

ACUMEN

academic careers understood through measurement and norms

FP7 Grant Agreement | 266632

Deliverable No and Title | D2.5 – Portfolio Model

Dissemination Level | PU (public)

Work Package | WP2

Version | 1.0

Release Date | 28 March 2014

Author(s) | Kayvan Kousha and Michael Thelwall

Project Website | <http://research-acumen.eu/>

European Commission
7th Framework Programme
SP4 - Capacities
Science in Society 2010
Grant Agreement: 266632



EUROPEAN
COMMISSION



Deliverable D2.5 Portfolio Model

Although assessing the value of research is still a complex issue, citation counting has commonly been used as a practical method to estimate the ‘formal impact of research’. However, there has been a long debate on using bibliometric indicators and in particular citation counting for research evaluation. Because the main sources of citation data are citation indexes (e.g., Thomson Reuters Web of Science and Elsevier’s Scopus), which contain a subset of the academic literature (e.g., mainly selected high impact English journals). In response, it has been suggested that citation data from wider types of information sources can help for monitoring research impact, especially in some social sciences and many humanities fields where journal articles and their citations are less important than in the sciences. In particular, there are some types of informal impacts of research that cannot be measured by citation counts. For instance, some research might be frequently used in teaching, highly discussed in scientific communities or frequently read by the public or by academics.

To fill this gap, ACUMEN Portfolio suggests a web metrics model for impact assessment of research as a supplement to, but not replacement for, traditional bibliometric indicators by including new types of web evidence about the value or use of research. The ACUMEN Portfolio web metrics model is based on previous research and ACUMEN research during the project. The proposed web metrics can provide a wide range of types of impact evidence from the web. The new metrics have been rated with a 5-stars system to reflect the quality of the evidence they provide in as clear and practical way as possible. This is designed to minimise the risk that the metrics are used unquestioningly. Hence, academics or reviewers can estimate scientific influence or value of research not only based on conventional citation metrics but also based on most high quality web metrics. Two major components of ACUMEN web metrics are (for details see *Guidelines for Good Evaluation Practice with the ACUMEN Portfolio available at <http://research-acumen.eu/portfolio>*):

- **Modeling web extracted citations:** This includes a series of methods and tools developed to extract citations from online documents indexed by Google Books, Google Scholar and commercial search engines (e.g., Google, Bing). This gives broader view of the impact of research based on citations from digitized books and, non-journal documents (e.g., online theses, research/technical reports, course syllabuses and preprint repositories) and PowerPoint presentations.
 - **Google Books citation:** Counting citations from books can help scholarly impact assessment and that Google Books (GB) is a useful source of such citation counts, despite its lack of a public citation index. To make it practical for ACUMEN Portfolio to extract citation metrics from online books, we introduced a method to automatically remove false and irrelevant matches from GB citation searches. The method was evaluated by manual checking of sampled GB results and comparing citations to about 14,500 monographs in the Thomson Reuters Book Citation Index (BKCI) against automatically extracted citations from GB across 24 subject areas. GB citations were 103% to 137% as numerous as BKCI citations in the humanities, except for tourism (72%) and linguistics (91%), 46% to 85% in social sciences, but only 8% to 53% in the sciences. In all cases, however, GB found substantially more citing books than did BKCI, with BKCI's results coming predominantly from journal articles. Result suggests that GB citation model could measure the different aspects of impact in ACUMEN Portfolio, filling the current gap for assessing research outputs in books based-fields, especially in arts and humanities. (see **Publication 1, Kousha and Thelwall, in press 2014a**).

