UKeiG
UK eInformation Group

eLucidate

Vol. 13 Issue 1, Spring 2016

ISSN: 1742-5921

eLucidate's editor is Gary Horrocks email: gary.horrocks@gmail.com

eLucidate is published by UKeiG, the UK eInformation Group. Membership and subscriptions are managed by the UKeiG Administrator, Val Skelton, 4 Friendly Street Mews, London, SE8 4DU, tel: 020 8691 0246, e-mail: val.skelton@ukeig.org.uk

UKeiG is a Special Interest Group of CILIP: the Chartered Institute of Library and Information Professionals. CILIP Registered Office: 7 Ridgmount Street, London, WC1E 7AE, Registered Charity No. 313014

Contents

Editor's Note

Welcome to eLucidate.

There is some fascinating and thought-provoking reading for you in this issue. Karen Blakeman argues that, while knowledge of alternative search tools is vital, it is no longer sufficient to have advanced search expertise; that information professionals require savvy appraisal skills and a better understanding of how search outputs are generated and manipulated. She also raises the issue of the "right to be forgotten" and the impact that is having on excluding potentially critical information from search results.

Michael Upshall introduces the great semantic Diffbot debate, yet another initiative to extract meaning from the web. But does it?

Martin White, playing devil's advocate, also touches on "search" asking the pertinent question: do intranet managers really need an information architecture, or is click and search the way to go? Search-driven versus menu-driven? A useful debate to revisit.

UKeiG has addressed many issues impacting on research activity over the years: most notably Open Access, copyright and intellectual property, but also search tools and advanced information retrieval skills. A key theme of this issue is scholarly digital publishing and research impact evaluation tools. The Library Research Services Team at the University of Hull provide a practical overview of bibliometrics and the associated commercial and public domain e-resources utilised to support this activity. Bibliometrics has immense value in academia and across all sectors, but has drawbacks as well as benefits. Altmetrics - the impact of social media on research visibility (who is talking about YOUR research?) – is featured and is one theme we will revisit in more depth in a future issue. Michael Upshall also touches on social networking, and highlights a key report about changing resource discovery behaviour amongst researchers. How do scholars find articles in 2016? Is traditional A&I on the decline? He also provides oversight of scholarly publication processes, and charts the rise of the library as publisher.

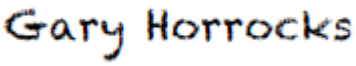
UKeiG was proud to be one of the supporters of the first LISDIS (Library and Information Science Dissertations) conference, held at the University of Huddersfield on 14th November 2015. The organisers explain the rationale behind the event and highlight some of the themes from the day. In future issues we hope to support this initiative by showcasing the fascinating e-information research undertaken by students across the UK LIS education sector.

In a time of escalating international crises, with the systematic destruction of cultural heritage in conflict zones, Chloe Menown, the recipient of UKeiG's Early Career Professional conference grant, reflects on the destruction of libraries. She left me thinking that there must be a greater urgency to drive forward digital preservation through advanced imaging technologies and digitisation projects. She also touches on the de-

professionalisation of staff in public libraries, raising how this could exacerbate the digital divide; the e-information have and have-nots. Again, food for thought for future issues of elucidate.

I'd like to thank all of our contributors. Each article is rich in cross-references to useful follow-up documents and more detailed web content. The UKeiG management committee is keen to receive your feedback and to hear your recommendations for future articles. We'd be delighted to hear about any projects or research you have underway. Drop me an email if you'd like to contribute any article. Notes for contributors are included at the end of this issue.

Enjoy!



Editor – Elucidate
gary.horrocks@gmail.com

Search 2016:
Human not Artificial Intelligence Needed

Karen Blakeman, RBA Information Services

karen.blakeman@rba.co.uk

One of the advantages of running search workshops is that it forces me to keep up to date with changes and new developments. For many people, weird search results are something that they have to cope with but do not necessarily have time to investigate. They are providing information and support to other people, and once one project is over another takes its place. There is very little time to look into why the search engine - often Google - is behaving bizarrely. For me, keeping abreast of what the search engine companies are doing is a large chunk of my job and what I discover is sometimes disconcerting and worrying.

Single, small changes in algorithms build up over time to effect bigger changes in the way a search is analysed, processed and presented. For example, Google recently stopped showing advertisements to the right of the results on desktop search. This was not altogether unexpected since Google and its competitors have been steadily moving towards a single, simplified interface that works on all types of devices. Don't think though, that there will be less advertising. Google is already pushing extra advertisements to the top of the search results, which means scrolling down further to get to the more reliable results. And it is all too tempting when using a mobile device to click on the first vaguely relevant link.

It is not just the advertising that one has to be wary of. A major trend with all of the major search tools is to offer "facts" and quick answers, extracted from one or more websites, both at the top and to the right of results. No need to click through to a document to find the answer to your query because it has already been found for you. The problem with these “facts” is that the source is not always given and the overall quality appears to be going downhill rapidly. Run a Google search for court fees for the UK small claims procedure and you'll probably see a four-row table that starts with a fee of £205 for claims up to £5000. A note at the bottom of the table tells you that there are five more rows, the implication being that there are higher fees for higher levels of claims. When you click through to the web page there are in fact more rows at the top of the table showing lower rates for claims below £5000. This is not an issue for those who take the time to click through to the website to see the full table, but those who do not could be deterred by Google's answer from pursuing their claim.

Where the search engines can seriously start to go wrong is when they combine information from multiple sources. Examples pop up with alarming regularity. Type *"frugivore"*, for example, into Google and it tells you that cats, lions and killer whales are examples of fruit eating animals! Some errors, such as this, are obvious howlers. It is the almost-but-not-quite-right answers that are potentially dangerous. Clicking through to the source, if it is provided, and double-checking the information with another site is time consuming but vital if one is to be sure of the accuracy of the data.

