

# SELECTION OF PAPERS FROM THE BIALL CONFERENCE 2014

## Data, Data Everywhere

**Abstract:** This paper that is written by Phil Bradley is based on his keynote lecture, which he gave on 12 June 2014 at the BIALL Annual Conference. He considers the growth of information on the internet, both in terms of type and amount. His article considers the difficulties that this flood of data brings with it, the challenges facing traditional search engines when faced with access by mobile devices and applications. The role of privacy and the ‘internet of things’ are also discussed. Finally, there is an overview of the role of the information professional in this new information environment.

**Keywords:** internet; social media; privacy; mobile technology; big data

### INTRODUCTION

Imagine if you will, that every word ever uttered by everyone who ever lived had been transcribed. That amount of data would be in the order of 5 exabytes.<sup>1</sup> If you prefer a library based equivalent, one exabyte could hold 100,000 times the printed material in the Library of Congress, or if we include all of its audio, video and digital material, between 500–3,000 times that content. In 2004 the global monthly internet traffic passed 1 exabyte for the first time, and it’s now estimated that the global IP traffic per month in 2016 will reach 91.3 exabytes and 131.9 exabytes by 2013.<sup>2</sup>

These figures are virtually incomprehensible to all of us, and they are so out of our realm of understanding that they really don’t help us to understand one of the greatest challenges facing not only the information industry, but in fact the entire world. The amount of data available is entirely out of control, and it’s not going to ever get any easier. This has many implications for how we do our jobs, find information, access and validate it, then make it available to people who need it.

### SOCIAL MEDIA BASED CONTENT

The vast majority of the data that we’re now seeing flooding on to the internet is user generated content; that is to say people are posting their own photographs, videos, tweeting, doing Facebook status updates, creating their own websites and so on. I don’t want to bore you with tedious collections of statistics, but it does help to get some idea of the sheer amount of activity people undertake. For example, on Facebook alone there are 243,055 photographs uploaded per minute, 3,125,000 likes, 323 days of YouTube video are viewed, and 500 new accounts added.<sup>3</sup> Elsewhere on social media platforms there are, per minute 433,000

tweets, 5 million videos viewed, 1,100 photographs uploaded to Flickr, and 100 hours of YouTube videos uploaded.<sup>4</sup> There is absolutely no sign that this is going to reduce in number; it will simply increase as more people use various different platforms. I think that it would only be the very foolhardy (or stupid!) who would try to convince us that social media is still a ‘flash in the pan’.

There are of course a huge number of implications to this for the information industry, and I just want to focus on a few of them. What is news? I have absolutely no idea any longer, and indeed neither does anyone else. Mark Zuckerberg is quoted as saying that ‘a squirrel dying in front of your house may be more relevant to your interests right now than people dying in Africa’.<sup>5</sup> News now comes to us from a much wider variety of sources than it ever did in the past, and we have to learn new ways to authenticate and verify that information. For example, in October 2012 rumour dominated Twitter to the effect that Fidel Castro was dead – at the rate of 250 tweets per minute, and confirmation of his death was announced in blogs. However of course that wasn’t the case, but if we took what a large number of people were saying at face value we would be looking in the newspapers for his obituary. On the other hand, recently a friend of mine told me that she’d heard an explosion near to where she lived, by the Dartford Tunnel. None of the mainstream media carried the story, but there were a large number of tweets to the same effect, including one from the Essex Fire Service; clearly there had been an explosion (at a chemical factory) but traditional media didn’t pick up the story as quickly as local citizen journalists who were immediately on the scene. Consequently I may well find myself trusting people that I know far more than media outlets. I certainly trust them more than I trust search engines such as Google for example; a search for “Martin Luther King” returns a hit very high up in the results which is actually the product of an American

white power organisation, and is full of falsehoods and inaccuracies. It's not Google's job to provide accurate information, it's there to match content to search terms while making money in the process.

