

Chapter 2

Synopsis

Abstract This chapter provides summaries of the main topics and conclusions of each chapter.

Keywords Altmetrics • Architecture of attention • Business interests • Citation-based indicators • Cognitive intelligence • Communication effectiveness • Constitutive effects • Correlation coefficient • Cross-disciplinarity • Economic indicators • Esteem • Evaluation strategy • Evaluative framework • Individual scholars • Informetric arguments • Journal internationality • Journal metrics • Minimum performance standards • Multi-dimensionality • Nobel Prize level • Ontology • Open Science • Patent-based indicators • Peer review • Performance based funding • Policy context • Precondition for performance • Predictor • Publication-based indicators • Random error • Reputation • Research infrastructure • Scientific collaboration • Scientific development • Scientific migration • Self assessment • Self-selection • Size-dependent • Societal impact • Socio-political context • Systemic Time horizon • Unit of assessment • Usage-based indicators • Webometrics • World university ranking

2.1 Part I. Introduction

2.1.1 *Chapter 1*

Chapter 1 highlights five strong points of the use of informetric indicators in research assessment. It helps to demonstrate one's performance; it gives information on shaping one's communication strategies; it offers standardized approaches and independent yardsticks; it delivers comprehensive insights; and provides tools for enlightening policy assumptions.

Five main points of critique are: Indicators may be biased and not measure what they are supposed to measure; most studies adopt a limited time horizon; indicators can be manipulated, and may have constitutive effects; measuring societal impact is

problematic; and when they are used, an evaluative framework and assessment model are often lacking.

The chapter describes a series of trends during the past decade in the domain of science policy: an increasing emphasis on societal value and value for money, performance-based funding and on globalization of academic research, and a growing need for internal research assessment and research information systems.

Due to the computerization of the research process and the digitization of scholarly communication, research assessment is more and more becoming a ‘big data’ activity, involving multiple comprehensive citation indexes, electronic full text databases, large publications repositories, usage data from publishers’ sites, and altmetric, webometric and other new data sources.

The above trends created an increasing interest in the development, availability and application of new indicators for research assessment. Many new indicators were developed, and more and more measures have become available on a large scale. Desktop bibliometrics is becoming a common assessment practice.

Two principal assumptions of the author are as follows.

- From what is cannot be inferred what ought to be. Evaluation criteria and policy objectives are not informetrically demonstrable values. Informetric research may study such values empirically, but cannot provide a theoretical foundation of the validity of the quality criteria or the appropriateness of policy objectives. Informetricians should in their informetric work maintain a neutral position towards these values.
- The future of research assessment lies in the intelligent combination of indicators and peer review. From the very beginning, bibliometric indicators stimulated peers to make the foundation and of their judgments more explicit. The notion of informetric indicators as a support tool in peer review processes rather than as a replacement of such processes still has a great potential, and this book aims to further explore it. A necessary condition for achieving this is a thorough awareness of the potentialities and limitations of both methodologies.

The main subjects of the book are:

- An overview of important new databases, methodologies, indicators and products introduced during the past 10 years in the field of informetrics and its application in research assessment.
- A comprehensive overview of the most important indicators or indicator families, the aspects or dimensions they are assumed to measure, and their potential and limits.
- A clarification of the relationship between the informetric, the evaluative and the policy domain.
- Possible new features that could be implemented in a research assessment process.
- New lines of research that are expected to lead to the development of new, useful indicators.

This book uses the term ‘*informetrics*’ as a generic term indicating the study of quantitative aspects of information. It does not only analyze bibliometric data based on publication and citation counts, but also altmetric and usage data, webometric data, economic data, research input data, and survey data on scholarly reputation.

The term ‘*assessment*’ is used as an *overarching* concept, denoting the total of activities in assessment or evaluation processes, or the act of evaluating or assessing *in general*. This book uses the term evaluation exclusively in relation to the setting of criteria for, and the formation of judgments on, the ‘worth’ of a subject.

This book focuses on *academic* research, primarily intended to increase scholarly knowledge, but often motivated by and funded for specific technological objectives such as the development of new technologies such as medical breakthroughs. It comprises both ‘curiosity-driven’ or ‘pure’ as well as ‘strategic’ or ‘application oriented’ research.

2.2 Part II. Informetric Indicators of Research Performance

2.2.1 Chapter 3

The multi-dimensional nature of research performance is highlighted. Four main components of research activity are distinguished: input, process, output and impact. *Input* includes funding, manpower and research infrastructure. Indicators of research infrastructure are not primarily performance indicators but relate rather to a *precondition* for performance. *Process* indicators focus on research collaboration and efficiency.

Output can be publication based, such as a journal article or monograph, or be delivered in a non-publication format, such as research dataset, and be directed to the scientific-scholarly community or to society and the wider public. A key distinction is made between scientific-scholarly and societal *impact*. Societal impact embraces a wide spectrum of aspects outside the domain of science and scholarship itself, including technological, social, economic, educational and cultural aspects.

Research performance is reflected in *all four* components. For instance, research funding, especially competitive funding, can be assumed to depend on the past performance and the reputation of the grant applicant, and therefore relates both to input and to impact. Efficiency is both a process and an impact indicator, as it aims to measure output or impact per unit of input. A table is presented that lists the 28 important indicators or indicator families and their pros and cons.

2.2.2 Chapter 4

Chapter 4 presents the main characteristics of the most important indicator families.

