ETH zürich

Infozine S·1

Special Issue 1 | December 12, 2016

Metrics in Research

For better or worse?

Editors:
Jožica Dolenc
Philippe Hünenberger
Oliver Renn

Infozine is published by the
Chemistry | Biology | Pharmacy Information Center, ETH Zurich

infozentrum.ethz.ch
Chemie | Biologie | Pharmazie

Imprint

Infozine Special Issue S1
Metrics in Research – For better or worse?

Editors:
Prof. Dr. Philippe Hünenberger (Guest Editor)
ETH Zürich, Laboratory of Physical Chemistry
Vladimir-Prelog Weg 2
8093 Zürich, Switzerland
Phone +41 44 632 55 03, phil@igc.phys.chem.ethz.ch

Dr. Jožica Dolenc
ETH Zürich, Chemistry | Biology | Pharmacy Information Center
Vladimir-Prelog Weg 10
8093 Zürich, Switzerland
Phone +41 44 632 29 48, dolenc@chem.ethz.ch

Dr. Oliver Renn
ETH Zürich, Chemistry | Biology | Pharmacy Information Center
Vladimir-Prelog Weg 10
8093 Zürich, Switzerland
Phone +41 44 632 29 64, renn@chem.ethz.ch

Copy-Editors: Oliver Renn, Philippe Hünenberger
Layout: Oliver Renn
Cover illustration: Oliver Renn, iPad Pro Drawing with the App Procreate and Apple Pencil

Download and use of full text
All contributions have been archived in ETH Zurich's e-collection and can be retrieved at http://e-collection.library.ethz.ch.
A PDF of the Special Issue can be retrieved from http://www.infozentrum.ethz.ch/en/downloads-icbp/brochures/.
DOI numbers have been assigned by the ETH Library, DOI-Desk.
Individual copyrights have been assigned to the contributions.

Please use the hashtag
#infozineS1metrics @infozentrum (Instagram, Facebook)) @icbpeth (twitter)
when mentioning articles in social media.

Infozine and its Special Issues are published by the Chemistry | Biology | Pharmacy Information Center,
a function within the Department of Chemistry and Applied Biosciences and the
Department of Biology at the ETH Zurich.
Infozine is published as an English and German edition, Special Issues in English only.
ISSN (English Edition) 2504-1851
ISSN (German Edition) 2504-1843
www.infozentrum.ethz.ch

Jožica Dolenc, Philippe Hünenberger, Oliver Renn

Editorial
Metrics in Research – For better or worse?

If you are an academic researcher but did not earn (yet) your Nobel prize or your retirement, it is unlikely you never heard about research metrics. These metrics aim at quantifying various aspects of the research process, at the level of individual researchers (*e.g.* *h*-index, altmetrics), scientific journals (*e.g.* impact factors) or entire universities/countries (*e.g.* rankings). Although such "measurements" have existed in a simple form for a long time, their widespread calculation was enabled by the advent of the digital era (large amount of data available worldwide in a computer-compatible format). And in this new era, what becomes technically possible will be done, and what is done and appears to simplify our lives will be used. As a result, a rapidly growing number of statistics-based numerical indices are nowadays fed into decision-making processes. This is true in nearly all aspects of society (politics, economy, education and private life), and in particular in research, where metrics play an increasingly important role in determining positions, funding, awards, research programs, career choices, reputations, *etc*...

In somewhat simplistic terms, numerical indicators allow to simplify the choice between two complex options A and B, associated with quality indices N_A and N_B, in two ways. First, the choice is *immediate*, as it boils down to solving an inequality (if $N_A > N_B$, pick A, otherwise pick B, no need to dive into the painful complexity of options A and B). Second, the choice is *objective* (as long as the procedures to derive N_A from A and N_B from B are deterministic and identical, the comparison itself is unbiased). In a society where public resources are tight, so that their efficient use and fair distribution must be justified, and where the time of decision-makers is precious, speed and objectiveness are clearly two major assets of metrics-assisted decision-making. And let us not forget a third psychological factor: the human brain (especially that of scientists and managers!) is by construction fascinated by numbers, and their strong power for classification and rationalization.

There are, however, two major downsides to metrics-based decision making. First, the reduction of a complex entity A (university, scientist, project, journal, publication) into a single number N_A representing quality is a projection from a high-dimensional space to a single number. Thus, it will always be *reductionistic* (incomplete, simplistic, distortive, dehumanized), and may even in some cases be entirely off-topic. In fact, most current research metrics do not measure a scientific quality, but rather a scientific output or impact (*i.e.* only – and arguably – one component of quality). Second, the systematic coupling of a reductionistic index N_A to decisions strongly influencing A induces a *feedback loop*, in which the entity A will start to optimize itself against N_A rather than against quality in a broader sense. At the extreme, this may result in a research community striving very competitively for output and impact, and considering collaboration, diversity, creativity, curiosity, risk-taking, education and ethics (definitely other components of scientific quality!) as dispensable virtues.

The "metrics system" is not a perspective for the future – it is already well-installed and gaining strength, because it is fundamentally compatible with the usual mechanisms and mainstream values of a modern society in the digital era. Yet, individual opinions diverge widely concerning how to weigh the above strengths and flaws of this system, and whether one should strive to reinforce it, to improve/refine it, or to abolish it.

Clearly, the debate is important (maybe vital!) for the future of academic research. For this reason, in this special issue of *Infozine*, we have collected 18 opinion statements concerning the topic of "research metrics". The potential contributors have been invited with the goal of providing a wide spectrum of opinions (supportive, moderate, or critical) and covering a wide spectrum of perspectives (including those of professors, students, publishers, editors, and metrics providers). This special issue is meant to provide a broad and unbiased spectrum of possible viewpoints and arguments on the topic, with the idea to feed into the thinking of the readers, and help them define lucidly their own position regarding the issue.

The "metrics system" is *de facto* already in place, and it is spontaneously self-reinforcing. You may decide to actively support it, or to accept and do the best out of it, or to fight against it … but, as always, it is extremely unwise to let others decide for you.

Citation: Dolenc J, Hünenberger P, Renn O. Editorial: Editorial
Metrics in Research – For better or worse?
Infozine **2016,** Special Issue 1, 1, DOI: 10.3929/ethz-a-010748857

Published: December 12, 2016

Contents

Oliver Renn, Jožica Dolenc, Joachim Schnabl
ETH Zürich

A brief visual history of research metrics

Research metrics started in chemistry

Research metrics are relatively recent considering that scientific journals, historically the main basis for these metrics, have been already established more than 350 years ago. The first research metrics introduced were bibliometrics, *i.e.* they relied on the statistical analysis of publications, books or journals (Figure 1). Among the methods used for bibliometrics, citation analysis is the most important. The first citation analysis was executed by Gross and Gross in 1927 [1]. They counted citations in articles (manually!) so as to produce a ranking of chemistry journals, aimed as a guide to help chemistry librarians decide which journals to subscribe to.

It was a chemist, Eugene Garfield, who suggested in 1955 [2] to systematically count and analyze citations in the scientific literature. At that time, he was publisher of *Current Contents* (CC), a directory of scientific journals' tables of contents. Scientists born in the sixties may still remember the times when one visited the library every week to browse through the latest CC issue, in order to stay tuned to new publications of interest.

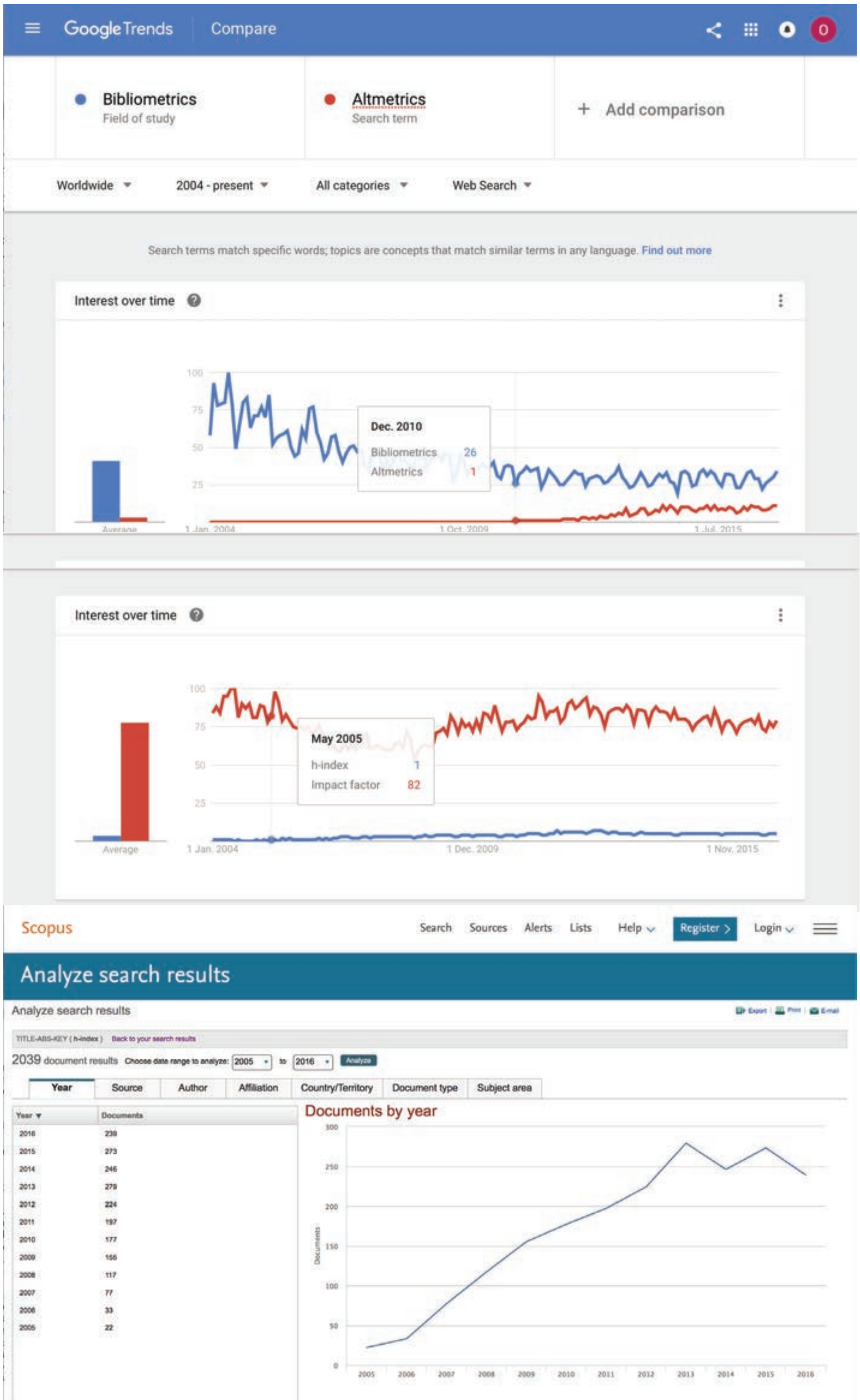


Figure 1: Top: Comparison of Google searches for bibliometrics *vs.* altmetrics and impact factor *vs.* *h*-index from 2004 to 2016. Source: Google Trends. Bottom: No. of articles published annually from 2004 to 2016 with "*h*-index" in the title, abstract or keyword list of the article. Source: Scopus

Journal-based metrics

In 1960, Eugene Garfield founded the *Institute for Scientific Information* (ISI), later acquired by Thomson Reuters in 1992 and sold to private equity investors in 2016. ISI offered bibliographic database services and in 1964, the *Science Citation Index* (SCI) was launched. As part of the SCI, the *Journal Citation Reports* (JCR) were published, based on which the *Impact Factor* (IF) was invented by Garfield. Since 1975, IFs are calculated yearly for all journals that are encompassed within the SCI. The IF of an academic journal is a measure reflecting the yearly average number of citations to recent articles published in that journal. More precisely, it is defined as follows:

> **Definition of the impact factor**
> **In any given year, the impact factor of a journal is the number of citations received by articles published in that journal during the two preceding years, divided by the total number of articles published in that journal during these years.**
> **For example, a journal had an impact factor of 4.7 in 2015 if its papers published in 2013 and 2014 received on average 4.7 citations each in 2015.**

Garfield intended to provide researchers with a tool to find the literature they needed to read and use, *i.e.* the CC and later the SCI, and to provide librarians with a tool to help them decide which journals to subscribe to, *i.e.* the IF. However, over the years, the IF evolved into a tool that ranked journals for publishers and researchers, *i.e.*, the publications of scientists became implicitly "evaluated" based on the IF of the journal they appeared in. Scoring scientists was possible by monitoring the total number of citations they received, but also by summing up the IFs of the journals accepting their articles.

Since the eighties, publishers anxiously await the new releases of the IF (Figure 2), usually in June, as the IF may largely influence the number and quality of the manuscripts they will receive in the following year. The IF thus ended up acting as a *circulus vitiosus* (a lower IF resulting in poorer submissions, likely less often cited manuscripts, and again a lower IF in the next year) or as a *circulus virtuosus* (a higher IF resulting in better submissions, likely more often cited manuscripts, and again a higher IF in the next year).

Eugene Garfield had frequently warned about the misuse in evaluating individuals because there is a wide variation from article to article within a single journal [3].

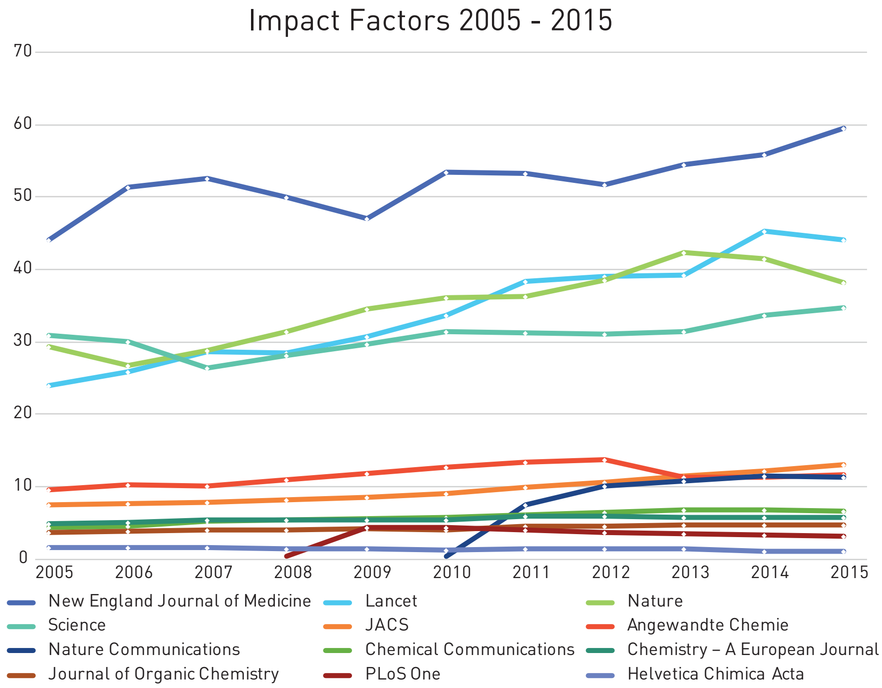


Figure 2: Development of the impact factor of selected journals in chemistry and life sciences

As the US-American SCI did not cover many European journals, and it was difficult for European publishers to get into the SCI so as to receive an IF, European scientists and publishers tried to establish alternative journal impact factors, like *e.g.* the *European Impact Factor* (EIF).

One would need an entire book to describe all existing citation-based (journal) impact factors; such a book has actually been published this year, the "Handbook of Bibliometric Indicators [4]", which is an encyclopedia describing all known research metrics [5], with special attention to the mathematics involved.

A few of the other, alternative citation-based journal metrics [6] can be found in *Scopus*, Elsevier's *Abstract & Indexing* (A&I) database that was launched in 2004:

SCImago Journal Rank (SJR): The SJR is weighted by the prestige of a journal. Subject field, quality and reputation of the journal have a direct effect on the value of a citation. Also, SJR normalizes for differences in citation behavior between subject fields. It is an indicator which ranks journals by their "average prestige per article" and can be used for journal comparisons in the scientific evaluation process. The SJR relies on a citation window of four years.

Impact per Publication (IPP): IPP is the average number of citations received in a particular year by papers published in the journal during the three preceding years. A citation window of three years is considered to be the optimal time period to accurately measure citations in most subject fields.

Source Normalized Impact per Paper (SNIP): The SNIP measures the contextual citation impact of a journal by weighting citations relative to the total number of citations in a subject field.

Yet another example for a journal metrics is the *Eigenfactor*, developed by Jevin West and Carl Bergstrom at the University of Washington [7]. Here, journals are rated according to the number of incoming citations, citations from highly ranked journals resulting in a higher score. The *Eigenfactor* score, and the closely related *Article Influence Score*, are calculated by eigenfactor.org, and are freely available as an alternative to the more standard IF. Originally, the *Eigenfactor* only ranked journals, but it has recently been extended to the author level [8].

Author-Based research metrics

The criticism that the IF is journal-based, and that poorer scientists may thus benefit from high IFs originating from the work of more cited (more talented?) researchers (and conversely for low IFs) has led to the introduction of the *h*-index.

The *h*-index was suggested by Jorge E. Hirsch in 2005, in an article [9] that started with

> I propose the index *h*, defined as the number of papers with citation number ≥*h*, as a useful index to characterize the scientific output of a researcher.

The *h*-index is an author-level metric that measures both the productivity and citation impact of the publications of a researcher. The index is based on the set of the most cited papers of the scientist and the number of citations they have received in other publications. Unlike the IF of a journal, which may fluctuate, the *h*-index of an author can only increase.

The A&I database *Scopus* made the *h*-index available quite early (Figure 3). A few years later it was also possible to look up the *h*-index in the ISI *Web of Science* (WoS), although less easily. Today, the *h*-index of an individual researcher can also be looked up in *ResearchGate* and *Google Scholar*. To see an *h*-index in *Google Scholar*, one needs to set up a *Google Scholar* profile, and one can use *e.g.* the browser add-ons developed by Giovambattista Ianni [10] (Figure 4).

Figure 4: Example of an Author Profile in Google Scholar

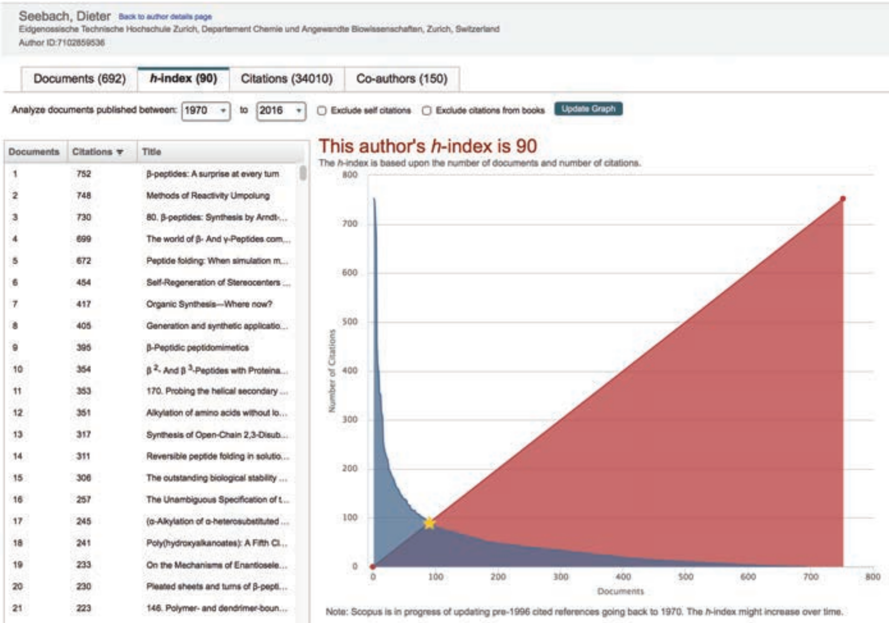


Figure 3: Citation overview and *h*-index of a researcher in Scopus

Why is the *h*-index of a researcher often different in ResearchGate, Scopus and Web of Science?

The *h*-index depends on the selection of journals used to count the citations entering into the calculation of the score (Table 1). As *Scopus* considers the largest numbers of journals, the *h*-index tends to be higher than in WoS, at least for younger scientists. Older scientists may have a higher *h*-index in WoS, as the database also covers their early publications. *Scopus* was first limited to post-1995 publications, but has recently started to add also earlier (pre-1996) publications, which results in an increased *h*-index also for seniors. *Google Scholar* often returns the highest *h*-index, as the basis of the calculation is not an A&I database but websites. *Google* has an efficient algorithm to detect URLs referring to a paper, but the system is not perfect and possibly duplicates are inappropriately counted multiple times in the *h*-index. Additionally, there is not always a clear distinction between the role of a scientist as *e.g.* book author, journal contributor, or editor.

The *h*-index can also be used to express the productivity and impact of a scholarly journal or a group of scientists in numbers, *e.g.* a department, a university, or a country.

Although it has some benefits over the use of journal IFs at the author level, the *h*-index was also criticized as soon as it was established, as it did and does not reflect truly the scientific importance of an author.

Table 1: Comparison of h-indices of three researchers A, B, and C in Google Scholar, Scopus, Web of Science, ResearchGate

	A (R)	B (G)	C (Z)
Google Scholar*	16	n/a	66
Google Scholar**	15	83	82
Scopus	13	65	57
Web of Science	14	62	56
ResearchGate	15	61	59

* calculated through browser add-on
** based on Author's Profile
n/a = no Google Scholar Profile

Webometrics

Before scholarly communication was transitioned into the web, citations could only be tracked based on references in peer-reviewed journals. Webometrics also includes citations and referrals in the web and as the social web evolved this led to *altmetrics*, an alternative to more traditional citation-impact metrics, such as the IF and *h*-index. Originally, altmetrics did not cover citation counts.

