Arch. Immunol. Ther. Exp., 2008, **56**, 363–371
PL ISSN 0004-069X
DOI 10.1007/s00005-008-0043-0

"Publish or Perish" as citation metrics used to analyze scientific output in the humanities: International case studies in economics, geography, social sciences, philosophy, and history

Audrey Baneyx

Institut Francilien "Recherche, Innovation et Société" (IFRIS)[1], Paris, France

Received: 2008.10.16, **Accepted:** 2008.10.27, **Published online first:** 2008.12.01

Abstract

Traditionally, the most commonly used source of bibliometric data is the Thomson ISI Web of Knowledge, in particular the (Social) Science Citation Index and the Journal Citation Reports, which provide the yearly Journal Impact Factors. This database used for the evaluation of researchers is not advantageous in the humanities, mainly because books, conference papers, and non-English journals, which are an important part of scientific activity, are not (well) covered. This paper presents the use of an alternative source of data, Google Scholar, and its benefits in calculating citation metrics in the humanities. Because of its broader range of data sources, the use of Google Scholar generally results in more comprehensive citation coverage in the humanities. This presentation compares and analyzes some international case studies with ISI Web of Knowledge and Google Scholar. The fields of economics, geography, social sciences, philosophy, and history are focused on to illustrate the differences of results between these two databases. To search for relevant publications in the Google Scholar database, the use of "Publish or Perish" and of CleanPoP, which the author developed to clean the results, are compared.

Key words: Web of Science, Google Scholar, citation analysis, bibliometrics, research evaluation in the humanities, Publish or Perish.

Corresponding author: Audrey Baneyx, Institut Francilien Recherche Innovation Société (IFRIS), Université Paris-Est Marne La Vallée, e-mail: baneyx@ifris.org

INTRODUCTION

We live in an age of metrics. Citation analysis now has important implications for grants, funding, and tenure decisions. It allows a researcher to follow the development and impact of an article through time by looking backwards at the references the author cited and forwards to those authors who then cite the article [2]. Citation analysis has become a strategic type of information for individuals, laboratories, institutions, and even countries. Eugene Garfield made widespread use of citation analysis in the academic world possible through his creation of three citation indices, the Science, Humanities, and Social Science Citation Indices, which were transformed into an electronic version called the Web of Science (WoS)[2], which is part of the Thomson Institute for Scientific Information (ISI) Web of Knowledge (WoK)[3] [6, 7]. Traditionally, these indices are the most commonly used sources of bibliometric data, in particular the (Social) Science Citation Index and the Journal Citation Reports, which provide the yearly Journal Impact Factors. Until recently, ISI databases were the only tools for locating citations and conducting citation analyses.

The WoS has proved itself in the natural sciences, but in the humanities, especially for scientists who do not publish in English, its use is not so advantageous. To report the activity and scientific production of scientists in the humanities, you have to take into account their specificities. Journals are various, heterogeneous, and

[1] http://www.ifris.org/
[2] http://scientific.thomson.com/products/wos/
[3] http://www.isiwebofknowledge.com/

distinct. Some are aimed at a broad, general, international readership, others are more specialized in their content and implied audience. It is necessary to be able to retrieve the publications in national languages. The publication of books is one of the most important means to spread knowledge in the humanities and social sciences, but they are not well indexed in the WoS. Coverage of the humanities is therefore difficult to assess as a whole, which is particularly prejudicial to researchers in this field.

Since November 2004, Google Scholar (GS)[4] has been considered a possible alternative to ISI WoK. It provides a means to search for scholarly literature broadly across many disciplines and sources: peer-reviewed papers, theses, books, abstracts, and articles from academic publishers, professional societies, preprint repositories, universities, and other scholarly organizations. GS sorts articles, weighing the full text of each article, the author, the publication in which the article appears, and how often the piece has been cited in other scholarly literature.

For obvious practical reasons, bibliographic databases can contain extracts of only scientific literature. The ISI citation databases are designed to cover scientific research journals with the greatest impact. The more common critics of ISI citation databases argue that they cover mainly North American, Western European, and English-language titles, the number of indexed journals is relatively weak, they do not count citations from books and most conference proceedings, and they provide different coverage among research fields [16]. GS also contains citation information, but it includes a less quality-controlled collection of publications from different types of Web documents [14].

The comparison of WoS and GS for the production of individual indicators in the humanities has not been investigated in a systematic way. In this preliminary study, I present elements of comparison on a small number of cases: authors recognized by their peers as being internationally renowned in the academic fields of economics, social sciences, philosophy, geography, and history. I focus this study on the differences between two databases, ISI WoK and GS, for citation analyses of researchers' scientific production.

