On Measuring the Publication Productivity and Citation Impact of a Scholar: A Case Study

Tefko Saracevic[1] & Eugene Garfield[2]

[1] Rutgers University, New Brunswick, USA
[2] ThomsonReuters Scientific (formerly ISI), Philadelphia, USA

Abstract. The purpose is to provide quantitative evidence of scholarly productivity and impact of Peter Ingwersen, a preeminent information science scholar, and at the same time illustrate and discuss problems and disparities in measuring scholarly contribution in general. Data is derived from searching Dialog, Web of Science, Scopus, and Google Scholar (using Publish or Perish software). In addition, a HistCite profile for Peter Ingwersen publications and citations was generated.

Keywords: Scholarly productivity; citation impact; quantitative measures.

Introduction

The paper is honoring the scholarly contribution of Peter Ingwersen, a scholar extraordinaire in information science. With his ideas, publications, presentations, and collaborations Professor Ingwersen attained a global reach and impact. The purpose here is to provide some numerical evidence of his productivity and impact with a further objective of using this data as a case study to illustrate and discuss the problems, difficulties and disparities in measuring scholarly contributions in general.

The essence of scholarship is proposition of ideas or explanation of phenomena in concert, at some time or another, with their verification. Since antiquity to the present day these were represented in publications – books, treatises, journal articles, proceedings papers etc. – in a variety of forms. Traditionally, their quality was assessed by peer review and recognition, critical examination, and verification of claims. The impact was the breadth and depth of these assessments and even more so their effects on scholarship that followed. Scholarly productivity and impact was a qualitative assessment.

In contrast, close to a century ago quantitative metrics associated with scholarly publications started to appear. Counting various aspects provided a further picture of productivity and impact. At first they were numbers such as publications per author, numbers of references and citations, and other indicators. Bibliometrics emerged in the mid of last century as an area of study of quantitative features and laws of re-

corded information discourse. Finally, a decade or so thereafter scientometrics focused on the scientific measurement of the work of scientists, especially by way of analyzing their publications and the citations within them – it is application of mathematical and statistical methods to study of scientific literature. Scholarly productivity and impact was also quantified.

Contemporary advances in information and communication technologies enabled innovative creation of large databases incorporating publication and citation data from which, among others, a variety of metrics are derived. Scholarly productivity and impact is being derived quantitatively from massive databases. Results are often used for a variety of evaluative purposes.

Thus, a distinction is made between relational bibliometrics/scientometrics, measuring (among others) productivity and evaluative bibliometrics/scientometrics measuring impact. In this paper we deal with both,

2 Problems, issues

A number of databases now provide capabilities to obtain comprehensive metrics related to publications of individual scholars, disciplines, journals, institutions and even countries. As to statistics related to publications, i.e. relational bibliometrics, they provide straight forward relational data. But as to impact, i.e. evaluative bibliometrics, they also compute a variety of citation-related measures or metrics. In other words, citations are at the base of evaluative bibliometrics. Three issues follow.

The first issue is about the very use of citations for impact studies. Numerous caveats are expressed questioning such use and warning of possible misuse. Leydesdorff [1] is but one of numerous articles addressing the problem. While fully recognizing the caveats and this problem we will not deal with them. Let it be said that such caveats should be applied to data presented here as well.

The second issue is operational and relates to the quality of citations from which evaluative data is derived. Citations are not necessarily "clean" data; ambiguities, mistakes, inaccuracies, inabilities to differentiate, and the like are present at times. Citation hygiene differs. White [2] is but one of numerous articles that discusses possible ambiguities in presentation and use of citation data. Again, while recognizing this issue and problem we will not deal with it here.

The third issue, the one that we will deal with here, is also operational, but relates to coverage and treatment of sources from which publication and impact metrics are derived. *Science Citation Index* appeared in 1963, compiled by the Institute for Scientific Information (ISI), followed a few years later by *Social Science Citation Index* and then by *Arts & Humanities Citation Index*. Using and enlarging on these indexes, in 1997 ISI, (now part of Thomson Reuters) released the *Web of*

Science (WoS) [3]. For four decades, - from 1960s till 2004 – these indexes, including WoS, were the sole source for citation studies and impact data. Thus, for a long while life for deriving and using such data was simple and unambiguous.

In 1972 the Lockheed Missiles and Space Company launched Dialog as a commercial search services, incorporating a number of indexing and abstracting databases for standardized access and searching. [4]. (After several owners, Dialog is now a part of ProQuest). Dialog became by far the largest and most diversified "supermarket" of databases available for searching. Among others, Dialog offered and is still offering ISA citation indexes for citation searches and analyses.

In 2004 Elsevier launched Scopus, a large indexing and abstracting database. At first Scopus covered science, engineering, medicine, and social sciences and later included humanities as well. But from the start, Scopus incorporated citation analyses of various kinds, including impact data. WoS and Scopus provide similar kind of citation analytic capabilities [5]. Suddenly, life was not simple any more. Two different sources for citation analyses became available.