Altmetrics

Webometrics including altmetrics arose when, in March 2009, the journal *PLoS* (Public Library of Science) introduced article-level metrics that measure how articles have been viewed, cited, and discussed. On October 26, 2010, *Jason Priem*, University of North Carolina-Chapel Hill, *Dario Taraborelli*, Wikimedia Foundation, *Paul Groth*, VU University Amsterdam and *Cameron Neylon*, Science and Technology Facilities Council, coined the term altmetrics by publishing the article "altmetrics: a manifesto" [11]. Similar to Eugene Garfield with the IF, their intention was not to provide a researcher ranking tool (although it evolved into one) as the first sentences of their manifesto indicates:

> No one can read everything. We rely on filters to make sense of the scholarly literature, but the narrow, traditional filters are being swamped. However, the growth of new, online scholarly tools allows us to make new filters; these altmetrics reflect the broad, rapid impact of scholarship in this burgeoning ecosystem. We call for more tools and research based on altmetrics. As the volume of academic literature explodes, scholars rely on filters to select the most relevant and significant sources from the rest. Unfortunately, scholarship's three main filters for importance are failing.

The three filters are peer-review (failing as most papers are eventually published somewhere), citation counting (failing because too slow and not taking into account the impact outside the academic environment) and impact factor (trade secret, gaming possible). According to Priem *at al.*, altmetrics are defined as "the creation and study of new metrics based on the social web for analyzing, and informing scholarship". The term was actually coined by Jason Priem, through a Twitter feed on September 29, 2010 (Fig. 5).

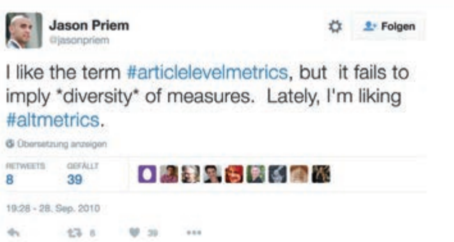


Figure 5: First tweet mentioning the term altmetrics

Altmetrics was conceptualized in 2012, when researchers, editors and publishers from the American Society for Cell Biology met in December 2012 during their annual meeting in San Francisco, and agreed that there was a need to improve the ways in which the outputs of scientific research were evaluated. The group subsequently circulated a draft declaration among various stakeholders, which resulted in the "*San Francisco Declaration on Research Assessment"* [12].

Soon, start-ups took up the idea of providing those alternative metrics. Among the first and best-known (the company name is indeed often mixed up with the concept) is *Altmetric*. *Altmetric* was founded by Euan Adie in 2011 and grew out of the burgeoning altmetrics movement. His team introduced an altmetrics app at the "*Elsevier's Apps for Science*" competition, and ended up winning. The first standalone version of the *Altmetric Explorer* was released in February 2012. Today, *Altmetric* is part of *Digital Science*, owned by *Springer Nature*. *Altmetric* has several products [13], including free applications like the *Altmetric Bookmarklet for Researchers*. Once installed as a browser extension, it gives you instant access to article-level metrics for any recent paper (Figure 6). There are numerous examples of research articles that receive attention mainly because of their titles (Figure 7).

Figure 6: Altmetric detailed view of an article with a high altmetric score

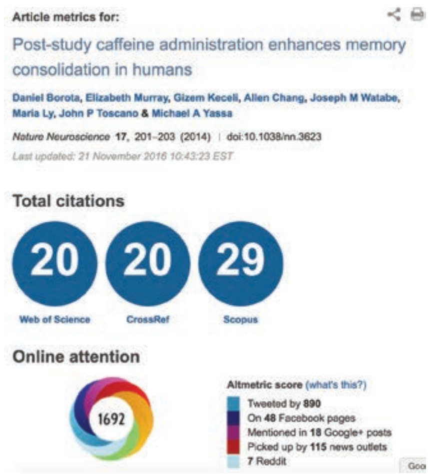
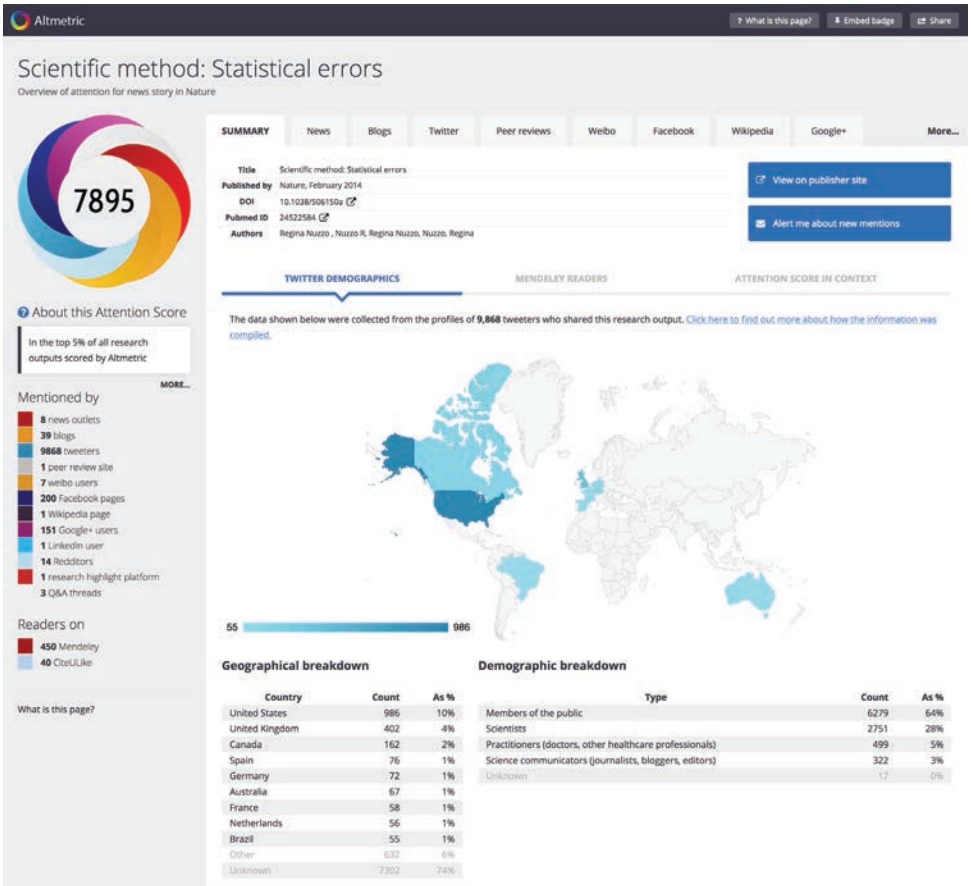


Figure 7: Altmetric score for a popular title



Another altmetrics start-up is *Impact-Story* [14], co-founded by Jason Priem. *Impact Story* (Figure 8) began as a hackathon project at the *Beyond Impact* workshop in 2011. As the hackathon ended, some continued working, eventually completing a 24-hours coding marathon to finish a prototype. In early 2012, *Impact Story* was funded by the *Open Society Foundation* and today it is funded by the *Alfred P. Sloan Foundation*.

PlumAnalytics [15] is the third example. It was founded early 2012, with the vision of bringing modern ways of measuring research impact, to be used by individuals and organizations analyzing research. In 2014, *Plum Analytics* became a part of *EBSCO Information Services*.

Sources tracked by altmetric providers are, for example,

- Public policy documents
- Mainstream media
- Online reference managers, like Mendeley
- Post-publication peer-review platforms, like Pubpeer and Publons
- Wikipedia
- Open Syllabus Project
- Blogs (over 9,000 academic and non-academic blogs every day)
- Citations
- Research highlights from *e.g.* F1000
- Social Media, like Facebook (mentions on public pages only), Twitter, Google+, LinkedIn, YouTube, Reddit

Scopus had shown article-level metrics that come from *Altmetric* until 2015, but then decided to display metrics from *Snowball*. *Snowball Metrics* [16] (Figure 9) is a bottom-up initiative, owned by research-intensive universities around the globe, to ensure that its outputs meet their own needs, rather than being imposed by organizations with potentially distinct goals, such as funders, agencies, or suppliers of research information. They are collaborating with a commercial publisher, *Elsevier*, to ensure that the methodologies are technically feasible before they are shared with the sector.

The recipes for *Snowball Metrics* can be found in the second edition of the "*Snowball Metrics Recipe Book*" [15]. These recipes can be used free-of-charge by any organization for their own purposes and, if applicable, under their own business.

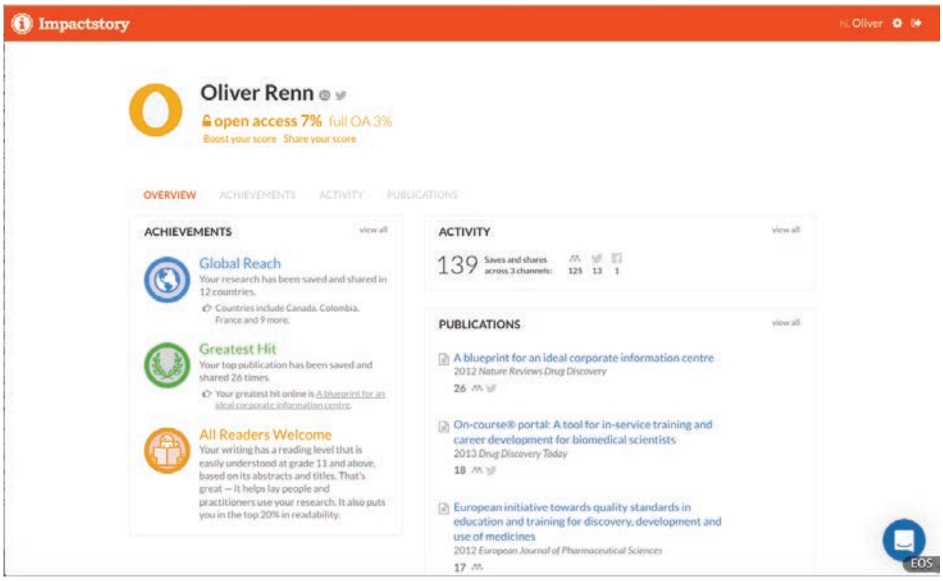


Figure 8: Top: ImpactStory Author dashboard, free to use with a Twitter log-in and ORCID synchronization

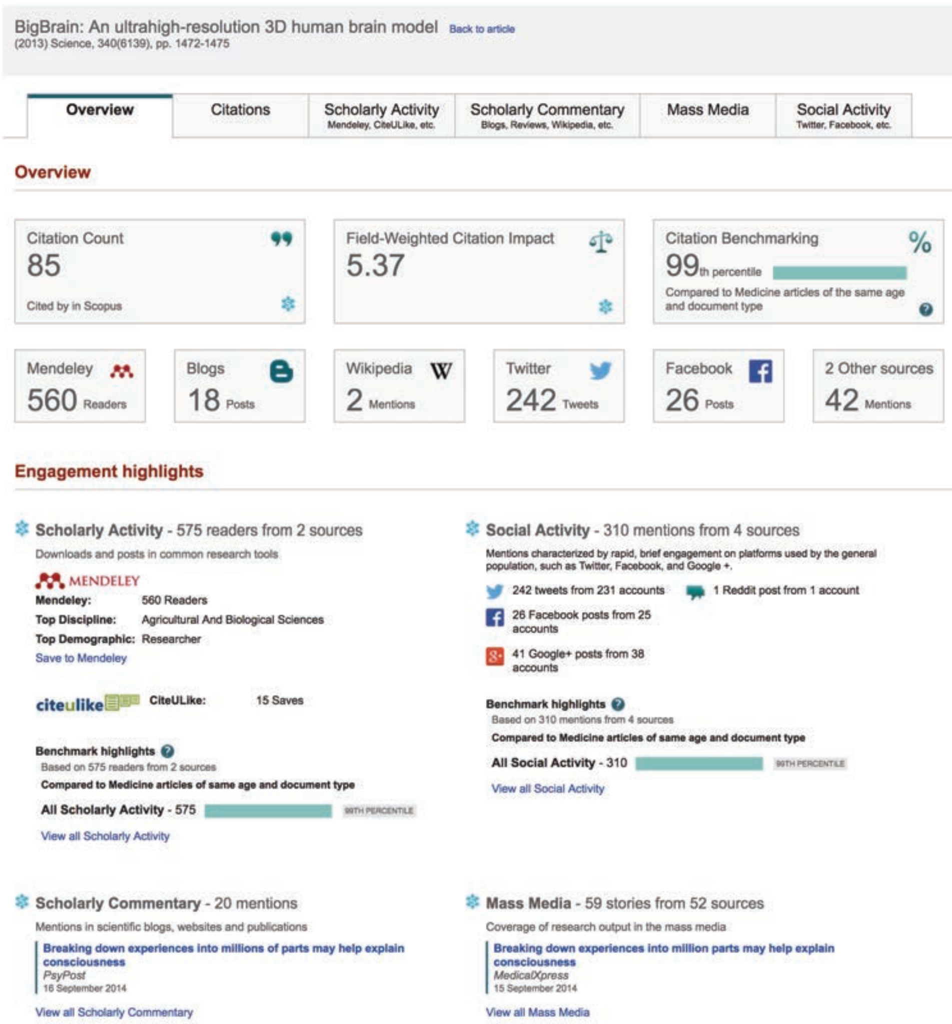


Figure 9: Bottom: Snowball metrics of a highly cited article from the D-CHAB, ETH Zurich in Scopus

Metrics integrated in researcher communities

Among the platforms that provide metrics is *ResearchGate* which, since March 2016, also presents the *h*-index, but has also a proprietary score, the RG score (Figure 10), which is based on how other researchers interact with your content.

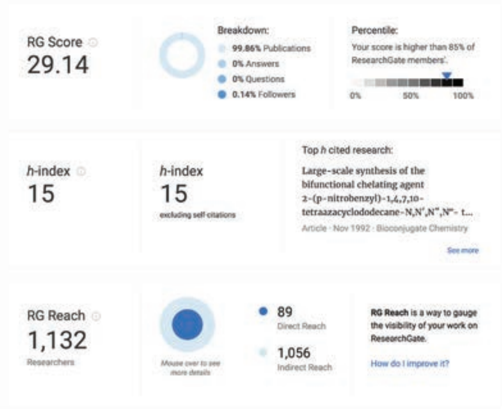


Figure 10: Author metrics in ResearchGate

Elsevier's Author Dashboard, where authors can view their metrics, was moved to Mendeley and is now available as *Mendeley Stats* (Figure 11).

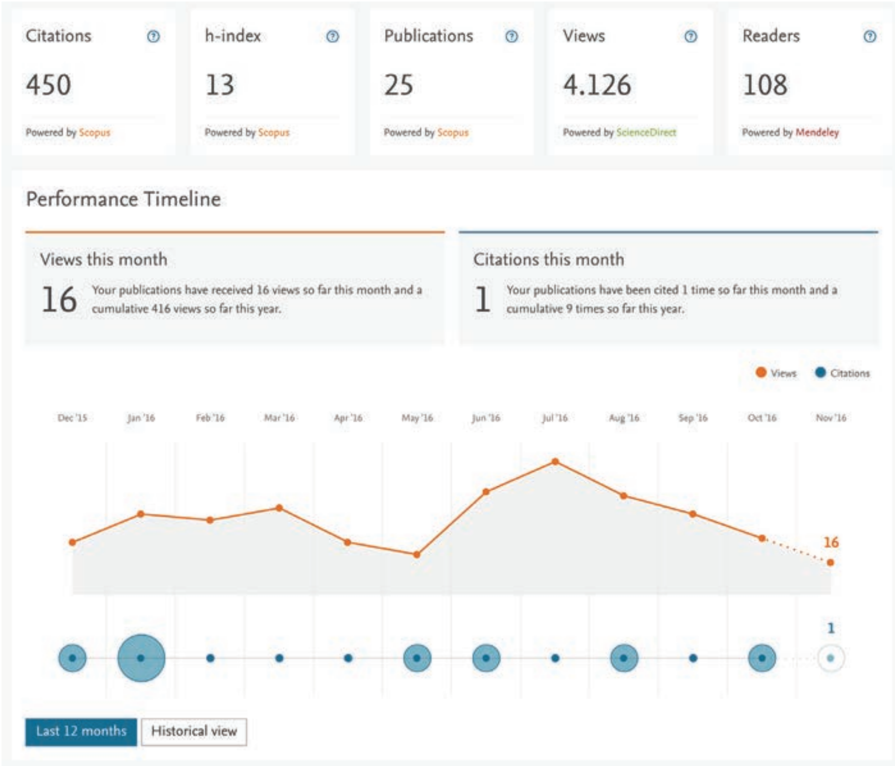


Figure 11: Author metrics in Mendeley

The future of research metrics

"*As data are increasingly used to govern science*", and this includes the new scores that are successors of the IF and the *h*-index, *Diana Hicks* and *Paul Wouters* published "*The Leiden Manifesto for research Metrics*" in *Nature* [17], with ten principles on how research should be evaluated. With the possible advent of Open Science and Open Innovation [18], new metrics will arise most likely once again, created by individuals who want to help researchers in selecting the right article or the right research information – in case journals will soon no longer exist as some believe.

References

[1] Gross PLK, Gross EM: College libraries and chemical education. *Science* **1927**, *66*, 385–389.
[2] Garfield E. Citation indexes for science: A new dimension in documentation through association of ideas. *Science* **1955**, *122*, 108–111.
[3] Garfield E: *Unfallchirurg* **1998**, *101*, 413–414, see also Eugene Garfield's website at http://garfield.library.upenn.edu for a translation of the article and search for „misuse" to find his statements addressing misuse
[4] Todeschini R, Baccini A: Handbook of Bibliometric Indicators: Quantitative Tools for Studying and Evaluating Research. Wiley-VCH, Weinheim, 2016.
[5] Reedijk J: Book Review. Handbook of Bibliometric Indicators. *Angew. Chemie. Int. Ed.* **2016**, *55*, 12144.
[6] For details, see www.journalmetrics.com
[7] See http://www.eigenfactor.org
[8] Bergstrom, CT, West JD, Wiseman MA: The Eigenfactor™ Metrics". *J Neurosci*. **2008**, 28, 11433–11434.
[9] Hirsch JE: An index to quantify an individual's scientific research output. *PNAS*, **2005**, *102*, 16569–16572. In fact, the astronomer Eddington proposed a similar criterion for cycling achievements much earlier, see https://en.wikipedia.org/wiki/Arthur_Eddington#E_number
[10] See https://addons.mozilla.org/en-US/firefox/addon/scholar-h-index-calculator
[11] J. Priem, D. Taraborelli, P. Groth, C. Neylon, Altmetrics: A manifesto, 26 October 2010. http://altmetrics.org/manifesto
[12] Published e.g. in Cagan R: The San Francisco declaration of research assessment. *Dis. Model. Mech.* **2013**, *6*, 869–870. DOI: 10.1242/dmm.012955
[13] See www.altmetric.com/products
[14] See https://impactstory.org
[15] See http://plumanalytics.com
[13] See www.snowballmetrics.com
[17] Hicks D, Wouters P, Waltman L, De Rijcke S, Rafols I: The Leiden Manifesto for research Metrics. *Nature*, **2015**, *520*, 429–431
[18] See e.g. European Commission, Open Innovation, Open Science, Open to the World. 2016. doi:10.2777/061652. Free-book download via http://bookshop.europa.eu

Citation: Renn O, Dolenc J, Schnabl J: A brief visual history of research metrics. *Infozine* **2016,** Special Issue 1, 3-8, DOI: 10.3929/ethz-a-010786351

Published: December 12, 2016

Philippe Hünenberger
ETH Zürich

Bibliometry: The wizard of O's

Imagine there is once a shortage of building space in Switzerland. This affects everyone, individuals and businesses, including restaurants. As the space must be reserved for the best places only, the authorities need an *objective* criterion to make an *optimal* choice, and an *obvious* definition for the quality of a restaurant is the frequency *F* at which the average citizen eats there. *Optimal*? *Objective*? *Obvious*? The three O's – I'll come back to that ... Anyway, *F*-factors are evaluated to five significant digits and analyzed using the most modern computer programs. Poorly performing places are closed and systematically replaced by better ones. Owing to this selection process, the cooks themselves end up considering the *F*-factor as the ultimate measure of professional success in the branch. And after a few years of this policy, one realizes with great surprise that the gastronomic landscape of the country has been reduced to cafeterias and fast foods. Indeed, the average citizen eats there more often than at four-star restaurants. Far-fetched? Well, I sometimes have the feeling that with the use and abuse of bibliometric impact factors to monitor academic research, we are letting the wizard of O's take us precisely down this path.

Optimal?

First "O" to question: In times of shortage, is *optimization* really the sole or even a good strategy and if yes, at which spatial (group) and temporal (planning) scales? Scientific research is a collective endeavor, and the best teams seldom consist of clones of an optimal individual. In addition, optimizing for short-term return is not the same as planning for sustainability and long-term effectiveness. Nature did very well 65 million years ago to have set aside some mammalians in case of, although they were definitely less performing than the dinosaurs at the time. In one word, diversifying is as important a strategy as optimizing. Thinkers, inspirers, nurturers and logisticians are as needed in scientific teams as pure individualistic communicators, recruited on the sole basis of their personal publication metrics. You do not make a winning soccer team with eleven top goal-scorers. And you do not make a successful "fellowship of the ring" with nine copies of Aragorn.

