Driving usage – what are publishers and librarians doing to evaluate and promote usage?

Based on a breakout session held at the 34th UKSG Conference, Harrogate, April 2011

Although a relatively recent phenomenon, measuring the usage of published research has rapidly become one of the most important ways to evaluate the relative value of different publications. Libraries and publishers are also investigating the impact of interface and technology provision in improving resource discovery and content usage. Demand for such data is increasing throughout the industry, partly in response to greater scrutiny of return on investment. As a result the techniques used by publishers and librarians to promote and evaluate usage are also developing. This paper looks at some of the methods currently adopted and examines the issues faced by the industry in driving forward the application of usage data.

SARAH PEARSON
E-Resources & Serials
Co-ordinator
University of Birmingham

CHRISTIAN BOX
Senior Publisher R&D
IOP Publishing

Introduction

The rapid nature of the switch from print to electronic has been well documented and a great number of changes have been experienced across the industry. One significant change has been the rising importance of the measurement of online usage. Whilst demand for the resulting usage data has historically been strongest from the library community, there is growing demand from authors, government bodies, funding agencies and readers. It has become abundantly clear that usage is now one of the most important measures across the entire academic research landscape.

The library perspective

For the library, analyzing decisions on collection development is not new or unique to electronic content. For years, libraries have been attempting to tie in book loans data with acquisitions records to look at the amount spent on material that is never borrowed (which does not, of course, mean that the material has not been read in the library). If a library is lucky enough to have a streamlined way of loaning journal volumes and issues, then it is possible to track loans of journals too. However, the same caveat applies on 'loans' versus 'use'. Of course, some have resorted to dust-checking print journals on the shelf for evidence of use – not exactly a cast-iron metric! Increasingly, print journal collections are becoming reference only as the electronic version becomes the main method of access. Therefore, online usage metrics have become much more important in allowing a library to monitor the success of their purchases and their collection development policy overall. These indicators are the first really concrete way of establishing whether a journal has been 'used'. Of course, what it does not tell us is what the journal has been used *for* in terms of research outcomes and return on investment for

the university (an area that Carol Tenopir has focused on extensively)[1]. Usage metrics such as the COUNTER JR1 also do not tell us how a user arrived at that full text (i.e. straight to the publisher website, or via a link resolver, or via Google, or PubMed, etc.).

What we are doing at the University of Birmingham is, at a very local level, looking at our eLibrary portal as a starting point and evaluating the traffic around the site to see where it is coming from and going to. Why are we doing this? Well, we are looking at the value of our eLibrary service as we prepare to procure and implement other library tools to support user access to electronic content. We are also keen to see whether the importance we place on eLibrary as a *starting* point for our users is borne out by evidence. Alongside this we are investing much effort in usage analysis of journal, e-book and database content on the publisher platform to ensure that we are making the most of our Information Resources Fund and to assess where there are gaps that might need filling.

Promoting usage

Each institution has its own unique set-up for providing access to a range of university resources for staff and students and the library needs to plug their interfaces into other university services such as the portal and the VLE as well as external sites in order to achieve maximum reach. Traditionally, library software tools have conflicted with the external authentication arrangements required by content providers. This has led to confusion over convoluted log-in arrangements to electronic content, which has alienated users. Librarians have invested a lot of effort in improving this in recent years and the advent of new resource discovery service layers has made access much more seamless and embedded. At the University of Birmingham we do many of the obvious things to promote electronic content, such as putting new resources on news feeds and our eLibrary home page, and promoting it through training sessions and meetings with academic departments. However, we also place emphasis on embedding links and integrating content with other sites. We are very committed to keeping our link resolver knowledge base as accurate and up to date as possible, and export MARC records from this to our library catalogue. We embed deep links to e-resources within our VLE so that students can access specific 'quicksets' of content from outside

of the eLibrary. We also embed our link resolver (SFX) as a source in all our A&I services and within Google Scholar. We try and make authentication as seamless as possible between the eLibrary and other university services and have a single-sign-on service within the eLibrary which logs into the portal itself, provides Shibboleth log-in and logs in to EZProxy if a user is detected off campus. As a result users should never need to log in again on that browser session. Finally, we have customized a search applet for SFX to also include e-resources from our metasearch and print content from our catalogue and are taking advantage of our metasearch product's central index to provide a quick 'supersearch' option. We have embedded this applet in our university portal page and in our Facebook page. (See Figures 1 and 2.)

