International Journal of Scientometrics, Informetrics and Bibliometrics
ISSN 1137-5019

October 8, 2012

VOLUME 16 (2012): ISSUE 1. PAPER 4

Repositories in Google Scholar Metrics or what is this document type doing in a place as such?

Emilio Delgado López-Cózar* and Nicolás Robinson-García

EC3 Research Group: Evaluación de la Ciencia y la Comunicación Científica
Facultad de Comunicación y Documentación, Colegio Máximo de Cartuja s/n, Universidad de Granada
18071 Granada, **Spain**
E-mail: **edelgado@ugr.es**; **elrobin@ugr.es**

**To whom all correspondence should be addressed*

Abstract

The present paper analyzes GS Metrics, Google's newest product aiming at ranking journals according to their H-Index. Specifically, we analyze GS Metrics' decision of considering journals and repositories as equal and therefore, including them in the product. In this sense, the authors position themselves against this decision and provide several arguments of different nature warning against the shortcomings this product has. The first one is of a conceptual nature and is related to the definition of journal and repository. Secondly, they refer at the methodological issues mixing repositories and journals can bring out. Then, they deepen on many other flaws GS Metrics presents. Finally, GS Metrics and its possible use as an evaluation tool are discussed and possible solutions to its shortcomings are provided.

Keywords

Google Scholar Metrics, Google Scholar, Scientific Journals, Repositories, Citation analysis, Bibliometrics, h-index, Evaluation, Ranking

1. Introduction

The emergence of the World Wide Web has led to new and unexpected possibilities for scholarly communication. The shift from a controlled and unique channel for communication among scientists to the wide scope of communication opportunities the Internet offers, unleashed in the early 1990s a fragmentation within the research community and publishers leading to the well-known Open Access movement (Albert, 2006). One of the first consequences derived from this fragmentation was the creation of subject-based repositories where research communities could concentrate their output and offer it for free, without any filtering or control. The development of **ArXiv** in 1991 by Paul Ginsparg, signifies the beginning of such a change which can now be considered as an expansive reality within the research community as this repository along with others such as **RePEc** (Krichel, 2000) for instance, are fully established as sources for academic information. However, such reinforcement is not only due to researchers' will, the adherence of some of the most renowned universities worldwide such as Cornell, Harvard or Massachusetts Institute of Technology; played a key role by creating institutional repositories and fomenting guidelines supporting Open Access (Albert, 2006).

In this context, the launch of Google in 1998[1] meant a milestone in the information retrieval field, and the development of new products such as Google Scholar, signified a breakthrough of the scholarly communication system as they provided the perfect path for easily accessing to academic information. Scholarly journals which were accessed via subscription and which had been until then the only providers of scientific knowledge; were now obstacles that prevented scholars from accessing to their contents. The effects these changes are having in the scholarly communication system are yet to be seen, but there is no doubt of the existence of a fragmentation of journals to papers similar to the transition experienced in the 17th Century from monographs to journals.

In this sense and despite still being the main scientific communication channel, journals are starting to suffer serious transformations which can be tracked via the relation between articles' impact and the traditional Journal Impact Factor (Lozano, Larivière & Gingras, 2012). These changes may have deep consequences in the research evaluation field as they influence not only researchers' behavior but also they endanger the use of journals as proxies for evaluating researchers' performance. Until recently, bibliometric studies have been subjected to the use of the Web of Science and Scopus databases as the only data sources for performing bibliometric studies. From the moment Google Scholar starts to offer citation data, it transforms into something similar to the traditional citation indexes (Torres-Salinas, Ruiz-Pérez & Delgado López-Cózar, 2009). Since this happens, it may well be used as another source to add which also may finish with many of the limitations of the others, such as the lack of coverage for Social Sciences and Arts & Humanities (Moed, 2005; Harzing, 2012a) or the English-language biases (van Leeuwen et al, 2001).

However, opinions differ on the capability of Google Scholar to offer the appropriate characteristics in order to assure that data is reliable. In this sense, Harzing (2012b) demonstrates how Google Scholar may well be used as an alternative source for citation analysis as it is stable over time, has a comprehensive coverage and seems less biased for the Social Sciences fields. On the other hand, Aguillo (2012) warns of the lack of control this database has, which makes it dangerous to be used for bibliometric purposes. In fact, this lack of control was tested by Beel & Gipp (2010) who manipulated real and faked papers adding words to its metadata in order to alter their ranking in the results page of Google Scholar in order to gain more visibility and, eventually, more citations. Although they acknowledge that this type of 'academic spam' may not be very likely as it requires technical skills and has no immediate benefits for researchers, a further study by Delgado López-Cózar, Robinson-Garcia & Torres-Salinas (2012) suggested how easy it may be for researchers to increase their citations by introducing faked papers in Google Scholar without requiring any technical skill.

But the emergence of Google Scholar's new products GS Citations and GS Metrics (Cabezas-Clavijo & Delgado López-Cózar, 2012) opens a window to the use of Google Scholar as a source for bibliometric studies, demonstrating the intention of the company to become a serious competitor to the other aforementioned databases. After a deep analysis, Jacsó (2010) shows a very critical position regarding GS Citations as it contains many and serious errors as a result of applying automatic procedures for processing the bibliometric data, which makes it highly unreliable. Although, Harzing (2012b) indicates most of these problems were rapidly corrected, the evidence given by Delgado López-Cozar et al (2012) warns over the lack of control denounced by Aguillo (2012) which has also contaminated these other products.

Regarding GS Metrics, no study has been published analyzing its reliability as an alternative to the journal rankings provided by Thomson Reuters' Journal Citation Reports, however some have just been published describing the tool (Jacsó, 2012). In a previous study, Cabezas-Clavijo & Delgado López-Cózar (2012) analyzed GS Metrics' characteristics and questioned the unreasoned decision taken by Google when mixing journals and repositories in their rankings. In this paper we offer new arguments against this decision and critically review the quality of the product. For this, the study is structured as follows. Firstly, we will analyze the conceptual differences between a scientific journal and a repository, deepening in the characteristics that make them unique products and, therefore incomparable. Then, we will conduct a further analysis deepening on the technical and methodological limitations that must be assumed when comparing repositories with journals. In section 4, we analyze the quality of the product. Finally, we conclude with a discussion over the results, Google's problems when establishing the correct control mechanisms and how these problems could be solved.

