Scientometrics (2013) 96:51–67
DOI 10.1007/s11192-013-0952-5

Can a personal website be useful as an information source to assess individual scientists? The case of European highly cited researchers

Amalia Más-Bleda · Isidro F. Aguillo

Received: 9 August 2012 / Published online: 29 January 2013

Abstract The web is not only the main scholarly communication tool but also an important source of additional information about the individual researchers, their scientific and academic activities and their formally and informally published results. The aim of this study is to investigate whether successful scientists use their personal websites to disseminate their work and career details and to know which specific contents are provided on those sites, in order to check if they could be used in research evaluation. The presence of the highly cited researchers working at European institutions were analysed, a group clearly biased towards senior male researchers working in large countries (United Kingdom and Germany). Results show that about two thirds of them have a personal website, specially the scientists from Denmark, Israel and the United Kingdom. The most frequent disciplines in those websites are economics, mathematics, computer sciences and space sciences, which probably reflect the success of open access subject repositories like RepEc, Arxiv or CiteSeerX. Other pieces of information analysed from the websites include personal and contact data, past experience and description of expertise, current activities and lists of the author's scientific papers. Indicators derived from most of these items can be used for developing a portfolio with evaluation purposes, but the overall availability of them in the population analysed is not representative enough by now for achieving that objective. Reasons for that insufficient coverage and suggestions for improvement are discussed.

Keywords Highly cited researchers · Personal website · Europe · Indicators · Assessment

A. Más-Bleda (✉) · I. F. Aguillo
The Cybermetrics Lab, Institute of Public Goods and Policies (IPP), Spanish National Research Council CSIC, C/Albasanz, 26-28, 28037 Madrid, Spain
e-mail: amalia.mas@cchs.csic.es

I. F. Aguillo
e-mail: isidro.aguillo@cchs.csic.es

Introduction

The web has become an outstanding tool for the collection and dissemination of scholarly information (Pitzek 2002; Chen et al. 2009a, b). Indeed, professors and researchers are increasingly reliant on electronic resources (Herring 2002) and it is practically mandatory for them to maintain some sort of online presence (Hyland 2011). However, the completeness of this presence, defined as the necessary and essential information for a specific academic topic on the Internet, is unsatisfactory (Chen et al. 2009b).

The web has also clearly changed the communication system of academics (Barjak 2006), and as Lawrence (2001b) underlines, the ability to locate relevant research quickly enhances communication and scientific progress. Moreover, the web is one of the popular sources for open access (OA) publishing. According to Kousha (2009), the "knowledge of characteristics shaping the OA citation networks can give a better understanding about their potential uses for open access scholarly research". Open access initiatives (OAI), defending free access to scientific production, are growing stronger. The association of All European Academies has recently described (ALLEA 2012) the optimal panorama for Open Science in the 21st century for Europe, listing the requirements to be fulfilled to achieve this goal.

Although OA academic information contributes to scientific progress and many researchers are making their work available (Swan 2007), measuring the online impact of web publishing is a highly problematic task (Thelwall and Harries 2004). Nevertheless, there are several studies that have examined the value of different web sources for scholarly impact assessment (Kousha and Thelwall 2007, 2008; Kousha, Thelwall, and Rezaie 2010; Thelwall, Kousha, Kayvan, et al. 2008). The fact that scientific impact is now related to online impact is commonly accepted by scientific journals, universities and departments of all countries; but some doubts exist about the relevance of web links for individual scientists (Barjak, Li and Thelwall 2007). Moreover, "the impact of scientists on the web may be influenced by the country in which they work", and the "national size of a discipline might imply greater impact in one country than in another" (Barjak, Li and Thelwall 2007).

For scientists, one of the ways to provide free access to their scholarly publications is the personal website. In fact, many scientists publish information about research online through personal websites and research group websites (Barjak, Li and Thelwall 2007; Dumont and Frindte 2005). Antelman (2004) showed that most freely available articles are found on personal websites, not in repositories or OA journals. It would be interesting to analyze whether this trend has continued in recent years or if this have changed. Along these lines, Chen et al. (2009a) report that OA repositories, while an important source of academic information, generally show a low proportion of citation. Barjak (2006) stresses that personal websites are a useful platform for publications already reviewed and publicized, but he expresses doubts about the effectiveness of this type of self-publishing, what type of scientist it appeals to, and how the work posted on the web might be characterized.

The personal website as a reflection of the scientist's activity

A personal website can have different social uses (Petric 2006) and the reasons behind creating it are different: a need to socialize (Papacharissi 2002b; Zinkhan et al. 1999), a desire to control or influence others (Schmitt, Dayanim, and Matthias 2008; Zinkhan et al. 1999), self-presentation (Döring 2002; Dumont and Frindte 2005; Papacharissi 2002a;

Weibel, Wissmath and Groner 2010), entertainment (Papacharissi 2002b; Weibel, Wissmath and Groner 2010) or identity expression (Döring 2002; Marcus, Machilek, and Schütz 2006; Schmitt, Dayanim, and Matthias 2008). At any rate, they generally reflect the identity of the author (Hyland 2011; Parks and Archley-Landas 2003; Thoms and Thelwall 2005; Vazire and Gosling 2004). Flanagin and Metzger (2003) advocate that personal websites be evaluated in terms of perceived credibility, depending on matters such as gender.

Besides these reasons, the scientist' personal website can be created to disseminate the different activities that they carry out, such as teaching, research, etc. Traditionally researchers disseminated their research results mainly through scientific journals and conferences; now, however, there are additional ways of divulgation such as OA journals and the self-archive of a copy of publications in OA repositories, personal websites, departmental websites, departmental digital archives or disciplinary archives (Bailey 2010). A self-archived copy of their research may be made available, either in its final form or in a previous version (post-print or pre-print). In this regard, some studies have showed that the more widely available a publication is, the more likely it is to be linked or cited (Aguillo, Ortega, and Fernández 2008; Antelman 2004; Kousha and Thelwall 2006; Kurtz 2004; Lawrence 2001a, b; Shin 2003; Swan 2007; Vaughan and Thelwall 2003), whereas other studies (Davis et al. 2008; Frandsen 2009a, b) do not find this correlation. Swan (2010) provides a compilation of these studies and other ones on the OA citation advantages.

The present study is focused on scientists' personal websites hosted on institutional domains (university, faculty, department, laboratory, hospital, etc.). There are others channels in addition to institutional personal websites for having a web presence, like a non-institutional personal website; however, it may be more difficult to find on the Web. Furthermore, the personal website's credibility is perhaps stronger when it is presented within the institutional realm to which it belongs. Other channels for having a web presence are: research group websites, institutional repositories, thematic repositories (ArXiv, SAO/NASA Astrophysics Data System, RePEc), scholarly databases (Google Scholar, Microsoft Academic Search, Mendeley, CiteULike) or social websites (Facebook, Linkedln), although they mainly provide information related to scholarly publications. An institutional personal website is perhaps a more appropriate means for providing other researcher's information that often is difficult to find, such as contact details, experience and expertise (education, past mobility, former and current positions, awards and honors), research interest, research project, the complete CV, learning materials or conference presentations.

