

# Living in a World of Data

Alan Dix

Talis  
43 Temple Row  
Birmingham, XXXX, UK

InfoLab21  
Lancaster University  
Lancaster, LA1 4AW, UK

alan@hcibook.com

<http://www.hcibook.com/alan/papers/assets2011-world-of-data/>

## ABSTRACT

The web is an integral part of our daily lives, and has had profound impacts on us all, not least both positive and negative impacts on accessibility, inclusivity and social justice. However, the web is constantly changing. Web2.0 has brought the web into the heart of social life, and has had mixed impact on accessibility. More recently the rise in API access to web services and various forms of open, linked or semantic data is creating a more data/content face to the media web. As with all technology, this new data web poses fresh challenges and offers new opportunities.

## Categories and Subject Descriptors

H.5.4 [Information Systems]: Hypertext/Hypermedia – *user issues*; I.2.0 [Artificial Intelligence]: General – *philosophical foundations*; K.4.1 [Computers And Society]: Public Policy Issues; K.4.2 [Computers And Society]: Social Issues – *assistive technologies for persons with disabilities*;

## General Terms

Design, Economics, Human Factors

## Keywords

web2.0, API, open data, physicality, embodiment, empowerment, long-tail, inclusiveness

## 1. THE RISE OF DATA

For most of us it is hard to imagine the pre-web world, even though the web is a mere 15 years old. However, the web is itself changing, five years ago many of us had never heard of Facebook or Twitter, yet now social media are supplanting the browser as being 'the Internet' for many users. Another move, that some have called web3.0 (or web4.0 depending on counting), is the move to a more data-centric web.

Many web sites (including Facebook, Twitter and Google) offer APIs so that third party developers can add functionality, or alternative interfaces. The Google Maps API was perhaps the first to capture the imagination of users beyond the highly technical developers, however the success of APIs is widespread. Arguably one of the reasons for Facebook's ongoing success is the proliferation of third-party apps enabled by its early API and more recent Open Graph Protocol [9]. Similarly, one of the drivers of

Twitter's growth has been mobile access, and yet it has only recently launched its own iPhone App, relying before on third party apps using its API. Many of the next generation start-ups are taking this to the extreme with API-only web services.

There is also a growing move to make data more available on the web. Tim Berners Lee has long been an advocate of the Semantic Web, making the human readable web available to computer interpretation [3]. While this dream has been long in coming and may turn out differently, there are clear signs of a more data-rich web. Web pages themselves are becoming more richly annotated, some with 'proper' semantic web markup such as RDFa [1], but perhaps more commonly with semi-semantic markup such as microformats or the recently agreed schema.org vocabulary, where the major search providers (Yahoo!, Google, Bing) agreed a common standard for semantic mark up.

As well as marking up web pages, raw data is increasingly available. Many governments, including the UK (data.gov.uk) and USA (data.gov) are making open data a key aspect of public policy, and some media organisations, notably the BBC [2] and the Guardian [11] have strong data-focused initiatives. Sometimes this data is in traditional formats, such as Excel or CSV spreadsheets, some is in more web formats such as RSS feeds, and some in Semantic Web RDF. Critically, many of these semantic data resources are linked through common identifiers leading to what is called the Linking Open Data (LOD) Cloud [4].

The web has had an impact on accessibility and inclusivity: opening up the world to some, and closing it off from others.<sup>1</sup> This new data web similarly both poses challenges and offers opportunities for a more inclusive world.

## 2. EMBODIMENT AND PHYSICALITY

Computers, and more critically the Internet, have enabled us to live in a parallel world of the virtual, where Negroponte's bits rather than atoms are most critical. Like the faerie lands in folk tales, this is not dissociated from the real world, in particular, the people of the virtual world are (usually) those of the real world; however, it does offer the potential to break some of the physical and, to a lesser, but still real, extent, the social and material constraints of day-to-day life. This is evident in chat rooms, multiplayer games, and second life, and epitomised in the character of Zona Rosa in Gibson's *Idoru* [10].

I have been interested for many years in the nature of human physicality. By understanding how our cognition is attuned to the physical world we are in a better position to design for digital environments, both in physical digital devices [8] and on the web

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*Conference '10*, Month 1–2, 2010, City, State, Country.

Copyright 2010 ACM 1-58113-000-0/00/0010...\$10.00.

<sup>1</sup> E.g. vast download sizes of software updates and Apple's decision to make Lion available principally through App Store, excludes or penalises those not in urban areas of developed countries, with fast broadband connections.

[7]. Philosophical theories of embodiment focus on the way we interact with the world directly and often emphasise that we do not need internal representations in our heads as the "world is its own best representation" [6]. However, there is also strong evidence that we do use internal representations, for remembering, planning and similar activities, but that these representations are strongly influenced by our physical being [5, 14]. This is important from an inclusivity perspective as not only the cultural aspects of the data structures we encounter, but potentially also their implicit cognitive structure may be biased by normative perceptual, social and motor assumptions.

As well as being virtually present in the web, increasingly our physical-world activities are sensed, tracked and distributed. Some of this, such as CCTV or credit card transactions, happens completely outside our control, some we are aware of peripherally, such as GPS and accelerometer data gathered on smart phones, and some is deliberately gathered, such as medical or fitness data. This personal sensed data offers medical and assistive opportunities both for individuals, and also, through visual analytics, to enable trends to be established, for example, between exercise and illness, or as in a recent Lancaster study, between children's school routes and pollution levels [13]. However, this same data may be used by insurance companies to limit benefits, or increase premiums.

