated researchers supervise their collected published output, ensuring the reliability and validity to the process. This is not necessary in the case of descriptive bibliometrics. Another inherent characteristic of evaluative bibliometrics is its participatory nature. This approach involves at least three main actors: policy makers, the peer review process, and bibliometrics experts. In a research context, this implies that evaluative bibliometrics has a deliberate focus on science policy.

We find many variants in the literature referring all to the concept of evaluative bibliometrics. For instance, Van Raan refers to it as "advanced bibliometric" as opposed to amateur bibliometrics (Van Raan, 1996), recognizing "contextualized scientometric analysis" as pertinent (Van Raan, 2019). Moed defends the use of "evaluative informetrics," a nomenclature that encompasses a broader spectrum, including aspects such as open access and altmetrics (Moed, 2017). Another term employed by Moed has been "quantitative studies of science and technology", although here he includes also links with disciplines such as sociology, history, economics, and information management.

In Spain, the predominant term is bibliometrics, a concept that most associate with scientific evaluation without further clarifications. The adoption of other terms, such as scientometrics or informetrics, which encompass different areas of interest and have distinct nuances (Brookes, 1990), has been confined to the academic sphere and is less familiar to the general public. Given the preference of our communities for the term bibliometrics, and considering its specific nuances, we have opted for this denomination, which, it should be noted, Cronin also recommended using (Cronin and Sugimoto, 2014).

Evaluative bibliometrics has a defined and incontestable origin. This origin lies in a report carried out for the National Science Foundation (NSF) by Francis Narin, founder of Computer Horizons Inc, published in 1976 under the title "Evaluative Bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity" (Narin, 1976). During the 1970s, Narin and his collaborators began a study on the relationships between basic research and patents. This inquiry led to the design of methodologies that enabled the quantification of the impact of research on technological innovation. However, Francis Narin's notoriety is not limited to his theoretical contributions; he is also recognized for his skill in developing applied techniques that provided the ability to measure scientific production more effectively. Specifically, Narin explored the structure of disciplines through citation analysis, offering a detailed perspective on how various domains of knowledge interconnect and transform over time. Narin emphasized bibliometrics as a novel measurement mechanism that allowed managers to make objective and rational judgments about scientific effectiveness, complementing peer review, which was the predominant evaluative mode until that time.

To bring his ideas to fruition, Francis Narin used the predominant tool at that time: the Science Citation Index (SCI). This groundbreaking database was conceived by another luminary in the field, Eugene Garfield, with whom Narin maintained a close professional relationship. Although Narin did not directly contribute to the creation of the SCI, he was among the first to recognize the intrinsic potential of citation indexes. Interestingly, in 1966, Nalimov in the Soviet Union also identified the value of these data and their statistical analysis (Rousseau, 2021), highlighting the universal significance of the SCI. The SCI represents the foundation of evaluative bibliometrics (Krüger and Petersohn, 2023) as it is reasonable to assert that without this database, advancements in this field would have been significantly delayed due to the technological restrictions of the tools available at the time. Garfield overcame these limitations by transitioning from a printed edition to a magnetic format, followed by CD-ROM, and finally to an online digital platform—a transition his competitors adopted after considerable delay.

Thus, in its early stages, Garfield monopolized the field of evaluative bibliometrics, controlling the main technologies and data, influencing the creation of a whole new family of citation indexes (SCI, SSCI, AHCI, JCR), the design of the first indicators used for evaluative purposes, i.e., the Journal Impact Factor (JIF), or information retrieval techniques like the Cited Reference Search. Hence, Garfield and Narin can be regarded as the foundational architects of the American school of evaluative bibliometrics. They played a key role on obtaining the recognition and validation of the scientific community by engaging with American science historians and sociologists of science, such as Derek de Solla Price or Robert K. Merton. But the debate surrounding the application of indicators within the framework of science policy intensified in the late 1970s (Garfield, 1979), eventually crossing the Atlantic and driving the europeanization of evaluative bibliometrics.

The legacy of CWTS

In parallel to the American school, a marxist tradition of science studies was developing in the Soviet Union (Bensman & Kraft, 2007). Garfield was well aware of it, and actually credited Nalimov for the term scientometrics. Successor of the marxist tradition was Tibor Braun, a chemist like Garfield, and a member of the Hungarian Academy of Sciences. He was responsible of the merging of these two schools through the foundation in 1978 of the journal Scientometrics. Tibor Braun played a key role in the field of bibliometrics, articulating around this journal a strong community of scholars worldwide which now had a venue for exchanging methodologies, advancements and practical case studies. He managed to circumvent the strong opposition from the Soviet Union and led, in the middle of the Cold War, an international journal edited by the Hungarian Academy of Sciences with an editorial board which included researchers from the US (such as Garfield and Price) and the USSR (e.g., Dobrov) among other countries.

The second milestone is the establishment of the Center for Science and Technology Studies (CWTS) at Leiden University (The Netherlands). The influence of CWTS has been crucial, positioning itself as a pioneer in the design of methodologies and professional guidelines, which, in parallel, led to a renewed academic community centred almost exclusively on evaluative bibliometrics. Its international recognition as the organization that expertly institutionalized the use of evaluative bibliometrics is evident, setting the pace in Europe (Petersohn & Heinze, 2018). The origin of the centre lies in a working group that, in the late 1970s, sought to use bibliometric indicators as a tool to guide the allocation of funds based on the publications of various faculties of Leiden University. In its early days, the CWTS shone with the prominent figures of Antony van Raan and his then pupil, Henk F. Moed. During these early years of the 1980s, their research focused on four aspects: (I) the normalization of indicators, (II) the comparison of researchers with international contexts, (III) the identification and evaluation of young and emerging groups, and, especially, (IV) the construction of local databases derived from the products of the Institute for Scientific Information.

Simultaneously, in the United Kingdom, Ben Martin and John Irvine, associated with the Science Policy Research Unit at the University of Sussex, tested the use of bibliometric indicators to assess radio astronomy institutes, concluding that bibliometric indicators and citations were adequate support measures to be used in the evaluation of basic research (Martin & Irvine, 1983). Returning to the Netherlands, in 1989, the group led by Van Raan adopted the official name of CWTS, and two years later, in 1991, this centre was awarded the first chair of quantitative science studies in the Netherlands, assumed by its director, Anthony Van Raan. CWTS established itself as a prominent research unit, expanding its team, refining its methodology, and developing its own database. During the 1990s, it was already recognized as a reference in evaluative bibliometric analyses for governmental agencies in the Netherlands and Flanders. The year 1994 marked an expansion for CWTS, as it participated in the evaluation protocols of the Netherlands and provided approximately 90% of the advanced bibliometric analyses requested by universities. During this period, CWTS reinforced its principles, advocating for bibliometrics as primarily a diagnostic tool.

In 2002, CWTS B.V. emerged as an autonomous spin-off, seeking greater flexibility. This advancement became evident in 2007 when the Higher Education Funding Council of the United Kingdom entrusted it with the responsibility of carrying out the bibliometric analyses for the British Research Assessment Exercise. Additionally, CWTS enriched the field through the publication of the "Handbook of Quantitative Studies of Science" (Moed et al., 2004). In 2010 it merged its “Leiden” STI series of conferences (started in 1988) with the ENID series (started in 2005), formalizing the creation of the European Network of Indicators Designers (ENID) association a year earlier. This allowed them to establishing a theoretical-practical framework and an academic exchange space. From 2008 onwards, with an annual funding of 1.5 million euros, CWTS consolidated itself by establishing a chair in science policy and a doctoral program in evaluative bibliometrics. Not free of controversies, such the heated debates around the field-normalized citation scores developed, it continued revolutionizing the field proposing novel solutions, including science mapping software. Competitors emerged both, from academic spheres as well as bibliometric data providers, namely Thomson Reuters’ Web of Science (now Clarivate) and Elsevier. Faced with this new scenario, CWTS diversified its areas of action: it integrated advanced analytics, experimented with new indicators such as altmetrics and social impact indicators, diversified its portfolio beyond bibliometrics, and strengthened its international presence.

In Spain, the influence of the European school and the leaders of CWTS is evident and reflected through two simultaneous and fundamental events. In 2006, the University of Granada and its Faculty of Library and Information Science hosted Henk F. Moed, Ed Noyons, and Clara Calero-Medina, invited by the EC3 Research Group (Science and Scientific Communication Evaluation) to deliver a seminar. This event triggered a long-standing and still ongoing collaboration link, both formal and informal, between the Spanish scientific community, especially with Granada, and CWTS. Additionally, the Dutch influence in the Spanish sphere was recognized and strengthened when Moed was appointed as the conference chair in charge of the scientific program and its proceedings at the 11th International Society for Scientometrics and Informetrics Conference (Madrid, June 25-27, 2007). This event marked a milestone in the international recognition of the Spanish research community at the time. For Daniel Torres-Salinas, a stay in the Netherlands, proved pivotal in his career. Many of the ideas discussed with Henk Moed were later consolidated into one of the earliest studies on the role of bibliometric units within Spanish universities (Torres-Salinas and Jiménez-Contreras, 2012). This work adhered to the theoretical framework of evaluative bibliometrics and strongly advocated for best practices, as we will elaborate next.

Professional perspectives

CWTS led to a professionalization of the field, mainly through the commercialization of standardized bibliometric reports. Although our origins are closely linked to the commercial world, as demonstrated by the Institute for Scientific Information and Computer Horizon, it has been the Dutch who have managed to develop an economic model characterized by its adaptability, affordability, and versatility, suitable for a wide variety of evaluative contexts. More interestingly, balancing this commercial interest with a thriving research environment. Many companies have emerged imitating this model such as Science-Metrix (now part of the Elsevier company portfolio), Technopolis, Digital Science, or in the Spanish case, EC3metrics and SCImago. These firms provide governmental and higher education organizations tailored evaluative and monitoring reports based mainly on bibliometric data. In this way, institutions ensure that indicators are provided by an independent external actor, complementing internal peer review processes. But there is an alternative model to this, that is, the internal management of indicators through the creation of agencies and departments within the organizational structure of these institutions. In this model, bibliometricians are integrated into their daily operations and organizational structure, with departments for evaluation and science policy being part of ministries of science, funding agencies and/or universities (Heinze and Jappe, 2020).

According to this model, research organizations such as universities and hospitals produce bibliometric reports through the creation of research support services. In some instances, these units are part of the library services and are explicitly named as scientific evaluation units or bibliometric units. One of the pioneers of this model was the Department of Bibliometrics and Publication Strategies at the University of Vienna, which participates in faculty hiring with its own methods and evaluative reports (Gorraiz et al., 2020). Its model has been a reference for European professionals in recent years, with its founder, Juan Gorraiz, promoting the dissemination of good practices through the European Summer School for Scientometrics (ESSS). Spain adopted a similar approach, although combined with CWTS practices, by integrating researchers and managers in these services, as reflected in the publication "Towards Bibliometric Units" (Torres-Salinas and Jiménez-Contreras, 2012). This text outlines the essential functions these units will carry out. Among others, two are directly related to evaluative bibliometrics, meaning they are directed at supporting decision-making.

The first function involves preparing bibliometric reports with two key objectives: supporting the development and management of strategic plans and assisting in the preparation of reports for funding calls, such as the establishment of institutes or the allocation of financial resources among research groups. The second function focuses on participating in the management and evaluation of research plans, which serve as mechanisms for allocating funding to researchers through various programs. These plans require the establishment of quantitative criteria aligned with national science policy to evaluate research promotion initiatives, including human resources, research projects, and incentive programs. This involves evaluating these initiatives and preparing reports that are subsequently reviewed by internal committees and, in some cases, the researchers being evaluated. Institutional plans generally lay the foundation of institutional scientific policies. Therefore, active participation in their management is essential to position bibliometric work within the strategic framework, ensuring it is seen as a decisive tool rather than merely a descriptive one.