So, a scientist's personal website can be the most complete channel for disseminating his/her work and career as well as for increasing his/her visibility. This leads us to think that one scientist's personal website, so long reflects the different activities undertaken by the researcher (teaching, research, popularization, internationalization, etc.) might be used as a complementary information source to evaluate that scientist. Thus, this study explores whether successful scientists use their personal websites to disseminate their work and career information and whether the contents provided on these sites reflect properly the different activities undertaken by those scientists, with the belief that if that happens, the personal website could be used as a complementary information source for evaluation.

Scientists could be evaluated, at least part of their activities, on the basis of their personal websites; for example, bibliometric methods could be used to evaluate the formal publications, webometric methods for the impact of teaching supporting resources and altmetric methods to evaluate the informal activities and the social impact. However, the assessment would be meaningless if the contents are not exhaustive enough. For this reason

this preliminary study focuses on the contents provided on scientists' personal websites and it not intend to address how to evaluate them, that is reserved for a further paper.

Highly cited scientists

It is known that scientists do not use the Internet in the same way and to the same extent (Barjak 2006), so it would be interesting to investigate in detail the behaviour of different groups. Particularly important for scholarly progress are those scientists with a relevant and internationally recognized research career. For this reason, the present study focused on highly cited researchers, particularly on those who work at European institutions. The population defined here will be also used in future papers focused on the modelization and conceptualization of new individual-level indicators (http://www.research-acumen.eu/).

The identification of this population is based on "ISIHighlyCited" database created by Institute for Scientific Information (ISI)/Thomson Reuters. The list consists of the most highly cited authors during the period 2000–2008 for each one of the 21 main disciplines of the ISI/Thomson Reuters citation databases. The number of authors per discipline is a maximum of 250. This database has been analysed in the past by other authors like Batty (2003a, b) that described the distribution of these scientists by country, place and institution and Basu (2006) that used this group to obtain an indicator of citation excellence at country level.

Research questions

In the context of our research (academic institutions), the personal website's credibility is perhaps stronger when it is presented within the institutional realm to which it belongs. For this reason, the present study looks only into websites hosted on the institutional domain (university, faculty, department, laboratory, hospital, etc.)

The specific questions to be addressed are as follow:

- Which are the characteristics (gender, country of affiliation and discipline) of European Highly cited (hereinafter EHC) researchers?
- Do these researchers have a personal website?
- What is the distribution by country and discipline of those researchers who have personal website?
- Do EHC researchers use the personal website to disseminate their work and careers?
- Do those researchers include enough contents on their personal websites to use it as a complementary source to evaluate them?
- Do those researchers who have no personal website, have research group website?
- Which countries and disciplines do researchers with research group websites pertain to?

Methods

Selection of highly cited scientists and their websites

The first step was the selection of highly cited researchers who worked at European institutions. The European countries were chosen from the official website of the European

Union (http://europa.eu/about-eu/countries/index_en.htm). The complete list of 45 European countries was chosen from the official website of the European Union (http://europa.eu/about-eu/countries/index_en.htm): the 27 member states of the European Union, 5 candidate countries and another 13 European countries. We listed the highly cited researchers working in these countries, based on the ISIHighlyCited.com database created by the ISI/Thomson Reuters. This database contained the 250 most highly cited researchers in each of the following 21 disciplines grouped into 5 broader areas: engineering (computer science, engineering, geosciences, materials science), hard sciences (chemistry, mathematics, physics, space sciences), health sciences (clinical medicine, immunology, microbiology, neuroscience, pharmacology), life sciences (agricultural sciences, biology & biochemistry, ecology/environment, molecular biology & genetics, plant & animal science) and social sciences (economics/business, psychology/psychiatry, general social science).

Only 22 of the 45 European countries were represented at least by one highly cited researcher. The last available edition of this database covered all the articles indexed during the period 1981–2008. For each researcher the following information, if available, were collected: researcher's name, main discipline, country of affiliation, country of birth, citizenship, year of birth, gender, and URL of his/her personal website.

Before the data collection was finished (September 2011), the ISIHighlyCited.com database was replaced by a new online directory (http://researchanalytics.thomsonreuters.com/highlycited/). The new directory only listed the researcher's name, affiliation and discipline, failing to provide the additional information. As both tools—database and online directory—were therefore incomplete and outdated, we adopted the following search strategies to fill any gaps: (1) the web of knowledge (WoK) database was used to find the full name of some researchers, (2) Google was used to locate the institution to which they belonged as well as their personal websites, and (3) researchers' personal websites were consulted to obtain personal information such as their birth date and country.

Classification of web contents

Having identified the highly cited researchers' personal websites, we proceeded to classify the web contents as indicated in Fig. 1, after the definition of (Fernández et al. 2009):

- An *institutional website* was understood to be a website hosted on a web domain of an academic institution (university, faculty, department, institute, laboratory, research group). We recognized two different types:
- *Personal website*: a website created by or for a researcher regardless of content. It typically includes biographical information, a Curriculum Vitae (CV), a list of publications, talks, and/or class materials.
- *Research group website*: a website hosted on the web domain of an academic institution and focused on a research group or laboratory.
- *A non-institutional website* was understood to be a site hosted on non-institutional web domains: a blog, a page in a community of experts, or a personal site with a particular domain.

We limited the scope of study to institutional websites alone and furthermore focused on highly cited researchers' personal websites. However, when we noted that some scientists used only research group websites, we decided to analyze these sites as well. For this reason we followed the specific research questions: 'Do those researchers who have no personal website, have research group website?' and 'Which countries and disciplines do