Another problem of optimization at all costs is that it is not compatible with risk-taking. As a rule of thumb, if you want a percentage *P* of true discoveries in research (and of orthogonal thinkers in scientific teams), you need to also accept a percentage *P* of unsuccessful efforts (and of poor scientists), the remaining being average incremental or fashionable research. By trying to optimize the percentage *P* towards zero in a no-risk strategy, one merely ensures that 100% of the research will be mediocre, while claiming very loud that it is top-level. Is this really what we want?

Objective?

Second "O" to question: Can scientific quality actually be measured by an *objective* criterion? As a scientist, I have the greatest respect for objective (reproducible) data. We are a theoretical-chemistry group, so we actually produce terabytes of it on a weekly basis. But this data alone does not make us any smarter. The real scientific talent is in the questions we formulate, in the design of clever experiments to address them, and in the analysis and interpretation of the results to formulate insightful answers. None of this is objectively measurable and, actually, none of this has anything to do with bibliometry whatsoever.

There is another interesting parallel between my work and bibliometric assessment. The interpretation of raw scientific results often relies on the reduction of very high-dimensionality problems (our terabytes of data) to one-dimensional indicators (a handful of functions shown in the figures of a scientific article). These projections must be selected carefully and are meant to facilitate an understanding of the process, given the limited capabilities of the human brain and language. However, both the selection and the interpretation of these indicators, two highly subjective processes, still require a deep knowledge of the mechanics and chemistry of the system. Treating these indicators as pure black-box outputs can be extremely misleading. The same holds for bibliometric indices. Although they represent some objective one-dimensional projection of the academic research process, their interpretation makes no sense and their use in decision-making is very dangerous for anyone who is blind to the underlying complexity. For this reason, it is essential that science managers keep in mind how scientific research works in practice, not on a flowchart but on the ground.

Let's return to the restaurant analogy. How would a gastronomic guide proceed to evaluate quality? They would send small teams of experts to taste the food, a procedure akin to peer reviewing. This procedure is tedious, time-consuming, expensive, demanding in terms of personal competence, and partly subjective. And (yes!) it does involve an emotional component. Yet, in many ways, it is far better to reason objectively based on subjective expert assessments, than to reason subjectively (without being aware of it) based on objective but irrelevant one-dimensional indicators. Ultimately, the probing instrument for a complex high-dimensional process must itself be complex and high-dimensional. When I hear an exciting scientific talk or I read a high-quality scientific article, I know it is good simply because I feel thrilled and inspired. And more often than the opposite, my colleagues feel just the same. But for some reason, although my computer can spit out a wealth of irrelevant bibliometric data about the author, it stubbornly refuses to share my enthusiasm.

Obvious?

Third "O" to question: Is there really an *obvious* relationship between scientific quality and bibliometric indicators? There are two aspects to this question: What do we consider to be quality in science and to which extent do bibliometric indicators characterize this quality? Already the first question is difficult, and there is a wide spectrum of opinions, from the most idealistic to the most utilitarian. Ultimately, we do science because it is in our genes of *Homo Sapiens Sapiens*: The urge to understand how the world functions and to apply this understanding for adjusting the world to our needs. So, maybe we can agree that scientific quality is related to the successful *acquisition* or *application* of new knowledge. This is already a two-dimensional space, *i.e.* beyond the realm of one-dimensional functions. And actually, bibliometric indicators do not even belong to this space as they exclusively focus on the *transmission* of knowledge, *i.e.* they are at best indirectly influenced by its acquisition and application. As a result, they probe scientific quality in a direct way neither from the idealistic nor from the utilitarian perspective.

Nowadays, the basic bibliometric currency unit (BCU) is one citation of one of your articles in an article of a peer scientist. And the basic assumption chain is something like: (1) your scientific quality is proportional to your number of quality papers; (2) the quality of a paper is proportional to the number of peer scientists who find it good; (3) the extent to which a peer scientist finds your paper good is proportional to the number of BCUs she/he gives you; (4) each time a peer scientist gives one or more BCUs to one of your papers, it means she/he has read it (I mean, past the title) and considered good; (5) all peer scientists have an equal probability to have seen any of your papers, before deciding whether they would give you a BCU or not. The first statement can arguably serve as a definition, with the already questionable corollary that a scientist who does not publish at all is automatically to be regarded as a bad scientist. For example, according to such a narrow definition and because he did not leave any writing of his own, Socrates would rank as an appalling philosopher. With the possible exception of this first one, no single statement in the list above is correct. Just check once in details where you (or a colleague) collect your BCUs. This is a sobering up experiment! And since there are many ways to generate BCUs artificially, I am wondering when it will become possible to buy them on the internet and what will be the resulting parity to the dollar. I will not detail the specific shortcomings of a given measure (*e.g.* *h*-index). All scientists who know how things work in practice can give you many examples of their shortcomings. My objection is not technical (how could we improve the index), it is fundamental: No numerical index whatsoever can measure scientific quality!

A self-reinforcing system

One of the main problems of the bibliometry-based evaluation system is that it is self-reinforcing. There is a well-known effect in sociology (self-categorization theory) called the social proof. In this particular instance, it states that if a certain *F*-factor, which may well be largely irrelevant, becomes the main criterion for accessing a given social elite (*e.g.* researcher position at a university, awardee of a prestigious grant), researchers, who are also humans after all, will spontaneously tend to first passively accept (compliance) and then actively believe (internalization) that the *F*-factor *is* the real measure of their fitness for this elite, *i.e.* of their true talent. As an older generation of scientists (those who also knew the pre-bibliometric times) gives way to a newer generation (those who obtained their positions thanks to their bibliometric fitness), compliance progressively gives way to internalization. More and more researchers show interest (and pride) for their bibliometric indices, compare their values to those of their peers, and work at boosting them as efficiently as possible. Questions about the goal of science and the true nature of scientific quality fade in the background, as they seem to be less immediately relevant.

To fight against the raise of the bibliometric dictatorship, I see a primary role for established scientists, those who no longer need to prove their quality and still have (some) freedom to comply or not with the current fashion. Comparatively, younger scientists are more on a tight leash, as it is made clear to them that bibliometric performance is the key to their academic future. But if one no longer finds critical thinkers in the universities, where will one find them? So, maybe we should all switch off our computers for a moment and take the time to think: what we do, why we do it, and whether it is good to keep doing it this way. The wizard of O's is no real wizard as everyone knows, merely an illusionist. And if we let him do his thoroughly absurd job till the end, we are going to be known to the future generations as the civilization of fast-food science.

Prof. Dr. Philippe Hünenberger
ETH Zürich
Laboratory of Physical Chemistry
HCI G233
CH-8093 Zurich
Switzerland

Phone +41-44-632 5503
phil@igc.phys.chem.ethz.ch
http://www.csms.ethz.ch

Citation: Hünenberger P: Bibliometry: The wizard of O's. *Infozine* **2016,** Special Issue 1, 9–10, DOI: 10.3929/ethz-a-010745184

Published: December 12, 2016

Matthias Tinzl
ETH Zürich

The grip of bibliometrics – A student perspective

With the growing numbers of researchers and the increasing pressure on these researchers to publish their findings, the volume of available scientific literature has reached incomprehensible levels. This is not only due to increased publishing in peer-reviewed journals, but also due to the advent of new publishing channels such as conference proceedings, open archives, e-papers and homepages. This vast amount of literature makes it impossible for an individual to read all publications, which automatically creates a demand for tools evaluating the quality of individual scientific contributions. Bibliometrics is an attempt to assess the quality of research articles based on measurable parameters such as the number of citations of an individual publication. When talking about bibliometrics, it is often forgotten that not only researchers, publishers and universities are affected by them, but also students. In the following it is highlighted how students are affected by bibiometrics, and whether this influence has a positive or negative impact on students.

Students are affected by research metrics even before they enter university, most of the time without even knowing it. University rankings such as the Times Higher Education Ranking (THE) or the Quacquarelli Symonds university ranking (QS) take into account parameters such as the total number of publications or the number of citations. Many students use the aforementioned rankings as an aid for choosing the university they want to study at, without considering which parameters are used to construct these rankings. This fact alone illustrates how powerful bibliometrics really is. Obviously a higher-ranking university will attract more students as well as more ambitious students, who are likely to produce more publications once they start their academic careers, which in turn improves the university's score in the rankings.

The downside to this is quite obvious. Universities which do not optimise their research output against measurable parameters will fall back in the rankings, even though the quality of teaching has not changed at these universities. Nevertheless, a worse position in a ranking might come hand in hand with a dropping number of applications and, possibly, decreased funding for research. To prevent getting caught in such a vicious circle, universities nowadays often try to optimise their bibliometric scores. However, it is not granted that research which ranks higher in terms of bibliometric indices such as impact factor or *h*-index is of higher quality. These parameters are strongly dependent on the field of research. For example, researchers who work in large fields generally have higher *h*-indices and publish in journals with higher impact factors compared to researchers who work in fields with a smaller research community. As a result, universities have an interest in funding researchers who produce many articles in an area of research with a large community, and tend to cancel funds for research with small communities. This is very problematic as it renders some research more important than other research, and could potentially – following the Matthew effect – lead to the complete eradication of some branches of research.

University rankings are not the only instance when students are affected by bibliometrics. Other examples are semesters abroad or the application for a PhD position. Usually (at least in the D-CHAB) students chose to do research projects during their semesters abroad, as the credits acquired for research projects are easier to transfer than those obtained for lectures. Therefore, one has to find a research group abroad. Obviously, this choice is mainly based on the compatibility of one's research interests with the research interests of the research group. However, if multiple options are available, one prefers to work in the group which has the highest quality of research. In order to determine which group's quality of research is higher, the easiest option is to look at the publication list of the groups, and search for the journals in which the most recent publications have appeared. Obviously, one assumes that the research is of higher quality if it was published in "high impact journals". Very often, this assumption is made without actually reading the publications, which should be an integral part of evaluating the quality of science. However, science which does get published in high-impact journals very often actually is good science. The peer-reviewing process ensures that only well proven facts are published, and editors should ensure that all articles are treated equally, *i.e.* that no advantages are granted to research done at more prestigious universities. With these mechanisms in place, the quality of research should be secured. Bibliometrics is therefore a somewhat valid tool for measuring the quality of research, but should not be trusted blindly. One limitation, for example, is the age of a professor. Very often, it is more difficult for a professor to get his articles published in journals with a high impact factor, because he has not previously published in journals with a high impact factor.

Altogether, students are greatly influenced by bibliometrics, especially in terms of their career choices. Despite all the flaws of bibliometrics, which need not be discussed in great depth at this point, bibliometrics is necessary for students to evaluate the quality of research carried out at other universities without having to spend hours reading papers. Completely rejecting bibliometrics would be unintelligent, as using bibliometrics to access the quality of the research is a trend that has come to stay. From a student point of view, it would be interesting to refine methods which produce bibliometrics in such a way that parameters like the *h*-index and the impact factor become less dependent on the field of research. Such an adjustment could be made by calculating different impact factors for different research disciplines, and explicitly stating the number of papers which were published by a scientist. As a last personal comment, I think that students should not primarily base their decisions on university rankings, impact factors and other similar indices as they have evident flaws. Personal judgement should be more important than bibliometrics in individual life-changing decisions.

Matthias Tinzl,
currently Master student Interdisciplinary Sciences, BSc Interdisciplinary Sciences at ETH Zürich
Rümlangstrasse 52
CH-8052 Zürich
Switzerland
mtinzl@student.ethz.ch

Citation: Tinzl M: Bibliometry: The grip of bibliometrics – A student perspective. *Infozine* **2016,** Special Issue 1, 11–12,
DOI: 10.3929/ethz-a-010745075

Published: December 12, 2016

Wendelin J. Stark
ETH Zürich

Honesty and transparency to taxpayers is the long-term fundament for stable university funding

The Swiss Society has been trusting the ETH domain for over 150 years: More than 2 billion CHF are annually given to research, engineering and technology transfer. This enormous sum of tax-derived money is spent by about 1000 faculty members leading teams at ETH, EPFL, PSI, EMPA and EAWAG, providing a globally unique funding situation.

On the more formal side, the Swiss Society has formulated the ETH Law (Art. 2.1.) specifying our job as teaching, research and technology transfer. Art 2.2 is very clear: "we must consider the needs of the country". Given the amount of money spent, the public is astonishingly patient in letting us do long-term investments of significant costs. This deep trust is based on ETH Zurich's past achievements and its standing as a "top university" – an unclear term, shaped through the ETH Zurich domain's view in the public.

We should therefore ask ourselves what determines this image? Most people will not understand the details of a faculty's work. They will use substitute parameters to evaluate the output of the ETH Zurich: Is this "average" or "outstanding" work? Are these people doing something useful with our investment?

Prominent researchers publicly condemn quantitative measurements of success and in particular university rankings, bibliographic parameters and outreach activity (e.g. altmetric or media coverage). Is this justified?

When evaluating the performance of a faculty, universities often rely on external evaluations. They are commonly performed by a committee delivering a report. A critically thinking layman, a politician or a tax-paying citizen may think about what is more honest or transparent:
(1) The application of well-defined, countable, openly-available parameters derived from publicly-available source data (e.g. scientific papers and their citations), or
(2) a collection of personal opinions gathered by a small group of invited experts from other universities?

The second procedure appears particularly questionable when the long-term relationship between faculty members at different universities is considered: The same groups of faculty share expert views in different committees, assist review panels, share or assign awards and grants, or organize keynotes at conferences. There are complex interwoven interests.

However, bibliometrics and other countable data also have severe flaws: Disturbing single-case events (e.g. highly-cited but retracted papers) can be identified in all areas of science. The opponents of objective measurement are quick in citing such flaws. What could we do instead?

Research ethics obliges us to use the best methods available to judge on a problem, i.e. the most objective methods. In addition, we should critically assess our choice of methods. Naturally, comparison depends on the focus of a research team. In all cases, however, a comparison is made with respect to others, considered equal or better. In our case, ETH Zurich should compare itself with world leading universities. At the level of individual researchers, we should also compare ourselves with leading scientists or engineers at leading institutions.

A fair measurement compares with a fair metric

Fundamental science. If a thought of an intelligent person is not published, it is lost. If a thought triggers generations of researchers to develop new thoughts, it has some impact. If two researchers at two good universities work on similar topics, scientific essays ("papers") become comparable, at least qualitatively.
Applied Sciences and Engineering. Solutions, processes and materials can similarly be irrelevant, if never used, and of no interest to others in the field. If solutions/thoughts are used, and create new research fields or products for companies, they have impact.
Societal implications. If the interactions of a person with society (through any media/means) are inexistent, there is no impact. If these interactions lead to new thoughts, changes in behavior, improvements etc., this scientist has some effect or impact.

Opponents to measurement may cite prominent cases where faculty members of different fields were compared in an unfair way: An artist cannot be compared to a chemist using papers in leading journals as a metric. Such cases, however, are no argument against an adequate effort to use fair parameters. The three following proposals may be used to illustrate such a procedure.
Evaluating a traditional scientific research group. Researcher *A* at ETH

Zurich works on the metabolites of maritime sponges, using chemical and pharmaceutical methods. She publishes in the leading chemical journals, and occasionally in a multidisciplinary journal. Her natural peers at MIT, Stanford, U. Cambridge, Harvard etc. publish in similar journals. What does this mean? If two groups make 10 or 15 papers in the leading journals, annually, they are at a comparable level. Another team contributes only 2–3 similar papers per year – that is clearly less productive. Second, what happens with that work? One organic chemistry group is cited 1000 times per year – clearly a leading position if compared to other similar groups. Another team is only cited 100 times – it is clearly noted less. Third, research financing is most relevant and clearly measurable (e.g. ERC grant vs. SNF grant).

Evaluating an engineering group. Researcher *B* works on new chemical processes, publishes papers and patents, and works with the chemical industry. His peers are at a number of leading universities (TU Delft, MIT, Stanford University, University of Minnesota, Harvard University, etc.). The output is at least in three areas, which can be counted separately: Papers in leading journals, patents and industry projects (amount of money; patents; licenses from the university's technology). Again, we can compare such an output: One group runs projects with large companies (e.g. 500 kCHF contract). That is clearly different from another researcher getting 20 kCHF for a sample analysis. A patent that is licensed and the basis for a 250 Mio CHF cash flow in a Swiss company is more valuable than a non-licensed patent that a university tries to commercialize for 5 years before dropping it. The output of an engineer in terms of papers can be compared at least partially with that of fundamental scientist A (above), since now both compete in the same category, using similar tools (papers, i.e. essays that might be cited and used by others). Ideally, one should look at the corresponding standing of these two persons in their field, e.g. top 1% of the field (excellent) vs. last third (not so good) instead of absolute numbers of citations only.

Evaluating a traditional mechanical engineer. Here, publications are less important and projects are measured against their use in industry (size of a process, cash flow, number of sites running that project etc.), commercial value (patent income, patent citations, licenses), or their impact in the fields (key conferences with formal abstracts, project presentations, contests won, etc.). The commercial fate of a process or product is important (e.g. over a spin-off company). Industry projects, measureable at least in their size, can be compared from one team to the other. The number and future jobs of educated students can also be compared.

Teaching. The success of a university faculty as a teacher, beyond course grade, is ultimately linked to the fate of her/his students: Do they find adequate jobs? Are they working in their field of education? Was their education of any use to their job? What do they earn? How long does it take to find a job? Following up on students a few years later is routinely done in American institutions, and may be significantly intensified at Swiss institutions.

Balancing individual contributions? Most faculties will contribute in several output metrics. How can such different parameters be combined, at least in a given discipline? The ETH tenure committee faces similar challenges, and ETH has developed a number of items where it evaluates an aspiring full faculty. These tenure criteria are a good starting point for faculty evaluation during department internal discussions, and during university-wide evaluations. However, they currently lack the "third part" of our job description (technology transfer), and are too heavily biased towards natural sciences.

The alternative is not to measure. A less objective judgement of performance is prone to personal bias. Bibliometrics and other measurements represent the action of thousands of actors and judges. As such, the outcome is at least less biased compared to the opinion of a small group of experts that partially know one another. If we do not measure, the best personally-connected faculty will win at evaluations done by small groups of evaluators. Opponents to measurement may add: "In a liberal, free society, everybody is free to think what she/he wants." Yes, but, most faculty members are employees of a university. Employment comes with a job description (see ETH law), hence, most researchers are not entirely free to do whatever they want. In the case of ETH Zurich, there is a well-defined purpose involved.

The above discussion becomes even more complex if one considers the following example: Researcher *C* uses 1 Mio CHF per year to hunt for a rare physicochemical effect, his papers are barely read and he does not take part in public outreach. At the same time, dozens of researchers at low-income countries try to improve treatments against diarrhea, using a combined budget of 1 Mio CHF. Here, the question of fair and morally acceptable use of finances becomes an unpleasant topic.

It appears difficult to explain to a layman why scientists should be rigorous with their object of study, but deny the same rigorous approach to their own performance. If we would live in a world of endless resources, this unpleasant discussion would not be needed. Being accountable and transparent is the basis for trust and (hopefully) a continued and generous financing of university research through the Public.

Recommended Reading

Whitesides G: Chemie neu erfinden. *Angew. Chem.* **2015**, *127*, 3238–3253, doi:10.1002/ange.201410884

Prof. Dr. Wendelin J. Stark
ETH Zürich
Institute of Chemical and Bioengineering
Vladimir-Prelog-Weg 1-5/10, HCI E 107
CH 8093 Zurich
Switzerland

wstark@ethz.ch
http://www.fml.ethz.ch

Citation: Stark W: Honesty and transparency to taxpayers is the long-term fundament for stable university funding. *Infozine* **2016,** Special Issue 1, 13–14, DOI: 10.3929/ethz-a-010745069

Published: December 12, 2016

Charlie Rapple
Co-founder, Kudos

Beyond metrics: Managing the performance of your work

Most researchers I know have a love–hate relationship with metrics. People hate the idea of their research performance being simplistically summarized with numeric proxies for quality. But they cannot resist checking out these numbers, and sharing them with other people if they seem to be doing well ("my *h*-index is bigger than your *h*-index" may be a joke, but we laugh because we recognize the reality it mocks!)

We are in a period of expansion, as far as metrics are concerned. Ten years ago, we counted citations and downloads, mostly at the journal level. Around five years ago, the focus began to shift to article-level metrics – increased tracking of citations and downloads for individual articles (and their authors), and the emergence of alternative metrics, "altmetrics", which attempt to track mentions of work in social media, traditional media, Wikipedia, government policy, clinical guides, and many more non-academic sources.

These new metrics are quickly becoming mainstream, with institutions, publishers and funders all working with providers such as Altmetric (known for its colourful "donuts" which signify the sources of attention for a work) and Plum Analytics (which offers similarly stylish "plum prints"). Organizations vary in their application of this new data but there is a clear trend – with movements such as DORA [1] – to move away from over-reliance on citations, and to broaden the range of metrics that are used to evaluate research and researchers.

This creates a challenge for researchers, who need to become familiar with a wider range of metrics – understanding how they are produced, by whom, based on what data. It's clear that while people want to keep on top of these changes, they have limited time for developing expertise in an area that is, after all, peripheral to their own research – however important it may be in terms of how that research is perceived or evaluated.

This is the challenge that my co-founders and I set out to solve when we started Kudos. We set ourselves the task of bringing together a range of metrics in one place, so that researchers wouldn't have to learn for themselves about different kinds of metrics, or take time visiting different sites to understand the performance of their work – by using Kudos, they would be able to see downloads, citations and altmetrics in one place.

