Open Access Publishing and Citation Impact - An International Study

Thed van Leeuwen, Clifford Tatum, and Paul Wouters

leeuwen@cwts.nl, c.c.tatum@cwts.leidenuniv.nl, p.f.wouters@cwts.leidenuniv.nl
CWTS, Leiden University, Wassenaarseweg 62a, Leiden (the Netherlands)

Abstract
This paper describes the analysis of open access (OA) publishing in the Netherlands in an international comparison. As OA publishing is now actively stimulated by Dutch science policy, similar to the UK, a bibliometric baseline measurement is conducted to assess the current situation, to be able to measure developments over time. For the study we collected data from various sources, and for three different smaller European countries (the Netherlands, Denmark, and Switzerland). Not all of the analyses for this baseline measurement are included here; the analysis presented in this paper mainly focuses on the various ways OA can be defined while using Web of Science, and the problems with interpreting these results. From the data we collected, we can conclude that the way OA is currently registered in various electronic bibliographic databases is quite unclear, and various methods applied deliver results that are different, although the impact scores point in the same direction.

Conference Topic
Journals, databases, and electronic publications

Introduction
Acceleration of open access goals in the Netherlands coincides with implementation of new current research information systems (CRIS) at Dutch universities and research institutes. This deployment of institutional CRIS systems provides an opportunity for national level tracking of open access through coordinated metadata schemes and common registration practices. As open access is notoriously difficult to measure, contemporary analyses often employ random sampling techniques (Archambault et al., 2014; Björk et al., 2010). All publication records in a given sample are tested to determine the proportion of full texts that are open access publications. National level coordination of research information provides an opportunity for improved, more precise assessment of open access publishing. In this study we use bibliographic data to establish a baseline analysis of the proportion of open access publishing in the Netherlands.

Assessment of open access publishing is complicated by a growing diversity of what counts as open access, the copyright restrictions for when a publication can be made openly accessible, and the lack of clear and consistent identification of open access publications in bibliographic data. To examine these challenges we begin with a definition from the Budapest open access Initiative (BOAI):

> Free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited. (BOAI 2002)

This definition highlights two distinct channels of access: (1) human access to read, download, and reuse the full text of published articles; and (2) machine access to crawl, index, or analyze the content of articles. The BOAI also proposes two operational paths to access through open access journals and self-archiving in repositories, subsequently referred to as

Gold open access and Green open access (Bailey, 2005). Hybrid open access generally refers to the situation whereby authors can pay to make their articles in subscription journals openly accessible on the Web (Björk, 2012).

In addition to the broad categories of Gold, Green, and Hybrid modes of open access, multiple versions of a manuscript may exist due to variations in publishers' licensing agreements. These agreements typically specify how, when, and under which conditions a manuscript may be openly accessible on the web. For example, a publisher may allow Green open access through self-archiving in an institutional repository. However, publishers' copyright restrictions differ on the stage of manuscript development that may be openly accessible, thus assigning different rights to different versions of the text. Commonly specified version types include the submitted manuscript (before peer review), the accepted manuscript (peer-reviewed but not formatted), and an exact copy of the published manuscript (Björk et al., 2013). This creates the possibility that the open access version of a manuscript is substantively different from the published version. In such instances, it is unclear whether the open access version has been sufficiently validated through the quality control measures such as peer review.

Another variation is delayed access, which is applied as an embargo period, after which a copy of the publication may be self-archived or the publisher may remove access restrictions on the journal website. Embargo periods are generally specified as a delay of 6, 12, 18, or 24 months after publication, with 12 months being the most common embargo period (Laakso & Björk, 2013). For Green open access, it is thus left to authors and institutions to track and manage a variety of self-archiving policies, which in itself has been shown to be a barrier to open access (Davis & Connolly, 2007). However, this kind of administrative overhead is largely absent from subscription journals that convert articles to open access after a specified delay (e.g. 12 months). In addition, a bibliometric analysis of 'delayed access' journals found journal and article impact factors higher than comparable averages from both subscription journals and direct (no delay) open access journals (Laakso & Björk, 2013).

A common refrain among proponents of open access is that open access publishing yields increased citation impact. While there are conflicting reports regarding an open access citation advantage (OACA), heightened attention to this issue has increased our understanding about citation behaviour more generally. Numerous bibliometric studies claim that open access publishing results in a significant increase in citations. In these studies the size of advantage varies widely based on a variety of issues, such as disciplinary differences, methodological approaches, variation in how open access is defined, and difficulty in determining when an article is made openly accessible (Swan, 2010). In addition, a number of confounding factors have been shown to influence citation frequency such as early exposure to draft versions of a manuscript (Moed, 2007), self-selection bias whereby an author may choose open access for only her best publications (Kurtz et al., 2007), the availability at multiple access points (Xia, Myers & Wilhoite, 2011), and physical proximity of researchers (Lee et al., 2010).

