To what extent does the Leiden Manifesto also apply to altmetrics? A discussion of the manifesto against the background of research into altmetrics

Lutz Bornmann* & Robin Haunschild**

* Division for Science and Innovation Studies

Administrative Headquarters of the Max Planck Society

Hofgartenstr. 8,

80539 Munich, Germany.

Email: bornmann@gv.mpg.de

** Max Planck Institute for Solid State Research

Heisenbergstr. 1,

70569 Stuttgart, Germany.

Email: R.Haunschild@fkf.mpg.de

Abstract

Hicks, Wouters, Waltman, de Rijcke, and Rafols (2015) have formulated the so-called Leiden manifesto, in which they have assembled the ten principles for a meaningful evaluation of research on the basis of bibliometric data. In this work the attempt is made to indicate the relevance of the Leiden manifesto for altmetrics. As shown by the discussion of the ten principles against the background of the knowledge about and the research into altmetrics, the principles also have a great importance for altmetrics and should be taken into account in their application. Especially important points in this connection are:

1) If altmetrics are really to be used for evaluations in the future, they should only be used in the framework of an informed peer review.

2) To obtain informative results, advanced altmetric indicators (normalized indicators) should be used.

3) In an evaluation with the help of altmetrics data, the context of the research unit should be taken into account. Analyses should only be undertaken if they make sense for the unit.

4) The altmetrics data used in an evaluation should be transparent and openly accessible.

5) The presentation of altmetrics data should avoid a misplaced concreteness and false precision, since the underlying data does not permit this.

6) The use of data in research evaluation should be accompanied by continuing scientometric research, in order to optimize altmetric indicators or to develop new ones. Attention must be paid to system effects of indicators, in order e.g. to recognize gaming.

Keywords

Altmetrics, Leiden manifesto, Bibliometrics, Scientometrics

1 Introduction

The basis of evaluative bibliometrics is a relatively simple procedure of determining the number of publications and their citations, and using this information for the assessment of research. The appropriate access to databases like Web of Science (WoS, Thomson Reuters) and Google Scholar, allows a bibliometric analysis to be performed for a person, an institute or a country in a short time. These analyses are correspondingly often undertaken by scientists or people from the science policy environment. However, it is sometimes not clear to these people that the production of a bibliometric analysis is beset by numerous pitfalls. Thus, for example, the average citation rates in the various specialist disciplines differ, which means that a bibliometric analysis should either be confined to one discipline or that so-called advanced indicators should be used which contain field-normalized citation scores.

In recent decades, a research field has been established for bibliometrics (or more widely: for scientometrics) which deals professionally with the pitfalls and not only develops advanced indicators, but also suggests methods for the assessment of bibliometric data (Moed, 2005; Vinkler, 2010). This research field has its own academic chairs, conferences (such as the International Conference on Scientometrics and Informetrics) and prizes (such as the Derek de Solla Price Memorial Medal). Scientists and people from the science policy background who use bibliometric data often do not know that this research field has been established and that suggestions for reliable, fair and valid application of bibliometrics exist (Bornmann et al., 2014). For this reason Hicks, et al. (2015) formulated the so-called Leiden manifesto, in which they assembled the ten principles for a meaningful research evaluation on the basis of bibliometric data. For instance, one principle indicates that a bibliometric analysis should take into account that there are variations by field in publication and citation practices (principle 6).

In recent years, a new research field has been established in scientometrics which is known as altmetrics (abbreviation for alternative metrics) (Das & Mishra, 2014; Glänzel & Gorraiz, 2014). "Altmetrics, short for alternative metrics … is a term to describe web-based metrics for the impact of scholarly material, with an emphasis on social media outlets as sources of data" (Shema, Bar-Ilan, & Thelwall, 2014). Altmetrics uses data in particular from social media applications (such as Facebook, Twitter or Mendeley), to determine the impact of publications. Since these social media are not only used by scientists, but also by other people from quite varied sections of society, it seems possible that altmetrics - unlike traditional metrics - can yield an impact measurement of research from areas outside science (Bornmann, 2014c). In addition, altmetrics have the advantage over traditional metrics of allowing a measurement of impact relatively soon after the publication of a paper (Priem & Hemminger, 2010) and that altmetrics data can also be produced for large data sets without excessive effort (Bornmann, 2014a).

Like traditional metrics, however, altmetrics also have a series of drawbacks and numerous pitfalls which should be taken into account in their use. A (growing) group of scientometricians who have specialized in altmetrics research are concerning themselves with the disadvantages and pitfalls (e.g. Bornmann, 2014b; Haustein, 2014; Mohammadi & Thelwall, 2014; Piwowar, 2013; Priem, 2014; Puschmann, 2014; Shema, Bar-Ilan, & Thelwall, 2012; Thelwall, Haustein, Lariviere, & Sugimoto, 2013; Zahedi, Costas, & Wouters, 2014). Even if this research is still at an early stage (Weller, 2015), altmetrics are nonetheless already being applied for evaluative purposes: Thus, there are reports that researchers provide them in publication lists (Piwowar & Priem, 2013), in the Scopus literature database (Elsevier) they are displayed alongside the search results for the publications, and they are mentioned in the Snowball Metrics Recipe Book (Colledge, 2014). This book consists of a collection of definitions of indicators for research evaluation agreed upon by a range of research organizations.

Even if altmetrics are not yet as established in research evaluation as the traditional metrics, we see the necessity of concerning ourselves with the principles of their application. Since the principles of Hicks, et al. (2015) were formulated very generally, we will undertake the attempt in the current work to transfer them to the area of altmetrics. Even if such principles have not yet been explicitly formulated for the area of altmetrics, works already exist which deal with the rules of their application (Liu & Adie, 2013; Neylon, Willmers, & King, 2014; NISO Alternative Assessment Metrics Project, 2014; Thelwall, 2014). As early as 2010 Priem, Taraborelli, Groth, and Neylon (2010) formulated a manifesto for altmetrics in which they do not provide any principles, but (1) propose the expression "altmetrics", (2) explain for what purpose altmetrics can be used, and (3) discuss which questions altmetrics research should address.

2 Principles of the Leiden Manifesto

The following briefly introduces the principles of the Leiden manifesto and discusses them in connection with altmetrics.