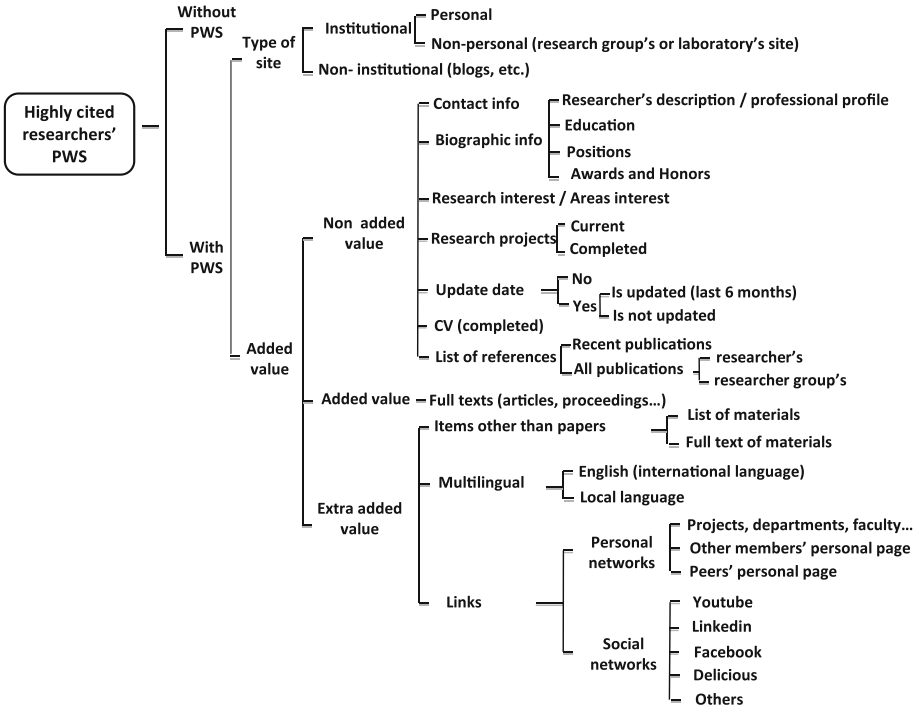


Fig. 1 Scheme used to classify and analyse the EHC researchers' personal websites

researchers with research group websites pertain to?' In this study the term "web presence" is considered in terms of having or not having personal website.

If the scientist had a personal website, the contents were classified according to Fig. 1:

- None added value information: basic personal information.
- Added value information: the full text of the researcher's publications is available.
- Extra added value information: the inclusion of items other than papers, such as additional teaching material or conference presentations, contents in several languages (local language and English), plus links to other sources and personal or social networks.

These contents were chosen because these could be candidate characteristics for the assessment of the scientist; for this reason, we checked if this group (EHC researchers) provided them.

Data collection and analysis was completed in February 2012.

One relevant shortcoming of our analysis can be attributed to the structural properties of the source used, on occasions not complete or fully updated. For example, we found deceased researchers (even several years ago), and others that had moved to another institution or even another different country (a few outside Europe). However, most of these problems were solved excluding the involved authors.

A second shortcoming is the disciplinary-biased coverage of database used. Of the 21 disciplines identified, 18 came under the area of Science and Technology, while only three disciplines belonged to the social sciences. Humanities were not included at all. A third

shortcoming is the language-biased coverage of database toward English speaking countries, especially toward the United States.

Results

Based on ISI/Thomson Reuters data of year 2011, we located 1,498 highly cited researchers working at European institutions, distributed in 22 different countries. These researchers represented 22.4 % of total of the highly cited researchers identified by ISI/Thomson Reuters.

European highly cited researchers' characteristics

As shown in Fig. 2, the distribution of EHC researchers was skewed, with a few countries (United Kingdom, Germany, France, Switzerland and Netherlands) accounting for most of these researchers. Almost half these scientists were in United Kingdom and Germany, especially in the first one. This can be due to role played by the language in the high citation levels of authors with bias toward English speaking countries. In "Others" were included the countries with less than 10 highly cited researchers: Ireland (9), Hungary (7), Russia (7), Greece (5), Poland (2), Romania (2), Cyprus (1) and Portugal (1).

We compared these data with Batty and Basu' studies (Batty 2003a, b; Basu 2006), based on data from December 2002 and April 2004 respectively, and we found that the number of EHC researchers had increased proportionally and therefore these researchers were distributed in the same countries. As an exception, Poland and Portugal had not increased the number of their highly cited scientists.

The findings regarding gender were noteworthy: 1,425 of this population (95 %) were males, while just 73 (5 %) were females. However, this pattern did not hold true in all the countries as for example the percentage of women was 14 % in Sweden and 12 % in Italy. This low presence of women among the highly cited scientists was not surprising, since previous studies had shown that female researchers tended to be less cited than men (Aksnes et al. 2011b; Kretschmer et al. 2012; Pudovkin et al. 2012) and that the higher the academic position, the lower the presence of women (Bordons et al. 2003; Mauleón and Bordons 2006; Prpic 2002; Torres-Salinas, Muñoz-Muñoz, and Jiménez-Contreras 2011). However, a study by Sánchez Peñas and Willett (2006) of five LIS departments showed

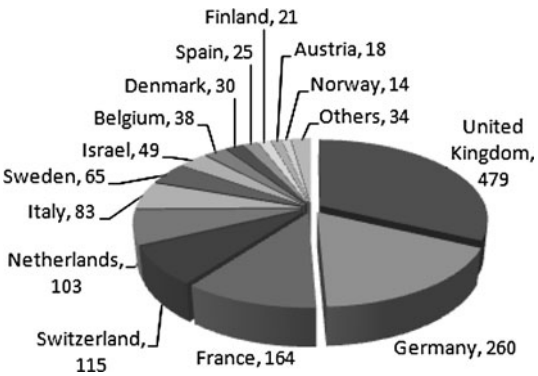


Fig. 2 Distribution by country of 1,498 EHC researchers

that although men published significantly more than women, no significant differences existed in the number of citations.

A substantial proportion of EHC researchers (117) were assigned to only one discipline, 77 to two disciplines, eight to three disciplines, and one researcher to four disciplines. Table 1 shows the distribution by discipline and country of these researchers. Disciplines are grouped under five larger areas: life sciences, health sciences, hard sciences, engineering and social sciences. The largest group (9 %) within Europe's highly cited researchers was that of pharmacologists, followed by botany & zoology (6.8 %), physics (6.6 %) and microbiology (6.3 %). The very low number of highly cited researchers in the social sciences (economics/business, psychology/psychiatry and general social sciences) reflects the bias in the coverage of the database used.

Concerning gender, male researchers were concentrated mainly in the health sciences (29 %) and life sciences (26 %), with a lesser presence in hard sciences (23 %) and engineering (17 %), and scarcely found in social sciences (3 %). Female researches were also mainly concentrated in life sciences (37 %) and health sciences (37 %), having a limited presence in hard sciences (15 %) and social sciences (8 %) and scarcely present in engineering (3 %). These data correlated with those of other studies (Aksnes et al. 2011b; Bordons et al. 2003; Mauleón and Bordons 2006; Pudovkin, Kretschmer, and Stegmann 2012), reporting that women were particularly concentrated in the medical sciences, and hardly appeared in engineering, for example.

The scientists' age was an item collected for just 771 (51 %) researchers of the population analysed, for reasons explained in the Methods section. We found that 4 % of these researchers were between 40 and 49 years in age; 25 % between 50 and 59; 42 % between 60 and 69; and 29 % over 70 years old. So, almost three quarters of the EHC researchers were over 60. This might suggest that most of them are people who have an established research career. However, two studies about Norwegian researchers (Aksnes et al. 2011a, b) found that the over 60 cohort were significantly less cited than their younger colleagues.