2. Repositories vs Scientific Journals

The main problem with comparing repositories and journals in a bibliometric product such as GS Metrics has to do with the different natures each of them have. For this purpose, it is necessary to understand the differences they have and why it is nonsensical to compare them. In this sense, a repository can be defined in its broadest meaning, as a digital storage facility for depositing academic publications. That is, from research papers (preprints, post-prints or final version), working papers, technical reports, dissertations, master theses or academic works, to teaching material (course programmes, lecture presentations, graphical content, etc.), emphasizing the various document types accepted which include not just written material but also audiovisual material. On the other hand, a scientific journal is a highly specialized publication which contains mainly research papers. Although it is true that journals may include other document types (editorial material, letters, notes, book reviews…), their core is formed by research articles. Repositories can fall into a wide scope of types, depending on the institution in charge, or to the research topics they cover. From the whole scientific universe to specific research fields (for example; **SSRN**, the Social Science Research Network for Social Sciences or ArXiv for Physics, Astrophysics, Mathematics, Computer Science and Statistics) or subject domains (RePEc and **NBER** for Economics, Business and Finance , for instance). However, scientific journals cover specific and narrow fields of endeavour. In fact, their level of specialization can vary from a whole discipline to a subject domain or a single topic.

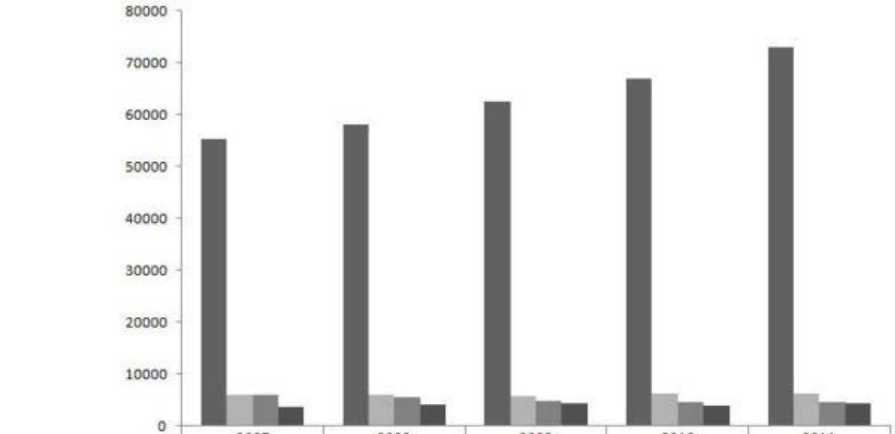
The main goals of repositories are storing and preserving documents on the one hand, and disseminating and retrieving the stored documents on the other. The objective of a journal is also to preserve and disseminate, it plays an important role as a public record and communication media of the scientific activity, but it also plays an important role validating and certifying it. Thanks to its severe and systematic process of selection and evaluation of manuscripts, it serves as a filtering and monitoring tool of scientific knowledge. Journals are vital for allowing scientific progress as they guarantee quality assurance by vouching for

the novelty, originality, relevance and validity of their publications. That is, they insure that their publications are scientifically viable and attend to the academic standards of the scientific method by means of submitting them to a just and impartial peer review process.

Therefore it is clearly stated that we are discussing two products with a completely different nature, purpose and outlook. Submitting a document to a repository is not the same as publishing in a journal. It is inadmissible for GS Metrics to put on a same level a 'container', - as repositories lack of any scientific control, - with a scientific registry where manuscripts are submitted to the research community's judgment. Repositories deliver visibility and promotion; they entail an enormous showcase where authors can promote their output. They assure promotion as they catalogue and classify documents with great precision and abundant metadata that ease their future retrieval by search engines. In other words, they play the role libraries have always played with their catalogue. In this context, to compare a repository with a journal is like comparing great department stores where all kinds of things can be bought, with a specialized shop where only one type of product can be bought. In science, as in other spheres of life, we must only compare what can be compared. It is as unreasonable treating unequally those which are equal as it is to treat equally those which are not.

3. The H-index: A size dependent indicator

The main technical reason warning against considering journals and repositories as equal has to do with their size. Many bibliometric indicators and especially the H-Index (Cabezas-Clavijo & Delgado López-Cózar, 2012), are size-dependent, that is, the number of papers contained may greatly influence their performance, rewarding the biggest ones (repositories) and not necessarily the best ones. This phenomenon can be easily tracked in figure 1. On the one hand, we selected a subject-based repository (ArXiv) which is mainly focused on the field of Physics. We retrieved from the repository the total number of submissions between 2007 and 2011 which is the time period analyzed by GS Metrics. Then, we selected the three most productive journals for the field of Physics according to the latest edition of Thomson Reuters' Essential Science Indicators. Although ArXiv also includes papers from other disciplines, the point in our analysis is not so much to focus on disciplinary differences but on the different nature of these two products, that is journals and repositories. We obtained the number of published papers for the analyzed time period by searching in the Web of Science database and we analyzed the average of papers published per year according to the journals retrieved in the Essential Science Indicators. For instance, we observe how ArXiv has indexed for this five-year period, 10.45 times more documents than Physical Review B, the most productive journal for the area. When comparing with the two second most productive journals, Applied Physics Letters and Journal of Applied Physics, the differences are of 12.45 and 15.25 times the size of these journals respectively. But the difference is even bigger when comparing with the median output of journals in the field of Physics according to the Essential Science Indicators (2108 papers). In that case, ArXiv has 149.72 times the size of the average journal. Clearly, using the H-index to rank such different sources considering their size is a mistake, as it biases results in favor of repositories.