The most significant, recent development in search is the use of artificial intelligence (AI) as part of the mix. Google, in particular, has invested heavily in AI and in October 2015 confirmed that RankBrain is now an integral part of its web search. As if to further emphasise its commitment to AI John Giannandrea, who had been leading the company's research into artificial intelligence, took over as Senior VP of search in February when Amit Singhal retired. There is much discussion as to how the AI component actually functions and some have attributed the increased variability in the quality of results to its activities (Google: RankBrain Doesn't Use New Signals But May Adjust Weights Of Existing Ranking Signals.) What is certain is that all bets are off when it comes to predicting how our searches are likely to turn out. Knowledge of advanced search commands will help to a certain degree but critical appraisal of what pops up on the screen is now more important than ever.

The impact of new technologies on research is a big enough headache in itself, but we now also have to consider recent developments in the so-called "right to be forgotten" legislation. To summarise what has happened to date: an individual in the EU/EEA has the right to request a search engine to remove links to information about them from search results generated by that search engine. Under the EU legislation, this affects any search engine that is based in the EU/EEA. It is up to the search engine to decide whether or not to comply with the request taking into account public interest as well the concerns of the individual. If agreed, the information remains on the original website but it is not visible to those viewing search engine results on a European version of the search engine. Until now there has been an easy way to circumvent the restriction, which was to use a non-European version of the search tool, for example Google.com. This is no longer possible for those of us identified as being located within Europe.

To comply fully with legislation, Google has announced that it will now "use geolocation signals (like IP addresses) to restrict access to the delisted URL on all Google Search domains, including google.com, when accessed from the country of the person requesting the removal." See Google Europe Blog: Adapting our approach to the European right to be forgotten and Search Engine Land: Google Agrees To Complicated Worldwide "Right To Be Forgotten" Censorship Plan has summarised it thus:

"Assume that someone in Germany files a Right To Be Forgotten request to have some listing removed for their name. If granted, the censorship will work like this for searches on that person's name:

- Listing censored for those in Germany, using ANY version of Google.
- Listing censored for those in the EU, using a European version of Google.

- Listing NOT censored for those outside Germany but within the EU, using non-European versions of Google.
- Listing NOT censored for those outside the EU, using ANY version of Google."

The obvious way around this is to use a VPN or proxy server that gives you an IP address outside of Europe. For many people this will probably not be an option. Alternative search engines such as StartPage.com and DuckDuckGo may be another solution, but is there a problem anyway? Does it really matter if links to some stories about an individual disappear? Yes, it might. Some of my research work involves due diligence on companies and individuals, and on two occasions I have discovered information that had been excluded from European searches as a consequence of the right to be forgotten. In one case, the extra information was deemed non-essential in the context of the enquiry but in the other it was critical. For both, I found the information by searching directly the databases and sources that held the original data. The data is there but not accessible via a general web search tool, again highlighting the danger of over-reliance on Google et al.

It is no longer enough to know how to use advanced search commands. We also have to understand how the results are generated and manipulated, and the restrictions that may be imposed on the output. Knowledge of alternative tools and the relevant, primary sources is vital. It is not artificial but human intelligence that is needed in 2016 and beyond to find information and appraise it so that it is fit for purpose.

Extracting Meaning from Web Content

Michael Upshall, Consult MU

Michael@consultmu.co.uk

For many years, experts have attempted in vain to extract meaningful information and content from web pages. HTML is a dumbed-down language, holding almost nothing of any meaning (except for generic headings such as H1, H2 and so on, that will vary in meaning from one site to another); so any attempt to improve the communication of meaning from a web page is usually based around adding additional codes and vocabulary to HTML to make it more structured and meaningful (schema.org is an example).

Now Diffbot claims to be able to identify site content better. According to a post in Marketing Land (perhaps not the most reliable source), Diffbot has gained $10m in funding (quite likely) because it is "creating semantic Web content – that is, information that is characterised by its meaning – even though the page hasn't been formatted in that way." This sounds very unlikely, in fact it sounds like magic.

Certainly Diffbot does some basic things with a Web page. It separates a Web page into pictures, title and story – all of which HTML does already. The magic is Diffbot's other initiative: the creation of a "Global Index", a collection of knowledge (or graph database, a rather fashionable term, although there isn't a graph anywhere in sight) that will be searchable. Their goal is "to categorise most of the business-valuable information on the Web".

The graph database is what is used to attempt to classify the web page, along the lines of if an article contains the terms "bridge", "trump" and "trick", then the chances are it is about bridge the card game rather than bridge the civil engineering structure. Most likely, although Diffbot don't reveal their exact methodology, the tool will be cleverer than that and will use what Amazon terms "statistically improbable phrases" – phrases that occur very often in a particular type of book, enabling Amazon to recommend other books (or content) like it, containing the same SIPs. Why statistically improbable? As one blog response points out, this means that, although the terms "magic" and "London" occur frequently in Harry Potter novels, other phrases such as "Hogwarts" and "Hermione Granger" are less likely to occur in non-Harry Potter titles.

All very creditable, but does it work? I tried putting one of my own blog posts through Diffbot's "Test Drive" page. It added two labels to my story: "publishing" and "library" – hardly rocket science. To be fair, it did recognise me as the author.

Why is Diffbot special? Well, they have a very impressive PR department! A glance at the headlines from the media suggests that the world outside believes Diffbot is something special, with comparisons to Google and Intel:

- TechCrunch: *Don't Read The Comments—Let Diffbot Analyse Them Instead*
- Xconomy: *Could a Little Startup Called Diffbot Be the Next Google?*
- GigaOm: *Diffbot Aims to Convert the Web Into One Big Database, One Page at a Time*
- VentureBeat: *DiffBot's New API Brilliantly Reveals What's Hiding Behind Any Link*
- Wired: *Diffbot helps machines to read web pages like humans*
- Wall Street Journal: *Investors Back Diffbot's 'Visual Learning Robot' for Web Content*

How successful is Diffbot?