- *Publication-based indicators.* In academic institutions, publications in all scientific-scholarly subject fields constitute an important format of academic output. However, in the private sector, and also in academic departments with a strong applied orientation, which aim primarily to produce new products or processes, publishing for the general public, peer reviewed literature often does not have the highest priority; in this case, other output forms must be considered as well. Publication counts may be used to define minimum performance standards.
- *Citation-based indicators.* Citation analysis offers a certain degree of standardization, and compares units of assessment against an independent yardstick, which makes an evaluator more independent from the views of the subjects of the analysis and of the commissioning entity. Citations can be interpreted as proxies of more direct measurements of intellectual influence, but they are by no means indicators of the validity of a knowledge claim. Citation impact and quality are not identical concepts. As all indicators, citations be affected by disturbing factors and suffer from serious biases.
- *Journal metrics.* The quality or impact of the journals in which a unit under assessment has published is a performance aspect in its own right. But the relationship between a journal's impact factor and the rigorousness of its manuscript peer review process is unclear. Journal metrics cannot be used as a surrogate of actual citation impact; they are no good predictors of the citation rate of individual papers. Moreover, their values can to some extent be manipulated, and may be affected by editorial policies.
- *Patent-based indicators.* Analyses of inventors of the findings described in patents reveal the extent to which scientists with academic positions contributed to technological developments. Patent-to-patent citations may reveal a patent's technological value, and patent citations to scientific papers a technology's science base. But the propensity to apply for patents differs across countries because of legislation or culture, and also across subject fields. In addition, patents are a very poor indicator of the commercialization of research results.
- *Usage-based indicators.* Data on downloads of an electronic publication in html or full text format enable researchers to assess the effectiveness of their communication strategies, and may reveal attention of scholarly audiences from other research domains or of non-scholarly audiences. Downloaded articles may be selected according to their face value rather than their value perceived after reflection. Also, there is an incomplete data availability across providers, and counts can be manipulated. It is difficult to ascertain whether downloaded publications were actually read or used.
- *Altmetrics* relates to different types of data sources with different functions. Mentions in social media may reveal impact upon non-scholarly audiences, and provide tools to link scientific expertise to societal needs, but cannot be used to measure scientific-scholarly impact. Their numbers can be manipulated, and interdependence of the various social media may boost figures. Readership counts in scholarly reference managers are potentially faster predictors of emerging scholarly trends than citations are, but results depend upon readers' cognitive and professional background.

- *Webometrics*. Indicators of Web presence and impact are extracted from a huge universe of documents available on the Web. They do not merely relate to institutions' research mission, but also to their teaching and social service activity. But there is no systematic information on the universe of Web sources covered and their quality; and an institution's Web presence depends upon its internal policies towards the use of the Web, and upon the propensity of its staff to communicate via the Web.
- *Economic indicators*. Indicators of economic value and efficiency are relevant measures in research assessment. But not all contributions to economic development can be easily measured. The relationship between input and output is not necessarily linear, and may involve a time delay. Accurate, standardized input data is often unavailable; and comparisons across countries are difficult to make, due to differences in the classification of economic data. Indicators of funding from industry are useful measures of economic value; but funding levels differ across disciplines, and data may be difficult to collect.
- *Reputation and esteem based measures*. Receiving a prestigious prize or award is a clear manifestation of esteem. But absolute numbers tend to be low; the evaluation processes on which the nominations are based are not always fully transparent; and there are no agreed equivalences that apply internationally and facilitate comparison across disciplines. Reputation can be measured in surveys using validated methods from social sciences. But response rates are often very low; mentions may be based on 'hear-say' rather than on founded judgement and may refer to performance made in a distant past.
- *Measures of scientific collaboration, migration and cross-disciplinarity*. Data on co-authorship and on how authors migrate over time from one institutional affiliations to another, provide useful indicators of intra-institutional, national, international scientific collaboration and migration. But instances that have not resulted in publications remain bibliometrically invisible. Indicators of cross-disciplinary measure the relevance of a piece of research for surrounding disciplines, or the cognitive breadth of research impact. Their calculation presupposes a valid, operational classification of research into disciplines.
- *Indicators of research infrastructure* are not primarily performance indicators, but focus on preconditions for performance. They measure basic facilities that support research, the scale of the research activities, and their sustainability. But research practices differ across disciplines; large research teams or laboratories are mostly found in the natural and life sciences. It is difficult to obtain reliable, comparable institutional data, as there is no agreement on the basis of a full cost calculation of research investment. There is no clear, generally accepted definition of being research active across universities, countries and disciplines.

The calculation of informetric indicators of research performance more and more becomes a '*big data*' activity. Not only the increasing volume of informetric datasets is of interest, but especially their *combination* creates a large number of new possibilities. For instance, the combination on an article-by-article basis of citation indexes and usage log files of full text publication archives, enables one to

investigate the relationships between downloads and citations, and develop ways to generate a more comprehensive, multi-dimensional view of the impact of publications than each of the sources can achieve individually.

There is an increasing interest in mapping techniques, and science mapping is to be qualified as one of the most important domains of informetrics as a big data science. It can be defined as the development and application of computational techniques for the visualization, analysis, and modeling of a broad range of scientific and technological activities as a whole.

2.2.3 Chapter 5

Although this book does not present details on the technical and statistical aspects of most informetric indicators, Chap. 5 does discuss three common misunderstandings as regards interpretation of particular, often used, statistical measures or techniques, related to journal impact factors as means of skewed distributions, errors in large datasets, and the interpretation of linear correlation coefficients. The conclusions are as follows.