A key inhibitor for us at present is the lack of a vertical search resource discovery service to enable us to increase this integration further, and this is something that we are planning to implement within the next 12 months. However, at present our philosophy on access to library-subscribed content is to embed links to content through whichever means available, get authentication right and let the user decide on the best route for them.

Analyzing content usage

We do invest a lot of effort in evaluating usage of our e-journals, e-books and non-full-text databases. We have decreased our administration effort through subscribing to tools such as Selection Support, which collects usage on our behalf for the larger platforms. We also use the Journals Usage Statistics Portal service (JUSP) developed by JISC, which provides some good custom reports for our NESLi2 packages. The aims for us in analyzing usage of our subscribed content are to review cancellation of poorly performing content; validate new subscription performance; identify issues such as single user access and turnaways; and overall, to demonstrate value for money of the Information Resources Fund. In terms of the 'big deal', we aim to make full use of cancellations allowances to save money and increase value; make full use of substitutions allowances between subscribed an unsubscribed content to increase value; and annually evaluate the merits of big deal (unsubscribed) usage over having individual subscriptions. Much of this work is still done manually and many institutions struggle to provide enough staff resource to support it.

Figure 1. University portal and eLibrary

Figure 2. eLibrary applet

Analyzing the eLibrary service

We have a very active and well-informed E-Resources Steering Group which analyzes and tests improvements to our interfaces and recommends rollout to live services as well as having a role in reviewing collections and usage. It has been six years now since we implemented our link resolver and metasearch tool. However, we have customized functionality quite a bit; especially in embedding e-book searching and customizing the GoFindIt applet to embed a 'SuperSearch' which is using the Primo Central quickset within MetaLib. We implemented Google Analytics on top of our metasearch engine and link resolver so that we could look at some basic information about our users and traffic through the site as well as sources and direction of traffic. We have not invested a lot of money in a commercial product to assist with this but have relied on Google Analytics and the excellent skills of Edward Craft in our Digital

Library Team in setting this up. We have a channel on Google Analytics for our eLibrary homepage which is where a user typically enters our metasearch product.

We also have linked channels for our link resolver and metasearch portal alongside the homepage.

So, what are our main findings and how have we responded to them?:

- the majority of access to eLibrary comes from referring sites rather than direct. Most popular is the University portal with 50% referrals. Second is the Library homepage and third is the Library catalogue, where there are links to the eLibrary homepage, and from individual e-journal catalogue records
- in addition to the above, 15% of referrals are from search engines
- approximately 15% is 'direct' traffic. This could be as a result of a bookmark link typed into the address bar, or the referring site having blocked Google Analytics from recording the referral
- there is a high 'bounce rate' from the eLibrary home page because users are then accessing the Metalib and SFX servers which have different

domains and so they are not actually all bouncing off again! The bounce rate within the link resolver and metasearch pages are much lower and indicate that users are spending some time within the eLibrary service before moving on to publisher platforms