There is little explanation on the site, but the achievements of Diffbot appear to be focused around a limited number of Web pages, such as product catalogues and retail sites. The nature of these sites makes it easier to extract some kind of understanding from them – it's not too difficult to work out from an online store Web page which are the products.

Diffbot makes claims for the success of its system by comparing their performance with other competitors (AlchemyAPI, Embed.ly, Goose, for example), which shows them having an improbably high F1 score of 0.94 (the F1 score measures precision - how accurate the results are, with recall – how well the tool finds the links. It is generally assumed that two different human indexers looking at the same material would score no more than around 0.8 on this measure, since humans disagree on what is a correct link). However, in the small print of this comparison it would appear that the quoted F1 score measures only the extraction of title and text from a web page. You could say if you were being uncharitable that any bot could do that.

Intranets:
Do You Really Need an Information Architecture?

Martin White, Managing Director, Intranet Focus Ltd

martin.white@intranetfocus.com

My career has been shaped by the staff of the library in Devizes allowing a 10-year-old schoolboy access to the adult section. Not only was I given six precious Browne book tickets I was also given a guided tour of the library. The librarian asked me about my hobbies. At that time I was a keen railway spotter, with the GWR main line just thirty minutes away in Swindon. He then revealed a deep mystery of libraries. Books on railway matters could be found in the section on Transport and also in the section on Engineering. As time went on I found other examples but my knowledge of library classification was immensely helpful until I reached the Library at the University of Southampton which used the Library of Congress classification and I had to start all over again.

The challenge of finding the "right" place for a content item is one that webmasters and intranet managers will be very familiar with. Many hours, and much patience and frustration, will have been devoted to organising the information architecture (IA) of the intranet, probably using card sorting. Always the outcome is a compromise, so usability testing is then carried out to justify the decision and persuade others that in the interests of time they should accept the consensus. For years I have been fascinated by the fact that IA designers see any attempt by a user to use "search" as indicating a failure on their part to optimise the IA, resulting in further work and delay to address the apparent failure. I should at this point highlight Chapter 9 in Information Architecture for the Web and Beyond, which sets out in fifty pages exactly why search is a core element of an information architecture. This is the 4th Edition of what is widely known as the Polar Bear book (in reference to the animal O'Reilly publications have included on the cover.) Every information professional should have their own copy. Of course the problem with any IA is that the moment it is launched something changes in the organisation to render it obsolete. Even if only a small element has to be changed the overall integrity of the intranet will probably be lost.

I am writing this column shortly after returning from the annual IntraTeam intranet conference in Copenhagen. Over the last decade this has become the meeting point for intranet managers from the Nordic area and beyond and has a community feel to it that the Intranet Now event in London has emulated with great success. (The 30th September has just been announced as the date for the 2016 Intranet Now event). This year at

IntraTeam over 210 delegates turned up. One of the themes this year is that the majority of the intranets on display were search-driven and had only a minimal menu system. It is easy to recognise a search-driven intranet – the home page has a large search box and little else. Two good examples were described by Philips (Netherlands) and DNV-GV (Norway). Typing the name of a person into the Philips search box brings up contact details as a preview using predictive search, without the need to press a search button. This does not need magical technology. DNV-GV uses SharePoint 2013 pretty much out of the box.

There are a number of advantages to this approach. The first is that an item of content only needs to be stored once, along with associated metadata, and can be displayed whenever the user needs it in response to a query. Secondly, it is very easy for a wide range of other applications and repositories to be integrated into the intranet. This enables Google-like search cards to be presented to a user who is looking for details about a customer. Contact details, billing status, office locations and the names of the sales team responsible can be garnered from a number of different applications and presented in a standard format. A third benefit is that it is possible to customise the presentation of news, corporate documents and other information for staff in specific roles and locations without the need to have very complex IA designs.

Although the technology is well established the requirements for high quality information management are perhaps higher than a conventional web architecture. The metadata has to be fit for purpose and so has to take into account the way that the organisation works and the business language that it uses. If someone queries [EMEA Sales] just which countries should be included in the European, Middle East and Africa acronym, or is there a pre-integrated sales report? It is not uncommon in global organisations to find that South Africa is not regarded as EMEA. User research still has to be undertaken but the focus has to be on how people work and not in the information they claim to be looking for. This is where personas have a very important role to play, but they can never summarise the requirements of every single employee. That is why search is so important. Either it can be used to deliver highly customised information or provide an effective application for people with very specialised needs.

Search-driven intranets are not for everyone, if only because it is far too common to find that there is no skill base in search to work with in their development. I cannot stress enough that these skills are not IT skills but information professional skills in applying metadata to business requirements and then tracking the performance of the system in many different ways to optimise the user experience and the business impact on the organisation. As well as a focus on search-driven intranets there were also very good examples of enterprise-wide search from PwC UK and Astra-Zeneca. It was interesting to note that this year the search-focused presentations were allocated the main conference room rather than one of the smaller side rooms, a further indication of the increased level of interest in search. This might be prompted by Google exiting the enterprise search business, leaving companies to find alternate solutions at fairly short notice.

The Digital Faculty:
What's Happening in Scholarly E-Publishing

Michael Upshall, Consult MU

Michael@consultmu.co.uk

This article explores some issues and developments in scholarly publishing revealing a complex landscape of online activity with multiple stakeholders – authors, librarians, publishers, researchers - all with different agendas, goals and activities.

How do scholars find articles?