2.1 Principle 1: Quantitative evaluation should support qualitative, expert assessment

Metrics should not be seen as an alternative to the evaluation of scientific work by peers, but as a supporting instrument. This can lead to an expectation of better decisions by the peers, and any bias tendencies in peer review can be counteracted (Bornmann, 2011b). This principle opposes attempts to establish a "bibliometric peer review" (Sanström, 2014), which (entirely) does without peer review.

In the area of altmetrics there are also voices calling for these data to be seen as an alternative to established methods of research evaluation. Thus, for instance, Fausto et al. (2012) write the following: "These new tools are based on a belief in the failure and

insufficiency of the three more traditional filters – peer-review, citation counting analysis, and Journal Impact Factor – to indicate the most relevant and significant sources in a context of an explosive growth of the volume of academic literature in today's internet-age science." Priem, et al. (2010) consider that one can crowd source peer review with altmetrics: "Instead of waiting months for two opinions, an article's impact might be assessed by thousands of conversations and bookmarks in a week. In the short term, this is likely to supplement traditional peer-review, perhaps augmenting rapid review in journals like *PLoS ONE*, *BMC Research Notes*, or *BMJ Open*. In the future, greater participation and better systems for identifying expert contributors may allow peer review to be performed entirely from altmetrics".

Even if these opinions on the replacement of the established research evaluation by altmetrics do exist, the dominant opinion in the discussion of altmetrics is that these metrics should only be used in the framework of a peer review process. "Although spontaneous reviews from readers and novel altmetrics are welcomed complementary evaluation tools, they will not replace a thorough scientific quality assessment of papers and scientists through a selected-expert peer review any time soon" (Anon, 2012, p. 907). According to Rousseau and Ye (2013) indicators are generally used " as support for peers who must perform evaluations of research groups, departments and universities, or make tenure decisions. Note that we explicitly use the word 'support.' Indeed, real scientific progress can only be acknowledged by peers who understand the research they have to evaluate … we propose that combining informetric data (via a multi-metric approach) and peer review (for the many aspects that are not quantitative, including the interpretation of quantitative data) is necessary for all forms of academic evaluation".

Altmetrics are especially well suited to the evaluation of publications, since publications can be relatively clearly identified with particular identifiers (e.g. DOIs) (Franceschini, Maisano, & Mastrogiacomo, 2015). But also other forms of output (non-

traditional output, such as datasets or exhibitions) are suitable for an evaluation by altmetrics. However, there is as yet relatively little research into this in the area of altmetrics. Besides the evaluation of non-traditional output, altmetrics has been applied to the assessment of the broad impact of research – that is the impact of research on other subsections of society (besides science): "Other motivations for administrator support of altmetrics – as tuition and other financial factors become increasingly important for some institutions – include attracting potential students and funding by showing a more complete picture of the institution's talent and the broad impact of its work" (NISO Alternative Assessment Metrics Project, 2014). On the question of the extent to which altmetrics are appropriate for a measurement of broad impact, a range of altmetrics studies has already been performed (Bornmann, 2014a, in press). Here, it is ultimately a matter of a targeted impact measurement of output, as described, for instance by Neylon, et al. (2014): "A signature of important policy and position papers is that they are highly viewed, have significant but not outstanding numbers of citations, and have large numbers of scholarly bookmarks" (p. 7).

2.2 Principle 2: Measure performance against the research missions of the institution, group, or researcher.

Scientists work in very varied contexts (e.g. subjects, projects, institutions, programs, cultures, countries). This context should be taken into account in the selection of indicators and the interpretation of results. Thus, for example, an institution performing research in the area of the humanities should not be evaluated on the basis of data from the WoS. The coverage of publications from this area is insufficient in the WoS (Marx & Bornmann, 2015).

In principle, it could be easier with altmetrics than with traditional metrics to take into account the specific context of an evaluated unit. Altmetrics allow a very wide range of different evaluation possibilities. "Altmetrics consists of a wide variety of data with different characteristics, linked by a common set of tools" (Taylor, 2013, p. 19). These evaluation

possibilities relate not only to the output investigated (see above), but also to the impact investigated. According to Priem (2014) "the breadth of altmetrics could support more holistic evaluation efforts; a range of altmetrics may help solve the reliability problems of individual measures by triangulating scores from easily-accessible 'converging partial indicators' … Altmetrics could also support the evaluation of increasingly important, non-traditional scholarly products like datasets and software, which are currently underrepresented in the citation record." A supplier of altmetrics data – Altmetric (www.altmetric.com) – has, for example, recently opened up a new source for altmetrics data: policy and guidance documents (Liu, 2014). With the help of this data it should be possible in the near future to measure adequately the impact of institutions which pursue research relevant to politics. However, no research evaluation should be performed with this data if the mission of the institution does not relate to politically relevant research.

It is especially in the social sciences and humanities that altmetrics offers the possibility of measuring the impact of their specific outcomes, which generally goes beyond papers in international peer reviewed journals. Even if the findings of Hammarfelt (2013) indicate that "many of the problems identified in research on the use of bibliometrics on the humanities are also relevant for altmetric approaches" (p. 720), the findings of Chen, Tang, Wang, and Hsiang (2015) show that "some new metrics were able to provide extra explanatory power not captured by the conventional measures, which suggests the complementary value of these new measures" (p. 108). According to the results of Hammarfelt (2014) it is Twitter, among the altmetrics, that seems especially suited to the impact measurement in the area of social sciences and humanities.

Peters, Jobmann, Eppelin, Hoffmann, and Künne (2014) have tested the use of altmetrics in a case study of a non-university research organization in Germany, which can only be poorly measured with classical bibliometric instruments. The study showed that for some institutes of the organization "altmetrics provide a real alternative for bibliometric

evaluations since more publications can be found on social media platforms than in databases traditionally used for research evaluation (e.g., Scopus)."

2.3 Principle 3: Protect excellence in locally relevant research

Even excellent research cannot always be identified by the usual literature databases which are used in bibliometrics. The two most important databases for bibliometrics (WoS and Scopus) index journals which are directed towards an international readership and are received by an international readership. Since the humanities and social sciences often write for a national readership in the appropriate national language, they are especially affected by this problem. Their publications and corresponding citations are insufficiently covered in WoS and Scopus. The problem of the identification of excellence in locally relevant research is mainly a problem of the data available in a database and the benchmarks which are applied to the evaluation of the publications. Against this background Bornmann, Thor, Marx, and Schier (2015) have suggested a bibliometric procedure for the evaluation of research in these disciplines - as long as their research results mainly appear in publications. Here, Google Scholar is used as the data source, and suitable reference sets are applied to measure performance in an appropriate context. Thus, for example, the citation impact of a paper which has appeared in a proceedings volume of a nationally relevant conference should be compared with the average impact of the other papers in the same volume (measured by citations found in the same database).