European highly cited researchers with personal website

Results showed that 1,030 (69 %) of the 1,498 EHC researchers had at least one website, while 468 (31 %) had none. Specifically, 925 researchers had a personal website, 117 had research group website, and 43 had both.

Figure 3 shows the percentage of highly cited researchers who had personal website in each European country. Countries are identified in the figure by their cTLD code and sorted by the number of highly cited researchers, thereby decreasing. Should be noted that in calculating these data was considered that 61 of EHC researches were in fact deceased (58 men and 3 women). We assumed that deceased researchers did not have personal websites or these were not updated, so, they were not take into account, decreasing to 1,437 EHC researchers the number of those analyzed in this part of the study.

Figure 3 reflects that United Kingdom and Germany were the countries with the highest number of researchers that had personal website, representing together the 54 %. This was expected, since as shows Fig. 1, these countries accounting for almost half of EHC scientists. However, we compared the number of highly cited researchers within a country and those who had personal website in the same country and we noted that Denmark, Israel and the United Kingdom had the best results: in Denmark worked 28 highly cited researchers, of which 89 % had a personal website; in Israel were are located 45 researchers, the 84 % had web presence and in United Kingdom for 479 researchers there are 346 (75 %) personal websites. Barjak (2006) reached a similar conclusion, stating that

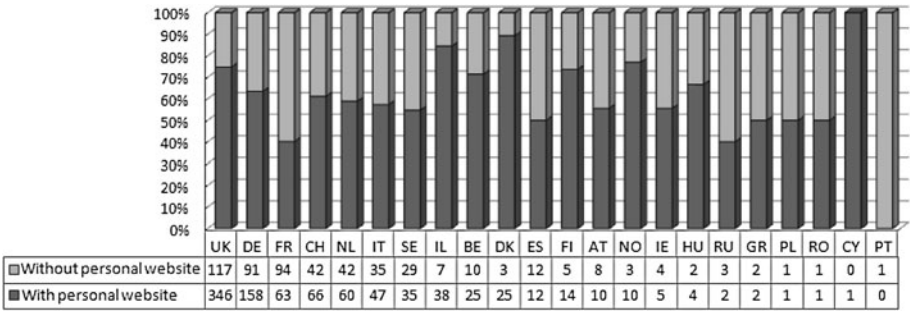
Table 1 Distribution by discipline and country of EHC researchers

Discipline	**No. total of researchers**	United Kingdom	Germany	France	Switzerland	Netherlands	Italy	Sweeden	Israel	Belgium	Denmark	Spain	Finland	Austria	Norway	Ireland	Hungary	Russia	Greece	Poland	Romania	Cyprus	Portugal
Agricultural Sci.	**89**	19	12	14	2	7	3	11	2	2	2	2	1		1	6		3	1		1		
Biology & Biochemistry	**59**	17	14	7	3	2	3	6		4						1			1	1			
Ecology / Environ.	**86**	28	9	3	8	9	2	6	4		3	2	6	1	4							1	
Molecular Biology & Genetics	**77**	18	22	10	8	8		1	3	1		4	2										
Plant & Animal Sci.	**109**	41	30	8	8	6		1	3	3	3				5		1						
Total	**420**	**123**	**87**	**42**	**29**	**32**	**8**	**25**	**12**	**10**	**8**	**8**	**9**	**1**	**10**	**7**	**1**	**3**	**2**	**1**	**1**	**1**	**0**
Clinical Medicine	**56**	29	2	2	2	7	6	2		1	1	1	2						1				
Immunology	**83**	17	7	14	14	6	11	2	2	5		1		2	1								1
Neuroscience	**92**	41	12	8	4	4	2	10			3	1	1	2	2		2						
Pharmacology	**143**	54	23	14	5	6	11	9	2	8	2	1	2	4		1				1			
Microbiology	**100**	26	25	9	8	14	1	2		10	1	2	1		1								
Total	**474**	**167**	**69**	**47**	**33**	**37**	**31**	**25**	**4**	**24**	**7**	**6**	**6**	**8**	**4**	**1**	**2**	**0**	**1**	**1**	**0**	**0**	**1**
Mathematics	**89**	27	9	21	4	3	3	2	8	1	4	4		1			2						
Physics	**106**	22	28	8	19	1	10	2	4		2	2	2	4				1	1				
Chemistry	**77**	22	27	4	8	6	4	2	1	1		1						1					
Space Sci.	**84**	32	15	2	6	9	13	1	3			1		1			1						
Total	**356**	**103**	**79**	**35**	**37**	**19**	**30**	**7**	**16**	**2**	**6**	**8**	**2**	**6**	**0**	**0**	**3**	**2**	**1**	**0**	**0**	**0**	**0**
Materials Sci.	**66**	19	18	8	6	2	2	3	2		1	1	2	1				1					
Computer Sci.	**71**	10	5	9	3	10	8	3	15	1		1		2		1	1		1		1		
Engineering	**47**	11	11	4	7		1	2	2	3	2		1	1	1				1				
Geosciences	**78**	27	12	24	3	3	1			1	4		1					1				1	
Total	**262**	**67**	**46**	**45**	**19**	**15**	**12**	**8**	**19**	**5**	**7**	**2**	**4**	**4**	**1**	**1**	**1**	**2**	**2**	**0**	**1**	**1**	**0**
Social Sci. General	**20**	12				2	1	4									1						
Economics / Business	**31**	14		6	2		2	1	1	1	2	2											
Psychology / Psychiatry	**31**	24	1		3				1	1			1										
Total	**82**	**50**	**1**	**6**	**5**	**2**	**3**	**5**	**2**	**2**	**2**	**2**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**
	1594	**510**	**282**	**175**	**123**	**105**	**84**	**70**	**53**	**43**	**30**	**26**	**22**	**19**	**15**	**9**	**8**	**7**	**6**	**2**	**2**	**2**	**1**

in Denmark the web is more important than in other countries for the dissemination of scientific information.

We also noted that in spite of the presence of women among EHC researchers was very low, the gender gap was smaller concerning web presence—65 % of men and 49 % of women had personal websites.