- Journal impact factors are *no* good predictors of the citation rate of individual articles in a journal.
- Only random errors tend to cancel out in large datasets; systematic biases may remain.
- When interpreting a correlation coefficient, never rely merely on its numerical value. Consider always a scatter plot of the underlying data.

2.3 Part III. The Application Context

2.3.1 Chapter 6

In Chapter 6 an analytical distinction is made between *four* domains of intellectual activity in an assessment process, including the following activities.

- *Policy or management*: The formulation of a policy issue and assessment objectives; making *decisions* on the assessment's organizational aspects and budget. Its main outcome is a policy decision based on the outcomes from the evaluation domain.
- *Evaluation*: A specification of the evaluative framework, i.e., a set of evaluation criteria in agreement with the constituent policy issue and assessment objectives. The main outcome is a *judgment* on the basis of the evaluative framework and the empirical evidence collected.

- *Analytics*. Collecting, analyzing, and reporting *empirical* knowledge on the subjects of assessment; The specification of an assessment *model or strategy*, and the *operationalization* of the criteria in the evaluative framework. Its main outcome is an analytical report as input for the evaluative domain.
- *Data collection*. Collection of relevant data for analytical purposes, as specified in the analytical model. Data can be either quantitative or qualitative. Its main outcome is a dataset for the calculation of all indicators specified in the analytical model.

A main objective of this analytical categorization is to distinguish between scientific-methodological principles and considerations on the one hand, and policy-related, political or managerial considerations on the other. Focusing on the role of informetricians, the chapter argues as follows.

- What is of worth, good, desirable, or important in relation to the functioning of a subject under assessment, *cannot* be established in informetric, or, more general, in quantitative-empirical research. The principal reason is that one cannot infer what *ought to be* from what actually *is*.
- What informetric investigators *can* do is empirically examine value *perceptions* of researchers, the conditions under which they were formed and the functions they fulfil, but the foundation of the validity of a value is *not* a task of quantitative-empirical, informetric research. The same is true for the formation of *evaluative judgements*. The latter two activities belong to the domain of evaluation.
- Informetricians should maintain a *methodologically* neutral position towards the constituent policy issue, the criteria specified in the evaluative framework, and the goals and objectives of the assessed subject. As professional experts, their competence lies *primarily* in the development and application of analytical models *given* the established evaluative framework.
- Obviously, informetricians are allowed to form and express ‘normative’ views while assessing a unit’s worth, but when doing so they should make these explicit and give them methodologically speaking a hypothetical status.

Several types of assessment models are distinguished: peer review based versus indicator based assessments, and self-assessment versus external assessment. It distinguishes four classes of *evaluation strategies*:

- Scientific-experimental focusing on impartiality and objectivity;
- Management-oriented systems based on systems theory;
- Qualitative anthropological approaches emphasizing the importance of observation and space for subjective judgement;
- Participant oriented strategies, giving a central role to ‘consumers’.

A genuine challenge is to combine the various models and create hybrid assessment models.

2.3.2 Chapter 7

Chapter 7 illuminates the influence of non-informetric or extra-informetric factors on the development of indicators, and in this way aims to *disentangle* informetric arguments and evaluative principles, one of the key objectives of this book. Typical examples are given as regards the following issues: size dependent versus size independent indicators; focus on the top or the bottom of a performance distribution; indicator normalizations; definition of benchmark sets; and the application of a short term or a long term perspective.

For instance, a series of citation impact indicators seek an ‘optimal’ combination of publication and citation counts, and address the issue whether this optimum should be size-normalized or not. Under the surface of this seemingly technical debate, a confrontation takes place of distinct views of what constitutes genuine research performance or ‘quality’.

- According to one view, a citation-per-paper ratio is the best indicator, because it helps to detect ‘saturation’, which occurs if a research group increases its annual number of published papers while the citation impact per paper declines.
- A second view holds that such a ratio penalizes groups with a large publication output, while a large publication output should be rewarded rather than penalized.
- A third view aims to reduce the role of absolute publication numbers by proposing an indicator counting only the ‘best’ papers in terms of citation impact.
- A fourth view claims that the only good performance indicator is an efficiency indicator dividing output or impact by ‘input’ measures.

It illustrates how in seemingly technical discussions on the construction and statistical properties of science indicators, ‘*evaluative*’, *theoretical assumptions on what constitutes research performance* play an important, though often implicit, role. Such values are denoted as extra-informetric, as their validity cannot be grounded in informetric research.

Chapter 7 also presents a brief history of some of the main lines in bibliometric indicator development from the early 1960s up to date. It focuses on the *wider socio-political context* in which indicators were developed. It describes the context of their launch, not so much in terms of the *intentions* of the developers, but, at a higher analytical level, in terms of how they fit into – or are the expression of – a more general tendency in the policy, political or cultural environment in which they were developed. A base assumption is that knowledge of the wider context in which specific indicators were developed contributes to a better understanding of how and under which conditions they can be properly applied.

It is argued that in the early decades newly proposed indicators primarily aimed to solve specific policy problems and fitted in with specific national or institutional policy contexts, but during the past 10–15 years the following two tendencies emerged: on the one hand, previously developed indicators were used in more and

more policy contexts, including contexts in which they are only partially or hardly fit-for-purpose; on the other hand, indicator development was more and more driven by an internal dynamics powered by mathematical-statistical considerations.