Viewing metrics is not enough

But we also realized that viewing metrics is not enough. People want to take more control of the visibility and impact of their work and not just leave its performance to chance. Every researcher I have ever spoken to feels burdened by the challenge of information overload – the struggle to keep up with the explosion of new research, and the resulting deluge of new publications. While people want to think that their work will stand on its own merits, in reality, the likelihood of it being found, read, applied and cited is diminished by the flood of other papers in the field. The increasing use of metrics is itself a response to the sheer volume of research being undertaken and published now, in terms of institutions' ability to evaluate all their research outputs, and the need for proxy measures to simplify this process.

This brings us back to the love–hate point, with researchers frustrated at the growing use of metrics – but grudgingly accepting that if your work is going to be measured in this way, you are sabotaging yourself if you don't take control of the performance of your work. There are many ways to do this, of course – from traditional approaches such as presenting work at conferences or sharing it with colleagues via email, to new options such as presenting work in academic networks or sharing it via social media. Again, my Kudos co-founders and I recognized a challenge here that we hoped to solve: with so many different ways to communicate your work, how do you know which is most effective when it comes to maximizing readership and impact?

Therefore, in addition to gathering a range of metrics together in one place, we also worked to provide researchers with a mechanism for tracking their communications so that the effect of outreach via e-mail, social media or academic networks could be easily mapped to "meaningful" metrics such as downloads, citations and altmetrics. By centrally managing your communications, across all your publishers and all the different places you might share them, Kudos enables you to see which efforts are correlated to improved metrics. From this, you can make more informed decisions about how and where to use your limited time for communicating.

The Kudos system has now been live for just over 2 years and over 100,000 researchers have signed up to use it. The Department of Chemistry and Applied Science (D-CHAB) at

ETH Zürich has been an early adopter of our institutional service, which provides staff with a view onto researchers' communications in order that they can better support and amplify these to further increase visibility and impact. A recent study has shown that usage of the Kudos toolkit is correlated to 23% higher downloads. It is a free service that you can try by signing up at www.growkudos.com/go/ethz.

In conclusion, not everyone is convinced that metrics are a useful way to evaluate research but it is clear that, one way or another, they are here to stay. Many researchers have embraced options for communicating their work and are seeing improved performance against metrics as a result – whether intentionally or as a by-product of their efforts. In an age of information overload, many argue that you are doing your work a disservice if you don't make efforts to ensure that it is found and applied by a broad audience. Ultimately, increasing attention, readership and citations to your work is a worthwhile outcome in and of itself, because of the increased opportunity that your work will be built upon by others, regardless of whether you improve the metrics in the process. So even if you don't care too much about metrics, I encourage you take action and give your work the best chance of finding its audience. And I hope you will experiment with Kudos in the process so you can communicate as efficiently as possible, and keep your focus on your research!

References

[1] DORA – San Francisco Declaration on Research Assessment, see http://www.ascb.org/dora for more information

Charlie Rapple
Sales & Marketing Director • Co-Founder
Kudos Innovations Ltd
2A Ashurst Court
London Road, Wheatley,
Oxford OX33 1ER
Great Britain

Phone: +44 1865 872527
charlie@growkudos.com
www.growkudos.com

Citation: Rapple C: Beyond Metrics: Managing the performance of your work. *Infozine* **2016,** Special Issue 1, 15–16, DOI: 10.3929/ethz-a-010745060

Published: December 12, 2016

Rafael Ball
ETH Zürich, ETH-Bibliothek

Scientific profiling instead of bibliometrics: Key performance indicators of the future

> *"Everyone is graded. Lovers by lovers under a veil of silence; traders by vociferous customer complaints; the media by quotas; doctors by patient flows; the elected by voter reactions."*
> Michel Serres [1]

A slice of history

Bibliometrics originally developed from the notion of supporting librarians in their task of selecting optimum literature and optimising holdings management. Not only was this the basic idea of the first bibliometric analyses, it was also the approach adopted by *Eugene Garfield*, the American chemist and founder of the first bibliometric index, the *Science Citation Index* (SCI), in the 1950s.

Cole and *Eales* gave us the first bibliometric analysis. In 1917 the authors studied which books on human anatomy had been published between 1550 and 1860 [2]. As this analysis purely measured the output on a particular topic, however, it was not yet a citation analysis.

The first bibliometric analysis to study citations was conducted by *Gross* and *Gross* in 1927 [3]. The authors analysed citations made in footnotes in the field of chemistry, which enabled them to compile a ranking of the key chemical journals of the time based on how frequently they were cited. On the one hand, the chemistry community used this information to assess the important publication organs, which is in keeping with the fundamental concept of journal rankings and the impact factor that is so important today. On the other hand, *Gross* and *Gross* were librarians and intended to help libraries in the procurement of journals with their study. In their analysis, they detected an irregular distribution of citations among the various journals and thus provided the basis for *Bradford's law*, which was developed in 1934 and according to which key scientific publications are concentrated on a handful of core journals.

Again, these analyses pursued the sole purpose of obtaining information on science and its processes rather than compiling quantitative rankings, for instance. Russian science philosopher *Gennady Dobrov* defined this kind of research in his book *Nazka o Nauke* ("The Science of Science") in 1966 [4].

Nothing changed in this bibliometrics objective until after the Second World War. It was not until the 1950s that the aforementioned *Eugene Garfield* systemised the quantitative measurement of scientific output by founding his Institute of Scientific Information (ISI), thereby paving the way for today's citation indexes.

This was the beginning of the age of the classic indicator canon in bibliometrics. The original aim of supporting libraries in managing their holdings was soon forgotten and the Science Citation Index initially developed into a research tool for content-based literature searches, then an instrument for the quantitative measurement of scientific output. This process took many decades. Thanks to the *Science Citation Index*, what *DeSolla Price* explained in his book *Little Science, Big Science* was now possible, namely to apply the tools of empirical science to the sciences themselves. "Why not apply the tools of empirical science to science itself? Why not measure, compile broad hypotheses and draw conclusions? [5]

Bibliometrics and the advent of performance-oriented funding

At first, politics had little interest in using the quantitative results on bibliometric analyses to assess performance or even allocate funding in science and research. However, this eventually changed in the wake of the so-called "Sputnik crisis", which revealed virtually overnight that the USSR had beaten the industrial nations of the West in the race into space based on scientific results. Politics began to become interested in managing the supply of scientific information and also exploited the quantitative results of bibliometric analyses so that citation-based procedures especially established themselves as the dominant instrument for performance assessment and research evaluation in the exact sciences in the 1980s.

In the process, the use of these indicators developed right down to micro-level for the assessment of individual scientists. Today, the majority of bibliometrists oppose this use on individual people due to the resulting inaccuracies at this aggregation level.

In the classic indicator canon of bibliometrics, which was valid for several decades, the measurement of output (number of scientific results) and its perception (essentially the number of citations ascertained) are at the forefront. These two parameters can then be used to produce rankings which provide a comparison between people, institutions or countries. Moreover, thematic focuses can be generated with the aid of bibliometric citations analyses.

Nonetheless, due to the indirectness of the assessment, these indicators only allow an approximation of the actual performance. Nor does the perception of a publication measured via the number of citations permit a direct conclusion regarding the quality of the scientific results. The actual problem of classic indicators lies in this extremely indirect approximation of the quality of scientific results.

However, this method is well established in the exact sciences and recognised in the scientific community. After all, due to the mass emergence of scientific publications in the last thirty years, decision-makers bank on quantitative support in science management. Nobody can rely on qualitative parameters of a person-based review these days.

This was also more than adequate in the expert communities and barely called into question in the inner circles of the respective disciplines.

The advent of the internet and the future of key performance indicators

The question of the significance of the indirectness of measuring performance indicators was only cast in a new light with the advent of the internet and the mass availability of digital data. At least four conditions have changed somewhat radically:

1. The mass availability of digital data on the internet enables many quantitative parameters to be evaluated automatically and provided in the form of pattern recognition.
2. The internet has created new public spheres that receive scientific results. Not only does a discipline's inner circle perceive the publications for longer, but also in different aggregation forms, and broad sections of the public can participate in the results from science and research via digital media. This widens the definition of the perception and the significance of the scientific publications and their authors.
3. In the internet age, scientific findings can be made available extremely swiftly and indirectly. The classic route of publishing in (printed) journals and books is supplemented with or substituted by the different paths in electronic publishing.
4. New communities are also emerging for scientists on a vast range of levels, which are all served and meet and perceive the findings with a varying depth and breadth.

For classic bibliometrics and its indirect indicators, usage statistics (metrics) that gauge the direct use of scientific results in the form of downloads and so-called alternative metrics (altmetrics), which indicates and renders accessible the perception of scientific results and those of the authors, such as via social media in the form of links, storage and recommendations, are now combined. The topic of indirectness (classic bibliometric indicators) is therefore not just nullified; it is also supplemented with direct indicators and might be replaced entirely with the direct visibility of the perception of the perception and use of scientific publications in future.

Moreover, the data source and media form of the scientific publications evaluated have changed considerably: for altmetric or usage measurements, not only do results become important in the written form, but also all forms of scientific "expression": research data, source texts, source codes, presentations, conferences, self-publications, weblogs, blog entries etc.

For the alternative measurement of scientific output, there are four distinctive forms of use:

1. "Viewed": activities that gauge the access to scientific articles.
2. "Saved": the uploading of an article onto a bibliographical programme, for instance.
3. "Discussed": a used article discussed via a wide variety of social media channels and supplemented by others.
4. "Recommended": exclusively an activity that recommends a paper for re-use.

This classification of usage results uses different altmetric systems and products, e.g. "article-level metrics [6] by the Public Library of Science (PLOS) or "Impactstory" [7] and more.

The development of bibliometrics clearly reveals that the variety and breadth of the indicators have increased over the decades and that completely new parameters have emerged in the wake of the variety and diversity of the media, which enable the performance, significance and quality of scientific results and their authors to be gauged.

In future, scientists and institutions will be given a whole series of scores, which not only yield a more complete picture of the scientific performance, but also the perception, behaviour, demeanour, appearance and (subjective) credibility. Whether we find this a good thing or not, it is in keeping with the kind and possibilities of evaluation in the digital web age of the twenty-first century.

The next development reveals a tendency towards comprehensive data acquisition and its evaluation. Under the umbrella term "analytics", it is possible to collect and analyse increasingly large and diverse amounts of data on the web. With big data, new nexuses are being uncovered that nobody had even conceived or called for before.

"As a consequence, an increasing amount of data on every single one of us is available – including from areas of our private lives. The image of the transparent customer and transparent citizen is certainly no longer a vision of the future; it has become a reality [8].

And the image of the transparent scientist, too.

A score like the one that has long existed for the evaluation of scientific efficiency, especially in allocating credits, can then be transferred to science.

The new *h*-index, which is supposed to determine the significance of a scientist's publications as a simple indicator, is obsolete and can be replaced by a digital "scientist score": a value that considers and combines a scientist's complete data available online. This kind of profiling is another trend, to which bibliometrics will greatly add. If vast amounts of (per sonal and institutional) information on scientists, which can be compiled and evaluated via a search algorithm, is available, before long this data will yield indications as to the output and performance of these individuals.

A series of analytical tools already exist on the market, such as *PLUM Analytics* [9], *Figshare* [10], *InCites* [11], or *SciVal* [12], which adopt an integrated management approach and offer performance, financial, personal and publication data for decision-makers in science and research.

Data from classical bibliometrics will then only be a small part of a comprehensive data evaluation of people and institutions.

References

[1] Serres M: Petite Poucette. Berlin, Suhrkamp 2013, p. 49
[2] Cole, FJ, Eales NB: The history of comparative anatomy. Part I: A statistical analysis of the literature. *Science Progress* **1917**, *11*, 578-596.
[3] Gross PLK, Gross EM: College libraries and chemical education. *Science* **1927**, *66*, 385–389.
[4] Dobrov G: Nauka o Nauke: Vvedenie v obshchee naukozanie. Kiev 1966
[5] Price, Derek de Solla: Little Science, Big Science. Suhrkamp, Frankfurt am Main,1974, p. 9
[6] URL retrieved October 22, 2016 http://article-level-metrics.plos.org/
[7] URL retrieved October 22, 2016 https://impactstory.org/
[8] Bachmann R, Kemper G, Gerzer T: Big data – Fluch oder Segen? Unternehmen im Spiegel gesellschaftlichen Wandels. mitp, Heidelberg, 2014, p. 20
[9] URL retrieved October 22, 2016 http://www.plumanalytics.com/press.html
[10] URL retrieved October 22, 2016 http://www.swets.com/figshare-for-institutions
[11] URL retrieved October 22, 2016 http://researchanalytics.thomsonreuters.com/incites/
[12] URL retrieved October 22, 2016 http://www.elsevier.com/online-tools/research-intelligence/products-and-services/scival

Dr. Rafael Ball
ETH Zurich
ETH Library
Director
Rämistrasse 101
CH-8092 Zürich
Switzerland
Phone: +41 44 632 21 25

rafael.ball@library.ethz.ch
www.library.ethz.ch

Citation: Ball R: Scientific profiling instead of bibliometrics: Key performance indicators of the future. *Infozine* **2016,** Special Issue 1, 17–19, DOI: 10.3929/ethz-a-010744972

Published: December 12, 2016

Carl Philipp Rosenau
ETH Zürich

More knowledge, less numbers

"*Do you know your* h-*index*?" was one of the introductory questions by Oliver Renn, Head of the ETH Chemistry | Biology | Pharmacy Information Center in his fall semester course entitled "Scientific information retrieval and management in chemistry and life sciences". I did not know mine even though I was already familiar with the famous Hirsch index (*h*-index), which was introduced in 2005. This index is one of the modern tools to enable an evaluation of the total number of publications versus the number of citations received by each publication. To give an example, if I have five publications and all are cited at least five times then I have an *h*-index of five. But is it true that a high *h*-index correlates with the relevance or importance of your research?

The need to measure scientific output and its relevance to a field may not be new. Certainly the online availability of scientific contributions opened new possibilities to have search engines and scripts going through databases to easily come up with a number like the *h*-index. But what is the effect and the message of those developments for doctoral students like myself? Well, the message is quite clear: Choose a field which is en vogue, and publish as much as you can in high-impact factor journals to increase the chances of getting cited. One could argue that this is not new and that the pressure to publish in prestigious journals has always been there – which is true. What clearly changed from 30 years ago is the instantaneous visibility and evaluation of scientific output fueled by the fast paced publishing industry.

I remember stories from my dad, who also is a chemist, and how he regularly spent time in the library during his PhD whenever new issues of certain journals were available. This happened on weekly, biweekly or monthly basis. Times have certainly changed! I receive email alerts by SciFinder whenever a publication appears that fits my search criteria. Furthermore, I get daily updates by e-mail of "just-accepted" publications in journals that I enjoy reading or have to read for my own research. Consequently, this results in dozens of daily e-mails illustrating nicely the change in the publishing industry. It is therefore not very surprising that there was and is a need for new ways to evaluate and rank the scientific output.

What is to conclude from the current state we are in? It seems more and more necessary to have bibliometric tools in order to sort and evaluate a scientist's work. From the pre-selection of candidates during academic and industrial job interviews, one hears stories of how the sheer numbers are more and more important. Can we decide, by looking at the number of publications, impact factors and *h*-indices who is the better scientist and whom to hire? To me the question is clearly to be answered with "No". Just looking at numbers may give you an impression of how productive (with respect to publishing) a researcher has been, but won't tell anything about an individual's actual skills. And, neither spoken language, social and soft skills nor hard skills can be extracted with bibliometric tools.

I do not want to come across as overly negative but to me the (amazing) new tools that are available to us are often misused or misinterpreted. To illustrate what I mean, take for instance statistics, a tool used in numerous areas of research and society. It is a powerful instrument, which can equally be misinterpreted. Famously, a study published in 2012 in the *New England Journal of Medicine* looked at the chocolate consumption per inhabitant and year versus number of the Nobel prize winners of a country [1]. There is a clear correlation between chocolate and the Nobel Prize (Switzerland being at the top of this chart). No one with some intelligence would conclude that starting to eat chocolate or move to a country with a high chocolate consumption will increase your chances for a Nobel prize. It would be a misinterpretation of correlation and causality. The interpretation of the numbers produced by bibliometric tools sometimes mirror this flaw.

As Plato stated "*a good decision is based on knowledge and not on numbers*". It is up to all of us to stay open-minded about new bibliometric tools, but be critical at the same time about their meaning and usage. In my opinion, one of the core purposes of science is research and the communication thereof for the betterment of society, not the number of publications you have, the impact-factors of the journals you published in, or your *h*-index.

References

[1] Messerli FH: Chocolate Consumption, Cognitive Function, and Nobel Laureates. *N. Engl. J. Med.* **2012**, *367*, 1562–1564. DOI: 10.1056/NEJMon1211064

Carl Philipp Rosenau
ETH Zürich
2nd year doctoral student, Laboratory of Inorganic Chemistry
Department of Chemistry and Applied Biosciences
CH-8093 Zürich
Switzerland
Phone +41 44 632 44 60
rosenau@inorg.chem.ethz.ch
http://www.tognigroup.ethz.ch

Citation: Rosenau CP: More knowledge, less numbers. *Infozine* **2016,** Special Issue 1, 20, DOI: 10.3929/ethz-a-010745064

Published: December 12, 2016

Rüdiger Mutz
ETH Zürich

Do we really need BIBLIO-metrics to evaluate individual researchers?

Frequently, metrics are applied as quantitative measurements in order to evaluate research performance, particularly in the assessment of individual researchers with regard to appointment procedures for professorships, research fellowships, and tenure tracks. Thereby, the primary focus are bibliometric indicators, measurements that are derived from the quantitative analysis of documents, such as those included in bibliographic databases (e.g. *Web of Science* or *Scopus*). Yet, in my opinion, it is neglected that the classical approach of research evaluation, i.e. peer review, is often based on quantitative measurements too (e.g. rating scales). In light of the strong criticism of the peer review procedure, especially of its low reliability in terms of a lack of agreement between the reviewers' ratings, possible alternatives are discussed intensely. Under consideration are, for instance, the post-publication peer review procedure, in which peers evaluate an article after its publication, and, as mentioned, bibliometric indicators.

Benefits and problems of bibliometric data

I think the reason why bibliometric indicators are preferred is mainly due of three reasons: First, bibliometric indicators are considered to be objective, as they are devoid of subjective evaluations and made available by independent database providers (e.g. *Elsevier*, *Thomson Reuters*). Second, bibliometric data is transparent and not anonymous. Both the publications of researchers as well as their citing references, which ultimately are countable citations, can be identified and verified in bibliographic databases. Thus, in principle, results of bibliometric analyses can be replicated. Third, bibliometric indicators allow an immediate and concrete interpretation of research performance. The information that a researcher has published 10 publications and 5 of those publications have been cited over 100 times will probably be more and immediately indicative of the research performance of said scientist than for example a rating scale point of 3.4 on a 5-point rating scale, which is the result of the average of several referees' ratings of the quality of research.

However, bibliometric databases should not be mistaken to be perfect. Indeed, not all scientific journals are covered in such databases, the documents may include bibliographic errors, there are name ambiguities, the affiliations of institutions are often incorrect, and any database update can lead to changes within the whole bibliographic database (for example by adding or removing journals). Furthermore, *Goodhart's law* is valid for bibliometric indicators as well: An indicator that is used for evaluation will itself become a target of a optimization and is, therefore, no longer a suitable measurement for evaluation. For instance, it is possible to increase citations of a publication if authors frequently cite their own work (problem of self-citations). Self-citations may serve as an example: Author(s) can increase the number of their citations by frequently citing themselves.

h-index and percentiles

A classic example of a bibliometric indicator that can be used for the assessment of individual researchers is the *h*-index, in the way it is for example implemented on Scopus by default. "A scientist has index h if h of his or her N_p papers have at least h citations each and the other $(N_p - h)$ papers have $\leq$ h citations each." ([1], p. 16569). On first sight the *h*-index appears to be very appealing; it is easy to calculate and combines quantity (number of publications) with quality (citation impact) [2]. However, the *h*-index is not without fault (e.g. [3], p. 78). For instance, it is field-dependent. Because of higher citation levels in the field, a researcher in the life sciences can be expected to have a higher *h*-index than a scientist in social sciences. Older researchers have the advantage that they were able to publish more than younger researchers. The *h*-index can only increase and can be influenced through self-citations. What it is that constitutes a high *h*-index is unclear, as there is no benchmark for comparison. This critical remark does not mean, however, that I question the application of bibliometric indicators in general, but rather that I suggest a critical usage of such indicators. In that way, the *h*-index could be complemented through the employment of percentile ranks [4, 5]. A percentile rank quotes the percentage of articles in a scientific field that are at or below the citation score of a given article. If, for example, a journal article ranks within the top 10%, it belongs to the 10% of the highest cited articles in the scientific field in which it has been published.

