2020

Altmetrics and societal impact measurements: Match or mismatch? A literature review

Iman Tahamtan
University of Tennessee, Knoxville, tahamtan@vols.utk.edu

Lutz Bornmann
Administrative Headquarters of the Max Planck Society

Recommended Citation

Tahamtan, Iman; Bornmann, Lutz (2020). "Altmetrics and societal impact measurements: Match or mismatch? A literature review". El profesional de la información, v. 29, n. 1, e290102. https://doi.org/10.3145/epi.2020.ene.02

Altmetrics and societal impact measurements: Match or mismatch? A literature review

Iman Tahamtan; Lutz Bornmann

How to cite this article:

Tahamtan, Iman; **Bornmann, Lutz** (2020). "Altmetrics and societal impact measurements: Match or mismatch? A literature review". *El profesional de la información*, v. 29, n. 1, e290102.

https://doi.org/10.3145/epi.2020.ene.02

Invited manuscript received on December, 15th 2019
Accepted on December, 27th 2019

Iman Tahamtan ✉
https://orcid.org/0000-0001-7750-0183

School of Information Sciences, College of Communication and Information
The University of Tennessee
Knoxville, TN, USA
tahamtan@vols.utk.edu

Lutz Bornmann
https://orcid.org/0000-0003-0810-7091

Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society
Hofgartenstr. 8
80539 Munich, Germany
bornmann@gv.mpg.de

Abstract

Can alternative metrics (altmetrics) data be used to measure societal impact? We wrote this literature overview of empirical studies in order to find an answer to this question. The overview includes two parts. The first part, "societal impact measurements", explains possible methods and problems in measuring the societal impact of research, case studies for societal impact measurement, societal impact considerations at funding organizations, and the societal problems that should be solved by science. The second part of the review, "altmetrics", addresses a major question in research evaluation, which is whether altmetrics are proper indicators for measuring the societal impact of research. In the second part we explain the data sources used for altmetrics studies and the importance of field-normalized indicators for impact measurements. This review indicates that it should be relevant for impact measurements to be oriented towards pressing societal problems. Case studies in which societal impact of certain pieces of research is explained seem to provide a legitimate method for measuring societal impact. In the use of altmetrics, field-specific differences should be considered by applying field normalization (in cross-field comparisons). Altmetrics data such as social media counts might mainly reflect the public interest and discussion of scholarly works rather than their societal impact. Altmetrics (*Twitter* data) might be especially fruitfully employed for research evaluation purposes, if they are used in the context of network approaches. Conclusions based on altmetrics data in research evaluation should be drawn with caution.

Keywords

Bibliometrics; Societal impact; Altmetrics; Citations; Research evaluation; Scholarly communication; Social media; *Twitter*; Review article.

1. Introduction

Science and technology have had an extensive influence on different aspects of society; they have improved and made drastic changes to the quality and length of peoples' lives, and have shaped our culture, beliefs and thinking (**Burke**; **Bergman**; **Asimov**, 1985). In the past, it was clear to everyone, specifically governments and science policy-makers, that science and technology should be financially supported because of their outcomes and impacts on society (**Godin**; **Dore**, 2005): over time, good science brings magnificent changes to the development of new ideas, technologies, discoveries, and makes its potential impact on society (**Kreiman**; **Maunsell**, 2011). Vannevar Bush proposed that high-level research is the best thing for a society (**Bush**, 1945). His thoughts and efforts meant that most governmental policies in the US should focus on providing universities and research institutions with the best possible facilities and resources in order to conduct high quality research (**Van-den-Akker**; **Spaapen**, 2017).

However, the climate has changed: over recent decades, universities and research institutions have been increasingly expected and pushed to evaluate the impact and success of their research projects that are funded by governments and public funders. All institutional actors have in response increased resort to indicators and methods for measuring the impact of their research. Universities have been facing financial and resource shortages for decades, making it extremely competitive for researchers to obtain scarce institutional funding for their research projects. Today, allocation of university finances and resources primarily depends on the extent to which researchers can communicate the benefits of their research to research managers and convince them. Research evaluation has become a core element in the management of research and scientific policymaking (**Wilsdon** *et al.*, 2015).

Over many years, measuring the impact of research referred to impact on academia (based on citations). Today, universities and other research institutions are required to pay more attention to the broader impact of their research, by thinking about how their research can address and solve real world problems (**Bornmann**, 2016). To do so, they have tried to go beyond the academic impact of research and investigate the environmental, health, economic, public, and social concerns of people, industries, and other sectors. They have also started investigating how these issues should be addressed by the research projects they design and propose. To achieve such practical goals, scholars and industry sectors may work together to initiate impact activities for research projects. University knowledge- or technology-transfer offices can help accelerate and facilitate this process by assisting scholars to

(a) brainstorm research ideas that could potentially be of interest and helpful to broader groups of people,
(b) write impact statements to describe the significance of research, and
(c) assist researchers collaborate in a more efficient way with industrial sectors (**Dance**, 2013).

The broader use of evaluative bibliometrics, a term proposed by **Narin** (1976), began in the 1980s for the purpose of measuring the performance of science (**Roemer**; **Borchardt**, 2015). Evaluative bibliometrics is concerned with indicators for measuring scientific productivity, research collaboration, and citation impact (**Belter**, 2018). Although these indicators are not truly perfect or unbiased (**Moed**, 2017), they have been used in a more or less standardized way in research evaluation. However, scientometrics has experienced important changes since the 2000s. Demands for measuring societal returns gave birth to other indicators by which the societal impact of research could possibly be measured (**Miettinen**; **Tuunainen**; **Esko**, 2015). In this paper we review the literature on societal impact measurement. We particularly address three major questions:

1) Which methods have so far been employed by universities, researchers, and research funding bodies in order to evaluate the societal impact of research?
2) Are altmetrics valid indicators for measuring the societal impact of research?
3) If altmetrics are not valid indicators for measuring the societal impact of research, what do they measure?

2. Societal impact measurement

Societal impact measurement has its roots in the early 1970s, when increasing pressure from policymakers and society was put on universities and research institutions to produce more relevant and practical research that would be of economic value (**Van-den-Akker**; **Spaapen**, 2017). Since then, Western governments have paid more attention to the societal benefits of their financial investments in academic research (**De-Jong**; **Smit**; **Van-Drooge**, 2016). Societal benefits and impact became especially highlighted as a criterion in supporting and funding research projects in many disciplines during the 1990s (**De-Jong** *et al.*, 2016; **Mostert**; **Ellenbroek**; **Meijer**; **Van-Ark**; **Klasen**, 2010).

> "In the 2000s, the EU and its member states started to develop frameworks for analyzing wider societal impacts of academic research, a task that was related to the introduction of the so-called third mission of universities" (**Miettinen** *et al.*, 2015, p. 258).

The aim of the third mission was to carry out applied research in accordance with the country's real needs. The third mission was proposed after the first and second missions of universities, which mainly focused on education and research. Third mission activities included both collaboration between universities and enterprises, and interaction between universities and society (**Göransson**; **Maharajh**; **Schmoch**, 2009). For instance, Denmark's third mission, initiated in 2003, emphasized that universities should closely collaborate with society and their educational programs, and that research findings should contribute to the further development of society (**Gregersen**; **Linde**; **Rasmussen**, 2009).

UK was also among the countries that made plans to capture and measure broader external impact of research (**Samuel**; **Derrick**, 2015; **Wilsdon** *et al.*, 2015). In the UK, the *Research Excellence Framework* (*REF*) took the responsibility of measuring the quality and societal impact of research in higher education institutions, since 2014 (known as *REF 2014*). *REF* evaluates the *ex post* (after the event) research impact by considering impact on different sectors such as economy, culture, public policy, health and welfare, environment, the quality of life, creativity, production, and international development (**Samuel**; **Derrick**, 2015). The *REF 2014* has provided

> "a rich picture of the variety and quality of the contribution that UK research has made across our society and economy" (*Department for Business, Energy & Industrial Strategy*, 2016, p. 21).

For the UK organizations, research is considered to have an external impact when it has a positive influence on a sector outside academic organizations (**Wilsdon** *et al.*, 2015). For these organizations,

> "evidence of external impacts can take a number of forms – references to, citations of or discussion of an academic or their work; in a practitioner or commercial document; in media or specialist media outlets; in the records of meetings, conferences, seminars, working groups and other interchanges; in the speeches or statements of authoritative actors; or via inclusions or referencing or weblinks to research documents in an external organisation's

websites or intranets; in the funding, commissioning or contracting of research or research-based consultancy from university teams or academics; and in the direct involvement of academics in decision-making in government agencies, government or professional advisory committees, business corporations or interest groups, and trade unions, charities or other civil society organisations" (**Wilsdon** *et al.*, 2015, pp. 44-45).

The Netherlands was one of the frontrunners in national research impact evaluation, with a focus on the economic benefits of publicly funded research. The Netherlands investigated the influence of research at larger scales beyond academia on developments in society (**Donovan**, 2008). "Societal quality" of research was first proposed in the Netherlands in the early 1990s (**Van-der-Meulen**; **Rip**, 2000), and was defined

"as the value that is created by connecting research to societal practice and it is based on the notion that knowledge exchange between research and its related professional, public and economic domain strengthens the research involved" (**Wouters**, 2016, annex, p. 25).

In the Netherlands, the *Standard Evaluation Protocol* (*SEP*) oversees measuring the impact of science since 2003. *SEP* considers three criteria, including

- scientific quality,
- societal impact, and
- viability of the research unit.

Productivity is not considered because it is believed that it would negatively affect the quality of research. *SEP* is a program similar to *REF* in the UK, in that research units provide case studies or narratives to highlight their impact on society. Research units in the UK have also been presenting case studies showing the impact of their research on the economy and society since 2014. Analyses of the UK case studies by the *King's College London* and *Digital Science* (2015) indicated that the societal effects in the UK case studies were multi-disciplinary, with more than 60 unique impact topics (**Van-den-Akker**; **Spaapen**, 2017).

In the current post-academic science, the focus of research is on application, known as Mode 2 research. Mode 2 research goes beyond Mode 1 research (the academic perspective on science). Mode 1 research was focussed on filling a gap in knowledge, contributing to science or building a theory. The main objective of Mode 2 research is rather to produce practical results based on real problems that are socially relevant, and which would have a tangible impact in society (**Ernø-Kjølhede**; **Hansson**, 2011). Researchers have determined different categories for Mode 2 research. The *RQF Development Advisory Group* (2005) classified the research impact of Mode 2 research into four broad categories:

(a) "social benefit" such as improving quality of life, solving social issues, and improving knowledge;
(b) "economic benefit" such as reducing costs improvements in service delivery and employment;
(c) "environmental benefit" such as decreasing environmental pollution;
(d) "cultural benefit" such as enriching the culture of a society.

The main differences between academic and post academic perspectives on research are shown in Table 1.

Table 1. The academic and post-academic perspectives on science and research

	The academic perspective on science	**The post-academic perspective on research**
Norms regulating scientist/researcher behaviour	CUDOS[a], although perhaps not always an accurate description of reality, is and should be the ideal	Norms vary and are context-bound, but may be characterised through applying a mix of *Place/Mitroff*'s norms[b] and CUDOS – with an emphasis on *Place/Mitroff*'s norms
The purpose of science/research	To accumulate certified, true knowledge as an end in itself	To produce knowledge for practical application
Quality is evaluated by:	Disciplinary gatekeepers referring to intra-scientific criteria (reliability, consistency, originality, objectivity)	Practitioners and peers using both intra- and extra-scientific criteria (relevance, utility, economic impact)
The individual scientist/researcher should:	Be independent and autonomous. The scientist 'owns' his or her work and publishes it in peer-reviewed journals to pursue an individual career	Be managed in accordance with societal/organizational/corporate objectives. Researchers may pursue both individual scientific and organizational careers, in which case publication is not necessarily of the essence
The prime source of control is:	Peers in the prestige hierarchy	The management of the employing organization
Best possible development of the institution of science/research takes place through:	Self-organisation	Design by institutional and political management
Typical exponents	Merton, Hagstrom, Barber, Popper, Bush, Storer, Polanyi	Fuller, Gibbons *et al.*, Etzkowitz and Leydesdorff, Ziman

Notes: [a] Proposed by **Merton** (1968a) and **Merton** (1968b), CUDOS refers to four social norms for scientific behaviour: communism, universalism, disinterestedness, and organized scepticism (**Ernø-Kjølhede**; **Hansson**, 2011).
[b] A set of norms (known as *Place* norms) proposed by **Mitroff** (1974) as opposed to Merton's social norms. These norms include rationality and non rationality, emotional commitment, particularism, solitariness, interestedness, and organized dogmatism (**Ernø-Kjølhede**, 2000).
Source: **Ernø-Kjølhede** and **Hansson** (2011, p. 133).

2.1. Possible methods for measuring societal impact

Academic researchers and experts, mainly economists, are supposed to measure the impact of research using a variety of indicators. Many efforts have been carried out in this area, yet there is a lack of straightforward mechanisms to measure output and outcome. Since the 1950s, economists have begun investigating the impact of science on economic growth and productivity. Economic evaluations for research impact include research cost-benefit, cost-effectiveness, and cost-utility analyses. Economic assessments refer to whether investing on a research returns the dollars that have been spent on it. The assessments also include whether the research has public and private returns. For instance, economists studied the impact of research on international trades. Results showed that science can have many economic impacts on areas such as production, financing, investments, commercialization, and budget (**Godin**; **Dore**, 2005).

> This overview of the literature discusses whether altmetrics data can be used to measure the societal impact of research

A classic and highly-cited study in this area was conducted by **Mansfield** (1991), who measured the economic benefits of basic research using a sample of 76 firms in the USA. **Mansfield** (1991) asked the companies' research and development managers what percentage of the company products and processes could not have been developed without relying on academic research. The results of this study indicated that 11% of the new products and 9% of new processes could not have been developed without academic research. The study also found that 2.1% of sales for new products relied on substantial aid from research and 3% of sales for new products would have been lost without academic research. Some years later, **Mansfield** (1998) conducted a follow-up study with 77 firms and found similar results. 15% of new products and 11% of new processes could not have been developed without relying on recent academic research.

The economic approaches to research evaluation have their own shortcomings. The most important one is that they focus on only one dimension of possible impact (which is an important dimension but does not reveal the full impact picture). For instance, economic approaches may not consider the impact of research on improvements in people's quality of life (**Searles** *et al.*, 2016).

Langfeldt and **Scordato** (2015) have described the various methods for measuring societal impact in general. One method is to conduct surveys and interviews with the target group that is supposed to benefit from the research output. Surveys and interviews can be used to determine the perceived benefits and usefulness of research from the viewpoints of the beneficiaries. The authors pointed to some limitations of using surveys and interviews. One major issue within measuring the societal impact

> "is defining and identifying both the potential target group and research that they have benefited from, within a specific time-frame in which the survey results are still relevant for decision-makers" (**Langfeldt**; **Scordato**, 2015, p. 25).

A major challenge is whether the target group can recognize the potential benefits of the research. The other issue is that although we might be able to define a specific target group, it would be difficult to measure the long-term impact of research on these groups. Surveys are also limited in terms of the number of people researchers can recruit and study, thus may not cover all potential users that are representative of the target group (**Langfeldt**; **Scordato**, 2015).

Searles *et al.* (2016) have overviewed three methods to measure societal impact: "payback", "economic evaluations", and "case studies" (we will discuss case studies in section 2.3, because it is one of the main methods used). **Searles** *et al.* (2016) noted that the payback or return investment on research was first proposed by **Buxton** and **Hanney** (1998) in the UK, and that the payback framework is frequently used in the health area (**Raftery**; **Hanney**; **Greenhalgh**; **Glover**; **Blatch-Jones**, 2016). Through interviews with researchers, the impact of their research can be understood (**Searles** *et al.*, 2016). **Buxton** and **Hanney** (1998) identified five categories of possible research payback, including "knowledge impact", "impact on future research", "political impact", "economic impact", and "impact on different health sectors." A promising way to measure these paybacks is to look at similar past projects and compare their impact on society (**Buxton**; **Hanney**, 1998). Payback methods, however, might sometimes become tedious and require evaluating many documents and resources (**Searles** *et al.*, 2016).

