The indexation of scientific journals and the bibliometry: examples with current tools

Research Note – IRD/HSM Montpellier, France – April 2017

Gil Mahé

Update of the previous paper Mahe, G. (2012). L'indexation des journaux scientifiques et la bibliométrie, bref aperçu. **Le Journal de l'Eau et de l'Environnement**, 20, 52-58.

Today there are 4 tools available on the internet to get bibliometric values for individuals, journals and institutions: Web of Knowledge-WOK (Thompson Reuters) and Scopus (Elsevier), the two oldest and most used ones, Google Scholar (GS)(best searched with Harzings-Publish or Perish), and the recent one ResearchGate (RG).

All of them can give the Impact Factors –IF of many journals, and the *h* individual index of many researchers. But each of them has limitations either of representativity or of access, which need a few explanations to ensure the better use of them, depending on the context one want to use the results.

Here are some examples to compare the searches with these different tools.

Example 1: search for a journal

You want to know if a journal has an Impact Factor, if it is indexed or not, how, where, which quartile, etc…?

Several sites allow to scan these statistics for the journals.

Search for Journal of Hydrology.

- http://www.scijournal.org/index.html is a site to search journals indexed in WOK. Journal of Hydrology is ranked into geosciences. The result is as follows:

2015/2016 Impact Factor : 3.043
2014 Impact Factor : 3.053
2013 Impact Factor : 2.693
2012 Impact Factor : 2.964
2011 Impact Factor : 2.656
2010 Impact Factor : 2.514
2009 Impact Factor : 2.433
2008 Impact Factor : 2.305

- http://www.scimagojr.com/ is a site to search journals indexed in SCOPUS. The result is as follows:

  This indicator counts the number of citations received by documents from a journal and divides them by the total number of documents published in that journal. The results shows the evolution of the average number of times documents published in a journal in the past two, three and four years have been cited in the current year. The two years line is equivalent to journal impact factor ™ (Thomson Reuters/WOK) metric.

Cites / Doc. (2 years)	2015 3.174	Cites / Doc. (2 years)	2011 3.100
Cites / Doc. (2 years)	2014 3.443	Cites / Doc. (2 years)	2010 3.000
Cites / Doc. (2 years)	2013 3.040	Cites / Doc. (2 years)	2009 2.844
Cites / Doc. (2 years)	2012 3.342	Cites / Doc. (2 years)	2008 2.487

- https://www.researchgate.net/journal/0022-1694_Journal_of_Hydrology, here is the site where one is redirected if one type this research "journal of hydrology ResearchGate". This is how to search metrics of journal indexed by ResearchGate. The values are calculated using ResearchGate data and is based on average citation counts from work published in this journal. The data used in the calculation may not be exhaustive.

2015 / 2016 RG Journal impact	3.85	2011 RG Journal impact	3.98
2014 RG Journal impact	3.99	2010 RG Journal impact	3.59
2013 RG Journal impact	3.62	2009 RG Journal impact	3.30
2012 RG Journal impact	3.77	2008 RG Journal impact	3.01

- https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=bio_hydrology here the site where one is redirected if one type this research "Google Scholar journal of hydrology". This is how to search metrics of journal indexed by GoogleScholar. There is no direct access to an Impact Factor on this site, but the *h* index over 5 years is used to rank the journals within the topic "*Hydrology and water resources*". *h5-index* is the *h-index* for articles published in the last 5 complete years. It is the largest number *h* such that *h* articles published in 2011-2015 have at least h citations each. It can not be compared to the impact factor, but the ranking is based on the same type of citations. The *h* index is more frequently used to assess the individual publications production and audience. The result is as follows for the 20 first journals the most frequently cited within the topic "*Hydrology and water resources*":

Publication	h5-index
1. Water Resources Research	64
2. Journal of Hydrology	64
3. Hydrology and Earth System Sciences	60
4. Advances in Water Resources	48
5. Hydrological Processes	46
6. Water Resources Management	42
7. Journal of Hydrometeorology	41
8. JAWRA Journal of the American Water Resources Association	32
9. Groundwater	30
10. Hydrogeology Journal	30
11. Vadose Zone Journal	30

12.	Ecohydrology	29
13.	Journal of Hydrologic Engineering	29
14.	Journal of Water Resources Planning and Management	28
15.	Hydrological Sciences Journal	26
16.	Journal of Hydroinformatics	21
17.	Journal of Irrigation and Drainage Engineering	21
18.	Hydrology Research	20
19.	Journal of Hydro-environment Research	18
20.	Proceedings of the Institution of Civil Engineers-Water Management	17

