Library technology in content discovery – evidence from a large-scale reader survey

Based on a paper presented at the 36th UKSG Annual Conference, Bournemouth, April 2013

It has long been understood that readers of scholarly journals use a wide variety of discovery resources to find relevant journal articles. In 2012, Simon Inger and Tracy Gardner undertook a large-scale international survey of journal readers[1], which followed on from smaller surveys undertaken in 2005 and 2008. The survey benefited from over 19,000 responses from across the world, with over 25% of the responses coming from each of North America, Europe and Asia, half the responses coming from the academic sector, but with still many thousands of responses making up smaller demographics. This paper studies the findings for library technology and other library-purchased discovery resources from the survey.

Readers have a wide range of discovery options open to them, and can use them to discover articles on publishers' websites or sometimes within an aggregation service. It is very hard to know how different demographics of readers, such as senior or junior academics, actually go about discovering articles. Through statistics, web logs and analytics, the library gets a certain partial view of what is happening and the publisher another (Figure 1).
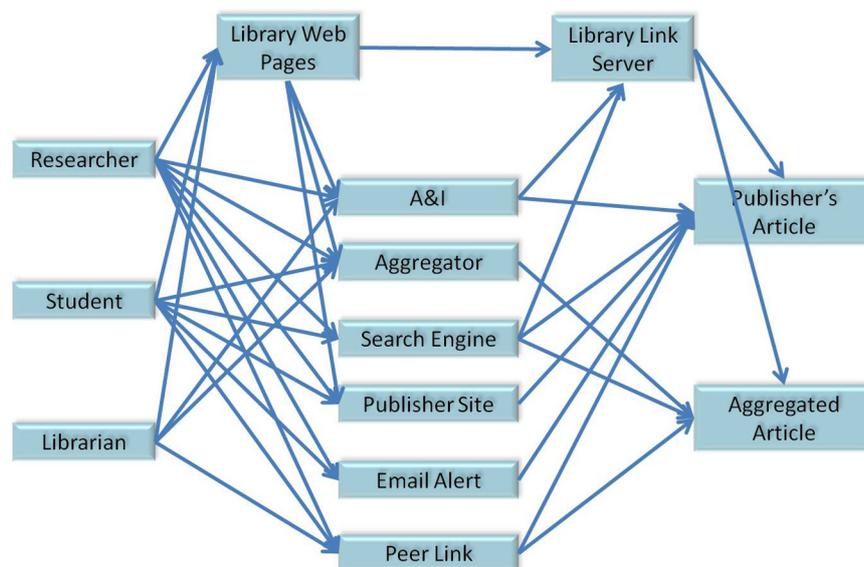

Figure 1. Possible navigation routes

SIMON INGER
Consultant
Renew Training

TRACY GARDNER
Consultant
Renew Training

The publisher website is the target for much of the content discovery and so the publisher gets to see how many readers come to them from within library-intermediated space, such as library web pages and link resolvers, how many readers come to them from abstracting and indexing services (A&Is), how many people come from search engines, bookmarked the publisher site or followed links in alerts or recommendations in peer e-mails. However, the

publisher does not usually know who those people are unless the individuals have got some kind of personalized presence on the publisher website. Publishers know in aggregate what these numbers are and which institutions they come from, but they do not know whether it was a researcher, a lecturer, or a student within that institution that gave a hit on their website (Figure 2).
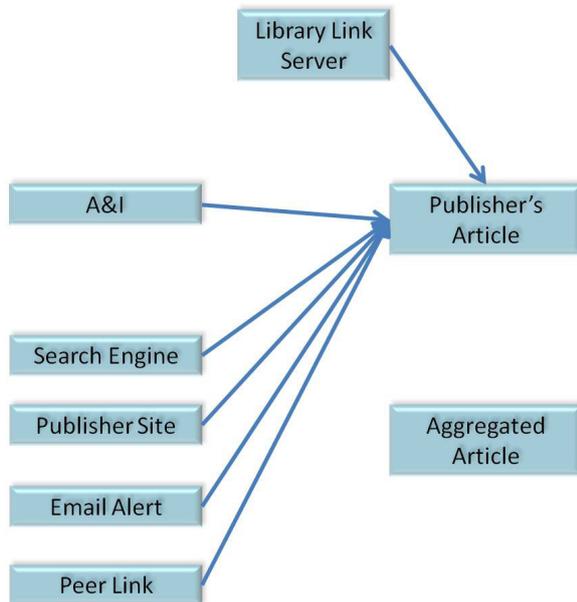
Figure 2. What the publisher can count