In developing indicators for societal impact measurement, **Molas-Gallart**, **Salter**, **Patel**, **Scott**, and **Duran** (2002, p. iv) demand that they should be

> "simple, measurable, actionable, relevant, reliable and reproducible, and timely."

However, these requirements are difficult to fulfil, because societal impacts include a variety of ethical, safety, economical, legal, and political issues (**Spaapen**; **Dijstelbloem**; **Wamelink**, 2007). This is the reason why solutions such as the payback framework are expensive. **Ovseiko**, **Oancea**, and **Buchan** (2012) conducted a survey on 139 clinical faculty members in order to review and evaluate the validity of the 20 impact indicators which were proposed by the *Higher Education Funding Council for England* (*HEFCE*) to help *REF* in allocating public research funding to higher education institutions (**Ovseiko** *et al.*, 2012). This study indicated that most indicators had some validity; however, concerns still exist regarding the operationalization and measurement of these indicators in a reliable manner. This study indicated that more valid and reliable indicators are required for measuring societal impact.

Willis *et al.* (2017) identified indicators for societal impact measurements of *Applied Prevention Research Centres* (*APRCs*) using the Delphi method. They asked stakeholders to identify the most important indicators. They collected a variety of indicators from existing research impact literature and asked the stakeholders to rate the importance and feasibility of these indicators, which resulted in eight indicators rated as highly important and highly feasible, such as being cited in public policy documents (see Table 2 for an overview).

Table 2. High importance indicators with high and low feasibility ratings

Capacity building indicators	**DM[c] indicators**
Items with high importance and high feasibility to measure[a]	
Stability of center funding for indirect/overhead costs	Use of research in informing public health programs and policy: input from public health practitioners / policymakers about what research has been used to inform their work
Number or percentage of center projects driven by expressed policy / practice needs of engaged policy / practice organizations	Citations in public policy documents
Evidence that research and policy and practice perspectives have jointly informed center priorities and activities	Evidence of center contributions to supporting decision-making processes and groups (e.g. steering groups, ministerial working groups, and government committees)
Number and type of knowledge exchange activities undertaken by the center	
Number of center projects involve co-production of knowledge with knowledge users	
Items with high importance and low feasibility to measure[b]	
Reputation of the center for producing relevant, credible, high-quality, and timely research	Center contributions to policy development, implementation, and enforcement for population-based prevention
Center influence on changes over time in knowledge, skills, and commitment / intention to use research findings among knowledge users in policy, program implementation, and administrative positions	Evidence of sustained impact of center research over time
Evidence the center has contributed to building the field of prevention research and its application.	
Impact of a center's partnership output/products on advancing relevant prevention research and its application.	
Influence of center knowledge exchange activities on relevant audiences, including center staff, students, researchers, decision makers, practitioners, etc.	

Notes: [a] Identified using importance/feasibility means across all groups: indicators with highest five importance ratings (falling above the mean) with feasibility ratings falling above the mean.
[b] Identified using importance/feasibility means across all groups: indicators with highest five importance ratings (falling above the mean) with feasibility ratings falling below the mean.
[c] DM means decision-making.
Source: **Willis** *et al.* (2017, p. 87).

Wilsdon *et al.* (2015) and **Thelwall** and **Kousha** (2015) have pointed to two important metrics for quantitative measurement of the wider societal impact, "clinical guideline citations" and "Google patent citations." **Thelwall** and **Kousha** (2015, p. 615) emphasized that among all indicators,

> "only *Google Patents* citations and clinical guideline citations clearly reflect wider societal impact and no social media metrics do."

We would like to add systematic review citations, since they are related to clinical guideline citations: systematic reviews are extensively cited in clinical guidelines. One way to measure societal impact using clinical guidelines is to investigate the citations mentioned in clinical guidelines and identify the functions of those citations: for instance, having an impact on clinical practices, contributing to the diagnosis of diseases and medical disorders treatment (see **Jones**; **Hanney**, 2016; **Taylor**, 2013a).

> "Being cited in a clinical guideline is direct evidence that a study has had a societal impact by guiding medical practice" (**Thelwall**; **Maflahi**, 2015, p. 960).

Some studies have already used this approach by investigating references cited in clinical guidelines (**Nelhans**, 2016). For example, **Grant** (1999) analysed the funding bodies in the 235 references cited in three clinical guidelines. The median time lag between the reference publication date and the guideline publication date was 5 years. 34% of cited references in clinical guidelines were supported by industry, 22% by government, and 17% by the private non-profit sector. Using clinical guidelines for measuring impact is not without biases. Guidelines may be spatially biased in citing references: for instance, a study indicated that the references from Edinburgh and Glasgow had a high frequency in the UK clinical guidelines (**Lewison**; **Sullivan**, 2008a).

> Three useful metrics for quantitative measurement of societal impact seem to be ‘clinical guideline citations’, ‘*Google* patent citations’, and ‘policy document citations’

Systematic reviews in medical sciences have become very important due to their great potential of improving health services. One source for publishing systematic reviews is *Cochrane* (see *https://www.cochranelibrary.com*), which has been shown to play a major role in improving health services such as introducing new treatments.

Cochrane is a good example of a publication-based approach of how research results transfer from research to practice. *Cochrane* reviews can have economic benefits by improving clinical quality and reducing unnecessary procedures (**Bunn** *et al.*, 2015). However, it is difficult to assess whether *Cochrane* reviews have a direct impact on health services or an intermediate impact through clinical guidelines (**Lewison**; **Sullivan**, 2008b).

According to **Wilsdon** *et al.* (2015) and **Thelwall** and **Kousha** (2015), patent citations are an important metric for quantitative measurement of the wider societal impact besides “clinical guideline citations” (see above). Many citations in patents are journal articles. Bibliometric analysis of citations in patents can be used to evaluate the impact of publicly-funded (medical) research on innovations in different industry sections (**Ovseiko** *et al.*, 2012). **Narin**, **Hamilton**, and **Olivastro** (1997) investigated the flow of knowledge from publicly funded research into industries by analysing the proportion of publicly funded research cited in the US patents. They indicated that 73% of the references in the US patents were supported by publicly funded academic research. This shows that the US industry relies a lot on publicly funded research. **Narin** *et al.* (1997, p. 330) concluded that

> “the science that is contributing to high technology is mainstream; it is quite basic, quite recent, published in highly influential journals, authored at major universities and laboratories, and supported by *NSF*, *NIH*, the *Departments of Defence* and *Energy*, and by other public and charitable institutions.”

Callaert, **Van Looy**, **Verbeek**, **Debackere**, and **Thijs** (2006) found that approximately 50% to 55% of non-patent references in patent documents were research papers indexed in the *Web of Science* (*WoS*, *Clarivate Analytics*). However, the coverage of scientific references in patents varies across disciplines. In the most recent study, **Poege**, **Harhoff**, **Gaessler**, and **Baruffaldi** (2019) matched 4.8 million patent families and 43 million publication records. The authors found a strong positive correlation between the quality of publications cited in patents and the value of the respective inventions. They ranked

> “patents by the quality of the science they are linked to. Strikingly, high-rank patents are twice as valuable as low-rank patents, which in turn are about as valuable as patents without direct science link” (**Poege** *et al.*, 2019).

2.2. Problems in measuring societal impact

The *Higher Education Funding Council for England* (2009) has proposed several key challenges to measuring impact, including time lags, attribution and limitations of metrics. Other researchers have confirmed these challenges (e.g. **Langfeldt**; **Scordato**, 2015).

The time lag between when the research is carried out and when the research comes to have its impact on products, processes or social practices is a major issue in measuring societal impact **(Spaapen**; **Van-Drooge**, 2011). Research needs time to become measurable (*Higher Education Funding Council for England*, 2009; **Oancea**, 2013; **Spaapen**; **Van-Drooge**, 2011). It is not clear which time frame should be considered in order for the impact of the research to become evident (*Higher Education Funding Council for England*, 2009). According to the *Health Economics Research Group* (2008, p. 7),

> “the issue of lags has often been ignored in the past, but from a policy point of view may be crucial, especially in the context of the current agenda for translational research.”

Measuring impact seems achievable when considering longer time frames. One can expect long time lags between when research is adopted and embodied in society and when its impacts can be measured (**Feller**, 2017). As a rule, societal impact is longer-term impact, such as research that contributes to improve the quality of life, decreases unemployment or improves a nation’s health and safety (**Ruegg**; **Feller**, 2003). In the literature, many authors have pointed to the problem of (long) time lags in measuring societal impact:

> "sufficient time must have passed to capture a major part of eventual impact, but not so much that traceability to research projects becomes difficult" (**Van-der-Meulen**; **Rip**, 2000, p. 13).

> "It may take years before knowledge is applied and has impact" (**De-Jong**; **Van-Arensbergen**; **Daemen**; **Van-der-Meulen**; **Van-den-Besselaar**, 2011, p. 62).

According to **Kok** and **Schuit** (2012, p. 2)

> "pathways from research to 'impacts' are very diverse: sometimes short and traceable, but often long, through multiple reservoirs, and via utilization at untraceable times and places."

Which estimated time frames have been presented for the process from research to practice? The Dutch physicist Hendrik Casimir proposed 15-year time windows for the research results to have practical applications. Other studies have offered to use 10- or 20-year time windows; considerable impact from research projects within short time windows cannot be expected, unless for exceptional breakthroughs in science (**Eisenstein**, 2016). According to **Morris**, **Wooding**, and **Grant** (2011), in biomedical fields the time lag from the publication of research to the manifestation of its impacts is almost 17 years,

> "whether that impact represented formal adoption of a medical intervention or marketing of a new drug" (**Eisenstein**, 2016, p. s21).

However, **Morris** *et al.* (2011, p. 510) have emphasized that in order to understand time lags, we should first agree on "models, definitions and measures; which can be applied in practice." There might be field differences in the time lag between a piece of research and its practical impact, since

> "different kinds of research achieve impact over different time scales", according to **Raftery** *et al.* (2016, p. 50).

Delaying measuring research impact for several years after the completion of the research project might raise some concerns (**Langfeldt**; **Scordato**, 2015).

First,

> "the evaluation may no longer be a relevant basis for decision-making [about current research]" (**Langfeldt**; **Scordato**, 2015, p. 25).

Second, in the long term, the outputs of research projects might be combined to different extents far beyond the initial individual projects, so that their impact is almost impossible to capture and measure (**Langfeldt**; **Scordato**, 2015). This problem can apparently be avoided by focusing on the time shortly after finishing a project. However,

> "most often *ex post* evaluations of research projects and programmes take place shortly after the completion of the projects/programme, and before impacts can be substantially identified" (**Langfeldt**; **Scordato**, 2015, p. 25).

One solution to the dilemma that the societal impact of research right after its completion cannot be substantially measured and long time frames fail because of the relevance problem, is to focus on the potential (i.e. not attained) impact of research (**Langfeldt**; **Scordato**, 2015).

Attribution is another key challenge in measuring societal impact. Attribution is

> "the extent to which an impact can be attributed to a particular research project [which] is a matter of judgement" (**Raftery** *et al.*, 2016, p. 50).

It's difficult to "attribute" specific impacts to certain research, because impact can be influenced by many intervening factors (besides the focal research). Thus, it cannot be reliably said whether the impact has been the result of the research activity or other factors (e.g. other research; see **De-Jong** *et al.*, 2011; *Higher Education Funding Council for England*, 2009). Even if we are able to determine which research the society has benefited from, it would be challenging to determine "what benefits can be attributed to what cause", which is known as the "causality problem" (**Martin**, 2007). Another problem with the impact attribution of research is that impact might exist, but one is not aware of it. Interviews with people involved in certain research activities and target groups such as stakeholders may discover these covert instances of impact. For instance, a "change in behaviour" may not be considered as being an impact, although in fact, it is; or improvement in "quality of life" may be considered as a simple change in behaviour, while it is so important (*Siampi*, 2011).

There are also difficulties in developing indicators for third-stream activities, such as differences across disciplines (see section 2.1). Disciplines may require more or less time to have an impact in society, such as industry, economy, and people's lives. Therefore, it might not be a good and practical idea to compare the performance of third-stream activities across universities and different disciplines (**Oancea**, 2013). Which further difficulties and problems exist in developing these indicators?

- First, some universities are finding it challenging to measure, track, and demonstrate the impact of their research, because they have not been trained and prepared to do so. Some universities do not have the required knowledge, infrastructure, experts, capability, and explicit policy to implement these initiatives (**Oancea**, 2013).

- Second, there is a lack of straightforward methods for measuring societal impact (**Langfeldt**; **Scordato**, 2015). Since different fields have different societal impact, the notion of societal quality and its operationalization must be broad. Operationalizing societal quality, therefore, depends on the nature of the research field (**Van-der-Meulen**; **Rip**, 2000).
- Third, the success of research projects is scarcely predictable, and serendipity may play a (major) role in the development of new products.
- Fourth, the nature of university-industry interactions has frequently been informal, based on personal connections between individuals, which makes it difficult to find appropriate methods for measuring this kind of impact (**Molas-Gallart** *et al.*, 2002).

2.3. Case studies

> Case studies in which societal impact of certain pieces of research is explained seem to be appropriate resources for capturing and demonstrating the societal impact of research

Case studies provide examples of how research has impacted different aspects of (specific) end users' lives. These studies are narratives and success stories that universities and research institutes provide in order to demonstrate their societal impact. Case studies are used either for prospective societal impact realizations of public research investments, or for conducting retrospective assessments. Every case study is more or less unique in terms of its explanation of impact on society. In the UK, case studies are brief documents

> "assessed by two criteria:
>
> (1) Reach – 'the spread or breadth of influence or effect on the relevant constituencies'; and
> (2) Significance – 'the intensity or the influence or effect'" (**Van-den-Akker**; **Spaapen**, 2017, p. 21).

One advantage of using case studies for measuring societal impact is that –similar to editorial peer-review– expert judgement is used to determine research impact: experts assess the societal impact dimension based on case studies. *REF* has taken a new approach in this regard, going beyond experts, and employing external users of research as part of impact assessment processes. External users provide their perspective regarding the impacts claimed by academics and experts. Case studies use a variety of indicators to accompany the narratives demonstrating the impact of research (**Wilsdon** *et al.*, 2015). For instance, **Wilsdon** *et al.* (2015) noted that the authors of almost 7,000 case studies that were submitted to *REF 2014* had used a variety of indicators (with little consistency) for showing research impact. The missing consensus in indicator use might reflect the broad spectrum of research for which case studies have been written in the past (**Wilsdon** *et al.*, 2015).

Marcella, **Lockerbie**, **Bloice**, **Hood**, and **Barton** (2018) have overviewed some disadvantages of case studies. Case studies "rely on expert advisory panels to review qualitative impact statements" (**Searles** *et al.*, 2016, p. 4). However, similar to interviews and self-reports, they have their own biases such as exaggerating the impacts of research (**Searles** *et al.*, 2016). Another critique of case studies is that expert panels might fail to recognize the "real" impact of research projects that are not presented well in the narratives, and they may focus on the impact story that is well presented. In other words, the practical importance of the underlying research for society may take a back seat. Furthermore, the scores given to case studies might be influenced by factors which should not impact those scores (**Marcella** *et al.*, 2018). For instance, case studies of established senior researchers with external funding might be scored higher than those of other researchers (**Kellard**; **Śliwa**, 2016).

An important problem with case studies is that preparing them is expensive and requires a lot of resources (*Department for Business, Energy & Industrial Strategy*, 2016). For example, the

> "*University College London* alone wrote 300 case studies that took around 15 person-years of work, and hired four full-time staff members to help, says David Price, the university's vice-provost for research" (**Van-Noorden**, 2015, p. 150).