Years ago it was simple. As a researcher, you walked into your institution library and consulted the journals relevant to your subject. After browsing the contents, you read a few articles, and your work was done.

Such a straightforward situation sadly no longer exists. Firstly, there are too many journals to search by hand! According to the 2015 STM Report, the number of researchers is growing by around 4%-5% per year, and the number of journals published increases each year by around 3.5%; in the past few years the rate has increased further, to around 6.3% since 2003. The reality of publishing today is exemplified by the journal PLOS ONE, which publishes around 100 new articles each day; and that's just one journal.

How do researchers find the content they are looking for today? They have several possible routes: they can get to a scholarly article via a library website, via a publisher's website, via an abstracting and indexing (A&I) service such as Inspec, or by Google (usually Google Scholar). Which do they prefer?

Simon Inger and Tracy Gardner have for several years carried out a major survey, How Readers Discover Content in Scholarly Publications, a study of over 40,000 users, with the latest report published just a few weeks ago in March 2016. Their findings are revealing. The most common discovery methods, in order of priority, are:

- The A&I service
- An academic search engine (Google Scholar)
- A general search engine (typically Google)
- Library web pages
- The publisher's website

Although A&I services have declined in popularity, they remain the single largest starting point for academic searches.

Since this survey has been running for some years, Inger and Gardner were able to look at trends over time, and they comment: "Whilst A&Is are marginally the most important search resource, their importance has consistently dropped since 2008. General search engines have lost some ground here to academic search engines ... All search resources that are under publisher control – publisher website, journal alerts, journal homepage and society webpage – have made gains."

They mention that social media tools such as Academia.edu and Mendeley are playing a steadily increasing role in article discovery, although in overall terms they represent a small proportion of total discovery.

This interpretation was criticised by Roger Schonfeld of Ithaka S+R in a Scholarly Kitchen post. He pointed out that Inger and Gardner carried out their survey using names supplied by publishers. These people had both registered at the publisher site and had agreed for their names to be used for further activities (they had "opted in"). Schonfeld suggested, with some justification, that such users were likely to be self-selecting as users of publisher systems. It was not surprising, therefore, that the publisher stats looked so high. Publishers will of course want usage of their platforms to be as high as possible. So the moral here is to be careful before interpreting results about what the assumptions behind those results might be. Who carried out the research, or more fundamental still, who selected the sample? Each party in the process has a vested interest in maximising their role.

What do Scholarly Publishers do?

This may seem a strange question, if you have worked for a scholarly publisher. But for anyone coming to scholarly publishing for the first time, compared with educational or trade (bookshop) fiction, the scholarly publishing business is quite different. Perhaps the most fundamental difference between trade publishing and scholarly publishing is that the authors aren't paid for their work. On top of that, the publisher doesn't do the selection of what is published – that is the work of peer reviewers, academics who review the work of others.

So, to summarise:

- The academic author writes a paper (free of charge)
- The paper is peer-reviewed (free of charge)
- A publisher collects peer-reviewed articles in a journal, then
- The publisher sells the journal to an academic institution (for lots of money, paid for by the very institutions where the authors work).

Delays in Publication of Published Articles

Not only does it appear that scholarly publishers earn a lot of money without paying any royalties, but they also (if you want to be uncharitable) take a long time to publish an article!

Daniel Himmelstein, a research biologist, asked why it took so long for his article to be published. On investigating many journals, he found (revealed in his blog) that the review time – the time between submission and acceptance of a scholarly paper – is typically around 100 days, and has been for the last 30 years. Worse still, figures from individual journals show that the review time for many of them has increased (at Nature it is 150 days).

Himmelstein followed up his research with a further blog post, where he classified delays into a) acceptance delays (the time between a proposed article being received by a journal and acceptance by that journal) and b) publishing delays (the time from acceptance to publication). Acceptance delay is around 100 days, as before, and publication delay has been reducing - it is around 25 days).

Overall, the time to publication, including both acceptance and production time, has not been substantially reduced since the advent of digital publishing, especially if you allow for submission to multiple journals. As a result, researchers are increasingly looking outside the standard scholarly publishing system, to alternative systems for making article available. Life scientists, for example, looked enviously at ArXiv, the dominant pre-press system for physics, following which bioRxiv for life science was launched 2013. Many of these pre-press archives are managed not by publishers but by not-for-profit institutions (a research laboratory for bioRxiv, Cornell University for arXiv).

Given the above, it's not surprising, therefore, that there are many calls for publishers to be replaced by libraries and institutions doing their own publishing. An example is a recent report by Ann Okerson and Alex Holzman (July 2015) published by Washington DC's Council on Library and Information Resources, which looks at libraries as publishers – not just of library catalogues, but of original scholarly works.

Okerson and Holzman go on to estimate how much it costs to publish one scholarly paper; they find a 2015 estimate that suggests the cost of creating a scholarly monograph to be $12,000, and they suggest this cost should be borne by institutions where the author works.

Of course, there are already some key examples of libraries acting as publishers. The initiatives they cite include some that would be described as publishing, Project MUSE, for example, (founded 1993, a collection of 600 peer-reviewed academic journals and 20,000 ebooks) and some that you might not define as publishing at all – more online content platform provision, like HighWire. It is not clear if they are suggesting that libraries are involved in the management of HighWire ("library staff have made appropriate

contributions to High Wire's activities" – what does this mean?) I would guess that HighWire runs independently of Stanford or of any other library. Whatever its achievements, it is not an example of libraries acting as publishers.