Many of these functions have been assumed by librarians, allowing them to expand their roles, actively participating in the development of new strategies and innovative services for both academic and administrative staff, and reinforce their position within the institution (Gumpenberger et al., 2012). This is particularly evident in Spanish universities, where most of these units fall under the library services. This phenomenon has had two main consequences: first, bibliometrics has been perceived as an additional competence within the library field; and second, it has led to a professional approach closely linked to the skills provided by Information & Library Science degrees. Thus, we have delineated our services largely based on our training and direct experience with users who are the main clients of libraries. This orientation has emphasized tasks such as bibliographic control and standardization, as evidenced by the management centred around Current Research Information Systems (CRIS) and research support services. As a result, specific functions associated with evaluative bibliometrics have been relegated, limiting our contribution to decision-making. From our point of view, it is essential that the skills of the bibliometrician expand to incorporate more intensive roles related to management, evaluation, and advice on science policies. We cannot continue operating merely as providers, as the success of these units will lie in their ability to actively engage in management and provide relevant and effective solutions.

Additionally, one of the most pressing challenges in our profession is the practice of bibliometrics without adequate training. This issue arises when individuals rely excessively on pre-computed indicators offered by bibliometric data providers without the necessary rigor or expertise. This frequently results in tables filled with numerous indicators that are easy to obtain but whose calculations and implications are not fully understood. Terms like "amateur bibliometrics" and "desktop bibliometrics" (Gläser & Laudel, 2007) have been coined to describe this non-professionalized approach, typically undertaken by managers, librarians, and researchers lacking specific training in this area. These superficial practices are often mistaken for evaluative bibliometrics. With the diversification of evaluative methods and the increasing number of actors conducting bibliometric analyses, it has become essential to clearly define the profile of a professional bibliometrician. This expert should combine a technical dimension, focused on generating reliable data for evaluations, with a more management-oriented component. In the context of evaluative bibliometrics, bibliometricians play a critical role in designing, conceptualizing and applying contextualized indicators. This includes preparing reports and evaluating plans and programs for R&D organizations, positioning themselves as key figures in shaping evaluative practices both with rigor and relevance.

Bibliometric indicators

Bibliometric indicators currently constitute the core of our work: they are our main output and an essential tool in the decision-making process. This means that most of our actions revolve around the use of indicators, including how they are defined, collected, validated and, more importantly, used. A poorly designed, computed, or used indicator will reflect a distorted reality, and will mislead the interests of whichever institution is using it. One of the challenges we currently face has to do with the proliferation of pre-calculated bibliometric indicators. While they facilitate their immediate implementation, they impede users to decompose or recalibrate them according to their specific needs and context. It is essential to understand how an indicator is formulated, how it can be reproduced, and above all, how to interpret it in an evaluative framework. Therefore, one of the bibliometrician's tasks is to propose a coherent set of measures to our users along with potential use cases and cautionary notes.

Maintaining a continuous balance between indicators' potential and limitations means having the necessary ability to select those that inform us of a specific aspect in the context of our evaluation. Indicators like the Journal Impact Factor or the *h*-index are not inherently more deficient than other bibliometric indicators. The key lies precisely in understanding when and how to deploy these indicators for a well-founded evaluation. We must bear in mind that no indicator is better or worse than another; no matter how established or novel it is, it simply offers a lens through which we observe a particular facet of the academic reality. None provide a definitive and monolithic diagnosis.

In any discipline, we must make inferences and judgments based on data that are, by nature, limited. This premise is not foreign even to the exact sciences, where measuring and calibrating instruments often capture secondary or partial aspects of the phenomena studied, and they do not lack precautions in their use (Robinson-Garcia et al., 2024). Van Raan offers a very illustrative metaphor: "*Talking about lamps and light, you better may say, bibliometric research is like astronomy. What is wrong about using the light of stars in galaxies to measure their intensities, to perform spectroscopic measurements?*" (van Raan, 1999, p. 417). Stars are not reduced to their light, just as scientists are not limited to their publications and the citations they receive. Although indicators give us an approximation to these realities, they only offer a partial view, not a replacement of reality itself (Martin & Irvine, 1983). Depending solely on these indicators without a critical analysis or disregarding contextual factors, we run the risk of falling into what has been Moed referred to as "magical thinking" (Moed, 2017). As in all scientific disciplines, bibliometric indicators are used as inferential and symbolic proxies, never to be considered definitive, and hence disregard magical attributes to what they measure.

The question that follows then is: What aspects of reality do bibliometric indicators reveal? Many authors have tried to answer this question talking about impact and outreach in specific audiences, leading us towards the concept of scientific interest or audience. Garfield already pointed out: "*we know that citation rates say something about the contribution made by an individual's work, at least in terms of the utility and interest the rest of the scientific community finds in it*" (Garfield, 1979, p. 372). Although Garfield links citations to scientific interest, the increasing openness of science to other platforms and measurements invites us to go beyond the scientific community and include non-scientific ones. Bibliometric indicators will reflect the influence of researchers in a particular domain, whether in a social or scientific context. This influence would be based both on the findings and results as well as on the ability to communicate them to the relevant public. In this way, by considering indicators as reflections of audiences, the importance of interpreting the indicators according to the audience they are directed at is highlighted. This means that a high number of citations in scientific journals does not necessarily translate to a high impact on society in general, and vice versa. It is important to understand what the indicators measure and determine whether they reflect the effort a researcher has made within an institution to meet its objectives.

The type of audience we want to capture and the adjustment to the measurement of objectives should help us decide on the indicators we will use. However, we must also consider technical issues here. When addressing the choice of indicators, there is an evident tendency for advocating for normalized indicators, which conversely works against the demand for simpler indicators. Managers often prefer simple measures because they can easily grasp their meaning; metrics like the *h*-index or the percentage of Q1 documents can be more understandable to a non-expert audience. Still, they do not capture the depth and precision required in an evaluation. It is essential to balance two factors: simplicity and accuracy. From a technical standpoint there are other aspects to consider, such as size-dependence, document types, or the attribution of credit, whether through full or fractional counting. These are just some of the factors to consider, for which deep technical knowledge is required, given the impact their use will have in people's lives.

The application of bibliometrics and its impact in research is complex and multifaceted. Indicators often generate unintended effects due to their reflexive capacity; that is, their mere existence can influence the direction and nature of the research that academics choose to undertake. There are multiple and well-known bad practices that are derived from the misuse of bibliometrics: from excessive self-citation, selecting topics based on the most 'citable' areas at the expense of innovative topics, to tactics like "salami publishing" where a single study is divided into multiple publications to increase the total number of articles. The problem has worsened and scaled up, with Saudi universities purchasing affiliations of highly cited researchers to boost up their position in rankings or the disgraceful expansion of journals citation cartels. Still, criticism should not be directed towards the field of bibliometrics, but towards those who misuse it. Hence, control mechanisms should be placed not only to point at those making an inadequate use of the metrics but also punishing those who engage in malpractice to manipulate indicators for their own benefit. The responsibility of bibliometricians is to consistently warn about and advocate for awareness of the limitations and consequences of using these indicators.

From the very first moment the Journal Impact Factor was introduced, pioneers such as Eugene Garfield have spent decades trying to emphasize that the JIF was never designed to be used as an instrument for individual researcher evaluation, a point further demystified by colleagues such as Wolfgang Glänzel who noted that "the possibility of measuring the scientific quality of individual publications through citations alone is a myth" (Glänzel, 2008, p. 6). Despite these repeated calls for caution, many administrators in the academic world have chosen to ignore them. The specialized bibliometric literature supports this concern, dedicating entire articles to highlight the restrictions on using these indicators. The history of bibliometrics is presented as one of a tool with great potential but frequently misunderstood. This misuse, often the result of uninformed decisions, has led to blaming the technique instead of critically analysing what led to a particular selection of indicators. It is essential that decision-makers approach bibliometrics with a constructive mindset, understanding that, far from being a magical wand, indicators, when used appropriately, provide significant observations, but if used incorrectly, can divert the course of science. This is why we must establish principles we should abide by, not only us, but more importantly, managers and decision makers.

Approach to the principles

The various manifestos we all know, such as Leiden (2007) and DORA (2012), as well as the new alliances and evaluative frameworks, like CoARA, were anticipated decades ago in the literature. Historically, we have observed that the modus operandi is established by CWTS, through countless publications led by Van Raan and Moed from the early 1980s to the present. We dare to assert that the work titled "The use of bibliometric data for the measurement of university research performance" (Moed et al., 1985) begins to configure the principles unequivocally. This publication should be considered one of the seminal works in evaluative bibliometrics. The original message can still be used with some adjustments for modern times. This is not a step backward; rather, it reflects the enduring truths that form the foundation of our field's tradition and serve as a basis for building future credibility. In the mentioned work, emphasis is placed on five basic principles:

- We must carry out a very careful data collection process that is also verifiable and guaranteed.
- We must work jointly in a coordinated fashion with managers and other evaluators, such as field experts.

- We must critically evaluate our own work, emphasizing the need to discuss methodologies and validate bibliometric indicators.
- We must produce reliable databases and tools, which means documenting every step and ensuring transparency.
- Evaluative bibliometrics should always be viewed as a complementary tool, not the central focus of evaluations.

Moed has been a key figure in this field, devoting most of his oeuvre to establish an appropriate evaluative framework (Moed, 2009). Below, we show another example, this time composed of seven recommendations related to three fundamental facets: open and transparent communication, academic and professional integrity, and the accuracy and relevance of the analysis. They are as follows:

- **Formality.** Before conducting an evaluation, evaluators, academics, and institutions must be informed about the use of indicators as sources of information.
- **Transparency.** Individuals subject to bibliometric analysis should have the opportunity to verify data accuracy and provide contextual insights to interpret quantitative results effectively.
- **Academic basis.** Bibliometric results must be presented within an academic framework, addressing their validity, theoretical assumptions, and explicitly defining their scope and limitations.
- **Expert focus.** Analyses should be complemented by a thorough understanding of the evaluated content, contextual factors and research objectives.
- **Policy clarity.** Indicators must be applied within a clearly defined evaluative framework with objectives transparently communicated to all participants.
- **Explicitness.** Users should explicitly define the concepts of academic quality and detail how these are considered in the evaluation process.
- **Interpretation.** Indicators provide partial insights into specific topics; they should be seen as tools for interpretation, and not mere formulas yielding automatic results.

Finally, we want to complement these statements with the proposals made by Van Raan (2019), who synthesizes the previous recommendations into three fundamental principles:

- **Reliable statistics.** Van Raan urges us to reflect on the need for normalizing publications and citations, considering the heterogeneity of scientific fields and carefully addressing outliers, especially when dealing with indicators that follow distribution patterns like the Pareto principle.
- **Precision of the data.** Van Raan emphasizes the need for following a meticulous data cleaning process, especially with regard to homogenizing author and institution names.
- **Diversity and consistency of indicators.** In this regard, van Raan states that transparency is not just a recommendation, but a mandate.

Van Raan stresses the importance of a rigorous understanding of the techniques, sources, and processes in bibliometrics, drawing a clear distinction between high-level evaluative bibliometrics and amateur bibliometrics. Revisiting this topic is crucial to discern the practices and principles that differentiate rigorous bibliometric approaches from superficial and poorly executed ones. In this context, Moed, in one of his last writings (Moed, 2020), offers a valuable insight into how evaluative bibliometrics, in its more form, contrasts with the so-called "desktop bibliometrics". Table 1 illustrates the main differences between these two approaches, emphasizing that while they may appear similar on the surface, their underlying principles are fundamentally different. Desktop bibliometrics focuses on pre-calculated decontextualized indicators lacking any kind of validation. In contrast, evaluative bibliometrics adopt a more adaptive and contextual approach, using validated and adjusted indicators according to their evaluative framework. While desktop bibliometrics focus on quantitative measures such as rankings, evaluative bibliometrics enable more nuanced analyses, incorporating institutional context, goals and expert opinion.