The semi-structured interviews by **Marcella** *et al.* (2018, p. 613) showed that writing case studies have been

> "time consuming, frustrating and iterative 'backwards and forwards, backwards and forwards'."

These efforts might decrease in the future, because universities and research institutes have taken actions to continuously capture the information required for case studies (*Department for Business, Energy & Industrial Strategy*, 2016). The fact remains, however, that great effort is required to write case studies.

What has research on case studies shown? The report by *King's College London* and *Digital Science* (2015) describes the results of an analysis on about 7,000 impact case studies submitted to the *REF* 2014. An example of such case studies is

> "research showing the importance of same-day diagnostic tests for tuberculosis led to improvements in access to care and reductions in costs incurred by patients in Malawi, Nigeria, Yemen, Ethiopia, Nepal and elsewhere" (*King's College London* and *Digital Science*, 2015, p. 12).

Results of the analysis of impact case studies revealed 3,709 unique research impacts; the UK research seems to have made a wide range of contributions to society (*King's College London* and *Digital Science*, 2015). **Terama**, **Smallman**, **Lock**, and **Johnson** (2017, p. 14) clustered the UK impact reports and found six general forms:

> "influence on education; public engagement; environmental technologies and solutions; enterprise; policy impact; and clinical applications."

Thus, case studies might result in very different reports on research, which makes their comparison difficult. In order to improve the future interpretation of case studies, it is important that case studies be written in a consistent manner (*King's College London and Digital Science*, 2015).

This recommendation is in line with **Heyeres**, **Tsey**, **Yang**, **Yan**, and **Jiang** (2019, p. 10), who emphasized

> "the need for more consistency in reporting through a case study approach, more systematic reporting of translation pathways and greater transparency concerning estimated costs and benefits of the research and its translation and impact assessment."

The authors analysed the characteristics of 25 impact case studies in order to identify their translation activities and quality of reporting. Results indicated that

> "24 papers reported intermediate impacts, such as advocacy, or the development of statements, tools, or technology. 4 reported on longer-term societal impacts, such as health outcomes and economic return on investment. 7 reported on translation activities" (**Heyeres** *et al.*, 2019, p. 10).

The reporting areas that obtained the weakest scores were

> "identification of stakeholder needs and stakeholder involvement, and ethics and conflict of interest" (**Heyeres** *et al.*, 2019, p. 10).

2.4. Societal impact measurements at funding organizations

Societal impact measurements are not only relevant after carrying out research, but also in the process of receiving research funds. The UK government-funded granting agencies, the *Swiss National Science Foundation* (*SNSF*), the *US National Science Foundation* (*NSF*) and other funding organizations require researchers to estimate the impact of proposed research in grant proposals (**Dance**, 2013). However, funding organizations ask for different kinds of impact to be included in grant proposals. For instance, the *National Institutes of Health* (*NIH*) cares more about the impact of research in one specific field rather than broad impact. Other organizations are interested in the economic implications of research. Judging the societal impact of research is something that scientific communities might be reluctant to do. A survey study in this regard, which investigated 428 people from funding agencies, suggested that such resistance in the scientific communities is possibly due to their unwillingness to conduct such activities, rather than a lack of confidence in their ability to judge the broader societal impacts (**Holbrook**; **Hrotic**, 2013).

Langfeldt and **Scordato** (2015) overviewed the impact measurements at funding agencies. They showed that funding agencies have different approaches to the broader impact measurement of research. For example, the *Research Council of Norway* rates impact criteria separately, *NSF* gives an overall rating, while at the *Natural Environment Research Council* (*NERC*) impact is specifically commented. The *EU Framework Program/Horizon 2020* uses an impact score threshold for research funding and gives weight to impact in the final decision. One criticism of impact measurements at funding organizations is that they do not let laypeople and potential users outside academia get involved in evaluating research proposals. Some organizations do get users involved (e.g. in project selection), but not in the assessment of the content of research (**Langfeldt**; **Scordato**, 2015).

2.5. What are the societal problems that should be solved by science?

Societal impact measurements should focus on societal needs. The measurements should be intended to show whether societal needs have or have not been (successfully) targeted by research efforts. Although it might be obvious that societal impact measurements start with societal needs, these measurements usually begin with the research that has been conducted at a research institution. In societal impact measurements, attempts are then made to capture the possible impact (e.g. using altmetrics, which are discussed in section 3). This process might not, however, be efficient in capturing and demonstrating pressing societal needs. The wide range of societal areas to which UK research can be assigned (see above), may be interpreted as a sign that the research is not (always) oriented towards pressing societal needs. In recent years, frameworks have been published that attempt to reveal these needs. These frameworks can be used to suggest which societal needs can (should) be addressed by science. **Hicks**, **Stahmer**, and **Smith** (2018) proposed a conceptual framework, based on the **Nussbaum** (2001, p. 78ff; 2009, p. 76ff) list of central capabilities, which can be used for categorizing (and focusing on) the societal goals of research. The list contains 10 items including "life", "bodily health", "bodily integrity", "sense, imagination and thoughts", "emotions", "practical reason", "affiliation", "other species", and "control over the environment."

Hicks *et al.* (2018) noted that many research activities are only useful for a particular scientific community. This type of research is known to have inward-facing goals. On the other hand, outward-facing goals of research focus on specific societal needs, such as solving agricultural issues. To evaluate the inward-facing value of academic research, traditional metrics such as citation counts can be applied. Although methods exist to evaluate outward-facing values of research (see above), they are not applicable in all disciplines, and it is not clear whether they measure what they intended to measure. The framework of **Hicks** *et al.* (2018, p. 5)

> "suggests that, because researchers typically have both inward- and outward-facing goals, both inward- and outward-facing metrics should be used. The list of central capabilities can help researchers communicate the goals of their research and identify areas where bibliometricians and evaluators need to develop new metrics."

Pollitt *et al.* (2016) proposed a similar list to **Hicks** *et al.* (2018), however, focusing on potential contributions of health research, outlining researchers' and the general public's preferences for different types of health research impact. For instance, the list included the suggestion that research should contribute to better care of patients. **Pollitt** *et al.* (2016, p. 1) indicated that

> "the general public and researchers provided similar valuations for research impacts such as improved life expectancy, job creation and reduced health costs, but there was less agreement between the groups on other impacts, including commercial capacity development, training and dissemination."

Although it is possible nowadays to acknowledge and measure the value of selected works of scientists who laid the groundwork for well-known technologies such as *Google* or certain communication technologies, it will not be possible for every research to foresee its usefulness for specific needs in the lists of **Hicks** *et al.* (2018) and **Pollitt** *et al.* (2016). The question is whether or not metrics (such as altmetrics) can help us measure the success and benefits of research projects (*National Research Council*, 2014). The conceptual frameworks proposed by **Hicks** *et al.* (2018) –based on the **Nussbaum** (2001, 2009) list of central capabilities– and **Pollitt** *et al.* (2016) might to some extent support the process of identifying the projects that would be beneficial to the society. High societal impact scores should point to research addressing listed needs by solving pressing societal problems.

A good example for a framework in which societal needs have been formulated, funds for reaching the goals provided, and research and indicators for evaluation proposed, are the *Millennium Development Goals* formulated by the *United Nations* in 2000 (see *https://www.un.org/millenniumgoals*).

Eight international development goals were established for the year 2015 (e.g. to eradicate poverty and hunger and to reduce child mortality). Detailed reports have been published which demonstrate –based on various sets of indicators– which goals have been reached, and which challenges are still faced (**Way**, 2015).

3. Altmetrics

Societal impact considerations in research evaluation can be seen as a scientific revolution in scientometrics, which have led to a broadening of the impact term (**Bornmann**, 2016; **Derrick**; **Samuel**, 2016): current impact measurements frequently include the whole society and are not restricted to science only. As the overview in section 2 reveals, no standard method has been established in scientometrics to measure societal impact; many methods, indicators, approaches, etc. have been proposed and used. In addition, frequently used methods have been targeted by fundamental critiques: for example, the case studies approach has the disadvantage that comparisons between institutions are scarcely possible and results are not generalizable. Against the backdrop of this situation in societal impact measurements, the so called "alternative metrics" or "altmetrics" came into play (**Priem**; **Taraborelli**; **Groth**; **Neylon**, 2010).

Altmetrics include new ways of measuring the impact of research (**Adie**, 2014) that consider non-traditional metrics such as number of page views, downloads, recommendations, shares on social media, social bookmarks, comments, ratings, and tweets (**Liu**; **Xu**; **Wu**; **Chen**; **Guo**, 2013; **Wilsdon** *et al.*, 2015). Altmetrics are expected to provide new insights into the impact of research, mainly the online impact of research (publications) via its appearance on *Facebook*, *Twitter*, blogs, and other web-based platforms (**Zahedi**; **Costas**; **Wouters**, 2014). The use of altmetrics assumes that impact can be measured beyond citation counts by demonstrating the overall usage of research (publications) on the web and social media platforms. Altmetrics seem to make broader interpretations of impact possible (**Waltman**; **Costas**, 2014), for instance by showing how many times an article has been downloaded, reflecting its usage (**Haustein**, 2014).

Many sources of altmetrics (e.g. *Twitter*) emerge faster than citations, often right after the research is published, and often come from open sources. Most sources seem to reveal other aspects of impact that differ from citation counts (**Robinson-García**; **Torres-Salinas**; **Zahedi**; **Costas**, 2014). According to **Konkiel**, **Madjarevic**, and **Lightfoot** (2016, p. 16),

> "citations are a useful indication of traditional scholarly influence, whereas altmetrics can tell us about public influence and non-traditional scholarly influence, which can occasionally predict later citations. You need both kinds of metrics to get the full picture of research's value."

The importance of altmetrics today is reflected by the fact that many publishers of scientific journals, such as *Elsevier*, *Springer*, *Wiley*, *BioMed Central*, *Nature* and *PLoS* report altmetrics for published papers (**Thelwall**; **Kousha**, 2015; **Wilsdon** *et al.*, 2015).

Several providers of altmetrics data exist which collect and process the data from social media such as *Altmetric* (see *https://www.altmetric.com*), *Plum Analytics* (see *http://plumanalytics.com*), and *Impact Story* (see *https://profiles.impactstory.org*) (**Sugimoto**; **Work**; **Larivière**; **Haustein**, 2017).

There are, however, inconsistencies and variations in altmetrics data received from these sources which make comparing datasets problematic and challenging (**Sugimoto** *et al.*, 2017). **Zahedi**, **Fenner**, and **Costas** (2014) and **Zahedi** and **Costas** (2018) found variations in the social media metrics for the same publications across different altmetrics sources.

Haustein, **Costas**, and **Larivière** (2015, p. 6) argued that

> "disciplines that have stronger ties to society (e.g. social sciences and humanities) or deal with specific concerns of people's everyday lives (health or environmental problems; i.e., biomedical and health sciences and life and earth sciences) have a higher probability of appearing on social media platforms than publications from more technical and applied disciplines (e.g. Natural Sciences and Engineering) or with a higher technical and complexity component (e.g. Mathematics and Computer Science)."

Wouters, **Zahedi**, and **Costas** (2019) classified available social media platforms used for impact measurements into the following categories (similar classifications have been proposed by **Moed**, 2017):

(1) Social bookmarking tools: Reference managers (e.g. *Mendeley*, see *http://www.mendeley.com*) can be used to count the number of times an item has been bookmarked or saved to a library (readership count) (**Wouters** *et al.*, 2019).

(2) Microblogging tools: *Twitter* is the main microblogging tool with the possibility of sharing and discussing scholarly documents online. *Twitter* indicators for altmetrics studies are the number of tweets and retweets, likes, and replies (**Wouters** *et al.*, 2019).

(3) Blogs: Blogs are a platform for disseminating scientific materials, often maintained by scientists or science journalists. Blog mentions (e.g. the mentioning of a researcher) and blog citations (of publications) are the two main data that can be obtained from blogs (**Wouters** *et al.*, 2019).

(4) Wikis: Wikis such as *Wikipedia* (an online encyclopaedia) are platforms for sharing information that can be edited by the general public. Citations of publications in wikis is a typical metric that can be obtained from *Wikipedia* (**Wouters** *et al.*, 2019).

These social media platforms cover a different amount of scientific papers. For example, the study by **Mas-Bleda** and **Thelwall** (2016) showed that among the altmetrics sources, *Mendeley* had the highest coverage of papers (80%), followed by *Twitter* (34%). Papers had received negligible mentions in other sources such as *Wikipedia* (2.6 %) and course syllabi (1.2 %) (**Mas-Bleda**; **Thelwall**, 2016).

After having presented some general information on altmetrics, we review the literature on specific altmetrics in the following sections.

3.1. Blogs

Blogs are one of the oldest social media platforms.

> "Blogs are websites that allow individuals to create personal webpages of text, pictures, graphics, videos, and other multimedia with the same ease as creating a word processing document. Unlike traditional websites, however, they provide a space where people can post comments and engage in online conversations" (**Boling**; **Castek**; **Zawilinski**; **Barton**; **Nierlich**, 2008, p. 504).

Some studies have noted that blog usage has declined over recent years, specifically among academics (**Fausto** *et al.*, 2012; **Shema**; **Bar-Ilan**; **Thelwall**, 2015). Blogs are used by a variety of groups such as journalists and activists for different purposes such reading about the most recent events (**Bonetta**, 2007) and seeking information (**Puschmann**; **Mahrt**, 2012). Bloggers might play an important role in the dissemination of scientific knowledge by simplifying and explaining complex scientific results to the general public (**Fausto** *et al.*, 2012). Bloggers are frequently scientists or people working in academia (**Shema**; **Bar-Ilan**; **Thelwall**, 2014) as evidenced in previous studies such as **Shema**, **Bar-Ilan**, and **Thelwall** (2012), who found that the majority of science bloggers (59%) were either students or researchers. **Puschmann** and **Mahrt** (2012) showed that 45% of the bloggers active on the German platform *scilogs.de* had a PhD degree. *SciLogs* is a website run by *Spektrum der Wissenschaft* which hosts over 60 scholarly blogs in total (**Puschmann**; **Mahrt**, 2012).

The efforts of scientific bloggers, besides eliminating the obstacles of gaining access to scientific knowledge (**Fausto** *et al.*, 2012), lead to another public service, i.e. giving advice and recommendations about public health and social issues and distributing rich information for practical use by the public (**Shema**, **Bar-Ilan**; **Thelwall**, 2015). Expert bloggers such as medical bloggers and skilled professionals use blogs to write about and discuss health issues. Medical bloggers are faithful to their audiences, and thus use valid resources to share information. One major motivation for medical blogging, besides providing people with health information, is to direct people's attention toward certain health issues (**Kovic**; **Lulic**; **Brumini**, 2008):

> "Medical blogs are frequently picked up by mainstream media; thus, blogs are an important vehicle to influence medical and health policy" (**Kovic** *et al.*, 2008, p. 1).

Most papers mentioned in blogs seem to be from prestigious journals (**Groth**; **Gurney**, 2010). **Fausto** *et al.* (2012) indicated that prestigious journals, such as *Science*, *Nature*, and *Proceedings of the National Academy of Sciences of the United States of America* (*PNAS*), were extensively covered in blogs in 2009 and 2010. Other studies have come up with similar findings. For instance, **Shema**, **Bar-Ilan**, and **Thelwall** (2015) showed that in 2009 21% of *Psychological Science* articles were mentioned in blogs followed by *Science* (18%), *Nature* (14%), and *PLoS Biology* (13%). In 2010, 31% of the papers published in *New England Journal of Medicine* were mentioned in blogs followed by *Psychological Science* (25%), *Nature* (23%), and *Science* (20%).

Shema, **Bar-Ilan**, and **Thelwall** (2014) investigated the correlation between being mentioned in blogs and being cited in papers later on. They found that blogged papers received more citations than other papers from the same journal. They noted that one possible reason is that bloggers opt to write about better papers which might receive more citations in the future, or because papers that are discussed in blogs receive more attention among the academic community which consequently results in obtaining more citations (**Shema**; **Bar-Ilan**; **Thelwall**, 2014).

