Wind power research in Wikipedia.

Does Wikipedia demonstrate direct influence of research publications and can it be used as adequate source in research evaluation?

Serrano-López, Antonio Eleazar; Ingwersen, Peter; Sanz-Casado, Elias

Published in:
Scientometrics

DOI:
DOI 10.1007/s11192-017-2447-2

Publication date:
2017

Document version
Peer reviewed version

Scientometrics, June 2017, DOI: 10.1007/s11192-017-2447-2; vol. (112): 1471-1488

Wind power research in Wikipedia: Does Wikipedia demonstrate direct influence of research publications and can it be used as adequate source in research evaluation?

Antonio Eleazar Serrano-López[2], Peter Ingwersen[1], Elias Sanz-Casado[2]

[1]*Royal School of Library and Information Science, University of Copenhagen, Denmark*

[2]*Carlos III University Madrid, Spain*

Aim: This paper is a result of the WOW project (Wind power On Wikipedia) which forms part of the SAPIENS (Scientometric Analyses of the Productivity and Impact of Eco-economy of Spain) Project (Sanz-Casado et al., 2013). WOW is designed to observe the relationship between scholarly publications and societal impact or visibility through the mentions of scholarly papers (journal articles, books and conference proceedings papers) in the Wikipedia, English version. We determine 1) the share of scientific papers from a specific set defined by Wind Power research in Web of Science (WoS) 2006-2015 that are included in Wikipedia entries, named data set A; 2) the distribution of scientific papers in Wikipedia entries on Wind Power, named data set B, captured via the three categories for the topic Wind Power in the Wikipedia Portal: Wind Power, Wind turbines and Wind farms; 3) the distributions of document types in the two wiki entry data sets' reference lists. In parallel the paper aims at designing and test indicators that measure societal impact and R&D properties of the Wikipedia, such as, a wiki reference focus measure; and a density measure of those types in wiki entries.

Methods: The study is based on Web mining techniques and a developed software that extracts a range of different types of Wikipedia references from the data sets A and B.

Results: Findings show that in data set A 25.4% of the wiki references are academic, with a density of 17.62 academic records detected per wiki entry. However, only 0.62 % of the original WoS records on Wind Power are also found as wiki references, implying that the direct societal impact through the Wikipedia is extremely small for Wind Power research. In the second Wikipedia set on Wind Power (data set B), the presence of scientific papers is even more insignificant (10.6%; density: 3.08; WoS paper percentage: 0.26 %). Notwithstanding, the Wikipedia can be used as a tool informing about the *transfer* from scholarly publications to popular and non-peer reviewed publications, such as Web pages (news, blogs), popular magazines (science/technology) and research reports. Non-scholarly wiki reference types stand for 74.6% of the wiki references (data set A) and almost 90% in data set B. Interestingly, the few WoS articles in wiki entries on Wind Power present on average 34.3 citations received during the same period (2006-2015) as WoS Wind Power publications not mentioned in wiki entries only receives on average 5.9 citations.

Conclusions: Owing to the scarcity of Wind Power research papers in the Wikipedia, it cannot be applied as a direct source in evaluation of Wind Power research. This is in line with other recent studies regarding other subject areas. However, our analysis presents and discusses six supplementary *indirect* indicators for research evaluation, based on publication types found in the wiki entry reference lists: Share of (WoS) records; Density; and Reference Focus, plus Popular Science Knowledge Export, Non-Scholarly Knowledge Export and Academic Knowledge Export. The same indicators are direct measures of the Wikipedia reference properties.

Introduction

The main goal of the WOW (Wind Power on Wikipedia) project is to measure the relationship between scientific publications and societal interest or impact on common citizens through the mentions of scholarly papers in Wikipedia, including co-referenced publications. The Wikipedia system is a crowdsourcing developed information system on the Web. As a «free encyclopedia» it is completely free in many ways: It is free to access; anyone with a device connected to the Internet can access to all contents of Wikipedia in many different languages. In addition, everyone can add, modify or delete entries, and no registration is required. The Wikipedia has some tools in order to prevent vandalism and fake information, as the «librarians» (users with watchdog permissions to revert changes, ban users, etc.) or bots, which are looking constantly for bad words, errors or fakes to correct these (Henderson, 2010;

Wikipedia, 2016). We are interested in the Wikipedia entries that 1) are associated with given scholarly publications, e.g. publications under research evaluation, and 2) contain references of scholarly and non-scholarly nature to be included in scientometric measurements of societal influence. Hence, since we deal with references and mention of scholarly records in the Wikipedia, we regard the measurements and indicators as belonging to Scientometric publication and citation analysis rather than to Webometrics as defined by Björneborn & Ingwersen (2004) or Altmetrics (Thelwall, 2016), even though the Wikipedia form part of the Web.

Wikipedia *mentions* are in general considered as an useful indicator to measure the diffusion and social divulgation of Science (Allen et al., 2013; Trueger et al., 2015). Added to classic scientometric indicators, Wikipedia mentions of given departmental publications and their distributions in wiki entries may serve as potentially valid indicators. The distribution pattern of WoS records over wiki entries is, like mentions, regarded as an *direct* research assessment approach. Co-reference patterns of a variety of (non-)scholar publication types, associated to given WoS records in wiki entries, are regarded *indirect* assessments of the original units to be evaluated, but seen as *direct* assessments of the Wikipedia itself.

Although everybody can edit Wikipedia entries, some studies demonstrate that it *is* possible to find academic experts and scholars contributing to Wikipedia in many different fields (Stein & Hess, 2007). Other works have analyzed the presence of Wikipedia in the general scholarly and scientific systems (Park, 2011). They looked at the impact of the Wikipedia on science by analyzing the Web of Science and Scopus systems for citations to Wikipedia entries.

More in line with the present study Teplitskiy, Lu & Duede (2016) identified the 250 most highly cited journals in each of 26 research fields (4,721 journals, 19.4M articles) indexed by the Scopus database, and tested whether topic, academic status, and accessibility make articles from these journals more or less likely to be referenced on the Wikipedia (English version). They found that a journal's academic status (impact factor) and accessibility (open access policy) both strongly increase the probability of it being referenced on the Wikipedia. However, Teplitskiy, Lu & Duede demonstrate (2016, Table 4) that in their large-scale sample the English-language Wikipedia's coverage of academic research varies immensely across the 26 academic fields and, most importantly, the coverage measured in mentions is scarce, between 0.04% (Dentistry) and 0.5% (Social Sciences). The Energy field has a Wikipedia coverage of 0.05% in their study.

Very recently, Kousha & Thelwall (2016) counted Wikipedia citations to 302,328 articles and 18,735 monographs in English indexed by Scopus from multiple subject areas in the period 2005 to 2012. On average 5% of the articles are cited (mentioned) by the English Wikipedia, with Environmental Science only reaching 3.4%. Energy Science does not take part in their study.