	2007	2008	2009	2010	2011
ArXiv	55246	58015	62485	66889	72978
Physical Review B	5917	5942	5829	6192	6313
Applied Physics Letters	5915	5560	4788	4555	4530
Journal of Applied Physics	3774	4214	4311	3942	4451

Figure 1. Number of documents published per year in ArXiv and the three most productive journals in the field of Physics

4. Analyzing the quality of GS Metrics

In order to demonstrate the poor quality of GS Metrics and how mixing repositories and journals is misleading, we will now conduct an exploratory analysis on a sample consisting on the top 40 most highly cited papers that contribute to feed the H-index of the top four repositories in English-language according to GS Metrics. Therefore we selected RePEc (4th position), ArXiv (5th position), Social Science Research Network (7th position) and NBER (34th position). We must point out that only these documents are selected as they are the only ones to which GS Metrics provides access (Cabezas-Clavijo & Delgado López-Cózar, 2012). Following, we emphasize our main findings:

Table 1. Documents per repository per source indicated in each repository
Sample of the top 40 most highly papers per repository between 2007-2011 according to GS Metrics

PUBLICATIONS	REPEC	NBER	ARXIV	SSRN	PUBLICATIONS	REPEC	NBER	ARXIV	SSRN
American Economic Review	6	6	-	-	**Open Access publ. from Tilburg Universit**	1	-	-	-
Journal of Economic Literature	5	5	-	1	Journal of Economic Growth	1	-	-	-
Journal of Economic Perspectives	3	5	-	1	Journal of Applied Econometrics	1	-	-	-
Books	-	3	-	6	Experimental Economics	1	-	-	-
The Review of Financial Studies	1	1	-	6	Econometric Reviews	1	-	-	-
Journal of Financial Economics	4	2	-	1	Social Science & Medicine	-	1	-	-
Physical Review Letters	-	-	7	-	Review of Economic Dynamics	-	1	-	-
The Quarterly Journal of Economics	5	1	-	-	Review of Agricultural Economics	-	1	-	-
Congresses	-	4	-	2	Journal of Human Resources	-	1	-	-
Journal of Political Economy	-	5	-	-	American Economic Journal: Microeconomics	-	1	-	-
Nature	-	-	5	-	American Economic Journal: Macroeconomics	-	1	-	-
Annual Review of Psychology	-	-	-	5	Statistics and Computing	-	-	1	-
Journal of Econometrics	2	1	-	1	Solid State Communications	-	-	1	-
Science	-	-	4	-	J. Phys.: Condens. Matter	-	-	1	-
Rev. Mod. Phys	-	-	4	-	Eur. Phys. J	-	-	1	-
Astrophys. J. Suppl.	-	-	4	-	Chin. Phys. Lett.	-	-	1	-
Review of Financial Studies	3	-	-	-	Annals of Statistics	-	-	1	-
Arxiv preprint	-	-	3	-	SIAM Review	-	-	1	-
American Economic Review	-	-	-	3	Physics Reports	-	-	1	-
The World Economy	1	-	-	1	Nature Materials	-	-	1	-
Journal of Finance	1	-	-	1	Political Analysis	-	-	-	1
Econometrica	1	1	-	-	Oxford Bulletin of Economics and Statistics	-	-	-	1
JHEP	-	-	2	-	New England Journal of Medicine	-	-	-	1
Astrophys. J.	-	-	2	-	Economic Journal	-	-	-	1
The Quarterly Journal of Economic	-	-	-	2	Academy of Management Review	-	-	-	1

CEPR Discussion Paper	-	-	-	**2**	World Bank Policy Research Working Paper	-	-	-	**1**
The Review of Economics and Statistics	**1**	-	-	-	Stanford University Graduate School of Business Research Paper	-	-	-	**1**
Staff General Research Papers	**1**	-	-	-	National Bank of Belgium Working Paper	-	-	-	**1**
Policy Research Working Paper Series	**1**	-	-	-					

Firstly, we find out that 95% of the documents stored in the repositories had also been published in scientific journals (89%), books (5%) or as proceedings papers (1%), although these percentages vary depending on the repository: REPEC (95%), ArXiv (92,5%), NBER (100%), Social Science Research Network (87,5%). That is, the most influential documents, those which determine repositories' impact, are in fact, papers published through the traditional scholarly communication channels (journals, editorials and congresses); only 5% of the documents can be considered as unique. If that is the case, one must wonder to whom where the citations referred: to the repository or to the journal? In table 1 we list the sources to which documents are referred. In bold we mark those which are not journals. As it can be observed only 4 documents from the sample were not published in any other place rather than the repository.

Secondly, many of the highly cited papers are duplicates. This is logical as a document can be stored in many repositories, while it can only be published in a journal. But it raises more doubts over how they have been processed. Are these duplicates of the same and final version of the manuscripts, or do they belong to different versions? The reader is referred to Appendix 1 where we show duplicates for a sample of papers formed by those with at least the same number of citations as the repository with the highest H-Index (which would be REPEC with 259), as it is the only way to establish comparisons. Funnily enough, we find great standardization differences but they all have the same number of citations. As shown in the example of table 2, some differ on the source or even in the publication year. Most of the duplicates include their journal's reference and they may be found in two repositories at the same time. Two curious cases are the one of "Psychology and Economics: Evidence from the Field" authored by Stefano DellaVigna which is found in three different repositories and "Market Liquidity and Funding Liquidity", authored by Brunnemeier & Pedersen, which appears in four different repositories. Do these duplicates correspond to the same version of the manuscript or are they different? And if they are different versions why do they receive the same number of citations? Can we trust in Google when assigning these citations? Also, the lack of transparency of the product which does not allow the reader to see the complete list of publications of each source but only the ones contributing to the h-index makes it impossible to see the percentage of duplicate manuscripts. There are too many questions for so few answers.