- We are able to see which browsers are being used and also the percentage of mobile devices accessing the site as well as the geographical location
- in our metasearch portal channel, goals have been set up for the number of users using specific parts of the site, for example 'Find Database', 'Quicksets', 'Find DB Results', and those taking advantage of personalization features
- we have implemented a new goal for the use of a specific quickset – our 'SuperSearch' quickset – to analyze the extent of usage that this gets as a percentage of overall quickset use
- on our link resolver, the direct traffic percentage is very high because users are going through our web.sso authentication, which blocks referring site URLs. For referring sites, the eLibrary homepage is highest (including
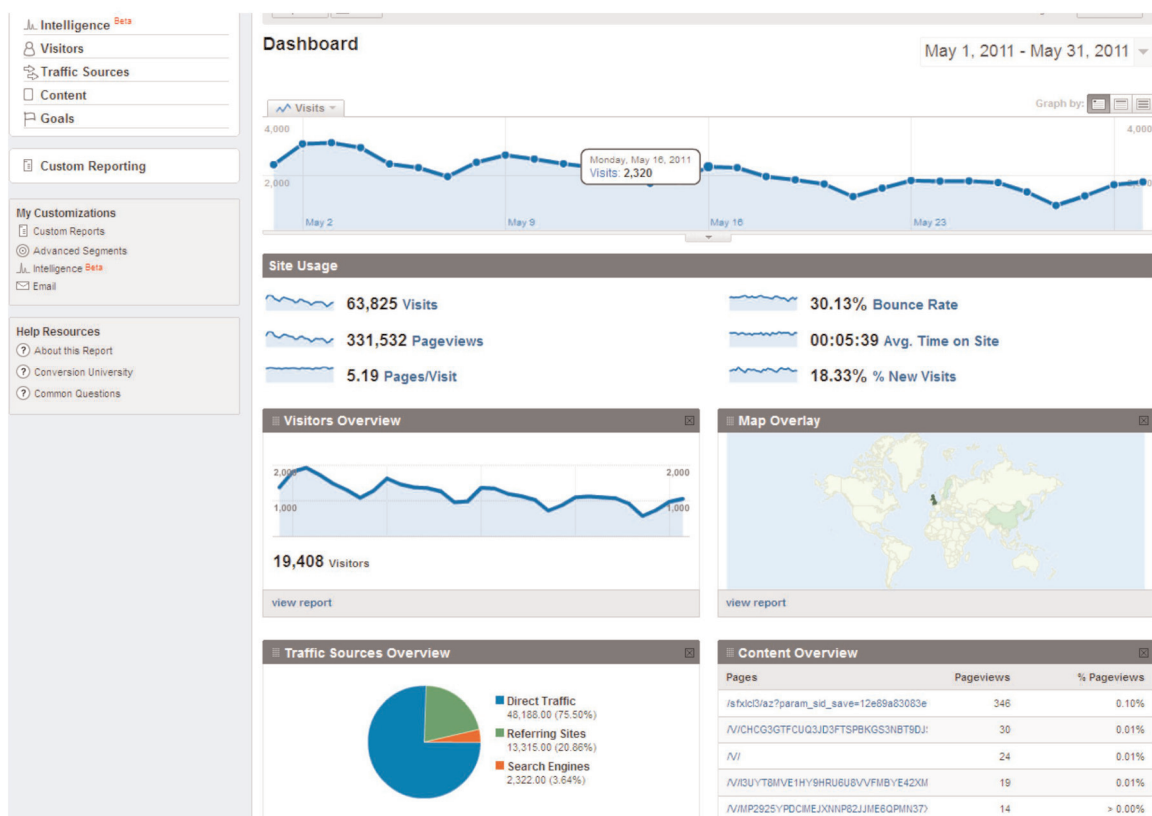