Earlier analyses have similar scarcity of Wikipedia citations in topically limited subject areas. Luyt & Tan (2010) analyzed 50 History Wikipedia entries for the types of wiki references they presented. 62% of the 480 wiki references detected were Web-based with 17.1% of all references originating from various governmental sources, 11.9 % from news media and 11.9% from internet sources. The most cited type among the non-Internet Wikipedia references were books, with 34.9% of all the 480 wiki references. Almost no journal articles were cited. Given the History subject this is not surprising. In a later investigation Koppen, Phillips &

Papageorgiou (2015) found that based on 21 drugs they extracted 601 references from corresponding Drug entries in the English Wikipedia. Like in the case of Luyt & Tan (2010) they were interested in the distribution of types of the references. 50% of all 601 references were journal articles, but only 20.4% of the same set belonged to crucial Core Clinical Journals, as defined by National Library of Medicine. Commercial Websites (11.1%), news media (10.5%) and government Websites (9.2%) were the most frequent types of wiki references observed. Further, in their study books ranked 5th with 6.5% of the distribution. Central meta-analyses and guidelines only counted for 2.2% and 0.8%, respectively.

However, among these studies only Teplitskiy, Lu & Duede (2016) and Kousha & Thelwall (2016) have investigated the amount of a given set of scholarly documents that Wikipedia mentions concerning a topic or institution, and their characteristics. It is our opinion that in order to use scholarly references from Web of Science or Scopus found in Wikipedia entries (from now on named 'scholarly wiki references') to form part of research evaluation or capture of societal impact, it is vital that the amount of such references in these entries is statistically fitting. According to Kousha & Thelwall (2016) the scarcity of academic references prohibits such analyses. An overall view of the analyses presented above indicates strongly that the distribution of types of wiki references is non-systematic and depends on the subject area in question. A trend seems to be that open access (OA) and free Web-based sources are more likely to be part of Wikipedia references (Teplitskiy, Lu & Duede (2016)), as books in humanistic subject areas (Luyt & Tan (2010); Kousha & Thelwall (2016)). The investigations, including the present one, can in addition bring answers to the question of how the information transfer between a scholarly and non-scholarly environments takes place, and what kind of information is used and useful to common citizens through the Wikipedia.

Following the Introduction this article describes the Methodology, followed by the Findings and Discussion sections. They are divided into results and indicators concerning the set of Wikipedia entries defined by references to WoS publications on Wind Power and the complementary findings in relation to the set of Wikipedia entries defined by the topic 'Wind Power'. Where relevant we compare and discuss outcome patterns and usability of the two data sets. A concluding section ends the article.

Methodology

The SAPIENS project (Sanz-Casado et al., 2013; Ingwersen et al., 2013; 2014) demonstrated interesting scientometric results concerning the development of Wind Power research 1995-2009, in particular the extensive use of conference proceeding papers in the scientific communication process (approx. 60 % of all research publications) and the very low citation impact of this document type. In the present analysis we apply the same retrieval profile as used in the SAPIENS Project, see Appendix I, to isolate a basic set of WoS records on Wind Power.

The present study makes use of two different methods when collecting data from the Wikipedia: Data set A: first to isolate Wind Power publications of all types in WoS and then detecting their occurrence together with other kinds of references in Wikipedia entries; Data set B: searching directly and isolating Wind Power entries in Wikipedia and detecting the occurrence of scholarly and other types of publications in this set of wiki entries. The reason for applying data set A is to ensure that the same research entities are evaluated in WoS and in Wikipedia, for instance a topic (as in the present study), a department or country output or an

author's research publications. Data set B illustrates a 'quick and dirty' way of collecting data directly in the Wikipedia on a topic.

For data set A we developed a script to extract scientific (and other types of) references from Wikipedia. The script was written in Python language (including urllib library), which includes many kinds of search patterns to find WoS records mentioned in Wikipedia entries: paper title (for long titles), paper title and source title, paper title and author, DOI, etc. This script returned a CSV file with one line for each WoS record and how often is it mentioned. Thus, we first isolated and downloaded a set of Wind Power records captured from WoS (25,540 records; publication window: 2006-2015), using the search strategy from the SAPIENS Project, Appendix I. Then the set of WoS records was cross-checked against the English Wikipedia version by means of the script. Each *mention* in Wikipedia entries, e.g. in the wiki Reference List, of record elements from the WoS set (title, title + authors, doi…) was detected and checked. Duplicates were removed. The script retrieved reliable elements pointing to 159 WoS records, representing *0.62 %* of the original set af WoS Wind Power records, and defining the set of 159 unique Wikipedia entries holding 258 mentions of the WoS records. Since each wiki entry does not repeat the same reference twice on its Reference List, some of the WoS records are co-occurring in the set of wiki entries, together with other kinds of wiki references, in total 11,027 wiki references. We may regard the 159 WoS records found in the Wikipedia as an (extremely small) sample drawn from the original 25,540 WoS records, seen as the population.

Data set B was generated by searching and isolating all Wikipedia entries directly on 'Wind Power' and then detecting the kind of references to publications found in this topic-defined set of Wikipedia entries. In praxis we used the Wikipedia Web portal related to Renewable Energy (http://en.wikipedia.org/wiki/Portal:Renewable_energy). From this portal we extracted (using web mining techniques) all wiki entries related to Wind Power (67 entries, Appendix II) and all their references to publications, including scholarly and non-scholarly articles and papers (in total 2,387 wiki references). A script similar to that above programmed in Python language was used for the wiki reference extraction. This script used the Wikipedia API to retrieve the information of each entry in JSON format. In this format, the script runs over the Wikipedia entries, goes to the reference list and extracts every reference to a CSV document containing each full reference, like item type, author, and publication year, title of the document referenced, URL, addition date and title of the Wikipedia entry. Also, the script takes into account the variety of ways which Wikipedia allows to use for references, and looks for these patterns in source code (using HTMLParser library), for example: <ref>, {{cite}}, {{citation}}, {{doi}}, {{ISBN}}, etc.

The problem is that not all of these fields are complete, for example, the information on publication year or authorship are often incomplete when Web pages are referenced.

The main difference between the two data sets is that WoS publications in set A on Wind Power can be mentioned in Wikipedia entries also dealing with other topics than the chosen topic. Also notice that most of the 67 Wikipedia entries from Data set B are included in the 159 wiki entries defined by the WoS records on Wind Power (Data set A). Other characteristics of the two data sets are discussed in the Findings sections.

Three simplistic research evaluation indicators

We have designed three indicators associated with the data sets A and B that determines how much scholarly information transfer occurs from traditional sources (peer reviewed journals, conferences, etc.) to the Wikipedia, and thus assumingly further transferred to common citizens.