Table 1. essential differences between the principles of desktop bibliometrics and evaluative bibliometrics

Aspect	Desktop Bibliometrics	Evaluative Bibliometrics
Availability	Uses standardized and pre-calculated indicators on platforms	Indicators are checked, validated, and customized
Framework adaptation	Little reflection on whether indicators are the most suitable for the framework	Indicators are designed based on the institutional context and goals
Unit	Considers evaluators (departments, researchers, ...) as isolated entities	Decision makers work hand-in-hand with data providers, using bibliometric expertise to support their decisions
Validity	Indicators are not cross-checked or compared with other sources or agents	Indicators are combined and validated with experts, managers, and researchers' opinion. If necessary, indicators are cross-checked with alternative sources
Ordering principle	Data are presented based on the highest value of an indicator, creating rankings and composite indicators	Does not rank. Results are presented according to different institutional and research dimensions
Decision making	Decisions are based purely on indicators' value, prioritizing those ranking the highest	Other criteria are considered to redistribute funding, such as career progression and emerging groups or strategic planning

Since its conception, evaluative bibliometrics has been concerned with ensuring a correct use of indicators in an evaluative setting, consistently denouncing malpractices and misuses. Hence, many of the criticisms to our field seem unwarranted or based on misunderstandings. Many of the contributions presented here have been intentionally ignored in the various manifestos. For instance, it is notable the lack of references to the contributions of these two key figures of the CWTS (van Raan and Moed) in the *Leiden*

Manifesto, despite its ten principles being clearly rooted in their contributions. are based on their contributions. Van Raan himself remarked in *Measuring Science*: "*In the Leiden Manifesto, principles are formulated to guide the use of research metrics, particularly bibliometrics. Most of the guidelines were already discussed 20 years ago (Van Raan, 1996), (Van Raan, 1997)*" (Van Raan, 2019, p.253).

Given that the *Leiden Manifesto* is the basis for all subsequent manifestos, this text aims to avoid a similar oversight. We have reformulated the principles while acknowledging their intellectual origins. To ensure this, we conducted a comprehensive review of the CWTS's historical and contemporary work, along with our own work, as documented in *"Bibliometric Journey"* (Torres-Salinas et al., 2023). This is the foundation of the following five fundamental principles of evaluative bibliometrics:

- Principle 1. Principle of support for decision-making
- Principle 2. Principle of collaboration with experts
- Principle 3. Principle of respect for contexts
- Principle 4. Principle of metric multidimensionality
- Principle 5. Principle of data verifiability and openness

Each of these principles are presented as conceptual groupings, i.e., open categories designed to evolve over time. They are structured sequentially, transitioning from a macro- to a micro-level. The first two principles address the philosophy of work and interaction with key stakeholders, including policymakers, establishing the framework for collaboration with other specialists. The next three principles delve more deeply into bibliometric practices, addressing issues such as how to adapt to different evaluative frameworks and institutional contexts, selection of indicators or issues related to proper data management. This structure reflects the book's overall logic: starting with theoretical reflections and concluding with practical applications, guiding readers from foundational principles to actionable insights.

3. FUNDAMENTAL PRINCIPLES

Principle 1. Principle of support for decision-making

The evolution of science policy, rooted in pivotal moments of the 20th century, such as the Manhattan Project, can be understood as a strategic framework of decisions and actions aimed at directing and stimulating research according to pre-established socioeconomic and cultural goals. This policy arises from the need of nations to integrate science and technology into their respective national agendas, recognizing their impact on economic growth, social progress, and national security. But science policy is not solely the domain of governments. Research institutions, universities, and industrial sectors also implement policies to align scientific output with contemporary demands and challenges. Thus, science policy emerges as an essential tool to translate scientific knowledge into practical applications. A cornerstone of science policy is the decision-making process, which relies heavily on objective criteria and evidence to guide actions. Here, bibliometrics plays an essential role, offering metrics to assess the effectiveness of such policies and support decision-making.

Evaluative bibliometrics is best understood as a tool available to policymakers rather than as a policy or the entirety of the evaluation process. It is a component within a broader system, influenced by multiple actors and factors. Table 2 illustrates the interactions between four main actors in science policy and the factors influencing decision-making. Political actors set the direction and orientation of science policies, while evaluative actors provide technical and specialized expertise. Within the evaluative group, experts and peer reviewers conduct a thorough evaluation leading to qualitative reports, while bibliometric specialists are responsible for collecting relevant data and generating their indicators. These groups, the political and the evaluative one, are interdependent: political and administrative decision-makers will establish and modify policies, while relying heavily on the evaluations and information provided by reviewers and bibliometricians.

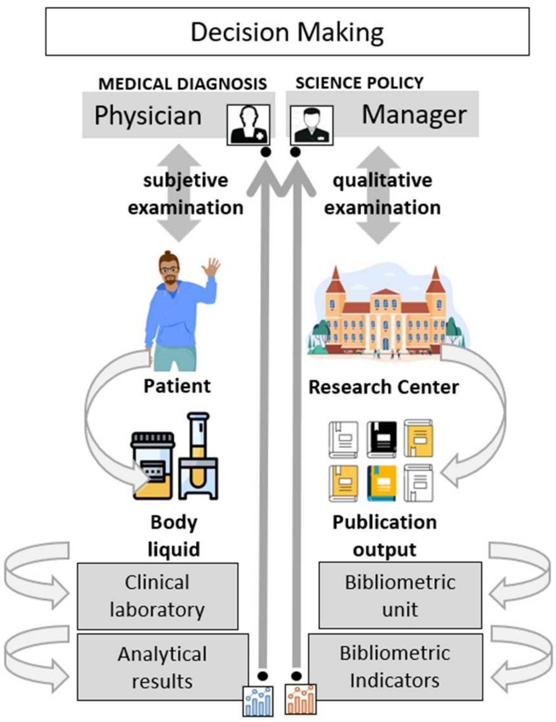
Table 2. main actors and conditioning factors in the evaluation framework

Involved Actors	Conditioning Factors
POLICY ACTORS	*+ Personal values + Social context* ↑↓ ↑↓
Politicians - Decision-making	Policies and promises Legislation at all levels regarding science
Administrators / Managers - Setting priorities - Preliminary decision-making	Institutional regulations Budget allocation
EVALUATIVE ACTORS	
Experts (peer review) - Final evaluation - Qualitative reports	Specific criteria for scientific evaluation Selection criteria
Bibliometrics Experts - Quantitative reports - Data collection	Methodological and analytical capabilities Information sources and data access

The management of science policies operates within an intricate structure involving diverse actors with distinct roles and responsibilities. Additionally, multiple factors will condition the decision-making process. Evaluative actors are not only subject to the methods and sources at their disposal but also to pre-established criteria and allocated resources. Political and managerial actors operate within a regulatory and legislative framework that determines the entire system. Within this system, bibliometric specialists are mere technicians in charge of providing accurate analyses and indicators within the framework outlined by policymakers. Political actors, as final consumer of this information, will weigh both the relevance and objectivity in the context of decision-making, assuming the responsibility of such decisions. While bibliometricians' contributions are largely technical and informative, their neutral stance enables them to maintain independence and objectivity. Ultimately, political actors hold the prerogative to determine which criteria to prioritize, underscoring their decisive role in the policy process. This dynamic ensures that evaluative bibliometrics remains a vital yet impartial component of science policy decision-making.

Tibor Braun (1999) proposed comparing the decision-making process in a research centre (university, hospital, institute, etc.) with a medical diagnosis to better understand the role of bibliometrics and how it intertwines with other actors in the evaluative process. Figure 1 illustrates how a medical diagnosis begins with a direct and superficial examination of the patient by the doctor. This first approach is enriched with the collection of various samples, such as blood or saliva, which are sent to a specialized clinical laboratory. From a blood test, for example, data such as the count of red and white blood cells are obtained. Once these results are obtained, they are sent back to the doctor who, combining the information from the initial examination and the laboratory results, determines a diagnosis and, if necessary, prescribes a treatment.

Figure 1. Decision-making process compared to a medical diagnosis, Braun (1999)



Within the university environment, bibliometric reports have become an essential tool for diagnosing the overall state of their scientific activity. It is in this context where bibliometrics can provide concise answers to fundamental questions. According to van Now imagine that our patient is a university seeking resources to support its research lines and allocate funds accordingly. The university administrators have formed a committee composed of experts from various fields to conduct an initial examination of the institution's research profile. However, the collected information is not entirely conclusive. Hence the need to analyse in-depth the scientific output, both in terms of number of publications and citation impact, in the various areas of the university. This task will be carried out using publications indexed in recognized databases. These publications are sent for analysis to a bibliometrics unit, which operates with the objectivity and precision of a clinical laboratory. This unit produces a bibliometric report, with a series of indicators reflecting the research status by areas of the university. With this information in hand, along with the committee of experts' recommendations, university managers would be able to make an informed decision.

Raan (Van Raan, 1999), ensuring the relevance of these indicators in decision-making requires addressing questions related to the quality of our organization regarding its scientific performance, such as scientific influence or "impact," the position of the organization in comparison with similar institutions and the type of knowledge it fosters. These are questions to answer but, Figure 1 shows, decisions are not based solely on indicators. Once the bibliometrician reports their findings, indicators serve a complementary role, functioning as an advisory resource rather than taking centre-stage in the decision-making process. The final responsibility for the decision must fall on the committees and managers, who must evaluate and synthesize all the available (quantitative and qualitative) information before taking action (Hammarfelt and Allanten, 2023).

When considering bibliometrics, the range of decisions it can inform is vast, raising the essential question: What types of decisions can it support? These can be classified according to the nature of the scientific evaluation undertaken, falling into two categories: ex-ante and ex-post evaluations. Ex-ante evaluations take place before the deployment of a program, focusing primarily on determining the feasibility or potential of the actors involved in the program. An illustrative example would be a projects funding call. Project proposals must be pre-selected for funding based on their feasibility and promise. Conversely, ex-post evaluations analyse and inform the performance of decisions that have already materialized. An example is the evaluation of individual’s research performance in periodical evaluation assessments during tenure track, where a set of expectations have been formulated and have to be meet during the evaluation process. While the distinction between ex-ante and ex-post evaluations may not be relevant at the unit level, understanding these different decision-making approaches is essential in the broader academic context. Each type of evaluation serves a unique role in guiding decisions, and so the role and importance of indicators will vary.

There are numerous examples in relation to ex-ante evaluation approaches. Table 3 shows some of them. But let us clarify the concept by looking at two specific cases and the consequent decisions that science policy managers can adopt:

- Selection of scientific personnel: Candidates' scientific curricula will be analysed in detail. Indicators are used at both the author level and the article level, applied to a pre-selected group of applicants. The primary decision in this scenario is the selection of a candidate who will be recruited as university staff.
- Design of strategic research plans: Here, we will carry out an analysis to identify the most relevant scientific fields based on a multidimensional selection of indicators. The critical decision here is to choose which specific areas will be funded in the coming years.