Conclusions

In my view, the initial question indicated in the title, whether bibliometrics should indeed be used for the evaluation of researchers, can be answered with "yes". Bibliometric indicators should be applied as a supplement to, not as a substitute for, peer review ("informed peer review"). In my view, following *Bornmann* and *Marx* [4], *Wouters*, *Glänzel*, *Gläser*, and *Rafols* [6], the *Leiden Manifesto* [7], and the *San Francisco Declaration on Research Assessment* (DORA) [8], the following factors have to be taken into account with regard to the usage of bibliometric indicators:

1. Journal-based metrics, such as the journal impact factor, should no be used as a measure of the quality of individual research articles ([8] p. 869).
2. Research assessment should consider more than one bibliometric indicator with regard to the target of the evaluation. In particular, basic measures (number of publications, number of citations) and percentiles should be used ([8] p. 869, [7], [6] p. 50, [4])
3. Research assessment should analyse scientific collaboration patterns and subject profiles of individual researchers as well ([6] p. 50).
4. It is preferable to use publication lists authorised by the authors in order to conduct bibliometric analyses rather than to solely trust information provided by databases.

In my view, the usage of bibliometric indicators will strongly depend on whether the validity of those indicators with regard to external criteria (e.g. peer review ratings, scientific career, post publication peer review, ex-post evaluation of funded projects) can be proven. „*They need to be jointly tested and validated against what it is that they purport to measure and predict, with each metric weighted according to its contribution to their joint predictive power. The natural criterion against which to validate metrics is expert evaluation by peers* …" ([9] p. 103).

References

[1] Hirsch JE: An index to quantify an individual´s scientific research output. *PNAS* **2005**, *102*, 16569–16572, DOI: 10.1073/pnas.0507655102

[2] Bornmann L, Mutz R, Hug SE, Daniel HD: A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. *J. Informetrics*, **2011**, *5*, 346–359. DOI: 10.1016/j.joi.2011.01.006

[3] Bornmann L, Marx W: The h index as a research performance indicator. *Eur. Sci. Ed.*, **2013**, *37*, 77–80.

[4] Bornmann L, Marx W: Evaluating individual researcher`s performance. *Eur. Sci. Ed.*, **2013**, 39, 39–40.

[5] Bornmann L, Mutz R: Further steps towards an ideal method of measuring citation performance: The avoidance of citation (ratio) averages in field-normalization. *J. Informetrics* **2011**, *5*, 228–230. DOI: 10.1016/j.joi.2010.10.009

[6] Wouters P, Glänzel W, Gläser J, Rafols I: The dilemmas of performance indicators of individual researchers. *ISSI Newsletter* **2013**, *9*, 48–53.

[7] Hicks D, Wouters P, Waltman L, de Rijks S, Rafols I: Bibliometrics: The Leiden Manifesto for research metrics. *Nature* **2015**, *520*, 429–431. DOI: 10.1038/520429a

[8] Cagan R: The San Francisco declaration of research assessment. *Dis. Model. Mech.* **2013**, *6*, 869–870. DOI: 10.1242/dmm.012955

[9] Harnard S: Validating research performance metrics against peer rankings. *Ethics Sci. Environ. Polit.* **2008**, *8*, 103–107.

Dr. Rüdiger Mutz
ETH Zürich
Professorship for Social Psychology and Research on Higher Education
Mühlegasse 21
8001 Zürich
Phone: +41 44 632 4918

mutz@gess.ethz.ch
www.psh.ethz.ch

Citation: Mutz R: Do we really need BIBLIO-metrics to evaluate individual researchers? *Infozine* **2016,** Special Issue 1, 21–22, DOI: 10.3929/ethz-a-010748893

Published: December 12, 2016

Peter I. Darroch, Lisa H. Colledge
Elsevier B.V.

Using research metrics responsibly and effectively as a researcher

Academic researchers today are faced with increasing competition for a limited number of academic positions and a decreasing pot of funding, all within a changing research policy environment. Furthermore, in the increasingly data-driven world, key performance indicators or research metrics are becoming more important than ever for a researcher's career prospects. Researchers are being challenged, and are under increasing scrutiny, to demonstrate both their academic and broader societal and economic impact.

Researchers therefore benefit from easy to use and effective qualitative and quantitative methods to help them demonstrate their excellence and manage their academic reputation. However, the number of metrics is increasing as research becomes more open and new data sources are indexed. Keeping track of the plethora of metrics with their benefits and limitations, the data sources behind them and the tools that present them, is becoming a significant challenge, especially bearing in mind that researchers are already over-stretched performing the many varied roles their academic positions demand of them.

Should a researcher show generated social media impact?

Publishing in good journals remains a key element of assessments for job, promotion and funding applications, and also in national assessments. Where a researcher publishes is not enough, though, to demonstrate or assess their excellence. Broader societal and economic impact of research seems to be requested in every application, for example. Which research metrics should a researcher use in their applications, and which should they not? Should a researcher show their h-index or a more complex, field-weighted indicator in a funding bid, alongside information about the journals they publish in? Should a researcher show the amount of social media activity their research has generated or mentions in the media to help demonstrate the broader impact of their research in their academic CV? Should they showcase their consultancy work with industry and the events they have organised for the general public?

Using Research Metrics: Two Golden Rules

In our roles within a research metrics team at a supplier of research data and metrics, we strive to make using metrics as easy as possible while ensuring that we encourage their use in a responsible manner. We promote two golden rules of using research metrics:

1. Always use quantitative metrics in combination with qualitative inputs such as peer review, as input into decisions.
2. Always use more than one metric, without exception, as part of the quantitative portion.

There are many ways that a researcher could choose to use research metrics. The most common are in a job or promotion application or a funding bid, to help the evaluator as they sift through the many applications they tend to receive nowadays. When building a view of their academic expertise or excellence, a researcher needs to demonstrate their qualifications in all areas of academic work: research, teaching and supervision, management and administration, as well as activities outside the university. Research metrics can play an important role in this across the more traditional output and citation based metrics to more recently developed non-citation based or so-called 'alternative metrics', which represent additional ways to demonstrate research impact and attention.

To help researchers navigate the research metrics landscape and use the best metrics for their decisions or demonstration of their academic excellence, there should be a broad basket of metrics, using the many different data sources available across the research workflow. Applying the extensive basket of metrics to models which will help researchers decide what metrics to use for a particular purpose or question, is one way to try and help ensure metrics can be used by everyone, responsibly and effectively.

Developing suitable models is not a straightforward task but we have drawn on our experience with the Snowball Metrics initiative [1] and our connections throughout the research community to build an initial version of a research metrics model for researchers, as illustrated in Figure 1. The model represents feedback we have gathered from stakeholders around the world, and highlights the many facets where researchers aim to demonstrate their excellence effectively and provide a more varied and nuanced view of their excellence. This extends from the demonstration of funding successes, productivity and quality, through to building a strong narrative about the broader impact and engagement of research through the use of alternative metrics such as media mentions and stimulation of activity in social networking services.

Facet	Theme	Metrics in areas of
Funding	**Awards** Can I support my research?	Number, value and duration of awards
Outputs	**Productivity** How productive am I?	Number, types and growth of outputs
	Visibility How prominent is my output in top outlets?	Impact of publication outlets
Research Impact	**Influence** How is my output used in academia?	Views, citations Reputation: awards, prizes, editorships
	Enterprise How is my output used in industry?	Commercial use (patents, licenses, spin outs, consultancy)
Engagement	**Network** How well linked am I within academia?	Collaboration: geographical, cross-disciplinary Network: number of collaborators, centrality, connectedness, geographical extent
	Connections How well linked am I outside academia?	Collaboration: cross-sector Celebrity: who's talking about me? Crowd-sourcing: collect and analyze data, raise funding
	Mentoring How do I transmit knowledge?	Who supervised me, and who have I supervised?
Social Impact	**Social Impact** What is my wider impact?	Direct and indirect impact on general public's well being, and understanding of research

Qualitative input

Figure 1: A model for a researcher's basket of metrics to help prove their excellence

By providing an extensive and accessible basket of metrics for all peers and entities, we hope to enable researchers to use research metrics responsibly as part of their day-to-day work, and so better prove their research quality and impact. In addition, it will hopefully help researchers effectively hold evaluators to account when evaluators are using research metrics to help them in their judgments.

References

[1] For more information, go to www.snowballmetrics.com

Peter I. Darroch
Senior Product Manager
Elsevier B.V.
Radarweg 29
1043 NX Amsterdam
The Netherlands

Phone: +44 7990 567 488
p.darroch@elsevier.com

Lisa H. Colledge
Elsevier B.V. /
Director of Research Metrics
Radarweg 29
1043 NX Amsterdam
The Netherlands

l.colledge@elsevier.com

Citation: Darroch PI, Colledge LH: Using research metrics responsibly and effectively as a researcher. Infozine 2016, Special Issue 1, 23–24, DOI: 10.3929/ethz-a-010744984

Published: December 12, 2016

Urs Hugentobler
ETH Zürich

Metrics in research: More (valuable) questions than answers

Research metrics are easier accessible than some years ago, are more sophisticated, but not always easy to understand – and these metrics are widely discussed and disputed. Many of the metrics provide an external view of an institution and on the impacts its researchers have in the scientific community. In international university rankings, these metrics are reduced to rank numbers that attract lots of attention.

Let's take for example indicators derived from publication and citation data: this information is used in a standardised way as proxy for productivity and for quality of research activities. At institutional level, it is often used to demonstrate research strength. But it also gains more and more importance in today's higher education landscape, as an aspect of accountability.

But how can we make use of research metrics in decision-making of a Higher Education Institution (HEI)? Does this metric help to understand the 'business' of a HEI? Or can those metrics support evaluating the impact of measures taken?

First a brief comment on the DORA discussion: DORA – the San Francisco Declaration on Research Assessment – asks '… that scientific output is measured accurately and evaluated wisely'. It therefore claims to refrain from using journal-based metrics such as the Journal Impact Factor (JIF) for evaluation purposes. Many organisations support this initiative by signing the declaration, including ETH Zurich [1]. However, we have to keep in mind that the past two decades have brought new developments in research metrics that go far beyond the JIF and its flaws. Article-based indicators such as field-based citation scores for example try to normalise the impact of a research output in order to make it more comparable. So, accepting DORA does not mean we have to ban any metrics from research evaluation, but we have to use "accurate measures and use them wisely".

Critical evaluation and self-evaluation is part of the academic value system. It is generally accepted that the citations received can be seen as a proxy for the usefulness of the results presented in the publication, an information that is valuable for the reader. This shows that the 'quality' of the information is not defined in absolute terms but is measured against the needs of the user of the information – in that sense quality is contextual.

An external view, based on standardised metrics, may help revealing patterns and findings for discussions and may provide different perspectives worth to be analysed and understood. But most importantly, any evidence the metrics show need to be interpreted in context of the relevant questions. Therefore, standardised and accepted research metrics in an informed peer review or in an appointment process can – as one aspect of the whole picture – support the decision-making with objective and accepted measures.

- A research profile based on publications and (normalised) citations received gives a picture in what scientific fields the unit is active and what impact the published work has compared to the peers. Such a profile at institutional level showing for example an over- all impact above world average may be adequate for monitoring reasons in an accountability reporting and for gaining attention. But it does not help if we want to use this outside view for verifying our quality assurance measures or for enhancing research quality within the institution. We need to link the outside view to internal organisational units and research groups in order to understand how these results develop.
- During a peer review process, one may find that a data-driven specific research profile of a unit *X* and its impact actually confirms the impression of the experts that were evaluating the unit. But does this help? Do those findings mean something for the strategy of unit *X*? Is the unit 'on the right path' to achieve its goals? A research profile alone, especially if it is only a snapshot instead of showing the development over time, cannot answer this question. But it can help providing indicators that show how the published research is perceived in the community.
- What if the research profile shows that the performance is below what is expected? Is the research of the unit ahead of 'main stream research' and therefore not yet visible in the metrics used? Or is nobody interested in the results published any more? Peers knowing the field should be able to answer questions like these.
- If the profile gives a wrong impression of the research activity and the performance, is it just a question of the publication strategy and of choosing the right journals? Are there major activities in this unit that are not captured by the chosen metric at all? What means would be needed to make it visible?

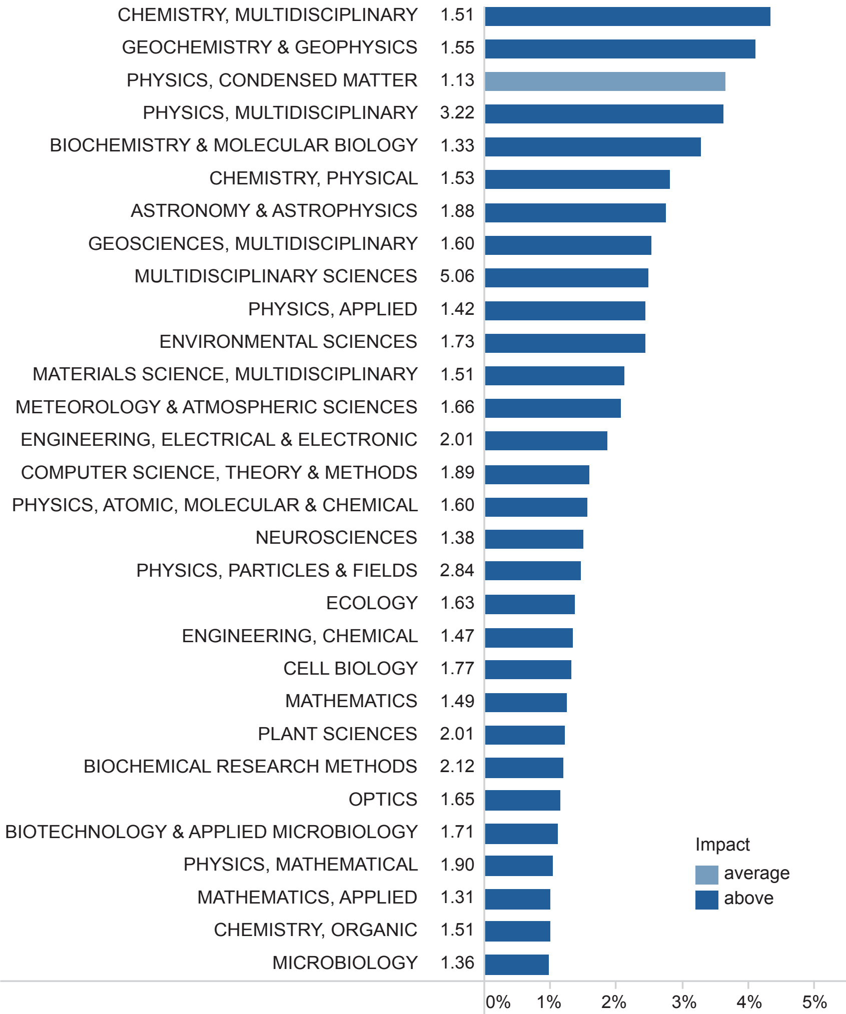


Figure 1:
An example of a Research Profile
Portion of publications per research field and weighted influence (MNCS); MNCS is above theworldwide average (> 1.2), at the average (< 1.2 and > 0.8), or below the average (< 0.8).

Research profile based on ETH Zurich publications 2003-2012. Shown in the chart are the 30fields holding >1% of the output each, covering 61% of all publications (of the total of 35700articles and reviews from Web of Science (WoS); Analysis by CWTS Leiden).

In this situation, alternative metrics such as views, downloads, mentions or shares in social media for example could add valuable information on the attention research results get. However, even though altmetrics in general provide more recent information, the interpretation of the findings is, again, more challenging. Compared to the above mentioned normalised citation based metrics at article level (where we evaluate the result against a well-defined total of similar articles), absolute numbers of downloads may be impressive but not very meaningful. Thus again, indicators need to be interpreted.

These are some of the many questions that arise when we want to make use of research metrics for internal decision support and for quality enhancement discussions. As HEIs worldwide are more than ever confronted with research metrics, we should at least try to understand why an institution shows up as it does, and make use of this information.

This is the approach we choose at the moment at ETH Zurich: analysing external views that are based on research metrics (e.g. from rankings) and bringing it in line with internal perspectives, but also providing detailed analyses to specific questions. Findings and analyses are discussed and the discussion reveals if measures have to be initiated or if the results serve the purpose of monitoring.

References

[1] As of October 23, 2016, 904 organizations signed DORA: http://www.ascb.org/dora-membertype-organizations

Urs Hugentobler
ETH Zürich
Head of Institutional Research Section, Controlling Department
Rämistrasse 101, HG F 65
CH-8092 Zürich
Switzerland

Phone: +41 44 632 22 97
urs.hugentobler@fc.ethz.ch

Citation: Hugentobler U: Metrics in Research: More (valuable) questions than answers. *Infozine* **2016,** Special Issue 1, 25–26, DOI: 10.3929/ethz-a-010748873

Published: December 12, 2016

Wilfred F. van Gunsteren
ETH Zürich

Publication of research results: Use and abuse

Why publish research results?

Scientific research lives of an exchange of ideas, opinions, methods, materials, equipment, software, and experience between researchers, both within a field of science and from different areas of science. Such an exchange can be oral or in written form, the latter mainly in the form of papers published in scientific journals. Because of the sheer size of the community of scientific researchers and its global spread, exchange in written form has gained weight compared to oral exchange. Publication of research results also allows every generation of researchers to stand on the shoulders of previous generations already deceased. This is why reporting of research results, be it positive or negative ones, in the scientific literature is of fundamental importance to the progress of our understanding and knowledge, *i.e.* of science. The scientific literature constitutes the repository of scientific knowledge. In addition, it offers the possibility to check and reproduce research data and results, a basic tenet of the scientific endeavour.

Quality checking of manuscripts

The integrity of published research is of fundamental value to the academic community of scholars. Since the 18th century quality control of research publications has been exerted by peer review: the judgement of scientific reports by academics with equivalent knowledge. Peer review can only function under the umbrella of the ethics of science. This assumes an unbiased examination of the opinion or data based on logical and empirical criteria. It also places trust in the competence and honesty of the reviewers, be they a colleague or a competitor. A reviewer should formulate an opinion on the quality of a manuscript:

1. Clarity of text, tables and figures.
2. Reproducibility of the results from the data reported.
3. Sound connection between the results and the conclusions (no overstatements).
4. Embedding of the results in the literature (proper referencing).
5. Relation to other methods addressing the same problem.
6. Novelty of the method or results.
7. Relevance of the results to the scientific community.

The quality of the reviewer's report is to be evaluated by the editor who requested it:

1. Apparent knowledge of a reviewer regarding the subject of the manuscript.
2. Validity and consistency of the arguments of a reviewer.
3. Possible bias because of a vested interest of a reviewer.

However, the increasing load of reviewing and editing manuscripts poses problems. Manuscripts are often only superficially read by reviewers and review reports are often only superficially read by editors. Instead of taking time to read a manuscript one observes an increasing reliance on simple so-called "quality" measures or indices when judging the quality of scientific research.

Are so-called "quality" measures or performance indices useful?

The time pressure on persons with the task to review research will inevitably induce them to rely on performance indicators rather than spending time to investigate in depth the research of a scientist. Yet, "quality" measures or performance indices are definitely not useful, they rather contribute to a degradation of the scientific endeavour.

1. A high-dimensional object (a research project, experiment, theory or model, or a person with its multiple tasks, activities, interactions, *etc.*) is projected onto or reduced to a single number or a one-dimensional object: a line of index values. Anyone who has looked at the projection of a 3-dimensional object (a chair or a house) onto a line (a 1-dimensional object) will understand that performance indices and rankings *etc.* are meaningless in regard to a characterisation of the original 3-dimensional object.
2. Numerical measures or indices measure quantity not quality. It is often assumed that quantity (number of papers, of citations, of downloads, *etc.*) is correlated with quality of research, quod non. Quality cannot be caught in a number.
3. Popularity or widespread use of a theory, method, software, or papers is not correlated with their quality. One only has to compare the quality of the *Bild Zeitung* with that of the *Frankfurter Allgemeine Zeitung*, of the *Blick* with the *NZZ*, or of the *Telegraaf* with the *NRC-Handelsblad*.

It is seductive to compare numbers. But, numbers lead to rankings, and rankings lead to competition. Excessive competition undermines care and rigour, encouraging activities close to or, ultimately, beyond the boundaries set by the ethics of science. The increasing pressure to violate academic principles is illustrated by the mount ing number of cases of plagiarism and scientific fraud. Focus on quantity as opposed to quality also leads to the aversion of risk: truly difficult and innovative research is shunned. A focus on competition will not enhance the quality of research. Quality measured by metrics alone is an illusion and the cost to society is growing inefficiency.

Is considering performance indices improving the quality of science?

The use of performance or citation indices sets the wrong incentives for researchers:

1. To consider one's own popularity to be more important than exchanging and criticizing research ideas and results of others;
2. To reference papers of oneself or of friends even if these are irrelevant;
3. To bias a review report towards one's own or friend's ideas of what must be correct;
4. To favour short-term (popular) simple research over addressing long-term basic scientific problems;
5. To bias against correct but unpopular theories, procedures or research results.

Thus the increasing use of performance indices tends to harm the quality of research. This implies that such indices are to be ignored. A research institution should not bother to provide data for their calculation. One should not mention, cite or consider them in any context. Indicators such as number of citations of publications, grant money gathered, number of successful students educated, or student satisfaction are only useful to detect extremes. A *curriculum vitae* with more than 1000 research publications must raise questions regarding the true involvement of the person in question in the research and the scope of the issues addressed. On the other hand, a lack of publication activity may indicate a lack of effort, the inability to finalise work, or reflect the difficulty of the research being executed.

If the *curriculum vitae* of an applicant for a professorship lists the number of citations, an *h*-index value or the amount of grant money gathered, one should regard this as a sign of superficiality and misunderstanding of the academic research endeavour, a basic flaw in academic attitude or, at best, as a sign of bad taste.

