

What do citations, downloads and readership data of an information systems journal have in common and where do they differ?

Juan Gorraiz,¹ Christian Gumpenberger,¹ Kris Jack,² Peter Kraker,³ Christian Schlögl⁴

¹*(juan.gorraiz/christian.gumpenberger)@univie.ac.at*

University of Vienna, Vienna University Library, Dept of Bibliometrics, Boltzmannngasse 5, A-1090 Vienna (Austria)

²*kris.jack@mendeley.com*

Mendeley, 144a Clerkenwell Road, London, EC1R5DF (UK)

³*pkraker@know-center.at*

Know-Center, Inffeldgasse 13, A-8010 Graz (Austria)

⁴*christian.schloegl@uni-graz.at*

University of Graz, Institute of Information Science and Information Systems, Universitätsstr. 15/F3, A-8010 Graz (Austria)

Abstract

In our article we investigate the commonalities and differences across citations, downloads and Mendeley readership data for The Journal of Strategic Information Systems. Our analysis shows a medium to high correlation between downloads and citations (Spearman $r=0.77$) and downloads and readership data (Spearman $r=0.73$) which is lower between citations and readership data (Spearman $r=0.51$). The difference among the three data sets is stronger with regards to obsolescence characteristics.

Introduction

There exist already several studies which have compared download and citation data. With the advent of the social web and its growing acceptance in academia, alternative metrics seem to be a further source for the measurement of science (Bar-Ilan et al, 2012). In particular, the social reference management system Mendeley seems to be an interesting candidate.

It is the goal of this contribution to investigate the commonalities and differences across citations, downloads and readership data for an information systems journal (The Journal of Strategic Information Systems). In particular, the following issues will be addressed:

- Are most cited articles the most downloaded ones and those which can be found most frequently in user libraries of the collaborative reference management system Mendeley?
- Do citations, downloads and readership data have different obsolescence characteristics?

Both citations and downloads were provided by Elsevier in the framework of the Elsevier Bibliometric Research Program (EBRP). For all documents published between 2002 and 2011 all monthly downloads were made available from ScienceDirect (SD) and all monthly citations from Scopus until mid of 2012. Furthermore, we received the number of occurrences of full length articles in user libraries together with the date of their particular data entry in Mendeley from 2002 to 2011.

Results

Download data

Though we received the download data for all SD documents, we concentrate on full length articles in the following. As Table 1 shows (since these are critical data for ScienceDirect, only reference numbers are displayed), there was a strong increase in the number of downloads between 2002 and 2011. An analysis of the obsolescence characteristics reveals that from the downloads of a certain year, most of them allot to articles either published in the download year or one year earlier (formatted in bold). However, also older articles are downloaded relatively often.

Table 1. Yearwise *relation*¹ of downloads per print publication year (2002-2011), (doc type: full length article, download year: ≤2011) (n=181)

Pub year	n	Download year											downloads per doc – <i>relations</i> ¹
		2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	All	
2002	13	1.0	2.3	1.7	1.3	1.2	1.4	2.4	2.8	2.8	2.7	19.6	7.4*x
2003	21	0.0	1.3	2.2	1.0	1.0	0.9	1.5	1.3	1.5	1.1	11.9	2.8*x
2004	17			1.7	2.6	2.1	2.2	2.4	2.7	2.9	2.3	18.9	5.5*x
2005	18				1.7	2.3	1.8	2.0	2.4	2.6	2.2	15.0	4.1*x
2006	14				0.2	2.4	2.1	1.8	2.1	2.0	2.0	12.5	4.4*x
2007	18					0.0	2.7	3.6	3.4	3.5	2.9	16.1	4.4*x
2008	16						0.0	2.9	3.5	3.0	2.4	11.8	3.6*x
2009	14								3.1	4.0	3.1	10.2	3.6*x
2010	21									3.9	4.4	8.3	2.0*x
2011	29									0.3	5.6	5.9	1.0*x
all	181	1.0	3.7	5.6	6.8	8.9	11.1	16.6	21.4	26.4	29.0	130.4	

¹ Since the download numbers are very sensitive, we did not provide the absolute figures but only the ratios among them.

Citation data

Table 2. Year-wise citations (2002-2011) per publication year (document types: article, review, conference paper), only cited documents (n=150).