3.2. *Mendeley*

Mendeley is both a tool for reference management and sharing research papers which functions as a scientific network platform in which scholars can connect to members with whom they share similar interests (**Rodgers**; **Barbrow**, 2013). An advantage of publication bookmarks in reference management as data source is that

> "as with citations and downloads, usage data are generated as a by-product of existing workflows. Unlike tweeting, for example, searching for and managing literature is an established part of the scholarly communication process" (**Haustein**, 2014, p. 335).

Mendeley has a great coverage of scientific fields and a large number of users and readers (**Hammarfelt**, 2014; **Haustein**, 2014; **Mas-Bleda**; **Thelwall**, 2016) –compared to other reference management software (e.g. *Zotero*). **Zahedi**, **Costas**, and **Wouters** (2014) reported that *Mendeley* had the highest coverage of articles (63% saved at least once) compared to other social medial platforms such as *Twitter* (13%), followed by *Facebook* (2.5%) and blogs (1.9%). **Priem**, **Piwowar**, and **Hemminger** (2012) found that 80% of *PLoS One* articles were bookmarked in *Mendeley*, followed by *Facebook* and *Twitter*.

It seems that *Mendeley* measures impact which deviates (moderately) from that which is measured by citations. According to **Mohammadi** and **Thelwall** (2014, p. 1627)

> "low and medium correlations between *Mendeley* bookmarks and citation counts in all the investigated disciplines suggest that these measures reflect different aspects of research impact."

In another study, **Mohammadi**, **Thelwall**, **Haustein**, and **Larivière** (2015, p. 1832) noted that *Mendeley* statistics might reveal

> "the hidden impact of some research articles, such as educational value for non-author users inside academia or the impact of research articles on practice for readers outside academia."

However, **Bar-Ilan** (2014) found fluctuations in the number of *Mendeley* readers for various publications to be a potential disadvantage of this data source. Her finding indicated that the number of readers for selected publications decreased over time. One possible reason is that *Mendeley* users have updated their publication lists, having removed publications already stored in the library (**Thelwall**, 2017b).

Bornmann and **Haunschild** (2015) investigated the interest of the various readership groups in certain scholarly papers. Their study was based on papers included in *F1000Prime* –a post-publication peer review system. In other words, they examined which *Mendeley* groups used or saved which type of documents in *F1000Prime*. Scholarly papers are tagged in *F1000Prime* by experts, as "good for teaching", "confirmation of previous results", "interesting hypothesis", etc. Almost all user groups (mainly PhDs, postdocs, professors, librarians, etc.) showed less interest in documents with the "confirmation" and "interesting hypothesis" tags. Almost all user groups (mainly postdocs, PhDs, and professors) were more interested in papers with the "new finding" tag than other papers. Papers with the "technical advance" tag were also popular among all user groups. **Bornmann** and **Haunschild** (2015) noted that the papers with the "good for teaching" tag were of interest to the people (lecturers and researchers outside academia) with less engagement with academic research.

In another study, **Bornmann** and **Haunschild** (2016c) proposed the use of overlay maps to demonstrate the impact of publications in terms of readership data. They visualized *Mendeley* reader counts of the 2012 publications from the *WoS*.

Their overlay maps can be used to see the disciplines in which publications of an institution have been read, and to what extent. They showed that, for example, for *TU Munich*

> "many papers have been read in the miscellaneous categories of biology, medicine, and chemistry and also the readership per paper is high in these categories compared to other subdisciplines" (**Bornmann**; **Haunschild**, 2016c, p. 3069).

Bornmann and **Haunschild** (2016c) also created overlay maps for some journals based on *Mendeley* reader count data. The map for the *Journal of the American Society for Information Science and Technology* (*JASIST*) indicated the scope of the articles published in this journal in 2012. As expected, library and information science have been a major focus of the articles published in *JASIST*.

3.3. *Wikipedia*

Wikipedia is an open access encyclopaedia that can be accessed and edited by anyone. Mentions of publications in *Wikipedia* are used as an altmetrics source. Since *Wikipedia* content can be manipulated by anyone, using the number of publication mentions in *Wikipedia* in research evaluation should be used with cautions (**Serrano-López**; **Ingwersen**; **Sanz-Casado**, 2017). Studies have indicated that the coverage of articles on *Wikipedia* (similar to blogs) is lower than their coverage on other platforms such as *Mendeley* and *Twitter* (**Priem** *et al.*, 2012). For instance, **Mas-Bleda** and **Thelwall** (2016) indicated that articles were rarely cited on *Wikipedia* (2.6%).

3.4. Policy documents

Daugbjerg *et al.* (2009, p. 806) defined policy documents

> "as written documents that contain strategies and priorities, define goals and objectives, and are issued by a part of the public administration."

Policy documents mentions (citations of publications in policy documents) are one of the most valuable altmetrics sources that can be used in target-oriented (the target is policies on different areas such as climate change or health) impact measurement. The references in policy documents can be analysed to determine which publications and to what extent have been cited in policy documents (**Bornmann**; **Haunschild**; **Marx**, 2016). For instance, **Bornmann** *et al.* (2016) showed that among the 191,276 climate change publications only 1.2% had been cited at least once in policy documents. This result shows that these 1.2% of publications have (possibly) had some impact on climate change policies. In other words, the majority of publications (98.8%) on climate change have had no impact on climate change policies. The authors discussed various reasons for this result. **Haunschild** and **Bornmann** (2017) also indicated that a low percentage (0.5%) of *WoS* publications in different subject categories were cited in policy documents.

Tattersall and **Carroll** (2018) analysed citations in policy documents and found that 1,463 papers from authors at the *University of Sheffield* (0.65% of *Sheffield* research) were cited by at least one policy document, most of which were journal articles (99%). The publications were cited in policy documents three months to 31 years after their appearance. The authors indicated that some topics were cited more frequently in policy documents such as medicine and dentistry than other topics such as social science and pure science. **Tattersall** and **Carroll** (2018) suggested several data problems with policy documents. For instance, they noted that it is not entirely clear how the company *Altmetric* identifies "policy documents", and that some of these documents might be journal articles and not policy documents (**Tattersall**; **Carroll**, 2018).

Vilkins and **Grant** (2017) investigated which papers were cited in policy documents. They analysed 4,649 references cited in 80 Australian policymakers' publications from 2010 to 2017. The most cited references in policymaking publications were journal articles (n = 1836), federal government reports (n = 1106), and Australian business information (n = 373). This is in contrast with previous studies which showed that policymakers were less willing to cite scholarly publications (**Vilkins**; **Grant**, 2017). **Vilkins** and **Grant** (2017) noted that references in policy documents can be used in research evaluation in order to obtain a better understanding of research utilization and impact in society.

Haunschild and **Bornmann** (2017) found that more than 100 sources of policy documents (e.g. the *European Food Safety Authority* and the *UK Government*) were tracked by *Altmetric*. There are, however, many other sources that are not tracked by *Altmetric* (**Bornmann** *et al.*, 2016). **Bornmann** *et al.* (2016) suggested that *Altmetric* should specify the types of policy making organizations and the name of organizations in each type. It is also mentioned that since *Altmetric* does not include all (or the majority of) policy-related sources, mentions in policy documents should be used with cautions for impact measurement (**Haunschild**; **Bornmann**, 2017).

3.5. *Twitter*

Twitter is a social networking site via which millions of people (including academics) are connected to each other worldwide. People can post 280-character tweets about different topics, respond to tweets, and retweet them. People can also post photos, videos, and links to other websites on *Twitter*. Academics and researchers (and the public) use *Twitter* to share scholarly contents (e.g. papers) and events (e.g. conferences) (**Mohammadi**; **Thelwall**; **Kwasny**; **Holmes**, 2018).

Twitter has been one of the most popular sources for altmetrics studies so far: many studies have been published based on tweets. Similar to *Mendeley*, *Twitter* seems to be the main source of sharing scientific papers. For instance, the study by **Haustein** *et al.* (2015) showed that the coverage of articles on *Twitter* was 21.5%, followed by other social media platforms such as *Facebook* (4.7%) and blogs (1.9%).

Robinson-García, **Costas**, **Isett**, **Melkers**, and **Hicks** (2017, p. 1) examined tweets linked to research papers from 47 dentistry journals

> "to assess the extent to which tweeting about scientific papers signifies engagement with, attention to, or consumption of scientific literature."

Three patterns were found in the top 10 most tweeted papers:

> "single issue tweeters, professional social media account management and broader tweeting, though with little original content. None of the patterns evidenced engagement with the journal article" (**Robinson-García** *et al.*, 2017, p. 4).

Only a few tweets reflected interest for the content of a paper. **Robinson-García** *et al.* (2017) suggested that the number of tweets about papers has little value for research evaluation. 74% of articles were automatically tweeted, possibly by bots mechanically retweeting, or humans who behave like bots, which means they tweet/retweet whatever they see without paying attention to the content of the tweet (**Robinson-García** *et al.*, 2017).

It has been mentioned that *Twitter* might not reflect the general public impact, because most tweets regarding scientific papers are often posted by researchers (**Sugimoto** *et al.*, 2017). Other studies have pointed out, however, that *Twitter* is widely popular outside academia and

> "thus seems to be a particularly promising source of evidence of public interest in science" (**Haustein**; **Larivière**; **Thelwall**; **Amyot**; **Peters**, 2014, p. 208).

Another study also suggests that most of the tweets are distributed by the public user (**Yu**, 2017). **Mohammadi** *et al.* (2018) showed that about 45% of tweets were posted by people who didn't work in academia. They investigated the demographic information and background of 1,912 users that had tweeted scholarly papers.

In the study by **Mohammadi** *et al.* (2018), many people had a social science or humanities background, and used *Twitter* to obtain and share information and develop professional networks. The study provides evidence that

> "*Twitter* plays a significant role in the discovery of scholarly information and cross-disciplinary knowledge spreading. Most importantly, the large numbers of non-academic users support the claims of those using tweet counts as evidence for the non-academic impacts of scholarly research" (**Mohammadi** *et al.*, 2018, p. 1).

3.6. Field-normalization

Altmetrics patterns –independently of the data source (e.g. *Twitter*, *Mendeley* or blogs)– might systematically vary depending on the publication time and discipline of the corresponding papers (**Taylor**, 2013a). Results show that general disciplines and topics related to a broader audience attract more social media attention than specialized disciplines such as physical sciences (**Haustein** *et al.*, 2014; **Ortega**, 2018). **Fausto** *et al.* (2012) investigated 19,000 blog posts that had cited 26,154 publications. The results indicated that 36% of the blog posts were in biology, followed by health (15%). Similarly, **Shema**, **Bar-Ilan**, and **Thelwall** (2014) showed that bloggers tended to post about the biological and medical disciplines in 2009 (67%) and 2010 (74%). **Mohammadi** and **Thelwall** (2014) indicated differences in *Mendeley* readership for disciplines: social sciences & humanities publications had the greatest number of readers among all disciplines, while mathematics & computer science had the lowest number of readers.

Based on these and similar findings, much evidence can be found in the altmetrics literature that points to the necessity of time- and field-normalizing altmetrics data, similar to citation data in bibliometrics. **Taylor** (2013b) recommended that field-specific differences of altmetrics should be taken into consideration. According to **Thelwall** (2017b, p. 10)

> "most indicators should not be compared between fields because of disciplinary differences. Most indicators should not be compared between years because of time differences."

Normalization can be achieved by benchmarking altmetrics

> "based on other articles within the same journal and from the same time period, as well as across the whole database" (**Liu**; **Adie**, 2013, p. 33).

Several researchers have attempted to propose the normalization of altmetrics data (**Bornmann**; **Haunschild**, 2016a, 2016b, 2017; **Haunschild**; **Bornmann**, 2016; **Noyons**, 2018; **Thelwall**, 2017a). We present some approaches in the following:

Bornmann and **Haunschild** (2016a) proposed a percentile-based indicator for normalizing *Twitter* counts for papers published in different disciplines and publication years. The authors mentioned that normalizing *Twitter* counts is essential, since papers from different disciplines receive a different number of tweets (see above). To normalize tweets, **Bornmann** and **Haunschild** (2016a) calculated percentiles based on tweet counts for a paper using a corresponding reference set. For example, a *Twitter* percentile of 90 for a focal paper means that 90% of the papers in the corresponding journal (or subject category) have received fewer tweets than the focal paper. *Twitter* percentiles can be used

> "for comparisons between units in science (researchers, research groups, institutions, or countries) which have published in different fields" (**Bornmann**; **Haunschild**, 2016a, p. 1410).

Bornmann and **Haunschild** (2016a) calculated *Twitter* percentiles for countries and showed that the most tweeted papers belonged to Denmark, Finland, and Norway.

Field- and time- normalization of *Mendeley* reader counts has been introduced by **Bornmann** and **Haunschild** (2016b, 2017) and **Haunschild** and **Bornmann** (2016). **Haunschild** and **Bornmann** (2016) proposed the Mean Normalized Reader Score (MNRS) which is based on the Mean Normalized Citation Score (MNCS) in bibliometrics (**Bornmann**, in press): the impact of a focal paper is divided by the mean impact of all papers in the corresponding journal (or subject category).

> "The MNRS enables us to compare the impact a paper has had on *Mendeley* across subject categories and publication years" (**Haunschild**; **Bornmann**, 2016, p. 62).

Haunschild and **Bornmann** (2016) found a high correlation (r = 0.70) between MNRS and MNCS for the studied journals. With the mean discipline normalized reader score (MDNRS), **Bornmann** and **Haunschild** (2016b) proposed a variant of the MNRS which normalizes impact on the receiving impact side (instead of the cited paper side). **Bornmann** and **Haunschild** (2016b, p. 776) tested the variants and came to the following conclusions:

> "(i) normalization of *Mendeley* reader counts is necessary,
>
> (ii) the MDNRS is able to normalize *Mendeley* reader counts in several disciplines, and
>
> (iii) the MNRS is able to normalize *Mendeley* reader counts in all disciplines."

Bornmann and **Haunschild** (2017, p. 230) extended MNRS to a target-oriented, field-normalized impact indicator which can be used

> "to measure the impact of scientific papers on certain groups –controlling for the field in which the papers have been published and their publication year."

For example, the extended MNRS can measure *Mendeley* reader impact of research papers on students, professors or journalists. The results of **Zahedi** and **Van-Eck** (2018) showed that the most active users in the social sciences & humanities are professors, students, and librarians, who are mainly interested in topics that are educational and theoretical. Researchers and professionals are the most active users in highly cited fields such as biomedical & health sciences, and tend to read more about practical, methodological, and technical topics. Thus, the results of **Zahedi** and **Van-Eck** (2018) demonstrated the necessity to normalize *Mendeley* data on the user-group level.

The most recent field- and time normalizing indicator for analysing altmetrics data (e.g. tweets and *Facebook* posts) was proposed by **Bornmann** and **Haunschild** (2018) and **Haunschild** and **Bornmann** (2018). This indicator is called the Mantel-Haenszel quotient (MHq).

> "The MHq is based on the MH analysis –an established method in statistics for the comparison of proportions" (**Bornmann**; **Haunschild**, 2018, p. 998).

MHq compares proportions of papers shared on social media (e.g. the number of times a paper has been shared on *Twitter* or *Facebook*) with proportions of non-mentioned papers (not shared on social media) (**Bornmann**; **Haunschild**, 2018). **Bornmann** and **Haunschild** (2018) examined the convergent validity of the indicator by studying whether MHq is able to distinguish between different quality levels (defined by the assessments by peers, i.e. *F1000Prime* recommendations, see above). If MHq gives us scores that are to a great extent similar to that of human judgements (assessments by peers), the indicator seems to measure what it is intended to measure. **Bornmann** and **Haunschild** (2018) demonstrated that MHq was able to distinguish between different quality levels. However, they noted that

> "the relationship between altmetrics (*Wikipedia*, *Facebook*, blogs, and news data) and assessments by peers is not as strong as the relationship between citations and assessments by peers. Actually, the relationship between citations and peer assessments is about two to three times stronger than the association between altmetrics and assessments by peers" (**Bornmann**; **Haunschild**, 2018, p. 998).