Recommended literature

Van Gunsteren WF: The Seven Sins in Academic Behavior in the Natural Sciences. *Angew. Chem. Int. Ed.* **2013**, *52*, 118–122, DOI: 10.1002/anie.201204076
Angew. Chem. **2013,** *125*, 128–132, DOI: 10.1002/ange.201204076

Van Gunsteren WF: On the Pitfalls of Peer Review. *F1000Research* **2015**, *4*, 1244, DOI: 10.12688/f1000research.7342.1

Van Gunsteren WF: Going for a PhD: Joys and Pitfalls. *Helv. Chim. Acta* **2016**, *99* 1–5.

Prof. Dr. Wilfried van Gunsteren
ETH Zürich
Laboratory of Physical Chemistry
Vladimir-Prelog-Weg 1-5/10,HCI G237
CH-8093 Zurich
Switzerland

Phone +41-44-632 5501
wfvgn@igc.phys.chem.ethz.ch
http://www.igc.ethz.ch/

Citation: Van Gunsteren W: Publication of research results: Use and abuse. *Infozine* **2016,** Special Issue 1,–27–28, DOI: 10.3929/ethz-a-010745085

Published: December 12, 2016

Eva E. Wille
Wiley-VCH Verlag

Wanted: Transparent algorithms, interpretation skills, common sense

Once upon a time

In 1988 the book "*The Timetables of Science*" by *A. Hellemans*, London and B. H. Bunch was published by *Simon and Schuste*r and in 1990 its German translation by *Droemer Knaur*. The authors presented about 10'000 science events from 2'400'000 B.C. to 1988 in chronological order for 10 subject areas. The index included ca. 3600 names; it still is a great overview of highest quality research.

In the same year I became head of the newly founded journals division of the society owned publishing house *VCH-Verlagsgesellschaft*, today known as *Wiley-VCH*; we published about 45 journals, many of which on behalf of societies, foremost the *Gesellschaft Deutscher Chemiker* (GDCh, German Chemical Society). Some of our time-honored journals like *Liebigs Annalen der Chemie* had a glorious history and of course the boards of GDCh/VCH discussed these journals a lot. Key Performance Indicators (KPIs) in those days were: Source of submitted manuscripts, types of manuscripts, their origin: industry, research institutes, West Germany/East Germany (!), number of pages and number of articles published, rejection rates, publication times, numbers of subscriptions inside/outside of Germany.

For most of the journals there was no peer review system, but a group of dedicated in-house desk editors worked very closely with a dedicated group of professors acting as "*Herausgeber*", primarily in Germany. No monitoring of Impact Factors/citation numbers, no *Nature Indexes*, no *Scopus*, no *Altmetrics* scores, no download numbers, no Hirsch (*h*) factors, no counting of app installations, no correlation with university rankings.

As a chemist by training I was always interested in experimenting and measuring the outcome in order to learn and gain more insight and also to inform our Editorial Boards. Thus, we soon started discussing impact factors, citations as well as later download numbers, and we realized for instance that such numbers can be correlated, but for good reasons sometimes they are not correlated at all; and they can be "gamed"!

Today

More research is done and published under enormous time and "return-on-investment" pressure in more countries by more scientists than ever. Big data is a buzz word for a growing group of scientists and companies, complex proprietary and confidential self-learning algorithms are influencing our daily lives: We see ourselves in the machine learning age and are more or less silently evaluated by software of various companies; we are in the hands of the GAFAs (*Google*, *Apple*, *Facebook*, *Amazon*) and more.

If you can't beat them, join and influence

Having served science and scientific societies from the publishing side for three decades, my experience is: What can be counted will be counted, what can be analyzed will be analyzed; it just varies how much we are prepared to pay for it and invest in it – time, energy and money wise. Realistically you can't stop this trend. Every responsible participant of "the publishing/science scenes" has to bring his or her knowledge, pragmatism, and common sense to the table to establish and maintain quality standards, a set of general values, and guard the ethics.

For the future

1. We need transparency about algorithms used to "calculate" metrics.
2. We need to understand their strengths and weaknesses, pros and cons – including how they can be influenced and "engineered".
3. We need to teach science administrators and other decision makers how the "metrics business" works, also students and early career researchers.
4. Ethical guidelines and standards have to be implemented by funding organizations, universities, and all other research institutions.
5. Basic rules like the Pareto principle or "less is more" as well as a culture of reading instead of downloading/importing references have to be applied.
6. Good research as well as good teaching have to be rewarded, not an increase in *h*-factors, for example.

When these or similar sets of rules are followed, distorting and distorted metrics will less distract from high quality research. It will be understood that:

1. Simple addition of impact factors for all articles published by members of an institution is not a metric for its innovation power.
2. Sex sells also science! High Altmetric scores are to be expected for scientific articles with headlines about the evolution of female orgasm or the length of penises, usually accompanied by press releases, social media storms by

3. authors and bloggers with many followers.
4. Some new services and their related KPIs are simply another layer of work and bureaucracy creating only additional noise without supporting science.
5. Some indexes and rankings are *l'art pour l'art*, at best good marketing following another "crazy" business idea. Less metrics is more, and what should rather be valued is common sense and risk taking. For this, reading and direct communication from scientist to scientist is the key.
6. High h-factors can be the result of publishing many innovative papers – but they can also be obtained by somebody writing many (mediocre) review articles about fashionable topics.

I would like to end with two quotes: "*Don't be dazzled by data or the latest technology. Big data has the aura of precision but often obscures the story*" (graphic designer Nigel Holmes), and "*We didn't know what we knew when history happened*" (historian Fritz Stern).

I would like to thank many board members of our journals, many journal editors as well as bibliometric experts like Hans-Dieter Daniel, Zürich, and colleagues like Iain Craig, Oxford, for many experiments and discussions.

Dr. Eva E. Wille
Wiley-VCH, VP, Executive Director
Wiley-VCH Verlag
Boschstr. 12
D-69469 Weinheim
Germany

Phone: +49 6201 606 272
ewille@wiley-vch.de
orcid.org/0000-0003-4045-7921

Citation: Wille EE: Wanted: Transparent algorithms, interpretation skills, common sense. *Infozine* **2016,** Special Issue 1, 29–30, DOI: 10.3929/ethz-a-010748933

Published: December 12, 2016

Renato Zenobi

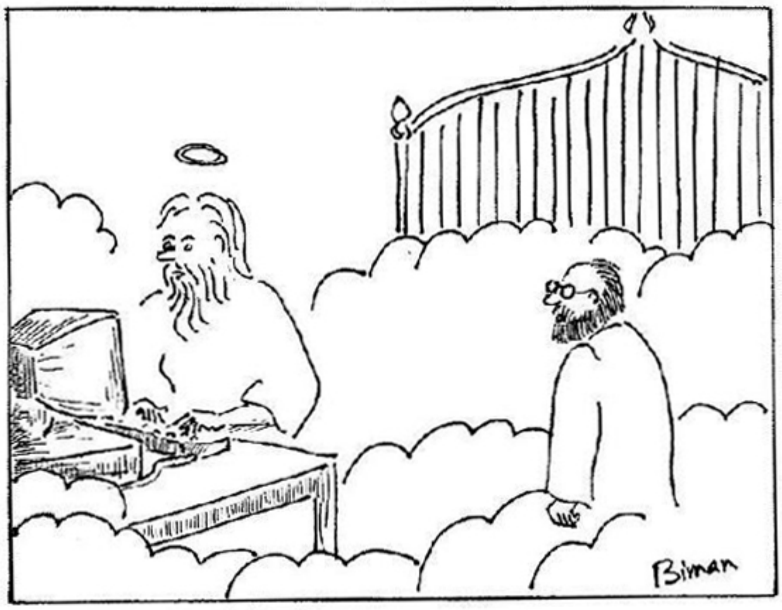
ETH Zürich and Associate Editor, Analytical Chemistry

Impact factors, the *h*-index, and citation hype – Metrics in research from the point of view of a journal editor

When we read a curriculum vitae, when we compare universities, when we decide where to publish a paper, when we discuss the journal landscape, a range of publication metric buzzwords surely but inadvertently pop up: should we choose the candidate with the highest *h*-index? Is ETH really not as good as MIT because it scores lower in the Shanghai ranking? Should we send our paper to the journal with the highest impact factor? Should the editors of a journal adjust their publication strategy to maximize the number of citations and, if so, short term or long term? I am serving as an associate editor for an American Chemical Society journal, Analytical Chemistry, and I am a member of several editorial advisory boards for other journals in my field. From these activities, I have some first-hand insight into the strategies of journals to improve their metrics. The thoughts in this short article express my personal opinion, but also come from the perspective and insights of a journal editor. Journals, publishers and editors are partly responsible for promoting and believing in metrics; some are downright obsessed with them.

Much has been said and written about metrics in research and in publishing. For example, I recommend reading the witty and lucid article written in CHIMIA by *Molinié* and *Bodenhausen* on this subject, "*Bibliometrics as Weapons of Mass Citation*" [1]. Although they have some limited value, I find many of the metrics irritating, and distracting from the main aim of science, which should be to advance knowledge and help translating this knowledge to improve the world we live in. There are a number of things that bother me:

- Metrics are often designed to reduce the value of a journal, a research program, a scientist, a university, or a paper to a single number. I find this totally insufficient. To me, this feels almost like "grading" your friends with numbers, such that the one with the highest grade becomes your best friend (or your spouse). There is much more to a friend than just a grade, and there is much more to the outcome of the scientific enterprise than some rank that is just a single number, because science is driven by people.
- As an editor handling a manuscript, I decide whether it eventually gets published or rejected. Rejection can happen at different stages: by editorial review alone, after consulting with another editor, or after one or several round of peer review. Many top journals, including *Analytical Chemistry*, have rejection rates well above 50%. One of the most difficult decisions an editor has to make is how to deal with manuscripts that contain very unusual approaches and ideas, which may be revolutionary but are destined to meet with resistance from the community. It is fairly easy to recognize high-quality mainstream



"I'll have to check your citation index first"

work from successful laboratories that will pass peer review without difficulties and will likely generate some impact in the field of the journal (in the form of citations, to use one metric). On the other hand, scientific "singularities" might either rightfully belong in the trash, or they might be paradigm-changing "gold nuggets" that are much more difficult to spot. A good journal and its editors will get such "gold nugget papers" occasionally and must do all they can to spot these and have the courage to publish them if they believe in the quality. It is not unusual that such a paper will at first not get any of the regular, immediate attention, that it will not be cited nor written about in the secondary scientific literature – but it could be that years after publication, such a paper will be recognized by everyone as the landmark paper that opened a new field. In other words, if measured by some of the customary metrics (e.g., the "immediacy index"), such a paper should have never have been published. Clearly, this would be a grave mistake!

- A large number of citations does not always mean that there is great science in the paper that is being cited. It can sometimes be the opposite, that it was a paper that was completely flawed (e.g. "cold fusion") or got a lot of attention due to some scandal.

- Metrics are often improperly used, sometimes in an amusing way. Consider the following: on hiring committees, I often see applications from candidates who list the impact factor of every journal they have published in. This is complete humbug! If anything, it's the impact of the particular paper (e.g., number of citations it has generated) that should be listed, rather than an average rating for the entire journal. I remember one candidate whose "ultimate rating" in the CV consisted of the sum of all impact factors of the journals where this particular person's papers had appeared …

- Metrics have the tendency to get uninformed and inexperienced researchers to behave a certain way, to maximize the value of whatever metrics they have in mind. This, I believe is shortsighted, almost equivalent to producing as many short lived "likes" as possible on a social media website for scientific output – rather than focusing on creating profound and sustainable impact. When members of my research group ask me "*could we publish in journal XYZ, because it has a higher impact factor than journal ABC?*", I explain to them that it is the value of their publication that counts, not the impact factor of the journal. For example the most highly cited paper from my own publication list has a rate of citations per year that is 26 times higher than the impact factor of the journal it appeared in. Every journal has a certain readership it reaches, in other words, a paper has to be well placed. It might not be seen by the target readership if "buried" in the wrong high-impact journal.

In summary, my opinion is that most if not all of the metrics are too simple to capture the true value of scientific advances. Metrics tend to be biased, and have to be interpreted with great caution. One should definitely not fall into the trap to adjust one's research, publishing practice, or even one's research field to optimize any given metric. Many of the metrics that gauge science may have been created for administrators or policy makers. The danger, of course, is that most administrators or policy makers have no other way of judging the value of scientific research, and may blindly take whatever metrics they see for face value. Science has become so diverse that even scientists can judge accomplishments outside of their specialty only with considerable effort, and may take the easy way out – to believe in some simplistic metrics. One of my favorite quotes on this aspect is one by Richard Ernst, Professor emeritus at the Department of Chemistry and Applied Biosicences, ETH Zurich, and winner of the 1991 Nobel prize in chemistry, who said: "*Very simply, start reading the papers instead of merely rating them by counting citations!*"

Acknowledgment: Author and publisher would like to thank Biman Nath for the permission of reproducing his cartoon http://www.rri.res.in/~biman/cartoon2.html

References

[1] Molinié A, Bodenhausen G: Bibliometrics as Weapons of Mass Citation. *Chimia* **2012**, *64*, 78–89. DOI:10.2533/chimia.2010.78

Prof. Dr. Renato Zenobi
ETH Zürich
Department of Chemistry and Applied Biosciences
Vladimir-Prelog Weg 3
CH-8093 Zürich
Switzerland

Phone +41 44 632 4376
zenobi@org.chem.ethz.ch
http://www.zenobi.ethz.ch

Citation: Zenobi R: Impact factors, the h-index, and citation hype – Metrics in research from the point of view of a journal editor. *Infozine* **2016,** Special Issue 1, 31–32, DOI: 10.3929/ethz-a-010748945

Published: December 12, 2016

Gabriella Karger
Karger Publishers

Rashomon or metrics in a publisher's world

The Rashomon effect is contradictory interpretations of the same event by different people.
Wikipedia, The Free Encyclopedia

The classic Kurosawa film deals with the nature of truth ... how the same events can be viewed in completely different ways by different people with different backgrounds, expectations, and experiences.
The Huffington Post

Einstein, in the special theory of relativity, proved that different observers, in different states of motion, see different realities.
Leonard Susskind (US physicist and mathematician)

These three quotes describe in the shortest and nicest possible way my personal opinion of metrics as an STM publisher: it is all a question of definition and interpretation, and this includes the person who does the defining and interpreting as well. In the following, I will expand on this within the context of my daily business as a biomedical publisher and in particular with regard to our main partners and clients.

What do we measure and who for?

More and more metric options, opportunities, and programs have surfaced in the last years to serve the diverging information needs of the different stakeholders in publishing. So what do we currently measure for these very distinct, diverse, and varied groups of partners?

Editors and Editorial Board Members
The classic Journal Impact Factor (JIF) and the related Journal Rankings reflect the relative importance of a journal within its predefined subject. To editors and editorial board members the JIF represents a visible and objective assessment of their efforts for the journal and hopefully shows that their endeavor has have been worthwhile. Furthermore, being affiliated with a high-IF journal does always look good on your CV.

The restrictions and limitations of the JIF are well known and have been extensively documented and discussed. We all know that being cited does not necessarily mean getting read and vice versa. However, as a long-term tool for observing and describing a journal's development within its field, the JIF is hard to beat. Its yearly report echoes the classic journal with its rhythm and growth cycle over years and decades.

Altmetrics, in contrast, refer to articles and how often they are cited, mentioned, or shared in social media. Based on this, one can conclude how well contents of a particular journal are distributed and discovered online. Altmetric data allow new insights and a fascinating, almost real-time tracking of an article's worldwide dissemination. Whether such a distribution can be equated to impact, however, is debatable. Altmetrics can thus be valuable feedback for editors: influential, much noticed articles may prove the value of a journal and confirm the decisions taken by the editors.

Authors
The JIF guarantees a certain quality standard and helps, together with the editor's and publisher's reputation, to build a secure foundation of trust for an author, influencing the decision of where best to submit a paper. But a JIF cannot be broken down from the journal to the article level in a meaningful way. Its per annum significance and its retrospective quality do not correspond to the generally much shorter life cycle of an article – and the author's expectations for immediate impact.

The *h*-index, in contrast, is a more recent initiative attempting to measure productivity and citation impact on the author's level.

Since the reception has shifted from print to online, STM publishers are focusing more and more on individual articles as independent units of information – rather than journals. The focus of dissemination is also shifting from traditional channels to online communities. For these reasons, altmetrics on article and author levels are steadily gaining importance. Aligned to the unique requirements on content level and independent of the publishing environment, altmetrics measure reach and impact of an article across web-based media and channels from shares on Facebook and Twitter to citations in Web of Science and Scopu*s*, as well as views, downloads, and bookmarks. Results are an immediate, continuously updated, and a highly visual assessment of success.

However, there is again a reason to be cautious: the mere fact of being quoted, mentioned, or shared and therefore counted in these metrics does not automatically confer quality or even correctness. This is nothing new: if it had not been for a misplaced decimal point in Erich von Wolf's notebook, Popeye would never have been created to promote spinach; the incorrect data was repeated and shared so many times it has become a myth still believed by many parents.

Readers
Nowadays most readers get their information from articles, book chapters, or even only abstracts. But how can you find the best information to read? How, where, and by whom a search is performed impacts the results that are shown. The all-important algorithms of search engines, based on their own intricate sets of metrics, help focus

search and bring back precise results so that there is little space left online for serendipity.

However, the list of potentially important material may still be too long. To separate the wheat from the chaff, factors defining the journal's standing such as the JIF, whether it is indexed in Medline, and the reputation of editors and editorial board members may be helpful but require time and effort.

Altmetrics on the article level are set up to help readers identify must-read material at a glance, by summarizing a set of quantifiable measurements in a nice colorful graph. Clearly, if so many others have read and shared this, I better read it, too. Well, again I would like to caution that a set of figures does not directly correlate with quality and substance. Otherwise why should Stephen Hawking's Facebook page show only 3'540'691 likes compared to Kim Kardashian West's 28'997'136?

Librarians

In times of tight budgets, tough choices must be made. Metrics such as the JIF, PubMed/Medline inclusion, and altmetric data can be helpful in identifying stable, reliable, consistently good products. However, straightforward key figures such as the number of downloads and turnaways or cost per download as well as general pricing policy and service seem to be much more central to librarians' decision making. This is why publishers supply usage statistics to our library partners and support initiatives such as COUNTER. No library will add a journal to its collection simply because it has a high JIF or its articles are often shared and commented on in Facebook. On the other hand, though, the subscription of a journal with a good reliable JIF, accepted by Medline and with other positive metrics, is less likely to be cancelled.

Publisher

As the JIF reflects the relative standing of a journal in its field, it yields important information over the years about the development of a journal and its field. Unfortunately, however, a high JIF is not an automatic guarantee for economic success. And neither does a relatively low JIF mean that a journal is automatically doomed (low submission rates are much more likely to spell the end for a journal).

Altmetric data may help you to get a more comprehensive picture and add another dimension to a journal's profile – in addition to established and essential figures such as the development of subscriptions, downloads, and submissions over time. In my personal view, another important dimension is longevity: quite a few of the journals my great-grandfather launched in the 1890s continue to play a role in research and clinical settings today (e.g., *Dermatology*, *Digestion*, *Gynecologic and Obstetric Investigation*, and *Ophthalmologica*). Is there an indicator for patience, for the ability of a publisher to allow journals to grow gradually and successfully over time? Quite often, it may take several years of engagement and commitment until the journal reaches a stage when metrics can be assessed in a meaningful way.

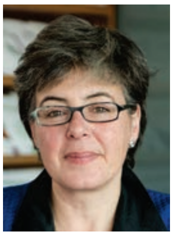
Key indicators used are based on data that can be counted, quantified and compared. However, in my opinion it remains open if quality can really be adequately expressed in a set of numbers, in any set of numbers.

We measure what is measurable – and useful

Part of our job as a publisher is to respond to our partners' and stakeholders' needs and demands for ratings and rankings, valuation, and validation. This is one of the services we provide to support our stakeholders in their specific roles. The JIF remains primarily important for editors, and altmetrics can best serve the needs of authors, while COUNTER statistics remain relevant for librarians.

Another part of our job is to make sure that everyone involved in science communication realizes that all data and the ensuing rankings are relative and in need of interpretation and contextualization. Measuring systems and methods measure what they were designed to measure within their defined parameters and on the basis of agreed assumptions. Nothing more, nothing less. Any attempts to draw other conclusions than those the systems are designed for or to change predefined criteria in the process will not work without a massive loss of meaningfulness. Like statistics, metrics should therefore always be taken with a grain of salt. Because, to quote Albert Einstein, in the end "all is relative" after all.

In my publishing world, we frequently rely on and assess qualities which cannot be adequately expressed in figures, numbers, and statistics. A good discussion with an editor, the supportive feedback of a customer, or an author's thank-you message cannot be expressed in any metrics. There are no indicators that can measure the excitement of launching a new journal or the fascination of a huge project like the new edition of *The Fabric of the Human Body* by Vesalius. But these factors greatly contribute to why we are doing what we are doing and have been doing for the last 125 years: serving and connecting the biomedical community to the best of our abilities.



Gabriella Karger
CEO
Karger Publishers
Allschwilerstrasse 10
CH-4055 Basel
Switzerland
Phone +41 61 306 11 11
Fax +41 61 306 12 34

karger@karger.com
www.karger.com

Photo: Copyright Christian Flierl.