The principle of protecting excellence in locally relevant research overlaps with the principle of taking account of the specific context of an evaluated unit (see section 2.2): Excellence in locally relevant research can only be identified after taking into account the context of the evaluated unit. As we have already explained above, altmetrics are particularly suited to accounting for the context of the evaluation, since they offer a wide spectrum of data and assessment possibilities. Besides researchers there are "journalists, policymakers, public

servants, public health professionals, civil society organizations and technology groups with an active presence on social media" (Neylon, et al., 2014, p. 9), whose access to locally relevant research can be measured specifically. Today, we have basically arrived at a situation where "for every single use of an electronic resource, the system can record which resource was used, who used it, where that person was, when it was used, what type of request was issued, what type of record it was, and from where the article was used" (Kurtz & Bollen, 2010, p. 4). These specific (new) possibilities of usage measurement should facilitate the identification of excellence in locally relevant research. It will be shown by the altmetrics research in future, whether these expectations are fulfilled by altmetrics,

2.4 Principle 4: Keep data collection and analytical processes open, transparent and simple

The bibliometric analysis should not be taken from a black-box evaluation machine. Those who undertake a bibliometric analysis, and those who receive this analysis, should have access to the underlying data and a transparent representation of the analyses undertaken. The methods of the analysis should be as simple as possible and thus transparent. But the simplicity should not lead to the neglect of other important principles like the normalization of citation impact in the framework of advanced bibliometric indicators (see section 2.6).

This principle can be transferred directly to the area of altmetrics: There too, the data collection and analytical processes should be open, transparent and simple. However, since the altmetrics cover a wide spectrum of data sources, which can be evaluated in several different ways, it will certainly be more difficult for altmetrics to be transparent and simple: "Attempts to capture such a wide range of research impacts require a toolbox of methods and approaches to track the reach, use, and reuse of research outputs such as journal articles, datasets, and software" (Dinsmore, Allen, & Dolby, 2014). In this connection, we see an

especially difficult problem with the inconsistencies which have been reported from many sides about altmetrics data (Hammarfelt, 2014): Even when data providers like Altmetric, ImpactStory, and Plum Analytics extract data from the same sources (such as Twitter), inconsistencies in the data are reported. According to the results of Zahedi, Fenner, and Costas (2014) it seems that although the three studied altmetrics providers "share some data sources (i.e. Facebook, Twitter and Mendeley) and also the date and time of altmetrics data collection from these three providers have been controlled in this study, altmetrics data reported for the same dataset of publications is not consistent among them and large differences have been observed" (see also Chamberlain, 2013). For instance, original tweets and retweets are handled differently by different altmetrics data providers.

Therefore, it is not only important for altmetrics research to investigate the quality, reliability and robustness of the altmetrics tools, but also to provide the origin of the data and the point in time when the data was extracted for an altmetrics study. In the interpretation of the results, one should take into account that the data can change more or less significantly (relatively soon) after it is extracted. In addition, the data of an evaluation study should be openly available - at least for the evaluated unit.

2.5 Principle 5: Allow those evaluated to verify data and analysis

All scientists who are part of a bibliometric analysis should be able to inspect, check and validate their data. Many bibliometric analyses are undertaken without the knowledge of the evaluated unit, which naturally then has no possibility of validating their data.

This principle should also apply for evaluation studies which are based on altmetrics data. It is particularly important in this area to allow an inspection by the evaluated unit, since the data quality is generally worse than it ought to be for bibliometric data. We have already noted above that one can expect different results from evaluations which may use the same source and time period but are based on the data of different data suppliers. In addition, there

are no rules for the way a paper on Twitter or in a blog should be mentioned (cited), or how an altmetric aggregator assigns a mention to a paper. “Although most tweeters, science bloggers and digitally native media outlets diligently include direct links to the journal articles they discuss, traditional news outlets have no such standard practice. As a result, a large number of science, health and technology news reports fail to include links to the research that they mention” (Liu & Adie, 2013, p. 32). An exact search for the mentions of a paper on Twitter or Facebook can therefore often lead to other results than those which one can find with the data suppliers of altmetrics data.

An additional problem which we also know from the area of bibliometrics is that altmetrics is a case of sparse data. “That is, the absence of an indicator cannot reliably be taken to mean an absence of activity. For example a lack of Mendeley bookmarks may not mean that a paper is not being saved by researchers, just that those who do are not using Mendeley to do it. A lack of tweets about an article does not mean it is not being discussed” (Neylon, 2014). A series of altmetrics is based on the tool of just one platform (like Twitter, Facebook or Mendeley). Even if the coverage of publications by the tools of these platforms is generally high, one can still assume that additional mentions of a paper would be found by the tools of other platforms. Therefore, a check of evaluation results should always refer to the data from other platforms: Thus, for instance, one could check whether the results based on Mendeley data can be confirmed with data from CiteULike (Mohammadi & Thelwall, 2014).

2.6 Principle 6: Account for variation by field in publication and citation practices

There are clear differences in the document types in which the research results in disciplines are published. For the computer scientists, for example, conference papers are very important and in the humanities books and book chapters. These differences should be taken into account in the measurement of the output. In the WoS, for example, books und book

chapters are hardly indexed, which means that this database is hardly suitable for evaluation in the humanities. In addition, the citation rates of the disciplines vary when they are measured with databases like WoS or Scopus. The responsibility for this seems to derive mainly from the different degrees of coverage of the literature in the databases (Marx & Bornmann, 2015). It is therefore important that an analysis use normalized indicators which take into account the different citation rates in the disciplines.

In bibliometrics a main procedure for normalization has been used for years, in which the citation impact of a specific publication is compared with the **average** citation impact of those papers which have appeared in the same subject area – the so-called reference set (Schubert & Braun, 1986; Waltman, van Eck, van Leeuwen, Visser, & van Raan, 2011). Hicks, et al.(2015) consider another, similar, procedure "the most robust normalization method" (p. 430): Here percentiles are calculated for the papers in a reference set (Bornmann, 2013a; Bornmann, Leydesdorff, & Mutz, 2013). For example, a percentile of 90 means that 10% of the papers in the set have achieved a higher and 90% a lower impact. The percentiles which were calculated for papers in question with the help of reference sets can be used for cross-subject comparisons. The advantage of percentiles, as opposed to the mean-citation based procedures mentioned above, is that the influence of a few, extremely highly cited papers on the normalization (and finally the bibliometric results) is minimized (Waltman et al., 2012).