4. The meaning of altmetrics: do they measure societal impact?

Research can have both inward-facing and outward-facing objectives. Inward-facing goals of research include the benefits of research for the scientific community, such as the production of new methodologies, techniques or conceptual models. Outward-facing goals refer to the benefit of research for society outside the scientific community (**Hicks** *et al.*, 2018). In recent decades, there has been an increasing demand for outward-facing objectives, which is also reflected in popular concepts such as Mode 2 knowledge production or the triple helix, a model of university-industry-government relations (**De-Jong** *et al.*, 2011; **Etzkowitz**; **Leydesdorff**, 2000). There has been increasing pressure on universities that

> "research must cut across scientific disciplines and theories and focus on problem-solving and practical use-value. In brief, the research must be Mode 2" (**Ernø-Kjølhede**; **Hansson**, 2011, p. 132).

Many funding organizations demand that the societal benefits of proposed research should be estimated in grant applications (**De-Jong** *et al.*, 2011).

The trend towards emphasizing the societal relevance of research is part of a bigger movement towards carrying out more applied research that is beneficial to private and public sectors for the purpose of addressing societal issues (e.g. climate change, education, and health care) (*ERiC*, 2010). Researchers should justify the societal benefits of their proposed research project in situations where the private sector is being held responsible for investing in research, regardless of their increasingly scarce financial resources (**Lähteenmäki-Smith**; **Hyytinen**; **Kutinlahti**; **Konttinen**, 2006). However, there has often been a lack of motivation among researchers to get involved in such activities, because they do not expect to receive any reward for their engagement in public outreach activities (e.g. seminars and workshops for stakeholders) (**Kassab**, 2019).

According to the *National Research Council* (2014, p. 70)

> "no high-quality metrics for measuring societal impact currently exist that are adequate for evaluating the impacts of federally funded research on a national scale."

Many metrics fail to provide an accurate measure of the knowledge generated by research and a proper picture of the societal impact of research (*National Research Council*, 2014). For example, most

> "conventional metrics for social impacts focus on economics or wealth creation, such as patents or technology transfer. These kinds of metrics are less appropriate for many scholarly fields, and miss the specific social concerns or needs that researchers aim to address" (**Hicks** *et al.*, 2018, p. 1).

The popularity of qualitative case studies for capturing and demonstrating research impact is a response to the lack of appropriate metrics for measuring societal impact. However, case studies are biased in that they only report success stories, while failure stories learned from research can also be useful, at the very least for future research activities (**Raftery** *et al.*, 2016).

According to **Moed** (2017), the increasing focus of governments and public funders on measuring the societal impact of research is an important driver of the use of altmetrics. Significant attention to altmetrics for societal impact measurement has been evidenced since 2010 (**Blümel**; **Gauch**; **Beng**, 2017). Two main lines of research in altmetrics exist.

- The first line of research investigates the extent to which scholarly publications are shared and used on social media.
- The second type of research deals with the comparison of altmetrics data with traditional metrics such as citations (**Blümel** *et al.*, 2017).

Although many studies have dealt with the question of the meaning of altmetrics (by correlating them with traditional metrics), these studies do not clearly reveal what is being measured by altmetrics. It is

> "not clear what general conclusions can be drawn when an article is frequently mentioned within the social web" (**Barthel**; **Tönnies**; **Köhncke**; **Siehndel**; **Balke**, 2015).

For example, it is not evident what conclusions can be drawn from a publication that has been saved or bookmarked on *Mendeley* libraries (**Zahedi**; **Costas**; **Wouters**, 2014). Altmetrics might measure attention (**Konkiel** *et al.*, 2016; **Moed**, 2017), popularity (**Xia** *et al.*, 2016) or public engagement (**Khazragui**; **Hudson**, 2015) rather than societal impact or academic quality (**Konkiel** *et al.*, 2016).

> Conclusions in research evaluation based on altmetrics data should be drawn with caution

Altmetrics may provide some perspectives about research popularity and the relevance of research to a broader public (**Moed**; **Halevi**, 2015). **Pulido**, **Redondo-Sama**, **Sordé-Martí**, and **Flecha** (2018) indicated that measuring impact based on tweets and *Facebook* posts may show people's vague interest in certain pieces of research. **Pulido** *et al.* (2018) analysed the content of posts on *Twitter* and *Facebook* and showed that the number of posts on *Twitter* and *Facebook* are not good indications of societal impact. For example, no evidence of societal impact was observed for a research project with 403 tweets and 423 *Facebook* posts, but another project with fewer number of tweets (n = 62) and *Facebook* posts (n = 43) had two pieces of evidence of societal impact (**Pulido** *et al.*, 2018).

Buttliere and **Buder** (2017) conducted a study to understand what meaning altmetrics and bibliometrics might indicate (e.g. quality, impact, and attention). They analysed a sample of approximately 33,000 papers from *PLoS* in 2014. Their results indicated

> "that there are at least two important underlying factors, which could generally be described as Scientific Atten-

tion / Discussion (citations), General Attention / Discussion (views, tweets), and potentially Media Attention / Discussion (media mentions). The General Attention metric is correlated about 0.50 with both the Academic and Media factors, though the Academic and Media attention are only correlated with each other below 0.05. The overall best indicator of the dataset was the total lifetime views on the paper, which is also probably the easiest to game. The results indicate the need for funding bodies to decide what they value and how to measure it (e.g., types of attention, quality)" (**Buttliere**; **Buder**, 2017, p. 219).

The difficulties in assigning meanings to altmetrics might point to a missing underlying theoretical framework (**Taylor**, 2013b). For citations, the normative theoretical perspective (**Merton**, 1973) has been proposed on which evaluative bibliometrics is based; however, it is not clear what the theoretical roots of evaluative altmetrics are (**Taylor**, 2013b). An absence of construct validity of altmetrics is a related problem:

> "Construct validity refers to the degree to which a test measure (e.g. a reference count or an *Altmetric* score) measures what it claims or purports to be measuring (e.g. quality or social engagement)" (**Rowlands**, 2018, p. 4).

According to the results of **Bornmann**, **Haunschild**, and **Adams** (2019), altmetrics do not measure what reviewers assess as societal impact. Altmetrics may capture unknown attention rather than societal impact.

Bornmann, **Haunschild**, *et al.* (2019) used the MHq to investigate the convergent validity of altmetrics. The authors studied

> "the convergent validity of altmetrics by using two *REF* datasets: publications submitted as research output (PRO) to the *REF* and publications referenced in case studies (PCS)" (**Bornmann**; **Haunschild**; **Adams**, 2019, p. 325).

Case studies, which are intended to demonstrate societal impact, should cite the most relevant research papers. The results of **Bornmann**, **Haunschild**, and **Adams**, (2019, p. 325) demonstrated

> "that news media as well as mentions on *Facebook*, in blogs, in *Wikipedia*, and in policy-related documents have higher MHq' values for PCS than for PRO. Thus, the altmetrics indicators seem to have convergent validity for these data. In the second part of the analysis, altmetrics have been correlated with *REF* reviewers' average scores on PCS. The negative or close to zero correlations question the convergent validity of altmetrics in that context."

Thus, the results of the study by **Bornmann**, **Haunschild** and **Adams** (2019, p. 325) revealed that altmetrics

> "may capture a different aspect of societal impact (which can be called unknown attention) to that seen by reviewers (who are interested in the causal link between research and action in society)."

Two things might be concluded from the results:

(1) altmetrics measure public (online) discussions, but not the societal impact of research. It seems that other indicators (e.g. patent citations) or qualitative approaches such as case studies are necessary.

(2) Simple counts of data might not be the right way of analysing data from altmetrics sources for receiving meaningful results. Since the meta-data contain many information, other ways of analysing the data should be found.

Some researchers have pointed to the possibility of using network-based approaches for altmetrics (instead of simple counts) to indicate and visualize interactions with scholarly documents among the general public. **Wouters** *et al.* (2019, p. 702) have pointed to the importance of networks for showing

> "the relationships and interactions among the different actors."

The authors noted that hashtag coupling analysis can be used to see how scholarly articles on *Twitter* have been linked to specific and broader hashtags. Two hashtags (e.g. #openaccess and #OA) can make a couple when they are linked to a similar group of scholarly articles (**Wouters** *et al.*, 2019). **Hellsten** and **Leydesdorff** (2020) introduced a network-based approach for analysing *Twitter* data by mapping the co-occurrences of hashtags and *Twitter* users (@usernames). They presented a network with three nodes including authors, actors (*Twitter* users), and topics. This approach has

> "the advantage of making it possible to map which users were addressed in connection with which topics" (**Hellsten**; **Leydesdorff**, 2020, p. 12).

In a follow-up study, **Haunschild**, **Leydesdorff**, **Bornmann**, **Hellsten**, and **Marx** (2019) proposed that the network analysis of *Twitter* data linked to publications can be used to discover public discussions about specific research topics. They used network analysis to compare the hashtags of tweeted publications (on climate change) with the author keywords of all publications. **Haunschild** *et al.* (2019) were mainly interested in understanding whether the topics discussed by people on *Twitter* about publications differed from the topics targeted in publications by authors (researchers). This study found that the shared topic on *Twitter* was about the consequences of climate change for people.

> "*Twitter* users are interested in climate change publications which forecast effects of a changing climate on the environment and to adaptation, mitigation and management issues rather than in the methodology of climate-change research and causes of climate change" (**Haunschild** *et al.*, 2019, p. 695).

Noyons and **Ràfols** (2018) proposed the use of overlay maps in order to demonstrate the societal engagement of research fields. They investigated

> "whether and how mapping bibliometric methods combined with other data sources (e.g., mentions in news and policy) can also be useful for mapping potential societal engagement of research fields" (**Noyons**; **Ràfols**, 2018, p. 1050).

To demonstrate this approach, **Noyons** and **Ràfols** (2018) overlaid the proportion of agriculture research cited in policy documents and news by using the *VOSviewer* software (see *https://www.vosviewer.com*).
Noyons and **Ràfols** (2018, p. 1055) indicated that

> "policy engagement is primarily observed in the social and behavioural sciences and health areas of agriculture, but also in soil and climate related areas."

The results of this study also demonstrated that

> "news interest mainly is focused on (mental) health research and food within agriculture research" (**Noyons**; **Ràfols**, 2018, p. 1055).

Another approach was proposed by **Noyons** (2018) and **Noyons** (2019), called area-based connectedness (ABC). It depends on the network analysis of publications of a journal or a set of articles. This method investigates the connectedness of research to different dimensions (**Noyons**, 2019, p. 11):

- "News (papers being mentioned in news items).
- Policy (papers being mentioned in policy documents).
- Industry R&D (industry authorship).
- Technological or commercial application (papers cited in patents).
- Local scope (papers in local languages, not in English)."

According to this method, in order to measure the connectedness of a specific publication set (journal), the number of publications from the set is counted that exist in each of over 4,000 research areas of science. These 4,000 research areas have been identified by a classification approach developed by the *Centre for Science and Technology Studies* (*CWTS*, *Leiden University*). Using the *VOSviewer* software a map is created for the set (the journal) in order to demonstrate the distribution of its publications across those 4,000 areas. The output of a journal will then be characterized

> "by the connectedness of the areas in each dimension, which means that the connectedness of a journal is the average of the sum of the product of the number of publications in an area and the score on a dimension in an area (weighted average). The final connectedness on each dimension will then be compared to the overall average of connectedness of that dimension" (**Noyons**, 2019, p. 15).

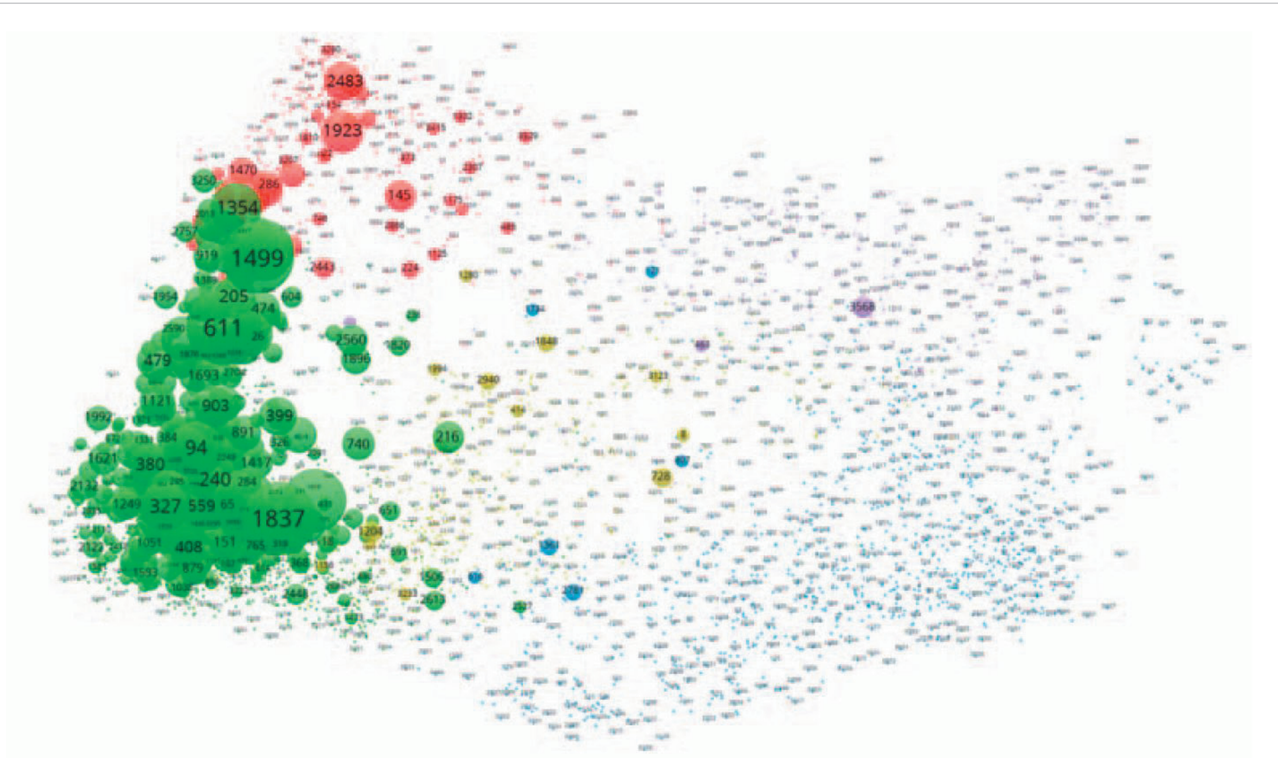

Figure 1. Landscape of science (data: *WoS* 2000-2017). The size of the nodes represents the number of publications in *The Lancet* (2014-2017). Source: **Noyons** (2019, p. 18).

For example, **Noyons** (2019) illustrated the distribution of *The Lancet* publications (n = 2815) across the 4,000 research areas (see Figure 1). In Figure 1, the colours represent four main science fields: Green: biomedical and health sciences; Yellow: life and earth sciences; Purple: mathematics and computer science; Blue: physical sciences and engineering, Red: social sciences and humanities (see *https://www.leidenranking.com/information/fields*)

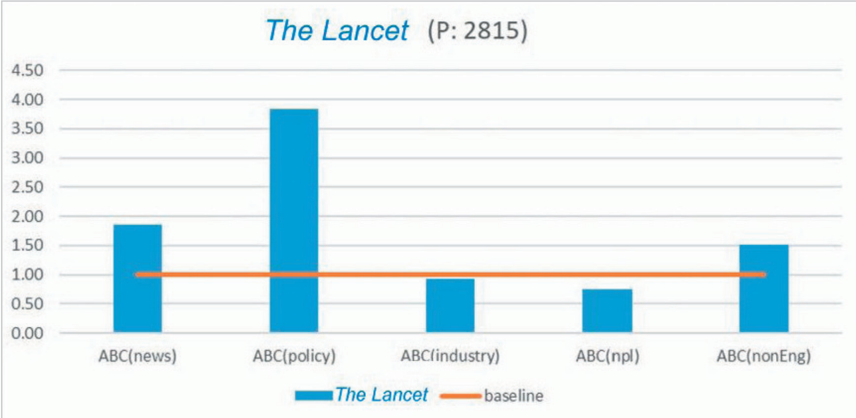

Figure 2. Area-based connectedness to society profile of *The Lancet* (2014-2017). Source: **Noyons** (2019, p. 19).