Finally, Chap. 7 discusses the influence of *business interests* of the information industry upon the development of indicators. It concludes that since Eugene Garfield introduced the Journal Impact Factor as an ‘objective’ tool to expand the journal coverage of his citation index independently of journal publishers, the landscape of scientific information providers and users has changed significantly. While, on the one hand, politicians and research managers at various institutional levels need valid and reliable fit-for-purpose metrics in the assessment of publicly funded research, there is, on the other hand, a tendency that indicators increasingly become a tool in the business strategy of companies with product portfolios that may include underlying databases, social networking sites, or even indicator products. This may be true both for ‘classical’ bibliometric indicators and for alternative metrics.

2.3.3 Chapter 8

Chapter 8 argues that in the design of a research assessment process, one has to decide which methodology should be used, which indicators to calculate, and which data to collect. To make proper decisions about these matters, one should address the following key questions, each of which relates to a particular aspect of the research assessment process.

- What is the unit of the assessment? A country, an institution, a research group, an individual, or a research field or an international network? In which discipline(s) is it active?
- Which dimension of the research process must be assessed? Scientific-scholarly impact? Social benefit? Multi-disciplinarity? Participation in international networks?
- What are the purpose and the objectives of the assessment? Allocate funding? Improve performance? Increase regional engagement?
- What are relevant, general or ‘systemic’ characteristics of the units of assessment? For instance, to which extent are they oriented towards the international research front?

The answers to these question determine which indicators are the most appropriate in a particular assessment process. Indicators that are useful in one context, may be less so in another. A warning is issued against a practice in which particular indicators are used in a given context simply because they are available, and because they have been successfully applied in *other* contexts.

Knowledge on general characteristics of the system of units of assessment plays an important role in the formulation of the objectives of an assessment. Such

assumptions do not focus on *individual* units, but on more general or systemic characteristics of these units *as a group*. Therefore, they can be denoted as ‘meta’ assumptions, and illuminate the assessment’s *policy context*.

For instance, if an analysis of the state of a country’s science system provides evidence that researchers tend to publish mainly in national journals without a serious manuscript peer review process, it is from an informetric viewpoint *defensible* to use the number of publications in the top quartile of journals in terms of citation impact as an indicator of research performance, not so much as an *evaluation tool*, but rather as an instrument to *change* certain communication *practices* among researchers.

However, if in internationally oriented, leading universities one has to assess candidates submitting their job application, it is questionable whether it makes sense comparing applicants according to the average citation impact of the journals in which they published their papers. Due to *self-selection*, the applicants will probably publish at least a large part of the papers in good, international journals, and in this group journal impact factors hardly discriminate between high and lower performance.

2.4 Part IV. The Way Forward

2.4.1 Chapter 9

This chapter discusses a series of problems in the use of informetric indicators for evaluative purposes. Its main conclusions are as follows. Their implications for the application of indicators and for future indicator development are further discussed in Chaps. 10–12.

- *The problem of assessing individual scholars.* Calculating indicators at the level of an individual and claiming they measure *by themselves* the individual’s performance suggests a façade of exactness that cannot be justified. A valid and fair assessment of individual research performance can be conducted properly only on the basis of sufficient background knowledge on the particular role they played in the research presented in their publications, and by taking into account also other types on information on their performance.
- *The effect of a limited time horizon.* The notion of making a contribution to scientific-scholarly progress, does have a basis in reality, that can best be illustrated by referring to an *historical* viewpoint. *History will show* which contributions to scholarly knowledge are valuable and sustainable. In this sense, informetric indicators do *not* measure contribution to scientific-scholarly progress, but rather indicate attention, visibility or short term impact.
- *The problem of assessing societal impact.* Societal value cannot be assessed in a politically neutral manner. The foundation of the criteria for assessing societal value is not a matter in which scientific experts have *qualitate qua* a preferred

status, but should eventually take place in the policy domain. One possible option is moving away from the objective to evaluate an activity's societal *value*, towards measuring in a neutral manner researchers' *orientation* towards any articulated, lawful need in society, as reflected for instance in *professional contacts*. Due to time delays, emphasis on societal impact on the one hand, and assessment focus on *recent* past performance on the other, are at least partially conflicting policy incentives.

- *The effects of the use of indicators upon authors and editors.* Studies on changes in editorial and author practices under the influence of assessment exercises are most relevant and illuminative. The issue at stake is *not* whether scholars' practices *change* under the influence of the use of informetric indicators, but rather whether or not the application of such measures enhances their *research performance*. Although this is in some cases difficult to assess without extra study, other cases clearly show traces of mere indicator manipulation with no positive effect on performance at all.
- *How to deal with constitutive effects of indicators.* A typical example of a constitute effect is when research quality is more and more perceived as what citation measures. If the tendency to replace reality with symbols and to conceive these symbols as an even a higher form of reality, are typical characteristics of *magical thinking*, jointly with the belief to be able to change reality by acting upon the symbol, one could rightly argue that the un-reflected, unconditional belief in indicators shows rather strong similarities with *magical thinking*.
- More empirical research on the size of constitutive effects is urgently needed. If there is a genuine constitutive effect of informetric indicators in quality assessment at all, one should not point the critique on current assessment practices merely towards informetric indicators as such, but rather towards any claim for an *absolute status* of a particular *way* to assess quality. If the role of informetric indicators has become too dominant, it does *not* follow that the idea to intelligently combine peer judgements and indicators is fundamentally flawed and that indicators should be banned from the assessment arena. But it does show the combination of the two methodologies has to be organized in a more sophisticated and balanced manner.
- *The need for an evaluative framework and an assessment model.* Chapter 6 underlines the need to define an evaluative framework and an assessment model. To the extent that in a practical application an evaluative framework is absent or implicit, there is a vacuum, that may be easily filled either with ad-hoc arguments of evaluators and policy makers, or with un-reflected assumptions underlying informetric tools. Perhaps the role of such ad hoc arguments and assumptions has nowadays become too dominant. It can be reduced only if evaluative frameworks become stronger, and more actively determine which tools are to be used, and how. To facilitate this, informetricians should make the normative assumptions of their tools explicit, and inform policy makers and evaluators about the potential and the limits of these tools.