Table 2. Reproduction of "Psychology and Economics: Evidence from the Field" by Stefano DellaVigna and "Market Liquidity and Funding Liquidity" by Brunnemeier & Pedersen shown according to GS Metrics in REPEC, SSRN and NBER

		TITLE	CITES	YEAR	SOURCE
REPOSITORY	**SSRN**	Psychology and Economics: Evidence from the Field	445	2007	NBER Working Paper
	NBER	PSYCHOLOGY AND ECONOMICS: EVIDENCE FROM THE FIELD	445	2007	NBER WORKING PAPER SERIES 13420
	REPEC	Psychology and Economics: Evidence from the Field	445	2009	Journal of Economic Literature 47 (2), 315-72
	REPEC	Market Liquidity and Funding Liquidity	958	2009	Review of Financial Studies 22 (6), 2201-2238
	SSRN	Market Liquidity and Funding Liquidity	958	2008	
	NBER	MARKET LIQUIDITY AND FUNDING LIQUIDITY	958	2007	NBER WORKING PAPER SERIES 12939
	CEPR	MARKET LIQUIDITY AND FUNDING LIQUIDITY	958	2007	DISCUSSION PAPER SERIES CENTRE FOR ECONOMIC POLICY RESEARCH LONDON 6179 958 2007

The lack of normalization of bibliographic data among the different repositories in GS Metrics exemplified in table 2, leads to many key points that should be discussed about the consequences it may have and the deficiency of the standardization process. Among them, we must emphasize three. Firstly, there are acute differences between the data brought by each repository. While REPEC identifies the journals in which all published documents were issued, ArXiv and NBER do not do so. On the other hand, SSRN does not always indicate the source where the documents were published. Secondly, the dates given by repositories do not always match with the publication date. Although this may not be an error and in fact correspond with the year when the document was deposited or with other versions of the manuscript, but then it is not understandable how the citations can be the same in all cases. Thirdly, some titles are presented in capital letters and some not. A minor error that signifies the lack of interest this product shows on presenting reliable data.

Finally, we indicate a surprising finding regarding NBER; many of the documents retrieved in Google Scholar as belonging to this repository are missing from GS Metrics, meaning a significant flaw in the count of its H-index. In order to deepen in this fact, we used the following query: "site:nber.org" and limited between 2007 and 2011. Then, we manually checked documents retrieved with those which are shown by GS Metrics. Because GS Metrics' results are fixed to the launch date (April 1, 2012) and our search was performed in June, 2012; we decided to restrict our search for documents with more than 200 citations, considering that NBER's H-index is 115 and two months had elapsed between both searches, citations rates would have increased. This way, we ensured that the documents found were above the 115 threshold in the date of GS Metrics release. In figure 2 we show some of the results that can be found in Google Scholar and are missing in GS Metrics, however, the reader is referred to Appendix 2 where more examples can be found. We found a total of 36 documents with more than 200 citations ranging from 1215 citations to 209; meaning that if these documents had been included, NBER could have positioned itself at least in the 15th position of the ranking for publications in English.

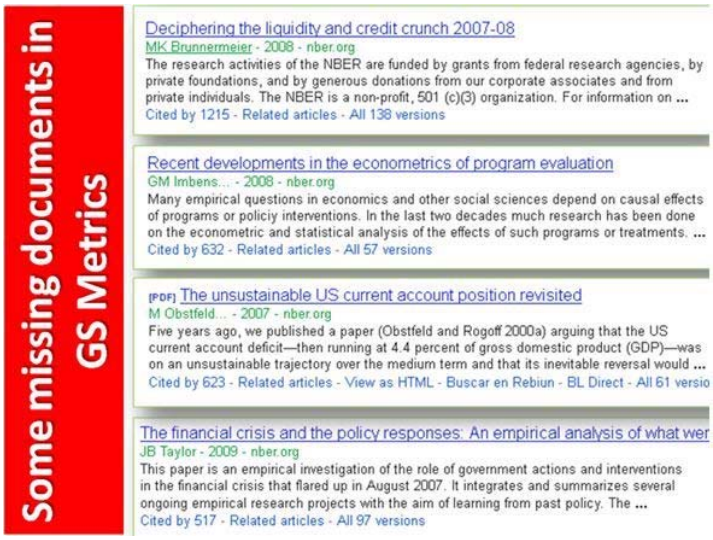


Figure 2. An example of documents retrieved in Google Scholar assigned to NBER and not in GS Metrics

But the most curious thing is that some of these missing documents are retrieved by other repositories and correctly assigned as papers stored in NBER. In figure 3 we observe how the paper entitled 'Deciphering the Liquidity and Credit Crunch 2007-08' is described as a NBER Working Paper when checking the documents contributing SSRN' H-index.

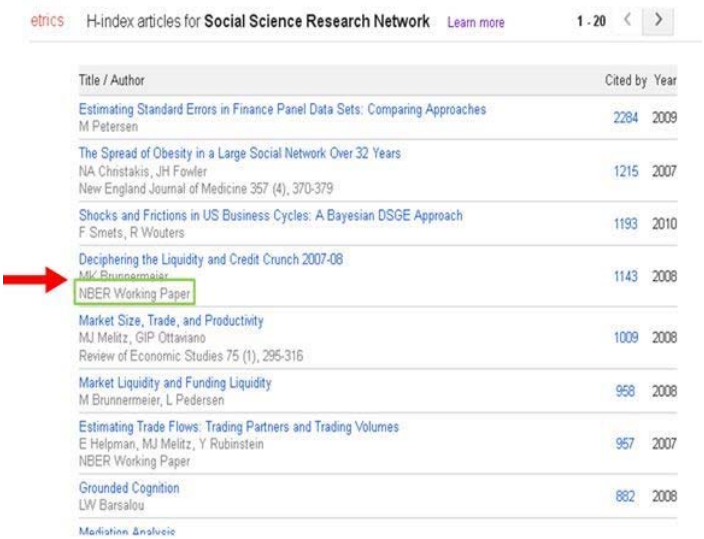


Figure 3. Detail of the paper 'Deciphering the Liquidity and Credit Crunch 2007-08' as displayed in GS Metrics for SSRN

5. Discussion and concluding remarks

In this paper we discuss and analyze Google's decision of including repositories as well as journals in the GS Metrics product, intended to ranking scientific publications according to their H-index. For this, we first discuss the changes that are taking place in the way researchers communicate and disseminate their scientific discoveries. Then, we briefly review the conceptual differences between repositories and scientific journals. Finally, we perform an exploratory analysis on the inconsistencies of the GS Metrics. We intend to emphasize the bibliographic and bibliometric contradictions committed by Google's decision of including repositories in their publications rankings, along with the many errors GS Metrics contains. Recently, Arlitsch & O'Brien (2012) discussed the difficulties Google Scholar has when processing correctly contents in repositories. Therefore, it is illogical for Google to venture in such a quest as releasing a bibliometric tool. Acknowledging possible errors by statements such as 'Dates and citation counts are estimated and are determined automatically by a computer program' is not enough when intending to elaborate a reliable and valid product.