Table 3. Examples of ex-ante evaluations and associated decisions

Type of ex-ante Evaluation	Potential Decisions
Evaluation of research project proposals	- Approve or reject funding - Allocate resources to each project - Determine the scope and duration of the project
Feasibility analysis of new scientific programs	- Establish or dismiss new programs - Define thematic priorities - Allocate budget and resources
Review of proposals for new scientific journals	- Approve creation based on editorial market - Determine thematic orientation - Decide on committee, format, and frequency
Evaluation of international scientific collaboration agreements	- Sign, modify, or reject agreements - Determine areas of collaboration - Allocate resources and responsibilities
Evaluation of emerging research groups	- Grant funding - Provide mentorship - Integrate into research networks

Hence, the fundamental purpose of ex-ante evaluations is: 1) to guide policies, 2) highlight priority focus areas, and 3) establish robust mechanisms for monitoring and assessment. Therefore, the decisions described in Table 3 will eventually require a verification process, as it is essential to discern the outcomes following the implementation of a specific action and determine whether that action has been successful or has encountered obstacles. In this sense, ex-ante and ex-post evaluations are not mutually exclusive. Rather, they often act as two sides of the same coin and frequently appear sequentially. Table 4 shows examples of evaluations where previous scenarios have transformed into ex-post evaluation scenarios. Let us go back to our previous cases:

- Tracking researchers' trajectories: The objective is to discern whether they have met the previously established goals. Through bibliometric analysis, it is possible to evaluate the degree of integration of the researcher into the international landscape. Additionally, the analysis of co-authorships provides insight into their competence in establishing productive collaborations. Based on such evaluations, administrative decisions could include salary adjustments or promotions to higher categories.

- Evaluation of strategic research plans: At the end of a plan, a retrospective examination is necessary. For example, this examination may look into significant changes in the funded fields with regard to impact and relevance during the period in question. Depending on these findings, subsequent measures could focus on optimal resource reallocation or reorienting efforts for future research strategies.

Table 4. Examples of ex-post evaluations and associated decisions

Type of ex-post Evaluation	Potential Decisions
Evaluation of the performance of funded research projects	- Decide whether to renew or cease funding - Review objectives and goals based on results - Make budget adjustments for the project
Study of the impact of implemented scientific programs	- Evaluate the program's return on investment - Readjust strategies based on impact - Reallocate resources to areas of greatest benefit
Review of the reception and impact of new scientific journals	- Choose to maintain current orientation - Adapt themes based on public interest - Optimize periodicity based on received feedback
Assessment of scientific agreements or partnerships based on results	- Expand collaboration to new areas - Incorporate new collaborators - Readjust agreement objectives
Retrospective analysis of the trajectory of research groups	- Strengthen groups with greater scientific impact - Provide additional resources to emerging groups - Optimize the size of the groups

Here we must note that any decision made on the selection of indicators or in the evaluation process will induce researchers to adopt undesirable behaviours. This could distort the evaluation system, derive on adopting wrong decisions, or even lead to a reform of the whole evaluation system. Indicators not only quantify but influence research practices especially when associated with economic incentives, indicators can become reactive measures (Weingart, 2005). For example, if the evaluation only makes use of articles, researchers could stop publishing books. If the evaluation focuses solely on obtaining funding, scientific dissemination may be neglected. Heads of bibliometric units and, especially, administrators, must weigh the consequences at both individual and institutional levels. Our mission is to inform about these issues transparently.

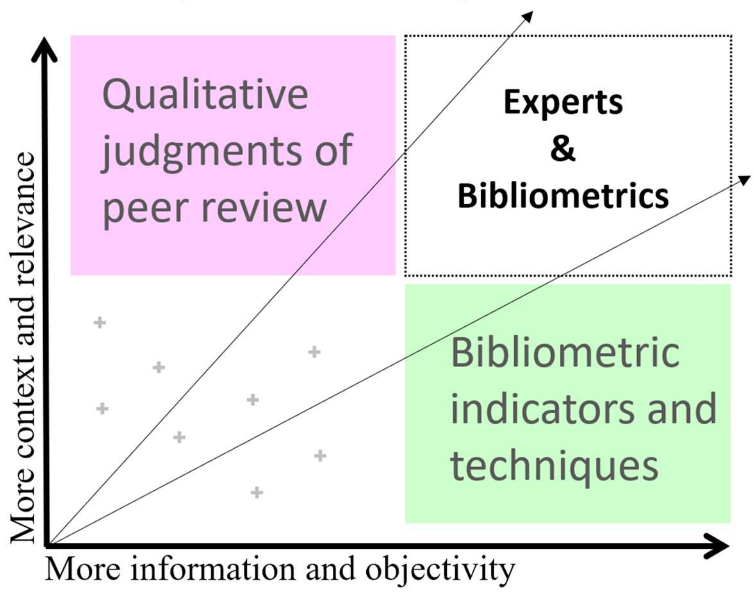
Transparency is a key element of the evaluative process, and it must be accompanied by value free evaluative practices (Moed, 2018). It is common to face pressures and criticisms from managers when they are presented with indicators that contradict their perceptions or decisions they intend to make, expressing their disagreement. Thus, when interacting with administrators, it is not uncommon for indicators to be instrumentalized, aligning them with certain prevailing interests. There have been instances of interference from administrative structures to omit results that highlight negative aspects, exclusively emphasize areas of greater impact, or simply ignore certain results. Furthermore, bibliometric reports can be conditioned by our own perspectives and preconceptions, which is why it is essential to ensure that evaluative bibliometrics are free of personal biases (Moed, 2020). Therefore, we should not allow our personal values compromise the objectivity of what we report on. Conclusions and observations must be neutral and not influenced by opinions. If indicators are biased, they can result in inappropriate decisions with serious repercussions for science policy. Therefore, it is sometimes necessary to abstain from participating in evaluative processes that show signs of corruption or conflict with our ethics and values.

Principle 2: Principle of collaboration with experts

The peer review process involves specialists in a discipline assessing the proposals or results of their colleagues with the aim of rendering a judgment on their adequacy. This review must necessarily find a balance between quantitative and qualitative approaches. Since the pioneering works of Francis Narin in 1976, this integrated vision, in which bibliometrics has played a complementary role to expert review, has prevailed in evaluative bibliometrics. Researchers at the CWTS have reinforced this principle. In practice, bibliometrics often enriches the decisions made by experts, rather than replacing them. However, in reality, bibliometrics many times overshadows the work of specialists, becoming the sole decisive criterion, a situation as undesirable as its opposite. The unilateral use of bibliometrics has led to policies that reject bibliometric indicators because scientific communities feel that such metrics undermine academic autonomy and exercise control over a field. Therefore, evaluative bibliometrics should advocate for a collaborative working framework.

Van Raan has examined the interaction between peer review and evaluative bibliometrics, emphasizing that indicators should not operate independently, but rather in conjunction with expert and peers. In "Measuring Science" (van Raan, 2019), he posits that *"[i]n any judgment there must be room for the intuitive insights of experts*" (p. 27) implying that it is the combination of expert judgment and bibliometric indicators what improves the evaluation mechanisms. This perspective is crystallized in his statement: "*I do not say that bibliometric methods can replace peer-review. We always apply bibliometric methods in the context of peer-review*" (van Raan, 1999, p. 418). He further reiterates that the combination of bibliometric analysis with peer review constitutes the ideal tactic to achieve a well-focused evaluation, warning of the risk of relying unilaterally on bibliometrics. The expertise of evaluators is not nullified by bibliometrics, but reconfigured, enhanced, and improved. The directive is clear: bibliometrics and peer review must act synergistically, mutually reinforcing each other, while one method brings relevance and context, the other one brings neutrality. This illustrated in Figure 2.

Figure 2. Combination of methods to achieve greater relevance and objectivity



Evaluative bibliometrics involves a hybrid approach that harmoniously integrates quantitative information into qualitative evaluation. The peer review process, being intrinsically qualitative, provides detailed judgments about the content of the research, allowing for the analysis of the relevance, originality, and theoretical contribution of researchers, ensuring that the content is pertinent and aligned with the current needs of the scientific community. On the other hand, bibliometrics offers an informed, neutral and contextualised dimension. Through this, patterns of collaboration, trends in scientific production, networks of collaboration, and influence in various communities can be identified, while also providing comparable data that reveals a broader reality often not perceived solely through qualitative judgment. The synergy between peer review and bibliometrics is essential for a robust and nuanced evaluation, truly reflecting the complexity of research. In this framework, the grounded information provided by bibliometrics can be essential to counteract the limitations of peer review.

It is notable how new evaluative methods question the validity of bibliometric indicators but do not criticize the limitations of peer review, despite the extensive literature. Studies have shown that a reviewer's decisions can be influenced by preconceptions about the researcher's gender or geographical origin (Severin et al., 2020). It can also act as a brake on innovation, as experts tend to favour research that aligns with established paradigms. Disruptive proposals may face greater obstacles when judged by traditional criteria. Reviewers, focused on their specific fields, may develop narrow views of science. This restricted perspective can lead to the dismissal of interdisciplinary research or those challenging conventional boundaries of a field. Often, evaluations favour predominant methodologies, relegating novel approaches that could offer fresh perspectives. This preference is largely conditioned by the reviewer's knowledge and training, who may not be familiar with or may be sceptical of emerging methodologies.

We cannot overlook the human factor. Academic rivalries, professional jealousy, and other conflicts of interest can interfere with objectivity. A scientist competing in a similar research niche could, consciously or unconsciously, make biased decisions to protect their position or prestige. To these limitations, we must add the enormous bureaucratic and administrative burden it entails. The process of selecting reviewers represents an intrinsic challenge, and it is not always easy to identify qualified, willing, and able experts without conflicts of interest. This is compounded by the limited budget of administrations for evaluation, resulting in insufficient compensation for reviewers. This often results in less meticulous reviews or prolonged review times. Another challenge in the administration of evaluations is variability: the appreciation of a work varies significantly depending on the reviewer. Bibliometric analyses, are not only perceived as cheaper solutions than peer review, but can evolve into information and monitoring systems, at a lower cost and with less bureaucracy involved. Bibliometrics emerges as a tool to rectify some of these issues, contributing to more agile, transparent, and economically viable processes.

Evaluation systems must support their decisions on bibliometrics, together with the specific criteria of expert committees and panels. The latter guide the final decision, while bibliometrics acts as a counterbalance, ensuring the integrity of the evaluative process. Therefore, instead of perceiving bibliometrics as a problem, it should be considered as a tool that allows identifying possible deviations. Bibliometrics has the potential to prevent and detect arbitrary decisions by reviewers, whose judgments, as we have seen, could be influenced by economic, political, ideological factors, and their personal values and perceptions. As Hammarfelt and Rushforth (2017) point out, in processes such as hiring staff, bibliometric indicators serve as judgment devices that mitigate the subjective component and individual conditioning. Therefore, it is essential to analyze discrepancies between the judgments of expert committees and bibliometric results. It is essential to collaborate, while maintaining their independence, and to ensure a robust and balanced evaluation.

To illustrate the complexity of the problem, Table 5 presents four elements or factors that influence an evaluation committee, highlighting the emphasis (high, low, moderate) placed on each of them and thus evidencing the variety of systems that can be configured. When considering only the expert committee, we must recognize that its implementation requires the intervention of a wide range of professionals with varied profiles, ranging from managers and technicians to scientists and administrative staff. It is also essential to define what we mean by "experts" or even “peers”, as these committees may include academics, researchers, industry professionals, among other specialists in the area. Subsequently, it is imperative that members are selected based on their experience and track record in relevant scientific fields, thus ensuring disciplinary diversity and impartiality. It is likely that we will need to evaluate the candidates to be part of the committee; this selection could be based on a thorough analysis of their curriculum vitae and interviews.

Table 5. Example of emphasis placed by different commissions on elements of a mixed evaluation system.