Noyons (2019) then showed the connectedness of *The Lancet* publications in five dimensions of news, policy, industry, technological and commercial application, and local scope (see Figure 2). Figure 2 shows a high connectedness of *The Lancet* publications on the policy dimension: it is approximately 4 times above the baseline (i.e. average). The news dimension and the local interest dimension are also above the average (**Noyons**, 2019).

5. Conclusions

This review began by presenting various efforts and developments during recent years with regard to societal impact measurements. We outlined that this area of research evaluation faces many problems (e.g. attribution and timeline), which differentiates it from the area of impact measurements on academia: for the latter, standard approaches have been established based on citation data, and new developments are more than welcome, such as indicators that measure the novelty (**Bornmann**; **Tekles**; **Zhang**; **Fred**, 2019) or disruptiveness of research (**Wu**; **Wang**; **Evans**, 2019). Since there is still a lack of standard methodologies, tools, metrics, and (data collection) processes for evaluating the societal impact of research, the use of case studies is the favoured approach in many research evaluation processes. Standard approaches in this area based on quantitative data, which might result in reliable and useful findings, refer to specific applications such as patent citations or citations in clinical guidelines.

Case studies seem to be appropriate resources for capturing and demonstrating the societal impact of research. With case studies, research can be evaluated by panel experts and/or the target groups who are the intended beneficiaries of the research findings. It is an advantage of case studies that a specific outline is given as to the area of society in which research was (or will be) useful. In many quantitative societal impact measurements based on altmetrics data, however, this is not clear: the measurements usually start with past research and a calculation of its broad impact, e.g. based on *Twitter* data. In these measurements, the areas of usefulness of research are scarcely of interest, what counts is broad impact reflected in corresponding data. This review indicates that it should be relevant for societal impact measurements to be oriented towards pressing societal problems. Thus, we propose that societal impact measurements do not follow the premise

> There is still a lack of standard methodologies, tools, metrics, and data collection processes for evaluating the societal impact of research

"it's all right as long as any broad impact (reflected in altmetrics data) is right" (measured by altmetrics)

but that pressing societal problems are used as basis for assessing the usefulness of research. In recent years, many lists have been published with societal needs, which can be used as basis for impact measurements in research evaluation (see section 2.5).

Although in this review we recommend going beyond simple counting of tweets, mentions, etc., in research evaluation, some altmetrics sources, particularly *Mendeley* data, might be useful in this regard. These sources are suitable for calculating indicators based on counts, since it is possible to measure impact on certain status groups (e.g. students and professors). When using these altmetrics counts, however, field- and user-specific differences should be considered (since systematic field- and user-specific differences can be observed in the data). Similar to citations in bibliometrics, altmetrics data such as *Mendeley* reader counts can be field- and user-normalized using standard approaches from bibliometrics. Applying these approaches, it is possible to measure the impact of research in a target-specific way. For example, the field- and time-normalized impact

> Societal impact measurements should be oriented towards pressing societal problems

of papers published by researchers from a university can be measured on students, journalists or professors. The resulting indicator scores show whether the impact is below or above the worldwide average.

> Altmetrics data such as *Mendeley* reader counts should be field- and time-normalized for research evaluation purposes

The identification of the target group is of high importance in societal impact measurement studies. **Bornmann** and **Haunschild** (2017) stated that in bibliometrics, the target group is clearly defined (researchers who publish scholarly works) and thus, bibliometrics is successful in the measurement of research impact. In contrast, the metrics used to measure the societal impact are intended to broadly measure all areas in a society, such as environment, culture, politics, economics, and health. Similar to citations that are target-oriented (citations measure the impact on scholarly works), the target group (the people who benefit from the results of scholarly works) for altmetrics can (should) also be clearly determined and defined (**Bornmann**; **Haunschild**, 2017).

Altmetrics data such as social media data, although already extensively used for research evaluation, might mainly show public interest, attention, and discussion of scholarly works rather than their societal impact. In our view, the literature in recent years makes it clear that altmetrics data in general are not able to reveal impact of research on society which is observable in concrete activities in specific areas. This depth in the interpretation of the usefulness of research cannot be achieved by using altmetrics data in research evaluation. Today, the favoured approach for assessing the usefulness of research for society is the case studies approach. Recent altmetrics research revealed that measuring interest, attention, and discussion does not have to be restricted to simple counts, but can be widened to network (or overlay) approaches showing relationships and dependencies (see **Haunschild** *et al.*, 2019). Since these approaches have been introduced very recently, future research will show whether or not they will be fruitfully for research evaluation purposes.

In our view, scientometrics research on societal impact measurements is still at an early stage (although these measurements are already part of many research evaluation processes). The necessity of research is especially pressing in the altmetrics area. Here, the danger today is that huge pools of data are available that are waiting to be analysed and applied in research evaluation. However, in many cases, the usefulness of the results for assessing countries, institutions or researchers based on the data is questionable. Thus, we would like to follow **Robinson-García** *et al.* (2014, p. 364) who suggested

> Altmetrics data may reflect the public interest, attention, and discussion of scholarly works rather than their societal impact

> "that more research is needed for understanding the methodologies for retrieving valid and reliable altmetrics data. In the same line, the selection of social media sources must be rigorous and critical, attending to its use within the different communities and audiences and avoiding potential discipline or language biases."

6. References

Adie, Euan (2014). "Taking the alternative mainstream". *El profesional de la información*, v. 23, n. 4, pp. 349-351.
https://doi.org/10.3145/epi.2014.jul.01

Bar-Ilan, Judit (2014). "Jasist@ Mendeley revisited". *Altmetrics14: Expanding impacts and metrics, Workshop at Web Science Conference 2014*.
https://doi.org/10.6084/m9.figshare.1031681

Barthel, Simon; **Tönnies, Sascha**; **Köhncke, Benjamin**; **Siehndel, Patrick**; **Balke, Wolf-Tilo** (2015). "What does Twitter measure? Influence of diverse user groups in altmetrics". *ACM/IEEE Joint conf on digital libraries (JCDL)*, Knoxville, TN, USA.
http://www.ifis.cs.tu-bs.de/node/3016

Belter, Christopher W. (2018). "Providing meaningful information: Part B - Bibliometric analysis". In: DeRosa, A. P. (ed.). *A practical guide for informationists*. Sawston, Cambridge, UK: Chandos Publishing, pp. 33-47. ISBN: 978 0 081020173

Blümel, Clemens; **Gauch, Stephan**; **Beng, Florian** (2017). "Altmetrics and its intellectual predecessors: Patterns of argumentation and conceptual development". *Proceedings of the Science, technology, & innovation indicators conference. Open indicators: innovation, participation and actor-based STI indicators*. Paris, France.
https://sti2017.paris/wp-content/uploads/2017/11/his-bluemel-beng-et-al.pdf

Boling, Erica C.; **Castek, Jill**; **Zawilinski, Lisa**; **Barton, Karen**; **Nierlich, Theresa** (2008). "Collaborative literacy: Blogs and Internet projects". *The reading teacher*, v. 61, n. 6, pp. 504-506.
https://doi.org/10.1598/RT.61.6.10

Bonetta, Laura (2007). "Scientists enter the blogosphere". *Cell*, v. 129, n. 3, pp. 443-445.
https://doi.org/10.1016/j.cell.2007.04.032

Bornmann, Lutz (2016). “Scientific revolution in scientometrics: The broadening of impact from citation to societal”. In: Sugimoto, Cassidy R. (ed.). *Theories of informetrics and scholarly communication*. Berlin, Germany: De Gruyter, pp. 347-359. ISBN: 978 3 11 030846 4

Bornmann, Lutz (in press). “Bibliometric indicators – methods for measuring science”. In: Williams, Richard (ed.). *Encyclopedia of research methods*. Thousand Oaks, CA, USA: Sage.

Bornmann, Lutz; **Haunschild, Robin** (2015). “Which people use which scientific papers? An evaluation of data from F1000 and Mendeley”. *Journal of informetrics*, v. 9, n. 3, pp. 477-487.
https://doi.org/10.1016/j.joi.2015.04.001

Bornmann, Lutz; **Haunschild, Robin** (2016a). “How to normalize Twitter counts? A first attempt based on journals in the Twitter Index”. *Scientometrics*, v. 107, n. 3, pp. 1405-1422.
https://doi.org/10.1007/s11192-016-1893-6

Bornmann, Lutz; **Haunschild, Robin** (2016b). “Normalization of Mendeley reader impact on the reader- and paper-side: A comparison of the mean discipline normalized reader score (MDNRS) with the mean normalized reader score (MNRS) and bare reader counts”. *Journal of informetrics*, v. 10, n. 3, pp. 776-788.
https://doi.org/10.1016/j.joi.2016.04.015

Bornmann, Lutz; **Haunschild, Robin** (2016c). “Overlay maps based on Mendeley data: The use of altmetrics for readership networks”. *Journal of the Association for Information Science and Technology*, v. 67, n. 12, pp. 3064-3072.
https://doi.org/10.1002/asi.23569

Bornmann, Lutz; **Haunschild, Robin** (2017). “Measuring field-normalized impact of papers on specific societal groups: An altmetrics study based on Mendeley data”. *Research evaluation*, v. 26, n. 3, pp. 230-241.
https://doi.org/10.1093/reseval/rvx005

Bornmann, Lutz; **Haunschild, Robin** (2018). “Normalization of zero-inflated data: An empirical analysis of a new indicator family and its use with altmetrics data”. *Journal of informetrics*, v. 12, n. 3, pp. 998-1011.
https://doi.org/10.1016/j.joi.2018.01.010

Bornmann, Lutz; **Haunschild, Robin**; **Adams, Jonathan** (2019). “Do altmetrics assess societal impact in a comparable way to case studies? An empirical test of the convergent validity of altmetrics based on data from the UK research excellence framework (REF)”. *Journal of informetrics*, v. 13, n. 1, pp. 325-340.
https://doi.org/10.1016/j.joi.2019.01.008

Bornmann, Lutz; **Haunschild, Robin**; **Marx, Werner** (2016). “Policy documents as sources for measuring societal impact: How often is climate change research mentioned in policy-related documents?”. *Scientometrics*, v. 109, n. 3, pp. 1477-1495.
https://doi.org/10.1007/s11192-016-2115-y

Bornmann, Lutz; **Tekles, Alexander**; **Zhang, Helena H.**; **Ye, Fred, Y.** (2019). “Do we measure novelty when we analyze unusual combinations of cited references? A validation study of bibliometric novelty indicators based on F1000Prime data”. *Journal of informetrics*, v. 13, n. 4.
https://doi.org/10.1016/j.joi.2019.100979

Bunn, Frances; **Trivedi, Daksha**; **Alderson, Phil**; **Hamilton, Laura**; **Martin, Alice**; **Pinkney, Emma**; **Iliffe, Steve** (2015). “The impact of Cochrane Reviews: a mixed-methods evaluation of outputs from Cochrane Review Groups supported by the National Institute for Health Research”. *Health technology assessment*, v. 19, n. 28.
https://doi.org/10.3310/hta19280

Burke, James; **Bergman, Jules**; **Asimov, Isaac** (1985). *The impact of science on society*. Washington, DC, USA: National Aeronautics and Space Administration.
https://trove.nla.gov.au/version/21856423

Bush, Vannevar (1945). *Science: the endless frontier. A report to president Truman outlining his proposal for post-war U.S. science and technology policy*. Washington, DC, USA: United States Government Printing Office.
https://www.nsf.gov/od/lpa/nsf50/vbush1945.htm

Buttliere, Brett; **Buder, Jürgen** (2017). “Personalizing papers using Altmetrics: Comparing paper ‘Quality’ or ‘Impact’ to person ‘Intelligence’ or ‘Personality’”. *Scientometrics*, v. 111, n. 1, pp. 219-239.
https://doi.org/10.1007/s11192-017-2246-9

Buxton, Martin; **Hanney, Steve** (1998). “Evaluating the NHS research and development programme: Will the programme give value for money?”. *Journal of the Royal Society of Medicine*, v. 91, n. 35 suppl, pp. 2-6.
https://doi.org/10.1177/014107689809135S02

Callaert, Julie; **Van Looy, Bart**; **Verbeek, Arnold**; **Debackere, Koenraad**; **Thijs, Bart** (2006). "Traces of prior art: An analysis of non-patent references found in patent documents". *Scientometrics*, v. 69, n. 1, pp. 3-20.
https://doi.org/10.1007/s11192-006-0135-8

Dance, Amber (2013). "Impact: pack a punch". *Nature*, v. 502, pp. 397-398.
https://doi.org/10.1038/nj7471-397a

Daugbjerg, Signe B.; **Kahlmeier, Sonja**; **Racioppi, Francesca**; **Martin-Diener, Eva**; **Martin, Brian**; **Oja, Pekka**; **Bull, Fiona** (2009). "Promotion of physical activity in the European region: Content analysis of 27 national policy documents". *Journal of physical activity and health*, v. 6, n. 6, pp. 805-817.
https://doi.org/10.1123/jpah.6.6.805

De-Jong, Stefan P. L.; **Smit, Jorrit**; **Van-Drooge, Leonie** (2016). "Scientists' response to societal impact policies: A policy paradox". *Science and public policy*, v. 43, n. 1, pp. 102-114.
https://doi.org/10.1093/scipol/scv023

De-Jong, Stefan P. L.; **Van-Arensbergen, Pleun**; **Daemen, Floortje**; **Van-der-Meulen, Barend**; **Van-den-Besselaar, Peter** (2011). "Evaluation of research in context: An approach and two cases". *Research evaluation*, v. 20, n. 1, pp. 61-72.
https://doi.org/10.3152/095820211X12941371876346

Department for Business, Energy & Industrial Strategy (2016). *Building on success and learning from experience. An independent review of the Research Excellence Framework*. London, UK: Department for Business, Energy & Industrial Strategy.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/541338/ind-16-9-ref-stern-review.pdf

Derrick, Gemma E.; **Samuel, Gabrielle N.** (2016). "The evaluation scale: Exploring decisions about societal impact in peer review panels". *Minerva*, v. 54, n. 1, pp. 75-97.
https://doi.org/10.1007/s11024-016-9290-0

Donovan, Claire (2008). "The Australian Research Quality Framework: A live experiment in capturing the social, economic, environmental, and cultural returns of publicly funded research". *New directions for evaluation*, n. 118, pp. 47-60.
https://doi.org/10.1002/ev.260

Eisenstein, Michael (2016). "Assessment: Academic return". *Nature*, v. 533, n. 7601, pp. S20-S21.
https://doi.org/10.1038/533S20a

ERIC (2010). *Evaluating the societal relevance of academic research: A guide*. Delft, The Netherlands: Delft University of Technology.
http://resolver.tudelft.nl/uuid:8fa07276-cf52-41f3-aa70-a71678234424

Ernø-Kjølhede, Erik (2000). *Scientific norms as (dis) integrators of scientists?* Copenhagen, Denmark: Copenhagen Business School.
https://research.cbs.dk/en/publications/scientific-norms-as-disintegrators-of-scientists

Ernø-Kjølhede, Erik; **Hansson, Finn** (2011). "Measuring research performance during a changing relationship between science and society". *Research evaluation*, v. 20, n. 2, pp. 131-143.
https://doi.org/10.3152/095820211x12941371876544

