Users, narcissism and control – tracking the impact of scholarly publications in the 21st century

She had a magic looking-glass, and she used to stand before it, and look in it, and say,

'Looking-glass upon the wall,
Who is fairest of us all?'

And the looking-glass would answer,

'You are fairest of them all.'

J. Grimm & Grimm, 2004

About this publication

Users, narcissism and control – tracking the impact of scholarly publications in the 21st century

SURFfoundation
PO Box 2290
NL-3500 GG Utrecht
T + 31 30 234 66 00
F + 31 30 233 29 60

info@surf.nl
www.surf.nl

Authors
Paul Wouters - *CWTS*
Rodrigo Costas - *CWTS*

Contributors
Wouter Boon - *Rathenau Institute*
Jeroen Bosman- *Utrecht University*
Mariken Elsen - *Netherlands Organisation for Scientific Research (NWO)*
Gert Goris - *Erasmus University Rotterdam*
Paul Groth - *VU University Amsterdam*
Henk van den Hoogen - *Maastricht University*
Wilfred Mijnhardt - *Erasmus University Rotterdam*
Ronald Snijder - *Amsterdam University Press*
Maurice Vanderfeesten - *SURFfoundation*
Ludo Waltman - *CWTS, Leiden University*

Editors
Marnix van Berchum - *SURFfoundation*
Keith Russell – *SURFfoundation*

Image on cover
The image on the cover is by 'LostBob'
http://www.flickr.com/photos/lostbob/3194719673/

SURF is the collaborative organisation for higher education institutions and research institutes aimed at breakthrough innovations in ICT (www.surf.nl/en)
This publication is online available through www.surffoundation.nl/en/publications

February 2012

Contents

Executive summary

What is the scientific and social impact of my research publications? This question has been of interest to scientists and scholars since the inception of modern science 400 years ago. But it was hard to answer. This may now be changing. Scholarship is transforming into a variety of digital networked forms. These developments have created new possibilities and challenges in the evaluation of the quality of research. This is of interest to research funders assessing the quality of research. It is also relevant to the individual researchers interested in assessing their career development.

This report explores the explosion of tracking tools that have accompanied the surge of web based information instruments. Is it possible to monitor 'real-time' how new research findings are being read, cited, used and transformed in practical results and applications? And what are the potential risks and disadvantages of the new tracking tools? This report aims to contribute to a better understanding of these developments by providing a detailed assessment of the currently available novel tools and methodologies. A total of 16 quite different tools are assessed.

The report concludes that web based academic publishing is producing a variety of novel information filters. These allow the researcher to make some sort of limited self-assessment with respect to the response to his/her work. However, this does not mean that these technologies and databases can also legitimately be used in research assessments. For this application, they need to adhere to a far stricter protocol of data quality and indicator reliability and validity. Most new tools do not (yet) comply with these more strict quality criteria.

The report therefore advises to start a concerted research programme in the dynamics, properties, and potential use of new web based metrics which relates these new measures to the already established indicators of publication impact. Its goal would be to contribute to the development of more useful tools for the scientific and scholarly community. This programme should monitor at least the following tools: F1000, Microsoft Academic Research, Total-Impact, PlosONE altmetrics, and Google Scholar. The programme should moreover develop the following key research themes: concepts of new web metrics and altmetrics; standardisation of tools and data; and the use and normalisation of the new metrics.

Managementsamenvatting

Wat is de impact van mijn publicaties voor de wetenschap en de samenleving? Deze vraag heeft wetenschappers bezig gehouden sinds de komst van de moderne wetenschap 400 jaar geleden. Deze situatie zou nu kunnen veranderen. De communicatie in de wetenschap vindt nu steeds meer plaats in een verscheidenheid aan digitale netwerken. Deze ontwikkeling biedt nieuwe kansen en uitdagingen voor de beoordeling van de kwaliteit van de wetenschap. Dit is interessant voor onderzoeksfinanciers die de kwaliteit van onderzoek evalueren. Maar het is ook relevant voor de individuele onderzoeker die nieuwsgierig is naar de voortgang van zijn of haar loopbaan.

Dit rapport verkent de ware explosie aan tracking tools die is ontstaan op de golf van tools die webgebaseerde informatie analyseren. Is het mogelijk om real-time te volgen hoe onderzoeksuitkomsten worden gelezen, geciteerd, gebruikt en ingezet worden in praktische resultaten en toepassingen? Wat zijn de potentiële risico's en nadelen van deze nieuwe tracking tools? Dit rapport probeert bij te dragen aan een beter begrip van deze ontwikkelingen door een gedetailleerde analyse en beoordeling van de nieuwe tools en methodologieën die nu beschikbaar komen. Hiertoe worden 16 heel verschillende tools beoordeeld.

Het rapport wijst uit dat webgebaseerd wetenschappelijk publiceren een verscheidenheid aan nieuwe informatie filters oplevert. Deze filters helpen de onderzoeker om beperkte zelf-evaluaties uit te voeren op de reacties op zijn of haar werk. Dit betekent echter niet dat deze technologieën en databases legitiem gereedschap vormen voor wetenschapsevaluatie. Om voor dit doel ingezet te worden moeten ze voldoen aan een veel strikter protocol voor de kwaliteit van de data en de betrouwbaarheid en validiteit van de indicator. De meeste nieuwe gereedschappen voldoen (nog) niet aan deze strengere kwaliteitscriteria.

Het rapport beveelt daarom aan om een samenhangend onderzoeksprogramma op te zetten naar de dynamiek, eigenschappen en potentiële toepassing van nieuwe webgebaseerde meetinstrumenten. Deze kan deze nieuwe meetinstrumenten relateren aan de reeds bestaande indicatoren voor wetenschappelijke impact. Het doel zou zijn om bij te dragen aan de ontwikkeling van beter bruikbare gereedschappen voor de wetenschappelijke community. Dit programma zou in ieder geval de volgende tools moeten volgen: F1000, Microsoft Academic Research, Total-Impact, PlosONE altmetrics, en Google Scholar. Het programma zou bovendien de volgende onderzoeksgebieden moeten ontwikkelen: de concepten van nieuwe webmetrieken en altmetrics, de standaardisatie van tools en data en het gebruik en normalisatie van de nieuwe meetinstrumenten.

1 Introduction

What is the scientific and social impact of my research publications? This question has been of interest to scientists and scholars since the inception of modern science 400 years ago. But it was hard to answer. During a large part of the history of modern science, most researchers could not know who was reading their work. The practical applications were also difficult to track, if at all possible. This may now be changing. Scientific and scholarly publishing is being pushed onto the internet (Jankowski, 2009). Scientists and scholars are routinely using web based applications in their research. In virtually all fields of research, digital web based tools have become indispensable (Dutton, Jeffreys, & Goldin, 2010). This is not the result of a technological revolution, but rather of an evolutionary interaction between scholarly practices, technologies and research infrastructures (Hine, 2006, 2008; The Virtual Knowledge Studio: et al., 2008; Wouters, Beaulieu, Scharnhorst, & Wyatt, 2012). The future is in many ways uncertain, but one aspect seems indisputable: the business models and practices that have become the dominant form of scientific publishing in the second half of the twentieth century can no longer be sustained, at least not in their current form (Roosendaal, Zalewska-Kurek, & Geurts, 2010). Scholarship is transforming into a variety of digital networked forms (Borgman, 2007). These changes are affecting the way researchers work, how they communicate their results and insights, and in what forms these results are codified in an archive of knowledge (Bulger et al., 2011; Gruzd, Goertzen, & Mai, 2011; Williams et al., 2009). The emergence of new paradigms of open access is an important example of how these changes have affected the way scholars think about the future of academic publishing (Willinsky, 2006). These developments have created new possibilities and challenges in the evaluation of the quality of research, also at the level of individual researchers and career developments (Wouters et al., 2010).

This report has been commissioned by SURFfoundation as part of the SURFshare programme and more specifically the work package focused on investigating new ways of measuring the quality and impact of scientific output. This report contributes to this work by exploring the explosion of tracking tools that have accompanied the surge of web based information instruments. Is it now possible to monitor in 'real-time' how new research findings are being read, cited, used and transformed into practical results and applications? At what level can these new tools be used? Will they have the potential to be incorporated in university-wide or even national systems of impact measurements? Are they useful for researchers themselves? And what are the potential risks and disadvantages of the new tracking tools? Do they deliver meaningful indicators of scientific, scholarly and societal performance by researchers? Can the measurements be applied across all fields or are they necessarily limited to particular areas of research? These are the questions which this report addresses.

These questions are not only urgent because of the transformed media of scientific and scholarly publications. They are also urgent because the research communities are under increasing pressure to make clear what their scientific and societal relevance and impact is. The process of quality control and research assessment has become much more important for researchers themselves and for stakeholders in science such as funding agencies, for-profit companies and governments. The increasing scale of scientific and scholarly research has made the research workforce a vital strategic asset for society. The transformation of industry-oriented societies to knowledge societies (in many different forms) has meant that researchers have to meet a larger variety of requirements in evaluation procedures. This is a global development which takes different forms in different countries and regions. In the Netherlands it has been institutionalised in an increasing emphasis on formal forms of research evaluation, in a protocolisation of assessments in a national 6 year cycle of disciplinary and programme assessments in the universities (the Standard Evaluation Protocol), and in an increasing demand for the evaluation of the societal impact of research (KNAW, 2010, 2011). In the context of evaluation, the balance between domain specific expertise and more generic formalised systems of quality control and assessment has been an important problem. It has fueled a debate about the relative importance of peer review versus bibliometrics or citation analysis (Moed, 2005; Smith, 2006). With the advent of new digital forms of publishing, new possibilities have been created to combine qualitative peer judgment with quantitative indicators and measurements.

The new possibilities in scholarly communication and methods also offer new opportunities for analysis and research assessment. This is particularly attractive for those disciplines and areas that have traditionally been excluded from bibliometric analysis since they are not well covered in databases such as Web of Science (WoS) or Scopus, such as the humanities and parts of the social sciences. Moreover, the new forms of metrics also hold the promise of measuring more than only scholarly impact, such as collaboration, societal impact, and valorisation. In short, will it be possible to overcome the limitations of the established forms of performance and publication metrics by alternative web based metrics?

According to information scientist Johan Bollen, we are witnessing ´a Cambrian explosion of metrics´ (Van Noorden, 2010). Over the past 20 years, bibliometrics has shown a ten-fold increase in publications. The field has produced a huge variety of measures and statistics, varying from simple citation and publication counts to sophisticated normalised impact indicators. On top of this, computer scientists have created a host of web based tracking tools which enable every web user to measure the extent to which other web users respond to his activities. Measuring one´s impact and influence has never been more popular. Although these tools have not been created to specifically monitor scientific impact, they can be readily applied in the context of research. For example, researchers may enrich their curriculum vitae and publication lists with impact and usage indicators, perhaps linked to discussions and comments from peers in social media and social bookmarking sites. Adoption of this type of use of metrics promises to diversify and enrich the concept of impact and usage of scientific and scholarly research, thus opening the door to the full development of ´influmetrics´, a concept already suggested in 1995 by Cronin & Weaver (1995).

User-driven innovation in performance metrics also creates a new challenge with respect to the quality of the statistical measurements and their interpretation. Does an increased availability of descriptive statistics also mean that the quality of these statistics is sufficient? How serious is the risk of distributing essentially meaningless numbers and indicators which start to lead a life of their own simply because they are numbers? Paraphrasing Albert Einstein, does everything that can be counted count? In other words, how serious is the risk of adverse effects of these more readily available impact indicators?

This report aims to contribute to a better understanding of these developments by providing a detailed assessment of the currently available novel tools and methodologies. This report does not include an overview of general developments in quantitative science studies and evaluation, such as webometrics, scientometrics, and peer review, since these reviews have already been published (see Moed, 2005; Moed & Glänzel, 2004; Thelwall, 2005; Van Raan, 1988). Instead, we offer a focused analysis of the potential of newly available tools in the context of the literature on established and alternative impact metrics. This assessment is based on the question to what extent these tools can overcome the present limitations of the measurement of publication impact. These limitations are sketched in the next chapter. On this basis, this report surveys and analyses the most important current initiatives in innovation of impact measurement and tries to assess their potential value for the future from a research assessment perspective. In the third chapter, we will present the most important new forms of impact measurement. In this chapter, we also analyse these forms on their strengths and weaknesses. Chapter 4 of this report uses this empirical analysis to assess the potential of the various impact measurement instruments and methodologies for the near future. This section will also draw conclusions.

2 The limitations of established and emerging impact metrics

Currently, peer review and bibliometrics provide the main paradigm for the measurement and evaluation of the impact of scientific and scholarly publications. Both forms of assessment have been created as a by-product of newly emerging systems of scientific and scholarly communication. Peer review has been created as a collegial form of quality control in the course of the establishment of science as a profession. Scientific disciplines have been established as the dominant form of the intellectual organisation of science. Science had an international orientation from the very start. Journals now base their publications on some form of peer review, although the exact form may differ considerably among fields, journals, and countries. Bibliometric evaluation of scientific work on the basis of the numbers of publications and citations became possible after the number of scholarly publications had reached a certain critical mass. After the number of journals and articles had created a demand for abstracting services, a number of database services were created. One of them was the Science Citation Index created by Eugene Garfield. The SCI made it possible for the first time to calculate the citation impact of research articles on a routine basis. Before the creation of the Science Citation Index in the early 1960s, these forms of impact measurement were done by hand. These studies were usually conducted by information scientists in scientific libraries or patent offices. Since the 1980s, citation analysis has become steadily more popular as a measurement tool in the context of science policy and research management.