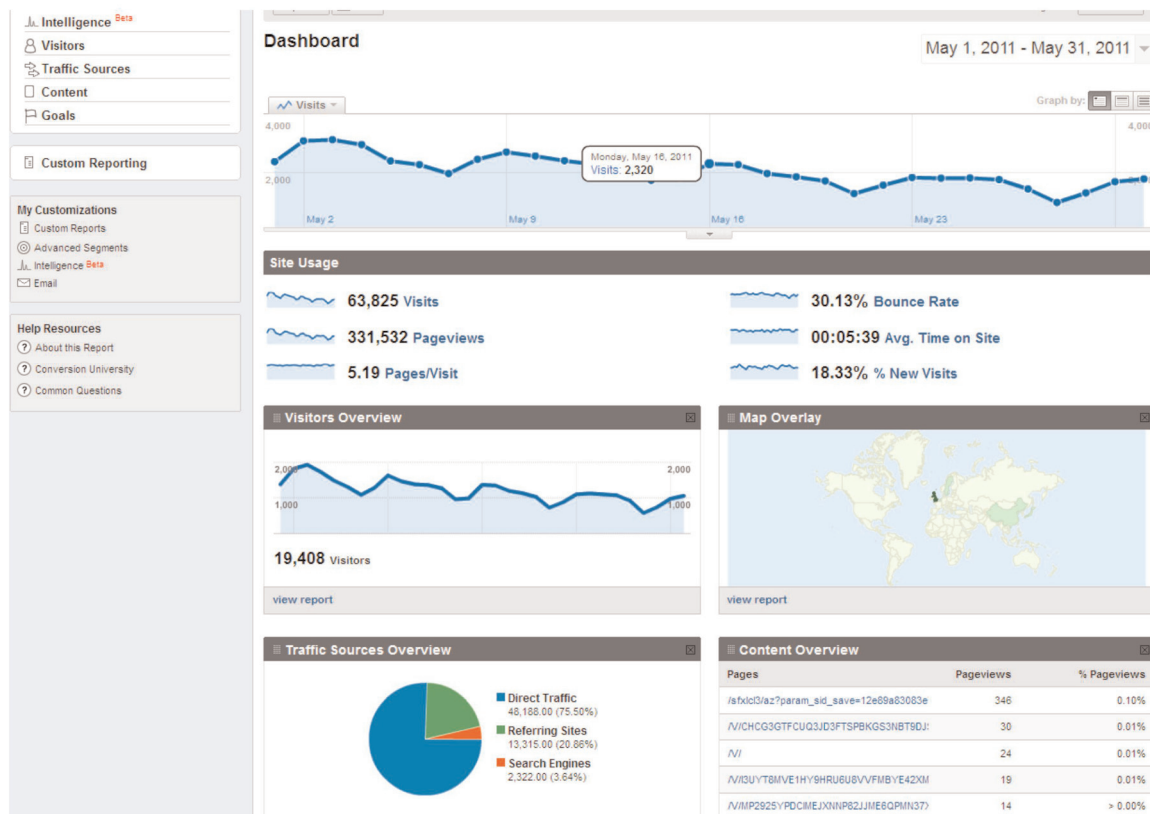*Figure 3. eLibrary Google Analytics dashboard*

Figure 4. Google Analytics goal conversions

our GoFindIt applet on the university portal). However, there is a high percentage of referrals from third-party A&I databases, such as Web Of Knowledge, which indicates that our SFX service, being embedded as a source within those databases, is utilized

■ Google and Google Scholar are reporting a lot of referrals. Again, we have embedded our link resolver as a source.

These are just some of the findings that we are starting to see from our use of Google Analytics on our eLibrary service. Members of staff have also set up the same analytics on other web pages, such as the Library catalogue, and we hope to build on this with some analysis between sites. We believe that the measures described above are invaluable in justifying our resources budget expenditure; evaluating the success of new functionality and recommending adoption of new services.

The publisher perspective

Project COUNTER added significant momentum to the increased use of and demand for usage data,

and the advances it made in standardizing production and presentation have clearly been a significant step forward. Of the many aspects that project COUNTER brought to the industry, one of the most interesting, was that it *shared* data between publishers and librarians. In theory, the resulting transparency drives us towards being a more efficient industry where the supply of published research should be in line with demand from readers in the research communities. I can think of few other industries that share this to such a blatant and open extent.

However, experience tells us that it is not quite so straightforward, and the variation across subject areas means that if all decisions were made purely on absolute usage, then some research communities would be unfairly disadvantaged. As we know, there is also a range of other metrics used throughout the industry that should be considered alongside measures of usage, but I think it is now widely accepted that usage is one of the most important.

Operationally, this means that every publisher needs to be highly skilled not only in driving usage of their content but also in interpreting and analyzing the resulting data effectively. One of the

publisher's key roles is to ensure that they are making authors' research as discoverable as possible, and so high usage can be seen as a measure of success.

Techniques that publishers adopt for driving usage can take many forms and it is essential to think about the success factors of different activities in different ways. There are broad-brush 'macro' approaches aimed at driving overall discoverability of content – a classic example being search engine optimization (SEO). But there are also more directed and subtle 'micro' techniques aimed at specific communities, using technological changes or marketing campaigns.