In the same way, when we compare the number of researchers within a discipline and those who had personal website in that discipline, results highlight mainly economics scientists (94 % of them had a personal website). Mathematics, computer science and



	UK	DE	FR	CH	NL	IT	SE	IL	BE	DK	ES	FI	AT	NO	IE	HU	RU	GR	PL	RO	CY	PT
Without personal website	117	91	94	42	42	35	29	7	10	3	12	5	8	3	4	2	3	2	1	1	0	1
With personal website	346	158	63	66	60	47	35	38	25	25	12	14	10	10	5	4	2	2	1	1	1	0

Fig. 3 EHC researchers with and without personal website in each country

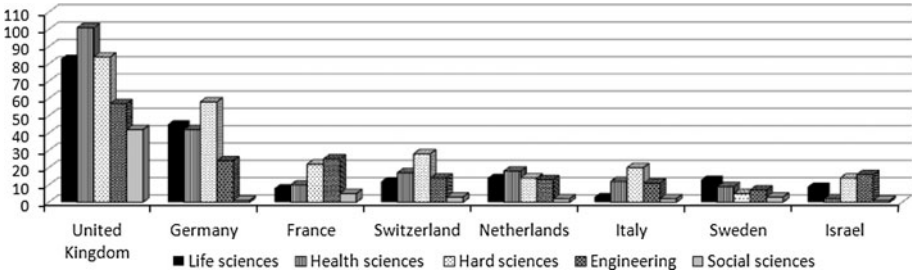


Fig. 4 Distribution by country and discipline of EHC researchers who had a personal website, grouped by country

space sciences researchers also belong to the same group: 89, 82 and 81 % respectively had personal websites. Agricultural sciences researchers were the group with the fewest (30 %) personal websites. These results support the Barjak's cited paper, who stated that "economists and computer scientists are more reliant than scientists from other disciplines on the WWW for obtaining and disseminating information" and that "economists use all types of electronic sources and computer scientists depend predominantly on personal web sites when searching for information" (Barjak 2006, p. 1362).

Figure 4 shows the disciplines of EHC researchers with a personal website, grouped by country (the eight countries having more researchers with personal websites). The most significant result was the lack of web presence of the social sciences with respect to the other disciplines. It does not mean that researchers in these disciplines did not have personal websites, but rather that they were hardly represented at all in the database used. Out of the 65 social science researchers having personal websites, 29 were economists, 23 were psychologists/psychiatrists and 13 worked in the general social sciences. Despite the small sample size, these results came to support those of Antelman (2006), Bergstrom and Lavaty (2007) and Batty (2009), who examined self-archive practices in economics and other social science disciplines.

Distribution by gender showed that male researchers with personal website worked primarily in hard sciences (27 %) and less in health sciences (25 %), life sciences (22 %), engineering (20 %) and social sciences (6 %); while women worked primarily in the research areas of life sciences (40 %) and less in health sciences (23 %), hard sciences (20 %) and social sciences (14 %). Only one woman appeared in engineering. These

differences were statistically significant for both men and women ($p < 0.0005$ and $p < 0.05$ respectively, Chi squared test).

A correspondence analysis (Fig. 5) revealed interesting differences by country as well. In United Kingdom, health science researchers had the highest web presence; yet in Germany, Switzerland and Italy, personal websites were more predominant in the hard sciences; in France and Israel engineering prevailed, albeit slightly. As we are talking of elite researchers probably these clusters are in fact showing specialised centres of excellence, not general patterns.

Contents provided on European highly cited scientists' personal website

Figure 6 shows the distribution of contents provided on personal websites related to personal information, experience and expertise. A total of 1,062 personal websites were identified as pertaining to 925 different researchers (890 males and 35 females). Contact information (supplied in 92 % of the sites) and research interests (61 %) were the most commonly provided data, being the research projects the least commonly provided information.

We also analysed whether researchers provided on their personal websites their CVs and lists of scholarly publications. We defined full CV presence as entries featuring a full list of publications in addition to biographical information, which meant that only 51 (5 %) sites provided it. Regarding format, 44 (86 %) full CV was in PDF, six (12 %) were in HTML and one (2 %) was available in both PDF and HTML formats. The short CV, only including biographical information, was included in 133 (13 %) of EHC researchers' personal websites; and of these, 77 (58 %) were in HTML, 48 (36 %) in PDF and eight (6 %) were in both HTML and PDF. Thus, there was a certain trend of presenting complete CVs in PDF and short CVs in HTML, and it was quite unusual to find the two types of CV on the same personal website.

Concerning scholarly publications, we checked whether the scientists had included on their personal websites a separate list of their recent publications, the full list of all their publications, or only a link to an external database or repository. Only 336 (32 %) websites

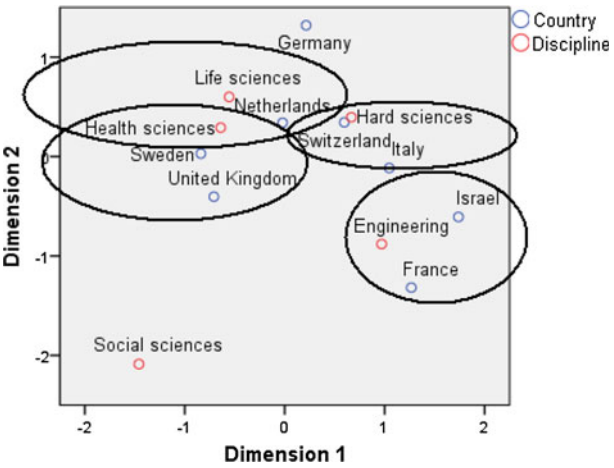


Fig. 5 Relationship between country and discipline of EHC researchers who had personal websites

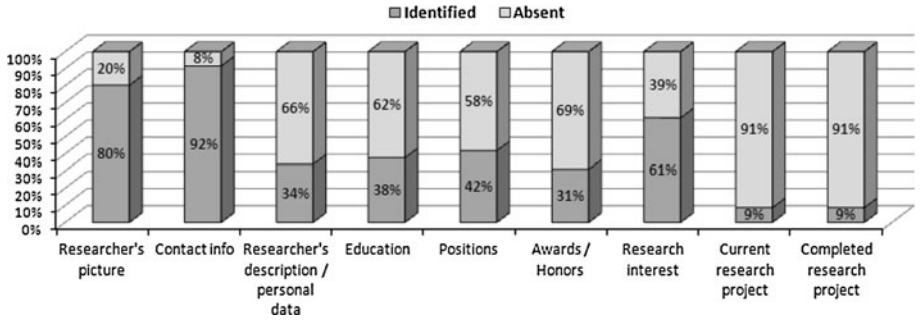


Fig. 6 Distribution by type of content of EHC researchers' personal websites, related to personal information, experience and expertise

had a list of recent or key publications; however, almost half the sites (40 %) included a list of all publications, the list usually being in HTML format. It is interesting to know if the list of publications is given in HTML of PDF because when this is in PDF, it hardly ever provides links to full text of those publications. On the other hand, when the list of publication is in HTML it is very often to find links pointing to the abstract or full text of some publications. These data suggest that EHC researchers prefer to provide complete information of their final formally published articles, chapters and books.