Because the citation indexes are now available in digital form, citation analysis has been embedded in software code and can be performed on a much larger scale than in the early days of scientometrics. These software codes allow at the same time for a more refined and advanced type of measurement. For example, it is now possible to compare the impact of a particular publication with similar publications, in the same field and of approximately the same age. In the era of printed citation indexes this type of analysis was simply not feasible. These advanced forms of citation analysis do require, however, the availability of the citation index in its entirety. The interface to the Web of Science which is used by most researchers, does not allow for this type of advanced citation analysis, in spite of several improvements in its analytical tools. This type of analysis can only be done by specialised centres of expertise where these indexes are available in the form of relational databases. These centres also provide the data cleaning procedures that are necessary because of the many errors in both the Web of Science and Scopus. Currently, a number of discipline-specific citation indexes are also available, but these do not easily allow for citation analysis across different fields. Moreover, the proprietary nature of the current interdisciplinary citation indexes also put limitations on the further development of open citation index practices. This frames the first question which this report will explore: will new forms of metrics be able to somehow address these current limitations to citation analysis?

The second question will be to what extent the new metrics will solve the limitations of citation analysis in those fields in which journal publication is not the most important form of output. For example parts of the humanities and social sciences publish mostly in the form of books, book chapters, etc. that are not covered by most of the databases used for bibliometric analysis. An important limitation of the current citation indexes is that they are focused on international scientific journals, although they also incorporate some of the national or local journals. The development of citation indexes to books has only very recently started. Will it be possible to address these limitations of citation analysis by using web based impact metrics? Moreover, references to the scientific literature are only one form of acknowledgement. There are other linkages between scientific articles embedded in the full text, such as in footnotes, acknowledgment sections and the main text of papers that are not incorporated in citation analysis (although recently WoS has started to collect data on 'Funding Acknowledgments' from research papers, cf. Costas & Leeuwen, 2011). Is it possible to data mine the scientific and scholarly literature in a more comprehensive way?

The third question which this report addresses is whether alternative metrics may enable to address the more fundamental limitations of impact measurement by citation analysis. Citations are not a direct measurement of the realised impact of a publication. There is a variety of reasons for researchers to cite a particular work. Works can be cited because they have influenced the author, but these citations may also be given for symbolic or social reasons. Moreover, the relative visibility and network position of an author may strongly influence the extent to which he is cited. Citation is also a probabilistic process: a work may remain uncited simply because a closely related publication attracts all the citations. Or the citations happen to flow to a review in which the article has been discussed. Presenting the decade-long debate about the meaning of the citation goes beyond this report, but it is clear that citations are no direct measurements of either quality or the realised impact. They are best defined as proxy measures of impact. This raises the question to what extent other metrics can supplement citations as impact indicators. For example, can impact be measured or predicted by usage statistics? A closely related question is whether it is possible to develop impact indicators that follow the most recent developments in a timely manner. An important limitation of citation analysis is that it takes time for an article to collect citations. As a result, reliable citation statistics are often available a few years after the publication has appeared. Is it possible to shorten this measurement window? And will new metrics allow for the measurement of more dimensions of impact than is currently possible through citation analysis?

This last question is especially relevant with respect to the impact of scientific research in society. This may take different forms, such as the use of scientific research to create new products and services (valorisation), the influence of research on the public opinion, the utilisation of research in professional practices (e.g. new clinical protocols and tools) and the way research is implicated in cultural heritage practices (e.g. museums). It is quite difficult to measure this type of impact. It may take many years before research bears fruition in these ways. It is often difficult to backtrack from particular innovations to the underlying research. And many different actors are involved in these interactions. Will monitoring these ´productive interactions´ in databases and on the web enable the more precise measurement of this form of impact?

This report addresses these questions by an empirical assessment of the currently available new information tools for the academic community. As far as we know, this is the first time such a thorough analysis of these tools has been made. As things go, the literature is littered with articles that praise the virtue of the newly available instruments. However, more often than not, these recommendations are more representative of the enthusiasm naturally surrounding inventions than of the real potential of the relevant tools (e.g. Priem & Hemminger, 2010). We will come back to this literature in the concluding chapter.

3 Novel forms of impact measurement – an empirical assessment

This chapter presents an overview of the different tools available for measuring the impact of scientific publications. We describe the main possibilities they offer to measure research performance, focusing on their main characteristics and potential limitations for research assessment. The main purpose of this analysis is to detect and describe the potential possibilities and problems of these tools.

We have applied the following method:

- We performed a comprehensive search for different online tools, starting with general search engines like Google, selected publications that have already tackled the problem, wikis, discussions with experts, etc.
- Next, we present a general description of the source/tools found, describing the main characteristics, limitations and possibilities that they have for research assessment.
- Regarding the possibilities for research assessment of the different tools we assume as positive assets the following properties:
    - The tools/sources should allow a certain extensive use for real existing research assessment processes. Thus, they shouldn't be limited only to authors, journals, countries, etc. but be open to use at several aggregation levels and with different analytical possibilities.
    - They should allow the download and proper management of data in order to permit the users to create their own frameworks, analysis and approaches.
    - Possibilities of normalisation of indicators (by disciplines, document types, etc.) are be regarded as very positive.

The potential of these tools for evaluations that have less strict requirements (such as a quick lookup of their impact by authors themselves) will be discussed in the last chapter, where we will explore how 'technologies of narcissism' may interact with 'technologies of control'.

3.1 F1000 – Faculty of 1000

http://f1000.com/

Faculty of 1000 is a commercial service launched in 2002 for researchers and clinicians that provides ratings of and commentaries on scientific research papers. The service acts as a filter, identifying and evaluating the most significant articles from biomedical research publications. A peer-nominated 'faculty' of scientists and clinicians rate the articles they read and explain their importance. These faculty members select and evaluate important papers in their areas of expertise, usually tagging them as well to further classify papers as clinical trials, novel drug targets, etc.

F1000 rates research articles on their own merits rather than according to the prestige or impact factor of the journal in which they are published. The Faculty of 1000 service also includes F1000 Biology Reports and F1000 Medicine Reports journals, which present peer-reviewed commentaries on emerging themes in biology and medicine. These open access journals are currently indexed and/or abstracted by Scopus, Embase, Global Health, CAB Abstracts and PubMed Central, among others.

In 2010, F1000 Posters (F1000Posters.com), an open access repository of posters and presentations from international conferences across biology and medicine was added to the F1000 service, and in 2011 F1000 launched a new journal ranking system. The F1000 Journal Factor (FFj) is a measure of how well an individual journal is performing on F1000. It is calculated from the individual F1000 Article Factor (FFa) values and normalised according to the total number of

eligible articles each journal has published. Journals are ranked by FFj down to subfield level. These new indicators still need to be properly studied and tested, especially regarding aspects such as consistency, size-dependency and other mathematical and statistical properties related to indicators (Huggett, 2012)

Characteristics

- F1000 publishes over 1500 evaluations each month. These cover all aspects and disciplines within biology and medicine. Although they evaluate articles published in popular and high-profile journals (Nature, New England Journal of Medicine, Science, Journal of Experimental Medicine, etc.), around 85% of their evaluations come from more specialised or less well-known journals. As of May 2011, articles from more than 3500 different publications have been evaluated.
- Search possibilities include author, journal title, issue, volume, DOI, etc. It is also possible to search by evaluator.
- It is possible to download the data and import it into EndNote, RIS, RefWorks, Bibtex, ProCite and Sente. But the information of the reviewers and their comments is lost in the data transfer.
- Rankings of journals and papers are presented based on their indicators (rankings include the top 10 rated papers, 'Hidden Jewels', annual ranking of journals, among others).

Limitations

- It is not a free tool. A free trial is possible but only for a limited period of time.
- The coverage of the database is limited to medical and biological journals.
- Most of the papers from the journals covered are actually not reviewed (Butler, 2011). This is something that F1000 explicitly mentions. In the rankings of journals the 'eligible' articles per journal are presented and in most cases they exceed the number of articles actually reviewed (the number of reviews that a paper receives is part of the formula for the rating of papers and journals). This raises the question whether it should be assumed that papers without reviews or recommendations are of lower quality or only of less interest or just overlooked by the experts involved in the system.
- There are also problems with the search of journals with abbreviated titles. The titles of the journals are not always standardised.
- It is not possible to filter and search by document type. Also, document types do not seem to play a role in the calculation of the indicators.
- It does not seem to provide for vocabulary control or author/address data standardisation.
- Normalisation of indicators by sub-fields (necessary to be able to compare the values of indicators across medical and biological subfields) is not provided.
- Problems related with the manipulation and gaming of the indicators, as well as with the selection of experts have been also pointed out for this tool (Butler, 2011).

Possibilities for research assessment

The data and indicators provided by F1000 are without doubt rich and valuable, and the tool has a strong potential for research evaluation, being in fact a good complement to alternative metrics for research assessments at different levels (papers, individuals, journals, etc.).

However, the fact that this information and the indicators are currently only limited to a few scientific areas (the medical and biological sciences) reduces its value for extensive research assessments. On the other hand, unlike other tools, the F1000 site gives a standard framework for analysis and research assessment as they consider all the publications from a given set of journals.

Another important limitation of the tool is that it is not a free tool and after the trial period a monthly fee of $9.95 has to be paid for access to the data.

Finally, we would like to mention the importance of other more conceptual problems related to this type of information and the potential help that they can provide in research assessments at all levels. Some questions that should be answered (or at least addressed) before this tool can be fully introduced as a standard element for research assessment are:

- From a research assessment point of view, what is better? To be reviewed (even if the outcome of the review is only mild) or not to be reviewed at all? (Thus making true the quote by Mae West that ´it is better to be looked over than overlooked´).
- In line with the previous statement, why do reviewers choose to review some papers and not others? Is it because they like them or are there other factors? Do papers that have been already reviewed attract more reviews?
- How does the reviewing process develop over time? Is it important to be reviewed positively soon after publication in order to attract more reviews and perhaps also citations? How long can it take for an average paper to be reviewed?
- Are the reviews and metrics based on this type of peer review actually a fundamental alternative measure or are they strongly correlated with more traditional measures of impact of publications? Could the number of citations received by papers be influenced by the reviews given before?
- How is the final selection of these ´faculty´ reviewers done? Can biases exist in the selection of the experts and their reviews (e.g. by countries, language, age, affiliation, scientific sub-specialty, network effects, etc.)?
- To what extent are the traditional limitations attributed to peer review assessment (e.g. subjectivity, ´invisible colleges´, ´old boys networks´, etc.) corrected for by this more open type of peer review?

In any case, we can conclude that F1000 is without doubt a very relevant tool that could support the research assessment at different levels of aggregation.

3.2 Peer Evaluation

http://www.peerevaluation.org

Peer Evaluation is a tool that seeks to provide open access to primary data, working papers, articles and media of authors allowing them to be reviewed and discussed by other peers.

Peer Evaluation supports qualified peer review. In this respect, its aim is to become a valuable supplement, inspiration and hub for peer reviewed journals and publications. The basic idea of the tool is to complement quantitative metrics by a new set of qualitative indicators that are comprehensive, transparent and immediately verifiable by researchers and funding institutions.

Characteristics

- A user must upload their publications in the system. Therefore it is of greatest interest to registered users (non-registered users will not appear).
- The system allows the user to search for other people that have included their profile in the system.
- Some metrics are presented at the individual level (e.g. ´trusted by´ , ´discussed by´, ´cited by´, ´downloads´, ´views´, etc.) although clear explanations on how these indicators and data are collected and calculated is not provided.
- Searches by persons are offered, which gives the possibility to freely access the publications that they have been uploaded to the system and ´view´, ´download´, ´disseminate´, etc. them. But no statistics at the paper level are provided. Indicators are only aggregated by person.

Limitations

- The interface is not always intuitive and sometimes it is difficult to work with. For example, if one incorporates a paper wrongly in one´s profile, this mistake cannot be corrected.
- It is difficult to search for data on individual papers, journals, affiliations, DOI, etc. The interface is mostly oriented towards the search for people rather than publications.
- There are no clear options to export data to other formats.
- The tool is oriented towards the individual level. This makes it hard to use for more integrated analysis and data collection.

Possibilities for research assessment

This system to some extent resembles the F1000 tool. In some respects it could be considered the free open version of the F1000 system. The main difference is that in this case all the other registered users of the system are potential reviewers for all the publications incorporated in the system. However, although the possibilities are similar to those of F1000, there are important limitations that do not allow using this tool extensively for research assessment in its current stage of development. Some of these critical limitations are the following:

- The fact that it is only centered on individuals and that data for other levels of analysis (e.g. papers, research organisations, journals, etc.) is not available, limits the use of the tool only for individual assessments.
- Most of the papers remain without any review or comments, although clear statistics on the number of papers not reviewed are not provided.
- There is an important lack of information on the number of users, their origin (e.g. country, field, language), etc.
- The lack of normalisation of the results jeopardizes the comparability and interpretation of the results.
- The problems associated with managing, downloading and working with the data prevent even more expert users (e.g. bibliometricians, information scientists) to access the data and use them in a more standardised framework.

In addition it can be added that the same conceptual limitations as suggested for the F1000 tool are also relevant for this tool.

3.3 PaperCritic

http://www.papercritic.com

PaperCritic aims to offer researchers a way of monitoring all types of feedback about their scientific work, as well as allowing them to easily review the work of others, in a fully open and transparent environment. It is related to the other two previously described systems (F1000 and Peer Evaluation).

Apart from helping the scientific community in rating and reviewing publications, PaperCritic also seeks to help researchers to organise their publication library, in this sense also working as reference management system.

Characteristics

- PaperCritic is a free system based on Mendeley's data (see the description for Mendeley below) combining its characteristics with elements such as tags, summaries and in-text notes. The system works with an API to access the data and it is acknowledged that it depends on the characteristics and validity of Mendeley's data.
- It allows for rating and reviewing of any publication in the library created by any of the users.
- Only registered users can post reviews or comments on the papers in Mendeley.

Limitations

- Most of the publications are not rated, reviewed or commented at all.
- Searches of publications are limited to authors and titles. Other searches (DOI, journal title, year of publication, document type, affiliation, etc.) are not allowed, thereby limiting the capacity of data retrieval.
- No downloading options are presented in the tool.
- It is not possible to search by reviewer although they are identified in the record of the paper.
- Normalisation possibilities are not available.