2.4.2 Chapter 10

Chapter 10 critically reflects on the assumptions underlying current practices in the use of informetric indicators in research assessment, and proposes a series of alternative approaches, indicating their pros and cons.

Communication Effectiveness as a Precondition for Performance

In academic institutions, especially in research universities, it is considered appropriate to stimulate academic researchers to make a solid contribution to scientific-scholarly progress. But is it defensible to require that they generate impact? What should be of primary interest to academic policy makers: importance (potential influence) or impact (actual influence)? An academic assessment policy is conceivable that rejects the claim that impact rather than importance is the key aspect to be assessed, and discourages the use of citation data as a *principal* indicator of importance.

Such an assessment process would *not* aim at measuring importance or *contribution to scientific-scholarly progress* as such, but rather *communication effectiveness*, a concept that relates to a *precondition for* performance rather than to performance itself. It expresses the extent to which researchers bring their work to the attention of a broad, potentially interested audience, and can in principle be measured with informetric tools.

Some New Indicators of Multi-dimensional Research Output

The *functions* of publications and other forms of scientific-scholarly output, as well as their *target audiences* should be taken into account more explicitly than they have been in the past. Journals with an educative or enlightening function are important in scientific scholarly communication, and tend to have a substantial societal value. Since their visibility at the international research front as reflected in citations and journal impact factors may be low, in a standard bibliometric analysis based on publication and citation counts this value may not be visible.

Scientific-scholarly journals could be systematically categorized according to their function and target audience, and separate indicators could be calculated for each category. In an analysis of research output in journals directed towards *national* audiences, citation-based indicators are *less relevant*. At the same time, in citation analyses based on the large international citation indexes focusing on the international research front, it would be appropriate to disregard such journals. It is proposed to develop indicators of *journal internationality* based on the geographical distribution of publishing, citing or cited authors. A case study shows that journal impact and internationality are by no means identical concepts. Whether or not a journal is indexed in one or more of the large citation indexes is not in all cases a good indicator of its international orientation.

Definition of Minimum Performance Standards

One possible approach to the use of informetric indicators in research assessment is a systematic exploration of indicators as tools to set minimum performance standards and define in this way a performance baseline. Important considerations in favor of this approach are as follows.

- There is evidence that citation rates are a good predictor of how peers discriminate between a ‘valuable’ and a ‘less valuable’ past performance, but that they do not properly predict within the class of ‘valuable’ performances, peers’ perception of ‘genuine excellence’. This outcome underlines the potential of informetric indicators in the assessment of the *lower part* of the quality distribution.
- Using indicators to define a baseline, researchers will most probably change their research practices as they are stimulated to meet the standards, but if the standards are appropriate and fair, this behavior will actually increase their performance and that of their institutions.
- Focusing on minimum criteria involves a shift in perspective from measuring performance *as such* towards assessing the *preconditions* for performance. Expert opinion and background knowledge will play a crucial role, not only in the definition of the standards themselves, but also in the assessment processes in which these are applied.
- The definition of minimum standards could also be applied to journal impact measures. Rather than focusing on the most highly cited journals and rewarding publications in this top set, it would be possible to discourage publication in the bottom set of journals (e.g., the bottom quartile) with the lowest citation impact.

At the *upper part* of the quality distribution, it is perhaps feasible to distinguish entities which are ‘*hors catégorie*’, or ‘*at Nobel Prize level*’. Assessment processes focusing on the *top* of the quality distributions could further operationalize the criteria for this qualification.

Policy Towards World University Rankings

Chapter 18 in *Part VI* of this book presents a comparative analysis of 5 World University Ranking Systems. Realistically speaking, rankings of world universities are here to stay. When university managers use their institution’s position in world university rankings primarily for marketing purposes, they should not disregard the negative effects such use may have upon researchers’ practices within their institution. They should also critically address the validity of the methodological claims made by producers of these ranking systems.

The following strategy towards these ranking systems is proposed. Academic institutions could, individually or collectively, seek to influence the various systems by formally sending them a request to consider the implementation of the following new features.

- Offer more advanced analytical tools, enabling a user to analyze the data in a more sophisticated manner than ranking systems currently offer.
- Provide more insight into how the methodological decisions of the producers influence the ranking positions of given universities.
- Enhance the information in the system about additional factors, such as teaching course language.

In addition, academic institutions could proceed as follows.

- Create a special university webpage providing information on a university's internal assessment and funding policies, and on its various types of performance, and giving comments on the methodologies and outcomes of ranking systems.
- Request ranking producers to make these pages directly accessible via their systems.

An Alternative Approach to Performance Based Funding

Major criticisms towards national research assessment exercises underline their bureaucratic burden, costs, lack of transparency, and their Matthew effect. Adopting an informetric viewpoint, an alternative assessment model is described that would require less efforts, be more transparent, stimulate new research lines, and reduce to some extent the Matthew Effect. The main features are as follows.