To control for these factors, Davis et al. (2008) employ randomized controlled trial methods, whereby randomly selected articles in subscription based journals are switched to open access. The resulting configuration is similar to hybrid open access, such that the article is made to be openly accessible and is listed among the non-open access articles on the journal's website. In the Davis et al. (2008) study a citation advantage was not present. However, the research design used to control for confounding variables (randomized controlled trial) also limited applicability of the findings to the hybrid model of open access. More recently, Archambault et al. (2014) show variation in the accumulation of citations associated with the different modes of open access. The authors find a citation *advantage* most prominently associated with the self-archiving mode of open access (Green OA) and a citation

disadvantage associated with full and immediate open access journals (Gold OA). This study also establishes a general ranking of citation accumulation on the bases of open access, listed in order of most to least: Green OA, Other OA, Not OA, and Gold OA." (Archambault et al., 2014, pp. 20, 24)

To address the variability of circumstances associated with open access publishing, recent studies invert the research design from top-down queries of bibliometric datasets to bottom-up testing whether a publication is an open access publication. This approach involves random sampling of a given publishing domain, harvesting full-texts from the Internet, and analysis of available metadata from harvested manuscripts (Björk et al., 2010). While this approach circumvents much of the variability noted above, it is nevertheless dependent on the presence and quality of metadata. (The potential for improved metadata practices is addressed in the discussion section below.)

The objective of our analysis is to show the challenges of bibliometrically analysing OA publications and associated impact scores. We use Web of Science (WoS) data, either directly retrieved from the database, or combined with article-level data extracted from journals listed in the Directory of Open Access Journals (DOAJ). As both data sources are incomplete with respect to open access publications, the analysis is focused on comparison of relative output and relative impact among three European countries of similar size and scientific production: the Netherlands, Denmark, and Switzerland, in order to show developments in time, as well as differences resulting from both approaches. It is important to note that Green OA articles are excluded from our analysis. While the Netherlands maintains a robust national repository for Green OA (NARCIS), there is not yet a reliable system of identifying the self-archived state of publications within bibliometric datasets. As such, the proportion of open access and associated impact comparisons are limited to the available data on Gold OA.

Data collection

In the study we make use of data from various sources. The Web of Science (WoS) database is used in its internet version, available to most Dutch researchers. We also used the CWTS version of the WoS, a tailor-made database based upon state-of-the-art bibliometric techniques and indicators. In this version, the functionality to search for OA output is not yet available. Finally, we make use of the journals and the publications listed in the Directory of Open Access Journals (DOAJ). From this data source, we will further focus on the digital object identifiers (DOIs), while leaving out other elements (such as the license types, as this information is unclearly defined as well as unclearly linked to the publications).

Method I: The first way of data collection from WoS starts from the desktop interface of the WoS database. The functionality to collect this information is not yet available in the in-house WoS database at CWTS, so therefore we had to collect these data from the internet version directly. This approach involved the following steps:

1) Collect the output of one of the selected countries for a particular year;
2) Within that set, further distinguish the OA part of that selected output;
3) Download these publications from the WoS database (including the so-called UT-code, a unique identifier within WoS that allows for linking to the CWTS WoS database);
4) Select within the CWTS database the output for the three countries;
5) Match the selected output from the Internet version of the WoS with the in-house CWTS version;
6) Create two sets within the CWTS database, an OA formatted set of publications, and a non OA formatted set of publications.

These steps were taken for all three countries, collecting publications from 2000-2013.

The definition of how the publications were defined as OA is based upon the following statement on the WoS database' website: "The Thomson Reuters Links open access Journal

Title List includes free journal content that are available for linking from the Web of Science."

Method II: The second method started from the Directory of Open Access Journals (DOAJ). This list contains journals that have implemented the Gold open access business model. CWTS has downloaded the complete list, and all publications published in the journals on the DOAJ list. By making use of this dataset, we could use a second approach to the OA output of the three countries taking the following steps:

1) First select within the CWTS database the output for the three countries;
2) Collect their Digital Object Identifiers (doi);
3) Match these with the doi's of the publications downloaded from the DOAJ list;
4) Create two sets within the CWTS database, an OA formatted set of publications, and a non OA formatted set of publications.

We focused on articles, letters and reviews only, excluding other types of documents such as editorials, meeting abstracts, book reviews, etc. The choice for these types is based upon the importance of these three types in communicating scientific findings among peers, and their relative homogeneity within the system.

Methods

In the study we present a number of indicators. In cases we present numbers of publications, this is indicated with a P. In case citation data are presented, we use MNCS (Mean Normalized Citation Score), as well as the MNJS, the field normalized journal impact indicator, to indicate the normalized impact scores in the study (Waltman et al., 2011a; Waltman et al., 2011b). While the output indicator can be used for the various electronic systems we use in the study, and P can relate to various document types analysed, the citation impact indicators are used only within the context of the WoS database. In case of the impact indicators, the length of the citation window is one year longer than the presented year block (so in case of the last block, 2009-2012, the citation impact is measured up until 2013, currently the last year fully covered in the CWTS WoS database).

Results

First we present the results from Method I, described above. The output numbers of the three countries according to the methodology I are found in Table 1 along with the two separate parts of the output, distinguished by openness. The analysis covers the period 2000 up until 2012 for publication data, and up until 2013 for citation impact data. In this analysis we use moving publication year windows, in order to create more solid and stable trend lines, as we are more interested in the trends than in variation from year to year.