Possibilities for research assessment

This tool presents similar functionalities and aims as F1000 and Peer Evaluation. However it has stronger limitations (e.g. lack of searching and filtering options, lack of downloading options, lack of metrics at other levels than the paper level, etc.) as compared to the other two tools. These limitations prevent its use as a valid research assessment tool in its current state of development. However, if some functionalities and more transparency are added to the system, it could become an interesting add-on tool to the data presented in Mendeley.

The conceptual problems highlighted for the two previous tools also apply for PaperCritic.

3.4 Google Scholar

http://scholar.google.com

In November 2004 Google Scholar (GS) was launched by Google Inc. Since then, its interest among the scientific community has grown increasingly. This tool provides a simple way to broadly search for scholarly literature. From the search tool it is possible to search across many disciplines and sources, including journal articles, theses, books, abstracts and other documents from academic publishers, professional societies, online repositories, universities and other web sites. For every document listed in GS, information about the citing documents is presented, allowing the user to detect other publications related to the first one.

Characteristics

- It is a free product powered by the well known search engine Google.
- Broad coverage:
    1. Broad typology of covered sources: databases, scientific societies, online library catalogs, repositories from research institutes and organisations, etc. (Broader than regular bibliographic databases).
    2. Broad typology of documents: preprints, journal articles, books, thesis, reports, conference proceedings, etc. (Broader than in regular bibliographic databases).
    3. Better coverage of non-English literature (than in regular bibliographic databases).
    4. Full text access if the document is free and available in open access.
    5. Linkage with library services in order to access the text of the publication.
    6. Collection of citations that are not covered by other bibliographic databases (citations given by preprints, thesis, etc.). That gives the advantage to GS in the analysis of humanities, social sciences and engineering sciences.
    7. GS is developing functionalities (search by author, by year, etc.) and solving problems reported in the past. For example ´introduction´ was an author in some publications in the past, now it has been solved for most of the cases).

- Simple interface for search and presentation of data:
    1. Easy interface for searching for a bibliography.
    2. Ability to detect different versions of a document and group them under the same title heading.

3. Harzing's Publish or Perish software[1] allows working with sets of up to 1000 documents and the calculation of a limited set of bibliometric indicators (mostly size-dependent and based on the h-index).
4. Possibility of paper per paper download to reference manager software (Endnote, BibTeX, RefMan, RefWorks).

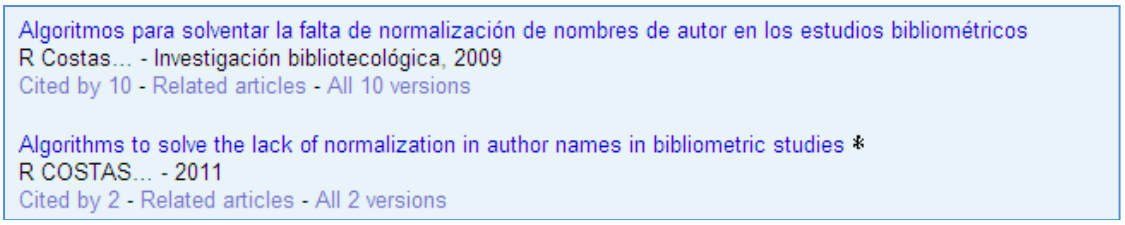
Limitations

- Coverage:
  1. Lack of transparency on the coverage and it is not clear if they follow a systematic coverage. For example, some of the journals and books covered are included under agreements with the publishers, agreements that are not explicitly mentioned by GS.
  2. Coverage of documents that are not purely academic or scientific: library guides, text books, teaching materials, etc.
  3. No quality control of the journals and sources covered in the system. Everything can be included. Besides, the user does not have the possibility to establish any quality control filters (e.g. selecting only peer reviewed articles, books, etc.).
  4. Deficient coverage of journals in humanities and social sciences that are covered in other databases (MLA Bibliography, Philosopher's Index, PsycInfo, Sociological Abstracts).

- Interface, search and results:
  1. Limited fields for searching bibliography (full text, title, author, journal and publication year), and no possibility of searching by address information.
  2. The results are presented in an already sorted way and this sorting cannot be changed.
  3. Exporting to other reference manager software is very limited (only paper by paper and not in an extensive way).
  4. Presentation of some duplicated results (e.g. the search of ´Scientometrics in 2008´, gives 131 results in WoS and 153 in Google Scholar).
  5. It has a limitation of 1000 documents (e.g. the Publish or Perish software with searches of more than 1000 documents does not work).
  6. No possibility of searching by cited documents and no possibility of downloading data in a systematic and extensive way.
  7. No API possibilities. This makes impossible to access, download, store and manage the information in a more ad-hoc and user oriented way.

- Normalisation & standardisation:
  1. Data and document types are not indicated (only books are identified).
  2. No possibility at all of obtaining normalised indicators (citation or references per field, etc.).
  3. Volatility. The database is continuously changing and there is no possibility (with the standard tools) of replicating previous results.
  4. No disciplinary organisation or scheme. Only 7 large disciplines are included for searching purposes.
  5. Self-citations are not tackled and they cannot be excluded.
  6. No data standardisation. For example, author names are not deduplicated, the address information is not collected, and journal titles can appear in different ways (e.g. JAMA and *Journal of the American Medical Association*).
  7. No vocabulary control, keywords are not available, etc.
  8. Presence of duplicates of papers in non-English language journals but with a copy of the title in English. See example:

[1] http://www.harzing.com/pop.htm



9. Key information present in the paper is not processed by GS (e.g. abstract, keywords, acknowledgments, cited references, email information), although the full text when available is searchable.
10. No unique identifiers of documents (at least not accessible) and no possibility of searching by DOI (this would help limit duplications).
11. Possibilities of manipulation through "engine spam" have been pointed out (Beel & Gipp, 2010).
12. Difficult and time consuming to clean the data and obtain any meaningful results (even if restricted at individual level).

Possibilities for research assessment

- In general GS is suitable for bibliography searches on one´s own topic in a much broader data set than in other databases (WoS, Scopus), and with the possibility of detecting related papers through the links of citations.
- GS also allows access to any free open access copy of the paper that the user is looking for.
- One can obtain rough indicators of production and impact of oneself, ones colleagues and other people one knows (´technologies of narcissism´). It is not possible to do other searches (and download) of data of some bigger sets of publications.
- Suitable for small-scale analysis under certain conditions (e.g. a few authors with very uncommon names).

However, GS is neither useful nor suitable for the following types of analysis:

- To collect the publications of a big unit of analysis or a big set of researchers in a systematic way. It is impossible to carry out a systematic and broad analysis of several units of analysis. This limitation also prevents the potential bibliometric user to download data and perform their ad hoc standardisation using normalisation methodologies.

- To perform broad and normalised studies. Given the fact that it is not possible to limit the search by document types, establish citation windows, detect self-citations, etc., it is not possible to establish a normalised framework of analysis that would allow systematic research assessment possibilities.

- To obtain field-normalised indicators. It is also not possible to obtain field-normalised indicators as it is not possible to consider the whole database and all the cited references by the papers, the lack of a classification of journals and/or of papers, apart from other key information, that would allow a minimum normalisation of the indicators.

- One cannot use GS for monitoring purposes due to the lack of stability of data (e.g. older documents can be incorporated in the web at any time, etc.).

Regarding all the previous limitations of GS, we highlight three main limitations that limit the potential of GS for research assessment purposes:

1. Impossibility of downloading and managing comprehensive data.

   Without doubt this is the most important limitation of GS as a potential research tool for research assessment purposes. It prevents any potential use of GS as a tool for research assessment in a systematic way (beyond the "narcissistic" approach). This limitation is very important, because even considering the other limitations of GS, if it would allow bibliometric experts to download and work with the data in a broader way, it would be possible for these experts to establish the limitations and possibilities of the data and the results (by limiting the data themselves, creating standard frameworks of analysis, etc.).

2. Improvement of the transparency of the coverage.

   The limited documentation of the coverage is the second most important limitation. The users of the database and the potential research assessment results must be aware of what they can expect to be included and what not. Of course the aim of GS is to be as comprehensive as possible, but explaining how the information is collected, the missing sources, the agreements that have been concluded with publishers, etc. would contribute to increasing the transparency of the coverage of the database and thus to establish its possibilities (and limitations).

3. Lack of parsing and quality control of the data.

   Third, an improved parsing of the bibliographic data contained in the documents covered by the database (authors, journal titles, address information, acknowledgments information, etc.) would increase the quality and usability of the data for bibliometric and research assessment purposes. It would also increase the filtering and normalisation possibilities of the data finally considered for the assessment and the creation of stable and comparable frameworks for the development of research assessment exercises.

3.5 Google Citations (GC) – Google Scholar Citations

http://scholar.google.com/citations

GC is a new service recently launched by Google. It is completely based on GS data. It seeks to provide a simple way for authors to keep track of citations to their articles, providing also some basic indicators (total number of citations, h-index, and i10-index). Basically it consists of the aggregation of GS publications into profiles of individual scientists. In order to have a profile, the author needs to create this herself and they must have a (freely available) Google account.

Characteristics

- Individual level based information. If the profile is publicly available it means that the authors have accepted to make it public, and some 'check' of the list of publications by the authors themselves can be assumed (although the thoroughness of this review may vary).
- Profiles are composed of GS publications and citations.
- The authors can check the publications themselves (it is possible to merge duplicates, remove wrong papers, correct mistakes and add missing publications if they are in GS).
- It is possible to export data to reference manager software (Endnote, RefMan and BiBteX).
- It is possible to update ones profile with photo, affiliation information, homepage, e-mail, etc.
- It automatically provides some standard indicators: h-index, citations (and a graph with its evolution over time), i10-index (the number of papers that received 10 citations in the last 5 years).
- It provides keywords of the main topics of research of the researcher, although it is not clear how these keywords are chosen (although the user can change the keywords if necessary).

Limitations

- Information is restricted to Google users (if one doesn't have a Google account one cannot create a profile). This may mean that many authors are currently not aware of this profile possibility.
- The information that can be accessed is limited to the persons (and profiles) that have been made public, otherwise they cannot be accessed. It is not possible to search for persons who don't have a Google account and/or who have not made their profiles publicly available.
- It is not possible to access and download large data sets. It is possible to download the full individual profiles of the authors that have made their profiles public.
- Some exporting limitations still remain, and although it is possible to export the data to RefMan, BibText and Endnote, the citation data in the information downloaded is excluded. Therefore it cannot be used to download several bibliometric profiles and to store the data for analytical purposes.
- Possibilities for filtering, sorting, etc. are not offered. Publications are sorted by number of citations, and they can only be sorted differently by publication year and title/author. Other filters and sorting possibilities are not available (e.g. by document type, journal title, co-author, etc.).
- The indicators offered in the website are very limited. They are size dependent and mostly h-index based. This makes these indicators problematic for most research impact measurements.
- Other indicators could be created based on the information provided by GC but they are difficult to get and need manual access (e.g. the total number of publications, the average number of citations per publication, etc.). In addition, it is explicitly stated by GC that they do not plan to include more metrics[2]. So, we don't expect changes in the indicators offered in the short-term.
- Lack of field normalisation (or any other type of normalisation) in the indicators.
- The quality and consistency of the data included in the individual profiles may vary. It all depends on the criteria of the user. If the user wants to merge some publications, nobody is going to check if this merge is correct. The same holds in the case the user splits the publications. This implies that there are some possibilities of "bibliometric engineering".

Possibilities for research assessment

The possibilities of GC as a tool for extensive, standardised and systematic research assessment are indeed very limited (in line with the lack of possibilities for GS). The tool could be useful to analyse individual researchers, but mainly for self-assessment (e.g. ´how good am I compared to some of my colleagues that have a public profile and whom I assume to have a similar scientific behaviour to mine?´).

Again, as the information is restricted to individual researchers, and it is not easily possible to download and manage (no API options), the use of this tool in a consistent and systematic way in a serious research assessment process is problematic.

Moreover, in general GC also suffers from all the limitations and restrictions mentioned for GS.

3.6 Microsoft Academic Search (MAS)

http://academic.research.microsoft.com/

Microsoft Academic Search (MAS) is a free academic search engine developed by Microsoft Research, which also serves as a test-bed for many research ideas in the areas of data mining, named entity extraction and disambiguation, data visualisation, etc. As a research prototype, the coverage of MAS is still limited to certain domains. It covers more than 36 million publications and over 18 million authors across a variety of domains with updates added each week. This large collection of data has also allowed users to create several innovative ways to visualise and explore

[2] According to the website: "while there's no shortage of other reasonable metrics, the incremental usefulness of adding each number generally goes down, while the user confusion generally goes up".

academic papers, authors, conferences, and journals (e.g. Academic Map[3], Co-author graph[4]and, Citation graph[5]).

Microsoft Academic Search services include full paper search, author search, conference search, journal search and a visual explorer. These are based on the search technology Microsoft has developed. The search results (including displaying, ranking) are the results of applying this technology to publicly available information. Microsoft does not warrant the accuracy of the search results. According to MAS one should exercise ones own discretion and judgment when using Microsoft Academic Search.