- The base unit of assessment is an *emerging group*, a small research group with a great scientific potential. Acknowledged as a 'rising star', the director has developed a promising research program, and has already been able to establish a small research unit.
- The profile of an emerging group should be further operationalized into a set of minimum quantitative criteria, taking into account the communication practices and funding opportunities in the group's subject field.
- In the assessment procedure, institutions submit groups rather than individual staff. Submissions provide information on the group's past performance and a future research programme, that are evaluated in a peer review process, informed by appropriate informetric indicators.
- The primary aim of the peer review is to define the minimum standards in operational terms and assess whether the submitted groups comply with these standards. These standards constitute a precondition for the group's future performance.
- There would be no need to rank groups, assign ratings, discriminate between 'top' and 'almost top' groups, or make funding decisions. Funding decisions take place within their institution.
- To stimulate the implementation of quality control processes within an institution, the availability of a certain amount of funding from internal,

performance-based allocation processes could be posed as a necessary condition.

- A part of public funding (block grant) could be allocated to institutions as a lump sum on the basis of the number of acknowledged emerging groups.

The practical realization of these proposals requires a large amount of informetric research and development. They constitute important elements of a wider R&D program in *applied evaluative informetrics*. These activities tend to have an applied character and often a short term perspective, and focus on the development side of R&D. They should be conducted in a close collaboration between informetricians and external stakeholders. Chapters 11 and 12 propose strategic, longer term research projects with a great potential for research assessment.

2.4.3 Chapter 11

Chapter 11 discusses the potential of altmetrics. A multi-dimensional conception of altmetrics is proposed, namely as traces of the computerization of the research process, and conceived as a tool for the practical realization of the ethos of science and scholarship in a computerized or digital age. Three drivers of development of the field of altmetrics are distinguished.

- In the policy domain: An increasing awareness of the multi-dimensionality of research performance, and an emphasis on societal merit.
- In the domain of technology: The development of information and communication technologies (ICTs), especially social media.
- From the scientific-scholarly community itself: The Open Science movement.

Four aspects of the computerization of the research process are highlighted: the computerization of research data collection and analysis; the digitization of scientific information; the use of computerized communication technologies; and informatization of research assessment.

Michael Nielsen's set of creative ideas constitute a framework in which altmetrics can be positioned. He argued that "to amplify cognitive intelligence, we should scale up collaborations, increasing cognitive diversity and the range of available expertise as much as possible". The role of altmetrics and other informetric indicators would not merely be, passively, to provide descriptors, but also actively, or proactively, to establish and optimize, Nielsen's "architecture of attention", a configuration that combines the efforts of researchers and technicians on the one hand, and the wider public and the policy domain on the other, and that "directs each participant's attention where it is best suited—i.e., where they have maximal competitive advantage".

It should not be overlooked that a series of distinctions made in 'classical' research assessment are most relevant in connection with altmetrics as well: scientific-scholarly versus societal impact; attention versus influence; opinion

versus scientific fact; peer-reviewed versus non-peer reviewed; intended or unintended versus constitutive effects of indicators.

2.4.4 Chapter 12

Chapter 12 proposes a series of alternative approaches to the development of new informetric indicators for research assessment that put a greater emphasis on *theoretical models* for the interpretation of informetric indicators, and on the use of techniques from *computer science* and the newly available *information and communication technologies*.

Towards New Indicators of the Manuscript Peer Review Process

A proposal to develop new indicators of the manuscript peer review process is based on the following considerations.

- Manuscript peer review is considered important by journal publishers, editors and researchers. But the process itself is still strikingly opaque.
- Reviewers tend to receive little training in what is one of the key academic activities, and there is little evidence of any standardization in how review reports are composed.
- There is little systematic, objective information on the quality of the process across journals and subjects, and on its effect upon the quality of submitted papers.
- With peer review largely a black box, proxies for its quality have grown up, most notably the journal impact factor (JIF) based on citation counts.
- But the digitization of scientific information offers great potential for the development of tools to allow peer review to be analyzed directly.

The objectives and set-up of the project are as follows.

- The aim of the project is to build up an understanding for each discipline of what is considered a reasonable quality threshold for publication, how it differs among journals and disciplines, and what distinguishes an acceptable paper from one that is rejected.
- The analysis consists of two phases, an explorative phase in which a classical-humanities approach is dominant, aimed to develop a conceptual model; and a data mining phase, applying techniques from digital humanities.
- Taking into account a journal's scope and instructions to reviewers, the various elements of a review report should be analyzed. Statements are categorized in terms of aspect and modality. Standards and assumptions applied by a reviewer are identified.

- Relevant concepts and their indicators are developed, including the formative content of a review report, and the distance a reviewer maintains towards his own methodological and theoretical views.

The project could have the following outcomes.

- It provides insight into the effects and ‘added value’ of peer review upon manuscript quality.
- It defines a set of minimum quality standards per journal and per discipline. This information enhances the transparency of the review process, and is useful for editors, reviewers and authors.
- It proposes indicators characterizing the various aspects of the process, for instance, its formative effect, and other tools to monitor the process.
- The results may be used to improve the quality of the peer review process.
- Ultimately, perhaps journal-level metrics can be validated that can supersede proxies such as journal impact factors.

Towards an Ontology-Based Informetric Data Management System

During the past decades, development and application of informetric indicators in research assessment have posed a series of challenges to the management—collection, handling, integration, analysis and maintenance—of informetric data, and in the design of S&T indicators. There are data-related, concept-related and maintenance-related issues.