Synthesis

The results show some significant variability between the impact factors of the Journal of Hydrology according to the 3 indexation sites, but the interannual evolution is quite similar (Fig.1)

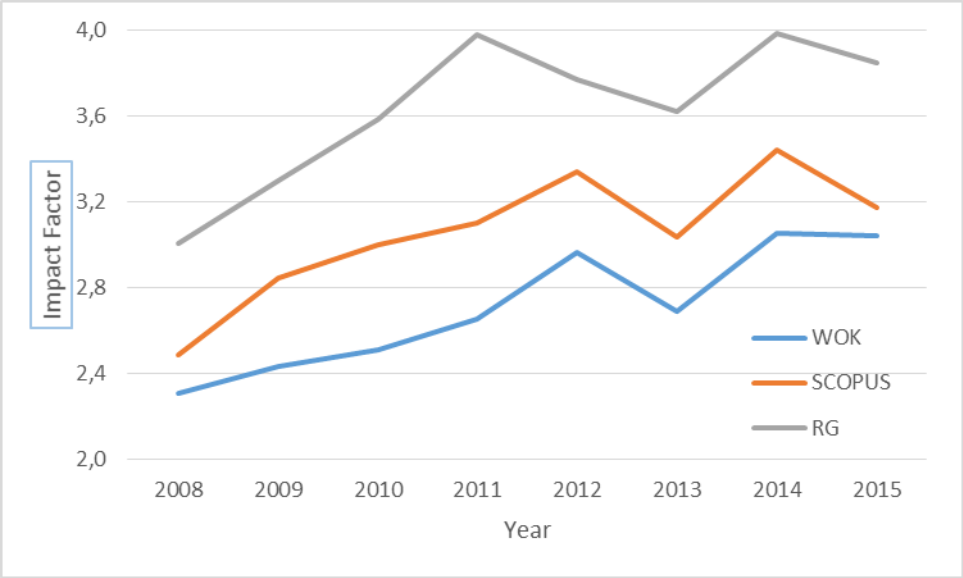

Figure 1: Impact Factor of Journal of Hydrology according to three different indexation sites.

The variability between the 3 sites is due to that all 3 indexation sites do not index exactly the same corpus of journals. From this it follows that the cited references are also different. As the Impact Factor is calculated from the number of citations in the indexed published papers, the IF is thus different for each site. Moreover, as there are less journals indexed in the WOK than in SCOPUS and than in RG, the total number of citations available to calculate the IF is also lower in WOK than in SCOPUS and higher in RG.

The quartiles

In several indexation sites the journals are classified according to their main topic. Within each topic, the journals are ranked according to their Impact Factor. It is often that the ranking is divided into 4 quartiles, separating the most often cited journals in the first quartile, and the least cited in the 4th quartile.
This ranking is revised annualy, and each journal can be classified into several topics.
The figure 2 presents the ranking and quartiles for “Ecohydrology”, a journal of Wiley in UK, according to SCOPUS. The journal was ranked in the second quartile at the beginning of the record in all 4 topics within which it was classified, and changed to the 1st quartile during different years since, depending on the topic concerned.