The library on the other hand gets a different picture again. They gets statistics, of course, from the publisher supplying COUNTER-compliant data, so they know how many of their readers ended up on the publisher website, but they do not know which of their readers it was, or other valuable information such as whether certain types of reader spend more time on the publisher website than others. If the reader chooses to go via library-intermediated space, then there is the opportunity to capture more detailed information, although not many libraries are actually doing that yet.

However, some libraries are starting to have to work out and attach a different value to each type of read-event by different groups of the people that they serve. At the moment, there is a disconnect between the need of the library to measure such usage, and the ability to do so, but, there are technologies coming along, such as Raptor, which may begin to provide some usage data to fill these gaps in our knowledge (Figure 3).

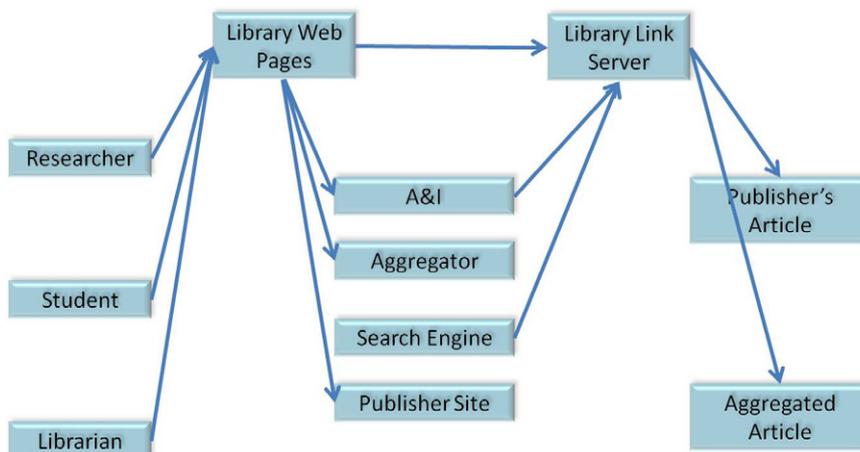"… there is a disconnect between the need of the library to measure such usage, and the ability to do so …"