Characteristics

- All authors can be searched. It is not necessary to have an account or being registered in MAS.
- APIs (application programming interfaces) are possible although with some restrictions.
- Several indicators can be obtained (Publications, Citations, g-index and h-index), as well as network figures of co-authors and citers.
- The navigation is attractive, although somehow limited to the search of authors or particular publications.
- It is possible to search by: authors, journals, DOI, and research organisations (addresses).
- There is some normalisation by organisations. It is even possible to compare research organisations in pairs. In such comparisons the persons working for those organisations are displayed as well.
- Data can be exported to reference manager software (EndNote, BibText and RefWorks), although the highest level for downloading publications is the author-level (it is not possible to download the publications of a research organisation or a country).
- Some indicators are displayed for the different authors, journals and organisation, most of them size-dependent (Publications, Citations, h-index, g-index).
- The coverage of journals is quite accurate: in a test search for ´Scientometrics – 2008´, we obtained 135 results in MAS compared to 131 results in the Web of Science; ´JASIST´ gave 221 results in MAS and 227 in WoS. This suggests that duplicates are tackled (at least better than in GS).
- Information on document types and keywords is provided.
- The profiles of the individual papers in some cases provide information on the "citation context" of the citing papers, which indicates the part of the text that was cited. This is a quite interesting feature because it enables the user to link bibliometric information to substantive information.

Limitations

- The identification of persons is less accurate than in GC. Authors tend to appear in more than one profile (e.g. Thed N. van Leeuwen; Marc de Jong), especially if the authors have used different names and/or have changed their research affiliation.
- Chinese and very common names show important mistakes. For example Wei Chen of the university of Eindhoven appears with more than 600 publications –which is not correct - and a wrong photography. In reality she is a lady and in MAS a man's photograph is shown.
- Inaccuracy of the author-address linkage (e.g. that Eugene Garfield is linked to the Royal Netherlands Academy of Arts and Sciences (KNAW)).
- The coverage of citations is lower than for GS (or even Web of Science). In general fewer citations are counted for the publications.
- The possibility of using APIs needs to be requested first. Anything that one builds based on MAS must be for non-commercial use only or it must be made available in a free version of one's product. Also, each API call will return only 10 items per call. MAS explicitly mentions that one cannot use the API to crawl the entire corpus (this hinders the development of normalised indicators).
- It is not clear if everything that appears in MAS is actually open access, or if they also include publications that are mentioned in other bibliographic websites (e.g. publications in DBLP[6]). If

[3] http://academic.research.microsoft.com/AcademicMap
[4] http://academic.research.microsoft.com/VisualExplorer#922315
[5] http://academic.research.microsoft.com/VisualExplorer#922315&citation

this is the case it would be more difficult (than in GS) to differentiate what is open access from what is not.
- Also the policy of coverage is not clear and the journals, publishers, repositories, etc. that are covered are not explicitly mentioned.
- It is not possible to carry out searches by countries.
- When the information is downloaded to the reference manager software, the information on citations is lost.
- The information on indicators is limited to size dependent indicators (publications, citations, g-index, h-index), although other indicators such as citing authors, number of co-authors, etc. are also displayed. Anyway, the calculation of other indicators (such as the number of citations per publication) would be quite straightforward.
- Only 16 fields are considered and it is not clear how the publications are assigned to these fields.

Possibilities for research evaluation

In spite of the limitations previously mentioned, MAS is the tool with more potential possibilities for research evaluation. It allows APIs that would permit bibliometric researchers to download big sets of data and treat them properly.

MAS seems also to be more open to collaboration with external partners than GS and they allow the tool to be researched and used. These options should be tested before final conclusions can be drawn.

There are also some other potential interesting possibilities in MAS that are not present in GS.

- MAS collects the data of individuals and research organisations. This would allow obtaining measures of average productivity at the level of research organisations (something that is not possible with any of the other tools, including WoS and Scopus).
- The inclusion of the DOI of publications makes the data provided more exchangeable and possible to match with other sets of documents and databases.
- The linkage of authors and organisations would allow interesting bottom-up comparisons that at present are not available in GS (or in any other commercial database). It seems that they haven't yet realised this potential, but is clearly there.
- The inclusion of the "citation context" for some of the publications is very promising. It would allow not only to count the citations given, but also to assess the reasons and the context of the citation (see example):

[6] http://www.informatik.uni-trier.de/~ley/db/welcome.html

However, currently MAS has three main limitations for a practical use as a research assessment tool:

1. Lack of transparency in how the data is collected
2. The coverage is less exhaustive compared to GS or WoS and still very limited in certain domains. This is a strong limitation of this tool.
3. Lack of standard tools for downloading data beyond the author and paper levels. Moreover, the number of citations is lost when the data is downloaded to reference manager software.

3.7 Arnetminer

http://arnetminer.org/

Arnetminer is a tool initially designed as a research project in social influence analysis, social network ranking, and social network extraction to search and perform data mining operations against academic publications on the Internet. It uses social network analysis to identify connections between researchers, conferences, and publications. This allows it to provide services such as expert finding, association search, course search, academic evaluation, and topic modeling. Assessment is one of the targets of the tool (especially through the development of the so-called "Academic ranks").

The system indexes 700,000 researchers, 6,000 conferences and 3,200,000 publications. The system attracts users from more than 200 countries and receives more than 200,000 accesses per day.

Arnetminer aims to be used in academia to identify relationships between, and draw statistical correlations about, research and researchers. The product was used in a study aimed at verifying the popular notion that no more than six degrees of separation connect any two people on Earth.

Characteristics

- Arnetminer is a mining tool that aims to provide comprehensive search and mining services for researcher social networks. In this system, the main focus is on:
  1. Creating a semantic-based profile for each researcher by extracting information from the distributed web;
  2. Integrating academic data (e.g., the bibliographic data and the researcher profiles) from multiple sources;
  3. Searching the heterogeneous network;
  4. Analysing patterns from the constructed researcher social network.

- Arnetminer offers the following possibilities of searching for information:
  1. Profile search: on the basis of the input of a researcher's name, the system will return the semantic-based profile created for the researcher using information extraction techniques. In the profile page, the extracted and integrated information include: contact information, photo, citation statistics, academic achievement evaluation, research interest, educational history, personal social graph, research funding (currently only US and China), and publication records (including citation information, and the papers are automatically assigned to several different domains).
  2. Expert finding: on the basis of the input of a query (e.g., data mining, bibliometrics), the system will return experts on this topic. In addition, the system will suggest the top conference and the top ranked papers on this topic. Users can also provide feedbacks to the search results.
  3. Conference analysis: with the input of a conference/journal name the system returns who are the most active researchers on this journal, and the top-ranked papers. However, it is not possible to get all the publications published in the journal (at least not in a straightforward fashion).
  4. Course search: input a query (e.g., 'data mining'), the system will tell you who are teaching courses relevant to the query.
  5. Associate search: input two researcher names, the system returns the association path between the two researchers. The function is based on the well-known 'six-degree separation' theory.
  6. Topic browser: it is possible to search for topics and keywords. Furthermore, the browser presents the most active researchers, the most relevant conferences/papers, and the evolution trend of the topic is discovered.
  7. Academic ranks: the tool provides eight measures to evaluate the researcher's achievement: 'h-index', 'citations', 'uptrend', 'activity', 'longevity', 'diversity', 'sociability', 'New Star'. For each measure, the tool outputs a ranking list in different domains.

8. Company search: it is possible to search for the name of a company (but also a research institute or organisation) and the system retrieves the number of patents that are assigned to this organisation. It is not evident from which source these patents are counted.

- The tool claims to be open source and APIs are possible.

- Roughly the same functionalities for completing the profiles of individual researchers are provided as in GC. Individuals can add to their own publication profiles.

Limitations

- Coverage:
  1. It is not clear how publications and citations are collected (Is it connected to Google Scholar?).
  2. The coverage is much smaller than in other sources (GS, MAS, etc.). Fewer papers per author are detected and also less citations.

- Downloading of the data:
  1. Data download is not possible with the available tools, although it seems possible to do it through their APIs (after contacting the developers).

- Searching & filtering functionalities:
  1. Searches are limited to authors, journals, conferences, courses.
  2. The search by affiliation seems to be linked to the idea of 'companies' and only provides information on patents.
  3. Searching by one journal does not give the total number of individual publications covered by the journal (only the 'top' publications).
  4. No document type classification and possibility of filtering are presented.
  5. Sorting options are not offered in the system.

- There is also an important limitation in the standardisation of author names as they appear with split and/or incomplete information. The quality of the data is quite low in general.

- There is no normalisation of indicators although several rankings are produced and presented in the website. The h-index is presented without clarification of its limitations and disadvantages.

Possibilities for research assessment

This is a very limited tool for research assessment purposes, even at the level of individual scientist. Some of the reasons why this tool is not very adequate for research evaluation are the following:

- Limited coverage. This is probably the most important limitation as the number of publications and citations is lower than in other tools.
- Lack of transparency on how the data is collected and incorporated into the data system.
- Low data quality with clearly deficient data cleaning (author names, journal titles, addresses, etc.).
- Lack of current tools for data downloading, management and standardisation.

This tool has only very limited possibilities for research assessment in its current form. Even for the individual level analysis it does not seem to be very useful as it is not clear how the publications are collected and the profiles of the researchers are mostly incomplete. Fundamental improvements are necessary for the tool to be a potentially reliable tool in evaluation processes.

3.8 JISC Open Citations

http://opencitations.net/

JISC Open Citations is a database of biomedical literature citations, harvested from the reference lists of all open access articles in PubMed Central that reference approximately 20% of all PubMed Central papers (approx. 3.4 million papers), including all the highly cited papers in every biomedical field. All the data are freely available for download and reuse.

The web site allows one to browse these bibliographic records and citations, to select an individual article, and to visualise its citation network in a variety of displays. Details of each selected reference, and the data and diagrams for its citation network, may be downloaded in a variety of formats, while the entire Open Citations Corpus can be downloaded from the source data page in several formats including RDF and BibJSON.

The project is completed and only the papers already incorporated into the system are accessible. However, it was not the aim of the project to become a service in itself. If the tool would be developed on an ongoing basis, it might allow the potential use of this tool for research assessment purposes.

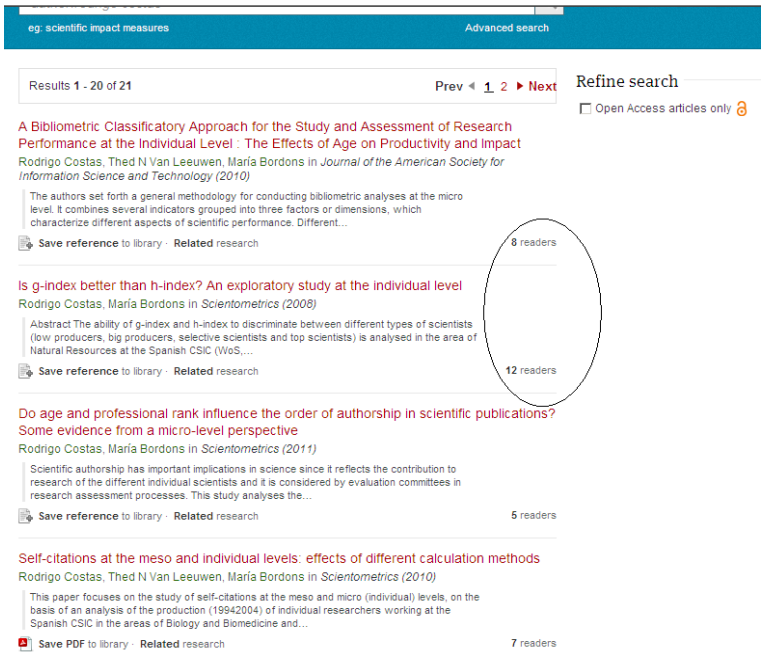
3.9 Mendeley

http://www.mendeley.com

Mendeley is a desktop and web programme for managing and sharing research papers, discovering research data and collaborating online. It combines Mendeley Desktop, a PDF and reference management application (available for Windows, Mac and Linux) with Mendeley Web, an online social network for researchers.

Characteristics

- The main characteristic of this source (from a research assessment point of view) is that after registration a user can search for publications where the number of 'readers' (i.e. other users of the system that have included the same articles in their profiles) is shown. See figure:



- Search options include title, authors, abstract, Medical Subject Headings (MeSH) terms, document type, publication year and up to 25 disciplinary fields.APIs seem to be possible, accessible and of good quality, especially considering that several other applications include data from Mendeley. The Mendeley team is also quite supportive of developers who want to use the Mendeley API.
- There are also individual profiles of the users of the system, which allows the analysis of some individual statistics for the users registered.

Limitations

- The number of 'readers' and the 'tags' received by a publication could be an interesting indicator of impact or interest of a publication. However, the information cannot be easily downloaded from the website (although it should be possible).

- The coverage of the database relies almost only on the new inclusions of publications by the users of the papers.

- It is also important to take into account that there are no explanations about the origin (mainly disciplinary) of the potential 'readers' of Mendeley. This could cause some bias in the population of readers of the papers. In any case non normalised data is presented and this does not seem feasible in the future either.

Possibilities for research assessment

- The idea of measuring the 'readers' and the number of tags that a publication receives is appealing, but more transparency and accessibility to the data should be necessary.
- If publications can be linked in a straightforward way with Mendeley and the number of 'readers' can be counted, then this could be an interesting add-on for research assessment purposes. However, at present this does not seem to be a possibility with the current software and facilities in the website of Mendeley.
- Mendeley can be combined with some of the other tools that have been developed based on it for assessment purposes. Some of them are discussed below (e.g. Total-Impact).

3.10 CiteUlike

http://www.citeulike.org/

CiteULike is a free service to help users to store, organise and share the scholarly papers they are reading, being a tool based on the principle of social bookmarking and aiming to promote and to develop the sharing of scientific references amongst researchers. The website is sponsored by the publisher Springer Science+Business Media.

The main interest of this tool is similar to Mendeley, it allows tracking of other uses of bibliographic literature by 'tagging' or 'readers'.

Characteristics

- Some download options are available (to some reference manager software packages), but the extra data (posts, notes by the readers, etc.) are lost when the data is downloaded.
- The main use of the tool is to create libraries, based on the selection of papers by authors, journals, titles, etc. But data cannot be downloaded extensively.
- It has a good exportability with the Zotero software.

Limitations

- The coverage is dependent on the users who are the ones who include new references of publications, normally extracted from other sites and bibliographies.
- Searches on the data show that it is difficult to search and download the data in a systematic way.
- In general it has similar limitations as Mendeley.