As stated before, the current changes undergoing in scientific communication are leading to faster and faster communication habits in which researchers' results have a greater degree of obsolescence. The link between journals and papers is weakening more and more (Lozano, Larivière & Gingras, 2012) as a consequence of researchers' ability to disseminate their work prior to publication by means of storing them in repositories. In this context, Google Scholar has positioned itself as the main information source for accessing this type of materials. These changes also affect to research evaluation and bibliometric analyses which have traditionally used the Journal Impact Factor as a proxy for measuring the importance of papers.

In this sense, we believe that the emergence of alternative metrics focused on article level such as usage indicators (Bollen et al, 2009) or the so-called Altmetrics (Priem et al, 2012) - which are not only focused on articles but on the social impact and interactions between researchers through web 2.0 tools such as Twitter, blogs or Slideshare and are becoming a hot-topic on research evaluation (Sugimoto, 2012) - are leading to a separation between the container (repositories or journals) and the content gaining on accuracy when analyzing research output. However, the emergence of this new generation of indicators leads to greater challenges, as they still lack of a stable theoretical background as citation analysis does. What does usage really measures? Are downloads a proxy of impact? Do these indicators provide the same information traditional bibliometric indicators did? Is this information complementary or different?

Although it is true that this new communication behavior leads to papers to be cited before or without being published in journals, GS Metrics' decision to consider repositories as equals introduces many inconsistencies. In this sense, we consider that there are two possible solutions to these problems. One answer would be to delete repositories from their publication rankings, maybe generating a new one specifically for this type of information source (as it is already being done elsewhere with "**The Ranking Web of World Repositories**"). The other one would be to clean and reprocess the bibliographic data retrieved from repositories, excluding duplicates when they correspond to the same version of a manuscript and distinguishing between citations directed to journals and citations referred to repositories.

Acknowledgements

This research was sponsored by the Ministerio de Economía y Competitividad under grant HAR2011-30383-C02-02. Nicolás Robinson-García is currently supported by a FPU Grant from the Ministerio de Economía y Competitividad of the Spanish Government.

References

Aguillo, I (2012). Is Google Scholar useful for bibliometrics? A webometric analysis. **Scientometrics**, 91(2): 343-351.

Albert, KM (2006). Open access: implications for scholarly publishing and medical libraries. **Journal of the Medical Library Association**, 94(3): 253-262.

Arlitsch, K.; O'Brien, P (2012). Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar. **Library Hi Tech** 30; 1: 60-81

Beel, J. and Gipp, B. (2010). Academic search engine spam and Google Scholar's resilience against it. **Journal of Electronic Publishing**, 13(3).

Bollen, J.; Van de Sompel, H.; Hagberg, A. & Chute, R. (2009). A Principal Component Analysis of 39 Scientific Impact Measures. **PLOS One** 4(6): e6022.

Cabezas Clavijo, Á & Delgado López-Cózar, E (2012). Scholar Metrics: the impact of journals according to Google, just an amusement or a valid scientific tool? **EC3 Working Papers**, 1. Accessible at: **http://eprints.rclis.org/handle/10760/16836**.

Delgado López-Cózar, E., Robinson-Garcia, N. & Torres-Salinas, D. (2012). Manipulating Google Scholar Citations and Google Scholar Metrics: simple, easy and tempting. **EC3 Working Papers**, 6. Accessible at: **http://hdl.handle.net/10481/20469**.

Harzing, A-W (2012a). Document Categories in the ISI Web of Knowledge: Misunderstanding the Social Sciences? **Scientometrics**, accepted for publication. Accessible at: **http://www.harzing.com/download/isi.pdf**.

Harzing A-W (2012b). A preliminary test of Google Scholar as a source for citation data: A longitudinal study of Nobel Prize winners. **Scientometrics**, accepted for publication. Accessible at: **http://www.harzing.com/download/nobelists.pdf**.

Jacsó, P (2010). Google Scholar Author Citation Tracker: is it too little, too late? **Online Information Review**, 36(1): 126-141.

Jacsó, P (2012). Google Scholar Metrics for Publications – The software and content feature of a new open access bibliometric service. **Online Information Review**, 36(4): 604-619.

Krichel, T (2000). Working towards an Open Library for Economics: The RePEc project. **PEAK, Conference Arbor ,** 23-24 March 2000. Accessible at: **http://openlib.org/home/krichel/myers.html**.

Lozano, GA, Larivière, V. & Gingras, Y. (2012). The weakening relationship between the Impact Factor and papers' citations in the digital age. **ARXIV**, **arXiv:1205.4328v1**

Moed, HF (2005). **Citation Analysis in Research Evaluation**. Springer.

Priem, J, Parra, C, Piwowar, H, Groth, P and Waagmeester, A (2012). Uncovering impacts: a case study in using altmetrics tools. **Second International Conference on the Future of Scholarly Communication and Scientific Publishing** (Sepublica2012), Heraklion, Greece, May 27.

Sugimoto, CR (2012). Talking the Measure of Metrics: Interviews with Four ASIS&T Members. **Bulletin of the American Society for Information Science and Technology**, 38(6): 33-38.

Torres-Salinas, D.; Ruiz-Pérez, R. & Delgado López-Cózar, E. (2009). Google Scholar como herramienta para la evaluación científica. **El profesional de la información**, 18(5): 501-510.

Van Leeuwen, TN, Moed, HF, Tijssen, RJW, Visser, MS & van Raan, AFJ (2001). Language biases coverage of the Science Citation Index and its

consequences for international comparisons of national research performance. **Scientometrics**, 51(1): 335-346.