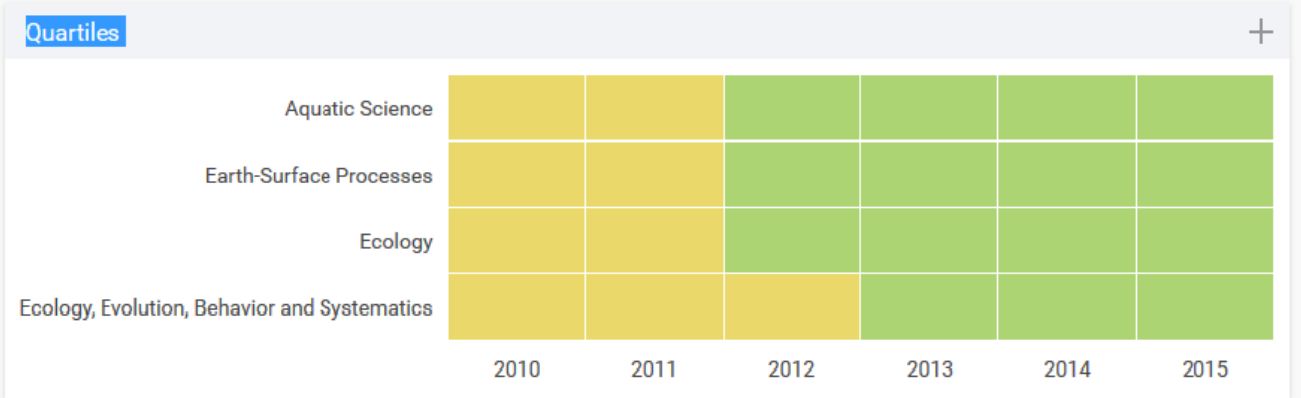

Figure 2: Quartiles ranking for Ecohydrology according to SCOPUS

Conclusion

All publications are not indexed, but many of them. To be indexed the journal must request an application, which is diffuclt to get for WOK and SCOPUS, but much easier for Google Scholar, which means that there are much more journals indexed by Google Scholar than in WOK and SCOPUS. Once a journal is indexed it is possible to process its Impact Factor, thus even **if a journal is** not **indexed** in WOK or SCOPUS, but in Google Scholar**, it means that it has an Impact Factor**. Reasearchgate can give the Impact Factors of a great number of journals which are not indexed in WOK or SCOPUS, as it is based on the papers loaded by the researchers themselves. **There is no limit to the number or journal indexed in RG**, except that of the researchers themselves when loading or not their papers. However it is not written from which indicators RG decides to open a page for a journal with its metrics, as there are still some journals with no metrics in RG (most of them being poorly cited).

One can question about the IF if it could not be calculated over more than 2 years, as the popularity of a paper might increase after a while. It is already done for instance by SCOPUS which proposes IF over 2, 3 or 4 years. The result is that the difference of IF between journals over time does not change much, thus leading to keep the initial IF calculation over 2 years (3 years in fact, as it is the number of citations during year 3 of all papers published in years 1 and 2).

Example 2: search for an Individual

To search for the scientific production of an Individual the results will also depend on the type of index researched. WOK and SCOPUS allow this of course, but with the same limitation already mentioned, ie it is limited to academic standardized journals, “accepted” by Thompson Reuters or Elsevier. **ResearchGate** will present all the production that a researcher will have loaded itself, or its co-authors, thus there is virtually no limit and absolutely all the productions can be loaded, even databases (table 1).

- Article
- Book
- Chapter
- Code
- Conference Paper
- Cover Page
- Data
- Experiment Findings
- Method
- Negative Results
- Patent
- Poster
- Presentation
- Project
- Question
- Raw Data
- Research Proposal
- Technical Report
- Thesis
- Working Paper

Table 1: The different categories of documents downloadable on ResearchGate

Google Scholar is based on the detection by google of all the productions visible on the web, thus it contents much more results than WOK and SCOPUS. But it is limited to the findings of GS on the web, and as it was mentioned before, all journals are not indexed in GS. Also a number of productions are not easily visible on the web as posters, presentations to conferences, master’s and Phd’s thesis and other academic reports, technical reports, etc… Yet Google Scholar is a good alternative to WOK and SCOPUS as what is found on the web does not depend strictly on individuals, as papers and some other contents are regularly posted on the web, which is not the case for

ResearchGate, to which a number of researchers are not registered. In this last case it does not mean that these researchers are invisible in RG, as if one of their co-authors has posted their common work, he can be visible. The difference is that a number of personal results can not be posted by someone else and thus are not visible.

The *h* index

The *h* index is the number of *h* publications which have received at least *h* citations. It is an indicator of both the number of papers and their popularity. But is has also limitations that have been discussed abundantly, thus leading to alternate indicators, like i10 or i20, ie number of papers cited at least 10 times or 20 times for instance, which gives more depth into a researcher's production, and several others, that are calculated by Harzings for instance. But the *h* index remains the most used of them. To illustrate the extremes of using this *h* index, one can cite the 2 cases where 2 researchers A and B have both 100 citations each of them, thus the same overall popularity. The researcher A has only 1 paper, thus cited 100 times it means a very high popularity, but his *h* index is only 1, because of 1 paper cited at least 1 time. Researcher B has 10 papers cited 10 times each, which means a medium popularity for each paper, but its *h* index is 10, for 10 papers cited at least 10 times. One might also discuss about the time elapsed since the publication of the papers, which could also be taken into account to compare individuals without regarding their age. But without regarding this, with 100 citations for the same publication, it is obvious that the results of the researcher A had a great impact on the community, much larger than the researcher B if one consider 1 result. Even if it is an extreme case, in a system that tends to look mostly at numerical ranking of individuals, one can question about the chances of both researchers to rise academic steps. Another point is to identify the number of papers published as a first author, or with former students as first author, and the papers where the role has not been leading. This is of some importance to assess the leading role of one researcher in a specific field.