Citation: Karger G: Rashomon or metrics in a publisher's world. *Infozine* **2016,** Special Issue 1, 33–34, DOI: 10.3929/ethz-a-010748951

Published: December 12, 2016

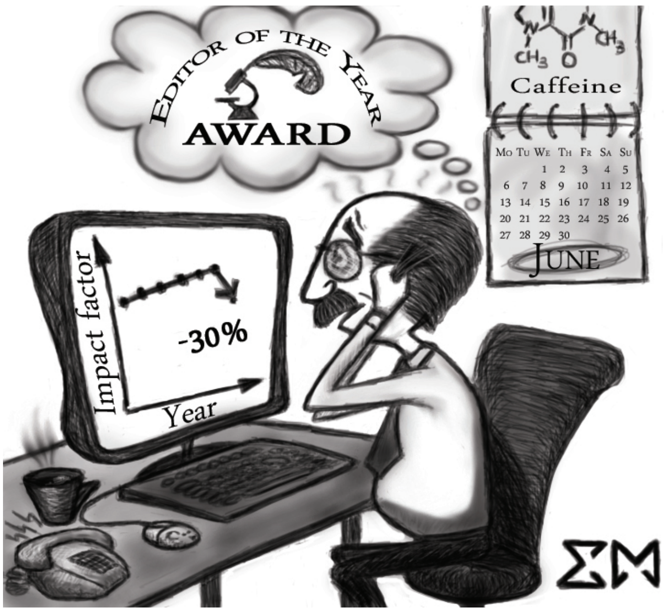
Jean-Christophe Leroux

ETH Zürich, and Associate Editor, Journal of Controlled Release

The impact factor and I: A love-hate relationship

The month of June is always awaited with feverishness by editors of scientific journals. It corresponds to the release of the new impact factors by the mass media and the information firm Thomson Reuters Corporation. Over the past 20 years, the impact factor has radically changed the way most journals are perceived by the scientific community. Editors become ecstatic when their journal has gained a few tens of a point and passably depressed when a small anticipated decline in the impact factor is confirmed. One could legitimately ask if this is reasonable and whether science is really benefiting from bibliometric indices.

Indeed, a couple of years ago, a quite recent journal in pharmaceutical sciences received an unexpectedly high impact factor, exceeding 11. At that time, I remember being quite skeptical about this value as the previous year it was barely above 3 and, to the best of my knowledge, I could not recall any outstanding and highly cited article from this journal that could have made its impact factor skyrocket. I contacted the editorial office to enquire whether this was a mistake or if this first-rate impact factor was real, but my query remained a dead letter. Surprisingly, after several weeks, the publisher issued a widely distributed flyer advertising the journal and its new incredibly high impact factor, to the great enthusiasm of our community. A few months later, in extreme confidentiality, the impact factor was corrected and brought down to a value closer to 4. Apparently, while calculating the impact factor, citations coming from articles published in another periodical had been erroneously attributed to this journal. The editor-in-chief was probably aware of this mistake from the beginning (most publishers calculate their own impact factors) but preferred not to immediately reveal it in order to benefit from a fugacious notoriety which could, during a short time, have attracted major contributions from reputed scientists.



Why am I sharing this anecdote? Initially, the impact factor was proposed as a tool for librarians to help them decide whether they should subscribe to a journal or not. After all, if a journal was rarely cited, it probably meant that it was not consulted very often. Today the impact factor is viewed as an indicator of the quality and relevance of the research. Authors want to publish in high-impact journals because their career (promotion, funding, notoriety, awards) will benefit from it, and publishers want their editors to make all possible efforts to increase the impact factor because it will augment the revenues generated by the journal. Editors are indeed magicians; they use all sorts of tricks to dope the impact factor.

As both an academic and associate editor, I have been playing this game for a number of years, but my belief is that this massive introduction of metrics in science is starting to harm the community. When the incentive to publish in a high-impact journal is too strong, publishing scientific findings is no longer a mean to disseminate and share discoveries, but becomes an objective in itself. The science reported in high-impact journals is generally claimed to be more attractive and pertinent than it actually is, and the reproducibility of the findings may turn out to be a secondary issue. This overall hysteria around the impact factor also bears a societal cost. It is nowadays frequent for scientists to have their papers rejected from prestigious journals multiple times before they are finally sent out for peer review. Once they receive the assessment of their cherished work, authors are usually asked to conduct a series of additional (and often unnecessary) experiments that do not change the conclusion of the article but make it appear "stuffed" enough to suit the journal's reputation. How much time and financial resources are wasted in this process? I do not know, but if it could be measured by a *frustration index*, I am convinced that it would be steadily growing.

The impact factor is not the only metric index that has infiltrated academics' daily life. Similarly, the *h*-index and total citation number are tools regularly used to assess the performance of researchers. These sprawling metrics are also giving birth to newer bibliometric mutants, and this is not even the end of it. As an ETH professor, I have the privilege to work in an institution that places the emphasis on the quality of the science rather than on numbers. Quality is difficult to measure, it requires time, discernment and a hint of vision, but isn't that what science is about? ... until the return of June.

Acknowledgment: Elena Moroz is acknowledged for drawing the cartoon.

Prof. Dr. Jean-Christophe Leroux
ETH Zürich
Institute of Pharmaceutical Sciences
HCI H 301
Vladimir-Prelog-Weg 3
CH-093 Zürich

Phone: +41 44 633 7310
jleroux@ethz.ch
www.galenic.ethz.ch

Citation: Leuroux JC: The impact factor and I: A love-hate relationship. *Infozine* **2016,** Special Issue 1, 35–36, DOI: 10.3929/ethz-a-010745051

Published: December 12, 2016

Ben McLeish
Altmetric

Personal experiences bringing altmetrics to the academic market

In late 2014 Altmetric released a platform (we call it the Explorer for Institutions) which allows institutions, academic or otherwise, to monitor, evaluate and take into account the online attention around scholarly materials which had been aggregating ever since the web became more of a socially-nuanced network than a simple series of websites. Attention has been collecting via social networks, blogs, within the stories on major (and minor) news sites, on *Wikipedia*, in policy sources (often as citations buried in thousands of PDFs) and even on deliberately academy-focused websites like *PubPeer*, *Publons* and *Faculty of 1000*. This attention has been accelerating as the world's scholars, journalists and the general public become more involved in the "great conversation" online around the fruits of research. The Altmetric Explorer helped bring all these disparate forms of attention together, not just on an article level, but on an author and institutional level too.

Since Altmetric decided to track this attention for everything which features a scholarly identifier from the very outset, our clients could effectively go back in time after implementing our system and see prior attention to items as far back as the beginning of 2012. Some clients found upon building their institutional profile that they had missed out on capitalising on huge online positive responses to certain publications, and had singularly failed to respond to legitimate negative feedback online for other research outputs. This was a self-evident area needing improvement, a place for communications, PR and media relations to bolster and reinforce its operations, and having access to this information has enabled changes to the engagement habits of these institutions in this brief interim.

Providing evidence of societal impact

I contend that the monitoring and measuring of engagement, re-use and commentary which collects around scholarly materials online is valuable for its own sake. These millions of interactions constitute important data around the impact and attention which research receives. They deserve to be factored into evaluations of institutional missions, help benchmark press and media success more meaningfully than mere downloads or page views can be expected to, and can also provide clear evidence of "societal impact" which is now required by many funding bodies around the world including *Horizon 2020* and the USA's *National Science Foundation*. Since I began working with libraries and companies in 2004, far and away the number one complaint I have heard from customers is regarding budget cuts imposed by institutions, company boards and governments. Surely Altmetric cannot be anything but a help to making a clear case for relevant, impactful and important research funding which should be defended vigorously, backed up by auditable data and real evidence of online engagement.

The reactions to this new arena of "altmetrics" have been both positive and negative, and for telling reasons in both sides. On the positive side of the balance sheet, authors, small institutions and less established voices in research have embraced the ability to gain an overview of the public data which collects around their works. Simply look at Altmetric's own Twitter account mentions to witness the number of authors thanking us for discovering huge news stories which mentioned their items, but which are not easily manually found; news organisations are not in the habit of exercising formal citation standards, and often mention the "*May 19 issue of Cell*" and a paper's co-author perhaps.

And yet clever data mining organised by the Altmetric team finds this needle of coverage in the internet's haystack; this offers the author an opportunity for real credit, another plank in their never-ending arguments for funding. Since each of our over 7 million details pages are freely available online for everything we track, we have handed authors and other interested parties real value about their research coverage. This openness places our efforts within the Open Data sphere as much as possible, while still operating as a business which was built to do the spadework and development required to avail the world of these data.

Institutions which already had a dim sense that there was an awful lot of discussion out there, discussion which is plainly invisible to bibliometricians who focus purely on academic citations, adopted Altmetric for Institutions at a speed I have not witnessed elsewhere in my 12 years in academic sales. Ask any of my clients, and they'll tell you I am not the pushy sort, either. These institutions understood instantly the benefits of these data, and were willing to develop some of their own use cases as well. They have been outstanding defenders of the broadening of the scope of attention tracking. Because Altmetric offers big data and an API, our clients have been as much the innovators behind new systems as our own agile development team.

But altmetrics as a field has met its fair few critics too, and I choose to outline these criticisms which I have come across at the close of this piece precisely because we can learn the most on the use and abuse of altmetrics from some of these attitudes.

High scores as an invitation for scrutiny

For one, there is a long-standing assumption that high attention scores in Altmetric correlate with or entail high quality. This is as untrue of Altmetric scores as it is of traditional citations (which is still by and large considered an indicator of academic success.) And yet some of the highest cited items are cited precisely for their ineptitude or flaws. Just as high citation counts do not automatically deliver praise to Andrew Wakefield's fraudulent and now retracted MMR/ Autism article in the *Lancet* many years ago [1], so a high Altmetric score only invites careful scrutiny of the actual coverage itself to gauge the nature of the attention and the impact of a particular article, and not an automatic endorsement.

Another criticism is the great worry of "gaming" metrics; essentially the claim goes that manipulation can inflate a score. While the point above on high scores not equating necessarily to good scores still answers some portion of this charge, two things are forgotten by those making this claim. For one, the sources Altmetric takes into account are hard to fabricate. Appearances in the online news media, policy PDFs from world-famous institutions such as the World Bank or the UK Government cannot be easily invented by researchers, and are some of the most compelling sources of attention outside of purely academic engagement. And secondly, it is much harder to game multiple metrics at once than to manipulate a single one – such as citations. One need only follow *RetractionWatch*'s excellent blog [2| to witness the multiple stories of citation stacking, faked peer review and other forms of manipulation to realise that if gaming were a problem in multiple metrics at once – such as Altmetric's multiple attention sources – it is an even more serious problem in just one metric – such as citations. Ultimately both citations and altmetrics are better off being evaluated together, providing a comprehensive, falsifiable and more complete picture of engagement.

There is a growing realisation now within research, either private or publicly funded, that relevance to society, or interest from society, is of vital importance for purposes of innovation and social development. At the same time, public literacy about research and science is a key factor in future funding, and in future recruitment of researchers from the general public. The steps I have seen taken by our brave and forward-thinking clients demonstrates a form of advocacy. An advocacy of a closer relationship between the academy and society, governments and industry.

On the research horizon, there should be as few conspicuous ivory towers as possible.

References

[1] RETRACTED: Wakefield, AJ et al.: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet*, **1998**, *351*, 637–641. DOI: 10.1016/S0140-6736(97)11096-0.
[2] http://retractionwatch.com

Ben McLeish
Product Sales Manager
Altmetric LLP
4 Crinan Street
London N19XW
Great Britain

Phone +44 1925321635
ben@altmetric.com
Twitter: benmcleish

Citation: McLeish B: Bibliometry: Personal experiences bringing altmetrics to the academic market. *Infozine* **2016,** Special Issue 1, 37–38, DOI: 10.3929/ethz-a-010748884

Published: December 12, 2016

Oliver Renn
ETH Zürich

Fatally attracted by numbers?

As a scientist, I am attracted by numbers as numbers tell me something. They are the result, the outcome of something I have planned – a quantitative proof of success. As an artist, I am absolutely not attracted by numbers as numbers do not tell me anything. Whether a piece of artwork I have created meets my expectations or not is solely based on my experience and gut feeling.

However, caution is recommended when gut feeling or experience is substituted by numbers and number-based metrics. This may lead to fairer evaluations – see below – but may also result in wrong decisions, as the algorithm behind the final score necessarily simplifies and standardizes the processes or outcomes to be rated.

Having occupied managerial positions for many years, including in STM (Science Technology Medicine) publishing, I am also attracted by numbers, and especially by key performance indicators (KPIs). Can numbers actually tell you which journals to continue or to discontinue publishing? When I was in scientific publishing 20 years ago, journal economics were analyzed by a multi-level contribution margin accounting. Level 1 accounts for costs and revenues directly related to a journal, while levels 2 and 3 subsequently consider overheads as well as general production and marketing costs. I will never forget the moment when I, as a publisher, analyzed my portfolio of journals. In my Excel spreadsheet I was trying to kick out those that had low performance on level 1, and suddenly realized that this had an unexpected and huge overall impact on the other journals, even resulting in negative metrics for journals that had been positively rated before. Playing with Excel, I finally found out that it was worth keeping journals that were poor on level 1, as those journals helped me to continue with others and to launch new ones. It helped me to understand the complexity and dependencies of my portfolio, including the various costs and showed me how to improve the entire portfolio. Fortunately, at that time, it was not senior management who played with the numbers but only me, combining experience, gut feeling, and the information provided to me by the numbers for achieving a sound and successful decision-making.

This was the start of my love for metrics. But I also had to learn quickly that metrics and KPIs can be misused. Misused by those, who do not understand the processes behind the metrics and simply define thresholds or percentage increases. Metrics are not for those who believe it is painful to think [1]. Metrics are a love-hate relationship. When analyzing metrics is solely up to me, it is love; but when others use the same figures, it may become dangerous.

Similar considerations apply to research metrics. The commonly used metrics, *e.g.* impact factor, *h*-index and now altmetrics, were actually created not as KPIs for third-parties but to ease the work of professional groups: *Garfield* invented the impact factor in the fifties [2] based on the work of *Gross & Gross* [3] with the goal of helping chemistry librarians to decide which journals to subscribe to. Today, however, the impact factor decides not only on the chemistry journal portfolio of a faculty, but on careers and success of researchers and publishers, and even societies – especially when they are also publishers.

In 2010 *Jason Priem et al.* coined the term altmetrics [4]. Like *Gross & Gross* and *Garfield* with the impact factor, their intention was to solve a problem related to the massive increase of literature. Not for journals but now for articles. The alternative metrics aimed at helping researchers to decide which articles they should read, as it has become impossible to read all the potentially relevant articles. However, also this metrics was quickly conquered by third parties, who started using it to decide on the careers of researchers.

Should we stop using metrics? I also recall a moment where I had the opportunity to sit behind a member of a search committee in a lecture hall. The candidate was somewhat red in the face, probably being nervous, as this was one of the most important talks of his life. The committee member noted "red face – potential cardiac problems – reject." This guy would probably have been better off with some sort of empirical metrics.

Metrics require brain. They are neither good nor bad *per se*, but require understanding and a balanced view. Metrics cannot be a quick substitute for a tedious decision-making process but can support your judgment, and less likely – as explained – those of third parties, such as senior managers and funders.

However, metrics are inevitable. In a data-driven society, there are numerical values that can be used as indicators and thus ratings. There are many systems out there that rate you and your performance. Apps track and rate your physical and social activities – if you want. You can, for instance, earn immediate "likes" when posting the view you have when writing at an outdoor office – like me when writing this article being on vacation [1]. And those "likes" are certainly motivating.

And also for researchers, there are meanwhile numerous platforms and tools where I can see how I am doing and how I compare to others. I get push e-mails from *Mendeley*, formerly *Elsevier's Researcher Dashboard*, informing me which of my articles are being read, cited, downloaded, or discussed. *ResearchGate* sends me e-mails telling me my score, same for LinkedIn. I can check my *Google Scholar* Profile and my *Altmetric* dashboard to see how my work is doing in the web, or my *Kudos* dashboard to see how I rank and where to improve, just to name a few. I can "altmetric" any article to see how this item is perceived. But you can use altmetrics also to see who is interested in your research and may find unexpected uptake, relations and may make interesting contacts.

Can I exclude myself from citations and mentions? It is definitely nice to see who read, mentioned and cited my articles, and to realize that my spreading-out activities have reached readers that otherwise would not have been aware of my work and of me. However, what will happen if everybody is using, *e.g.*, Kudos to raise awareness for their latest paper through LinkedIn, FaceBook, Twitter and the constantly increasing numbers of platforms? What will happen if every researcher or research group has a blog, a twitter or whatever site generating multiple and numerous push notifications? Most likely another group of individuals will come up with a tool that helps researchers finding the needle in the haystack. And there is not one needle. Actually, every blade of grass can mean a needle to someone, a important piece of information to build on and carry out new and important research.

I strongly believe that serendipity is a key ingredient in successful research. Metrics or KPIs (like alerts) that point you to the expected rarely help you developing fundamentally new ideas. It is the unexpected that takes your brain to new ideas, by connecting so far unconnected observations. We have to ensure that this important human ability is not getting lost by the use of metrics and KPIs, rankings and ratings. Although I have to confess that meanwhile I am receiving serendipitous information through *e.g.* Twitter, *i.e.* information that is out of my scope but interesting, and which I would have missed otherwise.

Lastly, being also in the position of an information scientist, I am naturally eager to explore and apply new tools. And I believe that it is better to be part of something new, to engage and ideally impact the developments, than to stay away and try to avoid the unavoidable. Metrics and user data-based analytics will be part of our future, in private life and in research. This is why we started early to work with Kudos and Altmetric and try to get our clients on board, helping them to explore the new possibilities, and ensuring that third-party use is as reasonable as possible. However, as often with new developments, uptake of new metrics is slow. As the opinion articles of ETH Zürich researchers show, the impact factor and now the *h*-index are the metrics researchers are familiar with. Altmetric scores are less known and used. Not every development is self-developing and self-promoting as it was with the smartphone and tablet, triggered by Apple.

Time will tell how the alternative metrics will develop, and get adopted by funders and researchers. Metrics could, in general, also lead to new appreciation models. *Bookmetrics* [5], for example, the metrics tools for books (Springer for the time being), could lead to a new revenue model. Authors could actually get paid according to the numbers of pages that are being read, rather than according to the number of books that are being sold (and often left unread). In an usage-based world, avid readers – especially those with a higher reading speed – would pay more as they could consume more information. Is that survival of the fittest?

In an open access and open science world, metrics and algorithms are suggested to serve as filters that direct you to relevant information in the post-journal era. However, "outcry" metrics are in favor of the loudest and not necessarily the brightest. A first step could be to measure whether a mention in social media is positive or negative [6] and, finally, the importance of the mention. This is already done based on the "value" of the source, but could also be brought a step further by content analyses. However, this makes you dependent on algorithms and also the *filter bubble* bias needs to be excluded.

References

[1] There is a book by Davis Rothenberg entitled Is it painful to think. Conversations with Arne Næss. Minnesota Press, Minnesota, 1992. Næss, the Norwegian philosopher used to write his books and essays in his cabin at Hallingskarvet. Writing this article also at a cabin at Hallingskarvet, I had to quote this phrase and the book.

[2] Garfield E. Citation indexes for science: A new dimension in documentation through association of ideas. *Science* **1955**, *122*, 108–111.

[3] Gross PL, Gross EM. College libraries and chemical education. *Science* **1927**, *66*, 385–389.

[4] Priem J, Taraborelli D, Groth P, Neylon, C: "Altmetrics: A manifesto. **2010.** http://altmetrics.org/manifesto (retrieved 30.9.2016)

[5] According to Martin Roelandse, SpringerNature, it is not the intention to set up a new model for the distribution of author revenues but a tool that motivates the authors by showing them the outreach of their work.

[6] On occasion of the 1:AM Conference in London in 2014 I had suggested to take into account whether a tweet is positive or negative. Publishing a bad paper, and getting lots of negative mentions, would nevertheless bring up the altmetric score. This was considered as (technically) not possible but was confirmed as doable by another attendee, married to a police officer.

Dr. Oliver Renn
ETH Zürich
Head Chemistry |Biology | Pharmacy Information Center
Head Science Communication D-CHAB
Vladimir-Prelog-Weg 10
CH-8093 Zurich
Switzerland
Phone +41-44-632 29 64
renn@ chem.ethz.ch
http://www.infozentrum ethz.ch
orcid.org/ 0000-0002-6966-7757

Citation: Renn O: Fatally attracted by numbers. *Infozine* **2016,** Special Issue 1, 39–40, DOI: 10.3929/ethz-a-010748927

Published: December 12, 2016

Gerd Folkers[1], Laura Folkers[2]
[1]*ETH Zürich*, [2]*Lunds Universitet*

On computable numbers

Recently, a Master student – obviously in his obligations as the head of an organizing committee – mentioned to the senior author of this paper that, in spite of his low publication performance as a director of the *Collegium Helveticum*, the organizing committee acknowledged a still acceptable *h*-index and hence expressed the wish to invite him for a lecture.

What had been meant as a compliment sounded at the end rather different. The semantics had been chosen in a strongly "scientific manner", by invoking quantification (*h*-index) and publications encompassing "articles in Science", but it came out somehow annoying. What he should have said, and very probably intended to say, was: "When you took over the new job, your contributions to Chemistry dropped remarkably; still, obviously, some people are interested in your old stuff. That's interesting, we want to hear something about that."

Why did he choose this emotionally distant wording? Because it sounded more objective? Because he wanted to be taken seriously as a scientist? Because he was nervous about a steep hierarchy to be expected?