I show that because of the broader range of data sources, the use of GS generally results in more comprehensive citation coverage in the humanities. Its use is particularly beneficial to academics publishing in sources that are not well covered in ISI, such as books. Another important practical reason for using GS is that it is freely available to anyone with an Internet connection and is generally praised for its speed [3]. The WoS is available only to those scientists whose institutions are able and willing to bear the (quite substantial) subscription costs of this and other databases in the Thomson ISI WoK. According to Pauly and Stergiou [17], free access to data provides more transparency in tenure reviews, funding, and other science policy issues and allows citation counts and their analyses to be performed and duplicated by anyone.

I have chosen to characterize scientific production in the humanities with three metrics: the number of papers, the number of citations, and the h-index. Since Hirsch's first publication of the h-index in 2005 [9], this new measurement of academic impact has generated widespread interest. The h-index is defined as follows: "A researcher has an index h if h of his/her Np papers has/have at least h citations each, and the other (Np-h) papers have no more than h citations each." It is designed to measure the cumulative impact of a researcher's production by looking at the number of citations that his or her work has received. The advantage of the h-index is that it combines an assessment of both quantity (number of papers) and visibility (citations of these papers, in other words, the impact on the community). In fact, a researcher cannot have a high h--index without publishing a substantial number of papers and these papers need to be cited by others in order to count for the h-index.

To calculate metrics we present two tools. The first one, "Publish or Perish (PoP)"[5], was developed by Prof. Anne-Wil Harzing of the University of Melbourne. The results of this tool (based on GS) are still not as clean as they could be. I have therefore developed an additional tool, "CleanPoP"[6], for the purpose of correcting and improving the results. This study is exploratory and will be extended in the near future.

MATERIALS AND METHODS

Materials

Selection of a sample of high-level scientists. The sample consists of twelve senior, internationally renowned researchers who are references in their communities: history (2 researchers), sociology (4), economics (2), geography (2), and philosophy (2). These authors are designated by their initials. This sample of personalities was selected by French experts in the field. Its size was deliberately limited so that the publications identified by the various tools used in this study could be verified manually. Finally, the aim was to characterize the maximum visibility of researchers and their publications in the humanities.

ISI Web of Knowledge and Web of Science. ISI WoK is an online academic database provided by Thomson Scientific. It provides access to many databases and other resources, such as WoS, which includes the Science Citation Index, Social Sciences Citation Index,

[4] http://scholar.google.fr/
[5] http://www.harzing.com/resources.htm
[6] http://cleanpop.ifris.org/

and Arts & Humanities Citation Index, which cover about 8,700 leading journals in science, technology, social sciences, arts, and humanities. The use of ISI WoK is licensed to institutions such as universities and the research departments of large corporations. I use the license granted for the University of Marne-la-Vallée (France).

Google Scholar. In November 2004, Google released the beta version of Google Scholar. GS is based on software that identifies and selects scientific papers from the Web by identifying common formats and then extracts the title, author(s), abstract, and references. GS searches "research publications such as journal articles, books, preprints and technical reports, putting the most pertinent articles at the top of its searches" [5]. Some researchers consider GS to be of comparable quality and utility to commercial databases [2], even though its user-interface is still in its beta version.

"Publish or Perish". "Publish or Perish" is a software program that retrieves and analyzes academic citations [8]. It was developed by Prof. Anne-Wil Harzing[7] of the University of Melbourne (Canada). It uses GS to obtain the raw citations, then analyzes these and presents a wide range of citation metrics in a user-friendly format. The principal utility of PoP is to list the results of GS and export them. However, this tool has two limitations. First, it does not allow for the merging of citations when an article appears several times in the data of Google Scholar, which is very frequent. Second, you can search for Audrey Baneyx's (search for "A Baneyx") papers and obtain, for instance, a result for François Baneyx because he will have published with A Bianchi. Likewise, you may also obtain a result for Alexandra Baneyx because the search "A Baneyx" is too large, but the search "Audrey Baneyx" is too limited. The use of PoP requires one to clean the list of publications obtained and calculate the indices for similar data again.

"CleanPoP". To exploit the results supplied by GS via PoP successfully, I developed CleanPoP, which is a web interface which cleans the Publish or Perish outputs. In particular, it allows the user to automatically detect and merge all entries for the same paper and then to work out the metrics. Moreover, the user can choose all the names and surnames that the system has automatically detected syntactically close to the author's name. CleanPoP's list of papers and metrics is therefore more accurate than that of PoP.

Methods

This section provides information to permit repetition of the experiments. The way to calculate metrics (number of papers, number of citations, number of citations by paper, and the h-index) is the same in WoS, PoP, and CleanPoP. The method used in September 2008 is described below in three steps.

Step 1: Publications search in ISI WoS. I used WoS as a search interface in WoK and searched for each author's publications in all the WoS databases. For each researcher I entered the family name and first-name initial in the "author" field. No time period was specified. The "refine result" interface enabled me to verify whether the publication domains ("subject areas") were indeed those associated with the researcher and to ensure that there were no obvious errors. The metrics were those produced by the "citation report" (cf. Image 1). Finally, I recorded and saved all the data.