One indicator deals with the *Share* (*R*) of WoS records found in Wiki entries (p_{WoS}) over a set of WoS records retrieved on a given topic, author, institution or country (P_{WoS}). Essentially, the indicator measures the amount of knowledge export or direct societal influence of WoS research publications onto Wikipedia entries – measured as ratio (*R*) in %:

$$R = 100 * (p_{WoS} / P_{WoS})$$

This indicator is a *direct* measure and a consequence of data set A, and is in line with the Kousha & Thelwall measure (2016, Table 1, p. 667). At *R* = 95-100 % almost all original WoS records (the population) are also found in Wikipedia entries and common knowledge export, influence or impact is almost total. Using the set of Wikipedia entries in alternative research evaluation is thus statistically sound. In the case of a very low ratio (< 1.5 %) too few scholarly records associated with the object to be evaluated exist in wiki entries. In that case, any alternative direct measure that involves that data set, including its use as societal impact indicator, cannot be statistically valid. Consequently, Ratio *R* may rather act as a predictor of probability of utility in research evaluation and should be calculated prior to other indicators (see discussion in Findings sections).

The second indicator concerns the *Density* (*D*) of number of wiki references (*n*), regardless publication type, or *mentions* (*m*) of WoS publications on the topic, in a given set of Wikipedia entries (*N*):

$$D = n / N \qquad \text{or} \qquad D = m / N$$

The Density indicator is an *indirect* evaluation measure and can be applied to data sets A or B described above. In the case of data set A, by logic its scholarly output can never go below 1.0. Various versions of the Density indicator are available. For instance, by applying data set A the *WoS Density* (D_{WoS-A}) – or average mention – concerns the number of WoS record mentions (m_{WoS-A}) in the set of wiki entries that is defined by retrieved WoS records on a given topic, author, institution of country, etc. (N_{WoS-A}): $D_{WoS-A} = m_{WoS-A} / N_{WoS-A}$. For instance, as shown above the number of WoS record mentions found in the WoS defined wiki entries is 258 (m_{WoS-A}) and the number of wiki entries is 159 (N_{WoS-A}); thus the Density $D_{WoS-A} = 1.6$.

By applying both data sets A and B it is possible to compare the density of the individual publication types, or their sum, thereby comparing the data capture methods. We define scholarly publication types found in WoS and Wikipedia as peer reviewed 'journal articles', 'review articles' and 'proceedings papers'. Other publication types referred to in Wikipedia entries on their List of References are, for instance, Web page, Popular Magazine article, R&D Report, News article. The categorization of wiki references is carried out semi-automatically. The density (D_{pop-B}) of Popular Magazine articles (n_{pop-B}) in a set of wiki entries on Wind Power (N_B): $D_{pop-B} = n_{pop-B} / N_B$ can be compared to the similar formula applied to data set A. The Density indicator is a good predictor for the usefulness of applying *indirect* alternative

assessments in a Wikipedia setting, such as the amount of non-scholarly publications co-occurring with original scholarly publications to be evaluated.

The third indicator, named *Wiki Reference Focus (F)*, is *indirect* and is the normalized Density of a particular (group of) wiki reference type(s). By this normalization it is possible to compare the Density indicator values across data sets and academic fields:

$$F = D_{type\text{-}B} \ / \sum (D_B)$$

where $\sum (D_B)$ signifies the sum of all document type densities in a given set (here Data set *B*) of wiki references. The value span of the Wiki Reference Focus is 0.0 – 1.0. The indicator displays the ranking of wiki reference types in a given set of wiki entries.

All three indicators are direct measures of properties of the Wikipedia and can be used for evaluation purposes of this particular source.

Findings and discussion

Below we discuss the central properties of the two Wikipedia data sets if they might be useful elements of assessments. This could be the distribution pattern of the 258 WoS publication mentions over the 159 wiki entries (Table 1), the co-occurrence of WoS records found in Wikipedia entries (Table 2), distributions of Document Types, Data Density and Wiki Reference Focus (Table 3), leading to knowledge transfer indications, indirect research assessments (co-referencing of original WoS records) and citation distributions (Figure 2). When relevant we compare to the original WoS set of records on Wind Power, because that is the set 1) to be observed for societal influence and 2) that forms the object of a research evaluation. Initially we apply the filter, ratio *R*, to the two data sets A and B to observe the societal impact of the original WoS set through Wikipedia.

Wikipedia data sets A and B – applying the ratio filter

With respect to data set A, based on the population of WoS records (25,540) on Wind Power and retrieved in the English Wikipedia (159 entries), one notices that this share (ratio *R*) of WoS records in the Wikipedia on the topic Wind Power is extremely low:

$$R = 100\ (p_{WoS\text{-}A} / P_{WoS}) = 100 \times (159 / 25{,}540) = 0.62\ \%$$

The Wikipedia data set A on Wind Power demonstrates small influence of the original set of WoS records. Further, in line with the findings by Kousha & Thelwall (2016) it does *not* possess sufficient statistical certainty to be applied as basis for alternative direct evaluation of Wind Power research through the Wikipedia. Statistically, with a standard confidence interval of 0.95 and an estimated error of 0.5 (5 %), the minimum sample size for a population of 25,540 records would be 379 records. With a sample size of 159 records the estimated error increases to 7.6 % - still keeping the confidence interval at 0.95. Keeping this insufficiency in mind we demonstrate below selected characteristics of Wikipedia data that eventually can be used as alternative indirect research evaluation tools. The *R*-value for 379 records is 1.5 % and thus serves as a borderline case. This implies that the 258 mentions of original WoS records in the 159 wiki entries also is below the threshold with a share of R = 1.01%. Note that in the case of Kousha & Thellwal (2016) their Environmental Science subject field, close to sustainable energy research, reaches 3.4%. The Energy field in the analysis by Teplitskiy, Lu & Duede (2016, Fig. 4, p. 5) demonstrates an 12 times lower *R*-value: 0.05%. Why this is the case may be because Teplitskiy, Lu & Duede (2016) rely on the 250 most high impact journals in each field

indexed by Scopus, where Kousha & Thelwall and we apply all articles in the selected subject fields.

Using the smaller data set B on Wind Power the *R* value is correspondingly extremely low: *0.26%*. Wikipedia data set B shows even less influence from the original scholarly WoS records, compared to data set A. It demonstrates hardly any societal influence and should not be applied as an alternative in direct research evaluation, e.g. using the amount of 'mention' as indicator. Nevertheless, certain *other* reference elements of selected wiki entries might prove useful as alternative *indirect measures* (Table 3), e.g. the amount of non-scholar co-references, Table 3.