## THE SPEED OF NEWS

The speed of news is also an important consideration now. When someone says that they want 'new' information, does that mean as in the last week, yesterday, or today? Might it even mean the last few seconds? A search engine such as Topsy<sup>6</sup> can index tweets in 150 milliseconds, and news breaks on social media far faster than it does on mainstream media. Since most of us started to use the Internet we tended to look at the key news websites, and pull other data from subject specific sites. Then when people started to blog people were getting their posts indexed within 20 minutes. Now we have to start to wonder at how effective websites are at getting news to us, and I think that the answer is 'not very'. Indeed, if you look at the results that you get from many major search engines they are increasingly linking to social media platforms; they are becoming stepping stones to help us get directly to the information that we need. As a result, the importance of the individual is increasing. I can run a search on Google and it may well give me a result from a particular author and Google can inform me that the author in question is followed by several million people on their Google+ network. Or I can find a result on Twitter and check to see how many people are following that person. Alternatively I can use any of the content curation sites such as Scoop.it to follow curated content as produced by experts that I actually know personally and trust. Indeed, Facebook in their attack on Google has added powerful new search options to their search offering. Google can give me a list of restaurants in Manchester for example, but without further research, how do I know they can be trusted, but on Facebook I can find out which restaurants in the city have been recommended by my friends; a more effective and trustworthy way to search perhaps. Search engines are attempting to counteract this by providing personalised results based on previous search histories, sites and adverts that have been clicked on, and the personal profiles that they have built up on us. Consequently the result that I see in 3<sup>rd</sup> place for a search may well not be the same one that you see, so we are in a situation that Eli Pariser describes as 'filter bubbles'.<sup>7</sup> It is therefore becoming increasingly difficult to trust the information that we find on the net; Eric Schmidt (Google executive chairman) said "*The technology will be so good it will be very hard for people to watch or consume something that has not in some sense been tailored for them.*"<sup>8</sup> There are of course solutions, but these require leaving the safety net of Google, and exploring other search engines which work differently, which do not track searchers or searches and which give people a greater sense of privacy. It's surely the role of the information professional

to explore these, experiment and then advise their clients or members on how and when it's best to use them.

## THE MOBILE ENVIRONMENT

We also need to consider the move away from static search and retrieval of information to a far more mobile environment. In 2000, just over half of UK adults said that they had a mobile phone, and that figure now stands at 94%. In the first quarter of 2013 49% of adults used their phones to access the internet, up by 10% on the previous year.<sup>9</sup> The sale of smart phones and tablets has exploded, leading to a very different internet experience. In the pc world, people share computers at home or use them at work, but in a mobile setting people can take search with them. The web and web search is limited in comparison to a mobile world where people can use apps to get the information they need. I can find out exactly where I am, where the nearest library is, I can use image recognition and augmented reality to get what I need quickly. My requirement to use traditional search and traditional website pages is reducing day by day. We can now take our wearable tech with us; not only do we have access to Google Glass, but we're now seeing the rise of wristwatches that we can use to obtain information, check emails, take photographs, connect to other Bluetooth devices and so on. We have reached a point where devices can monitor everything that we do, including our health and interact with other services on our behalf – often without us even knowing about it.

## THE INTERNET OF THINGS

The internet is swiftly becoming the internet of things; a system whereby many different items can be connected to the internet. These could be heart monitoring implants, biochips, cars, fridges and toasters. You may well laugh at some of these, and think what is the possible value of having a toaster connected to the internet. However, the manufacturer can use the connection to see how often the toaster is used, how well it performs, and if it breaks under warranty, they can send you another in the post. Your fridge will know what food you have in it, it can interact with your personal device, see that you're having a friend around for supper and can then alert you when you're passing a supermarket to buy the extra ingredients to make a particular meal. If traffic signals are connected to the internet they can see what the current situation on the roads is and adjust stop and go times to result in a smoother journey for commuters. According to Gartner there will be nearly 26 billion devices connected to the Internet of Things by 2020.<sup>10</sup>