Step 2: Publication search in GS via PoP. I used the PoP 1.9 version developed for Linux. Since GS limits its answers to the 1000 most-cited articles by an author, I chose to restrict the search. For each author, the fields "Medicine, Pharmacology, and Veterinary Science" and "Biology, Life Sciences, and Environmental Science" were excluded from the search. For each researcher whose name was entered, the surname was followed by the first-name initial in the "author" field, and no time period was specified (cf. Image 2). As no other screening was carried out, the results were to some extent biased by the presence of duplicates and authors with similar names. I exported the results in the CSV format to be able to import them into CleanPoP.

Step 3: Transferring the results from PoP into CleanPoP. For each researcher, I imported into CleanPoP the CSV file containing the results from PoP and thus from GS. In the first phase I chose the authors of interest from the names identified. In the second phase, duplicate articles were semi-automatically selected (cf. Image 3). Duplicates were identified by means of an algorithm that calculates similitude, taking into account titles, dates, and the network of coauthors for each publication. I then validated the duplicates which had been found to have a rate of similitude of between 40 and 50%. The system automatically validates similitude of between 50 and 100%. The metrics were then calculated and all the results recorded.

RESULTS

CleanPoP's utility

In Table 1 we see that there is a relatively large difference between the h-index calculated with PoP and with CleanPoP. This is explained firstly by the fact that CleanPoP seems to detect and to merge correctly similar articles. Moreover, articles with a strong rate of citations in GS are the ones which were well identified and parsed by the GS crawlers.

Concerning the comparison of the number of articles between CleanPoP and PoP, a relative regularity is observed: on average, CleanPoP halves the number of articles from PoP. The interesting fact is that, concern-

[7] http://www.harzing.com/index.htm

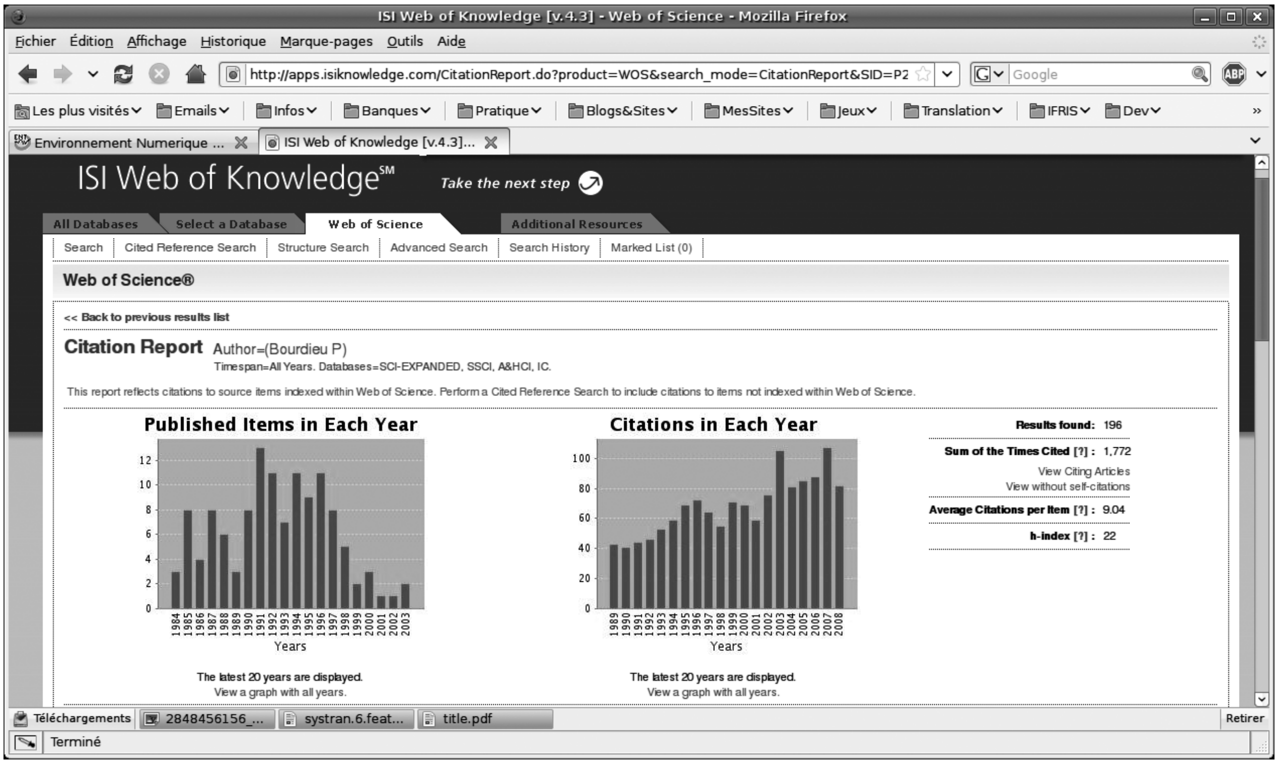

Image 1. Copy of the Web of Science’s Citation report for P. Bourdieu.