Distribution of Web of Science records mentioned in Wikipedia entries

Only 11 original WoS documents are mentioned 3 or more times (Table 1) in the Wikipedia data set A. Notice that the Wikipedia entries in the set (Table 2) may not be mainly about Wind Power but rather on other related generic issues, e.g. Renewable Energy; Energy Storage; Sensory Ecology; Climate Change Mitigation, in which Wind Power plays an aspectual role, as determined by the wiki entry authors in the set. The percentage of wiki entries (109 entries) in Data set A *not* dealing mainly with the topic at hand (Wind Power) indicates the *degree of transfer, spreading or association* of the topic to other fields or specialties in the Wikipedia: 100 * 109 / 159 = *68 %*. This measure is on normalized set level and can be compared to the captured set by Method B (Appendix II). In set B all the 67 wiki entries are directly on the topic at hand. So although set B is much smaller than set A the former is far more focused on Wind Power.

Since only 3 documents have 10 or more mentions across the set of Wiki entries (Table 1), it is very probable that the English Wikipedia follows a power law distribution like the Bradford or Lotka laws, with a few WoS documents accumulating a lot of mentions, while the long tail of documents obtains one-time-mentions of scholarly WoS publications. The same pattern can be observed for data set B with respect to the original WoS records mentioned. In data set A the total number of mentions of original WoS records is 258 (Table 1). On average each wiki entry in data set A contains 258 / 159 = 1.6 WoS records from the original set of 25,540 records. This small Density figure makes their application insufficient as a direct measure of scholarly references in Wikipedia associated with elements of academic records. Likewise, Table 3 demonstrates the distribution of wiki entry references from the two data sets, the density and the wiki reference focus. It is obvious that scholarly wiki references, including the original WoS records, constitute a quite small proportion of all wiki references.

Table 1. WoS titles mentioned in Wikipedia entries (>=3) – data set A. (Spring, 2016)

WoS Title (N=159)	Times mentioned (n=258)
Wind energy	25
Providing all global energy with wind, water, and solar power, Part I: Technologies, energy resources, quantities and areas of infrastructure, and materials	15
Providing all global energy with wind, water, and solar power, Part II: Reliability, system and transmission costs, and policies	10
Advances in solar thermal electricity technology	6

WoS Title (N=159)	Times mentioned (n=258)
Review of solutions to global warming, air pollution, and energy security	6
Supplying baseload power and reducing transmission requirements by interconnecting wind farms	5
KiteGen project: control as key technology for a quantum leap in wind energy generators	3
Life cycle assessment of two different 2 MW class wind turbines	3
Meteorologically defined limits to reduction in the variability of outputs from a coupled wind farm system in the Central US	3
The history and state of the art of variable-speed wind turbine technology	3
Towards an electricity-powered world	3
Peer Production and Desktop Manufacturing: The Case of the Helix_T Wind Turbine Project	2

Further, rather few Wikipedia entries mention several of the original WoS documents. In data set A only seven entries include four or more co-occurring references to documents in Web of Science as such (Table 2). They are often entries by authors about themselves, like Roy Billinton, Henrik Lund, Benjamin K. Sovacool, Mark Z. Jacobson or Martin J. Pasqualetti.

Table 2. Mentions of WoS records by Wikipedia entries, Data set A (>=3 mentions)

Wikipedia Entry Title (N=159)	Times Mentioned of WoS records (n=258)
Environmental impact wind power	8
Roy Billinton	8
Henrik Lund (academic)	6
Intermittent energy source	6
Wind power	5
Wind turbine design	5
Benjamin K. Sovacool bibliography	4
Feed- tariff	3
Life-cycle greenhouse-gas emissions energy sources	3
Mark Z. Jacobson	3
Martin J. Pasqualetti	3
Renewable energy	3
Renewable energy debate	3
Sensory ecology	3
Solar updraft tower	3
Sustainable energy	3
Wind farm	3
Wind turbine	3

Figure 1 demonstrates extracts from the reference list of the Wikipedia entry for “Wind Farm” – found in set A as well as set B. A glance demonstrates that most often the co-references to a Wind Power WoS record (Ref. no. 42, Figure 1) consists of press releases, newspaper articles, web pages, governmental or institutional reports, etc., i.e., non-scholarly

publications. See Table 3 for breakdown into document types of the sets A and B. As demonstrated in Figure 1 *other* scholarly WoS records (no. 6 and no. 132) also co-occur with the original WoS Wind Power record. These concurrent reference types can probably be applied as *alternative (indirect) measures* of scientometric nature associated with the scholarly objects of evaluation, e.g., author or institutional name, journal titles, a country, a topic or single or a set of WoS records. These kinds of measures, and foremost the ones based on non-scholarly wiki references, such as the *share*, *density* and *focus* of Popular Magazines, are probably the most valuable indirect research evaluation indicators provided by the Wikipedia in relation to scholarly output. Simultaneously, they provide direct indicators of the Wikipedia's mode of influencing its readers.

References

1. Watts, Jonathan & Huang, Cecily. Winds Of Change Blow Through China As Spending On Renewable Energy Soars, The Guardian, 19 March 2012, revised on 20 March 2012. Retrieved 4 January 2012.
2. Terra-Gen Press Release, 17 April 2012
3. Wind energy-- the facts: a guide to the technology, economics and future of wind power page 32 EWEA 2009. Retrieved 13 March 2011.
4. "WINData LLC - Wind energy engineering since 1991". WINData LLC. Retrieved 28 May 2015.
5. IEC61400-1 site assessment
6. *Meyers, Johan; Meneveau, Charles (2012-03-01). "Optimal turbine spacing in fully developed wind farm boundary layers". Wind Energy 15 (2): 305–317. doi:10.1002/we.469. ISSN 1099-1824*.
7. "Historic Wind Development in New England: The Age of PURPA Spawns the "Wind Farm"". U.S. Department of Energy. 9 October 2008. Retrieved 24 April 2010.

---

42. Garvine, Richard; Kempton, Willett (2008). "Assessing the wind field over the continental shelf as a resource for electric power" (PDF). Journal of Marine Research 66 (6): 751–773. doi:10.1357/002224008788064540. ISSN 0022-2402. Retrieved 30 November 2009.

---

132. Roy, Somnath Baidya. Impacts of wind farms on surface air temperatures Proceedings of the National Academy of Sciences, 4 October 2010. Retrieved 10 March 2011.

133. Takle, Gene and Lundquist, Julie. Wind turbines on farmland may benefit crops Ames Laboratory, 16 December 2010. Retrieved 10 March 2011

Fig. 1. Reference list (extract) from the Wikipedia entry 'Wind Farm', data sets A+B (Spring 2016). Legend : entries in *Italics : scholarly publ.* ; entries in ***bold+italics*** : ***WoS record*** from original WoS set on Wind Power.

