

The Value of Freeing Data

For years, governments, international agencies and corporations understood the **intrinsic value of their data and fiercely guarded it.** Now, as the tools become available for all of us to harness its power, more and more people are calling for it to be freed.

Tim Lyddiatt

Google, whatever else they might be, have been true to their word. At their inception just 11 years ago they set out to 'organise the world's data and make it useful.' Whatever else you think about them, and how seriously you think they can still claim to not 'be evil,' most will agree that they have achieved this aim in spades. As we come to the end of the first decade of the 21st century, we inhabit a world that is, in several key ways, fundamentally different from the world we inhabited 10 years ago. First is that ubiquitous broadband is a reality for far more of us than it was then: we have fast, always-on connections to the Internet and, as a result, it is far more intricately woven into the fabric of our daily lives. One of the key reasons that this has come to pass is that the vast quantities of data available to us has been organised – not just by Google, but by an increasingly vocal set of academic and entrepreneurial pioneers and campaigners who wish to free data from its commercial shackles and its academic obscurity – so as to be accessible, and therefore useful, to more and more people. On the back of data, these campaigners say, new, rich possibilities lie just around the corner. All we have to do is free the terabytes of non-commercial data currently held



A photo of the Google Inc. logo is shown on a computer screen

by governments and NGOs and, going further, watch as new commercial entities are built on the back of data suddenly freed.

Second is that our connection to the Internet is no longer tethered to our homes and offices: mobile devices afford us Internet connectivity practically everywhere we happen to be. As a result of this, and because more and more of these mobile devices utilise global positioning system (GPS) data, location based services are becoming the next big thing, with

services cropping up across the UK and other developed markets that can point, locate, recommend, and then direct you to whatever you happen to be looking for. Feeling hungry, let your iPhone, Yelp and Google Maps find, review and locate the perfect lunchtime destination. Looking for a cheap hotel near the conference you are scheduled to attend, let hotels.com help you choose from tens of thousands anywhere in the world. In a mobile world, certain datasets become more useful: geo-location data, overlaid

with commercial data is a heady mix and the number of possible commercial applications is almost unlimited.

These changes, and others, represent a seismic shift in the way in which we live our lives. No longer do we rely on prior knowledge of a location in order to navigate an area, nor are we reliant on blind discovery in order to be able to function within it. Instead, we now have the tools in order to discover, assess and evaluate an area long before we have ever set foot in the place. Similarly, a greater understanding of the world we inhabit is opening up huge changes in the way we view and interact with that world. If we can better visualise the intricacies of our world, will we not understand it better and be more sympathetic toward it?

Whilst commercial concerns have a vested interest in allowing their location data – and other datasets, such as details of which business the company operates, its services and even comparison information – to be accessed easily from anywhere, there is a growing call for government and public data to be freed in the same way, allowing who knows what to be built on top of it.

At the beginning of October, a rather extraordinary call for help was made by the Greater London Authority (GLA) to the city's IT developer community. They were inviting London's top geeks to attend a meeting, scheduled at month end, in order to help the authority free their data. Like every other public authority in the world, throughout the course of its daily operation the GLA generates and curates huge amounts of data – everything from planning and zoning details to policy and environmental data. They were inviting geeks to consult and advise them as to the best way of making this data accessible. This call for help in freeing their data by City Hall was recognition of two things: that data is more valuable when it is accessible to more people, and that there are right ways and wrong ways to make that data accessible.

The commercial aspects of freed data are huge, but there are also social and community benefits too that have arisen thanks to authorities such as the GLA freeing their data. And it is in pursuit of these that a movement has arisen in the UK and elsewhere, calling



Managing Director of Yahoo Mobile Asia-Pacific Kunze speaks during news conference at CommunicAsia trade show in Singapore

for all authorities to free data that, the movement claims, already belongs to us because its collection was funded by our tax payments. What the GLA was asking for was help in deciding on the best way for this data to be released; in what format and on what platform so that the most benefit can be garnered from its release.

**Data is more
valuable when it is
accessible to more
people, and that
there are right ways
and wrong ways
to make that data
accessible.**

The problem with data is that it's useless unless it can be used. In the mid 80s, in conjunction with the European Commission, several technology companies and the BBC came together in an ambitious project to mark the 900th anniversary of the publication *Domesday Book*. The original *Domesday*, published in 1086 was a survey and census of England undertaken and released by the conquering Norman king William, otherwise known as

William the Conqueror. The BBC's *Domesday* project was compiled over two years and included a new 'survey' of the United Kingdom, in which people, mostly schoolchildren, wrote about geography, history or social issues in their local area or just about their daily lives. This was linked with maps and many colour photos, statistical data, video and 'virtual walks'. Over one million people participated in the project and the project also incorporated professionally – prepared video footage, virtual reality tours of major landmarks and other prepared datasets such as the 1981 census.

Today, this fascinating insight into mid eighties UK life is all but inaccessible because it was created using technology that, whilst cutting edge at the time, is obsolete now. The modern *Domesday Book* was produced and stored on laserdiscs, a technology that was unique at the time for the many types of data it could store – text, video, voice recordings etc – and for the amount: each side of a laserdisc could store up to 300mb. Less than a quarter century later and the technology has been superseded and much of the data all but lost forever.

Who knows how the modern *Domesday* project could have been used by today's generation of school children, but at the very least, it would have provided a unique snapshot of a country at

a particular point in its history. It is in pursuit of avoiding such digital obsolescence that the GLA has sought the help of London's brightest geeks. They have recognised that for its data to be useful, it must remain accessible forever. So, assuming that data can be freed in a format that will survive history, what can be done with it, now and in the future?