In 2005 Google launched Google Scholar, with the goal to cover scholarly literature. The coverage is broad. As to citations, a "cited by" link is provided but citation analysis can not be done directly. Independently, enters Anne-Wil Harzing, a professor at the University of Melbourne, Australia, and in 2006 releases Publish or Perish (PoP), a free tool or app for deriving various citation analyses, including impact data, from Google Scholar [6]. With three large databases available for citation analyses and impact metrics life got really complicated.

Soon after appearance of Scopus and then Google Scholar a number of papers compared features of these two with WoS (e.g. [7]). But the more interesting question was not comparison of features, but of results. The issue is: How do citation results from these three giant databases compare? For instance, do publication data or impact metrics differ? If so, why and by how much? E.g. If we search for citation and impact data for an author – in this case Peter Ingwersen – are results from the three databases close? Or not?

Not surprisingly, a number of studies were launched trying to answer these questions, i.e. comparing results of citation searches from the three databases. A cottage industry developed addressing the issues and problems. This paper is one of them. Here is but a sample of more recent studies from various fields comparing citation results from WoS, Scopus, and Google Scholar (GS).

Meho and Yang compared ranking of 25 top scholars in library and information science and found that "Scopus significantly alters the relative ranking of those scholars that appear in the middle of the rankings and that GS stands out in its coverage of conference proceedings as well as international, non-English language journals...[and that] WoS, helps reveal a more accurate and comprehensive picture of the scholarly impact of authors."[8].

Kulkarni, et al. compared the citation count profiles of articles published in general medical journals and found that "Web of Science, Scopus, and Google Scholar produced quantitatively and qualitatively different citation counts for articles published in 3 general medical journals." [9].

Bar-Ilan compared citations to the book "Introduction to Informetrics" from the three databases and found that "Scopus citations are comparable to Web of Science citations ... each database covered about 90% of the citations located by the other. Google Scholar missed about 30% of the citations covered by Scopus and Web of Science (90 citations), but another 108 citations located by Google Scholar were not covered either by Scopus or by Web of Science." [10].

Taking it all together: there were differences in results from the three databases, but the magnitude differs from study to study and field to field.

3 Method

Four databases, - Dialog, Web of Science (WoS), Scopus, and Google Scholar (GS) (using Publish or Perish (PoP) software) - were searched for author "Ingwersen P" or "Ingwersen Peter" to identify:
- number of publications,
- number of citations including self-citations,
- number of citations excluding self-citations,
- the h-index,
- papers with highest citation rate, and.
- number of collaborators.

In addition, analysis of Ingwersen publications and citations was done using Hist-Cite, described below.

In **Dialog** the following four files were searched: Social SciSearch (file 7), SciSearch 1990 - (file 34), SciSearch 1974-1989 (file 434), and Arts and Humanities Search (file 439). These files are incorporated in WoS, but their organization and searching in Dialog is very, very different.

WoS was searched using the version available through Rutgers University Libraries – subscription in this version is restricted to WoS data from 1984 to present. Thus, this is a **partial WoS**, but it does contain most Ingwersen publications and citations that appeared in WoS covered journals, since Ingwersen started publishing in 1980.

Scopus was searched in its entirety. Scopus covers journals and other sources that substantially overlap with those in WoS, but also covers some additional ones.

PoP was used to extract data from Google Scholar. GS covers many types and sources of publications but it is not transparent what the coverage is as to sources or time period [7].

HistCite, developed by Eugene Garfield, is a software package that provides a variety of bibliometric analyses and mappings from data in WoS [11]. Input is generated form whole WoS but it also allows input of publications not in WoS (e.g. books, proceeding papers) to search for their citations. Here, the input (collection) for HistCite included: (a) papers by "P Ingwersen" downloaded from whole WoS; (b) papers that contained the cited author "P Ingwersen" also downloaded from WoS; **plus** (c) selected papers not in WoS from an Ingwersen bibliography of 126 publications supplied by Birger Larsen, Royal School of Library and Information Science, Denmark. In other words, papers from that bibliography not in WoS were added to HistCite collection.

All searches were done in the second week of May 2010.

4 Results

This section provides results from searches and analyses in a tabular form. The next section, Discussion, provides interpretation of these results linked to each table. In other words, results are presented all together in one section and discussion again all together in another one. In this way, a reader can look at the results alone and draw own interpretations, and then follow our discussion.

4.1 Publications, citations, h-index

Basic results related to Peter Ingwersen's publications, citations and h-index are presented in Table 1.

Database	No. of publications by P. Ingwersen	Total citations with self-citations	Total citations without self-citations	h-index
Dialog	53	902	859	NA
Scopus	55	1208	1123	14
Web of Science 1984-present	52	1101	663	16
Google Scholar	279	4639	NA	27
HistCite	85	1850	1696	20

Table 1. No. of publications, citations, and h-index for Peter Ingwersen from Dialog, WoS (1984-date), Scopus, Google Scholar (using PoP) and HistCite.