The data presented in Table 1 clearly show that OA publishing is becoming increasingly important, in all three selected countries. The Netherlands is lagging somewhat behind Denmark and Switzerland, albeit with only a small part of the total output.

In Figure 1, we have distinguished between the open access format output of the three countries (indicated by the 'Ex OA' label to the country names). What we observe are increasing trends for the parts of the output not published in OA format, which is also visible for the OA format of the output of these three countries, and as shown above in Table 1, increases somewhat faster for Denmark and Switzerland as compared to the Netherlands.

Table 1. Output (P) of Denmark, the Netherlands, and Switzerland, distinguishing OA and non-OA output, 2000-2012.

	NL Ex OA	NL OA	Share OA	DK Ex OA	DK OA	Share OA	CH Ex OA	CH OA	Share OA
2000 - 2003	75607	712	1%	30616	452	1%	53283	995	2%
2001 - 2004	78087	858	1%	31262	557	2%	54793	1220	2%
2002 - 2005	81849	1180	1%	31972	728	2%	56982	1836	3%
2003 - 2006	85386	1663	2%	33024	949	3%	60319	2217	4%
2004 - 2007	88745	2349	3%	34082	1244	4%	63205	2790	4%
2005 - 2008	92349	3265	4%	35273	1631	5%	65920	3517	5%
2006 - 2009	96278	4269	4%	36672	1997	5%	69518	3912	6%
2007 - 2010	101270	5587	6%	38726	2554	7%	72687	4981	7%
2008 - 2011	106560	7299	7%	41417	3264	8%	76658	6354	8%
2009 - 2012	111990	9504	8%	44264	4420	10%	80786	7990	10%

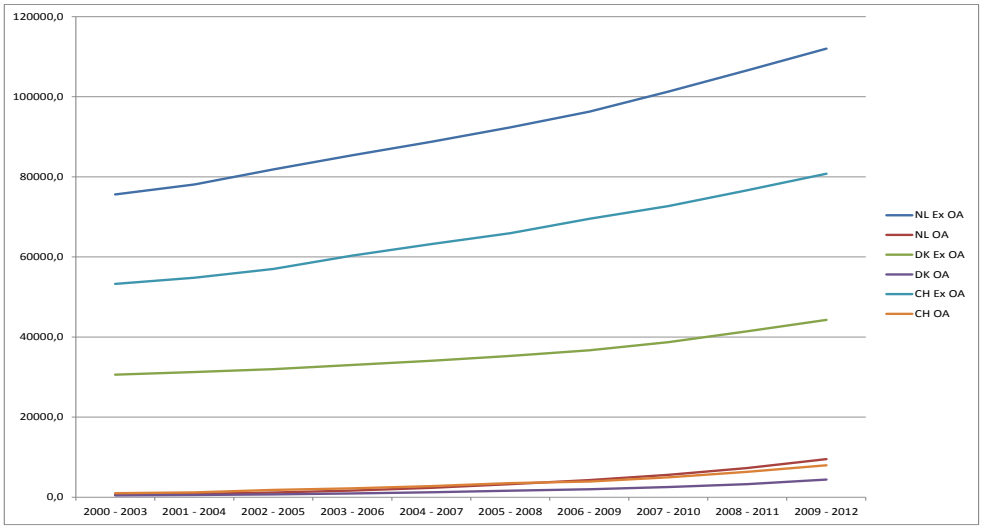

Figure 1. Output development (P) of Denmark, the Netherlands, and Switzerland, 2000-2012/2013.

In Table 2, we present the citation impact scores as represented by the MNCS indicator, the field normalized impact of the outputs of the three countries, again separated by the two types of publication output: open access and non-open access publications.

Figure 2 shows that for all three countries the non-OA part of the output has a citation impact well above world average, with Switzerland topping the other two countries, which have a nearly equal field normalized impact score. The impact of OA publications is lower for all three countries. The impact of the OA part of the national outputs of Denmark and Switzerland were initially well above world average. This is also the case for Swiss publications, as the OA format published output is lower on MNCS only from 2007-2010/2011 onwards. In case of Denmark, this drop started somewhat earlier, while in the case of the Netherlands, the OA output never got an impact higher than that of the non-OA format output. Another interesting phenomenon is the increase of the gap between the impact of OA and non-OA output. This is particularly the case for Switzerland and Denmark, where we observe a clear drop of the impact of OA format output compared to their non-OA formatted output, and to a lesser extent for the Netherlands, where the two impact lines are more slowly diverging.

Table 2. Citation impact (MNCS) of Denmark, the Netherlands, and Switzerland, distinguishing OA and non-OA output, 2000-2012.