It is also important with the altmetrics that a subject normalization be performed. A series of papers have already been able to show that there are large subject-dependent differences in the data (Costas, Zahedi, & Wouters, 2014; Hammarfelt, 2014; Haustein, Larivière, Thelwall, Amyot, & Peters, 2014; Zahedi, Costas, et al., 2014; Zahedi & Eck, 2014). With papers published by the Public Library of Science (PLoS), one can already look at the number of views of a paper in comparison with the average views of a subject-specific reference set (Chamberlain, 2013). Plume and Kamalski (2014) have recently presented the

calculation of a field-weighted download impact (FWDI) (along the lines of the mean-citation based procedure described above).

Haunschild and Bornmann (2015) have suggested a paper-side and reader-side normalization of reader counts for Mendeley data. The equivalents among the traditional indicators are the cited-side and citing-side normalization (see the overview in Bornmann & Marx, in press). In the reader-side method, each reader mention of a paper is divided by the average number of readers in the same Mendeley discipline. In the paper-side normalization – in line with the mean-based procedure described above – the number of reader mentions for a paper is divided by the average number of reader mentions in the (WoS) subject category where the paper in question is listed. Haunschild and Bornmann (2015) have applied these methods on the level of institutions and journals as an example. Even if the first results are very promising, the methods and results should be checked in additional comprehensive studies before being applied in research evaluation.

ImpactStory (2014) already offers altmetrics data which is normalized with the help of percentiles (Piwowar & Priem, 2013). If information on disciplines is available for a source of altmetric data (e.g. Mendeley), percentiles are calculated for these data - in a similar way to that described above for citations (http://feedback.impactstory.org/knowledgebase/articles/400281--highly-cited-and-other-impact-badges). If this information is not available for a source, a reference set is built instead from those papers which were published in the same year (i.e. no subject normalization takes place). The reference sets are updated weekly by ImpactStory. Roemer and Borchardt (2013) consider the calculation of percentiles for altmetrics meaningful; however, the data from ImpactStory – according to Roemer and Borchardt (2013) – can only be used for relatively small units: “While the latest version of ImpactStory now includes report feature with some comparative data, such as percentile scores for each metric compared to an appropriate baseline, the limited design of the tool's import feature still makes it impractical for anyone

beyond a small group of researchers to perform analysis on a joint collection of research, let alone the work of an entire department or institution" (p. 16).

Since we can assume that all altmetrics sources display more or less strong subject-specific differences, it would be desirable if corresponding solutions could be found for all (important) sources. "The challenge, therefore, is to create robust, informative standards of context that can withstand minor changes in technology and online scholarly communication. Much more research on the usage of particular publishing platforms and social media networks is needed in order to construct and refine typical threshold levels of attention according to specific groupings" (Liu & Adie, 2013, p. 34).

2.7 Principle 7: Base assessment of individual researchers on a qualitative judgement of their portfolio

Individual scientists should not be evaluated with just a single number, such as the h index, but on the one hand with a wider set of quantitative information and on the other hand by peers (i.e. qualitatively). This point connects with the first principle (see section 2.1): Bibliometric indicators should not be used without the judgement of peers. An example of a set of informative indicators which can be used for the evaluation of individual researchers has been proposed by Bornmann and Marx (2014b): Among other things they suggest determining the number of papers of a scientist that (1) have appeared in different document types, (2) were mainly written by them as the first author, and (3) were written without co-authors. This differentiated view is necessary to allow a fair assessment of the output (in comparison with other scientists). With regard to the citation impact, normalized indicators (besides non-normalized indicators, such as the h index) should be calculated, which permit an evaluation about the citation impact across subject boundaries.

We see it as important with altmetrics too, that the evaluation of a scientist is performed on a wide basis of different indicators (NISO Alternative Assessment Metrics

Project, 2014). Since altmetrics research has only just begun to develop informative indicators (see section 2.6), a meaningful evaluation of scientists will only become possible when these developments are completed in a first phase. As we have already explained in section 2.1, an evaluation on the basis of altmetrics should always take place in the framework of an informed peer review.

2.8 Principle 8: Avoid misplaced concreteness and false precision

The Journal Impact Factor, which is published annually by Thomson Reuters, is given to three decimal places. The advantage of this precision is that the journals can clearly be listed in the order of their average citation impact. The disadvantage of this precision is that the user receives a false impression of the correctness of the Journal Impact Factor. Bibliometric data cannot be correct, since – among other things – (1) the authors who embed citations in their papers make mistakes, (2) imperfect reference lists to the publications are produced by the journals and (3) the database suppliers (such as the WoS) have errors in their bibliometric data. One should therefore either avoid presenting bibliometric data as though they were precise, or publish the data with error bars. For example, the results of the Leiden Ranking (www.leidenranking.com) are shown with stability intervals (Waltman, et al., 2012) and Bornmann, Mutz, and Daniel (2013) suggest presenting bibliometric results with confidence intervals.

As we have shown above, altmetrics are also affected by diverse sources of error. Therefore, providing the results with misplaced concreteness and false precision should also be avoided here. An additional problem with altmetrics at the moment is that it is not clear what is actually being measured: “Since altmetrics is still in its infancy, at the moment, we don’t yet have a clear definition of the possible meanings of altmetric scores … for example in case of Mendeley: what does it reflect when an item is saved/added by several users to their libraries? Also, what does it mean that an item is mentioned in Wikipedia, CiteULike, Twitter

and any other social media platform? We need to study for what purposes and why these platforms are exactly used by different scholars" (Zahedi, Costas, et al., 2014). In a similar way Taylor and Plume (2014) write that it is currently less certain "the underlying nature of what is being measured by current indicators represented … and what can (and cannot) be read into them for the purposes of assigning credit or assessing research impact at the level of individual researchers, journals, institutions or countries" (p. 19, see also Adie, 2014). Therefore, it should also be stated in an evaluation study based on altmetrics, why particular data items are used for the evaluation (and not other data items).