4.2 Time span of publications and citations

Table 2 shows the number of publications per year by Ingwersen from 1984 to 2009. Table 3 shows the number of citations received by Ingwersen's papers per year from years 1984 to 2009. Both are derived by WoS (1984-present).
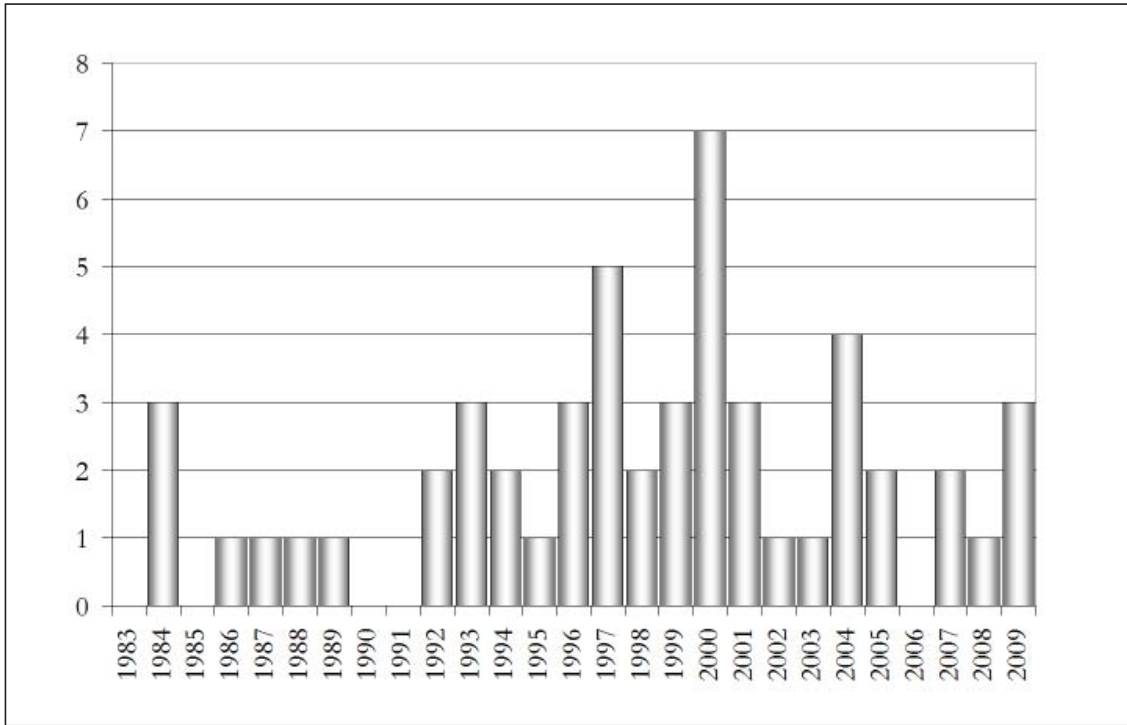
Table 2. WoS (1984-present): No. of publications by Peter Ingwersen published over the years.
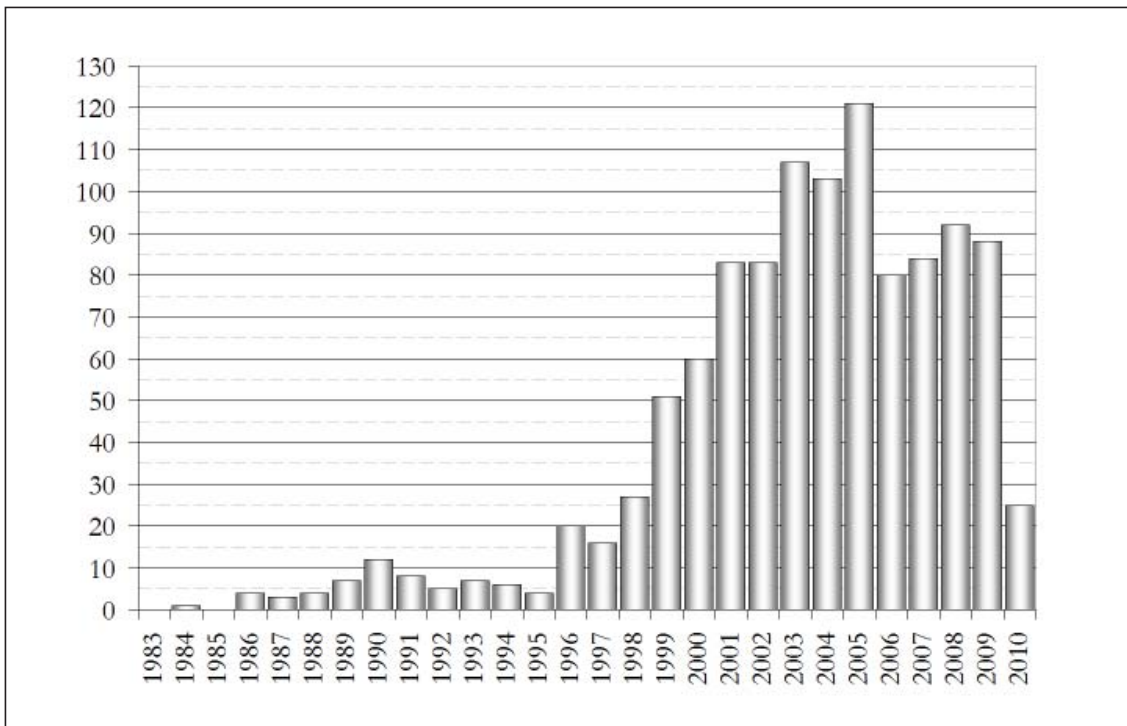

Table 3. WoS (1984-present): No. of citations to Peter Ingwersen's papers over the years.