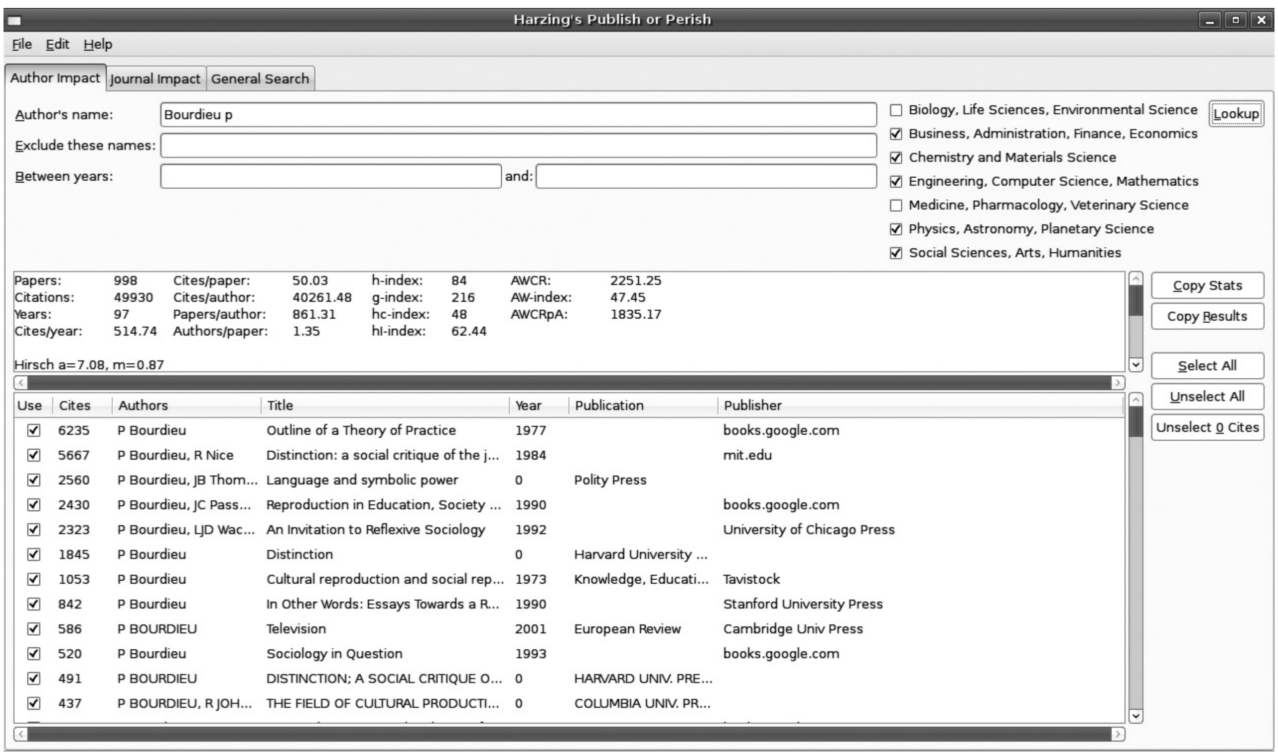

Image 2. Copy of the screen displaying results obtained for “P Bourdieu” by the tool PoP.