Pub year	n	Citation year											cites per doc
		2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	all	
2002	13	2	19	38	69	88	105	158	165	194	199	1037	79.8
2003	14		1	6	21	27	39	35	41	40	39	249	17.8
2004	17				15	40	56	74	78	88	107	458	26.9
2005	19					16	46	78	76	93	99	408	21.5
2006	14				1	2	14	31	31	53	49	181	12.9
2007	18						1	31	74	92	85	283	15.7
2008	15							3	30	69	83	185	12.3
2009	14								3	34	57	94	6.7
2010	18									5	40	45	2.5
2011	8										14	14	1.8
all	150	2	20	44	106	173	261	410	498	668	772	2954	

Since ScienceDirect and Scopus use different document types which are not compatible to each other, the Scopus analyses were performed for the three Scopus document types “article”, “conference paper” and “review” which mainly correspond to “full length articles” in SD. Table 2 shows the year-wise citation distribution of articles, reviews and conference papers between 2002 and 2011. As can be seen, in all citation years most citations (formatted in bold) accrue to articles from the publication year 2002. In contrast, only a few documents were cited in the year of publication. This shows a clear difference to downloads which have their maximum either in the year of publication or one year later.

Readership data

Since Mendeley started in 2008, there are only very few readership data for this year. The strong increase of the users – the end of 2012 Mendeley reached two million users – makes an obsolescence analysis difficult. As can be seen in Table 3, the maximum occurrences accrued in the last one and a half years. However, also older articles, for instance published in 2002, were added to user libraries in more recent years though, due to a much smaller time window, their total is not as high compared to the download and citation data.

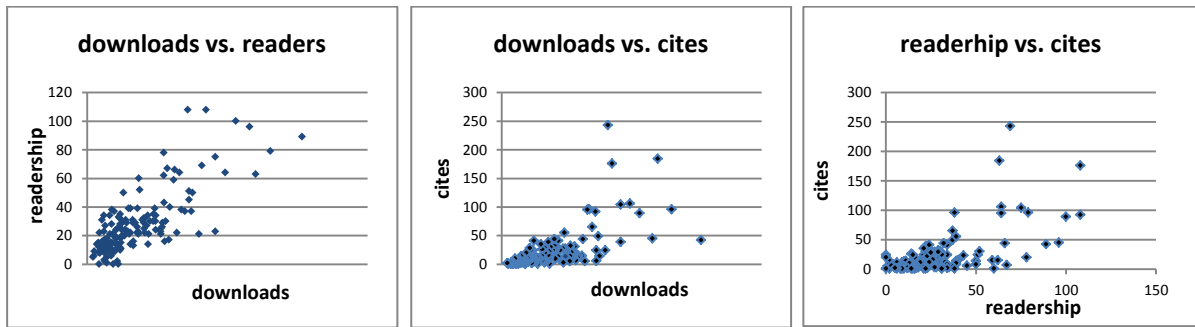
Table 3. Readership data per print publication year (2002-2011), (doc type: full length article, data extracted from Mendeley: October 2012) (n=181)

Pub year	n	Readership years						Downloads per doc
		2008	2009	2010	2011	- July 2012	all	
2002	13	7	30	126	245	183	591	45.5
2003	21	1	29	58	108	145	341	17.1
2004	17	11	36	107	158	165	477	28.1
2005	18	2	31	79	141	151	404	23.8
2006	14	6	39	88	128	148	409	29.2
2007	18	4	45	129	222	209	609	35.8
2008	16	7	36	99	182	164	488	32.5
2009	14	0	27	111	127	150	415	29.6
2010	21	0	0	84	238	191	513	24.4
2011	29	0	0	4	208	282	494	17.6
all	181	38	273	885	1757	1852	4741	