Table 4. WoS (1984-present): List of Ingwersen's co-authors. In this collection Ingwersen has 52 papers with 47 different co-authors (although Willett and Willet are the same author); as example, he co-authored 9 papers with Larsen.

Authors (Refine) (Exclude) (Cancel)   Sort these by: Record Count ▾

The first 100 Authors (by record count) are shown. For advanced refine options, use ▤ Analyze results .

☐ INGWERSEN, P (52)	☐ BORGMAN, CL (1)	☐ FUHR, N (1)	☐ RUSSELL, J (1)
☐ LARSEN, B (9)	☐ BORLUND, P (1)	☐ HARPER, D (1)	☐ SCHNEIDER, JW (1)
☐ WORMELL, I (7)	☐ BROOKS, HM (1)	☐ HYLDEGARD, J (1)	☐ SEIDEN, P (1)
☐ BJORNEBORN, L (3)	☐ BYLANDER, T (1)	☐ JEPSEN, ET (1)	☐ SKOV, M (1)
☐ CHRISTENSEN, FH (3)	☐ CHAVAN, VS (1)	☐ KEEN, M (1)	☐ SKRAM, U (1)
☐ JARVELIN, K (3)	☐ CLEVERDON, C (1)	☐ KUHLEN, R (1)	☐ SMEATON, A (1)
☐ JACOBS, D (2)	☐ COSIJN, E (1)	☐ LUND, B (1)	☐ THOMPSON, R (1)
☐ SPARCK-JONES, K (2)	☐ CROFT, WB (1)	☐ NIEMI, T (1)	☐ VANRIJSBERGEN, K (1)
☐ AGOSTI, M (1)	☐ DANIELS, P (1)	☐ NOYONS, E (1)	☐ VIBY-MOGENSEN, J (1)
☐ ALMIND, TC (1)	☐ DEERWESTER, S (1)	☐ RADA, R (1)	☐ WALKER, D (1)
☐ BEAULIEU, M (1)	☐ FOX, EA (1)	☐ ROBERTSON, S (1)	☐ WILLET, P (1)
☐ BELKIN, NJ (1)	☐ FREI, HP (1)	☐ ROUSSEAU, R (1)	☐ WILLETT, P (1)

Author Name

☐ Ingwersen, P. (55)	—
☐ Larsen, B. (13)	☐ Viby-Mogensen, J. (1)
☐ Wormell, I. (4)	☐ Walker, D. (1)
☐ Jarvelin, K. (3)	☐ Willett, P. (1)
☐ Skov, M. (3)	☐ Woods, C. (1)
☐ Jorgensen, H.L. (3)	☐ Zijlema, A.F. (1)
☐ Borlund, P. (3)	☐ Almind, T.C. (1)
☐ Bjorneborn, L. (3)	☐ Belkin, N.J. (1)
☐ Boyes, J.D. (2)	☐ Borgman, C.L. (1)
☐ Sudan, R.N. (2)	☐ Brooks, H.M. (1)
☐ Smith, D.L. (2)	☐ Bylander, T. (1)
☐ Rehfeld, J.F. (2)	☐ Chavan, V.S. (1)
☐ Bennett, L.F. (2)	☐ Christensen, F.H. (1)
☐ Greenly, J.B. (2)	☐ Cosijn, E. (1)
☐ Anderson, D.E. (2)	☐ Crofts, W.B. (1)
☐ Hjortgaard Christensen, F. (2)	☐ Daniels, P. (1)
☐ Ingwersen, P. (2)	☐ Deerwester, S. (1)
☐ Prætorius, L. (1)	☐ Fox, E.A. (1)
☐ Papaeconomou, C. (1)	☐ Hammer, D.A. (1)
☐ Rada, R. (1)	☐ Jacobs, D. (1)
☐ Rousseau, R. (1)	☐ Jepsen, E.T. (1)
☐ Noyons, E. (1)	☐ Jones, K.S. (1)
☐ Russell, J. (1)	☐ Kekalainen, J. (1)
☐ Schneider, J.W. (1)	☐ Lund, B. (1)
☐ Seiden, P. (1)	☐ Lynge, E. (1)
☐ Skram, U. (1)	☐ McAlpine, G. (1)
☐ Thompson, R. (1)	☐ Niemi, T. (1)

Table 5. Scopus: List of Ingwersen's co-authors In this collection Ingwersen has 55 papers with 52 different co-authors; he co-authored 13 papers with Larsen.

Publications by P Ingwersen
All-Author List (62)

Records: 85, Authors: 62, Journals: 45, Cited Re...
Yearly output | Document Type | Language | Ir...