For 133 (13 %) sites there was no list of publications at all, only links to a database or repository where they could be downloaded in full text, an option frequently found on the space science webpages. The main databases used by scientists were their own institutional repository, PubMed (medicine), SAO/NASA Astrophysics Data System, Arxiv (physics and related hard sciences), Google Scholar and, less frequently, Spires (High Energy Physics). This suggests that a proportion of scientists prefer to promote their research output through the institutional repository or recognized databases.

It also was analyzed the format in which the full text of the publication was provided. A total of 182 (17 %) sites were found to include the full text of at least one publication in PDF format, 79 (7 %) sites provided a link to the full text of at least one publication in HTML format, and 248 (23 %) featured links to an external site with a summary of the publication, giving the option of downloading the text. There was great heterogeneity, some scientists linking to full text of only a few publications (probably those with granted rights), others linking to the full text of most or all of their publications.

On the other hand, only 368 sites (35 %) indicated their date of last update. For 215 (58 %) of these sites, a date was given but the website was not updated (the last change was not made in the previous 6 months, taking as reference the day when the website was analysed). Meanwhile, 153 (42 %) of these sites indicated the date of update and they were actually updated.

We also looked into further characteristics that, in our opinion, enrich with extra added value the researcher's personal website—for example, including material other than publications, such as PowerPoint presentations for teaching support, conferences presentations, workshops lectures, etc. Only 172 (16 %) of personal websites studied included such list of items, and just (10 %) included these materials in full.

Very few personal websites were multilingual, whereas 153 (14 %) sites were published in both the local language and English, 9 % (98 sites) were published only in the local language and 76 % (811 sites) were only in English. It is important to bear in mind,

however, that half of English-only sites belonged to researchers who worked in countries where one of the official languages was English (United Kingdom, Ireland, and Cyprus).

When we focused our analysis on links, it was seen that 598 (56 %) sites linked to research projects, research groups, laboratories, departments and/or the university central pages; 108 sites (10 %) linked to other people' personal websites (members of their own research group, postdoctoral students, other colleagues working on similar topics, etc.) and 254 (24 %) linked to other websites. Only 102 (10 %) personal websites were linked to a social network of some sort, being Facebook the main one (links to Facebook on 95 sites). These were not links to the researcher's profile, but rather made it possible to reach the researcher's website through the social network.

We also checked if EHC scientists provided on their personal websites a section with the latest news related to personal issues of the researcher or about his/her working areas, finding that only 25 (2.4 %) sites featured a section called "latest news" and just six (0.6 %) sites facilitated subscription via RSS (really simple syndication) to contents of researcher's personal website.

In terms of usage and collaborative tools, 20 (1.9 %) sites provided statistics such as the number of visits/visitors or the number of document downloads, and only one showed a graph describing the collaborative relationships of that researcher with others, or the relationship of the researcher with university departments: Ole E. Barndorff-Nielsen, from the University of Aarhus (http://pure.au.dk/portal/en/publications/discretevalued-levy-processes-and-low-latency-financial-econometrics%28092bde8d-2dca-41ce-8a1e-a6e45d223e50%29.html).

Another issue to explore was whether those researchers who had no personal website, had alternatively a research group website. There were 117 EHC researchers (110 males and 7 females) that had a group website. All except two were the leaders of their research groups. Most (74 researchers) used this site as the only tool for disseminating their research results, while the others (43 researchers) clearly combine both personal and group web-pages in a complementary way.

As we saw with the personal websites, the European countries with the highest number of highly cited researchers that had a research group website were again United Kingdom and Germany. However, with regard to disciplines we found interesting discrepancies: although most researchers with personal websites pertained to the hard sciences (especially mathematics), the ones with research group websites were associated mainly with life sciences and health sciences, especially molecular biology and genetics.

Discussion and conclusions

Today a researcher can have a diverse and rich web presence through various channels such as research group websites, institutional repositories, disciplinary repositories, scholarly databases or social websites; however they mainly provide information related to scholarly publications. A researcher's personal website is an excellent tool to provide, in addition to his/her list of publications, other information that is often difficult to find, such as experience and expertise, research projects performed, teaching commitment (materials for onsite or distance learning), conference presentations, events involvement, etc. So, a scientist's personal website can become one of the most complete channel for disseminating his/her work and career and perhaps the preferred one given a stronger control the authors themselves have on its contents. Keeping this in mind, the present study attempts to know which contents are provided by a recognized group of scientists on their personal

websites, exploring the possibility that some of them could be used in research evaluation. However, we do not analyze in the current paper not what information is important for the evaluation purposes or how website contents should be used in evaluation.

We limited the scope of study to institutional websites, that is, those hosted on the domain of the institution to which the scientists belonged. However, it should be noted that some scientists create their personal websites outside the domain of the institutions to which they belong to, maybe due to these institutions do not have a policy of personal websites or they do not offer the necessary means. To solve this problem, these academic institutions should be aware of the benefits of creating a space within its domain for researchers' personal websites, and so to improve the visibility of both faculty members and the institution per se. Other solutions to try that scientists do not create their personal websites outside the domain of academic institutions might be to develop a friendly CMS environment, a supportive publication policy and some guidelines for creating personal websites, as well as to recognize and to compensate the effort and time required for its maintenance, which as Hess (2002) states, in some cases takes place during researchers' free time.

Other consideration could be to give freedom to researchers with regard to design and contents. We strongly feel that the scientist must be the unique responsible for the contents of his/her personal website, and the institution merely should provide the necessary infrastructure and support, so that researchers were not limited by the institutional web-master decisions or fancy useless designs. This freedom is important because, as Thoms and Thelwall (2005) indicate, the term "personal website" loses its meaning when truly personal content, and therefore identity, is sacrificed.

The study focused on highly cited scientists who work at European institutions. Their identification of them was based on "ISIHighlyCited" database created by ISI/Thomson Reuters. The biases of the ISI/Thomson databases are well known and have been addressed in bibliometrics literature since the 1960s. Because these shortcomings also affect the highly cited researcher database, they should be taken into account in any discussion involving citation. The database used has a language-biased coverage toward English speaking countries, especially toward the United States. It also has a disciplinary-biased coverage, and in this case the number of social sciences researchers and the absence of any authors specializing in Humanities. It also should be noted the difficulty to find the correct researcher's personal website due to the problems with author name homonymy and synonymy.

The EHC group was clearly biased towards senior male researchers working in large countries, especially in United Kingdom and Germany. The low presence of women among the EHC researchers seems to support previous studies (see Sect. Introduction) which have shown that female researchers tend to be less cited than men as well as the higher the academic position, the lower the presence of women. However, the increase of women in academia should change this trend over time.