Possibilities for research assessment
The possibilities for assessment and publication impact measurement are comparable to those of Mendeley. The limitations are also similar. The interoperability with the Zotero software makes it interesting for data collection and exploration. However, the tool still does not have strong possibilities for research assessment because it is not clear what the meaning is of the posts given by the users. There are no other influence metrics that can be recorded from this source.

3.11 Zotero

http://www.zotero.org/

Zotero is free, open source reference management software to manage bibliographic data and other research materials. Zotero is a tool that helps users gather, organise, and analyse sources (citations, full texts, web pages, images, and other objects), and lets users share the results of their research in a variety of ways. Zotero is produced by the Center for History and New Media of George Mason University (GMU). Features include web browser integration, online syncing, generation of in-text citations, footnotes and bibliographies, as well as integration with the word processors Microsoft Word, LibreOffice, OpenOffice.org Writer and NeoOffice. An extension to the popular open-source web browser Firefox, Zotero includes the best parts of older reference manager software, like EndNote. It has the ability to store author, title, and publication fields and to export that information as formatted references. However it also has the best parts of modern software and web applications, such as the ability to interact, tag, and search in advanced ways. Zotero integrates closely with online resources; it can sense when users are viewing a book, article, or other object on the web, and—on many major research and library sites—find and automatically save the full reference information for the item in the correct fields. Since it lives in the web browser, it can effortlessly transmit information to, and receive information from, other web services and applications; since it runs on ones personal computer, it can also communicate with software running there (such as Microsoft Word). It can be used offline as well.

Characteristics
- It allows for collecting bibliographic data directly from the web and from some of the previous tools (e.g. F1000). The problem is that when the data are collected the metrics are lost (they are not collected in the record in Zotero).
- It allows collection of data from Google Scholar, page by page, being possible to select all the papers from a page with one click, making the collection of data from GS a little bit faster (although not systematic).
- It works also with Google Chrome. It is not always compatible with other browsers (e.g. MS Internet Explorer).

Limitations
- When the data is collected the metrics are lost, thus you are not able to obtain indicators later.
- This makes it impossible to obtain any metrics of any type.

Possibilities for research assessment
This tool does not have true assessment capabilities as it does not provide any metrics and it does not record the metrics provided by other tools such as GS, CiteUlike, etc. Its main interest relies on its capability to collect and store data from some of the previous tools and sources (as well as other sources not discussed in this report) in a quite systematic, quick and relatively efficient way. An improvement in this respect would allow this tool to gain strong importance for research assessment purposes.

3.12 Readermeter

http://readermeter.org/

This is a website that, using readership-based metrics (based on Mendeley), can estimate publication impact on the basis of the consumption of scientific content by a population of readers. ReaderMeter adapts two popular impact metrics for authors (the h-Index and the g-Index) and redefines them using bookmarks instead of citations. It calls these the HR-Index and GR-Index respectively. Analysing readership data can help discover areas of real-time impact that may not be visible to traditional citation-based measurements.

Characteristics

- This is a tool based on Mendeley and allows searching for an author and obtaining the metrics mentioned before (HR-index, GR-index) based on the number of readers.

Limitations

- There is no way to limit the search by any other element than the name of an author. If the name is very common the program does not work properly.
- Other searches (journal, year, document type, etc.) are not possible.
- The downloading of data is not possible.
- The standardisation of the data, as well as the merging, addition, correction, etc. of data (even at the individual level) is not possible, making it impossible to correct some of the mistakes that are found (authors with different variants, errors in the names, etc.).

Possibilities for research assessment

- Only for limited benchmark and assessment purposes, particularly at the individual level and when the collection of data is not very problematic.
- No potential use as a standard tool given the limitations in the filtering and downloading capacities of the data.
- The fact that mostly only size-dependent indicators are presented (HR-index, GR-Index, etc.) also gives limited evaluative possibilities to this tool even at the individual level.

3.13 Total-Impact

http://total-impact.org/

Total-Impact is a website that allows for quick and easy viewing of the impact of a wide range of research outputs. It goes beyond traditional measurements of research output – citations to papers – to embrace a much broader evidence of use across a wide range of scholarly output types. The system aggregates impact data from many sources and displays it in a single report, which is given a perma-URL for dissemination and can be updated any time.

Characteristics

- The search of information is done by DOI, PubMED ID, URL, and other document identifiers. Search by author, title, journal title, year, etc. is not possible.
- Several groups of documents can be analysed and although there is no explicit mention, 250 is the maximum of documents that can be analysed at once.
- It is based on the aggregation of several tools (Mendeley readers and groups, CiteUlike, PlosAltmetrics, etc.)
- It is possible to export the results (the metrics) to a text file which offers a lot of possibilities for later assessment. As the DOI is kept, it is very straightforward to link it with other systems.
- There are API possibilities.

Limitations
- Search is limited to the document identifiers previously mentioned.
- Only 250 documents can be analysed at a time.
- No other searches are allowed (author, journal, title, etc.)
- It takes a few minutes (3 or 4) to calculate the indicators for 250 items.

Possibilities for research assessment
This is a highly interesting and powerful tool. It has clear potential for research assessment purposes. If the user has selected a valid set of publications (this set cannot be very big, but probably for bigger sets it could be automated) it is possible to attach all the altmetrics described above.

Besides, given the fact that they work with several other tools (Mendeley and CiteUlike for example), these other sources must be accessible through APIs.

At the moment the main problem is that the tool is still in a development phase and extensive use of the tool is currently not advised.

3.14 ScienceCard

http://sciencecard.org/

ScienceCard is a web service that collects all scientific articles published by an author and displays their aggregate article-level metrics. ScienceCard allows a researcher to create and maintain a researcher profile with minimal effort, and to export and reuse this information elsewhere. To make this as effortless as possible, ScienceCard relies on unique identifiers for authors (currently identifiers from Microsoft Academic Search and AuthorClaim) and articles (currently digital object identifiers or DOIs). Future versions will add more author identifiers services, including ORCID when the service launches in 2012, as well as other identifiers for scholarly content (e.g. from arXiv).

ScienceCard uses freely available sources for the article-level metrics. Some sources are only available from some publishers (e.g. HTML page views or PDF downloads), or when one is a publisher oneself (e.g. CrossRef).

Characteristics
- The tool is similar to Total-impact but focuses only on the author level.
- It gathers information from all the previous altmetric sources: Mendeley, CiteULike, etc.

Limitations
- It is dependent on the information provided by the authors themselves (Only ScienceCards of authors that have registered with ScienceCard are displayed). Similar limitation as with GC.
- Normalisation of indicators is not offered.

Possibilities for research assessment
- Although the information provided is complete, as it is based on the decision of the authors to create the ScienceCard, it can only be applied for individual assessment purposes and with many restrictions as not many authors have included their profiles.

3.15 PLoS ONE Almetrics (Article-Level Metrics Information)

http://www.plosone.org/static/almInfo.action

The PLoS ONE journals started a program to provide ´article-level metrics´ on every article in all the journals of the PLoS family in March 2009. The data can be downloaded in Excel files and used for different analytical purposes.

Characteristics

- PLoS ONE offers different types of metrics for all the papers published in any of their journals, combining citations, social bookmarking, downloads, HTML page views, ratings from the users, comments and blogging.
- These data can be downloaded in Excel format which enables one to sort, select, etc.
- The data are free.
- Different dimensions of indicators are observed (see factor analysis in the table below). This includes indicators of citations, social bookmarking, number of PDF downloads, HTML views, ratings, comments and threads by the readers, etc.

Rotated Component Matrix[a]						
	Component					
	1	**2**	**3**	**4**	**5**	**6**
Citations - CrossRef	.942					
Citations - Scopus	.927					
Citations - PubMed Central	.918					
Social Bookmarking - CiteULike		.746				
Social Bookmarking - Connotea		.684				
Total PDF Downloads	.511	.656				
Combined Usage (HTML + PDF + XML)		.633		.553		
Total HTML Page Views		.612		.562		
Number of Ratings			.897			
N. of 'Star Ratings' that also include a text comment			.843			
Average Rating			.801			
Number of Comment threads				.767		
Number of replies to Comments				.670		
Total XML Downloads						
Number of Note threads					.796	
Number of replies to Notes					.794	
Number of Trackbacks						.698
Blog Coverage - ResearchBlogging.org						.633
Blog Coverage - Nature Blogs						.507
Blog Coverage - Bloglines						
• Extraction Method: Principal Component Analysis. • Rotation Method: Varimax with Kaiser Normalisation. • 33,128 documents downloaded from PLoS. • 64% total variance explained. • Loadings higher than .500 are shown.						
a. Rotation converged in 7 iterations.						

Table 1: Factor Analysis covering various metrics

- The table shows the factor analysis of all the indicators for a group of more than 30,000 publications from PlosONE. As can be seen, the indicators relate to different dimensions, this means that they are actually measuring different aspects and different types of influence and impact.
- An interesting observation is also that the best ´predictor´ of the total number of citations of publications is the number of PDF downloads, while the number of HTML page views of publications is more related with comment threads and replies to comments. This seems to suggest that users download to cite later, and use the paper in the HTML version to comment on it.

Limitations

- The main limitation of this source is that all the information provided is limited to PLoS journals.
- There are still some major methodological problems (e.g. lack of normalisation, lack of indicators for authors, organisations, etc.) as well as some conceptual problems (e.g. what is the meaning and value of the different dimensions of indicators?) that need to be studied further.

Possibilities for research assessment

The information provided by PLoS ONE has strong research possibilities but is currently not suited for general assessment purposes (unless the analysis is limited to this journal). Right now the main value of these tools is for research purposes in information and library science and altmetrics development. If these types of metrics and its further development by PLoS ONE would be adopted by other journals and publishers, the value and capacity of all these altmetrics would become even more relevant.

3.16 Statics on the Usage of REpositories (SURE) 2

http://repositorymetrics.narcis.nl/

The Insight Repositories Dashboard shows the usage of repositories in a selection of Dutch research organisations. It shows the abstract views and full text downloads. The dashboard has been created in cooperation with all the participating institutions in the project called SURE2. This project was funded by SURFfoundation from the Netherlands.

SURE2 is a project that delivers one API and three 'views' for two user groups: repository managers and researchers. The data are collected from the log files at the different repositories from the participant universities. The usage data from all the repositories are treated in exactly the same way, so the data is comparable across repositories. The data is made available via an API, which is the core component of the tool. Widgets can be created to visualise the data at the single publication level and the aggregated level involving multiple publications.

Here we will review the Repository managers view. The widgets on the repository metrics dashboard visualise the usage data at the aggregation level of an institution and country, and changes when filters are manipulated. Also the widgets can be embedded in other websites of the users.

The institutions that are currently covered by the repository metrics dashboard are: Vrije Universiteit, Universiteit van Amsterdam, Universiteit van Twente, Wageningen Universiteit en Research, Centrum voor Wiskunde en Informatica, NCB Naturalis, TU Eindhoven and Erasmus Universiteit.

This project provides an API allowing querying for usage statistics at an article level granularity. Future developments of this project will provide for a metrics tab in NARCIS (http://www.narcis.nl/) for single publications and at other aggregation levels like the author level. For the latter they make use of the Digital Author Identifier – DAI.

Characteristics

- Information is displayed on the number of downloads and views of abstracts for the publications included in the institutional repositories of the eight research institutions previously mentioned.
- Figures and widgets for the visualisation of the data are included.
- Aggregated data can be downloaded in different formats and for different uses.
- API possibilities are available to access the data at the article level. This allows for the creation of services beyond the current options already available in the current view. APIs allow developers to enable aggregations of publications for journals and other aggregations like faculties and departments. This creates further possibilities for development of the tool.

Limitations

- No data is provided for other countries or organisations apart from the ones already included in the system. The information is at present limited to the Netherlands and is not available for all the universities.
- Initially, there are no search possibilities by author, journal, document type, etc. Thus, it is not possible to limit the query by any other element (e.g. DOIs, authors, journals, etc.) with the current options available on the website. However, most of these possibilities are available through the use of the API, although the user still needs to create a collection of persistent document identifiers to find the documents.
- It is not possible to download data at the document or author level with the current standard options presented in the tool, but again the API allows more extensive access to the data, thus enriching the analytical possibilities of the tool.

Possibilities for research assessment

SURE2 presents very rich information about the number of downloads and views of publications deposited in the institutional repositories of eight Dutch research institutions, thus giving valid and reliable information on the usage of the scientific production covered by the Dutch repositories.

With the current tools of accessing the data in the repository metrics view it is not possible to obtain data at the document level (extensively) or for other units of analysis (authors, departments, etc.). However, these options are possible through the API, which makes it possible to download the statistics for a single persistent identifier. This allows the analysis of different aggregations, for example of authors, research groups, etc. The future development of the project may include statistics of individual publications in a graphical user interface (GUI).

On the other hand, there are also some conceptual limitations, mainly related to the fact that not all the publications contained in the repositories are necessarily all the publications done by the researchers. Publications can be missing in the system (e.g. other publications in other institutions; publications that are published in journals that do not allow self-archiving, publications simply not included in the repository, etc.) thus partially limiting the evaluative power of the tool. However, it may be possible to minimise this problem because SURE2 is based on the COUNTER[7] standard for article level online usage statistics. This allows the incorporation of statistics from the publishers, in order to get a more complete picture on the usage of publications beyond the institutional repositories.

The most important limitations of this tool for research assessment purposes are:

- Coverage is limited to only eight Dutch institutions. This is probably the most important limitation. Not all research organisations in the Netherlands are included in the system. However, the increase of the institutions involved would significantly expand the analytical and evaluative possibilities of the tool.
- Lack of possibilities to select more detailed data and filtering options (year or date of publication/archive, document type, etc.). The current standard options of the tool do not allow identifying/downloading of the individual publications (although it is possible through the API). However, the availability of the API and the continuity of the development of the project will allow more analytical and evaluative possibilities, together with the linkage to other systems.