#	Author	Recs	TLCS	TGCS
1	Ingwersen P	85	146	1752
2	Larsen B	16	10	64
3	Wormell I	9	16	74
4	Jarvelin K	5	8	125
5	Bjorneborn L	4	4	131
6	Christensen FH	3	18	82
7	Schneider JW	3	1	3
8	Borlund P	2	2	85
9	Jacobs D	2	2	24
10	Jepsen ET	2	0	9
11	Seiden P	2	0	9
12	Skov M	2	1	3
13	Sparck-Jones K	2	1	31

*Table 6. **HistCite:** List of Ingwersen's co-authors up to co-authorship of two papers. In this collection Ingwersen has 85 papers with 62 different co-authors; he co-authored 16 papers with Larsen. Recs = number of records; TLCS = Total Local Citation Score, shows the count of cited papers within the collection; TGCS = Total Global Citation Score, shows the Citation Frequency based on the total count in the Web of Science.*

4.3 Co-authors

In doing research and publishing papers Ingwersen collaborated with a number of scholars. List of Ingwersen's co-authors as listed in WoS (1984-present) are shown in Table 4, as listed in Scopus in Table 5, and as listed in HistCite in Table 6; this table shows co-authors who published 2 or more papers with Ingwersen; single co-authorship list is not shown, because it is too long.

4.4 Highest cited papers

Five highest cited papers by Ingwersen as listed in WoS are shown in Table 7, in Scopus in Table 8, and in HistCite in Table 9.

Table 7. WoS (1984-present): Five highest cited papers by Ingwersen with number of citations for each.

4.5 HistCite

As mentioned, HistCite produces a variety of analyses and mappings using WoS data, but allows input of publications that are not necessarily in WoS, as was the

Document (sort by relevance)	Author(s)	Date	Source Title	Cited By
1. Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory Abstract + Refs | View at Publisher	Ingwersen, P.	1996	*Journal of Documentation* 52 (1), pp. 3-50	213
2. The calculation of web impact factors Abstract + Refs | View at Publisher | Show Abstract	Ingwersen, P.	1998	*Journal of Documentation* 54 (2), pp. 236-243	169
3. Informetric analyses on the world wide web: Methodological approaches to 'webometrics' Abstract + Refs | View at Publisher | Show Abstract	Almind, T.C., Ingwersen, P.	1997	*Journal of Documentation* 53 (4), pp. 404-426	139
4. Perspectives of webometrics Abstract + Refs | View at Publisher | Show Abstract	Björneborn, L., Ingwersen, P.	2001	*Scientometrics* 50 (1), pp. 65-82	85
5. The development of a method for the evaluation of interactive information retrieval systems Abstract + Refs | View at Publisher | Show Abstract	Borlund, P., Ingwersen, P.	1997	*Journal of Documentation* 53 (3), pp. 225-250	76

Table 8. Scopus: Five highest cited papers by Ingwersen with number of citations for each.

#		GCS
1	16 INGWERSEN P **Information Retrieval Interaction** INFORMATION RETRIEVA. 1992; : 1-246	269
2	32 Ingwersen P **Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory** JOURNAL OF DOCUMENTATION. 1996 MAR; 52 (1): 3-50	236
3	40 Ingwersen P **The calculation of Web impact factors** JOURNAL OF DOCUMENTATION. 1998 MAR; 54 (2): 236-243	177
4	36 Almind TC, Ingwersen P **Informetric analyses on the World Wide Web: Methodological approaches to 'webometrics'** JOURNAL OF DOCUMENTATION. 1997 SEP; 53 (4): 404-426	139
5	71 INGWERSEN P, JARVELIN K **The turn: integration of information seeking and retrieval in context.** TURN INTEGRATION INF. 2005;	114

Table 9. HistCite: Five highest cited papers with number of citations for each. GSC= Global Citation Score, shows the Citation Frequency based on the total count in the Web of Science.

case here where selected papers from Larsen's bibliography for Ingwersen that were not in WoS were added.[1] Only a sample of HistCite data is presented here; full array of data can be accessed as follows:

Publications by Peter Ingwersen are available at

http://garfield.library.upenn.edu/histcomp/ingwersen-p_auth/index-tl.html

Papers citing Peter Ingwersen are available at

http://garfield.library.upenn.edu/histcomp/ingwersen-p_citing/index-tl.html

---

1   HistCite data presented here is derived from data available online at mentioned sites and is on par with a static report. If one uses the actual HistCite software (available for a free trial at http://www.histcite.com/), the experience is different as more information is available and there are numerous ways to edit and define the collection to ascertain a variety of different statistics. Coupled with the ability to export to Excel, there are many different ways to use data through HistCite software.