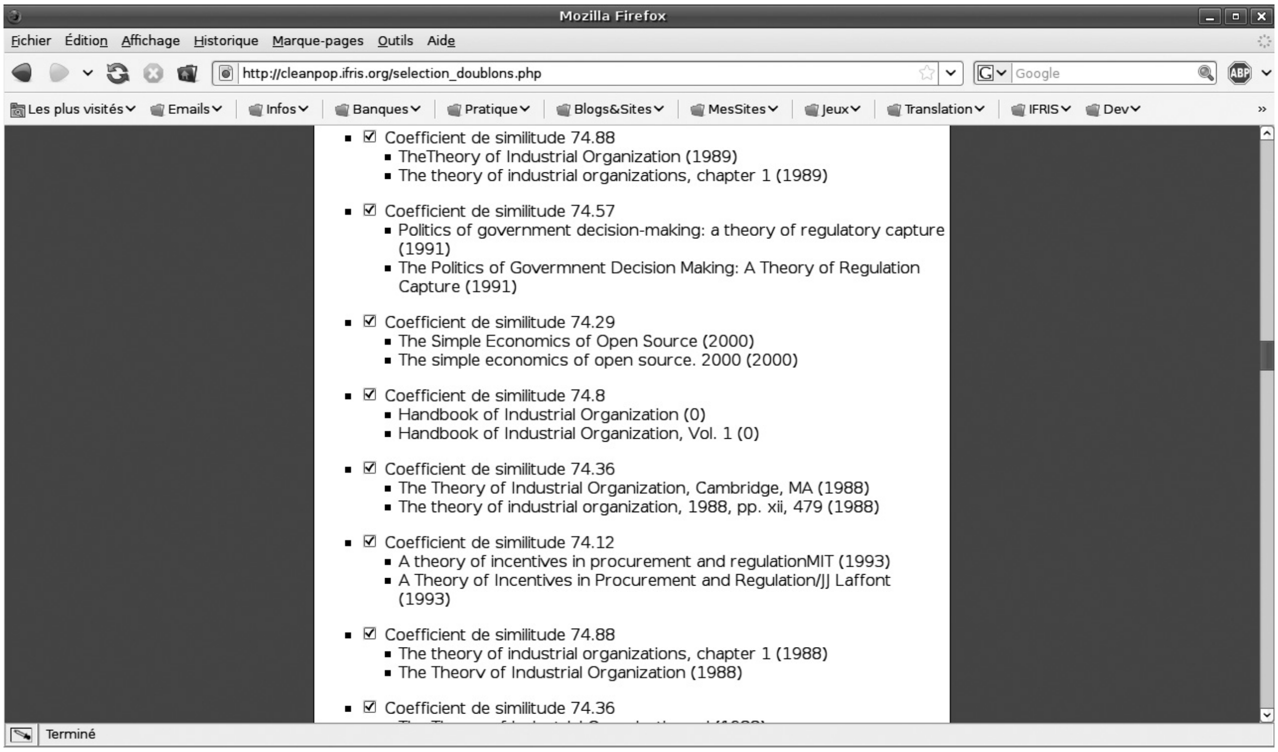

Image 3. Copy of the screen displaying the interface on which duplicates could be eliminated in CleanPoP.

ing the number of citations, it maintains around 80% of the PoP results. This implies that either the deleted articles are the less cited or most of the articles are not deleted but recognized as duplicates and merged. Note that merging two articles sums the citations in CleanPoP, so that the overall sum of citations is maintained. Regarding the standard error, the relatively weak values (17.17% for articles and 24.11% for citations) show that these averages seem more or less predictable for all the authors except those who undergo the strongest corrections, such as R. Br (see Table 1). Apart from these authors, both number of articles and number of citations are quite predictable. These extreme behaviors can be related to pseudo-homonymy from PoP regarding the weak first-name determination (only the initials) and the lack of association between first and family names. Concerning the R. Br citations, the importance of the phenomenon of pseudo--homonymy and its influence are apparent. This author has a common family name. If we continue to study this author we see that PoP finds 3707 citations, but CleanPoP keeps only 526, which represents around 14%. This means that around 86% of citations found by PoP are not correct. It also has some influence on the h--index of this author, who moves down from 25 with PoP to 9 with CleanPoP.

For citations, the results provided by PoP for J. H, P. B, and F. B are fairly good because CleanPoP's corrections are weak. However, caution is advised here because GS and, consequently, PoP easily attribute some publications to the search authors which are not theirs. It is therefore essential to carry out a check, either manually (but this can take a long time) or automatically, with CleanPoP.

Study of the scientific production of a sample of high-level scientists

Table 2 is a synthesis of the metrics which we obtained by questioning ISI WoS and GS via PoP and CleanPoP (in the table the column is GS/CP) in September 2008. Table 3 presents the ratio in terms of the number of papers and numbers of citations between these results. Tables 4, 5, and 6 are classifications of a subset of researchers in sociology and show the evolution of each one's rank according to the results obtained with the different tools.

The differences between the results of WoS and CleanPoP are very obvious (cf. Table 2).

Visibly, the major disadvantage of the WoS resides in underestimating scientific production and citation impact. This is true for all of the researchers that I studied, for both the number of papers and the number of citations, but with differences between disciplines (see the h-index for economy in Table 2). It is now well known that GS retrieves more documents than the WoS. On the other hand, the major disadvantage of GS resides in the fact that there is no distinction between a paper in a well-known journal, a book, a scientific report, or a paper published in proceedings. Concerning the numbers of articles, number of citations, or the h--index, the contribution of the GS database is clearly vis-