Etzkowitz, Henry; **Leydesdorff, Loet** (2000). "The dynamics of innovation: From national systems and "Mode 2" to a Triple Helix of university–industry–government relations". *Research policy*, v. 29, n. 2, pp. 109-123.
https://doi.org/10.1016/S0048-7333(99)00055-4

Fausto, Sibele; **Machado, Fabio A.**; **Bento, Luiz-Fernando J.**; **Iamarino, Atila**; **Nahas, Tatiana R.**; **Munger, David S.** (2012). "Research blogging: Indexing and registering the change in science 2.0". *PLoS one*, v. 7, n. 12, e50109.
https://doi.org/10.1371/journal.pone.0050109

Feller, Irwin (2017). "Assessing the societal impact of publicly funded research". *The journal of technology transfer* (first online).
https://doi.org/10.1007/s10961-017-9602-z

Godin, Benoît; **Doré, Christian** (2005). "Measuring the impacts of science; beyond the economic dimension". INRS Urbanisation, Culture et Société. *HIST Lecture*, Helsinki Institute for Science and Technology Studies, Helsinki, Finland.
http://www.csiic.ca/PDF/Godin_Dore_Impacts.pdf

Göransson, Bo; **Maharajh, Rasigan**; **Schmoch, Ulrich** (2009). "New activities of universities in transfer and extension: Multiple requirements and manifold solutions". *Science and public policy*, v. 36, n. 2, pp. 157-164.
https://doi.org/10.3152/030234209x406863

Grant, Jonathan (1999). "Evaluating the outcomes of biomedical research on healthcare". *Research evaluation*, v. 8, n. 1, pp. 33-38.
https://doi.org/10.3152/147154499781777658

Gregersen, Birgitte; **Linde, Lisbeth-Tved**; **Rasmussen, Jorgen-Gulddahl** (2009). "Linking between Danish universities and society". *Science and public policy*, v. 36, n. 2, pp. 151-156.
https://doi.org/10.3152/030234209x406818

Groth, Paul; **Gurney, Thomas** (2010). "Studying scientific discourse on the Web using bibliometrics: A chemistry blogging case study". In*: Proceedings of the WebSci10: Extending the frontiers of society on-line*, Raleigh, NC, USA: Web Science Trust.
https://www.researchgate.net/publication/228339919_Studying_scientific_discourse_on_the_Web_using_bibliometrics_A_chemistry_blogging_case_study

Hammarfelt, Björn (2014). "Using altmetrics for assessing research impact in the humanities". *Scientometrics*, v. 101, n. 2, pp. 1419-1430.
https://doi.org/10.1007/s11192-014-1261-3

Haunschild, Robin; **Bornmann, Lutz** (2016). "Normalization of Mendeley reader counts for impact assessment". *Journal of informetrics*, v. 10, n. 1, pp. 62-73.
https://doi.org/10.1016/j.joi.2015.11.003

Haunschild, Robin; **Bornmann, Lutz** (2017). "How many scientific papers are mentioned in policy-related documents? An empirical investigation using Web of Science and Altmetric data". *Scientometrics*, v. 110, n. 3, pp. 1209-1216.
https://doi.org/10.1007/s11192-016-2237-2

Haunschild, Robin; **Bornmann, Lutz** (2018). "Field-and time-normalization of data with many zeros: an empirical analysis using citation and Twitter data". *Scientometrics*, v. 116, n. 2, pp. 997-1012.
https://doi.org/10.1007/s11192-018-2771-1

Haunschild, Robin; **Leydesdorff, Loet**; **Bornmann, Lutz**; **Hellsten, Iina**; **Marx, Werner** (2019). "Does the public discuss other topics on climate change than researchers? A comparison of explorative networks based on author keywords and hashtags". *Journal of informetrics*, v. 13, n. 2, pp. 695-707.
https://doi.org/10.1016/j.joi.2019.03.008

Haustein, Stefanie (2014). "Readership metrics". In: Cronin, Blaise; Sugimoto, Cassidy R. (eds.). *Beyond bibliometrics: harnessing multi-dimensional indicators of performance*. Cambridge, MA, USA: MIT Press, pp. 327-344. ISBN: 978 0 262026796

Haustein, Stefanie; **Costas, Rodrigo**; **Larivière, Vincent** (2015). "Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns". *PLoS one*, v. 10, n. 3, e0120495.
https://doi.org/10.1371/journal.pone.0120495

Haustein, Stefanie; **Larivière, Vincent**; **Thelwall, Mike**; **Amyot, Didier**; **Peters, Isabella** (2014). "Tweets vs. Mendeley readers: How do these two social media metrics differ?". *IT – Information technology*, v. 56, n. 5, pp. 207-215.
https://doi.org/10.1515/itit-2014-1048

Health Economics Research Group, RAND Europe (2008). *Medical research: what's it worth? Estimating the economic benefits from medical research in the UK*. London, UK: Evaluation Forum.
https://www.rand.org/pubs/external_publications/EP20080010.html
https://www.brunel.ac.uk/__data/assets/pdf_file/0008/183455/TAP825EconomicBenefitsReportFULLWeb.pdf

Hellsten, Iina; **Leydesdorff, Loet** (2020). "Automated analysis of actor–topic networks on twitter: New approaches to the analysis of socio-semantic networks". *Journal of the Association for Information Science and Technology*, v. 71, n. 1, pp. 3-15.
https://doi.org/10.1002/asi.24207

Heyeres, Marion; **Tsey, Komla**; **Yang, Yinghong**, **Yan, Li**; **Jiang, Hua** (2019). "The characteristics and reporting quality of research impact case studies: A systematic review". *Evaluation and program planning*, v. 73, pp. 10-23.
https://doi.org/10.1016/j.evalprogplan.2018.11.002

Hicks, Daniel J.; **Stahmer, Carl**; **Smith, MacKenzie** (2018). "Impacting capabilities: A conceptual framework for the social value of research". *Frontiers in research metrics and analytics*, v. 3, n. 24.
https://doi.org/10.3389/frma.2018.00024

Higher Education Funding Council for England (2009). *Research Excellence Framework. Second consultation on the assessment and funding of research. September 2009/38*. Bristol, UK: Higher Education Funding Council for England (HEFCE).

Holbrook, James-Britt; **Hrotic, Steven** (2013). "Blue skies, impacts, and peer review". *RT. A journal on research policy & evaluation*, v. 1, n. 1.
https://doi.org/10.13130/2282-5398/2914

Jones, Teresa H.; **Hanney, Steve** (2016). "Tracing the indirect societal impacts of biomedical research: Development and piloting of a technique based on citations". *Scientometrics*, v. 107, n. 3, pp. 975-1003.
https://doi.org/10.1007/s11192-016-1895-4

Kassab, Omar (2019). "Does public outreach impede research performance? Exploring the 'researcher's dilemma' in a sustainability research center". *Science and public policy*, v. 46, n. 5, pp. 710-720.
https://doi.org/10.1093/scipol/scz024

Kellard, Neil M.; **Śliwa, Martina** (2016). "Business and management impact assessment in Research Excellence Framework 2014: Analysis and reflection". *British journal of management*, v. 27, n. 4, pp. 693-711.
https://doi.org/10.1111/1467-8551.12186

Khazragui, Hanan; **Hudson, John** (2015). "Measuring the benefits of university research: Impact and the REF in the UK". *Research evaluation*, v. 24, n. 1, pp. 51-62.
https://doi.org/10.1093/reseval/rvu028

King's College London and *Digital Science*. (2015). *The nature, scale and beneficiaries of research impact: An initial analysis of Research Excellence Framework (REF) 2014 impact case studies*. London, UK: King's College London.
https://www.kcl.ac.uk/policy-institute/assets/ref-impact.pdf

Kok, Maarten O.; **Schuit, Albertine J.** (2012). "Contribution mapping: A method for mapping the contribution of research to enhance its impact". *Health research policy and systems*, v. 10, n. 1, article 21.
https://doi.org/10.1186/1478-4505-10-21

Konkiel, Stacy; **Madjarevic, Natalia**; **Lightfoot, Amy** (2016). *Altmetrics for librarians: 100+ tips, tricks, and examples*.
https://doi.org/10.6084/m9.figshare.3749838

Kovic, Ivor; **Lulic, Ileana**; **Brumini, Gordana** (2008). "Examining the medical blogosphere: An online survey of medical bloggers". *Journal of medical internet research*, v. 10, n. 3, e28.
https://doi.org/10.2196/jmir.1118

Kreiman, Gabriel; **Maunsell, John H. R.** (2011). "Nine criteria for a measure of scientific output". *Frontiers in computational neuroscience*, v. 5, n. 48.
https://doi.org/10.3389/fncom.2011.00048

Lähteenmäki-Smith, Kaisa; **Hyytinen, Kirsi-Maria**; **Kutinlahti, Pirjo**; **Konttinen, Jari** (2006). *Research with an impact: Evaluation practises in public research organisations*. Espoo, Finland: VTT Technical Research Centre of Finland.
http://www.vtt.fi/inf/pdf/tiedotteet/2006/T2336.pdf

Langfeldt, Liv; **Scordato, Lisa** (2015). *Assessing the broader impacts of research: A review of methods and practices*. Oslo, Norway: Nordic Institute for Studies in Innovation, Research and Education (NIFU).
https://nifu.brage.unit.no/nifu-xmlui/handle/11250/282742

Lewison, Grant; **Sullivan, Richard** (2008a). "The impact of cancer research: how publications influence UK cancer clinical guidelines". *British journal of cancer*, v. 98, n. 12, pp. 1944-1950.
https://doi.org/10.1038/sj.bjc.6604405

Liu, Chun-Li; **Xu, Yue-Quan**; **Wu, Hui**; **Chen, Si-Si**; **Guo, Ji-Jun** (2013). "Correlation and interaction visualization of altmetric indicators extracted from scholarly social network activities: Dimensions and structure". *Journal of medical internet research*, v. 15, n. 11, pp. 17.
https://doi.org/10.2196/jmir.2707

Liu, Jean; **Adie, Euan** (2013). "Five challenges in altmetrics: A toolmaker's perspective". *Bulletin of the American Society for Information Science and Technology*, v. 39, n. 4, pp. 31-34.
https://doi.org/10.1002/bult.2013.1720390410

Mansfield, Edwin (1991). "Academic research and industrial innovation". *Research policy*, v. 20, n. 1, pp. 1-12.
https://doi.org/10.1016/0048-7333(91)90080-a

Mansfield, Edwin (1998). "Academic research and industrial innovation: an update of empirical findings". *Research policy*, v. 26, n. 7-8, pp. 773-776.
https://doi.org/10.1016/s0048-7333(97)00043-7

Marcella, Rita; **Lockerbie, Hayley**; **Bloice, Lyndsay**; **Hood, Caroline**; **Barton, Flora** (2018). "The effects of the research excellence framework research impact agenda on early- and mid-career researchers in library and information science". *Journal of information science*, v. 44, n. 5, pp. 608-618.
https://doi.org/10.1177/0165551517724685

Martin, Ben R. (2007). "Assessing the impact of basic research on society and the economy". In: *Rethinking the impact of basic research on society and the economy*, Vienna, Austria.
https://www.researchgate.net/publication/228894032_Assessing_the_Impact_of_Basic_Research_on_Society_and_the_Economy

Mas-Bleda, Amalia; **Thelwall, Mike** (2016). "Can alternative indicators overcome language biases in citation counts? A comparison of Spanish and UK research". *Scientometrics*, v. 109, n. 3, pp. 2007-2030.
https://doi.org/10.1007/s11192-016-2118-8

Merton, Robert K. (1968a). "The Matthew effect in science: The reward and communication systems of science are considered". *Science*, v. 159, n. 3810, pp. 56-63.
https://doi.org/10.1126/science.159.3810.56

Merton, Robert K. (1968b). *Social theory and social structure*. New York, NY, USA: Simon and Schuster. ISBN: 978 0 226520711

Merton, Robert K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago, IL, USA: University of Chicago Press. ISBN: 978 0 226520926

Miettinen, Reijo; **Tuunainen, Juha**; **Esko, Terhi** (2015). "Epistemological, artefactual and interactional–institutional foundations of social impact of academic research". *Minerva*, v. 53, n. 3, pp. 257-277.
https://doi.org/10.1007/s11024-015-9278-1

Mitroff, Ian I. (1974). "Norms and counter-norms in a select group of the Apollo moon scientists: A case study of the ambivalence of scientists". *American sociological review*, v. 39, n. 4, pp. 579-595.
https://doi.org/10.2307/2094423

Moed, Henk F. (2017). *Applied evaluative informetrics*. Heidelberg, Germany: Springer. ISBN: 978 3 319 60522 7

Moed, Henk F.; **Halevi, Gali** (2015). "Multidimensional assessment of scholarly research impact". *Journal of the Association for Information Science and Technology*, v. 66, n. 10, pp. 1988-2002.
https://doi.org/10.1002/asi.23314

Mohammadi, Ehsan; **Thelwall, Mike** (2014). "Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows". *Journal of the Association for Information Science and Technology*, v. 65, n. 8, pp. 1627-1638.
https://doi.org/10.1002/asi.23071

Mohammadi, Ehsan; **Thelwall, Mike**; **Haustein, Stefanie**; **Larivière, Vincent** (2015). "Who reads research articles? An altmetrics analysis of Mendeley user categories". *Journal of the Association for Information Science and Technology*, v. 66, n. 9, pp. 1832-1846.
https://doi.org/10.1002/asi.23286

Mohammadi, Ehsan; **Thelwall, Mike**; **Kwasny, Mary**; **Holmes, Kristi L.** (2018). "Academic information on Twitter: A user survey". *PLoS one*, v. 13, n. 5, e0197265.
https://doi.org/10.1371/journal.pone.0197265

Molas-Gallart, Jordi; **Salter, Ammon J.**; **Patel, Pari**; **Scott, Alister**; **Duran, Xavier** (2002). *Measuring third stream activities. Final report to the Russell Group of Universities*. Brighton, UK: SPRU, University of Sussex.
https://www.researchgate.net/publication/246796517_Measuring_Third_Stream_Activities

Morris, Zoë-Slote; **Wooding, Steven**; **Grant, Jonathan** (2011). "The answer is 17 years, what is the question: Understanding time lags in translational research". *Journal of the Royal Society of Medicine*, v. 104, n. 12, pp. 510-520.
https://doi.org/10.1258/jrsm.2011.110180

Mostert, Sebastian P.; **Ellenbroek, Stéfan P. H.**; **Meijer, Ingeborg**; **Van-Ark, Gerrit**; **Klasen, Eduard C.** (2010). "Societal output and use of research performed by health research groups". *Health research policy and systems*, v. 8, n. 1, article 30.
https://doi.org/10.1186/1478-4505-8-30

Narin, Francis (1976). *Evaluative bibliometrics: the use of publication and citation analysis in the evaluation of scientific activity*. Cherry Hill, NJ, USA: Computer Horizons, 459 pp.
http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.473.8004

Narin, Francis; **Hamilton, Kimberley S.**; **Olivastro, Dominic** (1997). "The increasing linkage between US technology and public science". *Research policy*, v. 26, n. 3, pp. 317-330.
https://doi.org/10.1016/S0048-7333(97)00013-9

National Research Council (2014). *Furthering America's research enterprise*. Washington, DC, USA: The National Academies Press. ISBN: 978 0 309 30522 8
https://www.ncbi.nlm.nih.gov/books/NBK253897

Nelhans, Gustaf (2016). "Professional impact of clinical research". In: Ràfols, Ismael; Molas-Gallart, Jordi; Castro-Martínez, Elena; Woolley, Richard (eds.). *Proceedings of the 21st International Conference on Science and Technology Indicators*. València, Spain: Universitat Politècnica de València.
http://ocs.editorial.upv.es/index.php/STI2016/STI2016/schedConf/presentations
https://www.researchgate.net/publication/308389908_Professional_impact_of_clinical_research