2.9 Principle 9: Recognize the systemic effects of assessment and indicators

Scientists adjust their behavior to the indicators which are used for their evaluation. Bornmann (2011a) has described this behavior adjustment as mimicry. On the one hand, this adjustment is desired: for evaluations, publications in peer-reviewed journals, for example, generally have an especially high weighting. If there is now a trend in science towards publication in these journals, science profits from the fact that more publications have gone through an (international) quality assurance procedure and thus should have a higher quality than those where this has not happened (e.g. book chapters). However, there are also undesired effects of the adjustment of the behavior of the scientists. If pure production figures play an important role in the evaluations, the so-called salami slicing can occur: Scientists distribute their research results across as many publications as possible, to increase the total number (Bornmann & Daniel, 2007).

As the use of indicators changes the system of research, these changes must be taken into account in the interpretation of results from bibliometrics. For example, if a bibliometric study based on the WoS indicates that the number of publications of a university has risen enormously in the social sciences and humanities as compared with the natural sciences, this will probably not have resulted entirely from increased productivity. The scientists from the

area of social sciences and humanities will probably have changed their publication behavior towards publication in internationally peer-reviewed journals (which are easily searchable in WoS).

As soon as altmetrics are relevant in research evaluation, they will produce system effects. For example, in the US National Science Foundation a principal investigator is invited to provide his or her research "products" rather than "publications" in the biographical sketch section. This means that, besides publications, "data sets, software and other non-traditional research products will count too" (Piwowar, 2013). In order to document the impact of these non-traditional research products, altmetrics are already mentioned as an option: "Outreach to the public through social media is one strategy for meeting broader impacts criteria required by the National Science Foundation" (Galloway, Pease, & Rauh, 2013). This relevance in research evaluation will probably expose the sources of altmetrics scores increasingly to gaming by the affected scientists and institutions (Das & Mishra, 2014). "One important aspect of data quality is the potential for gaming metrics, e.g., behavior that is meant to unfairly manipulate those metrics, generally for one's benefit. Many alternative assessment metrics are more prone to gaming compared to traditional citations. Before effects of gaming on alternative assessment metrics can be factored out in computations, the community needs consensus on what behaviors are considered cheating/gaming vs. what is considered acceptable promotion. A better understanding of the factors that make a metric more or less likely to be gamed is needed" (NISO Alternative Assessment Metrics Project, 2014).

With regard to gaming, the suppliers of altmetrics data should develop procedures which allow the identification of manipulated data. Particular kinds of gaming will probably be well able to be identified automatically. "Right now such gaming of the system is rare, but simple to spot both algorithmically; in the case of Twitter spam, where hundreds of fake accounts will suddenly engage in meaningless, random retweets, all of the accounts are quite new, follow each other and have never mentioned a scholarly article before" (Liu & Adie,

2013, p. 33). In time, such gaming attempts will certainly be better disguised, so that one will have to develop better means of identifying such gaming. On the side of the evaluating scientometricians, Priem and Hemminger (2010) regard it as important to undertake a cross calibration of the data in order to identify problematic data. "One particular virtue of an approach examining multiple social media ecosystems is that data from different sources could be cross-calibrated, exposing suspicious patterns invisible in single source".

To what extent system effects other than gaming will appear is difficult to estimate at present. For example, one might imagine that the researchers would adjust their publication behavior to the measurement of broader impact. Thus, for example, more frequent reports on (one's own) research could be written, which are not directed towards professional colleagues but to a broader audience (or a specific audience, like politicians). It can be expected from such reports that they - measured with altmetrics - can achieve a greater impact than with specialist articles (Bornmann & Marx, 2014a).

2.10 Principle 10: Scrutinize indicators regularly and update them

Science is exposed to ongoing change processes which should also be taken into account in the evaluation of research. Thus, for example, researchers of a subject area who have directed their publication behavior over the years to international, peer-reviewed journals, could be evaluated with the help of the WoS database from a particular point in time. If an indicator appears to be unreliable or invalid, it should be replaced by another indicator. As described above, percentiles are now preferred to the mean-citation based procedures, because they are hardly influenced by a few, very highly cited papers. If politics demands a new evaluation of research, the use of indicators must adapt to this: Thus, for example, in the UK Research Excellence Framework, the societal impact of research is also measured (Bornmann, 2012, 2013b; NISO Alternative Assessment Metrics Project, 2014), and reliable indicators are being sought in scientometric research.

We can already see the application of altmetrics for evaluative purposes as a development in the use of indicators, which is adapting to a changed scientific and social system (Bik & Goldstein, 2013). "In the Web 2.0 era, growing numbers of scholars are using scholarly social network tools to communicate scientific ideas with colleagues, thereby making traditional indicators less sufficient, immediate, and comprehensive. In these new situations, the altmetric indicators offer alternative measures that reflect the multidimensional nature of scholarly impact in an immediate, open, and individualized way" (Liu, Xu, Wu, Chen, & Guo, 2013). However, altmetrics do not just reflect the changed environment in which scientific communication takes place, but also - as already explained above - new requirements from the area of politics for a broader impact measurement (Bloch et al., 2014). "As tenure committees and funding agencies begin to demand science that informs policy or provides meaningful change and demonstrated outcomes, altmetrics may change the playing field of how we recognize and reward scientific outputs" (Darling, Shiffman, Côté, & Drew, 2013).

With altmetrics it is first of all necessary for "core altmetrics tools" to be developed (Roemer & Borchardt, 2013), which are regarded by scientometricians as standard. When these cored indicators are available, it will also be important with altmetrics to adapt the understanding and use of indicators on a regular basis to new changes. A good example of this is the (still) changing popularity of the use of social media: "One of the major problems in social media websites is shifting trends in popularity; for example, MySpace used to be the 'hottest' social media website, but it lost this title to Facebook" (Bar-Ilan, Shema, & Thelwall, 2014). If the popularity of the use of a social media tool reduces, one will have to ask the question whether it can and should continue to be used as an altmetrics data source.