Figure 3. What the library can count

Our research included asking a number of key questions about reader navigation, and many of them give us significant insight into how and when readers engage with library technologies. With such a large-scale survey, we were able to not only ask readers about how they believe they behave in aggregate, but also we are able to test their answers by asking them how they discovered the very last article they read. Figure 4 shows the results of this 'most recent article' question for academic researchers, broken down by the subject areas medicine, life sciences and social sciences. Researchers in life sciences say they spend nearly as much time using journal alerts as they do searching, whereas researchers in medicine spend way more time searching than they do following journal alerts. Figure 5 is another look at the same data, but this time making a comparison by job role: students, researchers and lecturers, in social sciences only. Students in social sciences use journal alerts much less than search, whereas for researchers and lecturers time is spent much more evenly between these two activities. While it is common for publishers to report that up to 60% of their traffic is referred from Google, what proportion of those referrals are actually researchers? Going further than that, what proportion of those referrals are readers who are licensed to access the content?
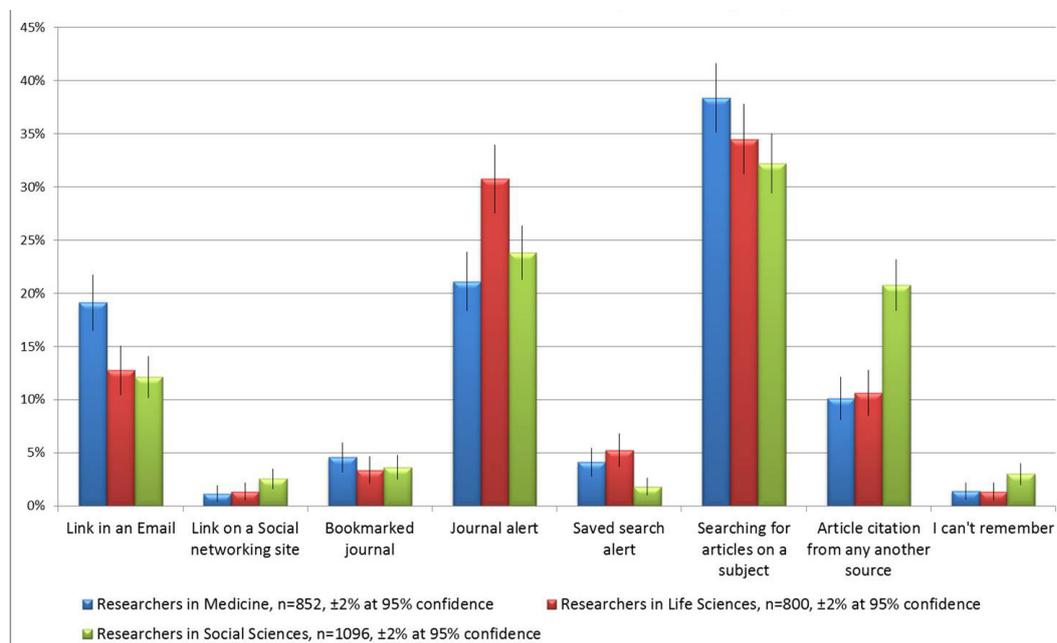
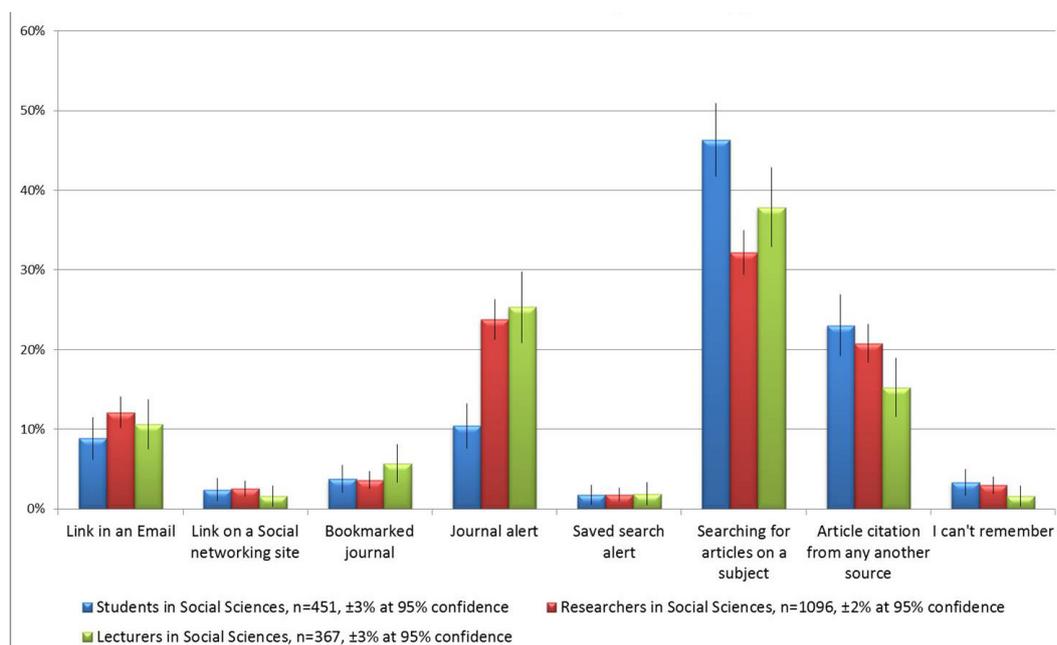Figure 4. Most recent article accessed – comparisons by subject