[7] COUNTER (Counting Online Usage of Networked Electronic Resources) is an international initiative serving librarians, publishers and intermediaries by setting standards that facilitate the recording and reporting of online usage statistics in a consistent, credible and compatible way (http://www.projectcounter.org/).

3.17 Summary of the empirical assessment

In general all the novel tools analysed present interesting and promising aspects from a research assessment point of view. However, in their current state, due to their limitations and restrictions in their use (mostly providing metrics on the paper and individual levels), it can be concluded that they seem to be more useful for self-assessment than for systematic impact measurements at several levels of aggregation.

A general summary of the main characteristics and functionalities of the main tools surveyed in this report is presented below (Table 2).

Characteristics	Tools/Sources														
	F1000	Peer Evaluation	Paper Critic	GC	GS	MAS	Arnet-miner	Mendeley	CiteULike	Zotero	Reader-meter	Total-impact	Science-Card	Plos-ONE	SURE2
Metrics for papers	Yes (d)	Yes (a)(d)	Yes (a)(d)	Yes	Yes	Yes	Yes	Yes (a)	Yes (a)(e)	No (c)	Yes	Yes (a)	Yes (a)	Yes	Yes (f)
Metrics for individuals	No	Yes (a)	No	Yes (a)	No	Yes (b)	Yes (b)	Yes (a)	No	No	Yes (a)	No	Yes (a)	No	Yes (f)
Metrics for institutions	No	No	No	No	No	Yes (b)	No	No	No	No	No	No	No	No	Yes
Metrics for countries	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes (a)
Metrics for journals	Yes	No	No	No (f)	No (f)	Yes	Yes (b)	No	No	No	No	No	No	No	No
Data download & management	Yes (b)	No	No	Yes (b)	Yes (b)	Yes (b)	No	Yes (b)	Yes (b)	Yes	No	Yes	No	Yes	Yes (b)
API possibilities	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Citations	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	No
Altmetrics - Readers	No	Yes	No	No	No	No	No	Yes	Yes (b)	No	Yes	Yes	Yes	Yes	No
Altmetrics - Bookmarks/Tags	No	No	No	No	No	No	No	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Altmetrics - Comments	Yes	Yes	Yes	No	No	No	No	No	Yes	No	No	Yes	Yes	Yes	No
Altmetrics - Downloads, views, etc.	No	Yes	No	No	No	No	No	Yes	No	No	No	Yes	Yes	Yes	Yes
Altmetrics - Others	No	Yes	Yes	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No
Peer review/Discussion by others	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No	No	Yes	No
Coverage - Transparency	Yes	No	No	No	No	No	No	No	No	No	No	No	No	Yes	Yes
Coverage - Multidisciplinary	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes (b)	Yes
Free access	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Registration necessary	Yes	Yes	Yes	Yes	No	No (b)	No (b)	No (b)	Yes	Yes	No	No	No	No	No
Searching/Filtering options	Yes (b)	No (b)	No (b)	Yes (b)	Yes (b)	Yes (b)	Yes (b)	Yes (b)	Yes	Yes	Yes (b)	No	No	No	Yes
Normalisation options	Yes	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Data standardisation/ cleansing	No	No	No	No	No	Yes (b)	Yes (b)	No	No	No	No	No	Yes	No	Yes
Easy/friendly/intuitive interface	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes (b)	No	Yes	Yes

Table 2. Main characteristics of the different tools surveyed

Notes:

a. Only items/persons/users included in the system. These are the cases when the tool does not create a universe of publications (e.g. GS, MAS) or personals profiles (e.g. MAS), but depends on the data collected/uploaded by the users.

b. But with restrictions/limitations. This means that the feature does exist in the functionality/option but can present limitations or difficulties for the given purposes (e.g. downloading big groups of data; calculation of indicators in a systematic way; indicators for several levels of aggregation; mistakes in the data; limited functionalities without registration of the users, etc.).

c. This is a tool more for data collection & management than for actual metrics.

d. Some publications don't have any rating/metric. This note refers to the situation when there are publications that although they are in the system they don't have any metric.

e. No real metrics but tagging, bookmarking, etc. information. These are cases when the data for the publications is not summarised in indicators.

f. Considering current options in the tool. This means that the functionality is not directly available in the tool, although it is possible with other means (e.g. APIs, ad hoc software, manual collection of data, etc.)

4 The interplay of technologies of narcissism, information filtering, and technologies of control – conclusions & recommendations

4.1 Summary of the empirical study

In the previous section, we have surveyed and reviewed the most important novel web based tools that carry some promise with respect to improved measurements of the influence, impact, or usage of scientific and scholarly publications. The main characteristics of these tools have been summarised in Table 2. A few patterns are striking. First of all, most of these tools have developed relatively friendly interfaces (with only two exceptions). Most of them (with only one exception) are also freely available, although most of them require a registration of the user for full access and use of the data. All of them provide metrics at the level of the article, manuscript or book. Taken together, these three characteristics make these tools attractive to individual researchers and scholars. It enables them to quickly see statistical evidence regarding impact, usage, or influence without too much effort. Without being pejorative, we can conclude that web based academic publishing is producing a variety of information filters and 'technologies of narcissism'[8] based on these information filters. With this we mean technologies that allow the researcher to make some sort of limited self-assessment with respect to the response to his/her work. However, this does not mean that these technologies and databases can also legitimately be used as 'technologies of control'. These are technologies that are being used by evaluating and managing actors to subject researchers to forms of research assessment. In order for impact measurements to be valid as technologies of control, they need to adhere to a far stricter protocol of data quality and indicator reliability and validity.

The main problem is that none of the new tools we reviewed meet crucial requirements for data quality and indicator construction. This prevents them from currently being usable in the context of research evaluation and assessment. As is visible in Table 2, only four of the tools provide a data cleansing and standardisation option, and two of these only in very limited form. This means that these web based tools may create statistics and indicators on incorrect data. We explained this in more detail in the previous section for each tool. A second relevant requirement is the possibility to normalise indicators (Moed, 2005). This is not an absolute requirement but non-normalised indicators can provide very useful information. The possibility to normalise does enable a comparison of impact across different fields. Scientific fields differ considerably in their communication characteristics. For example, the numbers of citations in clinical research are very high because a very large number of researchers is active, the lists of references per article are relatively long, and there are many co-authored articles, sometimes with tens of authors per paper. As a result the average clinical researcher has a higher citation frequency than the average mathematician. The latter operates in much smaller communities with relatively short lists of references and many solitary articles. As a consequence, it would be irresponsible to compare the raw citation data as a proxy measure of scientific impact among units with production from very different fields. In many evaluation contexts, it is therefore desirable to be able to normalise impact indicators. Table 2 shows that only one tool (F1000) enables some data normalisation (at least at the journal and sub-disciplinary levels). The third requirement for serious use of a tool in impact measurements is the data coverage. In this respect, the results are mixed. A large number of tools have a multidisciplinary coverage, which may enable the consistent use of these tools across fields. However, most tools are not transparent about the data coverage. Only three tools are clear about the way they build up their database. For example, although Google Scholar is clearly the most popular of the tools discussed, it is not completely clear and documented how GS builds up its database and how the citation frequencies are actually calculated. Applying GS based data in a research assessment context, without considering these limitations and drawbacks, would therefore be unreliable. After all, the researchers evaluated have no means to check either the

[8] With thanks to Diana Hicks who suggested the opposition of technologies of narcissism and technologies of control.

underlying data or the way the indicators are computed. The tools have some more limitations. There are almost no tools that provide metrics at other levels of aggregation such as research institutes, journals, etc. Most tools also do not provide easy ways for data downloads and data management. Although less severe than the crucial requirements, these limitations also diminish the usability of many of these tools.

This does not mean that the tools reviewed cannot play a useful role for scientists and scholars. The reason for this is that the context of use of these measurements is different. A researcher who checks, for example, her citation counts in Google Scholar knows everything about the context in which she works. She knows her field, her colleagues and competitors, and the journals or publishers through which she communicates her results and insights. This means that the citation numbers will be interpreted in this context and will not be taken at face value. If the citation data surprise her, she will either learn something about her work that she did not know yet, or attribute it to idiosyncrasies in the way the Google database has been set up or the citation counts are computed. The same holds for all the technologies presented in this report. As technologies of narcissism they may prove to be very useful because they may raise interesting questions for the individual researcher or lay to rest anxieties about one's work. It is quite a different thing to apply the same data and indicators as a technology of control (Beniger, 1986).

4.2 Four arguments for alternative metrics

The existing literature about emerging and alternative metrics tends to gloss over this important distinction between technologies of control, information filtering and technologies of narcissism. Enthusiasm about the new possibilities, which are indeed more impressive than the casual web user may realise, tends to be dominant. The main topic in the literature on alternative forms of impact measurement is not the quality of the data or the reliability of the indicators that can now be calculated with these tools. As far as we know, this report is the first to do this in such a rigorous way. Rather, the *leitmotiv* of the body of work on altmetrics is the promise of the availability of better tools to monitor, track, and measure other aspects of scientific and scholarly literature than is possible by citation analysis or peer review (the currently dominant paradigms of impact assessment). A good example of this is the Altmetrics Manifesto[9] . The need for information filters given the explosive growth of the literature is the context of this manifesto. This calls for tools beyond traditional peer review, citation analysis, and ISI's Journal Impact Factor. 'In growing numbers, scholars are moving their everyday work to the web. Online reference managers Zotero and Mendeley each claim to store over 40 million articles (making them substantially larger than PubMed); as many as a third of scholars are on Twitter, and a growing number tend scholarly blogs.' (Priem, Taraborelli, Groth, & Neylon, 2010). The manifesto calls for research and technology development for tomorrow's information filters. The manifesto notes that traditional forms of publication in the current system of journals and books are increasingly supplemented by other forms of science communication. These include: the sharing of 'raw science' like datasets, code, and experimental designs; new publication formats such as the 'nanopublication', basically a format for the publication of data elements (Groth, Gibson, & Velterop, 2010); and widespread self-publishing via blogging, microblogging, and comments or annotations on existing work (Priem et al., 2010).

The literature has developed four key arguments in favour of alternative forms of impact measurement by web based tools such as the ones discussed in this report.

The first argument is diversity and filtering. Because web based publishing and communication has become so diverse, we need an equally diverse set of tools to act upon these traces of communication. The altmetrics tools build on their use as information filters to also start measuring some forms of impact (often defined differently from citation impact).

The second argument is speed. It takes time for traditional publications to pick up citations and citation analysis is only reliable after some initial period (which varies by field). The promise of

[9] http://altmetrics.org/manifesto/

altmetrics is an almost instant measurement window. 'The speed of altmetrics presents the opportunity to create real-time recommendation and collaborative filtering systems: instead of subscribing to dozens of tables-of-contents, a researcher could get a feed of this week's most significant work in her field. This becomes especially powerful when combined with quick "alt-publications" like blogs or preprint servers, shrinking the communication cycle from years to weeks or days. Faster, broader impact metrics could also play a role in funding and promotion decisions.' (Priem et al., 2010).

The third argument is openness. Because the data can be collected through APIs (for limitation on these per tool see Table 2), the data coverage is completely transparent to the user. This also holds for the algorithms and code used to calculate the indicators. An important advantage discussed in the literature is also the possibility to end the dependency on commercial databases such as Thomson Reuters' Web of Science or Elsevier's Scopus. The difficulties that are entailed in the bottom-up creation of a completely new usage, impact, or citation index is however usually not mentioned. Still, this promise of a non-commercial index that can be used to measure impact or other dimensions of scientific performance should not be disregarded. In the long term, this may be the direction in which the publication system is moving.

The fourth argument is that many web based traces of scientific communication activity can be used to measure aspects of scientific performance that are not captured by citation analysis or peer review. For example, download data could be used to measure actual use of one's work. The number of hyperlinks to one's website might also be an indication of some form of impact. Indeed, since 2000 the fields of internet research, webometrics and scientometrics have developed a body of work comparing the roles of citations and hyperlinks and the possibility of building impact measurements on these analogies (Bar-Ilan & Peritz, 2002; Björneborn & Ingwersen, 2001; Hewson, 2003; C. Hine, 2005; Rousseau, 1998; Mike Thelwall, 2005). It is clear that the literature on the implications of web based publishing and communication is still in an early stage. This is not surprising given the recent birth of the web in its current form (1995) and the strong hold that traditional paradigms have on publishers and researchers alike. The literature provides glimpses of possible future tools that will need much more work before they can be used in routine ways. As the Altmetrics Manifesto states: altmetrics are in an early stage and much research is still needed. In the framework of the *Conference Series of the Association of Internet Research* since 2000, a serious body of work has already been presented on the characteristics of internet communication (Howard & Jones, 2004). More important in the framework of this report is the work done on the use of the web by scholars and scientists (Borgman, 2007; Hine, 2006, 2008; The Virtual Knowledge Studio: et al., 2008; Willinsky, 2006; Wouters et al., 2012). Assessing the extent to which the new tools can be used to measure publication impact does also require a thorough knowledge of the actual uptake of these new tools by the scientific communities. Recent surveys have shown that no universal trend exists, that actual information and communication practices differ strongly by field, and that assessment and career decisions are still mainly based on the traditional forms of academic output (Bulger et al., 2011; Gruzd et al., 2011; Harley, Acord, Earl-Novell, Lawrence, & King, 2010; Williams et al., 2009).