Is it possible today to assess people not only from their metrics? The international peer review system of papers, firstly does not prevent from many publications of low quality assessment, as with the internet it is easy to publish quickly, sometimes only provided you pay, and secondly can not be assimilated to an assessment commission in charge of evaluating the researchers. Thus if national assessment commissions give a major importance to the *h* index to evaluate the quality of a researchers' work, it means that the main thing that the commission takes into account is the work of the peer reviewers from journals approved by Thompson and Elsevier... However, we could suppose that the work a researcher is awaited to do is not mainly to publish plenty of papers abundantly cited, but also to develop innovative projects, create and develop communities and networks, teach its new results to students, explain its research to a large audience, etc… all this not being assessed at all by the peer reviewers, and thus being minored by the national commissions if mainly ranking people by the *h* index.

Knowing this, one of the impact of this is that researchers tend to develop works which can give results in a very quick time range, thus field work tend to decrease to the profit of modelling activities, as field work is time consuming. The overall impact of this is the reduction of field observations, while there is an increase of international databases, which quality of content is poorly debated …

Example of search: Z. Zidane

RG

Z Zidane has not registered to RG, thus his metrics are not available in RG.

WOK

1 paper is recorded in WOK, cited 0 times, thus *h* is 0.

Adaptive Minimum Variance Control of a DC motor

By: Zidane, Z.; Lafkih, M. Ait; Ramzi, M.; et al.

Book Group Author(s): IEEE

Conference: 18th Annual International Mediterranean Conference on Control and Automation (MED) Location: Marrakech, MOROCCO Date: JUN 23-25, 2010

SCOPUS

3 papers for Z. Zidane, one of them being cited once, thus *h* is 1. One can see that the paper cite once is not indexed in WOK, while that cited also in WOK is also cited 0 times.

1 Adaptive Minimum Variance Control of separately excited DC m... 2011 0 0
2 Adaptive generalized predictive control of a heat exchanger ... 2011 1 0 **1**
3 Adaptive minimum variance control of a DC motor 2010 0 0

Harzings

Hereafter (table 2) are the raw results after querying Z Zidane in Harzings, which searches within the Google Scholar bibliometric database. There are much more results than with WOK or SCOPUS, and visibly there are several different Z Zidane recorded. In general one have to separate within the results those belonging to different homonymous. This is quite easy as people rarely work on the same topic, and often co-publish with the same co-authors with different names. One can see quite quick that there are 3 Z Zidane:

1. Z Zidane from Mexico working on Sociology, with 5 occurrences, cited 10 times, and *h*=2
2. Z Zidane the football player, which 8 occurrences, but two of them correspond to the same book, thus 7 different occurrences, cited 13 times, and *h*=2.
3. Z Zidane working on motors with 10 occurrences, but two are the same publications, thus 9 occurrences, all co-published with the same two co-authors, except the Master thesis in 2008. 15 citations overall, and *h*=2.

Thus the same Z Zidane working on motors has either 1, 3 or 9 publications according to the different indexation sites, the number of citations is 0, 1 or 15, and *h* index is 0, 1 or 2, according to the sites.

The metrics with ResearchGate are often greater than that of SCOPUS, as the researchers can register plenty of their productions that are not strictly papers as in SCOPUS or WOK, but lower than that of Harzings, as Harzings registers everything visible on the web, and not only that being registered by the researchers themselves like in RG. This point is important. One can see that WOK or SCOPUS are very limitative in the kind of production that they register: it is only peer reviewed indexed journals approved by Thompson-Reuters or Elsevier.
All other non indexed products are they of non scientific interest? In the case of GS/Harzings, one can see that other products are cited, and sometimes most cited than indexed papers. Not taking into account these productions and citations during the researcher's assessment does it mean that these works are of non interest –so can we stop to do them? Does it also mean that it is unuseful to cite them when writing a paper?