Figure 5. Most recent article accessed – comparisons by job role

The study also asked for reader preference for the discovery resources they use for search. Figure 6 compares the responses for academic researchers in life sciences, social sciences and physics. Life sciences show a large reliance on professional abstracting and indexing services, and are presumably influenced by the free-to-use PubMed service. Library web pages are much more popular in search for social scientists than for life scientists and physicists, and we see that behaviour replicated in other job roles, not just researchers. Similarly, aggregated content databases are much more preferred as a starting point for social scientists than life scientists and physicists. The most important starting point for social scientists appears to be an academic search engine, such as Google Scholar, eclipsing even the general search engine, such as Google, whereas this is not true for the other subject areas studied.
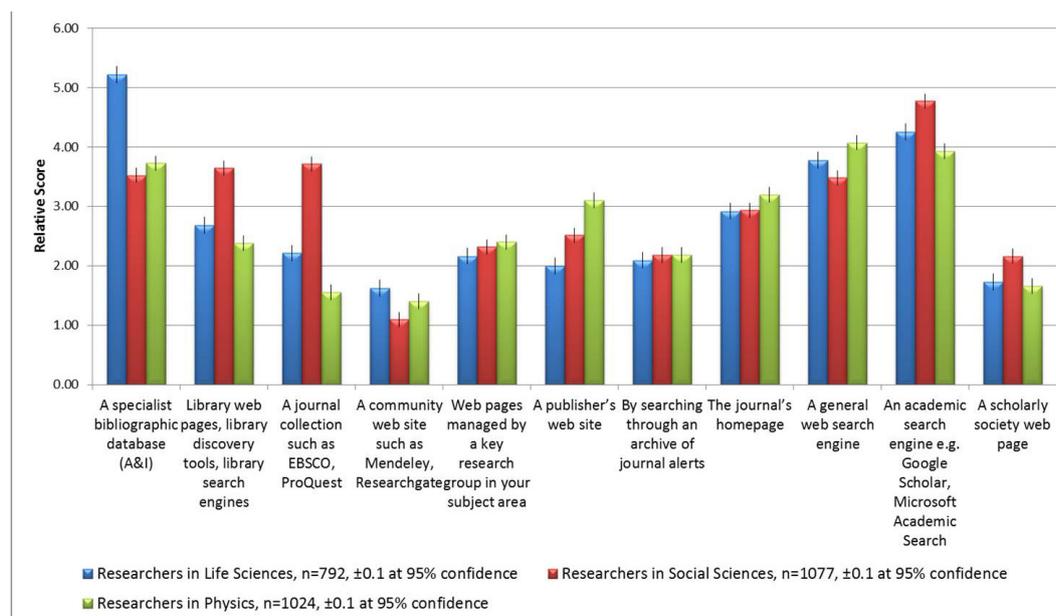

Figure 6. When you need to do a search for articles on a specific subject, where on the web do you start that search? Comparisons by subject area

Over the seven-year period of the three studies undertaken, we can see an increase in popularity of library web pages in search, but still most important overall is the professional A&I service. Library web pages have not shown the same upward trend when it comes to tracking down the content of the latest journal issues in readers' core title lists. In this regard, journal table of contents (TOC) alerts have seen a downturn in popularity, although deeper in the data it is evident that while this is true for most subject areas, chemistry is the exception. In chemistry, TOC alerts have stayed particularly strong when it comes to discovering latest journal articles (clearly the function for which they were invented). Publisher investment in their own websites seems to be paying off as more people seem to have bookmarked the journal homepage, or the publisher's website, and just go there straight away rather than going through any other discovery route, such as TOC alerts, in order to view the latest articles for a title. (See Figures 7 and 8.)