Noyons, Ed (2018). *Monitoring how science finds its way into society: Measuring societal impact through area-based connectedness (ABC)*.
https://www.cwts.nl/blog?article=n-r2u264&title=monitoring-how-science-finds-its-way-into-society-measuring-societal-impact-through-area-based-connectedness-abc

Noyons, Ed (2019). "Measuring societal impact is as complex as ABC". *Journal of data and information science*, v. 4, n. 3, pp. 6-21.
https://doi.org/10.2478/jdis-2019-0012

Noyons, Ed; **Ràfols, Ismael** (2018). "Can bibliometrics help in assessing societal contributions of agricultural research? Exploring societal interactions across research areas". In: Wouters, Paul (ed.), *Proceedings of the science and technology indicators conference 2018 Leiden. Science, Technology and Innovation indicators in transition*. Leiden, the Netherlands: University of Leiden.
http://hdl.handle.net/1887/65366

Nussbaum, Martha C. (2001). *Women and human development: The capabilities approach* (Vol. 3). Cambridge, UK: Cambridge University Press. ISBN: 978 0 521003858

Nussbaum, Martha C. (2009). *Frontiers of justice: Disability, nationality, species membership*. Cambridge, MA, USA: Harvard University Press. ISBN: 978 0 674024106

Oancea, Alis (2013). "Buzzwords and values: the prominence of 'impact' in UK research policy and governance. *Research trends*, n. 33, pp. 6-8.
https://www.researchtrends.com/issue-33-june-2013/buzzwords-and-values

Ortega, José-Luis (2018). "Disciplinary differences of the impact of altmetric". *Fems microbiology letters*, v. 365, n. 7.
https://doi.org/10.1093/femsle/fny049

Ovseiko, Pavel V.; **Oancea, Alis**; **Buchan, Alastair M.** (2012). "Assessing research impact in academic clinical medicine: A study using Research Excellence Framework pilot impact indicators". *BMC Health services research*, n. 12, article 478.
https://doi.org/10.1186/1472-6963-12-478

Poege, Felix; **Harhoff, Dietmar**; **Gaessler, Fabian**; **Baruffaldi, Stefano** (2019). "Science quality and the value of inventions". *Science advances*, v. 5, n. 12, eaay7323.
https://doi.org/10.1126/sciadv.aay7323

Pollitt, Alexandra; **Potoglou, Dimitris**; **Patil, Sunil**; **Burge, Peter**; **Guthrie, Susan**; **King, Suzanne**; **Wooding, Steven**; **Grant, Jonathan** (2016). "Understanding the relative valuation of research impact: A best–worst scaling experiment of the general public and biomedical and health researchers". *BMJ Open*, v. 6, article e010916.
https://doi.org/10.1136/bmjopen-2015-010916

Priem, Jason; **Piwowar, Heather A.**; **Hemminger, Bradley M.** (2012). *Altmetrics in the wild: Using social media to explore scholarly impact*.
https://arxiv.org/abs/1203.4745

Priem, Jason; **Taraborelli, Dario**; **Groth, Paul**; **Neylon, Cameron** (2010). *Altmetrics: A manifesto*.
http://altmetrics.org/manifesto

Pulido, Cristina M.; **Redondo-Sama, Gisela**; **Sordé-Martí, Teresa**; **Flecha, Ramon** (2018). "Social impact in social media: A new method to evaluate the social impact of research". *PLoS one*, v. 13, n. 8, e0203117.
https://doi.org/10.1371/journal.pone.0203117

Puschmann, Cornelius; **Mahrt, Merja** (2012). "Scholarly blogging: A new form of publishing or science journalism 2.0?". In: *Science and the internet*, pp. 171-182. Düsseldorf, Germany: Düsseldorf University Press. ISBN: 978 3 943460 16 2
http://dup.oa.hhu.de/id/eprint/149

Raftery, James; **Hanney, Steve**; **Greenhalgh, Trish**; **Glover, Matthew**; **Blatch-Jones, Amanda** (2016). "Models and applications for measuring the impact of health research: update of a systematic review for the Health Technology Assessment programme". *Health technology assessment*, v. 20, n. 76.
https://doi.org/10.3310/hta20760

Robinson-García, Nicolás; **Costas, Rodrigo**; **Isett, Kimberley**; **Melkers, Julia**; **Hicks, Diana** (2017). "The unbearable emptiness of tweeting -About journal articles". *PLoS one*, v. 12, n. 8, e0183551.
https://doi.org/10.1371/journal.pone.0183551

Robinson-García, Nicolás; **Torres-Salinas, Daniel**; **Zahedi, Zohreh**; **Costas, Rodrigo** (2014). "New data, new possibilities: Exploring the insides of *Altmetric.com*". *El profesional de la información*, v. 23, n. 4, pp. 359-366.
https://doi.org/10.3145/epi.2014.jul.03

Rodgers, Emily; **Barbrow, Sarah** (2013). *A look at altmetrics and its growing significance to research libraries*. Ann Arbor, MI, USA: The University of Michigan University Library.
http://hdl.handle.net/2027.42/99709

Roemer, Robin-Chin; **Borchardt, Rachel** (2015). *Meaningful metrics: A 21st century librarian's guide to bibliometrics, altmetrics, and research impact*. Chicago, IL, USA: Association of College and Research Libraries, A division of the American Library Association. ISBN: 978 0 8389 8755 1
http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/9780838987568_metrics_OA.pdf

Rowlands, Ian (2018). "What are we measuring? Refocusing on some fundamentals in the age of desktop bibliometrics". *FEMS Microbiology letters*, v. 365, n. 8, pp. fny059.
https://doi.org/10.1093/femsle/fny059

RQF Development Advisory Group (2005). *Research quality framework: Assessing the quality and impact of research in Australia. Research impact (report by the RQF development advisory group)*. Canberra, Australia: Department of Education.

Ruegg, Rosalie T.; **Feller, Irwin** (2003). *A toolkit for evaluating public R&D investment: Models, methods, and findings from ATP's first decade*. Gaithersburg, MD, USA: National Institute of Standards and Technology.
https://nvlpubs.nist.gov/nistpubs/gcr/2003/NIST.GCR.03-857.pdf

Samuel, Gabrielle N.; **Derrick, Gemma E.** (2015). "Societal impact evaluation: Exploring evaluator perceptions of the characterization of impact under the REF2014". *Research evaluation*, v. 24, n. 3, pp. 229-241.
https://doi.org/10.1093/reseval/rvv007

Searles, Andrew; **Doran, Chris**; **Attia, John**; **Knight, Darryl**; **Wiggers, John**; **Deeming, Simon**; **Mattes, Joerg**; **Webb, Brad**; **Hannan, Steve**; **Ling, Rod**; **Edmunds, Kim**; **Reeves, Penny**; **Nilsson, Michael** (2016). "An approach to measuring and encouraging research translation and research impact". *Health research policy and systems*, n. 14, article 60.
https://doi.org/10.1186/s12961-016-0131-2

Serrano-López, Antonio-Eleazar; **Ingwersen, Peter**; **Sanz-Casado, Elías** (2017). "Wind power research in Wikipedia: Does Wikipedia demonstrate direct influence of research publications and can it be used as adequate source in research evaluation?". *Scientometrics*, v. 112, n. 3, pp. 1471-1488.
https://doi.org/10.1007/s11192-017-2447-2

Shema, Hadas; **Bar-Ilan, Judit**; **Thelwall, Mike** (2012). "Research blogs and the discussion of scholarly information". *PLoS one*, v. 7, n. 5, e35869.
https://doi.org/10.1371/journal.pone.0035869

Shema, Hadas; **Bar-Ilan, Judit**; **Thelwall, Mike** (2014). "Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics". *Journal of the Association for Information Science and Technology*, v. 65, n. 5, pp. 1018-1027.
https://doi.org/10.1002/asi.23037

Shema, Hadas; **Bar-Ilan, Judit**; **Thelwall, Mike** (2015). "How is research blogged? A content analysis approach". *Journal of the Association for Information Science and Technology*, v. 66, n. 6, pp. 1136-1149.
https://doi.org/10.1002/asi.23239

Siampi (2011). *Final report on social impacts of research*. Social Impact Assessment Methods for research and funding instruments through the study of productive interactions (SIAMPI).
http://www.siampi.eu

Spaapen, Jack B.; **Dijstelbloem, Huub**; **Wamelink, Frank** (2007). *Evaluating research in context: A method for comprehensive assessment*. The Hague, the Netherlands: Consultative Committee of Sector Councils for Research and Development (COS).
https://www.qs.univie.ac.at/fileadmin/user_upload/d_qualitaetssicherung/Dateidownloads/Evaluating_Research_in_context_-_A_method_for_comprehensive_assessment.pdf

Spaapen, Jack B.; **Van-Drooge, Leonie** (2011). "Introducing 'productive interactions' in social impact assessment". *Research evaluation*, v. 20, n. 3, pp. 211-218.
https://doi.org/10.3152/095820211X12941371876742

Sugimoto, Cassidy R.; **Work, Sam**; **Larivière, Vincent**; **Haustein, Stefanie** (2017). "Scholarly use of social media and altmetrics: A review of the literature". *Journal of the Association for Information Science and Technology*, v. 68, n. 9, pp. 2037-2062.
https://doi.org/10.1002/asi.23833

Tattersall, Andy; **Carroll, Christopher** (2018). "What can Altmetric.com tell us about policy citations of research? An analysis of Altmetric.com data for research articles from the University of Sheffield". *Frontiers in research metrics and analytics*, v. 2, n. 9.
https://doi.org/10.3389/frma.2017.00009

Taylor, Mike (2013a). "Exploring the boundaries: How altmetrics can expand our vision of scholarly communication and social impact". *Information standards quarterly*, v. 25, n. 2, pp. 27-32.
https://doi.org/10.3789/isqv25no2.2013.05

Taylor, Mike (2013b). "Towards a common model of citation: Some thoughts on merging altmetrics and bibliometrics". *Research trends*, n. 35, pp. 19-22.
https://www.researchtrends.com/issue-35-december-2013/towards-a-common-model-of-citation-some-thoughts-on-merging-altmetrics-and-bibliometrics

Terama, Emma; **Smallman, Melanie**; **Lock, Simon J.**; **Johnson, Charlotte**; **Austwick, Martin-Zaltz** (2017). "Beyond academia - Interrogating research impact in the Research Excellence Framework". *Plos one*, v. 11, n. 12, e0168533.
https://doi.org/10.1371/journal.pone.0172817

Thelwall, Mike (2017a). "Three practical field normalised alternative indicator formulae for research evaluation". *Journal of informetrics*, v. 11, n. 1, pp. 128-151.
https://doi.org/10.1016/j.joi.2016.12.002

Thelwall, Mike (2017b). *Web indicators for research evaluation: A practical guide*. London, UK: Morgan & Claypool.
https://doi.org/10.2200/S00733ED1V01Y201609ICR052

Thelwall, Mike; **Kousha, Kayvan** (2015). "Web indicators for research evaluation. Part 2: Social media metrics". *El profesional de la información*, v. 24, n. 5, pp. 607-620.
https://doi.org/10.3145/epi.2015.sep.09

Thelwall, Mike; **Maflahi, Nabaeil** (2015). "Guideline references and academic citations as evidence of the clinical value of health research". *Journal of the Association for Information Science and Technology*, v. 67, n. 4.
https://doi.org/10.1002/asi.23432

Van-den-Akker, Wiljan; **Spaapen, Jack**; **Maes, Katrien** (2017). *Productive interactions: Societal impact of academic research in the knowledge society*. League of European Research Universities (LERU).
https://www.leru.org/publications/productive-interactions-societal-impact-of-academic-research-in-the-knowledge-society#

Van-der-Meulen, Barend; **Rip, Arie** (2000). "Evaluation of societal quality of public sector research in the Netherlands". *Research evaluation*, v. 9, n. 1, pp. 11-25.
https://doi.org/10.3152/147154400781777449

Van-Noorden, Richard (2015). "Seven thousand stories capture impact of science". *Nature*, n. 518(7538), pp. 150-151.
https://doi.org/10.1038/518150a

Vilkins, Samantha; **Grant, Will J.** (2017). "Types of evidence cited in Australian government publications". *Scientometrics*, v. 113, n. 3, pp. 1681-1695.
https://doi.org/10.1007/s11192-017-2544-2

Waltman, Ludo; **Costas, Rodrigo** (2014). "F1000 Recommendations as a potential new data source for research evaluation: A comparison with citations". *Journal of the Association for Information Science and Technology*, v. 65, n. 3, pp. 433-445.
https://doi.org/10.1002/asi.23040

Way, Catharine (2015). *The millennium development goals report 2015*.
https://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20rev%20(July%201).pdf

Willis, Cameron D.; **Riley, Barbara**; **Stockton, Lisa**; **Viehbeck, Sarah**; **Wutzke, Sonia**; **Frank, John** (2017). "Evaluating the impact of applied prevention research centres: Results from a modified Delphi approach". *Research evaluation*, v. 26, n. 2, pp. 78-90.
https://doi.org/10.1093/reseval/rvx010

Wilsdon, James; **Allen, Liz**; **Belfiore, Eleonora**; **Campbell, Philip**; **Curry, Stephen**; **Hill, Steven** *et al.* (2015). *The metric tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. Bristol, UK: Higher Education Funding Council for England (Hefce).
https://doi.org/10.4135/9781473978782

Wouters, Paul (2016). *Self evaluation report CWTS 2008 - 2015*. Leiden, The Netherlands: CWTS.

Wouters, Paul; **Zahedi, Zohreh**; **Costas, Rodrigo** (2019). "Social media metrics for new research evaluation". In: Glänzel, Wolfgang; Moed, Henk; Schmoch, Ulrich; Thelwall Mike (eds.). *Springer handbook of science and technology indicators.* Cham, Switzerland: Springer International Publishing, pp. 687-713. ISBN: 978 3 030 02511 3

Wu, Lingfei; **Wang, Dashun**; **Evans, James A.** (2019). "Large teams develop and small teams disrupt science and technology". *Nature*, n. 566, pp. 378-382.
https://doi.org/10.1038/s41586-019-0941-9

Xia, Feng; **Su, Xiaoyan**; **Wang, Wei**; **Zhang, Chenxin**; **Ning, Zhaolong**; **Lee, Ivan** (2016). "Bibliographic analysis of Nature based on Twitter and Facebook altmetrics data". *PLoS one*, v. 11, n. 12, e0165997.
https://doi.org/10.1371/journal.pone.0165997

Yu, Houqiang (2017). "Context of altmetrics data matters: An investigation of count type and user category". *Scientometrics*, v. 111, n. 1, pp. 267-283.
https://doi.org/10.1007/s11192-017-2251-z

Zahedi, Zohreh; **Costas, Rodrigo** (2018). "General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators". *PLoS one*, v. 13, n. 5, e0197326.
https://doi.org/10.1371/journal.pone.0197326

Zahedi, Zohreh; **Costas, Rodrigo**; **Wouters, Paul** (2014). "How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications". *Scientometrics*, v. 101, n. 2, pp. 1491-1513.
https://doi.org/10.1007/s11192-014-1264-0

Zahedi, Zohreh; **Fenner, Martin**; **Costas, Rodrigo** (2014). "How consistent are altmetrics providers? Study of 1000 PLoS one publications using the PLoS alm, Mendeley and Altmetric.com APIs". *Altmetrics 14.* Workshop, Web Science Conference, Bloomington, USA.
https://figshare.com/articles/How_consistent_are_altmetrics_providers_Study_of_1000_PLOS_ONE_publications_using_the_PLOS_ALM_Mendeley_and_Altmetric_com_APIs/1041821

Zahedi, Zohreh; **Van-Eck, Nees-Jan** (2018). "Exploring topics of interest of Mendeley users". *Journal of altmetrics*, v. 1, n. 1, p. 5.
https://doi.org/10.29024/joa.7