	NL Ex OA	NL OA	DK Ex OA	DK OA	CH Ex OA	CH OA
2000 – 2003	1,29	0,99	1,30	1,03	1,37	1,11
2001 - 2004	1,30	0,95	1,29	1,31	1,35	1,21
2002 - 2005	1,30	0,99	1,29	1,39	1,36	1,36
2003 - 2006	1,31	1,07	1,31	1,34	1,36	1,46
2004 - 2007	1,30	1,12	1,31	1,30	1,38	1,47
2005 - 2008	1,31	1,13	1,32	1,30	1,39	1,48
2006 - 2009	1,35	1,15	1,34	1,26	1,39	1,39
2007 - 2010	1,38	1,17	1,37	1,26	1,42	1,37
2008 - 2011	1,40	1,18	1,40	1,25	1,46	1,36
2009 - 2012	1,44	1,18	1,44	1,18	1,50	1,33

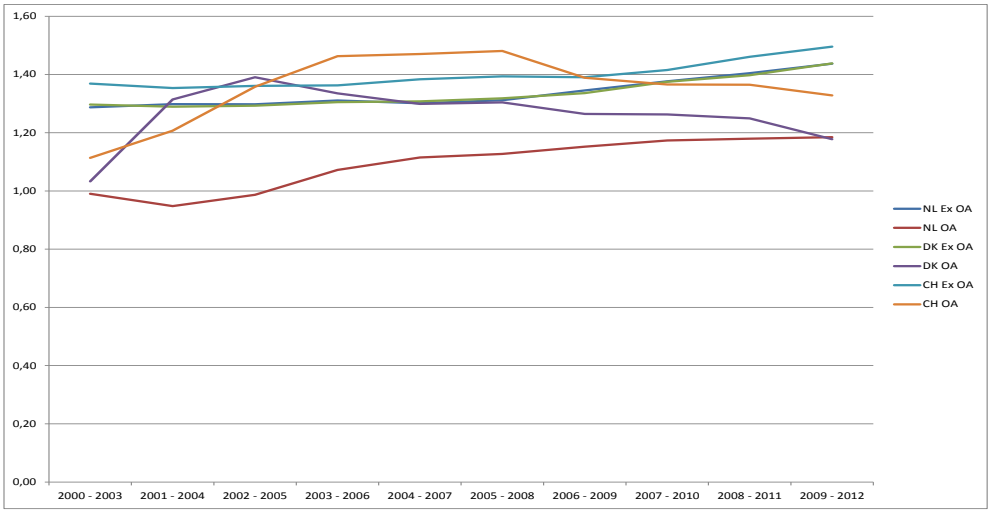

Figure 2. Impact development (MNCS) of Denmark, the Netherlands, and Switzerland, 2000-2012/2013.

If we shift our focus towards the journal impact analysis (see Table 3 and Figure 3), for which we use the indicator MNJS, we see an even more interesting phenomenon. While the output in non-OA format published journals shows a choice for journals with increasing impact scores, the OA format published outputs end up in journals with decreasing field normalized impact scores. We even notice a diverging trend in these two clusters of trend lines: non-OA format published journals tend to show increasing impact scores, while OA format published journals show decreasing impact trends. This is striking since these are three of the 'scientifically stronger' nations, as far as can be measured with bibliometric instruments.

Here we start with the results from methodology II. The results of the output analysis are shown in Table 4, which again covers a similar distinction between OA and non-OA format output, but now according to the definition described above under Method II. We combined the DOIs of journals on the DOAJ list with the DOIs available in the WoS. From the total set of 787,611 DOIs in the DOAJ list, we matched 226,641 publications in WoS on the basis of available DOIs. The reason for this seemingly low recall is twofold. In the first place, not all journals covered by the DOAJ list are processed for the WoS database, and secondly, not all publications in journals covered in WoS do contain DOIs. This means that for some journals that are both covered in the DOAJ list as well as in WoS, a match is impossible, particularly

for the earlier years in the analysis. Like the first methodology we followed, we separated the OA format published output from the Netherlands, Denmark, and Switzerland from the total set of publications for the three countries under study.

Table 3. Journal-to-field citation impact (MNJS) of Denmark, the Netherlands, and Switzerland, distinguishing OA and non-OA output, 2000-2012

	NL Ex OA	NL OA	DK Ex OA	DK OA	CH Ex OA	CH OA
2000 - 2003	1,18	0,95	1,15	0,84	1,19	1,06
2001 - 2004	1,19	0,97	1,16	1,02	1,20	1,03
2002 - 2005	1,19	1,00	1,16	1,08	1,20	1,19
2003 - 2006	1,20	1,06	1,16	1,11	1,20	1,20
2004 - 2007	1,22	1,09	1,18	1,12	1,22	1,11
2005 - 2008	1,24	1,09	1,20	1,10	1,24	1,14
2006 - 2009	1,26	1,11	1,22	1,07	1,26	1,11
2007 - 2010	1,29	1,11	1,25	1,06	1,29	1,11
2008 - 2011	1,30	1,10	1,26	1,05	1,31	1,11
2009 - 2012	1,32	1,09	1,28	1,00	1,33	1,09