Records: 85, Authors: 62, Journals: 45, Cited References: 1193, Words: 274
Yearly output | Document Type | Language | Institution | Institution with Subdivision | Country

#	Date / Author / Journal	LCS	GCS	LCR	CR
	1982				
1	INGWERSEN P SEARCH PROCEDURES IN THE LIBRARY - ANALYZED FROM THE COGNITIVE POINT OF VIEW JOURNAL OF DOCUMENTATION. 1982; 38 (3): 165-191	12	93	0	42
	1983				
2	INGWERSEN P INFORMATION IN ITALY JOURNAL OF INFORMATION SCIENCE. 1983; 6 (2-3): 91-94	1	1	0	0
	1984				
3	INGWERSEN P A COGNITIVE VIEW OF 3 SELECTED ONLINE SEARCH FACILITIES ONLINE REVIEW. 1984; 8 (5): 465-492	6	32	1	35
4	INGWERSEN P PSYCHOLOGICAL-ASPECTS OF INFORMATION-RETRIEVAL SOCIAL SCIENCE INFORMATION STUDIES. 1984; 4 (2-3): 83-95	1	18	1	27
5	INGWERSEN P INFORMATION TECHNOLOGY - WHICH APPLICATIONS SOCIAL SCIENCE INFORMATION STUDIES. 1984; 4 (2-3): 185-196	0	0	1	20
	1986				
6	INGWERSEN P, WORMELL I Improved subject access, browsing and scanning mechanisms in modern on-line IR 1986 ACM SIGIR C. 1986; : 68-76	1	10	0	0
7	INGWERSEN P, KAJBERG L, PEJTERSEN AM Information technology and information use : towards a unified view of information and information technology INFORMATION TECHNOLO. 1986;	2	9	0	0
8	INGWERSEN P Cognitive analysis and the role of the intermediary in information retrieval [Intelligent information systems : progress and prospects] INTELLIGENT INFORMAT. 1986; : 206-237	7	17	0	0
9	INGWERSEN P INTERACTION IN INFORMATION-SYSTEMS - A REVIEW OF RESEARCH FROM DOCUMENT-RETRIEVAL TO KNOWLEDGE-BASED SYSTEMS - BELKIN,NJ, VICKERY,A JOURNAL OF DOCUMENTATION. 1986 SEP; 42 (3): 197-200	0	0	0	4
	1987				
10	BELKIN NJ, BORGMAN CL, BROOKS HM, BYLANDER T, CROFT WB, et al. DISTRIBUTED EXPERT-BASED INFORMATION-SYSTEMS - AN INTERDISCIPLINARY APPROACH INFORMATION PROCESSING & MANAGEMENT. 1987; 23 (5): 395-409	1	31	0	34

Table 10. HistCite: Sample from 85 publications by Ingwersen; listed are 11 publications from 1982 to 1987. LCS= Local Citation Score, shows the count of cited papers within the collection; GSC= Global Citation Score, shows the Citation Frequency based on the total count in WoS; LCR= Local Citation Score, shows the Citation Frequency within the collection; CR= Cited References, shows the number of all cited references as given in the paper's bibliography.

1342	Chung WY **Web Searching and Browsing: A Multilingual Perspective** ADVANCES IN COMPUTERS, VOL 78. 2010; 78: 41-69
1343	Craven J, Johnson F, Butters G **The usability and functionality of an online catalogue** ASLIB PROCEEDINGS. 2010; 62 (1): 70-84
1344	Nolin J, Astrom F **Turning weakness into strength: strategies for future LIS** JOURNAL OF DOCUMENTATION. 2010; 66 (1): 7-27
1345	Savolainen R Source preference criteria in the context of everyday projects Relevance judgments made by prospective home buyers JOURNAL OF DOCUMENTATION. 2010; 66 (1): 70-92
1346	Palsdottir A The connection between purposive information seeking and information encountering A study of Icelanders' health and lifestyle information seeking JOURNAL OF DOCUMENTATION. 2010; 66 (2): 224-244
1347	Jowkar A, Didegah F Evaluating Iranian newspapers' web sites using correspondence analysis LIBRARY HI TECH. 2010; 28 (1): 119-130
1348	Guimaraes MCS **Geography of science makes a difference: an appeal for public health** CADERNOS DE SAUDE PUBLICA. 2010 JAN; 26 (1): 50-58
1349	Lee YO, Park HW **The Reconfiguration of E-Campaign Practices in Korea A Case Study of the Presidential Primaries of 2007** INTERNATIONAL SOCIOLOGY. 2010 JAN; 25 (1): 29-53
1350	Fu X **Towards a Model of Implicit Feedback for Web Search** JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY. 2010 JAN; 61 (1): 30-49
1351	Velasco F, Gonzalez-Abril L, Ortega JA, Alvarez JA **A STUDY OF THEMATIC AREAS IN ECONOMY BY A MEASURE OF SIMILARITIES BASED ON A KERNEL** INTERCIENCIA. 2010 MAR; 35 (3): 191-197

Table 11. HistCite: Sample of publications citing Ingwersen; listed are 10 (out of 23) publications that were published in 2010.