Table 1. Comparison between PoP and CleanPoP

DOMAINS	AUTHORS	NB of ARTICLES			NB of CITATIONS			H-INDEX	
		PoP	*CleanPoP*	*CleanPoP/PoP*	*PoP*	*CleanPoP*	*CleanPoP/PoP*	*PoP*	*CleanPoP*
Economy	J-J. L	823	498	60,5%	15683	14460	92,2%	55	49
Economy	J. T	1000	462	46,2%	36204	29315	81,0%	87	68
Geography	R. Br	811	103	**12,7%**	3707	526	**14,2%**	25	9
Geography	R. Bo	142	98	69,0%	1162	776	66,8%	15	12
History	P. R	96	72	**75,0%**	646	619	95,8%	11	10
History	F. B	427	196	45,9%	3785	3680	97,2%	40	25
Philosophy	R. G	548	202	36,9%	2977	2316	77,8%	19	13
Sociology and Philosophy	J. H	998	607	60,8%	34806	34664	**99,6%**	69	61
Sociology	P. B	998	604	60,5%	49519	49229	99,4%	83	81
Sociology	N. L	997	320	32,1%	10357	6697	64,7%	40	33
Sociology	B. L	930	462	49,7%	20094	18661	92,9%	48	43
Sociology	W. B	381	185	48,6%	4656	3475	74,6%	19	16
	Average			49,82%			79,68%		
	Standard error			17,17%			24,11%		
	Max			75,00%			99,59%		
	Min			12,70%			14,19%		

Table 2. Synthesis of metrics resulting from the study undertaken in September 2008

DOMAINS	AUTHORS	NB of ARTICLES		NB of CITATIONS		H-INDEX	
		ISI WoS	*GS/CP*	*ISI WoS*	*GS/CP*	*ISI WoS*	*GS/CP*
Economy	J-J. L	169	498	3137	14460	31	49
Economy	J. T	132	462	5920	29315	49	68
Geography	R. Br	17	103	189	526	7	9
Geography	R. Bo	5	98	8	776	2	12
History	P. R	25	72	8	619	1	10
History	F. B	22	196	16	3680	2	25
Philosophy	R. G [1]	96	202	304	2316	10	13
Sociology and Philosophy	J. H	133	607	687	34664	11	61
Sociology	P. B	196	604	1770	49229	22	81
Sociology	N. L	96	320	597	6697	13	33
Sociology	B. L	95	462	532	18661	15	43
Sociology	W. B	26	185	240	3475	7	16

[1] R. G has a common family name so I only made the search with PoP with the « Business, Administration, Finance, and Economics » and « Social Sciences, Arts, and Humanities » categories checked.

ible: the numbers for J. H speak for themselves. However, one has to bear in mind the current lack of visibility in the resources parsed by GS, which has published no official list.

Table 3 presents the ratio between the results from WoS and from GS via PoP and CleanPoP (GS/CleanPoP) considering the number of articles and the number of citations.

In terms of number of articles, the results of ISI WoS represent between 5 and 47.5% of the volume of GS/CleanPoP's results. For this sample of authors, there are, on average, four times more articles in GS/CleanPoP than in ISI WoS. The same applies when it comes to the number of citations, which varies from less than 1 to 36%. The ratio of F. B's citations between ISI WoS and CleanPoP shows that citations are not found in the journals indexed by ISI WoS databases. GS/CleanPoP finds 232 times more citations than ISI WoS does. There are probably very few documents in common in ISI WoS and GS for this author.

Is the ranking of researchers maintained with the three tools?

In Tables 4, 5, and 6 we want to ascertain whether ISI WoS, GS/PoP, and CleanPoP rank a specific group of researchers in a similar way; in other words, whether we can observe a correlation between databases. We do not want to compare the researchers' positions in relation to one another. In the three tables we observe the same ranking between the classification from GS/PoP and CleanPoP. There is a relatively constant evolution of the results when they pass from PoP to CleanPoP. There is no strong correlation between the ranking from ISI WoS and GS/PoP, but the differences are not large, either.

Table 3. Ratio between the different results obtained depending on the tools used

RATIO BETWEEN RESULTS OF METRICS			
		ISI WoS divided by GS/CleanPoP	
DOMAINS	**AUTHORS**	**ARTICLES**	**CITATIONS**
Economy	J-J. L	33,94%	21,69%
Economy	J. T	28,57%	20,19%
Geography	R. Br	16,50%	35,93%
Geography	R. Bo	5,10%	1,03%
History	P. R	34,72%	1,29%
History	F. B	11,22%	0,43%
Philosophy	R. G	47,52%	13,13%
Sociology and Philosophy	J. H	21,91%	1,98%
Sociology	P. B	32,45%	3,60%
Sociology	N. L	30,00%	8,91%
Sociology	B. L	20,56%	2,85%
Sociology	W. B	14,05%	6,91%
Average		*24,71%*	*9,83%*
Standard error		*11,99%*	*11,03%*
Max		*47,52%*	*35,93%*
Min		*5,10%*	*0,43%*

Table 4. Classification of the subset of sociologists by number of citations in GS/PoP