3 Discussion

According to Cronin (2013) “neither Twitter mentions nor Facebook ‘likes’ are, for now at any rate, accepted currencies in the academic marketplace; you are not going to get promoted for having been liked a lot, though it may well boost your ego. A robust h-index, on the other hand, could work wonders for your career” (p. 1523). Even if the diagnosis of Cronin (2013) still describes the current state of affairs, it can still be expected that the relevance of altmetrics for research evaluation will change in the next few years. “Altmetrics have already been considered as novel indicators for identifying influential research and for evaluation. However, to date, altmetrics are not being used to, for example, make funding decisions, although some in the altmetrics community are advocating for this approach” (Weller, 2015, p. 265). In a survey of bibliometricians “85.9% thought that altmetrics had some potential in author or article evaluation. The majority (71.8%), believed that the number of article downloads or views could be of use in author or article evaluation … Other sources such as citations in blogs (38.0%), Wikipedia links or mentions (33.8%), bookmarks on reference managers (33.8%), and discussions on Web 2.0 platforms (31.0%) were believed to have potential as altmetrics indicators as well” (Haustein et al., 2014).

It will, however, be important that the implementation and application of altmetrics in research evaluation is always accompanied by scientometric research (similarly to the past and present case in bibliometrics). Only when (1) research reveals what altmetric data really measures, and (2) research has arrived at the first standards for the application of altmetrics (such as the development of field-normalized indicators which are generally accepted), should altmetrics data regularly be applied to research evaluation. However, at present “large-scale studies of altmetrics are rare, and systematic evidence about the reliability, validity, and context of these metrics is lacking” (Haustein, Peters, Sugimoto, Thelwall, & Larivière, 2014, p. 657).

In the later application of altmetrics in research evaluation, it should always be kept in mind that these methods are complementary to and not a replacement for other methods of research evaluation. As explained in section 2.1, these quantitative data should only be used in the context of a peer review procedure. Only peers can recognize the relevance of the data and assess it appropriately against the background of their own experience. The use of altmetrics makes it possible above all to compensate for shortcomings in bibliometrics (Haustein, 2014): "Notable shortcomings include the following: (1) Citations do not measure readership and do not account for the impact of scholarly papers on teaching, professional practice, technology development, and non-academic audiences; (2) publication processes are slow and it can take a long time until a publication is cited; (3) publication practices and publication channels vary across disciplines and the coverage of citation databases, such as the Web of Science and Scopus, may favor specific fields; and (4) citation behavior may not always be exact and scholars may forget to acknowledge certain publications through citations or may tend to quote those papers that are already more visible due to a high number of citations" (Weller, 2015, pp. 264-265).

In this work, the attempt is made to indicate the relevance of the Leiden manifesto for altmetrics. As shown by the discussion of the ten principles against the background of the knowledge about and the research into altmetrics, the principles also have a great importance for altmetrics and should be taken into account in their application. Finally, we would like briefly to refer to or elaborate on some important points of the principles mentioned in section 2:

1) If altmetrics are really to be used for evaluations in the future, they should only be used in the framework of an informed peer review.

2) To obtain informative results, advanced altmetric indicators (normalized indicators) should be used.

3) In an evaluation with the help of altmetrics data, the context of the research unit should be taken into account. Analyses should only be undertaken if they make sense for the unit.

4) The altmetrics data used in an evaluation should be transparent and openly accessible.

5) The presentation of altmetrics data should avoid a misplaced concreteness and false precision, since the underlying data does not permit this.

6) The use of data in research evaluation should be accompanied by continuing scientometric research, in order to optimize altmetric indicators or to develop new ones. But here attention must be paid to system effects of indicators, in order e.g. to recognize gaming.

We consider this point very important, and it should be taken into account in the early stage of the application of altmetrics. Since the data quality and gaming of the data could play a greater role with altmetrics than with bibliometrics, these data in particular should be offered very transparently. “Making all altmetrics data openly available via a standardized API and/or download, a centralized altmetrics data clearinghouse, and audits for altmetrics data are some of the approaches discussed that could improve data quality. For broader acceptance, efforts are required to normalize the data across data providers” (NISO Alternative Assessment Metrics Project, 2014).

We hope that the current work can contribute to the achievement of a sensible and informative application of altmetrics.

References

Adie, E. (2014). Taking the Alternative Mainstream. *Profesional De La Informacion, 23*(4), 349-351. doi: DOI 10.3145/epi.2014.jul.01.

Anon. (2012). Alternative metrics. *Nat Mater, 11*(11), 907-907.

Bar-Ilan, J., Shema, H., & Thelwall, M. (2014). Bibliographic References in Web 2.0. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond bibliometrics: harnessing multi-dimensional indicators of performance* (pp. 307-325). Cambridge, MA, USA: MIT Press.

Bik, H. M., & Goldstein, M. C. (2013). An Introduction to Social Media for Scientists. *PLoS Biol, 11*(4), e1001535. doi: 10.1371/journal.pbio.1001535.

Bloch, C., Sørensen, M. P., Graversen, E. K., Schneider, J. W., Schmidt, E. K., Aagaard, K., & Mejlgaard, N. (2014). Developing a methodology to assess the impact of research grant funding: A mixed methods approach. *Evaluation and Program Planning, 43*(0), 105-117. doi: http://dx.doi.org/10.1016/j.evalprogplan.2013.12.005.

Bornmann, L. (2011a). Mimicry in science? *Scientometrics, 86*(1), 173-177. doi: 10.1007/s11192-010-0222-8.

Bornmann, L. (2011b). Scientific peer review. *Annual Review of Information Science and Technology, 45*, 199-245.

Bornmann, L. (2012). Measuring the societal impact of research. *EMBO Reports, 13*(8), 673-676.

Bornmann, L. (2013a). How to analyse percentile citation impact data meaningfully in bibliometrics: The statistical analysis of distributions, percentile rank classes and top-cited papers. *Journal of the American Society for Information Science and Technology, 64*(3), 587-595.

Bornmann, L. (2013b). What is societal impact of research and how can it be assessed? A literature survey. *Journal of the American Society of Information Science and Technology, 64*(2), 217-233.

Bornmann, L. (2014a). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics, 8*(4), 895-903. doi: 10.1016/j.joi.2014.09.005.

Bornmann, L. (2014b). Measuring the broader impact of research: The potential of altmetrics. Retrieved August 1, 2014, from http://arxiv.org/abs/1406.7091

Bornmann, L. (2014c). Validity of altmetrics data for measuring societal impact: A study using data from Altmetric and F1000Prime. *Journal of Informetrics, 8*(4), 935-950.