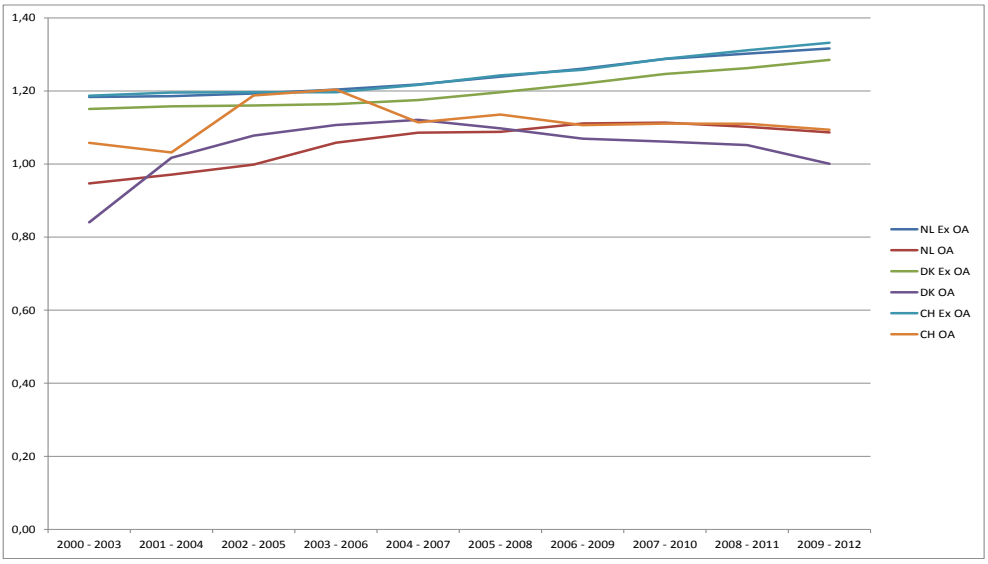

Figure 3: Journal impact development (MNJS) of Denmark, the Netherlands, and Switzerland, 2000-2012/2013.

First of all, we observe that the overlap between the DOAJ list/WoS combinations with Dutch/Danish/Swiss publications in WoS is much smaller compared to the previous analysis on Dutch/Danish/Swiss output in OA format, which is most likely the result of the missing DOIs in the WoS database. If we compare the results of Table 1 with those presented in Table 4, we find much lower shares of OA output compared to the overall output of the three countries. This is further underlined by Figure 4, in which the OA format output of the three countries is at the low end of the graph, while we simultaneously observe a strong increase in the output of the non-OA format output of the three countries.

Table 4. Output (P) of Denmark, the Netherlands, and Switzerland, distinguishing OA and non-OA output (based on DOI-matching), 2000-2012

	NL Ex OA	NL OA	Share OA	DK Ex OA	DK OA	Share OA	CH Ex OA	CH OA	Share OA
2000 - 2003	75607	10	0%	30616	4	0%	53283	2	0%
2001 - 2004	78087	35	0%	31262	25	0%	54793	30	0%
2002 - 2005	81849	136	0%	31972	83	0%	56982	97	0%
2003 - 2006	85386	344	0%	33024	170	1%	60319	232	0%
2004 - 2007	88745	648	1%	34082	312	1%	63205	420	1%
2005 - 2008	92349	1068	1%	35273	486	1%	65920	690	1%
2006 - 2009	96278	1531	2%	36672	664	2%	69518	972	1%
2007 - 2010	101270	2207	2%	38726	924	2%	72687	1461	2%
2008 - 2011	106560	3036	3%	41417	1231	3%	76658	2062	3%
2009 - 2012	111990	3896	3%	44264	1595	4%	80786	2608	3%

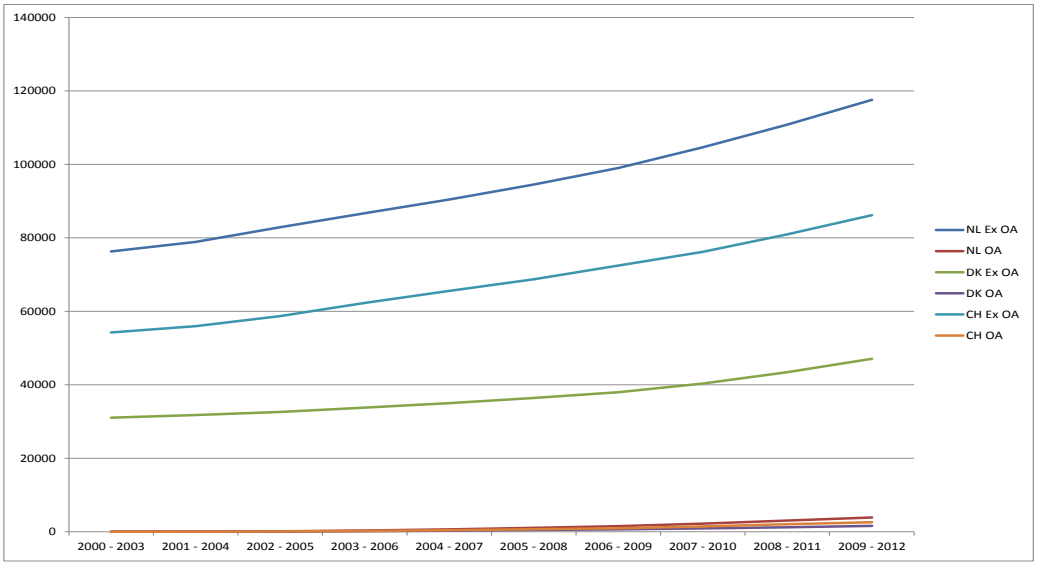

Figure 4. Output development (P) of Denmark, the Netherlands, and Switzerland, based on matching of DOI's, 2000-2012/2013.