Here are excerpts from main results. Table 10 shows publications by Ingwersen from 1982 to 1987 – listed are 11 out of a total of 85 publications in HistCite. In addition to date as shown here, data can be sorted by various parameters indicated in blue. Table 11 shows a sample of 10 papers published in 2010 that cite Ingwersen. Table 12 shows 20 most significant words in tiles of papers by Ingwersen. Table 13 shows an example of a historiograph – a map – generated by HistCite; in this case it shows connections of the listed paper in the center of the map. On the above site, connecting papers can be identified by scrolling over them.

Publications by P Ingwersen

Word(i) List (274) Word count: 585, All words count:

Records: 85, Authors: 62, Journals: 45, Cited Referenc
Yearly output | Document Type | Language | Instituti
Page 1 of 2: [1 2]

#	Word	Recs	TLCS	TGCS
1	INFORMATION	35	66	894
2	RESEARCH	21	21	107
3	RETRIEVAL	17	56	786
4	COGNITIVE	12	53	460
5	IMPACT	11	14	233
6	SCIENCE	10	13	75
7	CITATION	9	17	81
8	ANALYSIS	8	27	90
9	BASED	6	2	45
10	DATA	6	7	62
11	INTERNATIONAL	6	14	57
12	VISIBILITY	6	7	32
13	WORLD	6	7	148
14	CONTEXTS	5	2	7
15	ONLINE	5	29	124
16	POLYREPRESENTATION	5	5	30
17	PUBLICATION	5	12	59
18	SCANDINAVIAN	5	11	38
19	SCIENTIFIC	5	1	19
20	SOCIAL	5	13	40

272 Ingwersen P
The calculation of Web impact factors
JOURNAL OF DOCUMENTATION. 1998 MAR; 54 (2): 236-243
LCR: 2   CR: 7   LCS: 176   GCS: 177

Table 12. HistCite: Top 20 significant words (out of 274) used in titles of papers by Ingwersen. Recs = shows the number of records where the word appears; TLCS= Total Local Citation Score, shows the count of cited papers within the collection; TGSC= Global Citation Score, shows the Citation Frequency based on the total count in WoS.
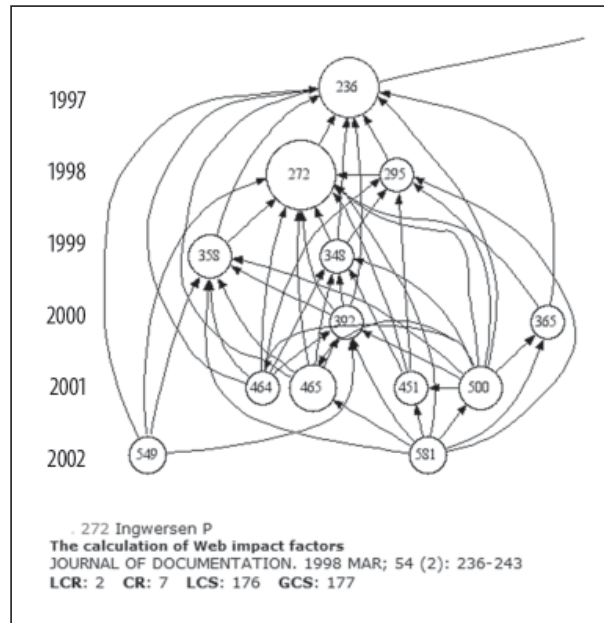
Table 13. HistCite: Map of connections (historiograph) for Ingwersen paper 272 The calculation of Web impact factors to demonstrate mapping. This is a part of a larger map. In the original hisoriograph connecting papers are identified by scrolling over.

5. Discussion

Publications, citations, and h-index (Table 1): Dialog, Scopus, and WoS contained a similar number of papers **by** Peter Ingwersen but produced differing number of citations **to** Ingwersen. (Note that the version of WoS used here is from 1982-present and not the whole WoS). It is not clear how Dialog, supposedly containing the same databases as WoS, produced a lesser number of citations. On the other hand, WoS produced the smallest number of citations without self-citations. Possibly, computing algorithms and rules may differ. The h-indexes were almost identical.

Google Scholar produced by far the largest number of publications by and citations to Ingwersen. This is partially a reflection of a much broader coverage than other databases, particularly in proceedings and non-English publications, and partially because of a lack of quality control e.g. counted are multiple versions of the same paper, ghost citations and the like as enumerated by Jacsó [12]. In general, data from Google Scholar are inflated.

As mentioned, HistCite included papers by Ingwersen from whole WoS plus those not in WoS added from Larsen's bibliography of Ingwersen. Thus, the base collection for searching for citations was larger. This produced probably the most realistic numbers of citations and h-index – simply because more of Ingwersen's publications were used. He has written quite a bit more than what is covered by WoS or Scopus but not as much as indicated by Google Scholar.

Time span of publications and citations (tables 2 and 3). Data here are generated from WoS only, even though a similar display can be obtained from Scopus as well. His highest productivity in publishing papers was the time span 1997-2000. His highest number of citations was in publications that were published in the time span of 2001 to 2009. His impact, as measured by the number of citations, is continuing to this day. In other words, the impact of his publications goes on unabated.

