

# Future of Data





## The Global Challenge

The future of data is a broad topic, which can cover a range of issues: some technical, some regulatory, some social and others philosophical. The web is still a young technology - it has only been twenty years since Sir Tim Berners-Lee and Robert Cailliau invented it at CERN: It will take many decades for us to fully understand its impact on our society. And the pace of change on the Internet, and that which is enabled by the Internet, is speeding up. Whatever happens, as it continues to develop, we'll be presented with more opportunities and more challenges. The web is a fundamentally democratic platform, and it reflects both positive and negative aspects of the offline world.

If we take the field of data to encompass all digital factual information, the current work of both leaders and emerging companies suggest the issues that will arise in years to come. Companies such as IBM, Oracle and SAS are making strides in data mining and database management. Their research shows that intelligent systems will become increasingly prevalent. Other organisations, like Amazon, Sun and even Google, are demonstrating the amazing benefits in scale and interoperability that come through moving data storage into the cloud. And, if one was to talk to the people who are driving the web forward, they anticipate a more powerful, flexible and useful web in the years to come. The much-touted 'semantic web' - in which the relationships between pieces of information will be both apparent and useable - may not be imminent, but it's certainly within sight. Its advent will drive further research, and it will also make the web more useful to people around the world.

As investment and regulation follows rapid development of potential technologies, they will have to adapt to the

new challenges of the online world. Google is involved in many of these, but for me the big issue at the heart of the future is that of access to information...to data. Today, anyone with an internet connection has access to more information, quickly and easily, than was available a generation ago to anyone not connected to a research library or university. That's an amazing development, but we should remember that less than a quarter of people globally have access to the web. New developments will increase the speed, scale and sophistication of the data we can use, but, for most people, there is still a high barrier to access.

Access to information is the great leveller. It empowers citizens and consumers alike. That's why it's imperative that access to data be fast, cheap, and ubiquitous, whether you are in the New York, Shanghai, Lagos or Patagonia. Right now, in many parts of Africa and Asia, internet connectivity is both expensive and slow. As such, the positive benefits of the information age have been unevenly enjoyed. Addressing this disparity is a clear and significant challenge for the future.

Access to information is the great leveller. It empowers citizens and consumers alike.

## Options and Possibilities

One certainty is that information discovery will continue to get better.

The beauty of the internet, and therefore by association access to data, is in its unpredictability. The web's openness means that new innovations appear online every day. Some succeed and others don't, and successes spawn further innovation. Two years ago, for example, very few people would have predicted the role that Twitter and YouTube played in the Iranian elections. Even so, as we look ahead, some things are clear.

One certainty is that information discovery will continue to get better. Wouldn't it be good to have a system that asks questions as well as answers them? A recent article in the Economist described how this could revolutionise innovation as we know it - citing a research chemist at Pfizer as an illustrative example. How cool would it be if he could find solutions to one of the mysteries of science, perhaps cure a disease, simply by asking the right question of the web? A semantic search engine that has read (and understood) all the relevant literature, interrogated the patent libraries and medical records, and studied the chemical theory, etc, might well suggest workable solutions. Science fiction? Perhaps, but imagine the value of a system that understands the relationships between information in different corpora, created with vastly different uses in mind.

It also seems clear that access to data will help to widen the pool of potential creative ideas - a step on from crowd sourcing towards democratized innovation. Think of Wikipedia, the online encyclopaedia written by its users, or iStockphoto allowing amateur photographers to earn money selling their pictures alongside professionals. Each combines cheap and widely available tools to allow talented people to make the most of the Internet's distribution efficiency, and this trend is only beginning.

Closer to hand is the migration of computer applications from the desktop to the web. In this shift to cloud computing, more and more of our personal and professional lives will be spent using our web browsers. That means browsers will have to be stable, powerful, and above all secure.

Also apparent today is the role mobile phones will play in improving access to the Internet (and therefore to data). There are already nearly 4 billion mobile phones in use today around the world, and over 80% of humanity lives within range of a mobile network. At the same time, the cost of web-ready phones continues to fall. Computers are getting smaller and cheaper, and the next generation of mobile networks will improve access speeds. Already, net-books can cost as little as \$200, making them cheap enough to be given away with mobile-broadband contracts in some countries.

Even when mobile access becomes universal, there will still be real challenges connecting some places to the larger Internet. There is reason for hope, though: a series of new cables are in the works to improve Africa's connectivity with the rest of the world, increasing capacity and reducing the cost of internet access. The first of these, the SEACOM cable, eastern Africa's first modern submarine cable, was completed in July 2009. In coming years, some places in Africa may well have higher speed connections than parts of Europe.

Doubling the number of people online will have an amazing impact on innovation. More people (with more diverse experiences than ever before) will be able to contribute to the innovation happening online. That's a very exciting prospect.



## The Way Forward

If we consider what has been achieved in the past ten years, over the next decade we have the opportunity to give more and more power to users. In the world of ubiquitous and uniform access, intelligent agents and the semantic web, we have the potential to enable even greater shifts in transparency and access to data than previous generations would have ever imagined. However, to achieve this we need to move forward on two key topics that will moderate the impact that can be achieved.

One pivotal issue will be identity online, as people become more comfortable managing what information they share about themselves, and with whom. Many services on the web improve quickly when they people give them access to personal information. An example: in the process of crawling the web, Google visits more than a trillion different pages. Several billion more are added every day. Finding the right information is like having a fraction of a second to find a needle in a haystack of astronomical proportions. The links between web pages are the first indicator of how important any given page is, but our search logs provide an excellent form of feedback on whether we're providing the best results. But if people are comfortable sharing their search history with us, we can use that as a valuable signal to provide them more relevant information more quickly.

The second and related issue, that I think needs to be taken several steps forward in the next couple of years, concerns regulation. There is an on-going discussion about how to limit the uses of personal information without compromising innovation or decreasing access to information. Different countries have significantly different views on this, but, as national boundaries become less significant in a world of digital natives, we need to decide what rules are necessary and how those rules should be formulated. We all need to understand the balance and consider the possibility for increasing transparency in both directions. If people are given access to data to re-use, the power of innovation shifts to the public, and the potential for sharing of more ideas increases exponentially.

If people are comfortable sharing their search history with us, we can use that as a valuable signal to provide them more relevant information more quickly.

## Impacts and Implications

Substantive research has already shown us that access to information has significant impact on quality of life from an economic, social and political perspective in many dimensions. For example, think first of the fishermen who can now identify in advance where they are most likely to get the best price for their catch and so sail straight to the port and thus improve their efficiency and also profitability. Or think of the student who can check online to see where friends a meeting up - and then decide whether to join in knowing who will be around, what the music will be like and, even get information about how to get there. Access to new data is already changing lives - what it will do in the future is pretty much only limited by our imagination.