In Table 5, we present the impact scores of the three countries, again distinguishing OA format output and non-OA format output. Again we observe lower impact scores for the OA format output of the three countries, except for the starting block of the analysis (please note that the output numbers are extremely low in this part of the analysis for the Netherlands and Denmark, respectively 10 and 4 papers). From the second year block onwards, we observe increasing trends in the impact of the OA format of the three countries, although we must stress that this is also the case for the non-OA format output of the three countries.

Figure 5 shows this stable development of both sets of publications in time, whereby the impact scores are increasing on both sets, although the 'difference' remains more or less the same between the two sets of scores.

In Table 6 we present the outcomes of the analysis on the journal impact scores, based upon methodology II. Here we observe, similar to the previous outcomes, fluctuations in the initials years of the analysis for the OA format output, followed by a more stable situation from 2005-2008 onwards. This finding is even more visible in the graphical representation of Table 6, as in Figure 6.

Table 5. Citation impact (MNCS) of Denmark, the Netherlands, and Switzerland, distinguishing OA and non-OA output (based on DOI-matching), 2000-2012

	NL ex OA	NL OA	DK ex OA	DK OA	CH ex OA	CH OA
2000 - 2003	1,28	1,65	1,29	1,32	1,36	
2001 - 2004	1,29	0,87	1,29	0,91	1,35	1,03
2002 - 2005	1,29	0,87	1,30	0,98	1,36	1,18
2003 - 2006	1,31	0,87	1,31	0,78	1,37	0,95
2004 - 2007	1,30	0,75	1,31	0,72	1,39	0,96
2005 - 2008	1,31	0,83	1,32	0,86	1,40	0,91
2006 - 2009	1,35	0,85	1,34	0,89	1,40	0,92
2007 - 2010	1,38	0,90	1,38	0,96	1,42	0,97
2008 - 2011	1,40	0,97	1,40	1,00	1,46	1,07
2009 - 2012	1,43	1,03	1,43	0,96	1,49	1,06

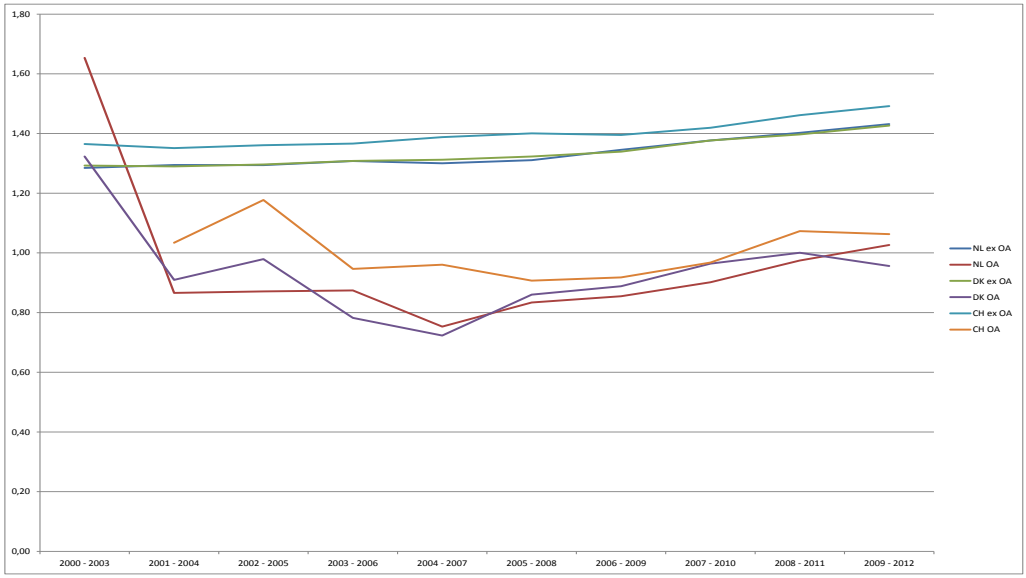

Figure 5. Impact development (MNCS) of Denmark, the Netherlands, and Switzerland, based on matching of DOIs, 2000-2012/2013.