Document types in the two data sets

Table 3 demonstrates the division of wiki references into document types for both data sets. In set A the share of scholarly references, including original WoS records, is 25.4%. The density of the original WoS records (1.62) is very low but that of scholarly references (17.63) is

substantial compared to the average density for data set A (69.35) and also compared to the corresponding lower density of data set B (35.63). In this comparison the *wiki reference focus* should be used (Focus, scholarly wiki references: Set A: 0.25; Set B: 0.11), as this indicator is set-normalized. The three predominant non-scholarly types of wiki references: Web pages; Popular/News magazines; Research reports constitute foci of 0.75 and 0.89, respectively for sets A and B. These types illustrate indirect indicators in two ways: 1) each concurrent *individual* type signifies a specific replacement of scholarly information (WoS records and books) into popular scientific magazines or factual information in the form of non-scholarly web pages and reports from institutions; 2) the three types *together* constitute a non-scholarly, non-peer reviewed *profile of information visibility* in (or potential popular influence on) society by the Wikipedia associated with the original academic WoS publication elements. Unfortunately neither the Kousha & Thelwall or the Teplitskiy, Lu & Duede (2016) study address the proportion of all scholarly as well as non-scholarly references that are used alongside the original set of selected scholarly references detected in the wiki entries. However, Koppen, Phillips & Papageorgiou (2015, Table 1, p. 142) demonstrate that in their analysis of Drug associated wiki entries the share of scholarly references reaches 58.8%, with journal articles amounting to 50.1%. In our study journal articles reach 16% in data set A and 6.1% in set B. The distributions within the non-scholarly wiki references are rather alike, with Web pages as the strongest type in both data sets (Table 3: set A: 42.2%; set B: 53.2%), followed by Magazine/News articles (set A: 20.9%; set B: 22.6%). When comparing the various study findings, the subject area seems to be the determining factor in the distribution of Wikipedia reference types.

Table 3. Document types for Wikipedia References on Wind Power, Data sets A & B

	Data set A				***Data set B***			
Document types	N = 159	%	Density	Focus	N = 67;	%	Density	Focus
Book	688	6.2	4.33	0.06	82	3.4	1.22	0.03
Conference Paper	196	1.8	1.23	0.02	14	0.6	0.20	0.006
Conf. Paper in journal issue	90	0.8	0.57	0.008	2	8.4-4	0.03	0.0008
Journal Article	1764	16.00	11.09	0.16	147	6.1	2.19	0.06
Magazine/News Article	2300	20.90	14.47	0.21	539	22.6	8.04	0.23
Patent	39	0.4	0.25	0.004	2	8.4-4	0.03	0.0008
Report	824	7.5	5.18	0.07	217	9.1	3.24	0.09
Web page	4650	42.20	29.25	0.42	1271	53.2	19.00	0.53
Other, *	476	4.3	2.99	0.04	111	4.7	1.66	0.05
Total/avg. wiki references (n):	**11027**	100	**69.35**	1.0	**2387**	**100**	**35.63**	1.0
No. of wiki entries with no refs.:	0				0			
Max. number of refs. in entry:	264	2.4	264		264	11.1	264	
Scholarly wiki references:	2801	***25.4***	***17.62***	***0.25***	254	***10.6***	***3.8***	***0.11***
Non-scholarly wiki references:	8226	74.6	51.73	0.75	2133	89.4	31.83	0.89
No. of WoS records mentioned in set:	258		1.62		64		0.96	

* Other, method A: citation needed (196); *review article (63)* , etc.

* Other, method B: citation needed (65); *review article (9)* , etc.

In addition, the distribution of wiki references is quite different in the two data sets on Wind power, manly owing to the proportions of scholarly references in the two sets. The

extraction method may thus also influences the distribution. Since data set A is defined by the occurrence of the scholarly WoS records in Wikipedia, we find it very likely that such a set contains additional scholarly references as well. In data set A journal articles constitute the dominant academic reference type (16 %; density 11.09; focus: 0.16); but in set B, defined by direct searching in the Wikipedia, the proportion of that type is only 6.1 % (density 2.19; focus: 0.06). These figures should be compared to the distribution of document types in the original Wind Power set in WoS, where conf. proceeding papers, incl. papers published in journal issues, constitute approx. 60 % of all publications (Sanz-Casado et al., 2013). In the Wikipedia the focus indicator for the same document type is extremely low (set A: 0.008; set B: 0.0008). The data set figures stress that the authors of the Wikipedia entries on Wind Power have made a conscious and dedicated selection of references to be displayed, targeting their presentation towards common users.

An example of alternative measures applying reference co-occurrence could be that a subset of data set A signifies a set on Wind Power, e.g. generated by a given university department under evaluation. Aside from use of traditional scientometric indicators, such as citations normalized for field and journals, such alternative *indirect* measures of Wikipedia references could be (scores from Table 3):

- *Popular science knowledge export/influence*: Share of Magazine articles: 20.9%, set A; the bigger the share the more potential influence.
- *Non-scholarly knowledge export/influence*: Share of all non-scholarly references: 74.6%, set A; the bigger the share the more potential influence of non-peer reviewed knowledge.
- *Academic knowledge export/influence*: Share of all scholarly references (types in *Italics*, Table 3): 25.4%, set A; the bigger the share the more potential export and direct influence of scholarly knowledge on Wikipedia readers.
- *Wiki reference focus*: the density figure of a publication type (or group of types) normalized by the overall average density of the set to which the type(s) belong: e.g. *Web page, set A: 29.25 / 69.35 = 0.53* (max. value: 1.0). The score can be compared to a different set of wiki entries' references, e.g. the Web page wiki reference focus calculated from the data set provided by the Koppen, Phillips & Papageorgiou analysis (2015, Table 1, p. 141-142): (24.7/100 share * 601 wiki references) / 21 Drug wiki entries = 7.07 (density) / 28.5 (total density in set) = *0.25* (Wiki Ref. Focus).

These four indicators rely on the degree to which the document type algorithm can detect correctly the variety of publication types presented in wiki entries. In dubious cases the correct classification is done manually by the researcher team. Typically, the categories of 'Other', 'Journal articles' and 'Conf. paper in journal issue' are difficult to separate by the algorithm. 'Citation needed' is a category signifying that the wiki entry author still wishes to add a reference to the list.

Figure 2 demonstrates the distribution of wiki references from data set B over their publication years and document types. The two large reference types, Web pages and Popular magazine/news articles, are mainly from 2008 to the present, while reports are of more recent nature.