Library web pages differ in popularity substantially by subject area. Figure 9 shows us the relative importance of library web pages by subject area. It seems that library technology is most popular in education research, humanities, and political & social sciences. Physics & astronomy shows much less reliance, with the other hard sciences not far off. How should libraries respond to this? Should they work harder to engage users in those areas with lower counts or, of course, it might be best to ignore these groups because they have already found better options.
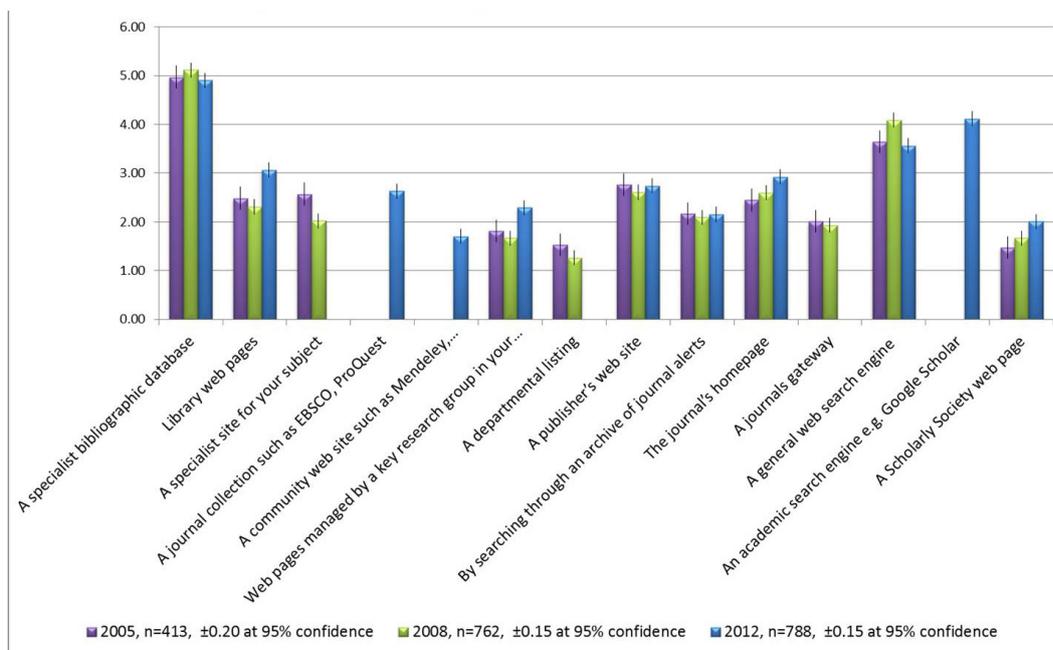
Figure 7. Starting points for searching for articles – trend from 2005 to 2012
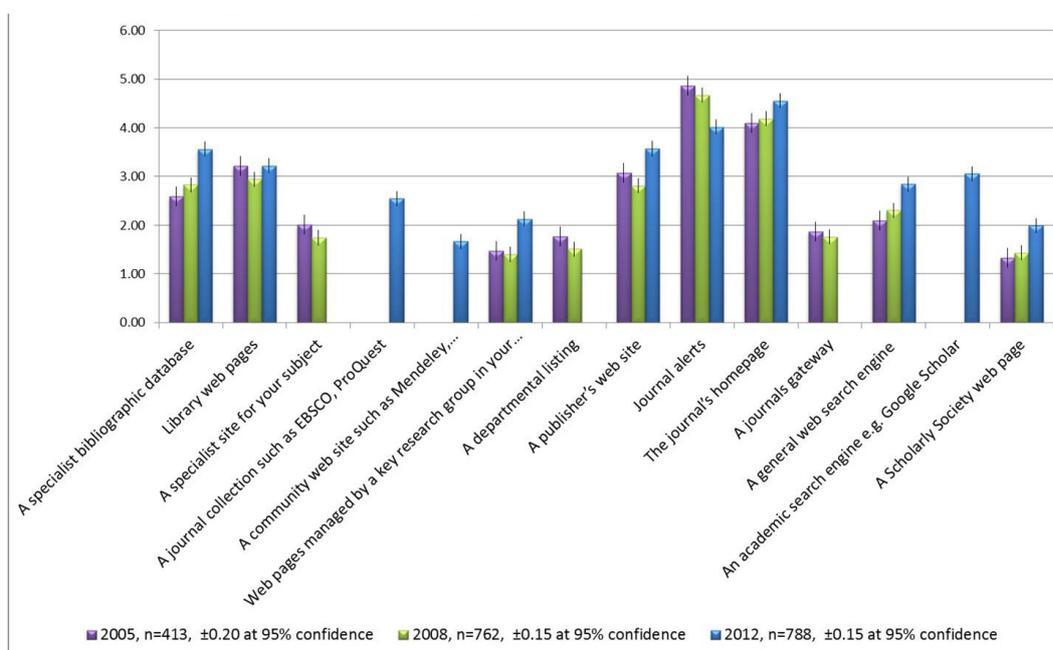