To solve these issues, it is proposed to develop an Ontology-Based Data Management (OBDM) system for research assessment, along the lines set out in Daraio et al (2016). The key idea of OBDM is to create a three-level architecture, constituted by a) the ontology; b) the data sources; and c) the mapping between the two.

An *ontology* can be defined as a conceptual, formal description of the domain of interest to a given entity (e.g., organization or community of users), expressed in terms of relevant concepts, *attributes* of concepts, *relationships* between concepts, and *logical* assertions characterizing the domain knowledge. The *sources* are the data repositories accessible by the organization in which data concerning the domain are stored. The mapping is a precise specification of the correspondence between the data contained in the *data sources* on the one hand, and the elements of the *ontology* on the other.

The main advantages of an OBDM approach are as follows

- Users can access the data by using the elements of the ontology. A strict separation exists between the conceptual and the logical-physical level.
- By making the representation of the domain explicit, the acquired knowledge can be easily re-used.

- The mapping layer explicitly specifies the relationships between the domain concepts in the ontology and the data sources. It is useful also for documentation and standardization purposes.
- The system is more flexible. It is for instance not necessary to merge and integrate all the data sources at once, which could be extremely time consuming and costly.
- The system can be more easily extended. New elements in the ontology or data sources can be added incrementally when they become available. In this sense, the system is dynamical and develops over time.

Towards Informetric Self-assessment Tools

The creation of an informetric self-assessment tool at the level of individual authors or small research groups is proposed. Such an application would be highly useful, but is currently unavailable. A challenge is to make an optimal use of the potentialities of the current information and communication technologies and create an online application based on key notions expressed decades ago by Eugene Garfield about author benchmarking, and by Robert K. Merton about the formation of a *reference group*.

The general concept is as follows.

- In a first step, the application enables an author to define a set of publications he/she wishes to take into account in the assessment. It is important that there is a proper data verification tool at hand.
- In a next step, a benchmark set is created of authors with whom the assessed author can best be compared, along the lines adopted by Eugene Garfield in his proposal for an algorithm to create for a given author under assessment, a set of ‘candidate’ benchmark authors who have bibliometric characteristics that are similar to those of the given author.
- Garfield’s idea could be further developed by creating a flexible benchmarking feature as the practical realization of Robert K. Merton’s notion of a reference group, i.e., the group with which individuals compare themselves, but to which they do not necessarily belong but aspire to.
- The calculated indicators should be the result of simple statistical operations on absolute numbers. Not only the outcome, but also the underlying numbers themselves should be visible.
- In addition, researchers must have the opportunity to decompose and reconstruct indicators. It should also be possible to insert particular data manually.

An adequate assessment of individual research performance can take place only by taking into account multiple sources of information about their performance. This does not mean that bibliometric measures in the assessment of individuals are irrelevant, especially when used as self-assessment tools. The proposed tool would be useful for the following reasons.

- In their self-assessments, researchers may wish to calculate specific fit-for-purpose indicators that are not ‘standard’ and therefore unavailable at the websites of indicator producers. The proposed self-assessment tool could be a genuine alternative to using journal impact factors or h-indices.
- Even if one is against the use of informetric indicators in individual assessments, one cannot ignore their availability to a wider public. Therefore, it would be useful if researchers had an online application to check the indicator data calculated about them, and to decompose the indicators’ values.
- In this way they would learn more about the ins and outs of evaluative informetrics, and, for instance, become aware of how the outcomes of an assessment depend upon the way benchmark sets are being defined. This would enable researchers to defend themselves against inaccurate calculation or invalid interpretation of indicators.

Towards Informetric Models of Scientific Development

Scientifically developing countries need tools to monitor the effectiveness of their research policies in a framework that categorizes national research systems in terms of *the phase* of their scientific development. Leaving out the dynamical aspects of a system gives an incomplete picture. The current section presents a model of a country’s scientific development using bibliometric indicators based on publications in international, peer-reviewed journals.

The model aims to provide a framework in which the use of informetric indicators of developing countries makes sense, as an alternative to common bibliometric rankings based on publication and citation counts from which the only signal is that such countries tend to feature in the bottom.

A simplified and experimental bibliometric model for different phases of development of a national research system distinguishes four phases:

- *Pre-development phase.* Low research activity without clear policy of structural funding of research
- *Building up.* Collaborations with developed countries are established; national researchers enter international scientific networks
- *Consolidation and expansion.* The country develops its own infrastructure; the amount of funds available for research increases
- *Internationalization.* Research institutions in the country start as fully-fledged partners, increasingly taking the lead in international collaboration

The distinction into phases is purely *analytical* and does not imply a *chronological order*. The model assumes that during the various phases of a country’s scientific development, the number of published articles in peer-reviewed journals shows a more or less continuous increase, although the rate of increase may vary substantially over the years and between countries.

It is the share of a country's internationally co-authored articles that discriminates between the various phases in the development. The model also illustrates the *ambiguity* of this indicator, as a high percentage of internationally co-authored papers at a certain point in time may indicate that a country is either in the building up or the internationalization phase.

The model is applied to empirical data on South-East Asian countries, and on Arab Gulf states and neighboring countries in the Middle East, and provided evidence, for instance, that while Saudi Arabia is in the building-up phase, Iran has already entered the internationalization phase.