Footnotes

[1] <http://www.google.com/intl/en/about/company/>

Appendixes

Appendix 1. Sample of duplicate documents with at least 259 citations in repositories REPEC, SSRN, CEPR and NBER

REPOSITORY	PAPERS	CITATIONS	YEAR	SOURCE
REPEC	Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches	2284	2009	Review of Financial Studies 22 (1), 435-480
SSRN	Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches	2284	2009	
CEPR	SHOCKS AND FRICTIONS IN US BUSINESS CYCLES: A BAYESIAN DSGE APPROACH	1193	2007	DISCUSSION PAPER SERIES-CENTRE FOR ECONOMIC POLICY RESEARCH LONDON 6112
REPEC	Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach	1193	2007	American Economic Review 97 (3), 586-606
SSRN	Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach	1193	2010	
SSRN	Deciphering the Liquidity and Credit Crunch 2007-08	1143	2008	NBER Working Paper
REPEC	Deciphering the Liquidity and Credit Crunch 2007-2008	1143	2009	Journal of Economic Perspectives 23 (1), 77-100
REPEC	Market Size, Trade, and Productivity	1009	2008	Review of Economic Studies 75 (1), 295-316
SSRN	Market Size, Trade, and Productivity	1009	2008	Review of Economic Studies 75 (1), 295-316
REPEC	Market Liquidity and Funding Liquidity	958	2009	Review of Financial Studies 22 (6), 2201-2238
SSRN	Market Liquidity and Funding Liquidity	958	2008	
NBER	MARKET LIQUIDITY AND FUNDING LIQUIDITY	958	2007	NBER WORKING PAPER SERIES 12939
CEPR	MARKET LIQUIDITY AND FUNDING LIQUIDITY	958	2007	DISCUSSION PAPER SERIES-CENTRE FOR ECONOMIC POLICY RESEARCH LONDON 6179 958 2007
REPEC	Estimating Trade Flows: Trading Partners and Trading Volumes	957	2008	The Quarterly Journal of Economics 123 (2), 441-487
SSRN	Estimating Trade Flows: Trading Partners and Trading Volumes	957	2007	NBER Working Paper
NBER	ESTIMATING TRADE FLOWS: TRADING PARTNERS AND TRADING VOLUMES	957	2007	NBER WORKING PAPER SERIES 12927
CEPR	FIRMS IN INTERNATIONAL TRADE	742	2007	DISCUSSION PAPER SERIES-CENTRE FOR ECONOMIC POLICY RESEARCH LONDON 6277
REPEC	Firms in International Trade	742	2007	Working Papers
SSRN	Firms in International Trade	742	2007	CEPR Discussion Paper No. DP6277
NBER	FIRMS IN INTERNATIONAL TRADE	742	2007	NBER WORKING PAPER SERIES 13054
REPEC	All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors	740	2008	Review of Financial Studies 21 (2), 785-818
SSRN	All that Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors	740	2008	
REPEC	Regression Discontinuity Designs: A Guide to Practice	686	2007	NBER Working Papers
SSRN	Regression Discontinuity Designs: A Guide to Practice	686	2007	NBER Working Paper No. W13039
NBER	REGRESSION DISCONTINUITY DESIGNS: A GUIDE TO PRACTICE	686	2007	NBER WORKING PAPER SERIES 13039
REPEC	The Economic Consequences of Legal Origins	632	2008	Journal of Economic Literature 46 (2), 285-332
SSRN	The Economic Consequences of Legal Origins	632	2007	NBER Working Paper
NBER	THE ECONOMIC CONSEQUENCES OF LEGAL ORIGINS	632	2007	NBER WORKING PAPER SERIES 13608
REPEC	Exports and Productivity: A Survey of the Evidence from Firm-level Data	596	2007	The World Economy 30 (1), 60-82
SSRN	Exports and Productivity: A Survey of the Evidence from Firm-Level Data	596	2007	
REPEC	Recent Developments in the Econometrics of Program Evaluation	589	2009	Journal of Economic Literature 47 (1), 5-86
SSRN	Recent Developments in the Econometrics of Program Evaluation	589	2008	NBER Working Paper No. W14251
REPEC	Measuring and Explaining Management Practices Across Firms and Countries	584	2007	The Quarterly Journal of Economics 122 (4), 1351-1408
SSRN	Measuring and Explaining Management Practices Across Firms and Countries	584	2008	
REPEC	The Technology of Skill Formation	553	2007	IZA Discussion Papers
SSRN	The Technology of Skill Formation	553	2007	NBER Working Paper No. W12840
NBER	THE TECHNOLOGY OF SKILL FORMATION	553	2007	NBER WORKING PAPER SERIES 12840
REPEC	Governance Matters VI: Aggregate and Individual Governance Indicators, 1996-2006	552	2007	Policy Research Working Paper Series
SSRN	Governance Matters VI: Governance Indicators for 1996-2006	552	2007	
REPEC	Is the 2007 US Sub-Prime Financial Crisis So Different? An International Historical Comparison	509	2008	NBER Working Papers
SSRN	Is the 2007 US Sub-Prime Financial Crisis so Different? An International Historical Comparison	509	2008	NBER Working Paper
NBER	IS THE 2007 US SUB-PRIME FINANCIAL CRISIS SO DIFFERENT? AN INTERNATIONAL HISTORICAL COMPARISON	509	2008	NBER WORKING PAPER SERIES 13761
NBER	THE CONSEQUENCES OF MORTGAGE CREDIT EXPANSION: EVIDENCE FROM THE 2007 MORTGAGE DEFAULT CRISIS	454	2008	NBER WORKING PAPER SERIES 13936
REPEC	he Consequences of Mortgage Credit Expansion: Evidence from the 2007 Mortgage Default Crisis	454	2008	NBER Working Papers
SSRN	The Consequences of Mortgage Credit Expansion: Evidence from the US Mortgage Default Crisis	454	2007	
SSRN	Psychology and Economics: Evidence from the Field	445	2007	NBER Working Paper
REPEC	Psychology and Economics: Evidence from the Field	445	2009	Journal of Economic Literature 47 (2), 315-72
NBER	PSYCHOLOGY AND ECONOMICS: EVIDENCE FROM THE FIELD	445	2007	NBER WORKING PAPER SERIES 13420
SSRN	Distributional Effects of Globalization in Developing Countries	406	2007	NBER Working Paper
REPEC	Distributional Effects of Globalization in Developing Countries	406	2007	NBER Working Papers
NBER	DISTRIBUTIONAL EFFECTS OF GLOBALIZATION IN DEVELOPING COUNTRIES	406	2007	NBER WORKING PAPER SERIES 12885
NBER	THE PRODUCTIVITY ARGUMENT FOR INVESTING IN YOUNG CHILDREN	400	2007	NBER WORKING PAPER SERIES 13016
REPEC	The Productivity Argument for Investing in Young Children	400	2007	Working Papers
SSRN	The Productivity Argument for Investing in Young Children	400	2007	NBER Working Paper
SSRN	The Macroeconomic Effects of Oil Shocks: Why are the 2000s so Different from the 1970s?	376	2007	NBER Working Paper
NBER	THE MACROECONOMIC EFFECTS OF OIL SHOCKS: WHY ARE THE 2000S SO DIFFERENT FROM THE 1970S?	376	2007	NBER WORKING PAPER SERIES 13368
REPEC	The Macroeconomic Effects of Oil Shocks: Why are the 2000s so Different from the 1970s?	376	2007	CEPR Discussion Papers
NBER	REASSESSING THE INS AND OUTS OF UNEMPLOYMENT	350	2007	NBER WORKING PAPER SERIES 13421
SSRN	Reassessing the Ins and Outs of Unemployment	350	2007	NBER Working Paper
REPEC	Reassessing the Ins and Outs of Unemployment	350	2007	Review of Economic Dynamics
NBER	THE ECONOMICS AND PSYCHOLOGY OF PERSONALITY TRAITS	347	2008	NBER WORKING PAPER SERIES 13810