So you could say, if scholarly publishers don't pay authors, and they don't choose the articles that are published, then what do they do? Kent Anderson, a former scholarly publisher himself, has been publishing a regularly updated blog for some years to answer this and other questions. In his latest, 2016 version, he lists 96 tasks carried out by a scholarly publisher. They are grouped into five areas:

- Editorial
- Marketing
- Community
- Technology (e.g. metadata tagging, XML conversion, social metrics)
- Finance and business

His is certainly an insider view, the view of the publisher - he talks about "we" publishers, and it is certainly true that publishers will feel comforted and encouraged to learn they are doing so many things in the course of publishing scholarly articles, and that far from being superfluous, their job is important and undervalued. But on looking more closely at these 80 activities, many of them are statements of intent rather than actual activities. Of course publishers vary widely in the services they provide – many publishers are more active in broadcasting to their authors than listening to them and establishing a community (the third of Anderson's five categories). In fact, John Sack, head of HighWire, stated at the recent Academic Publishing in Europe Conference in Berlin "Do we as publishers want to support the full interaction around the knowledge, or do we want that happening somewhere else?" There is an increasing recognition amongst publishers that much of the activity around scholarly publishing passes the publisher by.

But for the moment, perhaps that doesn't matter. Perhaps the best summary of what scholarly publishers do was given by David Nicholas of CIBER Research, at the same Berlin Conference mentioned above: "The main currency for the scholar is not power nor wealth, but reputation. However, reputation has been built upon one activity - research, one output – publication and one measurement - citation. It is a narrow view of reputation that has, so far, served publishers well."

In other words, the scholarly publisher provides the reputation on which an academic career depends. The publisher-controlled journals, for the most part, have the greatest reputation. It is generally agreed that an article that appears in Nature has greater kudos and credibility than if the same article appeared in pretty much any other journal. To gain that credibility, academics need publishers.

Bibliometrics: an Overview

Dr Michael Latham & Stuart Bentley
University of Hull

M.Latham@hull.ac.uk
S.Bentley@hull.ac.uk

Research support is an expanding area of activity for libraries in the HE sector. At the University of Hull, the recent reorganisation of Library and Learning Innovation involved a redistribution of expertise to meet the changing needs of the University, its staff and students. As part of this, a new Research Services Team was created to meet the needs of the research community and so contribute to a key strategic aim of increasing the quantity and quality of research outputs.

Research Support at the University of Hull

The Research Services Team supports the University research community in its broadest sense, meeting the needs of faculty staff and student researchers. Specific areas of support include Open Access (OA) publishing, Research Data Management (RDM), copyright, research resource discovery, and bibliometrics. Support is provided by different modes of delivery, including face-to-face teaching and advocacy, online learning tools and support guides, and other promotional initiatives.

Impact Factors

Impact factors measure how quickly and how often articles in a specific journal are cited by authors of other articles, allowing a comparison of how heavy an impact that journal has within a particular discipline.

The ability to assess the impact of research publications is a growing area of importance for researchers, in terms of securing funding, career development and research output dissemination strategies. Bibliometrics provides a quantitative analysis of the influence of research. It looks at the citation counts for articles to see how they have impacted on the research landscape and with the introduction of altmetrics looks at how the influence of the article can be measured. It stands alongside qualitative measures of excellence such as peer review. Available metrics can be differentiated as to what is being measured in terms of impact.

Authors and Institutions

Citation counts, publication counts, h-index

Articles

Citation counts, Facebook/Twitter mentions

Journals

Journal impact factor, SCImago journal rank, Eigenfactor

Subjects

Highly-cited papers, hot papers, journal impact factor

It is worth noting that citation measures are most developed for research in the sciences and social sciences. The arts and humanities, due to disciplinary differences, have fewer tools available for such analysis. It is also difficult to compare impact factors of journals in different discipline areas as citation practices vary between disciplines.

Bibliometrics have the following limitations:

- Not established for all disciplines
- Citation practices vary from one discipline to another
- High number of citations does NOT imply high value
- Potential manipulation, e.g. “group” citing, splitting research between multiple articles
- Coverage of sources other than journal articles can be poor

The following points should be born in mind when trying to raise a research profile:

- Use a consistent form of your name wherever possible
- Use an author ID system, e.g. ResearcherID, ORCID
- Ensure you include your institutional affiliation
- Promote your research via appropriate social media
- Use self-citation in a responsible way

There are a number of different tools available to collect bibliometrics, many commercial and some free to use. The free to use resources tend to be available on an individual registration basis; so can be recommended to researchers with the proviso that they will need to register themselves.

Web of Science

Web of Science is a platform provided by Thomson Reuters and includes a number of different databases and services that can help track bibliometrics. Two major products in the field of bibliometrics are only available through Web of Science: Journal Citation Reports (JCR) and Essential Science Indicators (ESI). These products have strong functionality for analysing bibliometrics - see below for more details.

Web of Science is multi-disciplinary but its bibliometric tools only cover titles in the sciences and social sciences. Also not all journal titles take part in JCR and ESI.

You can also get a quick snapshot from Web of Science itself, although you need to restrict your search to the Web of Science Core Collection. When the results are displayed, you will quickly see those papers which are identified in ESI as either Highly Cited Papers or Hot Papers. And if you click on the journal name, you will see a pop-up window which gives the 2-year and 5-year impact factors for the journal, as well as its ranking in the subject categories to which it has been assigned.

Journal Citation Reports compiles cited references to articles to measure the impact factor and citation rates at journal and category levels. It covers 10,800 journals in the sciences and social sciences from over 2,550 publishers in 232 disciplines from 83 countries. This breaks down to over 8,400 journals in the Sciences and over 3,000 journals in the Social Sciences edition.

An impact factor of 1.0 shows that on average the articles in the journal published one or two years ago have been cited once. It is also possible to include a 5-year impact factor and to look back at the impact factors for a number of years to see whether the journal's impact is increasing or waning.

Data in JCR is updated on an annual basis, meaning that in a given year, data for a previous year is added to the product.