Researchers and scholars are developing new activities such as blogging and twittering, but it needs to be better understood how these relate to their scientific and scholarly practice as a whole. For example, what does it mean that about one third of the researchers are reported to have an account on Twitter? Does this mean the same in different fields? And what do Twitter statistics mean: what kind of dimension of scientific activity or performance might they represent? And also, should all researchers 'tweet' as a part of their scientific activities? It is easy to be seduced by the large-scale availability of numbers and be misled by them. For example, in a recent study an editor of a journal claimed that the number of tweets could predict the number of citations (Eysenbach, 2011). As a result, this article itself was tweeted 527 times in a short time. However, a re-analysis of the paper showed that the results were far less groundbreaking, to say the least (Davis, 2012). Before jumping too quickly to the conclusion that some web based freely available measures are statistically related to existing measures of impact, it is necessary to study the specific characteristics of different forms of web based communication and publication. For example, a recent study compared research blogging in the area of chemistry with journal publications. It was found that blogging is more oriented towards the social implications of research, tends to focus on

high-impact journals, is more immediate than scientific publishing, and provides more context of the research (Groth & Gurney, 2010). We need much more of these studies while we are developing impact measurements that are relevant to the specific characteristics of the different formats of web based communication and publication. Indeed, this comprises an important element of the research agendas in the fields of internet research, science & technology studies, information science, communication science, social network analysis, webometrics and scientometrics.

Discussing these research agendas in their entirety would be beyond this report. Instead we focus on the four arguments in favour of new alternative metrics. To start with the fourth argument: can we expect to be able to measure more diverse aspects of scientific performance and communication by using novel web based metrics? Our answer is "yes, but not yet". We expect that the different forms of web activities will in the future be interpreted in the context of research evaluation. It is far less clear what the numbers might mean. To which dimension of science does the number of tweets relate? And what is the role of blogging in areas such as mathematics or clinical research? Moreover, a statistical framework is lacking to interpret the numbers quantitatively. For example, at what number of tweets should one be impressed? Does this vary by discipline or topic? Can they be cheated or manipulated? The same holds for downloads, usage data, number of incoming hyperlinks, trackbacks in blogs, etc. The advantage of these measures, their diversity, is also their disadvantage: it makes it harder to normalise and contextualise indicators based on these measures. To be sure, we would not want to claim that this is impossible. But it would require a serious ongoing effort to codify the collection of relevant data, their definition, and underlying fundamental research in the characteristics of novel forms of web communications.

This is also the state of affairs in measuring the societal impact of research. The current frameworks to analyse this dimension of academic research is either based on link relations between scientific and patent databases (Moed & Glänzel, 2004; Oppenheim, Cronin, & Atkins, 2000) or on the conceptualisation and measurement of 'productive interactions' between researchers and non-academic stakeholders (Spaapen & van Drooge, 2011). However, the state of the art is not yet at the stage where it is already clear how these interactions can be measured in the relevant research areas. We need a sustained effort for a number of years in the form of case studies as well as comparative analyses before we know how to measure and evaluate societal interactions of scientific and scholarly communities. Also, much more database work is needed before we can measure the evidence for societal outcomes of research in a reliable and valid way[10]. For example, in clinical research the number of citations in professional journals might be a valid indicator of societal impact, but this depends on the role of these journals. Perhaps the inclusion in clinical protocols would be a more valid indicator. In other fields, different measures and databases will be relevant. We simply do not yet know enough about these interactions to already use them in assessment exercises that may have a deep impact on research and research careers.

The argument of openness has a more political character and is less dependent on technical considerations. It is clear that the system of publications is generally moving towards a higher level of openness, albeit with quite different rates of change in different fields. But it is important to realise that openness has different dimensions (Tatum & Jankowski, 2012). The comparison of the open Google Scholar with the commercial Web of Science is informative. Although GS is free while the WoS requires a license, the construction of the dataset and the calculation of indicators are actually more transparent in the WoS than in GS. This has important implications because many web based tools are implicitly based on search engine results. This also holds for many webometric tools (Aguillo, Ortega, & Fernández, 2008; M. Thelwall, Wouters, & Fry, 2008; Thelwall, 2005). So although openness is an important asset, it is not always clear that web based tools are a better implementation of this ideal than for-profit information services. In the framework of research

[10] This is the reason that the measurement of societal impact has been included as a priority in the new research programme of CWTS.

evaluation, transparency and consistency of data and indicators may be more important than free availability[11].

The argument of speed also needs some qualification. Although it may be quite annoying that it takes time for publications to gather citations, this does not mean that faster indicators are always better. As we discussed above, the very fast twitter activity tends to be also very superficial. This is exacerbated by the fact that many tweets are generated by computers rather than by humans. It seems plausible to expect that the first waves of tweets about a scientific article may represent an entirely different dimension of science communication than the later wave of perhaps more carefully crafted citations. The evaluation and use of new scientific or scholarly insights simply takes time. Faster may not be better. We think that it still is not clear how fast research evaluation and assessment indicators should be. We cannot discard the possibility that 'Sleeping beauties' (van Raan, 2004) which pick up impact much later in their existence could also play a role in web based communication. Of course, this may be different when the tool is used as an information filter or as a self-appraisal tool. In those cases, speed may count more.

This brings us to the first argument in the literature: diversity. We agree with the proponents of altmetrics that this is a strong argument in favour of the development of new indicators on the basis of new media and activities in science communication. As we indicated, this means that more research work in this area should be encouraged in order to build up a critical mass of validated knowledge which can serve as the basis for new evaluation tools and information filters.

4.3 Recommendations: steps towards novel impact measurements

Among the tools reviewed in this report we can distinguish between tools for improved peer review, improved and expanded citation analysis, and altmetrics tools capturing different dimensions. In the following paragraphs the main limitations detected for these tools are summarised. With this we suggest that the complete or partial solution of these problems would make the tools much more relevant for the measurement of publication impact.

4.3.1 New tools for peer review analysis

The most important tools that focus on the review and discussion of publications by other reviewers and peers are Faculty 1000, Peer Evaluation and Paper Critics. Clearly, these new tools introduce a new dimension of assessing publications from the perspective of qualitative assessments beyond the more traditional bibliometric and citation analysis.

However, in their current state of development they still present some important limitations that need to be considered before the full introduction of these tools in research assessment activities. These limitations are:

- Limited in scope, normally focused on some fields or journals (e.g. F1000).
- The more interdisciplinary ones (Peer Evaluation and Paper Critics) have less transparency in the origin and typology of the reviewers and publications included. In addition, it is remarkable that most of the publications appear actually without any review.
- Conceptual limitations need to be tackled, including a better understanding of the meaning of the reviews and ratings, their value in research assessment, how they relate to other metrics, etc.

[11] A good example of this is F1000, which although it is a for-profit tool outperforms the free competitors (Peer Evaluation and PeerCritics) as the former presents a much more robust and transparent delineation of the dataset and indicators as compared to the latter.

4.3.2 New tools for citation analysis

Google Scholar and Google Citations are very promising tools, but for now they cannot be applied systematically and comprehensively in broad research assessment exercises (Aguillo, 2012)

The main limitations of these tools are:

- Lack of transparency of coverage. How are the publications collected? What sources and publications could be missing?
- Lack of systematic and comprehensive access to the data. As the data cannot be downloaded in order to be pre-processed by the users, it is virtually impossible to create standard frameworks for research assessment purposes.
- Restricted access to the individual profiles (in GC). This means that only individuals who made public their profiles can be compared and analysed.

On the other hand, Microsoft Academic Search seems to be the most accessible tool. Its promise of APIs suggests more openness and possibilities for a systematic use of the data provided. Moreover, MAS has a more advanced data infrastructure (the parsing of the data is more complete, document types are detected, etc.). The main limitations are the coverage of the database, which is lower than the coverage by GS, and the deficiencies in the standardisation of the data (e.g. author identification).

4.3.3 Altmetrics

As indicated above, altmetrics are booming and they are starting to be seen as alternatives to more conventional citation measures. Metrics on the number of readers, tags used, bookmarks, comments and threads, blogging, tweets, etc. are starting to be suggested as new tools to assess the impact and influence that researchers have over their colleagues and society-at-large.

However, for the time being, these tools present serious limitations that do not yet allow them to be considered systematically in research evaluation. These are some of the main limitations:

- The most important developments in altmetrics are done in one single family of journals (i.e. PLoS journals), thus reducing their scope and potential use. This may change in the future if other publishers and journals would join the initiative thus allowing the more extensive use of these metrics.
- More transparency and clarity in the data covered would be necessary. The detection of potential biases is very important in order to be able to frame them in a standardised context (e.g. from where are the readers/bloggers/taggers? How many are they? Are there potential country, field, age, or language biases?). In addition, normalisation of the metrics would be important in order to have benchmark references for the comparability of the data.
- Improving access to the data (downloading, filtering, searching, etc. options) and increasing the quality of the data (more parsing of data in fields, more depuration of information, improving exportability, matching data, etc.). Access to the data is still limited to authors and small sets of documents. Moreover, it is not possible to systematically download the data in ways that more expert users (e.g. bibliometricians) can treat the data themselves, thus limiting their potential use.

4.3.4 General recommendations

Keeping all the previous limitations in mind, we should not forget that the new web tools for peer review assessments, citations analysis and altmetrics are being developed rapidly (a good example is the profusion of different tools considered in this report) and they are continuously improving over time (e.g. GS). Their continuous development will present ongoing challenging and valid elements for research assessment in the future (e.g. broader coverage, more citations, other dimensions of 'impact', and 'influence', other fields such as Humanities that can benefit from them, etc.).

Although their current state of development means they are inadequate to use in a systematic and extensive way (i.e. with the same research assessment purposes as Web of Science or Scopus can be used nowadays), this situation could change in the future and they could become clear competitors and alternatives of the more traditional sources. Solving some of the problems and limitations described in this report would clearly reinforce their role as "technologies of control" with possibilities of use for research assessment in more standardised and broader contexts.

We therefore recommend to keep surveying and monitoring all the reviewed tools. Given their current state of development, some of the tools have a particular potential and deserve close attention:

- *F1000* is the most developed and standardised tool with peer review assessments. Although it is not a free tool it gives a very clear idea of the potential of this new type of review of scientific publications. It also presents some conceptual and methodological aspects that still need to be better clarified (e.g. how are the peers selected and if there are potential biases in their selection, how are the journals and papers selected for review, etc.). The expansion of the covered disciplines and the exchangeability of the data would also be very positive assets in the potential incorporation of this tool for future standard research evaluation processes.
- *Microsoft Academic Search* is the database that allows more accessibility to the data, it presents more structured data (journal titles, organisational information), more indicators and more analytical options. The main limitations are the coverage (smaller compared to GS and WoS) and the problems in data standardisation. Solutions in these issues would considerably improve the possibilities of MAS as a 'technology of control'.
- *Total-Impact* is a tool to get different altmetrics with a simple search. The work of integrating different tools in one simple interface is a strong advantage. The possibility of searching by DOIs makes this tool suitable for more extensive downloading of data and for application in research evaluation. The main limitations are the lack of normalisation of the indicators and the limit to 250 publications per download.
- *PlosONE altmetrics* is doubtless the most advanced and developed provider of altmetrics, with possibilities of downloading the data, linkage through DOIs, etc. The main limitation is that it is limited to this one family of journals. However, the extension of this initiative to other journals and publishers would clearly contribute to the popularisation and standardisation of all these metrics.

Considering that most of the tools surveyed provide metrics for individual scholars, it is important to highlight that research assessment at the level of the individual scientist is without doubt the most difficult type of evaluation, where even traditional bibliometric and citation indicators have many limitations and problems (Costas, van Leeuwen, & Bordons, 2010). From a conceptual point of view, the main limitations and problems observed for traditional bibliometrics at the individual level are also relevant in the case of the new metrics and altmetrics reviewed in this report.

To give a few examples:

- Statistical limitations and sensitivity to outliers. At the individual level the amount of quantitative information is smaller (i.e. smaller numbers of citations, publications, readership, etc.). This makes the indicators less reliable. In this sense, differences in relative positions of rankings of researchers are frequently not significant and the value of rankings is very limited at this level. Besides a much higher precision is required in the collection and cleaning-up of data while the problems in identification of scientists due to common names and/or different variants of their names can seriously limit the evaluations.
- Problems in the comparability of researchers. At the individual level it is very important that researchers are assessed with their most similar colleagues; however it is very difficult to determine what is the disciplinary domain and who are the most comparable colleagues to whom a given researcher can be benchmarked. This is especially relevant for all the new metrics as most of them do not include any field classification or normalisation. Besides, the consideration of the age of researchers is also relevant in order to avoid/minimise the influence of generational differences.
- Manipulation of indicators and changes of behaviour of researchers. The introduction of inappropriate research assessment methodologies and especially the misuse of metrics may

result in modifications of the behaviour of scientists (e.g. changes in their selection of research topics, massive publication strategies, 'salami slicing', etc.). Therefore it is very important to consider the incentives that are introduced by the new indicators.

Therefore, any use of the new metrics and altmetrics at the individual level must be considered with great care and caution, and always as complements and in combination with other types of indicators (e.g. 'informed peer review').

Bearing all these problems in mind, it is important that the developers of these new tools realise that being mainly providers of indicators for individuals could 'trivialise' their image among scholars and managers as only providing 'technologies of narcissism', thus jeopardising their real potential value as strong and standardised assessment tools. For all this, their evolution and transformation from 'technologies of narcissism' to 'technologies of control' may contribute to their broader acceptance among the scientific community.

4.3.5 Future lines of research

We see various challenging lines of research that should be developed in the field of altmetrics and web tools for the measurement of other types of influence and impacts of scientific publications. Although there are many, most of these lines can be summarised in three main topics:

1. Conceptualisation of the new web metrics and altmetrics

Creating a sound conceptual framework is probably the most important line of research. It should be the basis for the validation of the new metrics developed. Among the main research questions that should be answered are: what do these new metrics measure? What dimensions of the scientific communication process and impact do they represent? How do these indicators relate among them? Can some of them be replaced or complemented by others? Are some of them predictors of some of the others? Are there differences by fields?