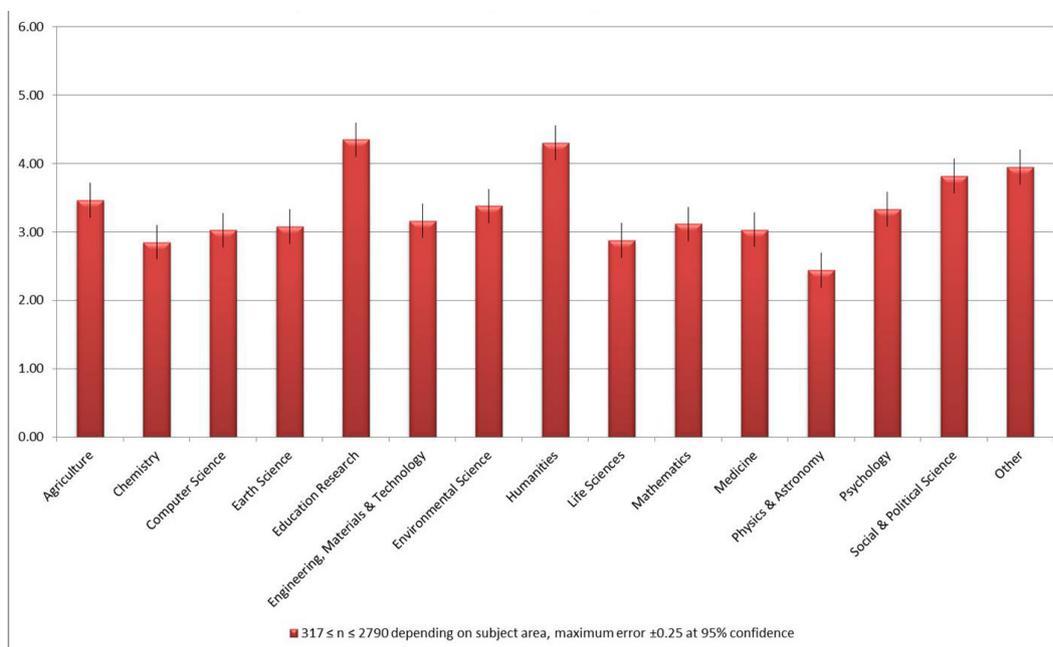Figure 8. Starting points for discovering latest articles – trend from 2005 to 2012

Figure 9. Relative importance of library web pages in search by subject area

There are, of course, other library-purchased discovery resources that readers are using when searching for articles as well. In Figure 10 we look at the relative importance by subject area for library web pages, abstracting and indexing services, and full text aggregations. For humanities, we see that full text aggregations seem to be more important than library web pages, which themselves are more important than A&I resources. Whereas, in medicine, we have the A&Is showing as more important than the library web pages, which in turn are more important than full text aggregations. (It can be argued that the A&I result for medicine is influenced by the presence of PubMed, which is not library-purchased, but rather is an open discovery resource.) As previously noted, workers in physics & astronomy seem to use library web pages very little, but the result is even lower for full text aggregations in this subject area. That may be a reflection of the amount of physics content in aggregations or it may be that people have other resources, such as arXiv, that they prefer to tap into.
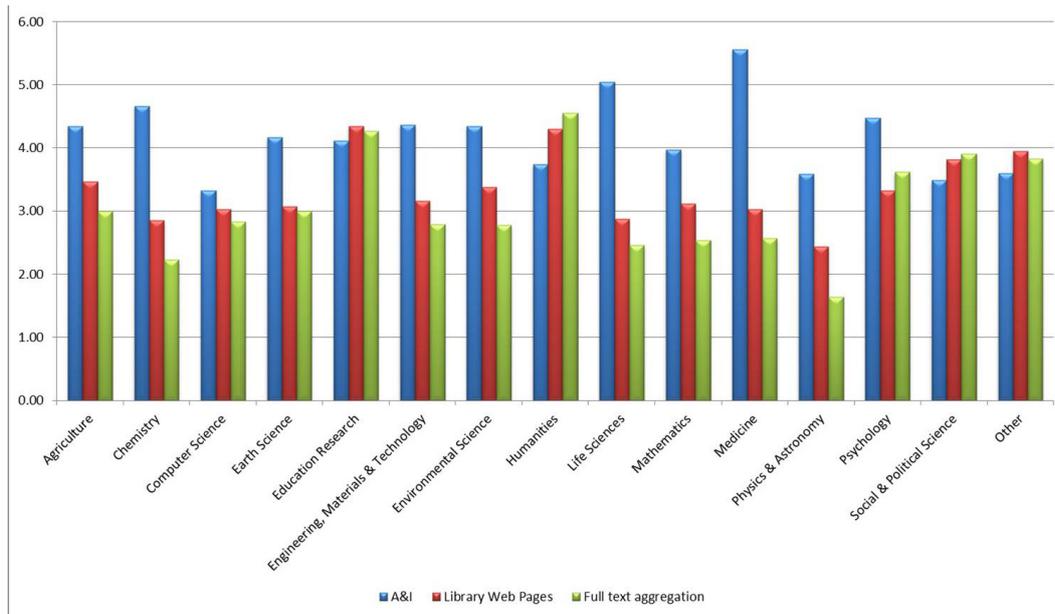