	Experts' Influence	Use of Indicators	Process Complexity	System Costs
Commission A	High	High	High	High
Commission B	High	Low	Moderate	Moderate
Commission C	Low	High	High	Low
Commission D	Moderate	Moderate	Moderate	Moderate

Once the profile is defined and the candidates are selected, a balanced commission generally consists of 5 to 10 experts. Within this commission, it is vital to designate specific roles, such as a chair and a secretary, who will be responsible for coordinating and documenting the activities. The commission's operation requires clear rules for the evaluation of proposals, always ensuring confidentiality. It is necessary to determine the relevance of expert judgment and the degree of use or importance given to indicators. At this point, it is very likely that we will need to hold preliminary meetings to review proposals for indicators that the commission deems appropriate. Therefore, it is crucial to ensure that the commission has knowledge of bibliometric indicators, as this will allow them to use, adjusted and select them properly. Additionally, it may require the complementary preparation of detailed materials on bibliometric methods or the provision of specific training. It is necessary to inform the commission of experts about at least the following issues:

- What indicators are available, how can they be used, and which are their advantages and limitations, noting the implications of certain misuses.
- What information sources, databases, and platforms have been selected, their coverage, scope, advantages, and limitations compared to others.
- What strategies can be implemented to mitigate the inherent biases of the various bibliometric methodologies and techniques employed.

Practical case

To conclude, let us now look at two concrete examples of everything discussed above. In the first case, we show a recruitment exercise. In the second case, an excellence program is evaluated to assess its performance.

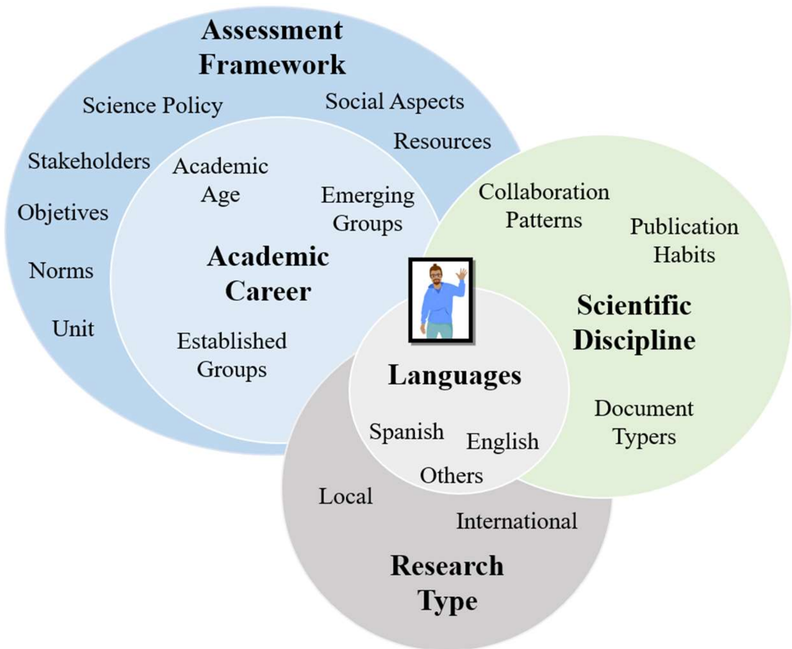
- **Case 1. Selection of candidates for a hiring process using peer review and indicators**. A combined strategy of peer review and indicators was chosen for selecting candidates. Initially, there were 121 candidates. Using indicators based on publications, citation volume, contribution to open science, and other professional merits, the list was filtered, and the 15 most promising candidates were highlighted. An expert committee evaluated the skills, abilities, and potential of each of these 15 applicants. Thanks to their deliberations, a consensus was reached on the five most suitable candidates for the position. These five finalists were provided with a more detailed bibliometric report on collaborators, citation networks, and impact outside academia. They then underwent a final round of interviews where both the indicators and the committee's assessments were considered, allowing for a justified selection of the ideal candidate.

- **Case 2. Detailed ex-post evaluation of research excellence units**. After funding excellence units at a university, an ex-post evaluation begins. First, using bibliometric techniques, the collaboration networks and their integration into the international scientific community are examined. Second, interdisciplinarity is analysed with science maps and thematic analysis with their global positioning. The relationship between the received funds and the resulting scientific production is highlighted, allowing the identification of the return on investment. Simultaneously, an expert committee conducts a peer review of the projects and innovations of each unit. These experts evaluate the practical applicability of the research, its social impact, and its consistency with the originally proposed objectives. By combining bibliometric data with the committee's perceptions, the units are ranked according to their effectiveness and contribution to the academic community, thus providing a comprehensive view of their performance after funding.

Principle 3: Principle of respect for contexts

In any evaluation, context plays a crucial role. Whether social, political, or economic, it determines and nuances the meaning of any criterion used. Therefore, bibliometric indicators cannot be considered as absolute, universal or static measures, but rather malleable categories that must be continuously reviewed, modified, and readapted and reinterpreted (Robinson-Garcia & Ràfols, 2020). These metrics, if applied universally and decontextualised, can give an inaccurate or distorted picture of reality. True bibliometric specialists can discern when to use certain indicators and how to adapt and interpret them according to different contexts and goals. Science does not operate in a vacuum but within a complex fabric of academic, social, cultural, and political interactions. This can be referred to as contextualized bibliometric analysis, highlighting the imperative need for evaluations to be genuinely adapted to their environment, to the changing dynamics of scientific research, and to the context in which the knowledge has been generated.

Figure 3. key contexts to consider in evaluation frameworks: influencing factors across academic career, research types, scientific disciplines, and language



Next, we will outline the five predominant contexts to which every bibliometric study must pay attention (Figure 3). First, there is the evaluation framework that includes the scientific policy dictating guidelines, overarching philosophy, and objectives to follow, and thus determines the decisions to be made. Second, closely linked to the previous one is the context and characteristics of the academic and research careers of the evaluated researchers or groups, which urges us to address essential issues such as age, career stage, or to establish policies for equality, diversity and inclusion. Thirdly, we find the disciplinary context, which addresses the intrinsic variations between specialties, in terms, for example, of publication and citation patterns. The linguistic context is equally fundamental, as the language of publication—whether Spanish, English, or regional languages—can influence the visibility and accessibility of works. Finally, the context of scope focuses on the distinctions between research oriented toward the local sphere and that of international scope, recognizing the importance of both in the global landscape of science. Each of these contexts, with their particularities and nuances, adds an additional dimension to our analyses. Below, we will develop each of them.

Assessment framework. A key aspect to consider here is the scientific policy framing the evaluation, as this will determine its normative and regulatory scope. It will also determine the unit for evaluation. This will not only affect data collection processes but also determine the structure of our analysis. Similarly, the agency or entity in charge of the evaluation process must share its objectives with the bibliometrician in order to select the most appropriate indicators. These indicators will not be chosen solely based on the characteristics of the evaluated unit and but may include external socio-economic conditions that could be affecting its performance. It is not the same assessing well-established research groups led by senior PIs, than those newly formed led by young PIs. Hence, the following questions must be answered:

- Unit of evaluation: Are we considering a country, an institution, a research group, an individual, a research field, or an international network, or is it a combination of some of them?
- Objectives of the evaluation: Is it about allocating funds, improving performance, increasing regional engagement, adjusting budgets, promoting collaboration or multidisciplinary research?
- Socio-economic profile: Is the evaluation aimed at an entity well-positioned in the international scientific community, or is it in a specific phase of scientific development?

There are further aspects to consider. For instance, the evaluated time period. Are we analysing a historical phenomenon, a current scenario, or making future projections? We will also have to clearly identify the stakeholders and understand what their expectations are and how they will be involved in the evaluation. Furthermore, how much time and resources can we devote to the evaluation process? It is crucial to define how much money will be dedicated to bibliometric evaluation and the team; the promoting entity must be aware of the cost of data and the personnel involved, as well as the time needed for conducting a thorough evaluation.

Academic career. When assessing a career trajectory, it is essential to consider both the length of the career and the experience of the candidate. It is important to understand that the most traditional and non-normalized indicators, such as the number of citations and the *h*-index, have enormous limitations and biases if applied to the careers of young researchers. An indicator should accurately reflect the work of emerging scientists, capturing their dynamism and recent contributions, especially in the early stages of their careers. A solution could be to consider shorter time periods or include other aspects, such as collaborations or the diversity of topics addressed, which can help understand a career more accurately. In this sense, selecting the right indicator is fundamental. For example, when evaluating young and emerging researchers seeking to strengthen their presence in international networks, it is justifiable to use measures such as the Journal Impact Factor in order to help us understand how candidates are strategically positioned in the global publishing landscape. Indicators must be adapted and selected based on the needs of the evaluation system but also ensuring fairness to all candidates based on their different stages and research profiles.

Scientific discipline. The field of research will be the most impactful factor determining researchers' scientific profile. It is imperative to recognize the inherent particularities of different disciplines and adapt the indicators to disciplinary practices and not the other way around. It is well documented how dynamics, publication patterns, collaboration, and knowledge dissemination channels vary between disciplines and specialties. For example, in the humanities, production cycles are longer, and a greater number of monographs and research chapters are published. These works are often aimed at a broader audience, and collaboration between authors is less common. In contrast, physics is characterized by rapid generation and dissemination of knowledge, extensive international collaboration, and the predominance of articles in recognized journals likely disseminated through preprints deposited in thematic and institutional repositories. In this sense, it is important to note that the type of media and platforms used for publication, as well as the speed of knowledge obsolescence, also differ. For example, conferences may be a primary means for disseminating results in Computer Science, while in Philosophy, essays may be of greater relevance. The choice of indicators cannot be uniform or arbitrary but must be informed by the intrinsic characteristics of each discipline.

Languages. One of the essential factors often overlooked is the language of publication. The dominance of English in databases, scientific journals, and other knowledge sources has led to a prevalence in academia not only of the language but also of an Anglo-Saxon scientific agenda. While English is a *lingua franca*, mediating international communication, it should not overshadow the importance and value of contributions in other languages representing large areas of population such as Spanish and Latin America. Even languages spoken by a smaller population like Catalan or Basque languages may be of crucial importance in certain fields and topics. It is essential to respect and value the research conducted in these languages, especially as they may address unique cultural and local issues that may not make sense to publish in other languages. The same applies to other regional and national languages, which have a rich corpus of academic research in the humanities and social sciences. These are fields where culture, science, and language are intrinsically linked. Ignoring contributions in languages other than English can be detrimental to whole areas of inquiry. In fact, excluding or minimizing these contributions can result in a biased and limited view of global academic knowledge. Bibliometric studies must recognize and incorporate linguistic diversity to ensure a complete and equitable representation of global knowledge.

Research type according to geographic scope. Finally, the local or global orientation of research can greatly explain differences in terms of citation impact which may have little to do with the quality or relevance of the research analysed. Both, global and local research, present inherent characteristics that reflect different approaches and objectives. Local research, rooted in the particularities of a specific site or addressing specific issues, offers solutions with a direct impact on a targeted community which may not be as relevant to other communities. In contrast, global research focuses on global-scale challenges, fostering the integration of knowledge and dialogue between various disciplines. While local research provides a detailed and contextualized approach, international research facilitates the transfer and adaptation of innovations in different settings. From a bibliometric perspective, identifying local research is still an unresolved issue (Di César & Robinson-Garcia, 2024), with approaches which range from affiliation-based and journal-based methods to the use of toponyms and demonyms, to the use of keyword co-occurrence analysis to identify predominant themes and their resonance in regional publications. Centrality indices in social media could be used to highlight relevant authors or institutions in the local context. For global research, it would be pertinent to analyse co-authorship networks and collaborations between institutions. Additionally, studying citation patterns at the international level would offer insight into the interconnection and impact of certain works on the global landscape. These are clear examples of how different research types may require different data and approaches for their evaluation.