- **Google Books in Webometric Analyst:** As an exclusive software to meet the ACUMEN project, a computer application was designed and tested for automatic citation extraction from Google Books. The programme is freely available and individuals, reviewers, institutions and universities across Europe and other countries can download and use the application for monitoring their research performance based on citations from millions of online books (**see Publication 2, Thelwall and Kousha, Google Books in Webometric Analyst**).
- **Modeling the web presence of European academics:** In order to investigate how the EU researchers (including highly cited researchers) use web to disseminate research, we conducted a large scale email survey of and analysed hyperlinks from academic web CVs.
 - **Disseminating Research with Web CV by European academics:** Some academic web CVs, including homepages and publication lists, link to open access (OA) papers, resources, abstracts in publishers' websites, or academic discussions, helping to disseminate research. To assess how common such practices are and whether they vary by discipline, gender and country, we conducted a large scale email survey of astronomy and astrophysics, public health, environmental engineering, and philosophy across 15 European countries and analysed hyperlinks from academic web CVs. About 60% of the 2,154 survey responses reported having a web CV or similar and there were differences between disciplines, genders and countries. A follow-up outlink analysis of 2,700 web CVs found that a third had at least one outlink to an OA target, typically a public eprint archive or an individual self-archived file. This proportion was considerably higher in astronomy (48%) and philosophy (37%) than in environmental engineering (29%) and public health (21%). There were also differences in linking to publishers' websites, resources and discussions. Perhaps most importantly, however, the amount of linking to OA publications seems to be much lower than allowed by publishers and journals, suggesting that many opportunities for disseminating full-text research online are being missed, especially in disciplines without established repositories. Moreover, few academics seem to be exploiting their CVs to link to discussions, resources or article abstracts, which seems to be another missed opportunity for research publicity (**see Publication 3, Kousha and Thelwall, in press 2014b**).
 - **Web presence of European highly cited researchers:** The online presence of about 1,500 highly cited researchers working at European institutions showed that about 70% of them have a personal website or other web contents, specially the scientists from Denmark, Israel and the United Kingdom. Nevertheless, was results biased towards senior male researchers working in large countries (e.g., The United Kingdom and Germany). The most frequent disciplines with high web presence were economics, mathematics, computer sciences and space sciences, suggesting the success of open access subject repositories like RepEc, Arxiv or CiteSeerX (**see Publication 4, Mas-Bleda & Aguillo, 2013**). A further study of 1,525 highly cited scientists working at European institutions also found that 61% of them had a personal website, although this was higher in social sciences with 79% and lower in health sciences with 49%. Webometric analysis was carried out for the 892 scientists with either a personal website or an online list of publications for which the crawler used worked, showing that 355 (40%) of them created at least one outlink to open access sources [OA repositories, pdf, doc, docx, rtf, ps, gz], being higher in hard sciences (59%) and social sciences (58%) and engineering (45%) than life sciences (26%) and health sciences (19%). Disciplinary and geographical differences for outlinking to OA research also were found for sampled highly cited

scientists (see **Publication 5, Mas-Bleda, Thelwall, Kousha & Aguillo, 2014**). It was also investigated if these highly cited researchers had social web presences and if these presences had a measurable impact. It was found a very low use of social sites, although researchers having one type of profile were more likely to have another in many cases. Most social web profiles had some evidence of uptake, if not impact, but the value of the indicators used is unclear (see **Publication 6, Mas-Bleda, Thelwall, Kousha, & Aguillo, 2013**).

- **Modeling using social web for research communication:** Nowadays, researchers use informal scholarly channels to disseminate and publicise their research through the web. To investigate this, we conducted three studies based on 1) *YouTube science videos* cited in academic publications 2) *ResearchGate* statistics which is a social network site for academics and 3) *Academia.edu* social network statistics.
 - **The role of online videos in research communication:** Although there is some evidence that online videos are increasingly used by academics for informal scholarly communication and teaching, the extent to which they are used in published academic research is unknown. This article explores the extent to which YouTube videos are cited in academic publications and whether there are significant broad disciplinary differences in this practice. To investigate, we extracted the URL citations to YouTube videos from academic publications indexed by Scopus. A total of 1,808 Scopus publications cited at least one YouTube video and there was a steady upward growth in citing online videos within scholarly publications from 2006 to 2011, with YouTube citations being most common within arts and humanities (0.3%) and the social sciences (0.2%). A content analysis of 551 YouTube videos cited by research articles indicated that in science (78%) and in medicine and health sciences (77%) over three quarters of the cited videos had either direct scientific (e.g., laboratory experiments) or scientific-related contents (e.g., academic lectures or education), whereas in the arts and humanities about 80% of the YouTube videos had art, culture or history themes and in the social sciences about 63% of the videos were related to news, politics, advertisements and documentaries. This shows both disciplinary differences and the wide variety of innovative research communication uses found for videos within the different subject areas (see **publication 7, Kousha, Thelwall and Abdoli, 2012**).
 - **Motivations for citing YouTube videos in the academic publications:** This article explores what are the common broad reasons for using online videos across major subject areas. A contextual analysis of citations to YouTube videos based upon 207 sampled full-text publications indicated that in sciences 70% and in medicine about half of the citations were created to visual information (e.g., real-time scientific demonstrations and laboratory experiments) as evidence to support or illustrate scholarly arguments. However, in arts and humanities YouTube videos were mostly used for citing artistic performances or visual outputs (43%) such as music, dance or movie parts and in social sciences direct or indirect citations to talks, speeches or lectures (45%) were common motivations for citing YouTube videos. This suggests that this new non-standard academic medium can be helpful for research communication, although different fields cite online videos for different reasons in scientific literature (see **Publication 8, Kousha and Thelwall, 2012**).
 - **Academia.edu as informal source of impact:** Academic social network sites Academia.edu and ResearchGate and reference sharing sites Mendeley, Bibsonomy,