5	Z Zidane, MA Lafkih, M Ramzi	Adaptive Generalized Predictive Control of a heat exchanger pilot plant	2011				
	Multimedia Computing and …	ieeexplore.ieee.org	http://ieeexplore.ieee.org/abstract/document/5945715/				
	https://scholar.google.com/scholar?cites=9033667624641275 34&as_sdt=2005&sciodt=0,5&hl=en&num=20				1	2017-04-08	5
3	Z Zidane, MA Lafkih, M Ramzi	Simulation studies of adaptive predictive control for small hydro power plant	2012				
	Journal of Mechanical …	article.sapub.org	http://article.sapub.org/10.5923.j.jmea.20120206.07.html				
	https://scholar.google.com/scholar?cites=5402215144615420703&as_sdt=2005&sciodt=0,5&hl=en&num=20				2	2017-04-08	3
2	Z Zidane, MA Lafkih, M Ramzi…	Adaptive minimum variance control of a DC motor	2010	Control & Automation (…			
	ieeexplore.ieee.org	http://ieeexplore.ieee.org/abstract/document/5547628/					
	https://scholar.google.com/scholar?cites=3175407702480710957&as_sdt=2005&sciodt=0,5&hl=en&num=20				3	2017-04-08	2
2	Z Zidane	Interview with Zinédine Zidane	2006	Canal Plus			
	https://scholar.google.com/scholar?cites=8051290210088500621&as_sdt=2005&sciodt=0,5&hl=en&num=20				4	2017-04-08	2
2	Z Zidane, MA Lafkih, M Ramzi	Adaptive Generalized Predictive Control of a Heat Exchanger Pilot Plant	2012				
	Journal of Mechanical …	article.sapub.org	http://article.sapub.org/10.5923.j.jmea.20120205.03.html				
	https://scholar.google.com/scholar?cites=16144556477412783021&as_sdt=2005&sciodt=0,5&hl=en&num=20				5	2017-04-08	2
1	Z Zidane, MA Lafkih, M Ramzi	Application of Multivariable Predictive Control in a Hydropower Plant	2013				
	Journal of Automation and …	pubs.sciepub.com	http://pubs.sciepub.com/automation/1/1/5/				
	https://scholar.google.com/scholar?cites=6698126294265938835&as_sdt=2005&sciodt=0,5&hl=en&num=20				6	2017-04-08	1
4	Z Zidane, D Franck	Zidane, le roman d'une victoire	1999	Paris, Robert Laffont			
	https://scholar.google.com/scholar?cites=5563976042368293085&as_sdt=2005&sciodt=0,5&hl=en&num=20				7	2017-04-08	4
4	Z Zidane	Modernidad y posmodernidad: la crisis de los paradigmas y valores	2000	Noriega			
	https://scholar.google.com/scholar?cites=11770624212529976274&as_sdt=2005&sciodt=0,5&hl=en&num=20				8	2017-04-08	4
3	Z Zidane, D Franck	Der mit dem Ball tanzt	2005	Bombus			
	https://scholar.google.com/scholar?cites=17127075911675876614&as_sdt=2005&sciodt=0,5&hl=en&num=20				9	2017-04-08	3
0	Z Zidane, MA Lafkih, M Ramzi	Design of Linear Quadratic Gaussian Controller for Sample Power System	0	pdfs.semanticscholar.org			
	https://pdfs.semanticscholar.org/d42f/79e04c79ec6a0624381d07cf0628c668a05b.pdf				10	2017-04-08	0