### 3. ECONOMICS AND THE LONG TAIL

A few years ago Web2.0 became the sexy buzzword capturing various technical and social changes in the web. At a technical level AJAX-based interfaces offer rich user experiences, yet often at the expense of accessibility. There are techniques to ameliorate this, or even make potentially more accessible web pages (e.g. 'plain' downloaded HTML that is later rewritten and reformatted using CSS and Javascript). An API, data driven world could improve this, by making the raw data available to accessibility software, but more often makes things worse as Javascript directly accesses data APIs. Underlying the vast majority of web sites today is some form of database or content management system; indeed over 20% of new domains are now being powered by WordPress alone. This has the potential to unlock the data to be re-presented in new forms, but this is by no means guaranteed.

Web 2.0 also created a focus on 'the long tail', rather than focusing on a few large market segments, instead finding ways to serve the very many, very small interests (e.g. Facebook Pages, Google AdSense). This offers hope for the more marginal in society for whom virtual connectivity can create groups large enough to be significant in a market-driven economy. A data-focus can help this, for example, allowing third-party interfaces both on multiple devices (important for the 'next billion' users in the developing economies) and potentially also for different perceptual and physical abilities. However, there is a corresponding danger that providing data APIs could be used as a 'get out' rather than providing truly accessible interfaces.

### 4. EMPOWERMENT AND OPEN ACCESS

The social and data web can be radical and empowering. Social media was at the heart of the Arab Spring, but also of course the recent London riots. The increasing availability of government open data is allowing community and activist groups to create mash-ups reflecting unique concerns. At its best open data means those who would otherwise have been information poor are able to access similar resources to large corporations or government. Furthermore, the web allows alternative data to be made available

alongside the official versions, for example [iraqbodycount.org](http://iraqbodycount.org), which has produced independently verified figures on civilian deaths since the invasion in 2003. Just as blogging sites, such as [wordpress.com](http://wordpress.com) and [blogger.com](http://blogger.com), and YouTube allowed anyone to contribute to the social media of web2.0, a growing number of data publishing sites are available allowing anyone to easily publish data both in more traditional table formats (e.g. Google Fusion Tables) or Semantic Web linked data (e.g. Talis' [Kasabi.com](http://Kasabi.com)). However, potential and opportunity do not mean inevitability. Statistics from the Indian open data initiative suggest that the majority of users are older, more highly educated and male – perhaps simply reinforcing existing power structures [12].

### 5. SUMMARY, POTENTIAL OR THREAT

In each of these areas, we have seen both threats and also potential for a more inclusive society. However, potential must always be actively realised. Standing at a cusp point in the development of the web, we can make the difference to the future.

### 6. REFERENCES

- [1] Adida, B. and Birbeck, M. 2008. *RDFa Primer*. W3C. <http://www.w3.org/TR/xhtml1-rdfa-primer/>
- [2] BBC. 2011. Backstage: Open Data and Resources from the BBC. <http://backstage.bbc.co.uk/>
- [3] Berners-Lee, T., Hendler, J. and Lassila, O., 2001, *The Semantic Web*, *Scientific American*, (17th May 2001), 35-43. <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>
- [4] Bizer, C., Heath, T. and Berners-Lee, T. 2009. Linked data - the story so far. *Int. J. Semantic Web Inf. Syst.*, 5, 3, 1–22.
- [5] Chatterjee, A. Disembodying Cognition. *Language and Cognition*. 2-1 (2010), 79-116
- [6] Clark, A. 1998. *Being There: Putting Brain, Body and the World Together Again*. MIT Press.
- [7] Dix, A. 2011. A Shifting Boundary: the dynamics of internal cognition and the web as external representation. In: *Proceedings of the ACM WebSci'11*, (Koblenz, Germany, June 14-17 2011). <http://journal.webscience.org/436/>
- [8] Dix, A., Gill, S., Ramduny-Ellis, D. and Hare, J. 2011. *Touch IT: exploring the physicality of the world and the design of digital products*. <http://physicality.org/TouchIT/>
- [9] Facebook. 2011. *Open Graph protocol*. <http://developers.facebook.com/docs/opengraph/>
- [10] Gibson, W. 1996. *Idoru*. Viking.
- [11] Guardian. 2011. *Data Store: Facts are Sacred*. <http://www.guardian.co.uk/data>
- [12] Gurstein, M. 2011. *Are the Open Data Warriors Fighting for Robin Hood or the Sheriff?*: <http://gurstein.wordpress.com/2011/07/03/are-the-open-data-warriors-fighting-for-robin-hood-or-the-sheriff-some-reflections-on-okcon-2011-and-the-emerging-data-divide/>
- [13] Pooley, C., Whyatt, J., Walker, M. Davies, G., Coulton, P. and Bamford, W. 2010. Understanding the school journey: integrating data on travel and environment. *Environment and Planning A*, 42, 4, 948-965: <http://dx.doi.org/10.1068/a41405>
- [14] Wilson M. 2002. Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625-36