Results showed that 64 % of this population had a personal website. One explanation for this relatively low percentage may be the fact that most researchers were more than 60 years old. Although they are researchers with established careers and a lot of knowledge to share, the reasons why they do not use the web to share that knowledge could be different: (1) they do not know to use these technologies; (2) they do not consider it appropriate, either because they disagree with OAI or because they do not need promote their work, since they have already achieved an international recognition; (3) they do not have enough time to create a personal webpage and/or maintain it updated; (4) they use other web-based ways to disseminate their research, for instance through thematic or

institutional repositories, blogs, social networks, etc. In fact, results in this study showed that EHC researchers that belonged to life sciences and health sciences, especially molecular biology and genetics, preferred to use the group research website rather than personal website to divulge their academic and research activities.

Many of the EHC researchers with personal website worked in Denmark, Israel and United Kingdom. Broken down by disciplines, economist scientists were the best represented in personal web space, followed by mathematics, computer science and space sciences researchers, which probably reflects the success of subject OA repositories in these disciplines like ArXiv, RepEc or CiteSeerX. The reason for the high web presence for economists and computer scientists might be the need to increase their visibility due to their concerns about the tendency to use publication databases that do not adequately cover these disciplines, such as Thomson/ISI web of science. This result supports the Barjak's study (Barjak 2006) in which he indicated that economists and computer scientists are more reliant than scientists from other disciplines on the web for obtaining and disseminating information, and that computer scientists depend predominantly on personal websites when searching for information.

In terms of gender, most highly cited men with personal websites worked in hard sciences, while most of the highly cited women were dedicated to life sciences. There were also differences between countries: in United Kingdom the researchers with the highest web presence pertained to health sciences; in Germany, Switzerland and Italy they pertained to hard sciences; and in France and Israel they worked in engineering.

We checked other information topics provided on personal websites and results suggested that those related to personal information was usually comprehensive, past experience and expertise should be improved and current activity was short of the information needed for the evaluation purposes. Although the topics included in most of the websites are appropriated, unfortunately the contents provided are insufficient for a proper evaluation exercise for now. It should be interesting to examine other populations to see if they show the same behavior.

Acknowledgments This paper is supported by ACUMEN (Academic careers understood through Measurement and Norms) project, grant agreement number 266632, under the Seventh Framework Program of the European Union. Thanks are extended to Statistical Analysis Unit of the CCHS-CSIC (Spain) for its assistance and the anonymous reviewers for their valuable comments.

References

Aguillo, I. F., Ortega, J. L., & Fernández, M. (2008). Webometrics ranking of world universities: Introduction, methodology and future development. *Higher Education in Europe, 33*(2), 233–244.

Aksnes, D. W., Rorstad, K., Piro, F., & Sivertsen, G. (2011). Age and scientific performance. A large-scale study of Norwegian scientists. In *Proceedings of 13th ISSI Conference* (pp. 34–45). Durban, South Africa: ISSI.

Aksnes, D. W., Rorstad, K., Piro, F., & Sivertsen, G. (2011b). Are female researchers less cited? A large-scale study of Norwegian scientists. *Journal of the American Society for Information Science and Technology, 62*(4), 628–636.

ALLEA—ALL European Academies. (2012). *Open Science for the 21st century. A declaration of ALL European Academies*, Rome. http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/allea-declaration-1.pdf. Accessed April 2012.

Antelman, K. (2004). Do open-access articles have a greater research impact? *College & Research Libraries, 65*(5), 372–382.

Antelman, K. (2006). Self-archiving practice and the influence of publisher policies in the social sciences. *Learned Publishing, 19*(2), 85–95.

Bailey, C. W. (2010). *Transforming scholarly publishing through open access: A bibliography*. Houston: CreateSpace. http://digital-scholarship.org/tsp/transforming.pdf. Accessed May 2012.

Barjak, F. (2006). The role of the Internet in informal scholarly communication. *Journal of the American Society for Information Science and Technology, 57*(10), 1350–1367.
Barjak, F., Li, X., & Thelwall, M. (2007). Which factors explain the web impact of scientists' personal homepages? *Journal of the American Society for Information Science and Technology, 58*(2), 200–211.
Basu, A. (2006). Using ISI's 'Highly Cited Researchers' to obtain a country level indicator of citation excellence. *Scientometrics, 68*(3), 361–375.
Batty, M. (2003a). Citation geography: It's about location. *The Scientist, 17*(16). http://jmichaelbatty.files.wordpress.com/2011/06/batty-scientist-2003.pdf. Accessed April 2012.
Batty, M. (2003b). The geography of scientific citation. *Enrironment and Planning A, 35*, 761–765.
Batty, L. M. (2009). Self-archiving of articles published in high-impact journals in the social sciences. A Master's paper for the M.S. in L.S. degree. http://ils.unc.edu/MSpapers/3469.pdf. Accessed February 2012.
Bergstrom, T., & Lavaty, R. (2007). *How often do economists self-archive?* http://www.escholarship.org/uc/item/69f4b8vz. Accessed February 2012.
Bordons, M., Morillo, F., Fernández, M. T., & Gómez, I. (2003). One step further in the production of bibliometric indicators at the micro level: Differences by gender and professional category of scientists. *Scientometrics, 57*(2), 159–173.
Chen, C., Sun, K., Wu, G., Tang, Q., Qin, J., Chiu, K., et al. (2009a). The impact of internet resources on scholarly communication: A citation analysis. *Scientometrics, 81*(2), 459–474.
Chen, C., Tang, Q., Huang, X., Wu, Z., Hua, H., Yu, Y., et al. (2009b). An assessment of the completeness of scholarly information on the internet. *College & Research Libraries, 70*(4), 386–401.
Döring, N. (2002). Personal home pages on the web: A review of research. *Journal of Computer-mediated Communication, 7*(3). http://jcmc.indiana.edu/vol7/issue3/doering.html. Accessed July 2011.
Dumont, K., & Frindte, W. (2005). Content analysis of the homepages of academic psychologists. *Computers in Human Behavior, 21*(1), 73–83. http://www.sciencedirect.com/science/article/pii/S0747563204000123. Accessed July 2011.
Fernández, M., Zamora, H., Ortega, J. L., Utrilla, A. M., & Aguillo, I. F. (2009). Género y visibilidad web de la actividad de profesores universitarios españoles: el caso de la Universidad Complutense de Madrid. *Revista Española de Documentación Científica, 32*(2), 51–65.
Flanagin, A. J., & Metzger, M. J. (2003). The perceived credibility of personal web page information as influenced by the sex of the source. *Computers in Human Behavior, 19*(6), 683–701.
Frandsen, T. F. (2009a). The effects of open access on un-published documents: A case study of economics working papers. *Journal of Informetrics, 3*(2), 124–133.
Frandsen, T. F. (2009b). The integration of open access journals in the scholarly communication system: Three science fields. *Information Processing and Management, 45*(1), 131–141.
Herring, S. D. (2002). Use of Electronic Resources in Scholarly Electronic Journals: A Citation Analysis. *College & Research Libraries, 63*(4), 334–340.
Hess, M. (2002). A Nomad faculty: English professors negotiate self-representation in university web space. *Computer and Composition, 19*, 171–189.
Hyland, K. (2011). The presentation of self in scholarly life: Identity and marginalization in academic homepages. *English for Specific Purposes, 30*(4), 286–297.
Kousha, K. (2009). Characteristics of open access scholarly publishing. *Aslib Proceeding, 61*(4), 394–406.
Kousha, K., & Thelwall, M. (2006). Motivations for URL citations to open access library and information science articles. *Scientometrics*, *68*(3), 501–517, http://www.springerlink.com/content/w84q15536062tx71/fulltext.pdf. Accessed March 2012.
Kousha, K., & Thelwall, M. (2007). The web impact of open access social science research. *Library & Information Science Research, 29*, 495–507.
Kousha, K., & Thelwall, M. (2008). Assessing the Impact of disciplinary Research on Teaching: An automatic analysis of online syllabuses. *Journal of the American Society for Information Science and Technology, 59*(13), 2060–2069.
Kousha, K., Thelwall, M., & Rezaie, S. (2010). Using the web for research evaluation: The integrated online impact indicator. *Journal of Informetrics, 4*(1), 124–135.
Kretschmer, H., Pudovkin, A., & Stegmann, J. (2012). Research evaluation. Part II: Gender effects of evaluation: Are men more productive and more cited than women? *Scientometrics*. doi:10.1007/s11192-012-0658-0. http://www.springerlink.com/content/fq78g70l6m22x202/fulltext.pdf. Accessed March 2012.
Kurtz, M. J. (2004). *Restrictive access policies cut readership of electronic research journal articles by factor of two*. http://opcit.eprints.org/feb19oa/kurtz.pdf. Accessed November 2011.
Lawrence, S. (2001a). Free online availability substantially increases a paper's impact. *Nature, 411*, 521.