Comparison among downloads, citations and readership data

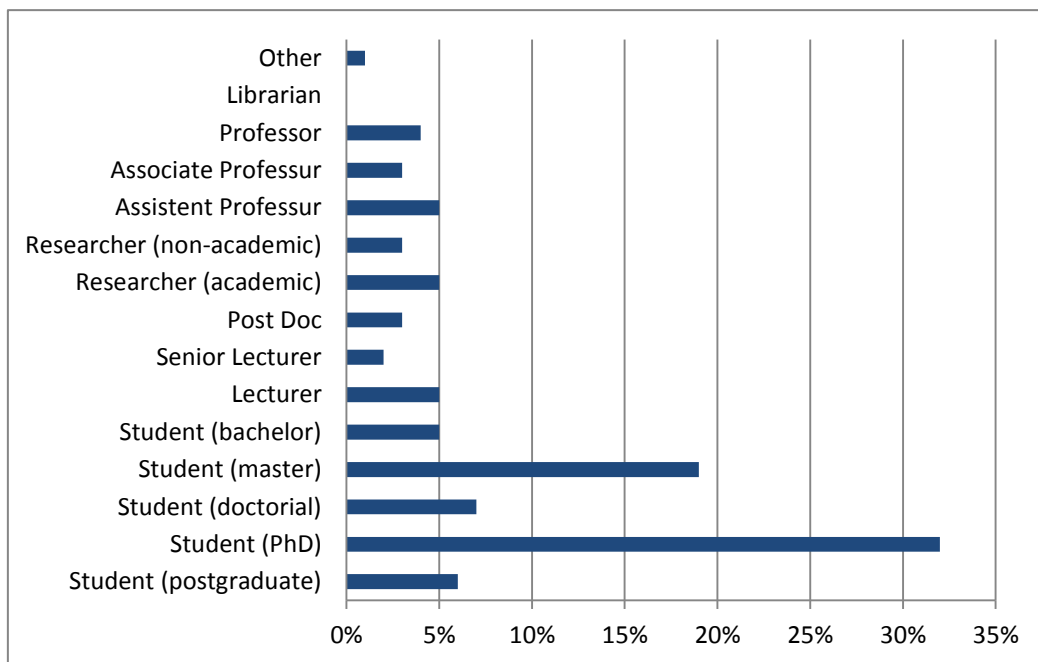
Figure 1 shows a medium to high relation among downloads, citations and readership data which is higher for downloads and citations (Spearman $r = 0.77$) and for downloads and readership data (Spearman $r = 0.73$). Among the ten most downloaded articles, seven (not the same) are in the top-10 readership and citation rankings. The correlation was lower between readership data and citations (Spearman $r = 0.51$) but in line with previous studies (for instance, Bar-Ilan 2012; Li, Thelwall and Giustini 2012).

Figure 1. Downloads vs. readers vs. cites, scattergram (publication year: 2002-2011, doc type: full length article, only articles cited at least once) (n=151)



One reason for the lower correlation between Mendeley readership and citation data could be that Mendeley users are younger. As Figure 2 exhibits, approximately two thirds of the readers of the analyzed information systems journal were students (most of them PhD and master students). This population clearly differs from that of the citing articles from the same journal.

Figure 2: Readership structure of the articles (2002-2011)



Conclusions and future research

Our study showed a medium to high correlation between downloads and citations and downloads and readership data which is lower between citations and readership data. In most cases, the most cited articles are also among the most downloaded and most read ones. An analysis of the readership data revealed that a perfect relation, at least between citations and Mendeley readership data, cannot be expected because of the different populations involved. There are clear differences between citations and downloads with regards to obsolescence characteristics. The obsolescence of readership data might be more similar to those of the downloads. However, due to the dynamic development of Mendeley, a definite prediction is not yet possible.

In the near future, we plan investigations with more journals also from other disciplines (e.g. economics, oncology, linguistics, and history).

Acknowledgments

This report is based in part on analysis of anonymous ScienceDirect usage data and/or Scopus citation data provided by Elsevier within the framework of the Elsevier Bibliometric Research Program (EBRP). Readership data were provided by Mendeley. The authors would like to thank both Elsevier and Mendeley for their great support.

References

Bar-Ilan, J. (2012). JASIST@mendeley. *ACM Web Science Conference 2012 Workshop*. Retrieved January 23 2013 from: <http://altmetrics.org/altmetrics12/bar-ilan/>

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H. & Terliesner, J. (2012). Beyond citations: scholars' visibility on the social web. In *Proceedings of the 17th International Conference on Science and Technology Indicators* (pp. 98-109). Montréal: Science-Metrix and OST.

Li, X., Thelwall, M. & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement. *Scientometrics*, 91(2), 461-471.