Co-authors (tables 4, 5, and 6). Here we can see significant differences among databases. WoS includes 52 Ingwersen papers with 47 different co-authors. Scopus has 55 papers with 52 co-authors. HistCite has 85 papers with 62 co-authors. Larsen is the highest placed co-author in all three databases, but WoS shows that Larsen co-authored 9 papers with Ingwersen, Scopus 13, and HisCite 16. This may be due to evident difference in coverage, but it could be also that method of processing – policies and/or algorithms may differ.

Highest cited papers (tables 7, 8, and 9). Again, here we can see significant differences. The order of five highest cited papers for WoS and Scopus are the same, but not for HistCite; however, the number of citations that these papers receive differs from one database to the other. And again this may be due to differences in coverage, algorithms, and policies, but with citations this may also be

due to type and intensity of quality control. As mentioned, citation data are not "clean" at all, thus the question is: how effectively are theyt cleaned?

HistCite (tables 10, 11, 12, and 13). Full and rich data from HistCite are provided at listed URLs, thus all the tables presented here offer only a set of static and very limited examples of what is available there. In dynamic HistCite, data can be sorter in numerous ways – date, authors, journals, papers, citations, references, yearly output, and more. This provides for a dynamic interaction with data and even discovery. One such example is analysis of significant words used in titles of papers by Ingwersen. The top 20 words almost describe his oeuvre of interest and research. The maps show connections that can be further explored.

6. Conclusions

The purpose of the paper was to provide quantitative evidence of scholarly productivity and impact of Peter Ingwersen; at the same time another objective was to illustrate and discuss problems and disparities in measuring scholarly contribution in general.

As to the main purpose: Data confirm a long and sustained effort by Ingwersen over the span of some three decades. Moreover, they also confirm several other aspects: the large impact he had on other work and research, the amazing breadth of his collaborations, and the international connections he nurtured.

But as to the second objective of the paper, data also confirm considerable difference in results from one database to another. This was observed in a number of other studies, thus, here is another confirmation. Reasons for such disparity were not investigated here – they could be due primarily to differences in coverage, but also in policies, algorithms, and quality control in handling of data. A number of larger questions related to use or even misuse of such data can also be raised [e.g. 13], but are not discussed here.

Still, with all the caveats and problems we are better off with such data then without.

References

Leydesdorff, L. Caveats For The Use Of Citation Indicators In Research And Journal Evaluations. J. Am. Soc. for Inf. Sc. and Techn., 59 (2), 278–287. (2008)

White, H. D. Citation Analysis. In: M. J. Bates and M. N. Maack (Eds.) Encycl. of Libr. and Inf. Sci. New York: Taylor & Francis. Third Edition, 1: 1, 1012–1026. (2010)

McVeigh, M. E. Citation Indexes and the Web of Science. In: M. J. Bates and M. N. Maack (Eds.) Encycl. of Libr. and Inf. Sci. New York: Taylor & Francis. Third Edition, 1: 1, 1027 — 1037 (2010)

Bellardo-Hahn, T. Pioneers Of The Online Age. Inf. Proc. & Mngmt. 32 (1), 33-48. (1996)

Dess, H.M. Data Base Reviews and Reports: Scopus. Issues in Science and Technology Librarianship. issue 45. (2006) http://www.istl.org/06-winter/databases4.html

Harzing, A.W. Publish or Perish. http://www.harzing.com/pop.htm

Jacsó, P. As we may search – Comparison of major features of the *Web of Science*, *Scopus*, and *Google Scholar* citation-based and citation-enhanced databases. Current Science, 89 (9), 1537-1547. (2005)
http://www.ias.ac.in/currsci/nov102005/1537.pdf

Meho, L.I., Yang, K. Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. J. Am. Soc. for Inf. Sc. and Techn., 58(13), 2105-2125. (2007)

Kulkarni, A. V., Aziz, B., Shams, I., & Busse, J. W. (2009). Comparisons of Citations in Web of Science, Scopus, and Google Scholar for Articles Published in General Medical Journals. JAMA – J. Am. Med. Assoc. *302*(10), 1092-1096. (2009)

Bar-Ilan, J. Citations to the "Introduction to Informetrics" Indexed by WOS, Scopus and Google Scholar. Scientometrics, , 1-12. (2010)

HistCite. Bibliometric Analysis and Visualization Software.
http://www.histcite.com/index.htm

Jacsó, P. Google Scholar Ghost Authors, Lost Authors and Other Problems. Libr. J. 134(18), 26-27. (2009) http://www.libraryjournal.com/article/CA6698580.html

Garfield, E. The History and Meaning of the Journal Impact Factor. JAMA – J. Am. Med. Assoc. (293): 90-93, (2006)
http://garfield.library.upenn.edu/papers/jifchicago2005.pdf

Addresses of congratulating authors:

Tefko Saracevic
School of Communication and Information, Rutgers University
4 Huntigton Street, New Brunswick, NJ 08901, USA
Email: tefkos[at]rutgers.edu

Eugene Garfield
Chairman Emeritus
ThomsonReuters Scientific (formerly ISI)
1500 Spring Garden Street, Philadelphia, PA 19130, USA
Email: Garfield[at]codex.cis.upenn.edu