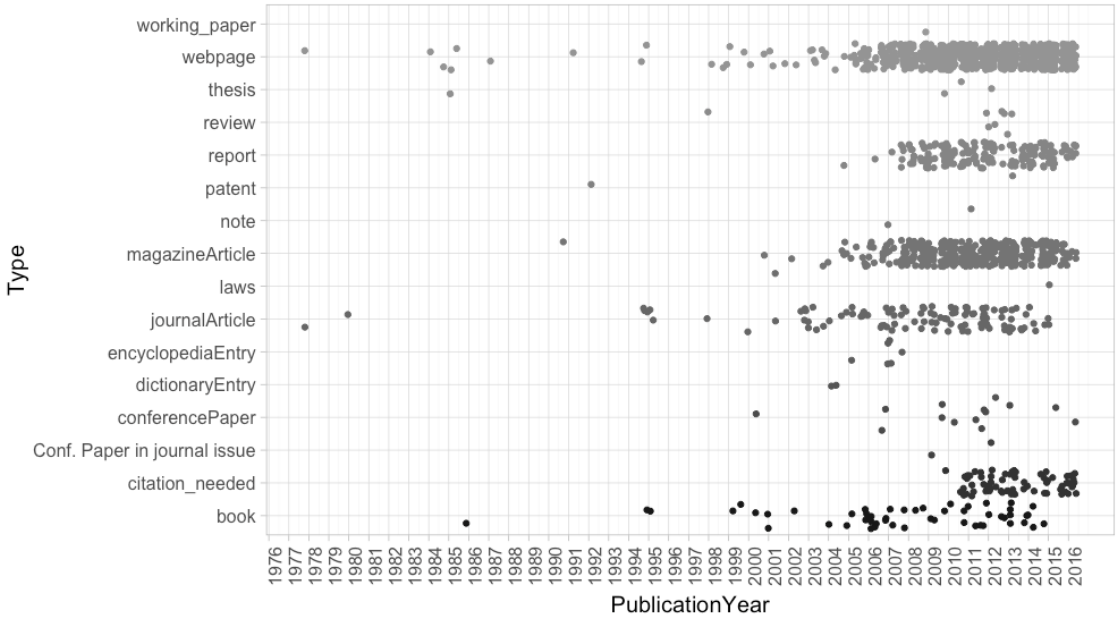

Fig. 2. Distribution of wiki references over publication year and document type from data set B (N=2387 wiki references).

The few journal articles are much more spread out, like books from 2000. This pattern demonstrates that the Wikipedia in Wind Power presents up to date references in the topical entries. Figure 3 shows the general distribution of wiki reference publication years, regardless document type versus Wikipedia entry publication year. The immediacy characteristics of the distribution is very clear for the topic 'Wind Power'.

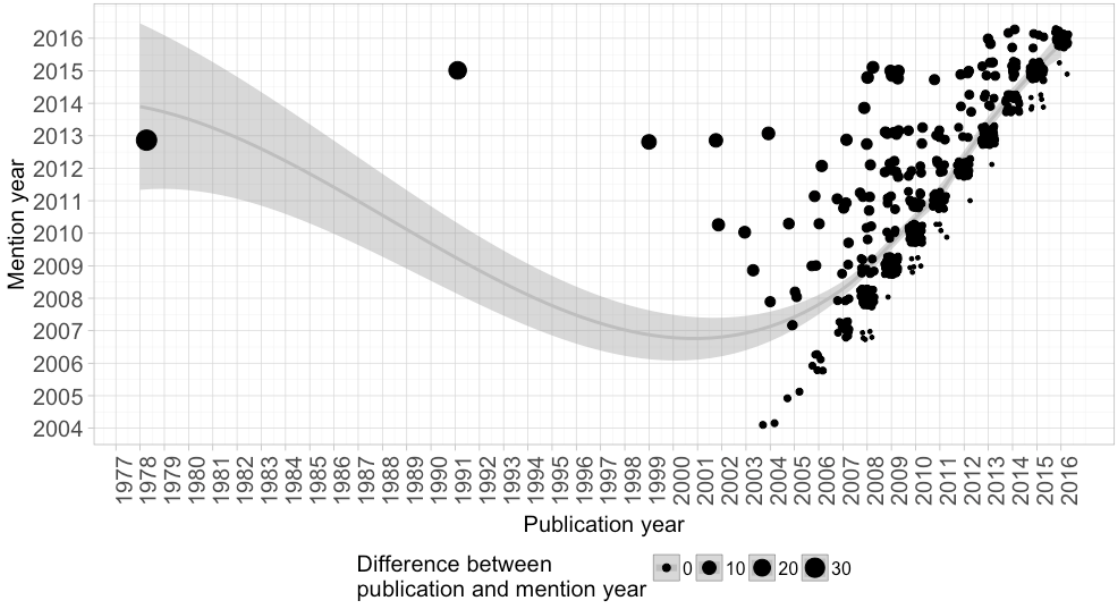

Fig. 3. Distribution of wiki reference publications years over Wikipedia entry publication years (mention years) from data set B (N=2387 wiki references).

Citations as indirect measure of Wikipedia entry impact

Instead of applying the number of mentions of the original WoS records (258) in data set A (or 64 in set B) we propose to use the citations to those records as an *indirect indicator of impact* guided by the Wikipedia entries holding those Wos Records.

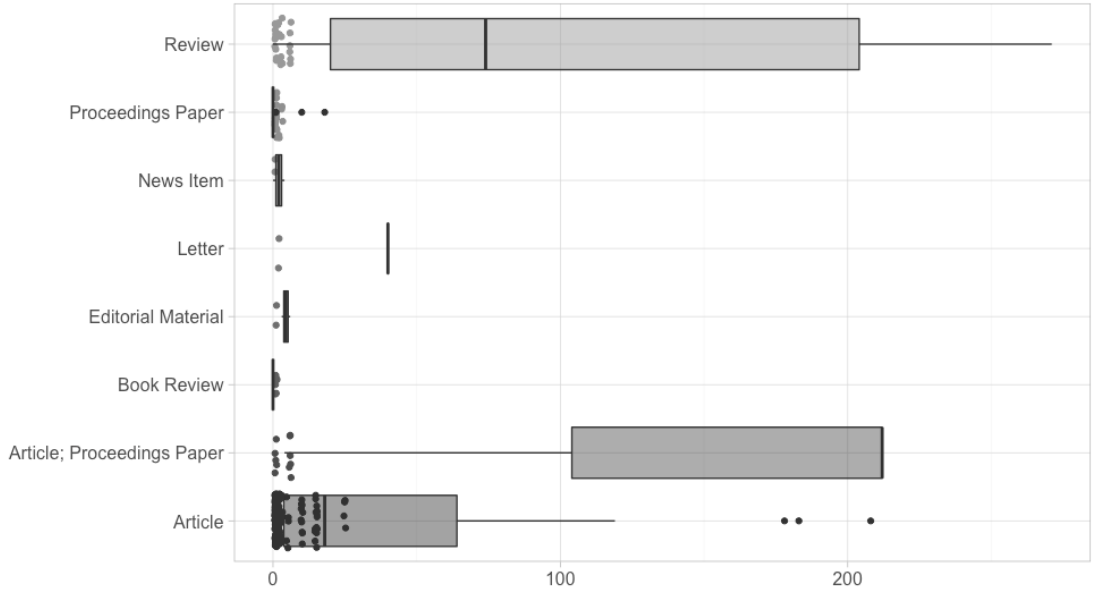

Fig. 4. WoS citations (boxplots) vs Wikipedia mentions of WoS records (dots ; N=258) in 159 Wikipedia entries by WoS document types, data set A.