Figure 10. Relative importance of library-purchased discovery resources in search by subject area

The survey also asked respondents to say how often they perceived library technology assisted them or affected them in their navigation to content. Figure 11 shows the impact of library technology for the top 15 countries (by number of responses). It shows the proportion of people who said that library technology affected their navigation more than half of the time. The chart shows that countries such as Malaysia have a highly visible library technology presence, but there may also be other factors at work here. On the one
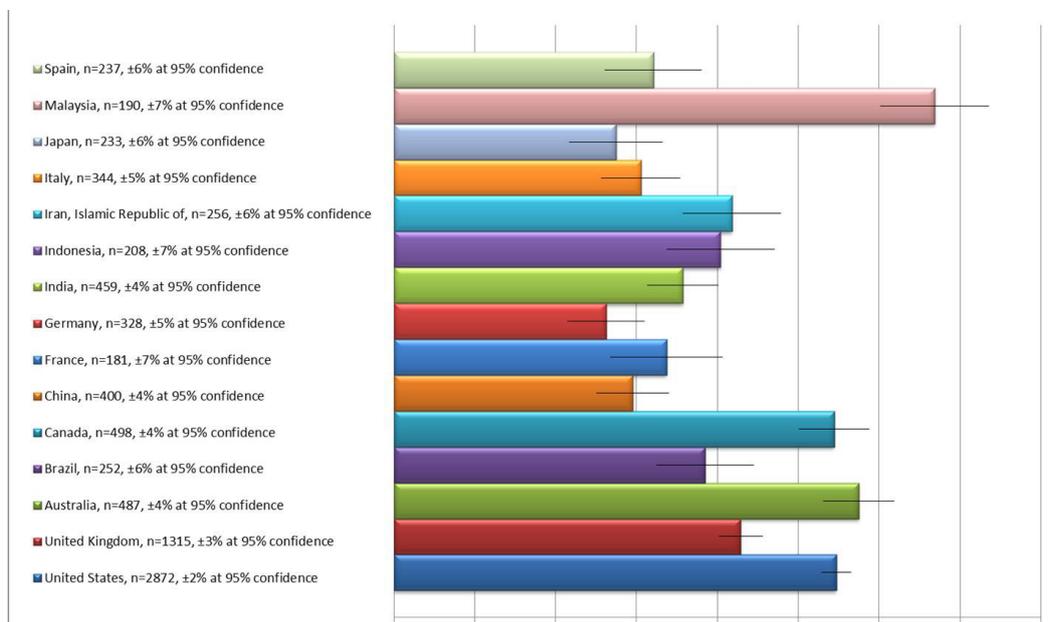