Essential Science Indicators (access via Web of Science) ranks scientists, institutions, countries and journals in 22 specific fields across 12 million articles from 12,000 journal titles. It includes baselines that allow researchers to analyse the benchmarks by which research impact is assessed. The authors of papers are matched to institutions and the institutions normalised. Information in Essential Science Indicators is updated more frequently than JCR.

ESI identifies Top Papers, which consist of two categories:

- Highly Cited Papers – These are chosen from 10 years of data and measure citation on the basis of the field the paper was published in and when it was published.
- Hot Papers – These are papers from the past two years that show an unusually high rate of citation in the current period.

Scopus

Scopus is a citation database provided by Elsevier. It features metrics that allow researchers to assess journals, articles and authors, which are gathered at http://www.journalmetrics.com. The assessment is based only on journals featured within Elsevier's database.

The major tool for this is Compare Journals, which compares up to ten journals on IPP (Impact per Publication) and SNIP (Source Normalised Impact per Publication.)

Scopus data is also surfaced in SCImago, which has a journal rank indicator, and CWTS - see below for more details. Alternatively, you can get a quick snapshot for a particular journal by clicking on the journal name in the list of results of a search. You will be presented with the SCImago journal rank, IPP and SNIP for the journal in question.

The 'Compare journals' function in Scopus allows you to compare up to ten journals and view various metrics in either graphical or tabular format.
The SCImago Journal & Country Rank includes indicators for journals and countries using the Scopus database.

CWTS journal indicators is a source of free access to bibliometrics for scientific journals, and has been calculated by the Centre for Science and Technology Studies (CWTS) at Leiden University using the Scopus database to provide indicators for over 20,000 journals. The two main indicators available are:

- RIP (Raw impact Per Publication) - equivalent to Scopus's IPP (Impact Per Publication) - average number of citations per publication.
- SNIP (Source Normalized Impact Per Publication) - average number of citations per publication, corrected for differences in citation practices between fields.

Google Scholar

An alternative and free source of metrics is Google Scholar. To use the citation analysis option in Google Scholar, you will need the Publish or Perish software, which is available for free download at http://www.harzing.com/pop.htm. Publish or Perish interrogates Google Scholar for raw citation details which are then analysed for the following:

- Total number of papers and total number of citations
- Average citations per paper, citations per author, papers per author, and citations per year
- Hirsch's h-index and related parameters
- Egghe's g-index
- The contemporary h-index
- Three variations of individual h-indices
- The average annual increase in the individual h-index
- The age-weighted citation rate
- An analysis of the number of authors per paper

Altmetrics

A recent development in the area of bibliometrics in altmetrics. Altmetrics measure impact in a different way. They take into account the rise of social media in the raising of research profiles through the analysis of activity in online tools and environments. As such they can be helpful in discovering where research has affected policy or been discussed in the media and therefore assessing the social impact of research.

- Altmetric.com provide a free tool for gathering altmetrics on specific papers
- Almetrics.org have a directory of altmetric apps and a manifesto
- Altmetrics are also being incorporated into databases and you can find altmetrics data in Scopus and in CINAHL Complete (nursing and allied health) and in some journal platforms like Taylor and Francis
- CINAHL Complete also features information on social media mentions for some articles. Where this information is available, the "Plum Print" icon will show and you can click on this to get further details
- For some articles in Scopus, you will see additional citation information provided by Altmetric - this appears when you display the full reference for the article you are interested in

Do It Yourself

It is also possible, with a bit of work, to create your own bibliometrics. However, it requires a lot of time, powerful computers and a good mathematical brain. We don't do this at the University of Hull but there are organisations with their own bibliometricians.

Navigating The Bibliometrics Landscape

As can be seen above there are a number of applications of bibliometrics and tools for their collection within the HE sector and beyond. Uses can range from benchmarking the quality of research in a department or institution, to helping researchers further their career.

It is important to recognise that if you use the commercial tools, they are calculated slightly differently in each tool and usually only on the basis of journals that feature in their respective databases. Therefore it is possible to come up with different scores for the same journal, article or author, depending on the source used. If you have access to multiple tools, you will need to decide which result you will use or if you will try to use the results from each tool to come to a normalised score.

It is also important to remember that bibliometrics are open to criticism as primarily a quantitative measurement of quality and should be considered alongside other qualitative measures to get a round picture of the actual quality of a resource. Altmetrics, a rapidly developing area, are a possible source of such data and are likely to have greater importance as they mature.

It is a rapidly changing landscape and one to keep an eye on. As such you might want to join *LIS-BIBLIOMETRICS* on JiscMail for all the current news.

Acknowledgements

The authors would like to express their thanks to Christopher Moll of the University of Hull, who had input into the creation of our Bibliometrics LibGuide which formed the foundation of this article.

Glossary

- Eigenfactor: A rating of the impact of a scientific journal by measuring the number of citations to a particular journal. Within JCR, eigenfactor scores for all journals sum to 100. Developed by Carl Bergstrom and Jenin West at the University of Washington
- h-index: Suggested by Jorge E. Hirsch, this is an index for quantifying scientific productivity of an author. A scholar with an index of h has published h papers that have each been cited h times
- g-index: Suggested by Leo Egghe, this is another index for quantifying the scientific productivity of an author.
- ORCID: ORCID ID is a persistent digital identifier that uniquely identifies a particular researcher and can be linked to other identifiers. Researchers can register for free, though many institutions are implementing ORCID on an organisational level
- ResearcherID: ResearcherID is Thomson Reuters' unique identifier scheme, which is ORCID compliant and integrates with Web of Science. Researchers can register for free
- SCImago: a research group from the Consejo Superior de Investigaciones Cientificas, University of Granada, Extremadura, Carlos III (Madrid) and Alcala de Henares who have produced a Journal and Country Rank portal