To understand the relationship between WoS citations and Wikipedia mentions of WoS records additional analyses were made. The comparison of WoS citations given to WoS records on Wind Power in data set A, and Wikipedia mentions of WoS publications in that set shows a distinct kind of distribution (Figure 4). Most mentioned documents (dots) receive few citations, while more highly cited documents only get one mention in the Wikipedia. The Spearman correlation coefficient shows similarly a very low correlation between both variables (0.06). The main difference occurs on review (articles) and proceeding papers published as articles in journals on Wind Power. They are highly cited in WoS but scarcely mentioned in Wikipedia.

Table 4. Citation average associated with the presence in Wikipedia, data set A.

	Mentioned in Wikipedia	**Not mentioned in Wikipedia**
Average of citations to WoS records per wiki entry, data set A	34.28 CITATIONS	5.89 CITATIONS

Nevertheless, Table 4 demonstrates that WoS documents on Wind Power mentioned in the Wikipedia obtain proportionally far more citations in WoS than documents not mentioned in the Wikipedia (5.8 times more). This might seem to be contradictory to the distribution of WoS publications in the Wikipedia, but, on the other hand, only 30 of the WoS records with at

least one mention in the Wikipedia have no citations in WoS. This could mean that documents mentioned in the Wikipedia obtain a higher visibility and sharing and thus receives more citations or, on the other hand, documents cited in WoS have a higher probability of being mentioned in Wikipedia, as shown by the t test, with a p-value lower than 8.137e-11. The latter explanation echoes the findings by Teplitskiy, Lu & Duede (2016) and the immediacy pattern, Figure 3. Citations to scholarly publications mentioned as wiki references (on the topic Wind Power) are thus feasible as basis for indirect research evaluation indicators using the Wikipedia.

Conclusions

Along this work some phenomena were observed relating to the flow of scientific information from the scientific sources to the Wikipedia:

The main conclusion is that the share of scholarly papers which are mentioned in Wikipedia entry references on Wind Power is not substantial (25.4% (set A); 10.4% (set B)). And if we look just at the mention of the 25,540 original WoS records on Wind Power only 0.66% is found in the Wikipedia (set A) and even less (0.26%) in set B, according to the present study. The Wikipedia entry authors in general prefer non-scholarly sources (like magazines, news, web pages, etc.) and non-peer reviewed research reports to support factual statements in the entries. However, observations from the other studies of the Wikipedia with respect to scholarly impact mentioned above strongly indicate that the subject area is a determining factor in the distribution of types of wiki references. In Drug related wiki entries the journal article is the dominant reference type (Koppen, Phillips & Papageorgiou, 2015, p. 142) whereas in History wiki entries (Luyt & Tan, 2010, p. 717-718), not surpricing books is the dominant type (34.4% of all wiki references) followed by news site items (11.9%). Only 2% are journal articles. In Teplitskiy, Lu & Duede (2016) this share is even below 1%. For humanistic fields Kousha & Thelwall (2016, p. 771) also point to books as an important wiki reference type. Due to this strong variation we must conclude that *no generalization* is possible about the Wikipedia reference distribution from findings based alone on a subject area analysis.

In addition, we *cannot* recommend the use Wikipedia for *direct* evaluation of scholarly work, e.g. by observing the amount of mentions of research publications or names detected in Wikipedia entries. This negative outcome of our analyses is also echoed by Kousha & Thelwall for most non-humanistic subject areas in their large scale analysis (2016, p. 770). Further, our analysis indicates that the method applied to isolate or extract wiki entries on a topic (here Wind Power) does influence the overall distribution of wiki reference types (Table 3). Our study is the only one that has addressed this issue.

On the other hand, the substantial amount of non-scholarly publications found on the wiki reference lists in all Wikipedia studies makes it possible to apply such publication types as *indirect* research evaluation measures that supplement traditional scientometric indicators. The three predominant non-scholarly types of wiki references: Web Pages; Popular Magazines/News; and Research Report constitute wiki reference foci of 0.70 and 0.85, respectively for sets A and B. These types illustrate indirect indicators in two ways: 1) each concurrent *individual* type signifies a specific replacement of (or supplement to) scholarly information (WoS records) by popular scientific magazines/news or factual information in the form of non-scholarly web pages and reports from institutions; 2) the three types *together*

constitute a non-scholarly, non-peer reviewed *profile of information visibility* in society through the Wikipedia *associated* with the original academic WoS publications. This leads to four indirect indicators of influence on readers through the Wikipedia references:

- *Popular science knowledge export/influence,* i.e. the share of Magazine/News articles.
- *Non-scholarly knowledge export/influence*, i.e. the share of all non-scholarly references.
- *Academic knowledge export/influence*, i.e. the share of all scholarly references.
- *Wiki reference focus*, i.e. the density of a publication type (or group of types) normalized by the overall average density of the set to which the type(s) belong.

These indirect indicators and the Density metric can indeed be regarded as supplementary to common Scientometric indicators, since they are constituted by publication references captured from a specific database (Wikipedia on the Web).

The same four indicators as well as Density are *direct measures* of the Wikipedia itself.

Indirectly, citations may play a role in the Wikipedia assessment by being citations to 1) the Wikipedia entries themselves or 2) to individual wiki entry references. As concerns the latter kind our analysis strongly indicates that the WoS papers mentioned in Wikipedia on Wind Power do obtain proportionally (5.8 times) more citations in Web of Science than papers on Wind Power not mentioned in the Wikipedia. Highly cited publications seem to be preferred by wiki authors in Wind Power, in line with findings by Teplitskiy, Lu & Duede (2016). We did not investigate the former kind of citations to entire wiki entries.

The study presented here is limited to the topic Wind Power. Like for other topical investigations the findings are not of general nature regarding the Wikipedia and scholarly publications. Further investigations of the Wikipedia in other topical areas may not contribute more generalized knowledge, but may contribute to emphasize the *variation* of wiki reference types in play. We have studied two overlapping but differently constructed data sets. We are aware that the smaller and topically focused set (B) is the most likely to be used by evaluators, since it is easy to operate in the Wikipedia and is directly centered on the topic in question. The larger and broader set (A) is much more cumbersome to capture since it involves the application of a second scientific information system (Web of Science or Scopus or a domain-specific database) *and* a software to match the retrieved records with wiki references. However, it is only through the data set A methodology that one may target an evaluation to, for instance, university or departmental research output, or to specific author profiles. Data set B may only satisfy topical investigations in the Wikipedia.