Figure 11. Library technology in navigation – regional variations

hand, this measure will be influenced by how many libraries have actively implemented discovery technologies and link servers locally and what impact that has had on navigation. But against that, some libraries perhaps make their technologies less obtrusive deliberately. One of the main categories of competitor to library web pages as a starting point for readers is the search engine, both Google and Google Scholar in particular (Figure 12). From the study, we calculated the relative preference for Google over Google Scholar by subject area. Google Scholar seems to be more popular for social science, psychology and education research than Google, and Google more popular for many of the remaining subject areas. Physicists say they much prefer Google over Google Scholar. There are lots of possible reasons for this, but perhaps the most compelling is the concept that although Google Scholar covers the academic material when physicists want to perform a search, they actually want to search a much broader base of material than just the academic articles – it could be data sets and other content which are indexed by Google but not indexed by Google Scholar. Conversely, one might well argue that social scientists really want to stick to the content indexed by Google Scholar because doing a general search on typical terms used in social science within Google might give you too many answers of unrefereed content.
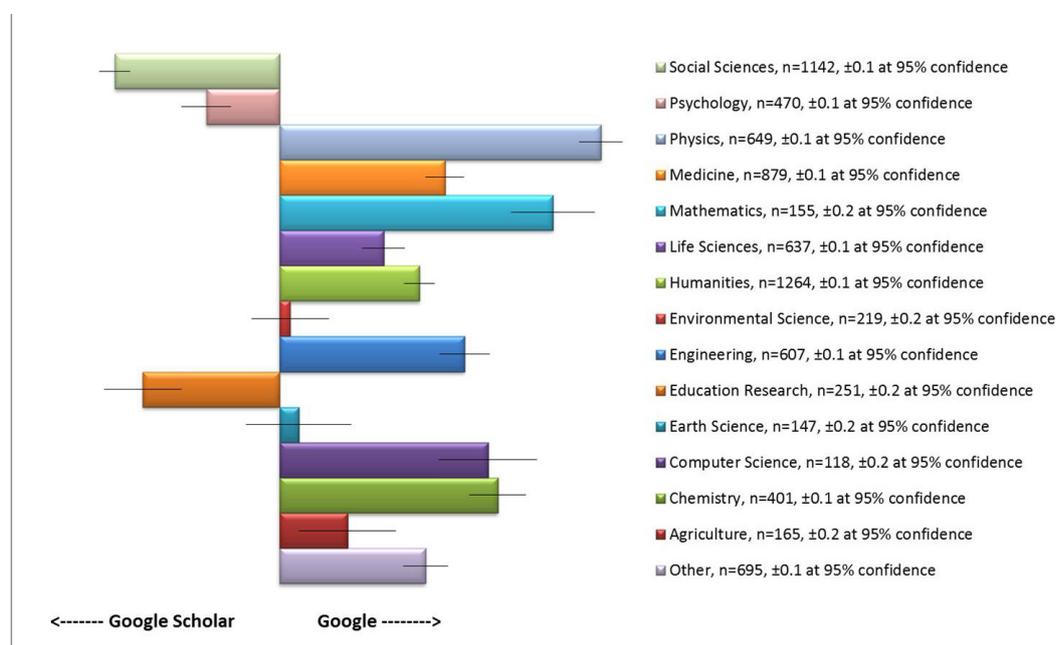

Figure 12. Relative use of Google over Google Scholar by subject in higher income areas

Conclusions

Library web pages are growing in popularity for search in many subject areas, but other tools still seem preferable to library patrons, particularly in some subjects, such as physics and life sciences. Libraries need to decide whether they should better engage with some subject areas that under-utilize library discovery or indeed, if the opposite is true, perhaps libraries should concentrate on those subjects where library technology is already providing the strongest service and leave the other subject areas to continue to use alternative discovery methods, not using the deployed library technology as much.

Publishers need to understand which content discovery resources readers use to discover content, and then actively work with the discovery resources of relevance to them. There are significant subject, sectorial, geographical and job function variations in the discovery data we have studied. Journal platform providers have their role to play in this too. As the technology providers underpinning a lot of those publisher websites, they are often the distributors of metadata to third-party indexers. Platform providers need to ensure that content discoverability, beyond standard search engine optimization (SEO), is at the forefront of their service offerings to publishers.

127 Library technology providers perhaps should address their underperforming subject areas. Is it coverage, awareness or better competing discovery tools in those subject areas that is the problem? Why is it that physicists are happier working in other search engine environments than web-scale discovery systems?

Illustrations all by courtesy of Renew Training

References

1. Gardner, T and Inger, S, *How Readers Discover Content in Scholarly Journals*, 2012, Renew Training: www.renewtraining.com/publications.htm (accessed 15 May 2013).

Corresponding author: Simon Inger, Consultant | ORCiD: 0000-0002-6367-6986
Renew Training
E-mail: simon@renewtraining.com