Bornmann, L. (in press). Usefulness of altmetrics for measuring the broader impact of research: A case study using data from PLOS (altmetrics) and F1000Prime (paper tags), from http://arxiv.org/abs/1409.2863

Bornmann, L., Bowman, B. F., Bauer, J., Marx, W., Schier, H., & Palzenberger, M. (2014). Standards for using bibliometrics in the evaluation of research institutes. In B. Cronin & C. Sugimoto (Eds.), *Next generation metrics* (pp. 201-223). Cambridge, MA, USA: MIT Press.

Bornmann, L., & Daniel, H.-D. (2007). Multiple publication on a single research study: does it pay? The influence of number of research articles on total citation counts in biomedicine. *Journal of the American Society for Information Science and Technology, 58*(8), 1100-1107.

Bornmann, L., Leydesdorff, L., & Mutz, R. (2013). The use of percentiles and percentile rank classes in the analysis of bibliometric data: opportunities and limits. *Journal of Informetrics, 7*(1), 158-165.

Bornmann, L., & Marx, W. (2014a). How should the societal impact of research be generated and measured? A proposal for a simple and practicable approach to allow interdisciplinary comparisons. *Scientometrics, 98*(1), 211-219.
Bornmann, L., & Marx, W. (2014b). How to evaluate individual researchers working in the natural and life sciences meaningfully? A proposal of methods based on percentiles of citations. *Scientometrics, 98*(1), 487-509. doi: 10.1007/s11192-013-1161-y.
Bornmann, L., & Marx, W. (in press). Methods for the generation of normalized citation impact scores in bibliometrics: Which method best reflects the judgements of experts? *Journal of Informetrics*.
Bornmann, L., Mutz, R., & Daniel, H.-D. (2013). A multilevel-statistical reformulation of citation-based university rankings: the Leiden Ranking 2011/2012. *Journal of the American Society for Information Science and Technology, 64*(8), 1649-1658.
Bornmann, L., Thor, A., Marx, W., & Schier, H. (2015). The application of bibliometrics to research evaluation in the humanities and social sciences: an exploratory study using normalized Google Scholar data for the publications of a research institute. Retrieved March 16, 2015, from http://dx.doi.org/10.6084/m9.figshare.1293588
Chamberlain, S. (2013). Consuming article-level metrics: observations and lessons. *Information Standards Quarterly, 25*(2), 4-13.
Chen, K.-h., Tang, M.-c., Wang, C.-m., & Hsiang, J. (2015). Exploring alternative metrics of scholarly performance in the social sciences and humanities in Taiwan. *Scientometrics, 102*(1), 97-112. doi: 10.1007/s11192-014-1420-6.
Colledge, L. (2014). *Snowball Metrics Recipe Book*. Amsterdam, the Netherlands: Snowball Metrics program partners.
Costas, R., Zahedi, Z., & Wouters, P. (2014). Disentangling the meaning of 'altmetrics': content analysis of Web of Science scientific publications. Retrieved July 9, 2014, from http://figshare.com/articles/Disentangling_the_meaning_of_altmetrics_content_analysis_of_Web_of_Science_scientific_publications/1041770
Cronin, B. (2013). The evolving indicator space (iSpace). *Journal of the American Society for Information Science and Technology, 64*(8), 1523-1525. doi: 10.1002/asi.23041.
Darling, E. S., Shiffman, D., Côté, I. M., & Drew, J. A. (2013). The role of Twitter in the life cycle of a scientific publication. *PeerJ PrePrints, 1*, e16v11. doi: 10.7287/peerj.preprints.16v1.
Das, A., & Mishra, S. (2014). *Genesis of altmetrics or article-level metrics for measuring efficacy of scholarly communications: Current perspectives* (Vol. 3).
Dinsmore, A., Allen, L., & Dolby, K. (2014). Alternative Perspectives on Impact: The Potential of ALMs and Altmetrics to Inform Funders about Research Impact. *PLoS Biol, 12*(11), e1002003. doi: 10.1371/journal.pbio.1002003.
Fausto, S., Machado, F. A., Bento, L. F. J., Iamarino, A., Nahas, T. R., & Munger, D. S. (2012). Research Blogging: Indexing and Registering the Change in Science 2.0. *PLoS ONE, 7*(12), e50109. doi: 10.1371/journal.pone.0050109.
Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2015). Errors in DOI indexing by bibliometric databases. *Scientometrics, 102*(3), 2181-2186. doi: DOI 10.1007/s11192-014-1503-4.
Galloway, L. M., Pease, J. L., & Rauh, A. E. (2013). Introduction to Altmetrics for Science, Technology, Engineering, and Mathematics (STEM) Librarians. *Science & Technology Libraries, 32*(4), 335-345. doi: 10.1080/0194262X.2013.829762.
Glänzel, W., & Gorraiz, J. (2014). Usage metrics versus altmetrics: confusing terminology? *Scientometrics*, 1-4. doi: 10.1007/s11192-014-1472-7.

Hammarfelt, B. (2013). *An examination of the possibilities that altmetric methods offer in the case of the humanities (RIP).* Paper presented at the Proceedings of ISSI 2013 - 14th International Society of Scientometrics and Informetrics Conference.
Hammarfelt, B. (2014). Using altmetrics for assessing research impact in the humanities. *Scientometrics*, 1-12. doi: 10.1007/s11192-014-1261-3.
Haunschild, R., & Bornmann, L. (2015). Normalization of reader impact for Mendeley reader statistics. Retrieved April 2015, 2015, from http://dx.doi.org/10.6084/m9.figshare.1334243
Haustein, S. (2014). Readership metrics. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond bibliometrics: harnessing multi-dimensional indicators of performance* (pp. 327-344). Cambridge, MA, USA: MIT Press.
Haustein, S., Larivière, V., Thelwall, M., Amyot, D., & Peters, I. (2014). Tweets vs. Mendeley readers: How do these two social media metrics differ? *it – Information Technology 2014; 56(5): , 56*(5), 207-215.
Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H., & Terliesner, J. (2014). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics*, 1-19. doi: 10.1007/s11192-013-1221-3.
Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology, 65*(4), 656-669. doi: 10.1002/asi.23101.
Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature, 520*(7548), 429-431.
ImpactStory. (2014). A new framework for altmetrics. Retrieved September 1, from http://blog.impactstory.org/31524247207/
Kurtz, M. J., & Bollen, J. (2010). Usage Bibliometrics. *Annual Review of Information Science and Technology, 44*, 3-64.
Liu, C. L., Xu, Y. Q., Wu, H., Chen, S. S., & Guo, J. J. (2013). Correlation and Interaction Visualization of Altmetric Indicators Extracted From Scholarly Social Network Activities: Dimensions and Structure. *Journal of Medical Internet Research, 15*(11), 17. doi: 10.2196/jmir.2707.
Liu, J. (2014). New Source Alert: Policy Documents. Retrieved September 10, 2014, from http://www.altmetric.com/blog/new-source-alert-policy-documents/
Liu, J., & Adie, E. (2013). Five challenges in altmetrics: A toolmaker's perspective. *Bulletin of the American Society for Information Science and Technology, 39*(4), 31-34. doi: 10.1002/bult.2013.1720390410.
Marx, W., & Bornmann, L. (2015). On the causes of subject-specific citation rates in Web of Science. *Scientometrics, 102*(2), 1823-1827.
Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht, The Netherlands: Springer.
Mohammadi, E., & Thelwall, M. (2014). Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the Association for Information Science and Technology*, n/a-n/a. doi: 10.1002/asi.23071.
Neylon, C. (2014). Altmetrics: What are they good for? Retrieved October 6, 2014, from http://blogs.plos.org/opens/2014/10/03/altmetrics-what-are-they-good-for/#.VC8WETI0JAM.twitter
Neylon, C., Willmers, M., & King, T. (2014). *Rethinking Impact: Applying Altmetrics to Southern African Research*. Ottawa, Canada: International Development Research Centre.
NISO Alternative Assessment Metrics Project. (2014). NISO Altmetrics Standards Project White Paper. Retrieved July 8, 2014, from