Based on the findings above and findings by Kousha & Thelwall (2016) and others, we can say that the Wikipedia, English version, is not a very comprehensive dissemination channel for scholarly information, because not many scientific papers (e.g. on Wind Power) reach the Wikipedia. As research evaluation tool the Wikipedia is consequently insufficient. But it is a useful channel, because those few papers that occur in Wikipedia co-occur with other relevant non-scholarly publication types, which help transform scholarly information into more common knowledge.

Acknowledgments: This research was funded by the Spanish Ministry of Economy and Competitiveness under the project CSO 2014-51916-C2-1-R. Titled “La investigacion en eficiencia energetica y transporte sostenibles en el medio urbano: analisis del desarrollo

cientifico y la percepcion social del tema desde la perspectiva de los estudios metricos de la información" "(Research on energy efficiency and sustainable transport in the urban environment: analysis of the scientific development and the social perception of the topic from the perspective of the metric studies of information)".

APPENDIX I: Wind Power search strategy in Web of Science

TS=("wind power" OR "wind turbine*" OR "wind energy*" OR "wind farm*" OR "wind generation" OR "wind systems")

Refined by: [excluding] Web of Science Categories=(ASTRONOMY ASTROPHYSICS)

Databases=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH

Time window: 2006-2015

Lemmatization=On

APPENDIX II: Wind Power entries from Wikipedia collected in Data set B

No.	Wikipedia Entry	No.	Wikipedia Entry
1	Airborne wind turbine	35	Wind power in India
2	Community wind energy	36	Wind power in Iran
3	Darrieus wind turbine	37	Wind power in Italy
4	Environmental effects of wind power	38	Wind power in Japan
5	Floating wind turbine	39	Wind power in Kenya
6	History of wind power	40	Wind power in Lithuania
7	List of offshore wind farms	41	Wind power in Malta
8	List of onshore wind farms	42	Wind power in Mexico
9	List of wind turbine manufacturers	43	Wind power in Morocco
10	Savonius wind turbine	44	Wind power in New Zealand
11	Small wind turbine	45	Wind power in Pakistan
12	Unconventional wind turbines	46	Wind power in Poland
13	Vertical axis wind turbine	47	Wind power in Portugal
14	Wind energy software	48	Wind power in Romania
15	Wind farm management	49	Wind power in Russia
16	Wind farms	50	Wind power in Serbia
17	Wind power	51	Wind power in Spain
18	Wind power consulting companies	52	Wind power in Sweden
19	Wind power forecasting	53	Wind power in Thailand
20	Wind power in Asia	54	Wind power in the European Union
21	Wind power in Australia	55	Wind power in the Netherlands
22	Wind power in Austria	56	Wind power in the Philippines
23	Wind power in Belgium	57	Wind power in the Republic of Ireland
24	Wind power in Brazil	58	Wind power in the United Kingdom
25	Wind power in Canada	59	Wind power in the United States
26	Wind power in China	60	Wind power in Turkey
27	Wind power in Croatia	61	Wind power industry
28	Wind power in Denmark	62	Wind resource assessment
29	Wind power in Estonia	63	Wind turbine aerodynamics
30	Wind power in Finland	64	Wind turbine design
31	Wind power in France	65	Wind turbines
32	Wind power in Germany	66	Wind-powered land vehicle
33	Wind power in Greece	67	Windmill
34	Wind power in Hungary		

References

Allen, H. G., Stanton, T. R., Di Pietro, F., & Moseley, G. L. (2013). Social media release increases dissemination of original articles in the clinical pain sciences. *PLoS One*, 8(7), e68914.

Björneborn, L. & Ingwersen, P. (2004). Towards a Basic Framework for Webometrics**.** *Journal of American Society for Information Science and Technology* (JASIST), 55(14): 1216-1227.

Hammarfelt, B. (2014). Using altmetrics for assessing research impact in the humanities. *Scientometrics*, 101(2), 1419-1430.

Henderson, W. (December 10, 2012). Wikipedia Has Figured Out A New Way To Stop Vandals In Their Tracks. Business Insider (http://www.businessinsider.com/pending-changes-safeguard-on-wikipedia-2012-12?r=US&IR=T&IR=T) – retrieved 03-08-2016.

Ingwersen, P., Larsen, B., Garcia-Zorita, J. C., Serrano-López, A. E., & Sanz-Casado, E. (2013, August). Contribution and influence of proceedings papers to citation impact in seven conference and journal-driven sub-fields of energy research 2005–11. In: *Proceedings of the 14th Conference of the International Society of Scientometrics and Informetrics*, Vienna (pp. 418-425).

Ingwersen, P., Larsen, B., Garcia-Zorita, J. C., Serrano-López, A. E., & Sanz-Casado, E. (2014). Influence of proceedings papers on citation impact in seven sub-fields of sustainable energy research 2005–2011. *Scientometrics*, 101(2), 1273-1292.

Koppen, L., Phillips, J.& Papageorgiou, R. (2015). Analysis of reference sources used in drug-related Wikipedia articles. *Journal of the Medical Library Association*, 103(3), 140-144.

Kousha, K. & Thelwall, M. (2017). Are Wikipedia citations important evidence of the impact of scholarly articles and books? *Journal of the Association for Information Science and Technology*, 68(3):762–779.

Luyt, B. & Tan, D. (2010). Improving Wikipedia's Credibility: References and Citations in a Sample of History Articles. *Journal of the Association for Information Science and Technology*, 61(4), 715-722.

Park, T. K. (2011). The visibility of Wikipedia in scholarly publications. *First Monday*, 16(8).

Sanz-Casado, E., Garcia-Zorita, J. C., Serrano-López, A. E., Larsen, B., & Ingwersen, P. (2013). Renewable energy research 1995–2009: a case study of wind power research in EU, Spain, Germany and Denmark. *Scientometrics*, 95(1), 197-224.

Stein, K., & Hess, C. (2007, September). Does it matter who contributes: a study on featured articles in the German wikipedia. In: *Proceedings of the eighteenth conference on Hypertext and hypermedia* (pp. 171-174). ACM.

Tananbaum, G. (2013) Article-Level Metrics: A SPARC Primer. Retrieved from http://sparc.arl.org/sites/default/files/sparc-alm-primer.pdf

Teplitskiy, M., Lu, G. & Duede, E. (2016). Amplifying the Impact of Open Access: Wikipedia and the Diffusion of Science. *Journal of the Association for Information Science and Technology*, 12 p. DOI: 10.1002/asi.23687.

Thelwall, M. (2016). *Web Indicators for Research Evaluation: A Practical Guide*. Morgan & Claypool Publ. doi:10.2200/S00733ED1V01Y201609ICR052. (Synthesis Lectures on Information Concepts, Retrieval, and Services).

Wikipedia (July 31, 2016). From (https://en.wikipedia.org/wiki/Wikipedia#History) – retrieved 03-08-2016.