Lawrence, S. (2001b). Online or invisible? *Nature*, *411*(6837), 521. http://ivyspring.com/steveLawrence/SteveLawrence.htm Accessed August 2011.
Davis, M. P. Lewenstein, B. V., Simon, D. H., & Connolly, M. J. L. (2008). Open access publishing, article downloads, and citations: Randomised controlled trial. *British Medical Journal, 337*. http://www.bmj.com/content/337/bmj.a568. Accessed March 2012.
Marcus, B., Machilek, F., & Schütz, A. (2006). Personality in cyberspace: Personal web sites as media for personality expressions and impressions. *Journal of Personality and Social Psychology, 90*(6), 1014–1031.
Mauleón, E., & Bordons, M. (2006). Productivity, impact and publication habits by gender in the area of Material Science. *Scientometrics, 66*(1), 199–218.
Papacharissi, Z. (2002a). The presentation of self in virtual life: Characteristics of personal home pages. *Journalism and Mass Communication Quarterly, 79*(3), 643–660.
Papacharissi, Z. (2002b). The Self Online: The Utility of Personal Home Pages. *Journal of Broadcasting Electronic Media, 46*(3), 346–368.
Parks, M., & Archley-Landas, T. (2003). *Communicating self through personal homepages: Is identity more than screen deep*? Paper presented at the annual conference of the International Communication Association, San Diego, CA. http://citation.allacademic.com//meta/p_mla_apa_research_citation/1/1/2/0/4/pages112040/p112040-1.php. Accessed September 2011.
Petric, G. (2006). Conceptualizing and measuring the social uses of the Internet: The case of personal web sites. *Information Society, 22*(5), 291–301.
Pitzek, S. (2002) *Impact of online-availability of science literature*. http://www.vmars.tuwien.ac.at/courses/proseminar/doc/paperserver.pdf. Accessed July 2011.
Prpic, K. (2002). Gender and productivity differentials in science. *Scientometrics, 55*(1), 27–58.
Pudovkin, A., Kretschmer, H., Stegmann, G. (2012). Research evaluation. Part I: Productivity and citedness of a German medical research institution. *Scientometrics.* doi:10.1007/s11192-012-0659-z.
Sánchez Peñas, C., & Willett, P. (2006). Brief communication Gender differences in publication and citation counts in librarianship and information science research. *Journal of Information Science, 32*(5), 480–485.
Schmitt, K., Dayanim, S., & Matthias, S. (2008). Personal homepage construction as an expression of social development. *Developmental Psychology, 44*(2), 496–506.
Shin, E. J. (2003). Do impact factors change with a change of medium? A comparison of impact factors when publication is by paper and through parallel publishing. *Journal of Information Science, 29*(6), 527–533.
Swan, A. (2007). Open Access and the Progress of Science. *American Scientist, 95*, 198–200.
Swan, A. (2010). *The open access citation advantage. Studies and results to date*. Technical Report, School of Electronics & Computer Science, University of Southampton. http://eprints.ecs.soton.ac.uk/18516/2/Citation_advantage_paper.pdf. Accessed March 2012.
Thelwall, M., & Harries, G. (2004). Do the web sites of higher rated scholars have significantly more online impact? *Journal of the American Society for Information Science and Technology, 55*(2), 149–159.
Thelwall, M., Kousha, & Kayvan, M. (2008). Online presentations as a Source of Scientific Impact? An analysis of PowerPoint files citing academic journals. *Journal of the American Society for Information Science and Technology, 59*(5), 805–815.
Thoms, L., & Thelwall, M. (2005). Academic home pages: Reconstruction of the self. *First Monday, 10*(12), http://www.firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1302/1222. Accessed August 2011.
Torres-Salinas, D., Muñoz-Muñoz, A. M., & Jiménez-Contreras, E. (2011). Análisis bibliométrico de la situación de las mujeres investigadoras de Ciencias Sociales y Jurídicas en España. *Revista Española de Documentación Científica, 34*(1), 11–28.
Vaughan, L., & Thelwall, M. (2003). Scholarly use of the web: What are the key inducers of links to journal web sites? *Journal of the American Society for Information Science and Technology, 54*(1), 29–38.
Vazire, S., & Gosling, S. D. (2004). E-perceptions: Personality impressions based on personal web sites. *Journal of Personality and Social Psychology, 87*(1), 123–132.
Weibel, D., Wissmath, B., & Groner, R. (2010). Motives for Creating a Private Website and Personality of Personal Homepage Owners in Terms of Extraversion and Heuristic Orientation. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *4*(1).
Zinkhan, G. M., Conchar, M., Gupta, A., & Geissler, G. (1999). Motivations underlying the creation of personal web pages. *Advances in Consumer Research, 26*, 69–74.