0	Z Zidane, MA Lafkih, M Ramzi	Application of Multivariable Linear Quadratic Gaussian Control and Generalized Predictive Control in a Hydropower Plant	0	sta-tn.com	http://www.sta-tn.com/IJ_STA/Papers/volume_7N1_april_2013/P7_IJSTA_V7N1_13.pdf	11	2017-04-08	0
0	Z Zidane, MA Lafkih, M Ramzi	Adaptive Generalized Predictive Control of a Separately Excited DC Motor	0	sta-tn.com	http://www.sta-tn.com/IJ_STA/Papers/volume_5N1_jun_2011/P4_IJSTA_V5N1_11.pdf	12	2017-04-08	0
0	Z Zidane, MA Lafkih, M Ramzi	Design of Multivariable Adaptive Generalized Predictive Control for the Part Turbine/Generator of Micro-Hydro Power Plant	2016	International Journal of Advanced …	earticle.net	http://www.earticle.net/Article.aspx?sn=270729	13	2017-04-08 0
2	Z Zidane, F Mesquine	Commande d'un moteur à courant continu (Mémoire de Master)	2008	Faculté des Sciences, Semlalia, Marrakech, Maroc	https://scholar.google.com/scholar?cites=2086772479193508165&as_sdt=2005&sciodt=0,5&hl=en&num=20	14	2017-04-08	2
2	Z ZIDANE	Modernidad y postmodernidad	2001	México: Noriega	https://scholar.google.com/scholar?cites=4622950343732794651&as_sdt=2005&sciodt=0,5&hl=en&num=20	15	2017-04-08	2
2	Z Zidane	Mon geste n'est pas pardonnable	0	Le Figaro	https://scholar.google.com/scholar?cites=9706063539110741964&as_sdt=2005&sciodt=0,5&hl=en&num=20	16	2017-04-08	2
2	Z ZIDANE	2001	2001	Modernidad y postmodernidad. MÃ© xico: Noriega	https://scholar.google.com/scholar?cites=7810618510920988290&as_sdt=2005&sciodt=0,5&hl=en&num=20	17	2017-04-08	2
2	Z Zidane, D Franck	Le roman d'une victoire: Le roman d'une victoire; vol. 2, Zizou, mes photos	1999	Laffont	https://scholar.google.com/scholar?cites=2685637293334952770&as_sdt=2005&sciodt=0,5&hl=en&num=20	18	2017-04-08	2
2	Z Zidane	Qué es la paradiplomacia	2008	Safe Democracy	https://scholar.google.com/scholar?cites=4351478740082827343&as_sdt=2005&sciodt=0,5&hl=en&num=20	19	2017-04-08	2
0	Z Zidane	Real Madrid Club de Fútbol	0	taringa.net	https://www.taringa.net/posts/deportes/10330632/Real-Madrid-Club-de-Futbol.html	20	2017-04-08	0
0	CMF Santana, MG Roberto, Z Zidane	Arab Immigration in Mexico in the Nineteenth and Twentieth Centuries. Assimilation and Arab Heritage	2004	JSTOR	http://www.jstor.org/stable/20140142	21	2017-04-08	0
0	I Sánchez, Z Zidane	Quiero devolver todo lo que el futbol me ha dado	2008	Corresponsables Media Responsable, SL		22	2017-04-08	0
0	…, É Carrière, Z Zidane…	Football européen. Les transferts: une réalité mal connue	2013	Jurisport: … Centre de Droit et d'Économie du …		23	2017-04-08	0

Table 2: Result of the search of "Z Zidane" with Harzings (searching google scholar base)

Conclusion

Harzings is the most comprehensive search tool for bibliography and bibliometrics in science, as it searches within the Google Scholar bibliometric database. A great majority of the scientific work is registered here, thus tracing a wide and quite complete panorama of the work realized by an individual during his or her career. It must be noted that if a production is not visible from the web, it will not be registered, and about the journals, they have to ask for their registration into Google Scholar, thus some of them are not visible in GS, except the papers that have been cited and thus are recorded, but as individual production, not as part of a journal's issue.

The second richer indexation site is now ResearchGate. But RG is based on individual registration, and thus people not registered might be low or not visible on RG, if their potential co-authors have not register their work either. From this RG is interesting when available, but its metrics are not comparable from one individual to another.

SCOPUS seems to give metrics often a bit higher than that of WOK, because there are slightly more journals registered in SCOPUS than in WOK. But the difference between both indexation sites tends to disappear since some years, and there is even a competition between them to propose more journals, more tools, etc…

Both SCOPUS and WOK metrics are used by individuals to present their work.

The best to do is to present the metrics according to these 4 indexation sites, or at least 3 of them (if RG is not available for an individual), and to describe which work is important and visible for instance in GS and not in SCOPUS nor in WOK, while more standardized publications are well covered by these last two indexes.

Disseminating his work is a duty for a researcher, and WOK and SCOPUS do not allow enough flexibility for this. GS is much more interesting, as it is the richer base, but a bit difficult to use even with Harzings due to homonymous and double occurences. ResearchGate seems to be for now the best communication tool between researchers to communicate their results and exchange results, ideas, projects, etc…

Thus it is much more than an indexation and bibliographic site, it is a tool that helps research management and prospective, and accelerates the diffusion of results to the wider audience concerned by specialized topics.

Reference

Harzing, A.W. (2007) **Publish or Perish**, available from http://www.harzing.com/pop.htm