Zotero, and CiteULike give scholars the ability to publicise their research outputs and connect to each other. With millions of users, these are a significant addition to the scholarly communication and academic information seeking eco-structure. There is thus a need to understand the role that they play and the changes, if any, that they can make to the dynamics of academic careers. This article investigates attributes of philosophy scholars on Academia.edu, introducing a median-based time-normalising method to adjust for time delays in joining the site. In comparison to students, faculty tend to attract more profile views but female philosophers did not attract more profile views than did males, suggesting that academic capital drives philosophy uses of the site more than friendship and networking. Secondary analyses of law, history and computer science confirmed the faculty advantage (in terms of higher profile views) except for females in law and females in computer science. It also found a female advantage for both faculty and students in law and computer science as well as for history students. Hence, Academia.edu overall seems to reflect a hybrid of scholarly norms (the faculty advantage) and a female advantage that is suggestive of general social networking norms. Finally, traditional bibliometric measures did not correlate with any Academia.edu metrics for philosophers, perhaps because more senior academics use the site less extensively or because of the range informal scholarly activities that cannot be measured by bibliometric methods (see **Publication 9, Kousha and Thelwall, in press, 2014a**).

- **ResearchGate metrics for research evaluation:** ResearchGate is a social network site for academics to create their own profiles, list their publications and interact with each other. Like Academia.edu, it provides a new way for scholars to disseminate their publications and hence potentially changes the dynamics of informal scholarly communication. This article assesses whether ResearchGate usage and publication data broadly reflect existing academic hierarchies and whether individual countries are set to benefit or lose out from the site. The results show that rankings based on ResearchGate statistics correlate moderately well with other rankings of academic institutions, suggesting that ResearchGate use broadly reflects traditional academic capital. Moreover, while Brazil, India and some other countries seem to be disproportionately taking advantage of ResearchGate, academics in China, South Korea and Russia may be missing opportunities to use ResearchGate to maximise the academic impact of their publications (see **publication 10, Kousha and Thelwall, in press, 2014b**).

A theoretical discussion on the impact of web-mediated scholarly communication in ACUMEN:

Quantitative research impact assessment has relied primarily upon analyses of citations between published academic documents in citation indexes (e.g., Scopus, Web of Science) which can only cover a limited subset of the scientific literature. There are also other academic activities and outputs that are essential to the progress of science but rarely result in formal publications—such as teaching, scholarly discussions, science videos, presentations—which may sometimes cite research in ways that would be undetectable through traditional citation indexes.

Web sources are increasingly used in research and scholarly communication and it may be possible to extend bibliometric methods beyond conventional citation indexes to the web. In a book chapter we discuss the use of web metrics for assessing the impact of academic research—whether artifacts, articles, researchers or institutions and argues that web impact metrics could potentially

supplement conventional impact metrics by including new or unique types of sources of impact (e.g., presentations, syllabi or digitised books), emerging types of scientific outputs (e.g., online videos or science blogs).

We described different methods to capture web impact, including hyperlinks, web citation, URL citations and hybrid approach (Web/URL citation). These methods have been used to collect web impact evidence by a web crawler or by queries to commercial search engines and have commonly been compared with conventional counterparts (e.g., WoS or citations) to assess whether these web metrics are relevant for research evaluation.

We reviewed previous attempts to extract and use formal citations from web databases and digital libraries including CiteSeer, Google Scholar, Google Book and reports previous results about using them as an alternative citation impact metrics. It has been discussed that Google Scholar citation metrics (citations counts, h-indexes, etc.) and Google Books citations from a huge number of digitised books can be used for monitoring research performance, especially in the social sciences, arts and humanities, when traditional citation indexes are not available or have insufficient coverage.