REPEC	The Economics and Psychology of Personality Traits	347	2008	UNU-MERIT Working Paper Series
SSRN	The Economics and Psychology of Personality Traits	347	2008	IZA Working Paper No. 3333
NBER	THE IMPACT OF UNCERTAINTY SHOCKS	341	2007	NBER WORKING PAPER SERIES 13385
SSRN	The Impact of Uncertainty Shocks	341	2007	NBER Working Paper
REPEC	The Impact of Uncertainty Shocks	341	2007	NBER Working Papers
NBER	THIS TIME IS DIFFERENT: A PANORAMIC VIEW OF EIGHT CENTURIES OF FINANCIAL CRISES	341	2008	NBER WORKING PAPER SERIES 13882
REPEC	This Time is Different: A Panoramic View of Eight Centuries of Financial Crises	341	2008	NBER Working Papers
SSRN	This Time is Different: A Panoramic View of Eight Centuries of Financial Crises	341	2008	NBER Working Paper
NBER	THE MACROECONOMIC EFFECTS OF TAX CHANGES: ESTIMATES BASED ON A NEW MEASURE OF FISCAL SHOCKS	331	2007	NBER WORKING PAPER SERIES 13264
REPEC	The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks	331	2007	NBER Working Papers
SSRN	The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks	331	2007	NBER Working Paper
NBER	FOREIGN CAPITAL AND ECONOMIC GROWTH	330	2007	NBER WORKING PAPER SERIES 13619
SSRN	Foreign Capital and Economic Growth	330	2007	IZA Discussion Paper No. 3186
REPEC	Foreign Capital and Economic Growth	330	2007	NBER Working Papers
NBER	ACCOUNTING FOR GROWTH: COMPARING CHINA AND INDIA	314	2007	NBER WORKING PAPER SERIES 12943
REPEC	Accounting for Growth: Comparing China and India	314	2007	NBER Working Papers
NBER	POLICY DISTORTIONS AND AGGREGATE PRODUCTIVITY WITH HETEROGENEOUS PLANTS	314	2007	NBER WORKING PAPER SERIES 13018
SSRN	Policy Distortions and Aggregate Productivity with Heterogeneous Plants	314	2007	NBER Working Paper
REPEC	Policy Distortions and Aggregate Productivity with Heterogeneous Plants	314	2007	NBER Working Papers
NBER	FINANCIAL INTEGRATION, FINANCIAL DEEPNESS AND GLOBAL IMBALANCES	309	2007	NBER WORKING PAPER SERIES 12909
REPEC	Financial Integration, Financial Deepness and Global Imbalances	309	2007	NBER Working Papers
CEPR	FINANCIAL INTEGRATION, FINANCIAL DEEPNESS AND GLOBAL IMBALANCES	309	2007	DISCUSSION PAPER SERIES-CENTRE FOR ECONOMIC POLICY RESEARCH LONDON 6149
SSRN	Financial Integration, Financial Deepness and Global Imbalances	309	2007	NBER Working Paper
NBER	BANK GOVERNANCE, REGULATION, AND RISK TAKING	290	2008	NBER WORKING PAPER SERIES 14113
SSRN	Bank Governance, Regulation, and Risk Taking	290	2008	
REPEC	Bank Governance, Regulation, and Risk Taking	290	2008	NBER Working Papers
NBER	HOUSING AND MONETARY POLICY	289	2007	NBER WORKING PAPER SERIES 13682
REPEC	Housing and Monetary Policy	289	2007	Discussion Papers
SSRN	Housing and Monetary Policy	289	2007	NBER Working Paper No. W13682
NBER	A BLACK SWAN IN THE MONEY MARKET	276	2008	NBER WORKING PAPER SERIES 13943
SSRN	A Black Swan in the Money Market	276	2008	NBER Working Paper
REPEC	A Black Swan in the Money Market	276	2008	NBER Working Papers
NBER	IS WELL-BEING U-SHAPED OVER THE LIFE CYCLE?	273	2007	NBER WORKING PAPER SERIES 12935
REPEC	Is well-being U-shaped over the life cycle?	273	2007	Social Science & Medicine 66 (8), 1733-1749
SSRN	Is Well-Being U-Shaped Over the Life Cycle?	273	2007	NBER Working Paper No. W12935
NBER	THE TERM STRUCTURE OF REAL RATES AND EXPECTED INFLATION	267	2007	NBER WORKING PAPER SERIES 12930
SSRN	The Term Structure of Real Rates and Expected Inflation	267	2007	NBER Working Paper
REPEC	The Term Structure of Real Rates and Expected Inflation	267	2007	Journal of Finance 63 (2), 797-849
NBER	THE ECONOMICS, TECHNOLOGY AND NEUROSCIENCE OF HUMAN CAPABILITY FORMATION	264	2007	NBER WORKING PAPER SERIES 13195
REPEC	The Economics, Technology and Neuroscience of Human Capability Formation	264	2007	IZA Discussion Papers
SSRN	The Economics, Technology and Neuroscience of Human Capability Formation	264	2007	NBER Working Paper
NBER	CENTRAL BANK COMMUNICATION AND MONETARY POLICY: A SURVEY OF THEORY AND EVIDENCE	259	2008	NBER WORKING PAPER SERIES 13932
SSRN	Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence	259	2008	ECB Working Paper No. 898
REPEC	Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence	259	2008	NBER Working Papers
NBER	STOCKS AS LOTTERIES: THE IMPLICATIONS OF PROBABILITY WEIGHTING FOR SECURITY PRICES	259	2007	NBER WORKING PAPER SERIES 12936
REPEC	Stocks as Lotteries: The Implications of Probability Weighting for Security Prices	259	2007	American Economic Review 98 (5), 2066-2100
SSRN	Stocks as Lotteries: the Implications of Probability Weighting for Security Prices	259	2007	AFA 2005 Philadelphia Meetings Paper