Teaching should be encouraging

Especially the latter is much of a problem from the perspective of a student. Hierarchies are normal and common, simply because of the asymmetries in knowledge and understanding. The "art of teaching" is to handle them in a proper way. If "the egos are sitting in the ranks, and waiting for failures" (quoting a visiting post-doc's view) all frankness and originality is taken away from the young scientists, as *e.g.* the courage to ask questions without having them labeled "excellent question, thank you". They will be streamlining their lectures, preparing sophisticated PowerPoint presentations weeks in advance and hiding some of the most exciting results in their drawers for surprise, instead of discussing them immediately. Of course, good teachers know all about that. So what went wrong that makes the unpleasant cases?

Towards quantification

We think that "massification", and in its consequence "quantification", are the causes. Social control seems to work in small groups only and it comes in different flavors (see above). As soon as the lab meetings, the audience in the lectures, or the keynote speeches at a conference exceed a certain amount, numeric measures are often used for assessment. In a competitive environment, *e.g.* too many PhD students in a research group, the fight for attention ends up in counting the number of experiments rather than the sophistication of a single one, the number of manuscripts submitted instead of the clear thoughts in a single one, the number of credits per time instead of attending a lecture in a far distant field of interest for pure intellectual challenge.

Time and space have to be given to the young researchers. They should be allowed to follow their fascination, without loosing their focus. They should be able to broaden their knowledge, without becoming superficial. Both become impossible when framing research projects into small fragments and trying to publish all of them, without always and repeatedly rethinking the concept of the whole.

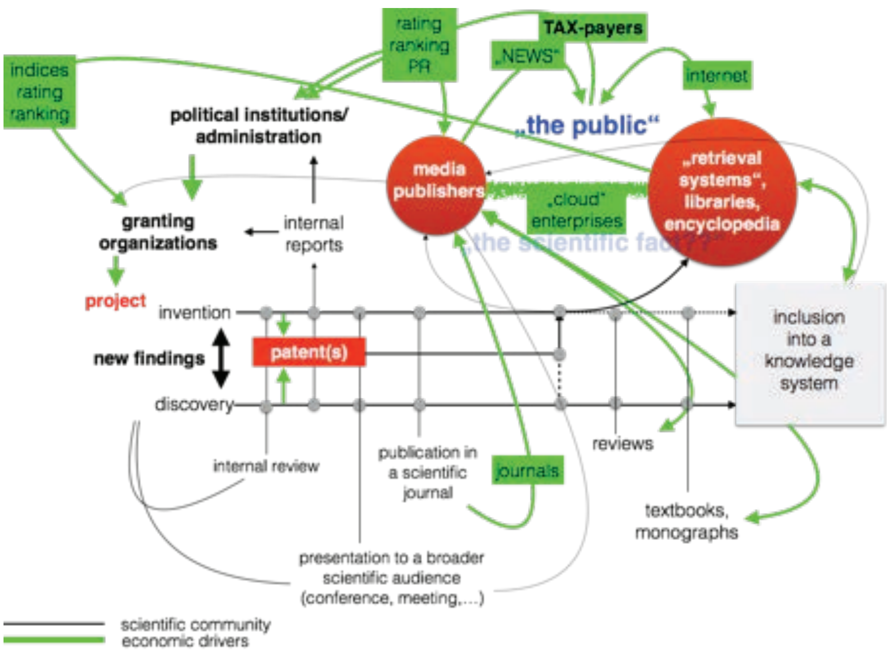


Figure 1: The "public life of scientific facts" (certainly not complete) seems to be a rather chaotic. Economic drivers interfere for the good or the bad at nearly every spot.

Economic behavior

As long as institutes or departments will ritualize their individual achievements, *e.g.* by demanding "*x* manuscripts accepted" for a PhD, "*y* papers published" for a habilitation, and "*z* first-author positions" in these papers, simple economics will come into play and the "Salami"-strategy will apply: if we need more papers, then we write more papers, hence we will have more journals, that will publish more papers, which will create higher numbers of publications for the individual, whose number of "first" or "last" author positions will increase simply by statistics, and who will eventually as a senior scientist demand more papers from others, which will … and so on.

A university is not an enterprise

Publishers parallel the tendencies by creating more and more journals. The most prestigious ones – by whatever perspective – specialize in a dozen sub-journals, cleverly marketing their "brand" along both ways: attention in the scientific community and market value for the commercial customers. Markets do not come without competition. The journal market, being especially tough is – matter of common knowledge – prone to error, manipulation and fraud. Occasionally it brings forth strange blossoms which could serve as case studies for undergraduate economics: A journal grants bonuses of USD 100 to authors the paper of which gets 20 citations within a year. This hamster wheel has to be stopped. Nobody knows exactly how to do that for the moment, but we need to establish an ongoing discussion about the ideals of higher education and the means to handle them. In contrast to an enterprise, a university can never be run "by numbers" and a university will never "make money". Too much quantification will on the contrary repel the creative minds, the real talents that we are seeking for.

What sounds very "*Schöngeist*" is in fact Hollywood's dream of education. Movies like "*A Beautiful Mind*" or "*The Dead Poets Society*" are in praise for fostering the development of individuals to its optimum by intellectual challenge. What their plots have in common is a charismatic and capable mentor who accompanies the students. That is what our students deserve and hopefully expect. That, by the way, is also in terms of economics the best choice a university can make.

Prof. Dr. Gerd Folkers
ETH Zürich
Department of Humanities, Social and Political Sciences
Rämistrasse 101
CH-8092 Zürich
Switzerland

Phone +41 44 633 87 07
gerd.folkers@gess.ethz.ch

Laura Folkers
MSc Chemistry (ETH Zurich)
PhD student Inorganic Chemistry
Lunds Universitet
Sweden

laura.folkers@chem.lu.se

Citation: Folkers G, Folkers L: On computable numbers. *Infozine* **2016,** Special Issue 1, 41–42, DOI: 10.3929/ethz-a-010745043

Published: December 12, 2016

Lawrence Rajendran
Founder, ScienceMatters

ScienceMatters – Single observation science publishing and linking observations to create an internet of science

End of 2013, Nobel Prize winner *Randy Schekman* publicly complained about the inefficiency of the academic publishing industry [1]. He claimed that from now on, he would not publish in the top-tier journals like *Nature*, *Science* and *Cell* anymore, because they distort the scientific process and are a "*tyranny that must be broken*". A few other Nobel Laureates including Sydney Brenner and Peter Higgs have also raised similar concerns. While it is the first time a Nobel laureate was so vocal about the science publishing platform, and it isn't without its own controversy (many observed that *Schekman* had himself published several papers in *Nature*, *Science* and *Cell* which undoubtedly paved the way to his Nobel prize, and that he was also using this occasion to promote his own newly founded journal *eLife*), numerous scientists – from graduate students to full professors – feel a deep dissatisfaction when the system demands that we publish science in high-impact journals.

Clearly, the fight might seem for one thing: that we should publish our scientific findings without artificial barriers put up by the publishers. This fight has become so ugly and widespread that its consequences include the appearance of more and more predatory journals and alternative ways of publishing. Of course, the question as to what is predatory remains up for debate – for many, these are mushrooming journals with no or low quality control, which thrive on the authors' paid article processing charges (APCs). There are a gazillion ones in recent years – and Jeffrey Beal, the librarian associated with the University of Colorado Denver who maintains a list of predatory publishers, has very recently been embroiled in a dispute with the Frontiers publishers, after listing some of their journals in the predatory list. For some, even journals that are reputed and have a decent review system, but charge the authors an APC of more 5000 USD for just a single paper (with 3 or 4 figures), also come close to being predatory, given that these journals are open access and don't have any printed version – so, the question is why such high costs for publishing on-line a content that the researchers entirely provide. The publishers claim that, by making the scientific content of the paper open-access, they have to charge the authors such an exorbitant amount.

This brings me to the other dark side of the science publishing – the closed or paywalled access of scientific contents. Publishers place artificial barriers not only in selecting the best, the sensational and the incredible (sometimes, literally so) stories that are expected to be of interest to the audience and, to some extent, also to be cited more and hence contribute to the increase in the impact factor, but they also actively place barriers for accessing published knowledge – in other words – there are barriers for creating knowledge as well as for accessing knowledge. This, in 2016, is just neither possible nor justifiable. Hence, ScienceMatters.

ScienceMatters was created to provide a free path to publishing solid and well validated observations, be they orphan, negative, confirmatory or contradictory ones, and to keep them openly accessible to anyone in the world. With single observations, we aim to democratise science and, to some extent, also to de-elitise it. What science and researchers need is a quick and easy way to publish their findings, and with ScienceMatters it is now possible. When a researcher makes an observation, he/she can publish it provided that it passes a relatively fast but thorough peer-review process. The publication criterion is the soundness of the observation rather than solely its significance. We believe that all robustly validated scientific observations should be published, regardless of their immediate or perceived impact. Hence we allow all observations to be published. However, the observations are ranked on account of their peer-review scoring (1–10, 10 being the highest), the score being based on three different measures: technical quality, novelty and impact. For example, for a paper to be to be eligible for publication in *Matters*, it needs to have at least a 4/10 on the technical quality and this alone is sufficient. If the observation scores 8/10 or above on all the categories, then it will be published in Matters Select. In this way, the nature/quality of the observation alone determines where it is published.

Furthermore, we introduce real-time publishing, which is making science in steps, akin to the lego^{TM} building of science. The single observations submitted by authors are developed into stories in real-time, allowing the story to develop progressively instead of demanding that a full story be submitted all at once from the start. After publishing a single observation, authors submit subsequent/related observations as *horizontal links*, *i.e.* linked observations provided by the same authors as the original core observation and that continue to build the story in real-time. They develop the original observation into a full story, but each observation is published independently of the story context and immediately citable. Other researchers may also pursue the observation with confirmatory, contradictory or extending data as *vertical links*. As a result, a narrative emerges that is more truthful, more collaborative and more representative of the complexity of scientific phenomena – similar to the network of an internet of science, except that the observation nodes are all peer-reviewed. This is what ScienceMatters aims at – creating an internet of science where all the observations are reviewed but also quantitatively scored based on the technical quality and impact.

Now, one can imagine a metric based on ScienceMatters – not depending on where the data are published (as is currently done) but depending on what kind of observation is published:

1. By allowing both confirmatory (positive) and contradictory (negative) data to be published next to the original observation (seeding node), the seeding node gets extended. As illustrated in Figure 1, our visualization algorithm enables the seeding node to be linked through edges that can either be positive (green edge) or negative (red edge). If a particular seeding observation can be reproduced by, say, five different groups (and not by one single group), then this seed has a high confirmatory score indicating that it is reproducible. However, if the seeding observation cannot be reproduced by many groups and has mainly contradictory links, then it has a low confirmatory score and a high contradictory score. We believe that this is important, as such identification measures could enable or even predict the success or failure of clinical trials or the translatability of the findings. In addition, we created Matteric™, the metric that measures the "seeding potential" and the "extension potential" of both the author and the observation itself. Seeding potential refers to how powerful a seeder an author/ observation is, *i.e.* how many further links were based on that particular node. For example, imagine that a rural researcher in Indonesia, without much knowledge about molecular biology or mechanistic insight, discovers that a certain herbal extract has the potential to reduce psoriasis, and publishes this "single but robustly validated observation" in Matters. Assume that this observation is extended by others in terms of mechanistic insights, identifying molecular and immunological basis, industry trying to replicate the observation in other cohorts, and pharmaceutical chemists isolating the very compound or a mixture of compounds responsible for the activity. Then, this particular study from the rural Indonesian researcher and her/his single observation would qualify as a great "seeder". And how far that researcher extended the observation is another measure of focus and persistence. We combine these two factors in the Matteric™, which we believe is a much better and a more direct measure of impact than the existing "journal's impact factor".

2. And the post-publication public review: In addition to the pre-acceptance peer-review and the post provisional acceptance pre-publication public review, all articles published in Matters and Matters Select will also have a possibility for the public to post comments, reviews, shares or likes (upvote).

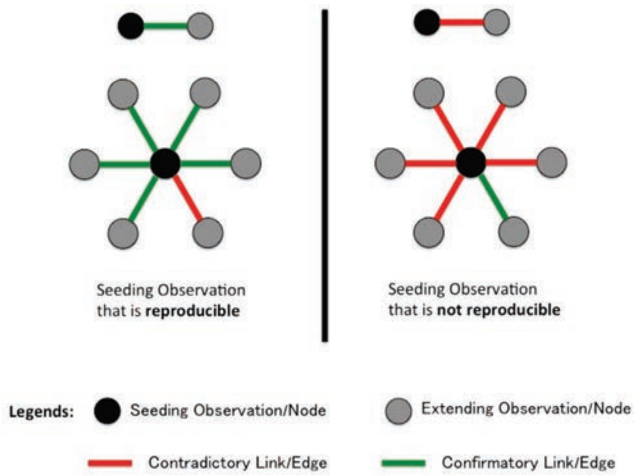


Figure 1: Principles of Matteric, the metric of ScienceMatters.

Thus, ScienceMatters comes in timely with its innovative concept of single-observation publishing to address many aspects of the crisis in science – however, two things are capital in bringing this innovation to be implemented: Money and mindset. While money can be bought, borrowed and even made, changing mindset is a challenge, and particularly that of scientists. But we remain optimistic as this is the only way to bring in the change.

References

[1] Sample I. "Nobel winner declares boycott of top science journals". The Guardian. December 9, 2013, retrieved October 22, 2016 www.theguardian.com/science/2013/dec/09/nobel-winner-boycott-science-journals

Prof. Dr. Lawrence Rajendran
Founder, ScienceMatters
Velux Stiftung Professor, Systems and Cell Biology of Neurodegeneration, University of Zurich
Wagistrasse 12
CH-8952 Schlieren
Phone: +41 44 634 88 63
lawrence.rajendran@irem.uzh.ch
www.sciencematters.io

Citation: Rajendran L: ScienceMatters - Single observation science publishing and linking observations to create an internet of science. *Infozine* **2016**, Special Issue 1, 43–44, DOI: 10.3929/ethz-a-010748900

Published: December 12, 2016

Quick Reference Cards *for* Research Impact Metrics

Find a printable version of larger cards at https://goo.gl/GXwqRN

Metrics illuminate the impact of your research outputs. Promotion and tenure committees, funders, advisors, research team leaders and potential collaborators are all interested in information about impact. **But where to start?** Your library can advise you on metrics — found on Elsevier products or via other sources — that can help you to:

DECIDE WHERE TO PUBLISH
Impact per Publication
SJR: SCImago Journal Rank
SNIP: Source Normalized Impact per Paper
Journal Impact Factor

ADD TO ONLINE PROFILE
h-index
percentile benchmarks
scholarly activity online
scholarly commentary online
social activity online
mass media stories

ENRICH PROMOTION & TENURE PORTFOLIO
h-index
percentile benchmarks
scholarly activity online
scholarly commentary online
citation count
mass media stories

APPLY/REPORT TO FUNDERS[1]
h-index
percentile benchmarks
scholarly activity online
scholarly commentary online
citation count
mass media stories
journal metrics (e.g., SNIP, SJR)

BENCHMARK A COLLECTION OF RESEARCH OUTPUTS (for team leaders)
percentile benchmarks
Field-Weighted Citation Impact
h-index (if in the same field)
Field-Weighted Download Impact[2]

DOCUMENT | AUTHOR | JOURNAL

"Document" in the definitions refers to primary document types such as journal articles, books and conference papers. See Scopus Content Coverage Guide (page 9) for a full list of document types: https://goo.gl/bLYHov

Indicates that the Snowball Metrics group agreed to include as a standardized metric, which is data source and system agnostic. https://www.snowballmetrics.com

CITATION COUNT

# of citations accrued since publication

A simple measure of attention for a particular article, journal or researcher. As with all citation-based measures, it is important to be aware of citation practices. The paper "Effective Strategies for Increasing Citation Frequency"[3] lists 33 different ways to increase citations.

DOCUMENT COUNT

# of items published by an individual or group of individuals

A researcher using document count should also provide a list of document titles with links. If authors use an ORCID iD – a persistent scholarly identifier – they can draw on numerous sources for document count including Scopus, ResearcherID, CrossRef and PubMed. Register for an ORCID iD at http://orcid.org.

FIELD-WEIGHTED CITATION IMPACT (FWCI)

# of citations received by a document / expected # of citations for similar documents

Similar documents are ones in the same discipline, of the same type (e.g., article, letter, review) and of the same age. An FWCI of 1 means that the output performs just as expected against the global average. More than 1 means that the output is more cited than expected according to the global average; for example, 1.48 means 48% more cited than expected.

h-INDEX

# of articles in the collection (h) that have received at least (h) citations over the whole period

For example, an *h*-index of 8 means that 8 of the collection's articles have each received at least 8 citations. *h*-index is not skewed by a single highly cited paper, nor by a large number of poorly cited documents. This flexible measure can be applied to any collection of citable documents. Related h-type indices emphasize other factors, such as newness or citing outputs' own citation counts.[4]

IMPACT PER PUBLICATION (IPP)

# of citations in present year for journal documents from past 3 years / total # of papers published in past 3 years in that journal

This score can be used for any serially published collection of publications. It is similar to the Journal Impact Factor in the way it is calculated, but allows for more comparison across disciplines, since disciplinary citation practices mean that not all journals will reach their peak citation rate within the Journal Impact Factor's 2-year period.

SCIMAGO JOURNAL RANK (SJR)

average # of weighted citations received in a year / # of documents published in previous 3 years

Citations are weighted – worth more or less – depending on the source they come from. The subject field, quality and reputation of the journal have a direct effect on the value of a citation. Can be applied to journals, book series and conference proceedings.

Calculated by Scimago Lab (http://www.scimagojr.com) based on Scopus data.

SOURCE NORMALIZED IMPACT PER PAPER (SNIP)

journal's citation count per paper / citation potential in its subject field

The impact of a single citation will have a higher value in subject areas where citations are less likely, and vice versa. Stability intervals indicate the reliability of the score. Smaller journals tend to have wider stability intervals than larger journals.

Calculated by CWTS (http://www.journalindicators.com) based on Scopus data.

JOURNAL IMPACT FACTOR

citations in a year to documents published in previous 2 years / no. of citable items in previous 2 years

Based on Web of Science data, this metric is updated once a year and traditionally released in June following the year of coverage as part of the Journal Citation Reports®. JCR also includes a Five-year Impact Factor.

PERCENTILE BENCHMARK (ARTICLES)

compares items of same age, subject area & document type over an 18-month window

The higher the percentile benchmark, the better. This is available in Scopus for citations, and also for Mendeley readership and tweets. Particularly useful for authors as a way to contextualize citation counts for journal articles as an indicator of academic impact.

OUTPUTS IN TOP PERCENTILES

extent to which a research entity's documents are present in the most-cited percentiles of a data universe

Found within SciVal, Outputs in Top Percentiles can be field weighted. It indicates how many articles are in the top 1%, 5%, 10% or 25% of the most-cited documents. Quick way to benchmark groups of researchers.

SCHOLARLY ACTIVITY ONLINE

# of users who added an article into their personal scholarly collaboration network library

The website How Can I Share It? links to publisher sharing policies, voluntary principles for article sharing on scholarly collaboration networks, and places to share that endorse these principles including Mendeley, figshare, SSRN and others.
http://www.howcanishareit.com

SCHOLARLY COMMENTARY ONLINE

# of mentions in scientific blogs and/or academic websites

Investigating beyond the count to actual mentions by scholars could uncover possible future research collaborators or opportunities to add to the promotion and tenure portfolio. These mentions can be found in the Scopus Article Metrics Module and within free and subscription altmetric tools and services.

SOCIAL ACTIVITY ONLINE

# of mentions on micro-blogging sites

Micro-blogging sites may include Twitter, Facebook, Google+ and others. Reporting on this attention is becoming more common in academic CVs as a way to supplement traditional citation-based metrics, which may take years to accumulate. They may also be open to gaming.[5]

MEDIA MENTIONS

# of mentions in mass or popular media

Media mentions are valued indicators of social impact as they often highlight the potential impact of the research on society. Sources could include an institution's press clipping service or an altmetric provider. Mendeley, Scopus (Article Metrics Module), Pure and SciVal (coming in 2016) also report on mass media.

1. Metrics selected will depend on the funders' interests and project strengths.
2. Plume, A. & Kamalski, J (March 2014). "Article downloads: An alternative indicator of national research impact and cross-sector knowledge exchange," *Research Trends*, http://www.researchtrends.com/issue-36-march-2014/article-downloads/
3. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2344585
4. See a good explanation at http://www.harzing.com/pop_hindex.htm
5. http://www.altmetric.com/blog/gaming-altmetrics/

V2.0_researcher_2016

Content by
Elsevier Library Connect
https://libraryconnect.elsevier.com
& Jenny Delasalle
Freelance librarian & consultant
@JennyDelasalle

Scopus is the world's largest abstract and citation database of peer-reviewed literature with content from over 5,000 publishers.
https://www.scopus.com

IPP, SNIP and SJR are provided free at:
https://www.journalmetrics.com

SciVal offers data-based insights into 7,500 research institutions and 220 nations worldwide to visualize research performance, benchmark relative to peers, develop collaborative partnerships and analyze research trends.
https://www.scival.com

MENDELEY is a free reference manager and academic social network where you can organize your research, collaborate with others online and discover the latest research.
https://www.mendeley.com

infozentrum
Chemie | Biologie | Pharmazie

Want to know more about research metrics and how to apply them?
Contact your: Chemistry | Biology | Pharmacy Information Center, www.infozentrum.ethz.ch/en/team/staff or Science Communication D-CHAB, www.chab.ethz.ch/en/outreach/scientific-communications

ELSEVIER