Practical case

Let us illustrate how this principle is put into practice with an example. The regional government, following EU scientific policy recommendations, wants to fund projects focused on artificial intelligence with a dual approach: (A) Computer Science projects employing advanced AI techniques and (B) projects examining the impact and implications of AI on society directed at fields from the Social Sciences and Humanities. Given the dual nature of the call, the process of selecting research groups and researchers must be rigorous and adapt to the singularities of each discipline. To do so, we will proceed as follows:

- **Projects A.** We will prioritise research teams with clear technical capability and experience in applying AI for complex data analysis. Traditional bibliometric metrics, such as the number of publications, normalized citation indicators, and first authored papers are used. Conference papers will also be considered. For emerging themes, indicators based on contributions to thematic repositories, including usage indicators such as number of downloads will be used. Additionally, the relevance and novelty of their developed algorithms as well as any economic impact derived from their commercial exploitation will be considered.
- **Projects B.** Teams that have demonstrated a deep understanding of the social and humanistic dimensions of technology are selected. The evaluation procedure is will not use bibliometric indicators exclusively; dissemination and discussion in academic blogs or impact on social networks and audience analysis are also considered. The production of outreach material is positively valued, as are publications in Spanish and leadership in academic networks in Latin America and other Spanish speaking countries.

Finally, in line with the socio-political context and equality policies, it is established that at least 40% of the principal investigators of the selected groups must be women. Furthermore, to ensure that not only the most established and long-standing groups monopolize the funding, 20% of the funding will be allocated to young and emerging researchers, that is, those with less than 10 years since obtaining their doctorate. This would be a case in which all context illustrated in Figure 3 affect the design and implementation of the evaluation procedure and the way in which bibliometrics are used and integrated.

Principle 4: Principle of metric multidimensionality

Scientific impact has been traditionally evaluated based on indicators centred around journal articles, i.e., citations, Impact Factors. But this approach neglects the range and multidirectionality of pathways by which science influences and permeates both the scientific community and society (Ramos-Vielba, Robinson-Garcia & Woolley, 2022). Research results not only feed the corpus of scientific knowledge but also shape and are shaped by the cultural, educational, economic, and political formation of a society. These are all interconnected dimensions that often overlap. It is crucial for evaluators and research managers to adopt a broader and multidimensional approach in order to understand the holistic impact of science. Adopting such a perspective will enable the design of more inclusive and efficient policies and will more faithfully reflect the real value of research in the social fabric. The principle of multidimensionality is fundamental, as understanding clearly what is to be evaluated will determine whether it can be measured and the best suited indicators framework needed for such purpose. This principle requires a broad knowledge of information sources and indicators.

Although this principle is now manifested more clearly, it has always been very present in evaluative bibliometrics. Perhaps Moravcsik (1984) in the 1980s was one of the first to mention it when he warned against unidimensional thinking. In the 90s, the discussion continued; however, the variety of measures was still very limited, as most were still derived from citations and publications in indexed journals. Technical limitations would restrain the bibliometric community from adopting a multidimensional approach. The computerisation of science (Moed, 2916) and the methodological and technical advancements derived from it, allowed to track research trails that were hidden up to then. One of the most interesting contributions in those early years was the article by Bollen et al. (2009), in which they presented a classification using statistical techniques of 39 indicators. This list included metrics that captured the interaction and use of publications in the digital environment (downloads, visits, uses, etc.). This was the first milestone towards other propositions that looked for traces of scientific outputs beyond the academic realm, such as webometrics or more recently, altmetrics.

Current evaluation processes should consider at least four dimensions of impact: scientific-academic, educational, economic-technological, and social-cultural (Moed, 2020). Each of these dimensions offers a wide range of metrics. As shown in Table 6, nowadays we can measure a vast range of contributions beyond journal publications, also identifying many other associated metrics of impact beyond citations. For example, in the scientific-academic dimension, beyond articles, the value of research data sets or software developments is recognized. Concerning educational impact, the creation of educational materials or the offering of online courses, such as MOOCs, is also gaining importance. The economic-technological dimension goes beyond patents, also encompassing the founding of spin-off companies and industrial rights. In the realm of social or cultural impact, we consider not only formal documents but also events, exhibitions, and, of course, communications on social media and traditional media. Science has a reach that extends publications, and it is our duty to integrate them in our bibliometric toolkit.

Table 6. General dimensions of scientific impact and contributions, publications and non-publications, that can be considered according to Moed

Impact Dimension	Based on Publications	Not Based on Publications
Scientific-Academic	Scientific journal article; book chapter; academic monograph; conference paper; editorial; review	Research data set; software, tool, instrument; experiment video; registered intellectual property
Educational	Course textbook; syllabus; manual or textbook	Online course; graduated students; degrees obtained, e.g., doctorates
Economic or Technological	Patent; commissioned research report	Product; process; device; design; image; spin-off company; registered industrial rights; intellectual property commercialization revenue
Social or cultural	Professional guidelines; policy documents; newspaper article; encyclopaedia article; popular book	Interviews; events; performances; exhibitions; scientific advisory work; social media communication, i.e., blog posts, microblogging.

These indicators provide the necessary framework to construct multidimensional matrices adapted to the particular evaluative contexts. Different bibliometric providers may offer different taxonomies or groupings of indicators, facilitating the application and usefulness of these indicators. Still, we must emphasize that these classifications need to be interpreted and adjusted to our specific conceptual framework, whether by choosing the relevant dimension for our study or recognizing that each dimension can be subdivided into different facets. For illustration, five categories or facets, along with their corresponding metrics, can be identified, which are commonly used to quantify and define the dimension of scientific-academic impact.

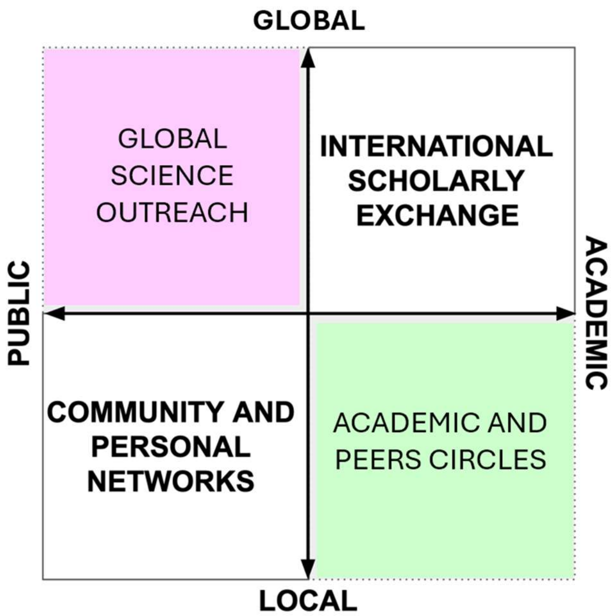
- **Research output.** The total number of publications reflects the ability of the researcher or research group to generate results.
- **Observed scientific impact.** The normalized citations that an article receives indicate the reception and relevance of the work in the scientific community.
- **Scientific collaboration.** The collaborative dimension of a researcher or group, demonstrating their ability to establish connections and form networks.
- **Leadership in authorship.** Publications where the researcher appears as the first or corresponding author can indicate leadership.
- **Open Science.** Considering the number of publications in Green or Diamond Open Access to verify compliance with open access policies.

The bibliometric universe extends beyond the usual indicators thanks to alternative metrics. Altmetrics, more than just social media metrics, encompass a wide range of measures that reflect how science is perceived, consumed, and mentioned in digital environments of different natures (Torres-Salinas et al., 2024). We do not refer just to counting mentions on X (Twitter) or Facebook, but to a great diversity of digital manifestations. For example, mentions issued in policy documents by national and supranational bodies, mentions in news media or citations in Wikipedia articles should also be considered. Although it is possible to calculate these metrics on their original platforms, there are so-called altmetric providers which aggregate these metrics, offering an integrated view that can greatly facilitate their data collection.

Due to its development and conceptual ambiguity, altmetrics can be defined as metrics based on mentions coming from beyond the academic realm. Depending on their origin, the meaning of these metrics will differ, however, they will always point towards some sort of attention to science coming from non-academic audiences and sectors. Advanced bibliometric methods can help us link research topics with social interests and needs. Through community detection methods, we can characterise such audiences (Arroyo-Machado, Torres-Salinas & Robinson-Garcia, 2021), pointing towards broader dimensions of impact beyond the traditional citation impact. Hence, altmetrics can help us find hot spots of potential educational, economic, social or cultural interest.

However, it is important to note that dissemination through these channels does not inherently mean wider public engagement. It is crucial to consider the type of audience before conducting a raw quantification to accurately assess the true reach and impact of science (Arroyo-Machado & Torres-Salinas, 2023). Understanding whether the audience is closer to the academic community or the broad public, and whether the orientation is global or local, is essential (Figure 4). For example, on platforms like X, there can be academic accounts engaging with research, including departmental colleagues, international peers, science communicators, local media, or friends, among others. This distinction helps to better comprehend the actual audience and the real influence of the content shared.

Figure 4. Categorization of social media audiences by reach and engagement type



The model proposed by the Influscience initiative[1] provides a structured framework as it identifies four dimensions and selects specific platforms and indicators to quantify each of them. The underlying basis of the model is that the influence and social transfer of scientific research can be objectively evaluated through the mentions and attention that publications receive on digital platforms. These mentions can also be interpreted differently depending on the nature of the platform and the audience accessing the information. The proposed dimensions and their operationalization are detailed below:

- **Social Influence.** This dimension captures the impact on a general non-specialized audience. An effective tool to measure this aspect is X or Bluesky, which acts as a digital mirror of significant segments of our society. Mentions or reposts of a scientific article on this platform can indicate the interest or discussion it generates among the broad public.
- **Political Influence.** It evaluates the impact of research in the political realm. It is essential to observe how organizations like the EU and the OECD use articles in their reports. Mentions in these reports reveal which research plays a crucial role in policy formulation and support.
- **Media Influence.** It focuses on how digital media present science to the public. A metric is the number of times articles are mentioned in digital newspapers, contextualizing the data with the characteristics of the media's audiences.
- **Educational Influence.** It represents how results are incorporated into educational contexts, reflecting effective transfer. A representative platform is Wikipedia, as mentions of articles in its entries can indicate their capacity to influence the educational world.

Therefore, it has become clear that science has different dimensions and impacts that can be nuanced, adapted, and measured through indicators.

Practical case

Next, we discuss with an illustrative case, how to apply the principle of multidimensionality. On this occasion, we are faced with an evaluation committee specializing in the fields of Political Sciences and Economics. This committee will assess the contribution and influence two business schools have had on national and international economic policies and decisions. Following the multidimensional model presented in Table 6, there are two dimensions that fit into the committee's evaluative framework:

- **Scientific dimension.** From a traditional bibliometric perspective, the scientific impact of these experts will be measured not only by their total publication output and the number of citations these receive, but will also include their collaboration rate, quantified throughout their co-authorship patterns, which sheds light on their ability to weave international academic networks. Furthermore, the committee

[1] InfluScience is a research project by the University of Granada that measures the social impact of scientific production beyond traditional citation metrics. Part of a national research plan funded by the State Research Agency, the project analyzes scientific publications' influence across multiple dimensions, creating classifications of researchers and articles in 22 scientific fields.

will assess their commitment to Open Science, reflected through the number of open access publications, especially working papers and shared datasets.

- **Sociocultural dimension.** The committee will pay special attention to the influence the outputs of these business schools has had in the political realm. To do so, they will be focus on citations issued from policy reports. This innovative approach seeks to detect how their research is considered relevant in the design and implementation of policies and public recommendations. A detailed analysis of the institutions citing the two schools' output will be conducted, allowing us to discern which organizations, whether national or supranational, are using their knowledge to inform, structure, or reinforce policies in the political and economic realms.