		Number of citations		
		ISI WoS	*GS/PoP*	*CleanPoP*
Sociology	P. B	1770	49519	49229
Sociology and Philosophy	J. H	687	34806	34664
Sociology	B. L	532	20094	18661
Sociology	N. L	**597**	10357	6697
Sociology	W. B	240	4656	3475

Table 5. Classification of the subset of sociologists by number of articles in GS/PoP

		Number of articles		
		ISI WoS	*GS/PoP*	*CleanPoP*
Sociology and Philosophy	J. H	133	998	607
Sociology	P. B	**196**	998	604
Sociology	N. L	96	997	320
Sociology	B. L	95	930	462
Sociology	W. B	26	381	185

Table 6. Classification of the subset of sociologists by h-index in GS/PoP

		H-index		
		ISI WoS	*GS/PoP*	*CleanPoP*
Sociology	P. B	22	83	81
Sociology and Philosophy	J. H	11	69	61
Sociology	B. L	**15**	48	43
Sociology	N. L	13	40	33
Sociology	W. B	7	19	16

DISCUSSION

For better or for worse, citation analysis now has important implications. Citation is considered in grants, hiring, and tenure decisions by laboratories, institutions, and governments. For many reasons, researchers may want to demonstrate the impact of their work, and citation analysis is one, albeit controversial, way of doing so. For many years, WoS was the most influential provider of citation tracking [18]. In 2004, the competitor Google Scholar emerged and proposed a broader coverage of publications that seems more interesting for the humanities [1]. However, a significant problem with GS is the secrecy concerning its coverage. Some publishers do not allow it to crawl their journals. Elsevier journals, for example, were not included until mid-2007, when Elsevier began to make most of its ScienceDirect content available to Google Scholar and Google's web search [4]. GS refuses to publish a list of scientific journals crawled and there is no information about its actual size and coverage [10–13]. The frequency of its updates is officially unknown, but Anne-Wil Harzing ran some experiments and suggests that it is updated at least every two to three months [8]. It is therefore impossible to know how current and/or exhaustive searches actually are in GS. For instance, Jacsó claims that GS provides more citations for certain academic domains, but some large-scale studies report the opposite in the social sciences and humanities [3, 14]. GS does, nevertheless, allow for easy access to published articles without the difficulties that are encountered in some of the most expensive commercial databases.

Interestingly, at least two studies report that GS provides unique citing material (which is not in the ISI database), although the exact composition of this citing material should be examined more carefully to establish clearly what is and is not included in GS searches [1]. Kousha and Thelwall [14] took a sample of 882 articles from 39 open-access ISI-indexed journals in 2001 in biology, chemistry, physics, and computing and found 70% of GS unique citations[8] in GS. Apparently, these citations were from full-text sources and there were large disciplinary differences between the types of citing documents, suggesting that a wide range of non-ISI citing sources is accessible by GS. Thus the use of GS/PoP restores some equity in the evaluation of an author whose editorial practice deviates from frequent modes of evaluation (e.g. publications in languages other than English or in books).

In spite of numerous and diverse errors in bibliographical databases, it seems certain that if the work of a researcher in the humanities is well cited, it is because he or she is influential in his or her research domain. However, if a researcher has weak metrics, it could be because he or she works in a narrow field, publishes little in English, or writes mainly books. Citation metrics in the humanities are greatly underestimated as

[8] K. Kousha and M. Thelwall define "Google Scholar unique citations" as those retrieved by Google Scholar which are not in the ISI database.

researchers in these domains tend to publish in their own language, and books are more important than in the "hard" sciences. Results show that bibliographical metrics are highly dependent on the quality and completeness of the bibliographical databases.

Meho and Yang [16] conducted a large-scale comparison between WoS, Scopus (Elsevier's), and GS. This study covered citations of over 1000 academic works published between 1996 and 2005 by all 15 faculty members of the School of Library and Information Science at Indiana University of Bloomington. They found that the overlap in citations between the three databases is relatively small, which is a very important and interesting conclusion. To sum up, the overlap between WoS and Scopus was 58.2% and only 30.8% between GS and the union of WoS and Scopus. This small overlap is due to the fact that GS produced more than twice as many citations as WoS and nearly twice as many citations as Scopus. Many of those additional citations came from conference papers, scientific reports, master's theses, books, and book chapters. The conclusion of Meho and Yang's study [16] is that GS can help to identify a significant number of unique citations. These citations might not significantly alter a scientist's citation ranking in comparison to others in the same domain and they might not all be of the same origin as those found in the WoS or Scopus [8]. However, the results provided by GS can be very useful in demonstrating evidence of a broader intellectual and international impact than is possible with WoS.

I am well aware that my conclusion may need to be revised following the development of my experiments. All results and conclusions in this study are preliminary, based on sample tests, and were valid in September 2008. Because of the broader range in data sources, we have seen that the use of Google Scholar results in more comprehensive citation coverage in the humanities. In particular, it benefits academics publishing in sources that are not well covered in ISI, such as books, conference papers, non-USA journals, etc. To calculate metrics in the humanities, the most interesting way seems to be to search Google Scholar with Publish or Perish and then to clean the results using CleanPoP, while keeping in mind that a part of these results represents "extensive" scientific production and impact [15].