As a publisher, we at IOP have adopted a range of approaches to drive usage over recent years. In fact, it is probably fair to say that almost all of our activities are directly or indirectly and completely or partially aimed at driving usage. The following five examples are a selection of just some of the approaches we have used over the last 12 months.

The first and most effective of them all is *search engine optimization*. This may seem like a dated concept and a dark art which is difficult to influence, but a brief glance at any website's usage data will show just how important it remains. There is no doubt that after our content was indexed by Google in 2003, we saw the largest single increase in usage. SEO is a continual process though, and we have recently engaged in a fresh audit to ensure that we remain current. It is essential that this is constantly monitored.

It is difficult to imagine that there will be another factor as influential as Google in driving usage, but the growth of *social networking/media* does offer some potential. The success factors from social networking are complex and clearly not as straightforward as just increasing full-text downloads, but the size of the audiences and methods of communication present a great opportunity for driving usage. As a result, this is an area where we have become more and more actively engaged and is likely to be a prominent area in stimulating usage in the future.

As a truly global industry, we have also looked to widen the reach of our content by launching *local-language websites* focused on engaging communities; initially in Japan, China and Latin America. Early levels of web traffic have been moderate, but growing. These websites offer a different type of visibility and we are yet to fully determine how effective they are, but the increasing flow of traffic suggests that they will prove to be a good method for increasing usage.

On the technological front, the recent growth in the use of *mobile devices* has been significant. This presents a great opportunity to drive usage by offering an alternative means for users to access content and address one of the researcher's biggest challenges today – time. We have launched a number of apps and a mobile site, both of which have seen strong demand and use. Currently usage from mobile devices comprises a relatively small proportion of our overall usage but we see wide variations between content types – particularly journal versus magazine content – and the increased take-up of tablets is likely to change this again. This is without doubt a key area for change over the short and medium term.

Finally, a slightly different approach has been the addition of *video abstracts* to some of our journals. As the name suggests, these are videos submitted with an article, allowing the author an additional medium to describe their research. The responses have been very positive but one particular piece of feedback encapsulates how this kind of addition can be crucial in stimulating interest in research: *"By featuring the people behind the science, video abstracts have the potential to convey inspiration and enthusiasm, and thereby the significance of scientific results, beyond the concise text of articles."*

I think that the ability to add enthusiasm and inspiration to published research will be a vital feature of how science is presented in the future, and we have already seen that it is a feature in driving usage.

The combined effect of this activity creates an abundance of data and, as an analyst, I love this! However, it does present a number of challenges. The sheer volume of raw data from our journal platform, and the wide range of interested parties both internally and externally, mean that we have to take a well-structured approach to delivering information. We have developed a suite of usage analysis techniques that we broadly group under three headings:

- reporting and management information
- web analytics
- data mining.

The reporting and management information elements (which include COUNTER reports) are essential for keeping our customers and our

business informed of the usage of our content. However, in this area we are generally concerned with individual usage events, i.e. an abstract view, a full-text download. Whilst this is useful, we realized that we needed to develop a far more in-depth understanding of usage.

I think there are three key advances that we have made from that point. Firstly, was the need to think beyond just full-text downloads. Even simply adding together downloads and turnaways to give 'full-text demand' gives a very useful alternative metric, but also considering requests – successful and unsuccessful – to all content on a site builds a much deeper understanding of usage. Where applicable we also adapt full-text download counts to show an alternative view, an example being that we now calculate a usage half-life (in a similar vein to the citation half-life published in the *Journal Citation Reports*) to gain a better understanding of how current research is. By focusing purely on full-text downloads you dramatically limit your understanding of usage.

Secondly, where possible we try to understand usage in terms of visits and visitors rather than individual page views or 'usage events'. Understanding usage at the visit and visitor level

provides a much richer view. We have taken this a step further and created customized metrics to measure desired visit outcomes such as 'number of full-text downloads per visit' and 'number of visits that contained a search'.

Understanding complete visits also opens up a range of opportunities in path analysis. Standard path analysis is useful for understanding how users navigate between individual pages. However, when looking at broader patterns we have developed a system that maps paths by page type.