Principle 5: Principle of verifiability and openness of data

Documents such as the DORA declaration, the Leiden Manifesto or reports like The Metric Tide have placed special emphasis on all aspects related to accountability in measurement. This is also the foundation of what we now know as the responsible metrics movement, which is based on the requirement for data verification and transparency. Evaluative bibliometrics must not only be limited to mere counting and quantification of research activity. Research management officials, and especially those being evaluated, require an environment based on trust in the methods and confidence in the proper use of individual and institutional data. It is essential to ensure that those being evaluated have access to all the bibliometric data we collect about them, allowing for the verification of indicators. Bibliometrics should not merely provide a set of metrics; its development entails a series of ethical and procedural aspects. Table 7 summarizes the technical and ethical aspects we consider most relevant.

Table 7. Technical and ethical aspects to consider when working with data

Aspect	Conceptualization	Definition	Brief Example
Reliable sources	Reliability of information	Use of recognized and validated tools and databases in the bibliometric field.	Use of multidisciplinary citation indexes that declare their coverage.
Data management	Responsible Processing	Detailed data management plans from data collection to data exploitation	Use of conventional data formats such as CSV for download and organization.
Use of APIs	Integration and accessibility	Application of tools that enable data communication and transfer	Development of an application that verifies data using a database's API.
Transparency	Full disclosure	Clear and complete disclosure of the methods, sources, and processes used.	Publication of a detailed report on how the analyses were conducted.
Interpretative ethics	Conscious data use	Reflection and responsibility when interpreting and presenting data.	Ensuring that metrics are not presented in a way that could distort reality.
Privacy	Information protection	Ensuring that personal and sensitive data are protected.	Anonymizing names in a public report to protect small entities.
Validation	Data verification	Ensuring that the collected information is accurate and correct.	Direct consultation with scientists to refine preliminary publication lists.
Ethical commitment	Integrity and responsibility	Ensuring that bibliometric practices align with general ethical principles.	Rejecting the manipulation or misrepresentation of data to achieve certain results.
Data coherence	Consistency of data	Avoiding the combination of data from different sources for the same metric.	Using citation counts from a single source to ensure consistent metrics.

The foundation of this principle, which addresses verifiability and openness, is as old as evaluative bibliometrics itself. As Thed van Leeuwen (2004) noted, the degree of validity and reliability of the data constitutes the fundamental pillar of any analysis. It is essential to emphasize the importance of always working from a bottom-up data collection approach, which involves compiling scientific contributions from the individual level and then ascending to broader levels, such as departments and centres. In the early work of the CWTS, it was detailed how they collected data in local databases and provided researchers with lists of their publications to verify if there were errors or omissions. Therefore, it has always been emphasized that the data should be validated by the authors. This practice must be maintained despite the advanced development of name and entity disambiguation algorithms. Fortunately, we have various author identification codes, such as ORCID, and CRIS systems that can assist us in these tasks, offering a more sophisticated and personalized validation experience.

Another aspect to consider is not so much what or how to collect, but where we should collect the data to build bibliometric databases. We are immersed in an unprecedented proliferation of sources and, consequently, metrics. This abundance, far from simplifying the work, requires a deep knowledge of the scientific information landscape. It is vital to understand how each of the bibliometric platforms and products is constructed, as it within its internal structuring that lies an intricate web of technical and commercial decisions that determine, in one way or another, aspects such as coverage, interfaces, and, of course, the evaluative options they offer, such as categories or classification systems. Infrastructures are not neutral. They are often shaped by assumptions and, in turn, influence and reshape our own practices. For instance, it is common to tailor evaluations to the indicators and data available in a given product. Furthermore, in this fluid landscape of data and metrics, some platforms may become obsolete or disappear, creating a dependence on specific systems. Another challenge is the reliability of products, which may suffer from poorly curated metadata or inconsistencies in the same indicator across different platforms, as seen with altmetric aggregators that use varying methods to track activity. This raises a critical question: how do we ensure that the data are not only reliable but also accurate and free of substantial origin errors?

These problems are overcome through proper management and curation of internal databases. Planning the entire data lifecycle with a long term and holistic vision, from acquisition to processing, curation, and exploitation, will ensure their openness. Fortunately, many indexes allow mass downloads of records and author profiles, which greatly facilitates the task. The proliferation of standard formats, whether CSV, JSON, or RIS, simplifies the creation of an initial database. Still, an intense curation of errors in author and institution disambiguation still needs to be done. There are many freely accessible tools, such as Bibliometrix, that can assist us in these tasks. However, the complexity behind their apparent simplicity of bibliographic records should not be underestimated: data management, especially in the long term, requires constant attention. Bibliometricians must also be aware of the legal implications when handling data, particularly proprietary and personal data. Data protection policies are therefore vital to safeguard the confidentiality of those being evaluated. All these aspects can help generate trust and prevent errors.

Another question relates to the management and storage of the information itself. In this regard, the usefulness of additional tools in the process, such as APIs and CRIS, can be considered. Let us start with the former. APIs (Application Programming Interfaces), provide bridges between different platforms and systems for data transfer. Their significance lies not only in metadata exchange but also in their capacity to create more open bibliometric products. These interfaces can automate processes and interconnect data, helping us create applications aimed at data verification and enrichment without exorbitant costs, making them technologically viable. APIs are not just technical tools; they are vehicles that promote a culture of integrity and openness in evaluative processes, maximizing the accuracy of bibliometric data. In these contexts, the evaluated individual can verify and explore how and with what data each of the indicators affecting their assessment has been calculated.

Another key aspect for the effective use of APIs and the ability to gather a robust set of indicators lies in the use of standardized unique identifiers. Each type of entity—such as authors, publications, or institutions—can be associated with standardized identifiers like DOIs, ORCIDs, or RORs, as well as internal identifiers used by various databases. These identifiers play a crucial role in ensuring that the same resources can be uniquely identified across different platforms, which significantly enhances interoperability. For instance, a DOI can be used to retrieve citation data from one source and altmetrics from another, streamlining the process of data collection and analysis. Additionally, standardized identifiers help in maintaining consistency and accuracy in data. By providing a common framework for identifying and linking data, these identifiers simplify the integration of diverse datasets and support more robust and extensive research analytics.

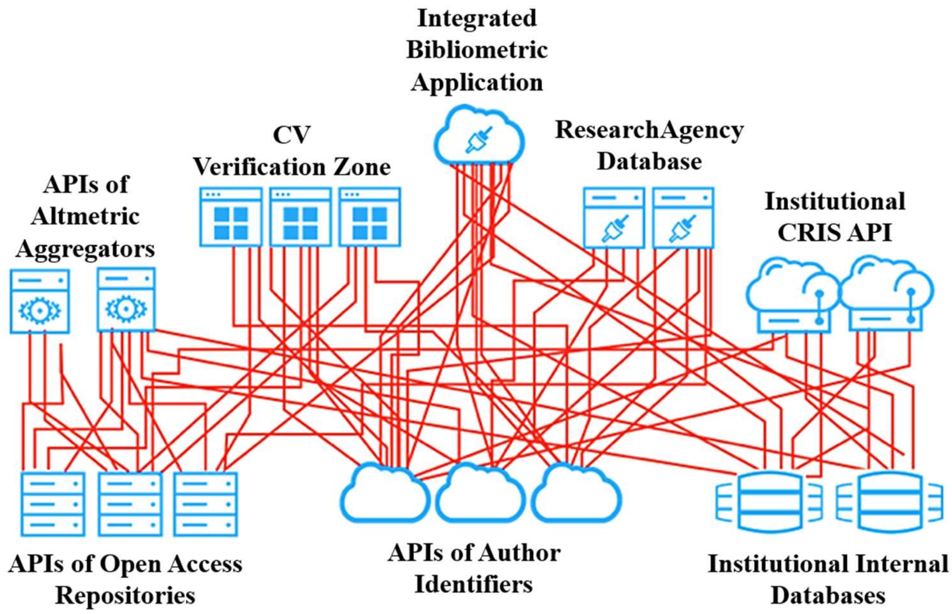
Despite their potential, APIs have two significant limitations. First, integrating data from different sources with varying data quality can introduce noise and complicate the processes of data disambiguation and cleaning. This challenge arises because each data source might have its own standards and practices, leading to inconsistencies that need to be meticulously addressed to ensure reliable outcomes. The task of harmonizing such disparate data can be time-consuming and requires sophisticated methods to manage and rectify discrepancies. Second, APIs can pose a barrier for bibliometricians who are less experienced in programming, as they must transition from using visual interfaces to writing and understanding code. This can be particularly daunting for those without a technical background, limiting the accessibility and usability of APIs for a broader audience. Nonetheless, to mitigate this issue, major bibliometric software increasingly offers user-friendly features that integrate the connection with these applications into their interfaces. These advancements aim to bridge the gap by providing intuitive tools that allow users to harness the power of APIs without needing extensive programming skills.

To better illustrate the potential and workflow of APIs, Figure 5 presents a scheme that serves as a case study of an application designed to evaluate groups in a university, integrating different types of sources, but especially bibliometric APIs. In this model, sources are selected that best adapt and resolve an evaluative framework at a given moment. From there, the bibliometric application is built ad-hoc and adjusted, for example, this is constructed based on the following principles:

- **Integration and connection**: The image emphasizes the essentiality of interconnecting various sources through APIs, facilitating a broad and accurate evaluation.

- **Diversity of sources**: Sources, from altmetric aggregators to open repositories, provide crucial data on impact and dissemination.
- **Verification process**: The "CV Verification Zone" allows researchers to check their contributions and indicators.
- **System flexibility**: The inclusion of "author identifiers" points to obtaining data from open sources generated by the evaluated individuals themselves.
- **Centralized platform**: The "Integrated Bibliometric Application" is the core of the system, centralizing and processing the information.
- **Integration challenges**: The density of connections points to challenges in data integration, emphasizing the need for coherence and consistency.

Figure 5. Design of an application for bibliometric purposes based on the interconnection of different information sources through APIs



Here we must mention the role played by Current Research Information Systems (CRIS). These systems consolidate information related to academic activities, from publications and projects to educational resources, and can provide a holistic perspective of institutional performance. Beyond their integrative capacity, CRIS centralise information and facilitate the data verification process. Ideally, researchers would feed these systems with their contributions, but we must be realistic and facilitate this work for them. Thanks to the interconnection provided by APIs, the loading process can be more fluid and many parts of it can be automatised, leaving researchers with the task of verifying and completing potential information gaps. CRIS can function as a bottom-up self-evaluation tool, illustrating how metrics and indicators fluctuate depending on factors such as publication lists or thematic boundaries. However, for specific evaluations, they are often not the best option due to their high management costs, complex architecture, heavy maintenance burden, and because they involve too many people. This reduces their dynamism and flexibility, which is sometimes necessary.

There are many aspects that encompass data management, and we will mention one more that we have intentionally left for the end: the ethical issues related to its use. We are in an era where global access to information is palpable, and bibliometrics has not been left behind, allowing non-experts to use sources to discover underlying patterns and trends in academic production. Citizen Bibliometrics, a term coined by Leydesdorff, Wouters and Bornmann (2016), should involve more than merely democratizing bibliometric knowledge; it should involve ethical action in the field. The massive use of information should be accompanied by responsibility to ensure the correct use of data. This means adopting a Numberethics that invites us to reflect on our relationship with numbers and the inherent responsibility in their use. These concepts, taken together, advocate for a more informed bibliometric practice based on solid ethical principles. Many of the decisions made with our information can affect researchers' career paths and, therefore, could have effects on their personal lives, so methodological care and rigor must be governed by the highest standards.

4. EPILOGUE

This brief immersion into evaluative bibliometrics has revealed an important tradition in the field of Documentation. Despite its nearly 50-year history, it has been clearly highlighted how its foundational principles, once reformulated, are perfectly applicable today, emerging as a guide to unify the practices of a heterogeneous group of librarians, information professionals, and evaluators. Our goal has not only been to rescue and revalue this body of knowledge but also to raise awareness about the complexity affecting our area, at the intersection of several disciplines. Therefore, we have highlighted great challenges, but also numerous opportunities for our work to be recognized naturally and respectfully in various scientific evaluation processes. In this book, we have sought to offer a professional framework that can be integrated into any evaluation process.