## THE VISUAL WEB

However, let's come back from that look into the future and once again consider where we are now. It's often

been said that a picture is worth a thousand words, and we're certainly moving towards a far more visual web than we've ever had before. YouTube for example is the second largest search engine in the world as defined by the number of searches per month, with over 3 billion searches.<sup>11</sup> Pinterest, which debuted in May 2011 saw an increase of 4,225% between July 2011 and July 2012 on the amount of time mobile web users spent on the site, and it's now the fourth largest driver of traffic worldwide.<sup>12</sup> Infographics are becoming a very common way to display information and having infographics in blog posts increasing the change of them being shared by up to 832%.<sup>13</sup> Searchers are increasingly using a wide variety of infographic search engines in order to find the content that they need. Social media has transformed into real time, visual social media – we are drawn to that type of content.

## PRIVACY

With all of the information that's now available, and will be so increasingly in the future, we have to reconsider what exactly privacy means. The new Facebook search options let me slice and dice the content that I'm looking for in over 40 different ways. The recent 'right to be forgotten' ruling has many implications on all of us. Do I have more of a right to know that a person convicted of a sex offence lives next door to me, or does the offender have the right to their privacy? While Google is considering requests to remove links to certain content, that doesn't remove the content itself, and nor does it always remove the content from Google; it's just harder to find. Currently Google only removes content on its European based search engines, but in the future it may be required to remove it from the .com version as well – but is it right that one country or regions law affects the rest of the world? Furthermore, who is overseeing what Google does? There appears to be no right of appeal, and Google has been very careful to inform mainstream media when their stories have been unindexed from their database. Should we really be accepting of the fact that what we can see should in part at least, be decided for us by a panel of staff from an American conglomerate? Of course in practice, it's nonsensical, because people can just move to different search engines based in different countries, and it's going to be a virtually impossible task for anyone to contact all of them and request that their details are removed.

We are increasingly moving towards a system of search on the internet that demands private searching which cannot be tracked or traced. There are many reasonable and legitimate reasons why people may decide to use browsers such as the Tor browser<sup>14</sup> which takes your communications around a set of relays run by volunteers, preventing sites from learning your physical location, and allowing access to sites that might otherwise be blocked. The 'dark web' is becoming more widely known; 4% of the information available on the internet is the

visible web – the material that you can find using a traditional search engine, while the other 96% of content is the dark or invisible web, which is not indexed and is very hard to keep track of.<sup>15</sup> For whatever reason, privacy, protection or for less wholesome reasons people are increasingly using the web in an entirely different way, and attempts to control the web and the content on it have merely pushed it further underground, requiring specialist knowledge to get appropriate content. We are moving towards a two tier net, or a new digital divide, between those who have the skills and expertise to locate data, manipulate it and perhaps pay for it with bitcoins (a virtual, untraceable currency) and those that don't.

## OUR CHANGING ROLES

Since the way we access information is changing, and the type of information is also vastly different now to what it was even 5 years ago, I believe that we need to rethink our approach to how we do our jobs, and what's important. The first point, and I cannot stress this strongly enough, is that we need to become social librarians. Access to social media platforms isn't a 'nice thing to have' it's becoming an increasingly important part of the way that we can do our job – it would be insane to assume that a librarian could work adequately without access to the internet (in most jobs at least) and we need to inculcate the same view towards Facebook, Twitter, Pinterest and the rest of them. We have to go to where the conversations are, in order to monitor them, correct mistakes, inform people within our organisations; in essence to take on the role of social media managers. We need to curate the information that we find – to see what people are talking about in our subject areas of interest on places such as Twitter or professional groups on LinkedIn. We need to be in a position to authenticate that information and then act as a beacon to the people that we work with so that they can see it. We need to use our existing skills to further monitor and sieve out the useful material from the dross, and with the amount of data flooding out every day that's becoming an ever more essential role. Our role is also becoming much more of an educator; we're all tired of hearing the 'it's all on Google, why do we need libraries and librarians' line, so it's never been more important that we're able to move people that we work with to a new level of understanding when it comes to information, finding it, validating it and authenticating it, then using it. Don't worry, because we're not going to put ourselves out of a job – the internet is continually changing, growing and developing, and we're the ones who have to keep tabs with what it's doing. I have been teaching a course on advanced internet searching since 1995, and I expect to be teaching it up to when I retire. The content changes, I talk about different search engines, but there is still an absolute need for those skills, and it's really up to us to teach them. We also need to be the people in the organisation who try out new things. Sure, some of them will not do