This line would also benefit from qualitative research involving the analysis of the perceptions and values that researchers actually attach to these metrics. Examples of potential research questions are: what do researchers prefer: being cited or being commented/tweeted (Ponte & Simon, 2011)? Would they review online papers that they do not like (especially if they have to make it public)? How is the process of selection of publications for online commenting or reviewing? Do they critically read everything that they tweet? How do they perceive all these new metrics? How do they perceive all these new activities related with the new metrics (blogging, tweeting, commenting, reviewing online, etc.)? And last but not least: what are the fundamental criteria for the quality of research and what metrics are the closest translations of those criteria?

Also, inevitably the question of 'exchange rates' among all these metrics would arise (Li, Thelwall, & Giustini, 2011), bringing questions such as: how many 'reads' in Mendeley equate a citation? Are readers or bookmarks the same as citations? How many posts in a blog can be equated to a paper? Is it the same being cited 10 times by other 10 WoS peer-reviewed papers or by 10 non-peer reviewed papers in GS? etc. However, if it can be proven that these different metrics actually capture different dimensions of scientific performance (which proponents of these metrics often argue and the factor analysis of the indicators in PLoS journals presented in this report also suggests), then calculating exchange rates would not really be so important. They could then be seen as related to different aspects of performance.

2. Standardisation of the existing tools and data

This would be a line of a more methodological nature. Aspects related to the proper availability and formatting of data, the detection of problems and limitations regarding the tools and data, as well as the establishment of general frameworks for analysis and benchmarking would be essential (Weller & Puschmann, 2011). For example, the determination of the number of different users, readers, bloggers, tweeters, etc., and their origin (country, fields, age group, etc.) would be necessary in order to be able to compare and contextualise the indicators. The detection of

potential ways of manipulation as well as the classification and proper normalisation of all the indicators are topics that will deserve much more attention as the available tools will continue their development and growth.

3. Production, normalisation and use of the new metrics

Research on how the new indicators should be calculated and normalised, their mathematical and statistical properties, their consistency and proper normalisation would be topics of high interest comparable to how these currently are in the field of bibliometrics and scientometrics.

The study of how these tools should be applied and considered for different research purposes (e.g. which metrics can be more informative in a hiring process, in a prospective exercise, or in monitoring tasks of research managers, etc.) and at different aggregation levels (e.g. how should they be applied for the analysis of individual researchers, for departments or universities, etc.?) would be also topics of great concern. The proper combination of all these new tools with other more traditional measures in order to obtain meaningful and valid analytical information would be also a topic of great interest. Since 'perfect' tools for evaluating the quality of research will not be found, important questions that must be addressed would be: how do we properly use the available methods and metrics? How do we interpret the indicators that are produced? How do we prevent rankings and evaluations to be seen as 'the truth and nothing but the truth', and to be used as a base for important political, policy and career decisions?

Finally, the use of these metrics to detect other types of research collaboration among scientists, as well as other types of networks of authors and ideas would be also possible, together with the possibility of enlarging the analysis of different types of co-occurrences (e.g. co-readership, co-blogging, co-tweeting, co-bookmarking, etc.) thus complementing and maybe also challenging the more traditional bibliometric developments in network research.

To summarise, we think that a concerted research programme in the dynamics, properties, and potential use of new web based metrics which relates these new measures to the already established indicators of publication impact may contribute to the development of more useful tools for the scientific and scholarly community. As Armbruster already noted in 2008:

> *'The growth and increasing complexity of global science poses a grand challenge to scientists: how to best organize the worldwide evaluation of research programmes and peers? For the 21st century we need not just information on science, but also meta-level scientific information that is delivered to the digital workbench of every researcher. Access, usage and citation metrics will be a major information service that researchers will need on an everyday basis to handle the complexity of science.'* (Armbruster, 2008)

5 References

Aguillo, I. F. (2012). Is Google Scholar useful for bibliometrics? A webometric analysis. *Scientometrics*, *In press*.

Aguillo, I. F., Ortega, J. L., & Fernández, M. (2008). Webometric Ranking of World Universities: Introduction, Methodology, and Future Developments. *Higher Education in Europe*, *33*(2-3), 233-244. doi:10.1080/03797720802254031

Armbruster, C. (2008). Access, Usage and Citation Metrics: What Function for Digital Libraries and Repositories in Research Evaluation? *Social Science Research Network Working Paper Series*. Retrieved from http://ssrn.com/abstract=1088453

Bar-Ilan, J., & Peritz, B. C. (2002). Informetric Theories and Methods for Exploring the Internet: An Analytical Survey of Recent Research Literature. *Library Trends*, *50*(3), 371-392.

Beel, J., & Gipp, B. (2010). Academic search engine spam and Google Scholar ' s resilience against it. *Journal of electronic publishing*, *13*(3). Retrieved from http://www.scienstein.org/publications/2010-Academic_search_engine_spam_and_Google_Scholars_resilience_against_it_-_preprint.pdf

Beniger, J. R. (1986). *The Control Revolution. Technological and Economic Origins of the Information Society*. Cambridge. Massachusets, and London, England: Harvard University Press.

Björneborn, L., & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics*, *50*(1), 65-82.

Borgman, C. L. (2007). *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. The MIT Press. Retrieved from http://www.amazon.com/dp/0262026198

Bulger, M., Meyer, E. T., Flor, G. de la, Terras, M., Wyatt, S., Jirotka, M., Eccles, K., et al. (2011). *Reinventing research? Information practices in the humanities*. *Network* (pp. 1-83). Retrieved from http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/information-use-case-studies-humanities

Butler, D. (2011). Experts question rankings of journals. *Nature*, *478*(7367), 20. Nature Publishing Group. doi:10.1038/478020a

Costas, R., & Leeuwen, T. N. V. (2011). Unraveling the complexity of thanking : preliminary analyses on the " Funding Acknowledgment " field of Web of Science database. *16th Nordic Workshop on Bibliometrics and Research Policy*. Aalborg: Royal School of Library and Information Science. Retrieved from http://itlab.dbit.dk/~nbw2011/index.php?s=programme

Costas, R., van Leeuwen, T. N., & Bordons, M. (2010). A bibliometric classificatory approach for the study and assessment of research performance at the individual level: The effects of age on productivity and impact. *Journal of the American Society for Information Science and Technology*, *61*(8), 1564-1581. doi:10.1002/asi.v61:8

Cronin, B., & Weaver, S. (1995). The praxis of acknowledgement: from bibliometrics to influmetrics. *Revista Española de Documentación Científica*, *18*(2), 172-177. Retrieved from http://www.libsearch.com/view/1144111

Davis, P. M. (2012). Tweets, and Our Obsession with Alt Metrics. *The Scholarly Kitchen*. Retrieved January 8, 2012, from http://scholarlykitchen.sspnet.org/2012/01/04/tweets-and-our-obsession-with-alt-metrics/

Dutton, W. H., Jeffreys, P. W., & Goldin, I. (2010). *World wide research: reshaping the sciences and humanities*. (William H. Dutton & P. W. Jeffreys, Eds.) (New., p. 408). The MIT Press. Retrieved from http://www.amazon.com/dp/0262513730

Eysenbach, G. (2011). Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact. *Journal of Medical Internet Research*, *13*(4). Journal of Medical Internet Research. doi:10.2196/jmir.2012

Grimm, J., & Grimm, W. (2004). *The Annotated Brothers Grimm*. (M. Tatar, Ed.) (p. 416). W. W. Norton & Company. Retrieved from http://www.amazon.com/Annotated-Brothers-Grimm-Books/dp/0393058484/ref=pd_sim_b_1

Groth, P., & Gurney, T. (2010). Studying Scientific Discourse on the Web using Bibliometrics: A Chemistry Blogging Case Study. Retrieved from http://journal.webscience.org/308/2/websci10_submission_48.pdf

Groth, P., Gibson, A., & Velterop, J. (2010). The anatomy of a nanopublication. *Information Services & Use*, *30*, 51-56. doi:10.3233/ISU-2010-0613

Gruzd, A., Goertzen, M., & Mai, P. (2011). *Survey Research Highlights: Trends in Scholarly Communication and Knowledge Dissemination in the Age of Online Social Media*. *Knowledge Creation Diffusion Utilization*. Halifax, Canada.

Harley, D., Acord, S., Earl-Novell, S., Lawrence, S., & King, C. J. (2010). *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines*. Retrieved from http://escholarship.org/uc/cshe_fsc

Hewson, C. (2003). *Internet research methods: a practical guide for the social and behavioural sciences*. London etc.: Sage.

Hine, C. (2005). *Virtual Methods: Issues in Social Research on the Internet*. Berg.

Hine, Christine. (2006). *New Infrastructures for Knowledge Production. Understanding e-science*. Hershey, USA: Information Science Publishing.

Hine, Christine. (2008). *Systematics as Cyberscience: Computers, Change, and Continuity in Science*. Cambridge, USA: MIT Press.

Howard, P. N., & Jones, S. (2004). *Society Online. The Internet in Context*. Thousand Oaks, London, New Delhi: Sage.

Huggett, S. (2012). F1000 Journal Rankings: an alternative way to evaluate the scientific impact of scholarly communications. *Research Trends*.

Jankowski, Nick (Ed.). (2009). *E-Research: Transformation in Scholarly Practice (Hardback) - Routledge*. Routledge. Retrieved from http://www.routledge.com/books/details/9780415990288/

KNAW. (2010). *Quality assessment in the design and engineering disciplines*. Retrieved from http://www.knaw.nl/Pages/DEF/27/160.bGFuZz1FTkc.html

KNAW. (2011). *Quality indicators for research in the humanities*. *Humanities*.

Li, X., Thelwall, M., & Giustini, D. (2011). Validating online reference managers for scholarly impact measurement. *Scientometrics*, 1-11. Akadémiai Kiadó, co-published with Springer Science+Business Media B.V., Formerly Kluwer Academic Publishers B.V. doi:10.1007/s11192-011-0580-x

Moed, H. F. (2005). *Citation analysis in research evaluation* (Vol. 9). Dordrecht: Springer.

Moed, H. F., & Glänzel, W. (2004). *Handbook of quantitative science and technology research : the use of publication and patent statistics in studies of S&T systems*. Dordrecht etc.: Kluwer Academic Publishers.

Oppenheim, C., Cronin, B., & Atkins, H. B. (2000). Do patent citations count? (pp. 405-432). Metford, NJ: Information Today Inc. ASIS Monograph Series.

Ponte, D., & Simon, J. (2011). Scholarly Communication 2.0: Exploring Researchers' Opinions on Web 2.0 for Scientific Knowledge Creation, Evaluation and Dissemination. *Serials Review*, *37*(3), 149-156. Elsevier Inc. doi:10.1016/j.serrev.2011.06.002

Priem, J., & Hemminger, B. H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social Web. *First Monday*, *15*(7). Retrieved from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2874

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). altmetrics: a manifesto – altmetrics.org. Retrieved January 8, 2012, from http://altmetrics.org/manifesto/

Roosendaal, H. E., Zalewska-Kurek, K., & Geurts, P. A. T. M. (2010). *Scientific Publishing: From Vanity to Strategy* (p. 190). Chandos Publishing (Oxford) Ltd. Retrieved from http://www.amazon.co.uk/Scientific-Publishing-Strategy-Hans-Roosendaal/dp/1843344904

Rousseau, R. (1998). Sitations: an exploratory study. *Cybermetrics*, *1*(1), 1. Retrieved from http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html

Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, *99*, 178-182.

Spaapen, J., & van Drooge, L. (2011). Introducing "productive interactions" in social impact assessment. *Research Evaluation*, *20*(3), 211-218. doi:Article

Tatum, C., & Jankowski, N. (2012). Beyond Open Access: a Framework for Openness in Scholarly Communication. In Paul Wouters, A. Beaulieu, A. Scharnhorst, & S. Wyatt (Eds.), *Virtual Knowledge*. MIT Press.

The Virtual Knowledge Studio:, Wouters, P., Vann, K., Scharnhorst, A., Ratto, M., Hellsten, I., Fry, J., et al. (2008). Messy Shapes of Knowledge-STS Explores Informatization, New Media, and Academic Work (Vol. 3, pp. 319-351; 14). Cambridge, Mass: MIT Press.

Thelwall, Mike. (2005). *Link Analysis: An Information Science Approach*. San Diego: Academic Press.

Thelwall, M., Wouters, P., & Fry, J. (2008). Information-centered research for large-scale analyses of new information sources. *Journal of the American Society for Information Science and Technology*, *59*(9), 1523-1527.

Van Noorden, R. (2010). Metrics: A profusion of measures. *Nature*, *465*(7300), 864-6. Nature Publishing Group. Retrieved from http://www.nature.com/news/2010/100616/full/465864a.html

Van Raan, A. (Ed.). (1988). *Handbook of Quantitative Studies of Science and Technology*. Amsterdam: Elsevier Science Publishers.

Weller, K., & Puschmann, C. (2011). Twitter for Scientific Communication: How Can Citations/References be Identified and Measured? *Proceedings of the ACM WebSci'11*. Retrieved from http://journal.webscience.org/500/2/Fig1.png

Williams, R., Pryor, G., Bruce, A., Macdonald, S., Marsden, W., Calvert, J., Dozier, M., et al. (2009). *Patterns of information use and exchange : case studies of researchers in the life sciences* (pp. 1-56). Retrieved from http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/patterns-information-use-and-exchange-case-studie

Willinsky, J. (2006). *The Access Principle: The Case for Open Access to Research and Scholarship*. The MIT Press. Retrieved from http://www.amazon.com/dp/0262232421

Wouters, Paul, Bar-Ilan, J., Thelwall, M., Aguillo, I. F., Must, Ü., Havemann, F., Kretschmer, H., et al. (2010). *Academic Careers Understood through Measurement and Norms (ACUMEN)* (pp. 1-39).

Wouters, Paul, Beaulieu, A., Scharnhorst, A., & Wyatt, S. (2012). *Virtual Knowledge*. The MIT Press.

van Raan, A. F. J. (2004). Sleeping Beauties in science. *Scientometrics*, *59*(3), 461-466.