LISDIS Conference

University of Huddersfield, 14th November 2015

Michelle Bond - Faculty Librarian, Liverpool Hope University and LISDIS co-organiser, *bondm@hope.ac.uk*
Jess Haigh – Subject Librarian, University of Huddersfield and LISDIS co-organiser, *j.m.haigh@hud.ac.uk*
Rosie Higman – Trainee Liaison Librarian, University of Reading and LISDIS co-organiser, *r.a.higman@reading.ac.uk*
Emily Wheeler – Library Learning Advisor, University of Leeds and LISDIS co-organiser, *e.wheeler@leeds.ac.uk*

Early on a dark Saturday morning in November 2015 librarians, information professionals and students from around the country battled through wind and rain to make their way to Huddersfield for the first ever LISDIS (Library and Information Science Dissertations) conference. LISDIS was developed, as we believe that too much librarianship and information science research is hidden away in dissertations and not used in practice. As newly qualified professionals we were aware of how much work goes into a Masters dissertation and were inspired after seeing excellent dissertation research presented at a New Library Professionals Network event. We hope that by showcasing a selection of research at LISDIS other information professionals will think about how we can better use this evidence base to improve information services and advocate for change.

The day consisted of a poster session, keynote speaker and three themed sessions with three speakers each. The first session on Collections and Discovery began with Sarah Hume's talk on the difficulties of classifying a women's studies collection using the outdated classification schemes available. It was a really engaging talk offering both abstract questions (if you cannot find yourself in the collection can you use it?) and practical solutions in the form of proactive cataloguing to improve the representation of historically marginalised groups. Lizzie Sparrow used ethnographic research methods to look at the use of a discovery layer in an academic library. Lizzie's research has already prompted some changes to the Senate House Library, London discovery layer exemplifying the impact that LIS dissertation research can have and which LISDIS aims to promote. Lucy Saint-Smith's talk on female book collectors in the 18th and 19th centuries closed the first session. Library history was a new area for many of us but Lucy brought it to life with an overview of four key collectors and insights from her extremely thorough research.

Later we heard about different aspects of Public Libraries and the Community. Ian Clark started us off looking at how the move toward community libraries run by volunteers is drastically reducing the ability of those libraries to address the digital divide. He highlighted how economic and educational inequalities can be reinforced once public services are removed, particularly once the professional expertise of qualified librarians is lost. After this overarching look at the effect of cuts to public libraries two speakers spoke about how public libraries are failing to serve particular groups. Alanna Broadley considered the provision of lesbian fiction in public libraries in Scotland, describing how collections can all too frequently focus on "classic" texts and fail to classify recent books as lesbian fiction. Martyn Greenwood's research looked at the availability of graphic novels in English public libraries suggesting that better cataloguing and displays by genre can improve their discoverability. All three talks in this section highlighted the varied ways in which public libraries are crucial to the communities they serve and the severity of the threat facing them.

Emma Coonan, our fantastic keynote speaker, refocused everyone after lunch and provided invaluable advice on how to get published and share your research with a wider audience. Emma, editor-in-chief of the Journal of Information Literacy, explained the publication process, highlighted the differences between a dissertation and a journal article, and gave some excellent advice about how to deal with the inevitable knockbacks. The audience was a mixture of new professionals, current LIS students and those who have been working in the sector for some time and Emma's advice engaged attendees at all stages in their careers.

The final session of the day focused on valuing the library with presentations about measuring value in a corporate library, the effect of tuition fees on attitudes in academic libraries, and the destruction of libraries in conflict zones. Natasha Chowdory's research demonstrated the value of the service she provides to her users and despite differences in the corporate sector her message that librarians are there to build relationships with their users resonated with attendees from all sectors. The relationship between librarians and users was also central to Marion Harris' presentation, which prompted much debate about the use of the term "customer" in academic libraries following the increase in tuition fees

in England. Sonja Kujansuu took a broader approach in the last presentation of the day where she shared her research on the destruction of libraries in conflict situations and how the international community can help. Her research felt particularly pertinent given recent events in Syria and it was fascinating to hear more about the organisations who attempt to protect libraries of cultural significance.

Many thanks to UKeiG for their generous sponsorship, which funded travel for several of our speakers enabling a variety of sectors and regions to be represented. Following the excellent response from attendees on the day and on Twitter at *#LISDIS2015* we are hoping to organise a second LISDIS conference next year. If you would like to view the slides from the day they are all available from our website.

"When they burnt our library to ground, they destroyed the heart of the city…"

Chloe Menown, Anglia Ruskin University

chloe.menown@anglia.ac.uk

What do we lose when we lose a library? This was the subject matter for the conference held at KU Leuven in September 2015. I was lucky enough to attend this three-day conference (9th-11th September) in the delightful city of Leuven, thirty minutes outside of Brussels.

In 1914, the Germans famously sacked the city of Louvain (Leuven) and set fire to the main parts of the town. They purposefully targeted the University and the 500-year-old library was burnt to the ground along with all its manuscripts. Europe's shock at the attack on a cultural institution lost sympathisers for the German cause and paved the way for the Allied forces propaganda. The "Flames of Louvain" became an iconic image of the Great War. The title of this article is a quote by our guide during the tour of the rebuilt University Library. This thought resonated with me and was echoed throughout the conference by different speakers time and again.

The tour of the University library (rebuilt in 1928 with American money by Herbert Hoover) provided a strong foundation to the history of the area and was followed by a visit to Louvain-la-Neuve in the French part of Belgium. In 1968 the University split into two separate institutions, the Flemish speaking half of the university stayed in Leuven and became Katholieke Universiteit (KU) Leuven whilst the French speaking side was relocated to Louvain-la-Neuve and became Université Catholique de Louvain (UCL). We attended a champagne lunch at UCL and the conference officially began with a talk by Alan Kramer (Trinity College, Dublin) on the Culture and Mass Killing in the First World War. The subject of book burning during WW1 was continued by the next two speakers with the main example being the University Library of KU Leuven.