what we want, and we might fail miserably, but we have to have a different view of what we're doing – if you want things to change, you have to explore and experiment, and finding that something doesn't do what you want isn't an error or a mistake, it's a positive learning outcome. Without exploration, understanding and communication it's not only us as individuals who will suffer, but so will the organisations that we work for.

## SUMMARY

We need to focus on our value, impact and our roles within the organisation. We need to view information and our access to it in a very different way. Pay less attention to the older traditional approaches of search engine,

website, webpage and content, and move towards a faster, more free flowing viewpoint. We need to set up our own circles of trust to share information, and to use our Twitter and Facebook contacts to keep us up to date. We also need to educate as many people as we can. Facebook is not about 'friends'; it's an increasingly important aspect of the internet and it's key to bringing an organisation to the attention of 1.25 billion users. Twitter, despite its trivial sounding name is somewhere that news breaks first, and quickly. It's also an excellent way of communicating with other professionals. Move the conversations away from the social media platforms, and towards the activities that they support. The approach and access to information is changing rapidly, and so must we.

## Footnotes

- <sup>1</sup> <http://web.archive.org/web/19990508062723/http://www.ccsf.caltech.edu/~roy/dataquan/>
- <sup>2</sup> [http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI\\_Hyperconnectivity\\_WP.html](http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.html)
- <sup>3</sup> <http://www.insidefacebook.com/2014/06/09/infographic-what-happens-in-a-facebook-minute-2/>
- <sup>4</sup> (<http://www.digitalinformationworld.com/2014/07/what-happens-in-just-60-seconds-on-social-media-infographic.html>)
- <sup>5</sup> <http://www.brainyquote.com/quotes/quotes/m/markzucker416520.html>
- <sup>6</sup> <http://www.topsy.com>
- <sup>7</sup> [http://www.ted.com/talks/eli\\_pariser\\_beware\\_online\\_filter\\_bubbles](http://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles)
- <sup>8</sup> <http://www.businessinsider.com/best-eric-schmidt-quotes-2014-7?op=1#ixzz39W6eNjNz>
- <sup>9</sup> <http://www.mobilemastiinfo.com/stats-and-facts/>
- <sup>10</sup> <http://www.gartner.com/newsroom/id/2636073>
- <sup>11</sup> <http://www.mushroomnetworks.com/infographics/youtube---the-2nd-largest-search-engine-infographic>
- <sup>12</sup> [http://www.mediabistro.com/alltwitter/social-referral-traffic-2014\\_b58664](http://www.mediabistro.com/alltwitter/social-referral-traffic-2014_b58664)
- <sup>13</sup> [http://www.mediabistro.com/alltwitter/infographics-on-twitter\\_b26840](http://www.mediabistro.com/alltwitter/infographics-on-twitter_b26840)
- <sup>14</sup> <https://www.torproject.org/projects/torbrowser.html.en>
- <sup>15</sup> <http://bgr.com/2014/01/20/how-to-access-tor-silk-road-deep-web/>

## Biography

Phil Bradley is an internet consultant who works with librarians and other information professionals, particularly in the areas of internet search, social media and multimedia. He runs courses on these subjects and also offers a consultancy service to companies that want to better understand the role of social media and how to profit from it. His clients include major government departments, law firms, banking companies, academic libraries, school and public libraries. Phil is the author of a number of books on aspects of the internet, and he is currently the Immediate Past President of the Chartered Institute of Library and Information Professionals (CILIP).