More specifically, the most important challenge is the assimilation and implementation of current evaluative trends, which depict a transforming landscape. This landscape must be built on the foundation of peaceful coexistence with experts and qualitative methods. In this framework, adaptability and collaboration with entities and agencies are fundamental. Many of us face the challenge of integrating these criteria into our research centres and universities. For those who have doubts about how to do this, we believe that this text has provided guidance. In any case, bibliometricians should not fear disruptive proposals in scientific evaluation; evaluative approaches such as the narrative curriculum do not limit or discriminate against us, but rather inspire and project us toward our goals and challenges. These changes push us toward a more analytical aspect of bibliometrics, which we have dared to call narrative bibliometrics, the younger sister of evaluative bibliometrics applied at the micro-level, meaning to curricula and individual contributions. This orientation values contextualization over quantification. By this, we mean that it is an optimal time to reconceptualize and redefine our practices.

There are more challenges to face, such as the complexity of sources, indicators, and methods. Hence, there is a great need for proper training to face the information avalanche. Education, starting with theory considerations and basic methodological awareness, followed by training in the correct understanding and application of bibliometric techniques, is essential. A significant issue lies in the new open sources, whose data are much less curated in their effort to be extensive, adding another layer of complexity. Another major challenge is the emerging role of AI in bibliometrics. AI has the potential to revolutionise data collection, processing, and evaluation processes. Its ability to analyse large volumes of information, identify patterns, and make predictions could catalyse the evolution of bibliometrics. However, it is essential to consider the ethical and methodological challenges that this integration could present, such as the proper interpretation of algorithm-generated results and potential biases in these tools.

Given these considerations, a future is envisioned where evaluative bibliometrics, supported by advanced tools and a deeper understanding of its practice, continues to be a key discipline in academic and scientific evaluation, prioritising ethics, precision in its application, and collaboration with other actors in the system.

5. BIBLIOGRAPHY

Arroyo-Machado, Wenceslao; Torres-Salinas, Daniel; Robinson-Garcia, Nicolas (2021). Identifying and characterizing social media communities: a socio-semantic approach to altmetrics. *Scientometrics, 126*, 9267-9289. https://doi.org/10.1007/s11192-021-04167-8

Arroyo-Machado, Wenceslao; Torres-Salinas, Daniel (2023). Evaluative altmetrics: Is there evidence for applying research evaluation? *Frontiers in Research Metrics and Analytics*, *8*. https://doi.org/10.3389/frma.2023.1188131

Bollen, Johan; Sompel, Herbert Van de; Hagberg, Aric; Chute, Ryan (2009). A Principal Component Analysis of 39 Scientific Impact Measures. *PloS One*, *4*(6), e6022. https://doi.org/10.1371/journal.pone.0006022

Braun, Tibor. (1999). Bibliometric indicators for the evaluation of universities—Intelligence from the quantitation of the scientific literature. *Scientometrics*, *45*, 425-432.

Brookes, B. C. (1990), Biblio-, Sciento-, Infor-metrics??? What are we talking about? In: L. Egghe, R. Rousseau (Eds), *Informetrics 89/90. Selection of Papers Submitted for the Second International Conference on Bibliometrics, Scientometrics and Informetrics,* Amsterdam, Netherlands, Elsevier, pp. 31–43.

Cronin, Blaise; Sugimoto, Casssidy R. (Eds.). (2014). *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact*. The MIT Press.

Di Césare, Victoria; Robinson-Garcia, Nicolas (2024). What is local research? Towards a multidimensional framework linking theory and methods. *Zenodo* https://doi.org/10.5281/zenodo.14033473

Garfield, Eugene. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics*, *1*, 359-375.

Glänzel, W. (2008). Seven Myths in Bibliometrics About facts and fiction in quantitative science studies. *Collnet Journal of Scientometrics and Information Management*, *2*(1), 9–17. https://doi.org/10.1080/09737766.2008.10700836

Gläser, J., & Laudel, G. (2007). The Social Construction Of Bibliometric Evaluations. In R. Whitley & J. Gläser (Eds.), *The Changing Governance of the Sciences* (Vol. 26, pp. 101–123). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6746-4_5

Gumpenberger, C., Wieland, M., & Gorraiz, J. (2012). Bibliometric practices and activities at the University of Vienna. *Library Management*, *33*(3), 174–183. https://doi.org/10.1108/01435121211217199

Hammarfelt, Björn; Hallonsten, Olof (2023). Are evaluative bibliometrics neoliberal? A historical and theoretical problematization. *Social Science Information*, 1-23. https://doi.org/10.1177/05390184231158195

Heinze, Thomas; Jappe, Arlette (2020). Quantitative science studies should be framed with middle-range theories and concepts from the social sciences. *Quantitative Science Studies*, *1*(3), 983-992. https://doi.org/10.1162/qss_a_00059

Krüger, Anne K.; Petersohn, Sabrina (2023). 'I want to be able to do what I know the tools will allow us to do': Practicing evaluative bibliometrics through digital infrastructure. *Research Evaluation*, *31*(4), 475-485. https://doi.org/10.1093/reseval/rvac009

Leydesdorff, Loet; Wouters, Paul; Bornmann, Lutz (2016). Professional and citizen bibliometrics: complementarities and ambivalences in the use and development of indicators—a state-of-the-art report. *Scientometrics, 109*(3), 2129-2150. https://doi.org/10.1007/s11192-016-2150-8

Martin, B. R., & Irvine, J. (1983). Assessing basic research. *Research Policy*, *12*(2), 61–90. https://doi.org/10.1016/0048-7333(83)90005-7

Moed, Henk F. (2016). Altmetrics as traces of the computerization of the research process. In C.R. Sugimoto (ed.). *Theories of Informetrics and Scholarly Communication* (pp. 360-371). Berlin: De Gruyter.

Moed, Henk F. (2017). *Applied Evaluative Informetrics* (1st ed. 2017). Springer International Publishing : Imprint: Springer. https://doi.org/10.1007/978-3-319-60522-7

Moed, Henk F. (2018). Towards a multi-paradigmatic, value free informetrics: A reply to Paul Wouters' book review "The failure of a paradigm". *Journal of Informetrics, 12*, 541-546. https://doi.org/10.1016/j.joi.2018.03.003

Moed, Henk F. (2020). *Appropriate Use of Metrics in Research Assessment of Autonomous Academic Institutions* (1). *2*(1), Article 1. https://doi.org/10.29024/sar.8

Moed, Henk F.; Burger, W. J. M.; Frankfort, J. G.; Van Raan, Anthony (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy*, *14*(3), 131-149. https://doi.org/10.1016/0048-7333(85)90012-5

Moed, Henk F.; Glänzel, Wolfgang.; Schmoch, Ulrich. (Eds.) (2004). *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of SyT systems*. Kluwer Acad. Publ.

Moravcsik, M. J. (1984). Life in a multidimensional world. *Scientometrics*, *6*(2), 75–85. https://doi.org/10.1007/BF02021280

Nalimov, V. (1966). Kolichestvennye metody issledovaniya protsessa razvitiya nauki. *Voprosy Filosofii*, *12*, 38–47.

Narin, Francis (1976). *Evaluative Bibliometrics: The use of publication and citation analysis in the evaluativon of scientific activity*. Computer Horizons.

Petersohn, S., & Heinze, T. (2018). Professionalization of bibliometric research assessment. Insights from the history of the Leiden Centre for Science and Technology Studies (CWTS). *Science and Public Policy*, *45*(4), 565–578. https://doi.org/10.1093/scipol/scx084

Ramos-Vielba, Irene; Robinson-Garcia, Nicolas; Woolley, Richard (2022). A value creation model from science-society interactions: Archetypal analysis combining publications, survey and altmetric data. *PLOS One, 17*(6), e0269004. https://doi.org/10.1371/journal.pone.0269004

Robinson-Garcia, Nicolas; Ràfols, Ismael (2020). The Differing Meanings of Indicators Under Different Policy Contexts. The case of Internasionalisation. In C. Daraio and W. Glänzel. *Evaluative Informetrics: The Arts of Metrics-Based Research Assessment: Festscrift in Honour of Henk F. Moed* (pp. 213-232). Springer. https://doi.org/10.1007/978-3-030-47665-6_10

Rousseau, R. (2021). Naukometriya, Nalimov and Mul'chenko. *COLLNET Journal of Scientometrics and Information Management*, *15*(1), 213–224. https://doi.org/10.1080/09737766.2021.1943042

Severin, A., Martins, J., Heyard, R., Delavy, F., Jorstad, A., & Egger, M. (2020). Gender and other potential biases in peer review: Cross-sectional analysis of 38 250 external peer review reports. *BMJ Open*, *10*(8), e035058. https://doi.org/10.1136/bmjopen-2019-035058

Torres-Salinas, Daniel; Jiménez-Contreras, Evaristo. (2012). Hacia las unidades de bibliometría en las universidades: Modelo y funciones. *Revista Española de Documentación Científica*, *35*(3), Article 3. https://doi.org/10.3989/redc.2012.3.959

Torres-Salinas, Daniel; Robinson-García, Nicolás; Jiménez-Contreras, Evaristo (2023). The bibliometric journey towards technological and social change: A review of current challenges and issues. *Profesional de la Informacion*, *32*(2). Scopus. https://doi.org/10.3145/epi.2023.mar.28

Torres-Salinas, Daniel; Docampo, Domingo; Arroyo-Machado, Wenceslao; Robinson-Garcia, Nicolas (2024). The many publics or science: using altmetrics to identify common communication channels by scientific field. *Scientometrics, 129*, 3705-3723. https://doi.org/10.1007/s11192-024-05077-1

Torres-Salinas, Daniel (2024). Principios de Bibliometria Evaluativa. *Editorial EPI-UOC, Barcelona*. 9788411660631

van Leeuwen, Thed (2004). Descriptive Versus Evaluative Bibliometrics. In H. F. Moed, W. Glänzel, and U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research* (pp. 373-387). Springer Netherlands. https://doi.org/10.1007/1-4020-2755-9_15

van Raan, Anthony (1997). Scientometrics: State-of-the-art. *Scientometrics*, *38*(1), 205-218. https://doi.org/10.1007/BF02461131

van Raan, Anthony (2019). Measuring science: Basic principles and application of advanced bibliometrics. In W. Glanzel, H. F. Moed, U. Schmoch, and M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 237-280). Springer.

van Raan, Anthony (1999). Advanced bibliometric methods for the evaluation of universities. *Scientometrics*, *45*(3), 417-423.

van Raan, Anthony (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics*, *36*(3), 397-420. https://doi.org/10.1007/BF02129602

Weingart, Peter (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics*, *62*(1), 117-131. https://doi.org/10.1007/s11192-005-0007-7

“Rather than love, than money, than fame, give me truth.”
HENRY DAVID THOREAU

This book was published and released online on January 16th, coinciding with the 41st anniversary of the death of Ramón J. Sender (1982), a renowned Spanish writer known for his works such as "Requiem for a Spanish Peasant". His literary legacy continues to inspire readers with its profound exploration of Spanish history and human nature.

We sincerely thank you for taking the time to read this modest contribution. We hope that it has brought you inspiration and new perspectives. If you have enjoyed it, we kindly encourage you to share it with your colleagues and peers through your professional networks. When sharing or citing this book, please use the following APA citation format:

Torres-Salinas, D., Arroyo-Machado, W., & Robinson-García, N. (2025). Principles of Evaluative Bibliometrics in a DORA/CoARA Context. Published in Granada by InfluScience Editions.

Thank you for contributing to the dissemination of knowledge.