After dinner, (back in Leuven) we were treated to a special keynote speech by Abdel Kader Haïdara from Timbuktu, you may recognise the name, as he is an internationally famous for moving over 30,000 manuscripts out of Timbuktu when they became a target for rebel groups in the area. His heroic efforts saved the manuscripts from certain destruction; Haïdara's speech was attended by ambassadors from America, Germany, Great Britain and Mali. Those of us who could not understand French listened to a translation of his speech on a headset; it was a rare experience for me to be a part of the international community, surrounded by so many different languages and cultures.

The second day of the conference started with a bang with a keynote lecture by Michael Suarez from the University of Virginia. Anyone who can talk about potatoes for forty

minutes and keep his audience raptured is an incredible presenter and Michael's showmanship became a talking point for the rest of the conference. Another talking point was the lecture by Colin Higgins (St Catharine College, Cambridge University) who argued for the non-existence of the Library of Alexandria. As a former Classics student, I thoroughly enjoyed his theory on how it was never destroyed in a dramatic battle, as the library never existed as a singular building.

There were also presentations on the practical considerations of losing a library and how to prevent damage to your collection. Paul Garside (British Library) gave a talk about treating fire damaged parchment but all I seemed to take away from his talk was that they set fire to a shelf of books in the car park behind the British Library to see how the books burnt. From speaking to Paul afterwards I think he was surprised that so many Librarians had suggestions for different ways to burn the books, for educational purposes of course.

By the final day my day head was exploding with knowledge and my understanding of how important a library is to a community was cemented by Marica Šapro-Ficovic (Dubrovnik Libraries). She spoke about the interviews she conducted with Librarians and survivors of the Croatian war (1991-1995), the number of people visiting the library was at its highest during the worst part of the war. The ability to provide hope and knowledge in times of darkness is a library's greatest asset. In the closing speech Michael Suarez made a passionate plea for the word "we" in the conference title. What do *we* lose when *we* lose a library? Every person in the world loses a culture, a history, and a community when a library is destroyed.

KU Leuven libconf2015 – What do we lose when we lose a library?
https://kuleuvencongres.be/libconf2015/website

Notes for Contributors

eLucidate is the journal of the UK Electronic Information Group. It is usually published four times each year, around March, June, September and December. It aims to keep members up to date with developments and innovations in the digital information industry, considering the impact on information professionals and consumers of e-information.

UKeiG encourages the submission of articles, reports and reviews about any of the topics covered by the journal. These include: electronic resource awareness, information management, information literacy, effective information retrieval and search technologies, intranets, social media, open access, e-publishing and e-industry research and development. UKeiG can't pay contributors, but you will retain your copyright and will be able to republish your work elsewhere.

Please follow these simple guidelines:

About our members

Our membership is eclectic and includes information professionals at all levels of the UK workforce involved in digital content management and awareness, information dissemination, training and service delivery.

The UKeiG demographic comprises academia, but also the private, commercial and public sectors, embracing schools, further and higher education, the NHS, healthcare and pharmaceutical industries, science, law, finance, arts, humanities, archives, museums and libraries.

UKeiG's most popular training courses include search tools and strategies, intellectual property, e-books, intranets and content management.

A key benefit of membership is that the training courses, meetings and networking forums provide "crossover" insight from one discipline to another. Members see UKeiG as a way of keeping up to date with trends and developments outside of their core, day-to-day business. Few other organisations provide this kind of cross-sectoral context and oversight.

Technical level

Although members rate themselves highly for technical awareness, they are typically users rather than creators of technology. Articles should not assume understanding of technical terms without explanation.

Length of article

Feature articles should be in the region of 1500-2500 words. Each article should be prefaced by a short summary (around 50 words.)

What to write

The world is your oyster in terms of suggested themes and subjects as long as they reflect the disciplines and membership base articulated above. You should never assume that readers will be entirely familiar with your topic, so anything you can do to offer definitions, explanations, examples and context would be welcome. You should always link to suggested reading and alternative resources to enable readers to explore your article further.

While the obvious focus of the group is the UK electronic information sector, the industry, by its very nature, is global and international developments should be reported when they impact on the UK landscape.

The most valuable viewpoint you can give is that of a practitioner. While UKeiG welcomes theoretical debate, we are primarily a forum where peers can share their practical experiences and understanding. So, if something worked for you, tell the readership. If something didn't, tell the readership why not.

How to submit

Please e-mail your copy to the editor gary.horrocks@gmail.com Articles should be delivered in a simple Word format. Hyperlinks to alternative/suggested content/further reading should be embedded in the text. Images are welcome if they illustrate a point or clarify a statement. Please send them separately, and also place them in the Word document in the appropriate sections. They may be in gif or jpeg formats.

Rights

By submitting an article to eLucidate, authors grant UKeiG the non-exclusive right to publish the material in any format in perpetuity. However, authors retain full rights to their content and remain the copyright owner.

About you

Please provide a 10-20 word biographical summary about yourself, alongside an email address and job title.

Editorial process

Your article will be copy-edited for spelling and for sense. If there are major changes to the article we may return it to you for your comments and approval, but most articles require only light corrections before appearing in eLucidate, and do not need a further review by the author.

Brief for book reviews

Book reviews are typically 600-1000 words. Because UKeiG is independent of any publisher, we are not obliged to have favourable reviews. If you think a book is poor, then by all means explain why. Members and non-members alike are welcome to suggest books for review or to submit reviews.