

# How Readers Discover Content in Scholarly Publications

*UKSG eNews 371*

13 May 2016

## **Tracy Gardner, Simon Inger Consulting**

One question I am often asked when discussing the discovery research Simon Inger and I have carried out over the last 10 years is "Why do you need this research when publishers know where their users are coming from through the analysis of web logs?". There are two answers to this question, one fairly basic and the other a bit more complex.

Firstly, publishers know from their logs the last place a user visited before arriving at their site. This, however, doesn't tell them where the user started their journey. A researcher could have started in PubMed but the link to full text could have been re-intermediated by library technology. The publisher would therefore record a referrer hit from the library or their service provider (such as ProQuest's Ex Libris) but would have no idea where the reader started out on their journey. The knowledge of where users start is crucial to publishers when developing metadata distribution strategies and marketing campaigns.

The second reason is that publishers cannot altogether trust their weblogs. This is slightly more complex and there are two parts to the problem:

- IP access with optional proxy: let's look at an example of a university in Australia with a remote campus in South Africa. If the publisher uses its access control system to drive its analytics, it will place a user from South Africa in Australia instead. If it simply uses the IP address of the client, then the analytics will place the user in South Africa, or potentially in the country where a proxy has been remotely operated, often the USA or UK (and soon to be Holland for some ezproxy installations). So depending on the sophistication of the analytics set up, our reader in South Africa could be recorded as being from South Africa, Australia or Holland! For many publishers this renders their analytics inaccurate and misleading.
- IP error: we know that through accumulated procedural error, and in some cases fraud, IP address databases held by many publishers are deeply inaccurate. Often users notice that a publisher website places them in more than one location, ie when they arrive at a publisher website the website will list two, often completely unconnected, institutions that they are part of. If the user goes on to download an article, then the publisher's COUNTER statistics often increment a download for each account, invalidating the publisher's own data, and the data it supplies to libraries.

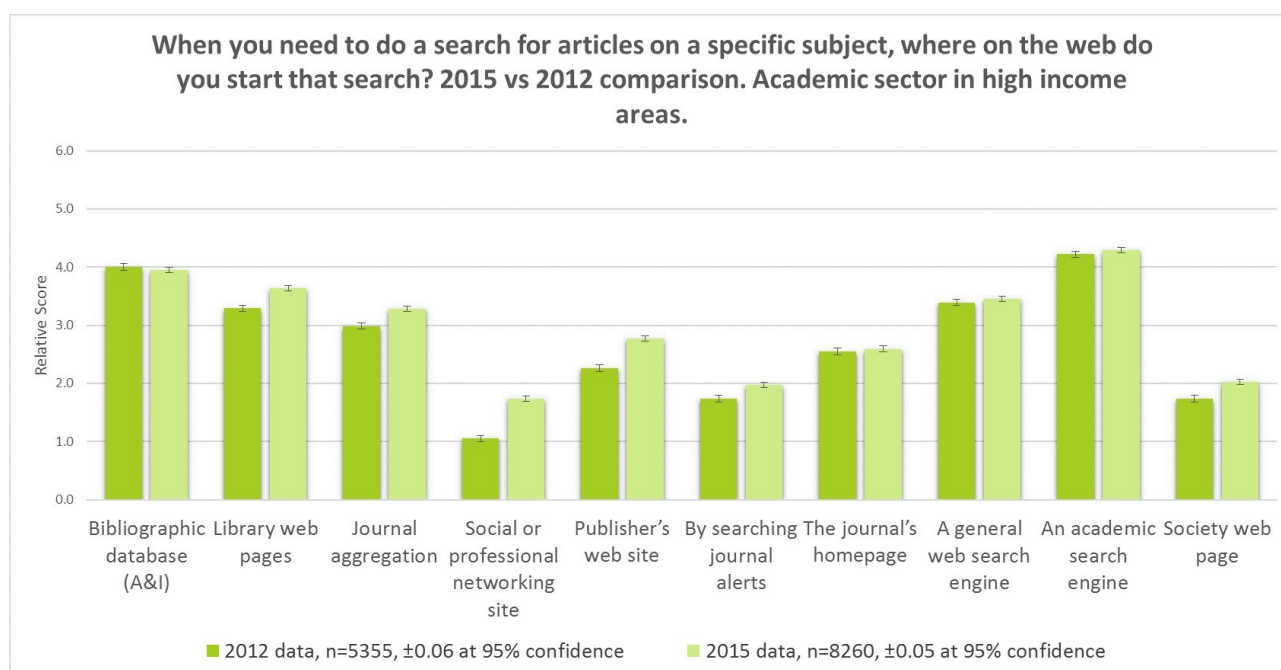
There is a proposed NISO project to address some of these issues, the development of a [Recommended Practice for Tracking Link Origins in a Networked Information Environment](#).