Table 6. Journal-to-field citation impact (MNJS) of Denmark, the Netherlands, and Switzerland, distinguishing OA and non-OA output (based on DOI-matching), 2000-2012

	NL ex OA	NL OA	DKex OA	DK OA	CH ex OA	CH OA
2000 - 2003	1,18	0,54	1,15	1,28	1,19	0,24
2001 - 2004	1,18	0,84	1,16	0,92	1,19	1,22
2002 - 2005	1,19	0,77	1,16	0,84	1,20	1,00
2003 - 2006	1,20	0,84	1,16	0,79	1,20	0,90
2004 - 2007	1,22	0,86	1,18	0,83	1,22	0,88
2005 - 2008	1,24	0,88	1,20	0,86	1,24	0,86
2006 - 2009	1,26	0,90	1,22	0,87	1,26	0,87
2007 - 2010	1,29	0,94	1,24	0,91	1,29	0,91
2008 - 2011	1,30	0,97	1,26	0,93	1,31	0,96
2009 - 2012	1,31	0,97	1,27	0,92	1,32	0,97

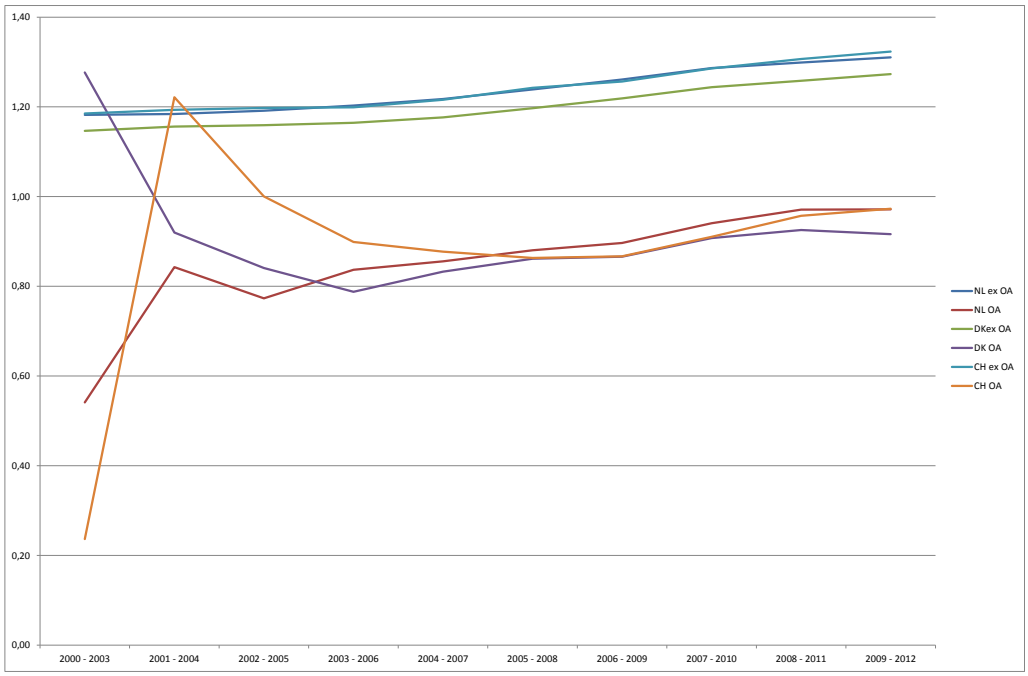

Figure 6: Journal impact development (MNJS) of Denmark, the Netherlands, and Switzerland, based on matching of DOI's, 2000-2012/2013

Conclusion and Discussion

In this final part of the paper, we will summarize the main bibliometric findings, and then move towards limitations in the ways OA is now disclosed in electronic systems supporting bibliometric analyses. Finally, we will discuss the need to improve identification of open access publications and the use of bibliometric techniques to measure OA.

Please note that our conclusions are mainly related to the domains in which journal publishing is the dominant way of communication (the natural, life and medical sciences, and to a lesser extent the social sciences and humanities (van Leeuwen, 2013). We observe for the three countries that the share in output in OA journals is lagging behind as compared to the journals that maintain the non-OA format. We observe a divergence in the development of citation impact for (Gold) OA and non-OA publications with consistently lower impact for the OA publications.

Second, we observe that OA journals have lower journal impact scores than non-OA journals. This may mean that they still struggle to find their position within the total 'reputational hierarchy' of the domain, and as such also within the WoS database. This is a common problem for new journals, and OA journals are no exception. It should be noted however, that our findings associated with OA impact are consistent with what others have found: Gold OA is associated with no citation advantage or a disadvantage (e.g. Archambault et al., 2014). With the inclusion of the various forms of Green OA, we would expect to find a larger proportion of open access articles and a more nuanced outcome related to impact. That Green OA has been found to have increased accumulation of citations (Archambault et al., 2014), may be associated with the circumstances identified above as confounding factors (e.g. early exposure, multiple access points, and proximity of researchers).

Third, we may need to worry about the role of peer review in the journals that are part of the expansion of the WoS database in the last couple of years, many of which are in the OA segment of the database. The Institute for Scientific Information, the predecessor of the current owner of the WoS database Thomson Reuters, always clearly indicated that a properly functioning peer review system within a journal was one of the conditions for a journal to be included in the system (next to other criteria, such as international focus, regular appearance,

preferably in the English language, etc.). We do not know whether this is still such a strong criterion, particularly given the fact that so many new journals appeared around the OA development.