We introduced new types of web impact including citations from online syllabi and course reading lists which reflect the educational impact of research, download counts of academic publications which may be an indicator of reading and usage rather than formal citation counts. We briefly discussed emerging social web impact metrics or Altmetrics which can potentially be used outside standard academic sources and indicators such as social bookmarks, tweets or online readership of scientific publications or views of online academic videos.

We concluded that there are many ways in which research impact can be assessed using the web and the practical applications of web extracted metrics for research assessment include calculating indicators for objects outside of traditional citation indexes, from scientific publications to scholars and institutions. However, web impact indicators suffer from a generic lack of quality control compared with scholarly citations, and hence should be used cautiously in research evaluation (**see publication 11, Kousha & Thelwall, 2014c**).

Publications associated with D2.5 ACUMEN Portfolio Model

1. Kousha, K. & Thelwall, M. (in press, 2014a). An automatic method for extracting citations from Google Books. *Journal of the American Society for Information Science and Technology*. Full-text available online: <http://www.scit.wlv.ac.uk/~cm1993/papers/AutomaticGoogleBooksCitationsPreprint.pdf>
2. Thelwall and Kousha. (2013). Google Books in Webometric Analyst. *Software*. Available: <http://lexiurl.wlv.ac.uk/searcher/googlebooks.htm>
3. Kousha, K. & Thelwall, M. (in press, 2014b). Disseminating Research with Web CV Hyperlinks. *Journal of the American Society for Information Science and Technology*. Full-text available online: <http://www.scit.wlv.ac.uk/~cm1993/papers/WebCVHyperlinks.pdf>
4. Mas-Bleda, A., & Aguillo, I. F. (2013). Can a personal website be useful as an information source to assess individual scientists? The case of European highly cited researchers. *Scientometrics*, 96(1), 51-67. Full-text available online: <http://digital.csic.es/bitstream/10261/90892/1/Mas-Bleda%20%26%20Aguillo%202013%20-%20Postprint%20of%20EHC%20researchers%27%20personal%20websites.pdf>

5. Mas-Bleda, A., Thelwall, M., Kousha, K., & Aguillo, I. F., (2014). Successful researchers publicizing research online: An outlook analysis of European highly cited scientists' personal websites, *Journal of Documentation*, 70(1), 148-172. Full-text available online: <http://www.scit.wlv.ac.uk/~cm1993/papers/SuccessfulResPublicizing-preprint.pdf>
6. Mas-Bleda, A., Thelwall, M., Kousha, K., & Aguillo, I. F., (2013). European Highly Cited Scientists' Presence in the Social Web. In: *14 th International Society of Scientometrics and Informetrics Conference*. Vienna, Austria, pp.1966-1969. Full-text available online: http://www.issi2013.org/Images/ISSI_Proceedings_Volume_II.pdf
7. Kousha, K., Thelwall, M. & Abdoli, M. (2012). The role of online videos in research communication: A content analysis of YouTube videos cited in academic publications, *Journal of the American Society of Information Science and Technology*, 63(9), 1710–1727. Full-text available online: <http://www.scit.wlv.ac.uk/~cm1993/papers/TheRoleOfOnlineVideosInResearchCommunication-preprint.pdf>
8. Kousha, K. & Thelwall, M. (2012). Motivations for Citing YouTube Videos in the Academic Publications: A Contextual Analysis. 17th International Conference on Science and Technology Indicators (STI), 5-8 September, 2012 in Montreal, Quebec, Canada. Full-text available online: http://sticonference.org/Proceedings/vol2/Kousha_Motivations_488.pdf
9. Thelwall, M. & Kousha, K. (in press, 2014a). Academia.edu: Social network or academic network? *Journal of the American Society for Information Science and Technology*. Full-text available online: http://www.scit.wlv.ac.uk/~cm1993/papers/Academia_preprint.pdf
10. Thelwall, M. & Kousha, K. (in press, 2014b). ResearchGate: Disseminating, communicating and measuring scholarship? *Journal of the American Society for Information Science and Technology*. Full-text available online: <http://www.scit.wlv.ac.uk/~cm1993/papers/ResearchGate.pdf>
11. Kousha, K. & Thelwall, M. (2014c). Web Impact Metrics for Research Assessment. In: B. Cronin & C.R. Sugimoto, (Eds), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, MIT Press. ISBN: 978-0262026796. Summary of book chapter available online: http://www.kousha.tripod.com/summary_mit_kousha_thelwall.pdf