Publisher weblogs therefore do not contain all the answers and, whilst libraries know a great deal about their users' behaviour, they only can capture what happens within the library intermediated environment.

So, we decided back in 2005 the only way to really understand readers' discovery habits was to ask them. Our latest research attracted 40,000 responses from people based all over the world, in all sectors, working and studying in a vast range of roles and subject areas.

So that covers the 'why'. Now I would like to delve into two key areas of our research which have sparked a number of discussions since the report was published.

Whilst people were perhaps not surprised that academic search engines and library discovery tools had both experienced a growth in importance for search, the rise in the importance of publisher websites was a surprise to some people. However, there is solid reasoning behind this result and several publishers have since told us that it marries with their own findings. The question we actually asked was "When you need to search for articles on a specific subject, where on the web do you start that search, please rate the following in terms of importance". The graph below shows the trend from 2012-2015 in the academic sector in high income countries. As you will see, whilst there is a growth in publisher websites, their importance in discovery still does not challenge the dominance of academic search engines and A&Is (abstracting and indexing services). This growth is definitely more marked outside of the academic sector and in lower income countries, and there are several reasons why this could be.



The overall growth in importance could simply be put down to better visibility and more sophisticated marketing from the publisher – after all, publishers spend a lot of money promoting themselves and their own websites, some of this marketing must have an effect. People may simply be more aware of the publisher website than before. Respondents could have felt that other services have become less reliable or affordable. If the user does not know about Google Scholar (as many in lower income countries and sectors outside academia do not appear to) they may have noticed Google displaying fewer scholarly results since its change of indexing policy for subscribed content several years ago and so have sought out alternative resources. For people with less knowledge or access to alternative resources the publisher website may well be the next best point of call. And whilst people may like the publisher website as a starting point they might not actually be hitting the search button when they get there - after all, there are other ways to use a publisher website than to hit the search button. This would explain the fact that later on in the survey only around half of

the people we asked found the search functionality useful.

The next key area of discussion was on the question of delivery. We asked people to indicate what proportion of their articles came from the following resources: the publisher website, journal website, full text aggregation or journal collection; a free subject repository; an institutional repository; Researchgate, Mendeley, or other scientific networking site; a copy emailed by the author. We reported that readers perceive that over half of their downloads are from free incarnations of content. PubMed Central has clearly become a significant provider of content for people in the medical sector and people in the academic sector indicated that 20% of their articles came from an institutional repository. Given that PubMed indexes PubMed Central and that A&Is are still a crucial discovery resource for people working in medicine and life sciences; and given that Google Scholar has grown in importance and both indexes institutional repositories and provides a link to the PDF in the institutional repository if they find one, this result can be explained. However, whether this figure is completely accurate or not is perhaps a somewhat moot point. One of the key messages is that if readers perceive that the majority of their needs are met through free resources then this has to have an impact on how much they are willing to support library budgets and subscriptions in the future.

The full report, '[How Readers Discover Content in Scholarly Publications](#)', is available free to download on the Simon Inger Consulting website.



This UKSG Editorial is taken from the industry newsletter *UKSG eNews*, published every two weeks exclusively for UKSG members. The newsletter provides up-to-the-minute news of current issues and developments within the global knowledge community.

To enjoy *UKSG eNews* and other member benefits [become a UKSG member](#). To submit an editorial suggestion for *UKSG eNews*, contact the editors: [seneditor@uksg.org](mailto:seneditor@uksg.org).