A very great deal, according to a rather dry report released at the same time as the UK's 2008 budget report. In Models of Public Sector Information Provision via Trading Funds three eminent professors at Cambridge University assessed the value to the UK economy of freeing up data produced by six UK institutions including the Ordnance Survey (maps and geographical data), the Meteorological Office (weather statistics and future modelling data), Companies House and the Land Registry. Their findings, printed over 150 pages found that "the case for pricing basic data at no more than the marginal cost of a copy is 'strong;'" in contrast, "the case is 'weak' for public enterprises not subject to regulation and providing monopoly services without fear of competition."

The report's analysis centres on comparison between the current model – where each agency uses commercial infrastructure and pricing in order to recoup its costs and turn a profit – with a free model where costs are recouped through taxation on revenues made by companies and individuals newly able to exploit the data generated by

these bodies. This alternative revenue generation is at the heart of an increasingly broad and varied collection of thinkers who believe, that when it comes to government data, 'free' is the only price that it is worth paying.

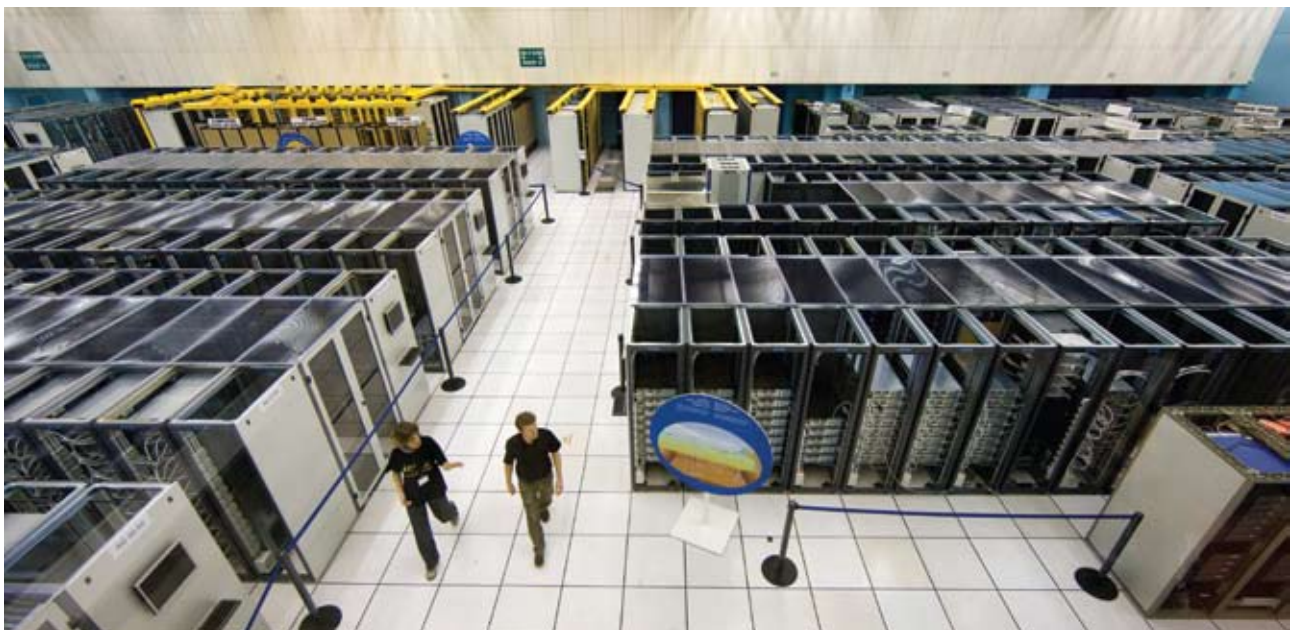
One of the report's authors, Rufus Pollock, is a founder of the Open Knowledge Foundation (OKF) which campaigns against restrictions to knowledge-sharing. Around the time that the GLA made its call to arms, Pollock, writing on the OKF blog (www.okfn.org), announced with some considerable glee that the beta phase of the government's data.gov.uk site had commenced: "it shows such a clear commitment to working collaboratively with the wider open data community into the future," he said. The project, similar in scope and ambition to the data.gov site introduced earlier in the year by President Obama, has more than a 1000 existing data sets, from seven departments, "all brought together for the first time in a re-useable form," Pollock informs us before posting some very cool graphics, built from the datasets about how UK tax money gets spent by region and function.

In these examples we can see how freeing data can help make the data more accessible, and therefore useful to all of us. Another example is Gapminder.com that collates terabytes of data from the World Bank, the UN, the IMF, Unicef and others into a single location and make it available in a graphically pleasing and intuitive format. As a result, previously

incomprehensible historical data is brought to life. The site's creator, Hans Rosling says that the site is "a modern 'museum' that helps make the world understandable, using the Internet." In this respect, he is echoing Google's mantra: 'organising the world's data, and making it useful.'

Whilst it seems obvious – with the benefit of hindsight – that all government data should be made available for free as soon as the technology became available to make it useful, some people want to take it even further. Jamie Heywood is co-founder of PatientsLikeMe, an online community that encourages the sharing of personal medical data. In October, he spoke to Wired Magazine saying, "Privacy has been used as an excuse by those who have a vested interest in hoarding this information. People have the right to take possession of a complete copy of our individual health data, without delay, at minimal or no cost."

Heywood is a driving force behind the Declaration of Health Data Rights, a four-point manifesto that supporters can publicly endorse through HealthDataRights.org, and is campaigning for the freeing of medical data that, arguably even more than any of the other datasets discussed here, truly belongs to individuals and not to faceless corporations or authorities: the data that details our own health. When you look at it like that, it seems incredible that anyone must campaign for this or any non-commercial data.



Scientists walk in the CERN LHC computing grid centre in Geneva