http://www.niso.org/apps/group_public/document.php?document_id=13295&wg_abbrev=altmetrics
Peters, I., Jobmann, A., Eppelin, A., Hoffmann, C. P., & Künne, S. (2014). Altmetrics for large, multidisciplinary research groups: A case study of the Leibniz Association. Retrieved July 8, 2014, from https://www.alexandria.unisg.ch/Publikationen/nach-neuen-Publikationen/231796
Piwowar, H. (2013). Altmetrics: Value all research products. *Nature, 493*(7431), 159-159.
Piwowar, H., & Priem, J. (2013). The power of altmetrics on a CV. *Bulletin of the American Society for Information Science and Technology, 39*(4), 10-13. doi: 10.1002/bult.2013.1720390405.
Plume, A., & Kamalski, J. (2014). Article downloads: An alternative indicator of national research impact and cross-sector knowledge exchange. *Research Trends*(36), 13-16.
Priem, J. (2014). Altmetrics. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond bibliometrics: harnessing multi-dimensional indicators of performance*. Cambridge, MA, USA: MIT Press.
Priem, J., & Hemminger, B. M. (2010). Scientometrics 2.0: toward new metrics of scholarly impact on the social Web. *First Monday, 15*(7).
Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: a manifesto. Retrieved March 28, from http://altmetrics.org/manifesto/
Puschmann, C. (2014). (Micro)Blogging Science? Notes on Potentials and Constraints of New Forms of Scholarly Communication. In S. Bartling & S. Friesike (Eds.), *Opening Science* (pp. 89-106). Cham, Switzerland: Springer International Publishing.
Roemer, R. C., & Borchardt, R. (2013). Institutional altmetrics & academic libraries. *Information Standards Quarterly, 25*(2), 14-19.
Rousseau, R., & Ye, F. Y. (2013). A multi-metric approach for research evaluation. *Chinese Science Bulletin, 58*. doi: 10.1007/s11434-013-5939-3.
Sanström, U. (2014). *Bibliometric evaluation of SEPA-funded large research programs 2003–2013*. Stockholm, Sweden: Swedish Environmental Protection Agency.
Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics, 9*(5-6), 281-291.
Shema, H., Bar-Ilan, J., & Thelwall, M. (2012). Research Blogs and the Discussion of Scholarly Information. *PLoS ONE, 7*(5), e35869. doi: 10.1371/journal.pone.0035869.
Shema, H., Bar-Ilan, J., & Thelwall, M. (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, 65*(5), 1018-1027. doi: 10.1002/asi.23037.
Taylor, M. (2013). Towards a common model of citation: some thoughts on merging altmetrics and bibliometrics. *Research Trends*(35), 19-22.
Taylor, M., & Plume, A. (2014). Party papers or policy discussions: an examination of highly shared papers using altmetric data. *Research Trends*(36), 17-20.
Thelwall, M. (2014, January 2). Five recommendations for using alternative metrics in the future UK Research Excellence Framework. Retrieved from http://blogs.lse.ac.uk/impactofsocialsciences/2014/10/23/alternative-metrics-future-uk-research-excellence-framework-thelwall/
Thelwall, M., Haustein, S., Lariviere, V., & Sugimoto, C. R. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. *Plos One, 8*(5). doi: 10.1371/journal.pone.0064841.
Vinkler, P. (2010). *The evaluation of research by scientometric indicators*. Oxford, UK: Chandos Publishing.
Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., van Eck, N. J., . . . Wouters, P. (2012). The Leiden Ranking 2011/2012: data collection, indicators,

and interpretation. *Journal of the American Society for Information Science and Technology, 63*(12), 2419-2432.

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: some theoretical considerations. *Journal of Informetrics, 5*(1), 37-47. doi: 10.1016/j.joi.2010.08.001.

Weller, K. (2015). Social Media and Altmetrics: An Overview of Current Alternative Approaches to Measuring Scholarly Impact. In I. M. Welpe, J. Wollersheim, S. Ringelhan & M. Osterloh (Eds.), *Incentives and Performance* (pp. 261-276): Springer International Publishing.

Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics*, 1-23. doi: 10.1007/s11192-014-1264-0.

Zahedi, Z., & Eck, N. J. v. (2014). Visualizing readership activity of Mendeley users using VOSviewer. Retrieved July 8, 2014, from http://figshare.com/articles/Visualizing_readership_activity_of_Mendeley_users_using_VOSviewer/1041819

Zahedi, Z., Fenner, M., & Costas, R. (2014). How consistent are altmetrics providers? Study of 1000 PLOS ONE publications using the PLOS ALM, Mendeley and Altmetric.com APIs. Retrieved July 8, 2014, from https://speakerdeck.com/mfenner/how-consistent-are-altmetrics-providers-study-of-1000-plos-one-publications-using-the-plos-alm-mendeley-and-altmetric-dot-com-apis