PERSPECTIVES

This preliminary study which I have just presented is going to lead to two developments during the coming months. First I would like to find a new metric to report faithfully on total scientific production by considering both scientific articles and books (based on the results supplied by both WoS and GS), in English or in other languages. This raises, moreover, the problem of the status of book translations: is it a good solution to count them as an entirely new scientific production or a new way to expand the visibility of scientific production? If not, what place should we make for them? What solution can we come up with to sort the GS data and clearly identify between books, conference proceedings, scientific reports, etc. Concerning books, a solution may lie in the identification of the large scientific publishers by recovering and processing their data and then searching in GS for citations corresponding to these works.

Second, several French laboratories working in the social sciences wish to report their scientific production to their institution of evaluation in a collective way. How can scientific production be characterized at the collective level? It seems evident that to calculate a simple average or a median of h-indexes is not satisfactory. What indicators can we use? Should we create new ones? Finally, because institutions wish to evaluate their researchers, notably by using the h-index, we would like to be able to define and characterize the position of a researcher in the social sciences and other domains.

REFERENCES

1. Bakkalbasi N. and Bauer K. and Glover J. and Wang L. (2006): Three options for citation tracking: Google Scholar, Scopus and Web of Science. Biomedical Digital Libraries, **3**, no. 7 (available at http://www.bio-diglib.com/ /content/3/1/7).
2. Bauer K. and Bakkalbasi N. (2005): An Examination of Citation Counts in a New Scholarly Communication Environment. D-Lib Magazine, **11**, no. 9 (available at http://dlib.org/dlib/september05/bauer/09bauer.html).
3. Bosman J., Mourik I., van Rasch M., Sieverts E. and Verhoeff H. (2006): Scopus reviewed and compared. The coverage and functionality of the citation database Scopus, including comparisons with Web of Science and Google Scholar. Utrecht University Library (available at http:// igitur-archive.library.uu.nl/DARLIN/2006-1220-200432/ /Scopus doorgelicht & vergeleken – translated.pdf).
4. Brantley P. (2007): Science Directly into Google. O'Reilly Radar (available at http://radar.oreilly.com/archives/ /2007/07/science-directl.html).
5. Butler D. (2004): Science searches shift up a gear as Google starts Scholar engine. Nature, **432**, 423.
6. Garfield E. (1955): Citation indexes for science – new dimension in documentation through association of ideas. Science, **122**, 108–111.
7. Garfield E. (1979): Citation Indexing – Its Theory and Application in Science, Technology, and Humanities. John Wiley & Sons, New York.
8. Harzing A.-W. and van der Wal R. (2008): Google Scholar as a new source for citation analysis? Ethics Sci. Environmental Politics, **8**, 62–71 (available at http://www. harzing.com/pop_gs.htm).
9. Hirsch J. E. (2005): An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. USA, **102**, 16569–16572 (available at http://arxiv.org/ /abs/physics/0508025 v5 29 Sep 2005).
10. Jacsó P. (2004): Google Scholar Beta. Thomson Gale (available at http://www.galegroup.com/servlet/HTMLFile Servlet?imprint=9999&region=7&fileName=/reference/archive/200412/googlescholar.html)
11. Jacsó P. (2005): As we may search – comparison of major features of the Web of Science, Scopus, and Google

Scholar citation-based and citation-enhanced databases. Curr. Sci., **89**, 1537–1547.
12. Jacsó P. (2005): Google Scholar Beta (Redux). Thomson Gale (available at http://www.gale.com/servlet/HTMLFileServlet?imprint=9999&region=7&fileName=/reference/archive/200506/google.html).
13. Jacsó P. (2005): Google Scholar: the pros and the cons. Online Information Rev., **29**, 208–214.
14. Kousha K. and Thelwall M. (2008): Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. Scientometrics, **74**, 273–294.
15. Mayr P. and Walter A.-K. (2007): An exploratory study of Google Scholar (available at http://arxiv.org/abs/0707.3575).
16. Meho L. I. and Yang K. (2007): A new era in citation and bibliometric analyses: Web of Science, Scopus, and Google Scholar. J. Am. Soc. Information Sci. Technol., **58**, 2105–2125.
17. Pauly D. and Stergiou K. I. (2005): Equivalence of results from two citation analyses: Thomson ISI's Citation Index and Google Scholar's service. ESEP, 33–35.
18. Roth D. L. (2005): The emergence of competitors to the Science Citation Index and the Web of Science. Curr. Sci., **89**, 1531–1536 (available at http://www.ias.ac.in/currsci//nov102005/1531.pdf).