A fourth conclusion relates to the messy situation around the various manners by which open access is defined in electronic databases. The two different ways open access can be operationalized within the world of WoS is an example of this unclear and somewhat messy situation. The fact that the Scopus database did not have the functionality to clearly define open access for users of the system is another instance of the situation around open access. Further examples of this lack of clarity are the various ways open access is operationalized by the publishing industry. There is no clear way of operationalizing in the larger databases of the various business models (such as Gold, Green, and Hybrid open access). Yet another example relates to the various license types related to open access.

A recently published metadata standard for open access holds some promise for improving both human and machine identification of open access publications (Carpenter, 2013). Here, too, stakeholders involved in the new standard were unable to agree on a precise definition of open access. Instead, the standard specifies metadata elements for *free to read* and *license reference*, the latter of which should point to copyright information publicly accessible on the Web (NISO 2015). Increased attention to national research assessment and increased use of institutional CRIS systems together provide a potentially welcoming context for implementing new metadata practices. This would ideally include the possibility of tracking open access among the diversity of research outputs maintained by CRIS systems and considered in assessment events. In this context, it becomes important to assign openly accessible, persistent identifiers to all research objects (Tatum & Wouters 2014). This would increase the potential use of institutional research information for tracking open access as part of regular research assessment practices, rather than relying solely on estimation derived from random sampling of commercial datasets.

References

Archambault, E., Amyot, D., Deschamps, Nicol, A., Provencher, F., Rebout, L. & Roberge, G. (2014). Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013. *Rapport, Commission Européenne DG Recherche & Innovation; RTD-B6-PP-2011-2: Study to Develop a Set of Indicators to Measure Open Access.*

Björk, B.-C., Welling, P., Laakso, M., Majlender, P., Hedlund, T. & Guðnason, G. (2010). Open Access to the Scientific Journal Literature: Situation 2009. *PLoS ONE,* 5 (6): e11273. doi:10.1371/journal.pone.0011273.

Björk, B.-C. 2012. The Hybrid Model for Open Access Publication of Scholarly Articles: A Failed Experiment? *Journal of the American Society for Information Science and Technology,* 63 (8): 1496–1504. doi:10.1002/asi.22709.

Björk, B-C., Laakso, M., Welling, P. & Paetau, P. (2013). Anatomy of Green Open Access. *Journal of the Association for Information Science and Technology,* 65 (2): 237–50. doi:10.1002/asi.22963.

BOAI. (2002). Budapest Open Access Initiative. *The Open Society Foundations*. http://www.opensocietyfoundations.org/openaccess.

Carpenter, T. (2013). Progress Toward Open Access Metadata. *Serials Review,* 39 (1): 1–2. doi:10.1016/j.serrev.2013.02.001.

Davis, P.M., & Connolly, M.J. L. (2007 March). Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace. http://hdl.handle.net/1813/5195.

Kurtz, M.J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E. & Murray, S.S. (2005). The Effect of Use and Access on Citations. *Information Processing & Management*, Special Issue on Infometrics, 41 (6): 1395–1402. doi:10.1016/j.ipm.2005.03.010.

Laakso, M., & Björk, B.-C. (2013). Delayed Open Access: An Overlooked High-impact Category of Openly Available Scientific Literature. *Journal of the American Society for Information Science and Technology,* 64 (7): 1323–29. doi:10.1002/asi.22856.

Lee, K., Brownstein, J.S., Mills, R.G. & Kohane, I.S. (2010). Does Collocation Inform the Impact of Collaboration? *PLoS ONE,* 5 (12): e14279. doi:10.1371/journal.pone.0014279.

van Leeuwen, T.N. (2013). Bibliometric research evaluations, Web of Science and the Social Sciences and Humanities: a problematic relationship? *Bibliometrie - Praxis und Forschung*, 1-18. http://www.bibliometrie-pf.de/article/viewFile/173/215

Moed, H.F. (2007). The effect of "Open access" on citation impact: An analysis of ArXiv's condensed matter section. *Journal of the American Society of Information Science & Technology*, 58 (13), 2047-2054

NISO. 2015. *Access License and Indicators NISO RP-22-2015*. National Information Standards Organization.

Swan, A. (2010). The Open Access Citation Advantage: Studies and Results to Date. Technical Report. http://eprints.ecs.soton.ac.uk/18516/.

Tatum, C. & Wouters, P.F. (2014). Next Generation Research Evaluation: The ACUMEN Portfolio and Web Based Information Tools. *OpenAIRE-COAR Conference: Open Access Movement to Reality - Putting the Pieces Together*. Athens. doi:10.6084/m9.figshare.1033681.

Waltman, L., van Eck, N.J., van Leeuwen, T.N., Visser, M.S. & van Raan, A.F.J. (2011a). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, 5(1), 37-47.

Waltman, L., van Eck, N.J., van Leeuwen, T.N., Visser, M.S. & van Raan, A.F.J. (2011b). Towards a new crown indicator: An empirical analysis. *Scientometrics*, 87 (3), 467-481.

Xia, J., Myers, R.L. & Wilhoite. S.K. (2011). Multiple Open Access Availability and Citation Impact. *Journal of Information Science,* 37 (1): 19–28. doi:10.1177/0165551510389358.