This allows us to understand, for example, the difference in subsequent usage when users land on an abstract page rather than the full text of an article, and how their resulting page-type path differs. We can then segment based on referrer type, customer type or content type to see if there is variation.

Thirdly, we analyze everything in context. In virtually all applications that we now use, we have incorporated some kind of data-model extension to ensure that we are looking at actual customer information and actual content groups rather than IP addresses and URLs. By integrating offline data sources such as our customer and content database we are able to build a much richer understanding
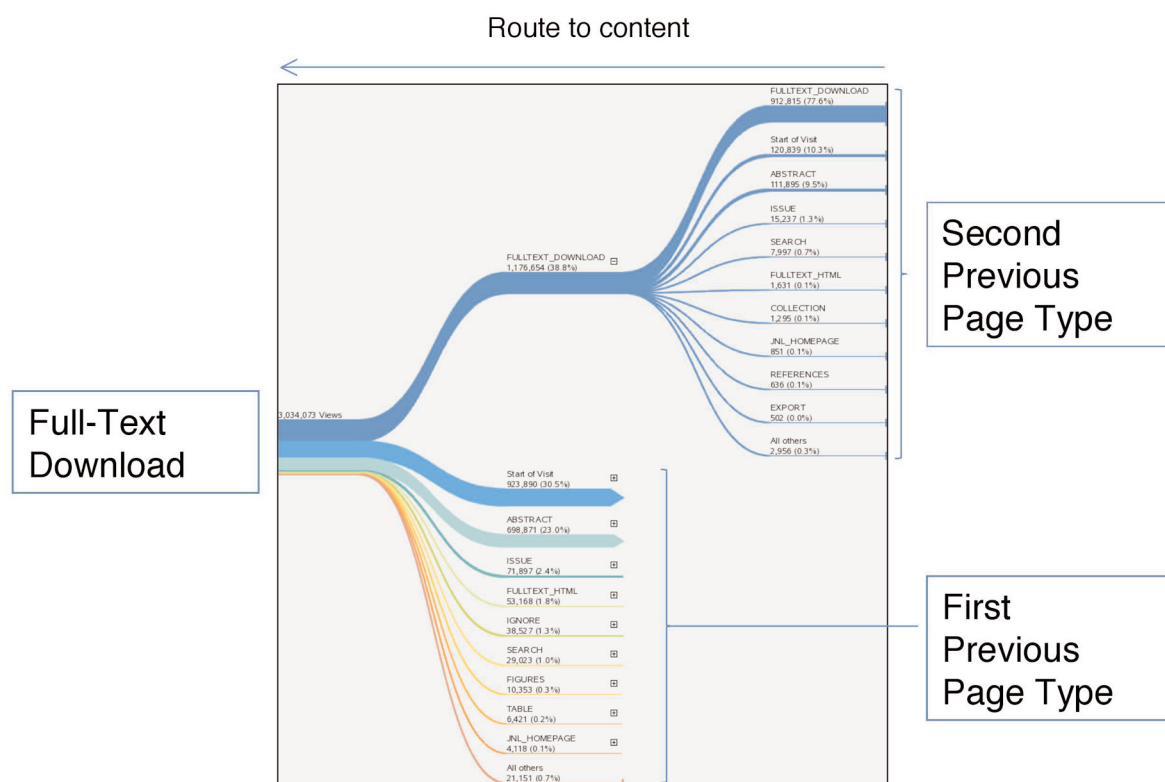