Appendix 2. Sample of papers belonging to NBER but not included in GS Metrics

PAPER	AUTHORS	CITATIONS
Deciphering the liquidity and credit crunch 2007-08	MK Brunnermeier - 2008 - nber.org	Cited by 1215
Deaths: final data for 2005	HC Kung, DL Hoyert, J Xu, SL Murphy - Natl Vital Stat Rep, 2008 - nber.org	Cited by 695
The unsustainable US current account position revisited	M Obstfeld... - 2007 - nber.org	Cited by 623
Recent developments in the econometrics of program evaluation	GM Imbens... - 2008 - nber.org	Cited by 632
The aftermath of financial crises	CM Reinhart... - 2009 - nber.org	Cited by 535
The financial crisis and the policy responses: An empirical analysis of what went wrong	JB Taylor - 2009 - nber.org	Cited by 517
Economic growth and subjective well-being: Reassessing the Easterlin Paradox	B Stevenson... - 2008 - nber.org	Cited by 436
Regression discontinuity designs in economics	DS Lee... - 2009 - nber.org	Cited by 455
From world banker to world venture capitalist: US external adjustment and the exorbitant privilege	PO Gourinchas... - 2007 - nber.org	Cited by 405
The panic of 2007	GB Gorton - 2008 - nber.org	Cited by 380
Deaths: final data for 2004	AM Miniño, MP Heron, SL Murphy... - National vital statistics ..., 2007 - nber.org	Cited by 381
An anatomy of international trade: Evidence from French firms	J Eaton, S Kortum... - 2008 - nber.org	Cited by 369
Social security programs and retirement around the world: The relationship to youth employment, introduction and summary	J Gruber, K Milligan... - 2009 - nber.org	Cited by 359
A new data set of educational attainment in the world, 1950–2010	RJ Barro... - 2010 - nber.org	Cited by 361
The Chinese approach to capital inflows: patterns and possible explanations	E Prasad... - 2007 - nber.org	Cited by 291
Importers, exporters and multinationals: A portrait of firms in the US that trade goods	AB Bernard, JB Jensen... - 2009 - nber.org	Cited by 284
Causes and Consequences of the Oil Shock of 2007-08	JD Hamilton - 2009 - nber.org	Cited by 325
When is the government spending multiplier large?	L Christiano, M Eichenbaum... - 2009 - nber.org	Cited by 302
A black swan in the money market	JB Taylor... - 2008 - nber.org	Cited by 287
Growth in a Time of Debt	CM Reinhart... - 2010 - nber.org	Cited by 311
Understanding crude oil prices	JD Hamilton - 2008 - nber.org	Cited by 271
The role of boards of directors in corporate governance: A conceptual framework and survey	R Adams, BE Hermalin... - 2008 - nber.org	Cited by 269

Civil war	C Blattman... - 2009 - nber.org	Cited by 266
An international comparison of capital structure and debt maturity choices	JPH Fan, S Titman... - 2010 - nber.org	Cited by 251
Credit constraints, heterogeneous firms, and international trade	K Manova - 2008 - nber.org	Cited by 255
Income, aging, health and well-being around the world: Evidence from the Gallup World Poll	A Deaton - 2010 - nber.org	Cited by 249
No-arbitrage Taylor rules	A Ang, S Dong... - 2007 - nber.org	Cited by 235
Banking crises: an equal opportunity menace	CM Reinhart... - 2008 - nber.org	Cited by 233
Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development	AS Deaton - 2009 - nber.org	Cited by 224
A global perspective on external positions	PR Lane... - 2007 - nber.org	Cited by 214
The leverage cycle	J Geanakoplos - 2009 - nber.org	Cited by 230
Monetary policy and stock market boom-bust cycles	L Christiano, C Ilut, R Motto, M Rostagno... - 2008 - nber.org	Cited by 213
Will the euro eventually surpass the dollar as leading international reserve currency?	M Chinn... - 2007 - nber.org	Cited by 214
Capital controls, sudden stops, and current account reversals	S Edwards - 2007 - nber.org	Cited by 209
Leveraged buyouts and private equity	SN Kaplan... - 2008 - nber.org	Cited by 224

Received 29/June/2012
Accepted 30/August/2012

WEBQL

Copyright information | | Webmaster | Sitemap
Updated: 10/08/2012

TOP