2.5 Part V. Lectures

2.5.1 Chapter 13

Chapter 13 presents two visionary papers published by Derek de Solla Price, the founding father of the science of science. It presents his view on the scientific literature as a network of scientific papers, and introduces important informetric concepts, including research front and immediacy effect. Next, the chapter shows how his pioneering work on modelling the relational structure of subject space evolved into the creation of a series of currently available, advanced science mapping tools.

2.5.2 Chapter 14

A comparative analysis of three big, multi-disciplinary citation indexes is presented in Chap. 14. It starts with a presentation of the basic principles of the Science Citation Index (SCI, later Thomson Reuters' Web of Science, currently Clarivate Analytics), a multi-disciplinary citation index created by Eugene Garfield in the early 1960s. Next, it presents a study conducted in 2009 comparing the Web of Science with Scopus, a comprehensive citation index launched by Elsevier in 2004, and a recent study comparing Scopus with an even more comprehensive citation index, Google Scholar, also launched in 2004.

2.5.3 Chapter 15

Chapter 15 discusses studies on the relationship between science and technology. It starts with presenting the pioneering work by Francis Narin and co-workings on the citation analysis of the linkage between science and technology. Next, it discusses

several theoretical models of the relationship between science and technology. As an illustration of an analysis of the development of a technological field, it presents key findings from a study on industrial robots. The chapter ends with illustrating the limitations of citation analysis of the scientific literature for the measurement of technological performance.

2.5.4 Chapter 16

Chapter 16 deals with of journal metrics. It starts with a discussion of the journal impact factor, probably the most well-known bibliometric measure. It shows some of its technical limitations and dedicates in an analysis of editorial self-citations special attention to its sensitivity to manipulation. Next, a series of alternative journal citation measures is presented, SJR, Eigenfactor, SNIP, CiteScore, and indicators based on usage.

2.5.5 Chapter 17

Definitions and properties of a series of informetric indicators discussed in earlier chapters of this book, are presented in Chap. 17: relative citation rates, h-index, Integrated Impact Indicator, usage-based indicators, social media mentions, and research efficiency or productivity measures. It highlights their potential and limits, and gives typical examples of their application in research assessment.

2.6 Part VI. Papers

2.6.1 Chapter 18

To provide users insight into the value and limits of world university rankings, Chap. 18 presents a comparative analysis of 5 World University Ranking Systems: ARWU, the Academic Ranking of World Universities, also indicated as ‘Shanghai Ranking’; The Leiden Ranking created by the Centre for Science and Technology Studies (CWTS); The Times Higher Education (THE) World University Ranking; QS World University Rankings; and an information system denoted as U-Multirank created by a consortium of European research groups.

As all ranking systems claim to measure essentially academic excellence, one would expect to find a substantial degree of consistency among them. The overarching issue addressed in this Chap. 6 is the assessment of this consistency-between-systems. To the extent that a lack of consistency is found,

what are the main causes of the observed discrepancies? A series of analyses is presented, from which the following conclusions were drawn.

- Each ranking system has its proper orientation or ‘profile’; there is no ‘perfect’ system. There is only a limited overlap between the top 100 segments of the 5 rankings.
- What appears in the top of a ranking depends to a large extent upon a system’s geographical coverage, rating methodologies applied, indicators selected and indicator normalizations carried out.
- Current ranking systems are still one-dimensional in the sense that they provide finalized, seemingly unrelated indicator values rather than offer a dataset and tools to observe patterns in multi-faceted data.
- To enhance the level of understanding and adequacy of interpretation of a system’s outcomes, more insight is to be provided to users into the methodological differences between the various systems, especially on how their institutional coverage, rating methods, the selection of indicators and their normalizations influence the ranking positions of given institutions.

2.6.2 Chapter 19

A statistical analysis of full text downloads of articles in Elsevier’s ScienceDirect covering all scholarly disciplines reveals large differences between disciplines, journals, and document types as regards their download frequencies, their skewness, and their correlation with Scopus-based citation counts. Download counts tend to be two orders of magnitude higher and are less skewedly distributed than citations. Differences between journals are discipline-specific.

Despite the fact that in all analysed journals download and citation counts per article positively correlate, the following factors differentiate between downloads and citations.

- *Usage leak.* Not all full text downloads of a publisher archive’s documents may be recorded in the archive’s log files.
- *Citation leak.* Not all relevant sources of citations may be covered by the database in which citations are counted.
- *Downloading* the full text of a document does not necessarily mean that it is fully read.
- *Reading and citing populations may be different.* For instance, industrial researchers may read scientific papers but not cite them as they do not publish papers themselves.
- *Number of downloads depends upon type of document.* For instance, editorials and news items may be heavily downloaded but poorly cited compared to full length articles.

- *Downloads and citations show different obsolescence functions.* Download and citation counts both vary over time, but in a different manner, showing different maturing and decline rates.
- *Downloads and citations measure different aspects.* Short term downloads tend to measure readers' awareness or attention, whereas citations result from authors' reflection upon relevance.
- *Downloads and citations may influence one another in multiple ways.* More downloads may lead to more citations, but the reverse may be true as well. Articles may gain attention and be downloaded because they are cited.
- *Download counts are more sensitive to manipulation.* While citations tend to be regulated by the peer review process, download counts are more sensitive to manipulation.
- *Citations are public, usage is private.* While citations in research articles in the open, peer reviewed literature are public acts, downloading documents from publication archives is essentially a private act.



<http://www.springer.com/978-3-319-60521-0>

Applied Evaluative Informetrics

Moed, H.F.

2017, XXI, 312 p. 165 illus., 148 illus. in color.,

Hardcover

ISBN: 978-3-319-60521-0