Figure 5. Page-type path to a full-text download

NB this image is included for illustrative purposes only and is not intended to be fully readable
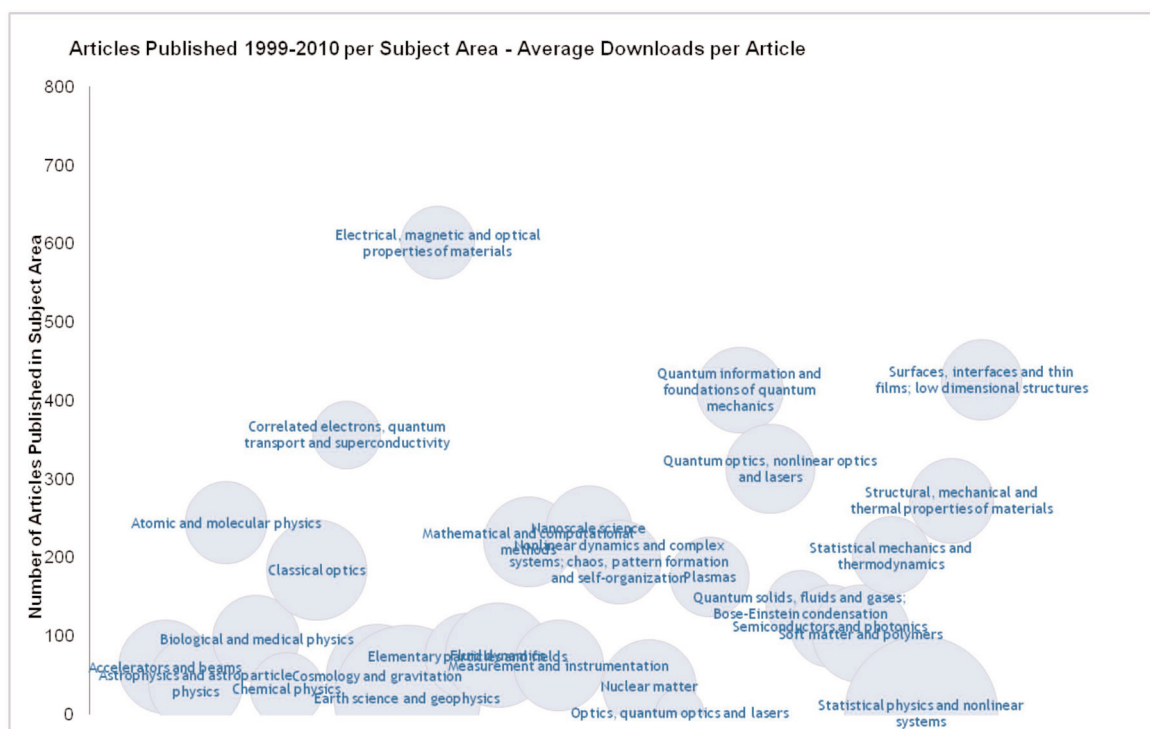
Figure 6. Usage per subject area for a single publication

of how different users interact in different ways.

In all areas we take a highly visual approach when sharing the data and ensure that the delivery is appropriate for the end user. A great example of where this has been invaluable is in trying to 'visualize the scientific landscape'. To enable this, we built a view of the number of downloads per article for different subject areas driven by the article-level Physics and Astronomy Classification Scheme subject identifiers.

The resulting maps enabled us to identify the areas that were driving usage for a particular journal and also to identify growing and shrinking areas of 'demand' beyond the constraints of raw data files.

Conclusion

So from an analytical perspective, I think we have made huge strides forward in our understanding of usage over recent years and we have learnt some important lessons along the way. Most notably, that the way users access content constantly changes and it is vital that your thinking stays current. We have also learnt that it is essential to understand the limits of what usage data can tell you: whilst it can provide insight into areas that were previously difficult to understand, it is not a replacement for the wide range of other metrics and knowledge that exist throughout the industry. However, clearly a failure to recognize its importance carries significant implications.

Reference

1. King, D W, Tenopir, C, Choemprayong, S and Wu, L, 'Scholarly Journal Information Seeking and Reading Patterns of Faculty at Five U.S. Universities', *Learned Publishing*, April 2009, 22 (2): 126–144. DOI: 10.1087/2009208

Article © Sarah Pearson and Christian Box

■ **Sarah Pearson**
E-Resources & Serials Co-ordinator
Library Services
University of Birmingham
Edgbaston
Birmingham B15 2TT, UK
Tel: +44 (0)121 414 5825
E-mail: s.pearson.1@bham.ac.uk

■ **Christian Box**
Senior Publisher R&D
IOP Publishing
Dirac House, Temple Back
Bristol BS1 6BE, UK
Tel: +44 117 930 1159
E-mail: christian.box@iop.org