

# Web 2.0 and Official Statistics: The UK Perspective

Alan Smith<sup>1</sup>

<sup>1</sup>Data Visualisation Centre, Office for National Statistics, e-mail:  
[alan.smith@ons.gov.uk](mailto:alan.smith@ons.gov.uk)

## Abstract

The World Wide Web is changing and the way in which official statistical outputs are produced and communicated needs to be revised as a reaction to both new opportunities brought about by innovations in web technology and new expectations for UK public sector agencies to engage with users over the internet. The UK Office for National Statistics (ONS) aims to meet this challenge by developing appropriate access mechanisms for data and developing engaging and meaningful visualisations to help users interpret statistical data.

**Keywords:** Web 2.0, dissemination, communication

## 1. Introduction

1.1. There is widespread acceptance among producers of statistics that the internet has now become their primary dissemination channel. This can be seen both in the declining numbers of traditional paper publications and the increasing number of websites and ‘web-only’ releases in National Statistics Institutes (NSIs), the UK included. The move to internet-focussed publication is not a surprising one given user demand and the cost effectiveness of web publishing (versus traditional paper publication). But as the World Wide Web continues to evolve, are statistical producers taking advantage of the evolving opportunities offered by maturing technology and more sophisticated users? This paper considers some of the issues raised and outlines plans by the UK Office for National Statistics to implement a more progressive dissemination strategy.

## 2. The Evolving Internet: ‘Web 2.0’

2.1. The evolution of the World Wide Web away from a disparate collection of static, linked documents and sites towards social and collaborative computing is commonly and collectively referred to as “*Web 2.0*” (O’Reilly, 2004). The ‘2.0’ does not refer to a particular technical upgrade or incremental version of the World Wide Web – instead the term is used to refer to the increasing usage of the web as a genuinely interactive read/write platform.

2.2. Web 2.0 is typified by mass collaboration and the ability of users to add or change content themselves, using various techniques to bring together information ideas and data from different sources in order to produce integrated, meaningful content. Web 2.0 embraces a number of technologies and services to encourage an open, participative environment. These include

blogs, wikis, mashups (a website or application that combines data from multiple sources), social networking, file-sharing, RSS news feeds and even virtual worlds.

- 2.3. Tapscott and Williams (2007) believe that creating an open, participative environment is key to a successful internet strategy. They reflect that the criteria which define success for web publication have now changed: “*the losers launched Web sites. The winners launched vibrant communities. The losers built walled gardens. The winners built public squares. The losers innovated internally. The winners innovated with their users. The losers jealously guarded their data and software interfaces. The winners shared them with everyone*”.
- 2.4. Another significant trend has been driven by the increasing sophistication of web browser software. The term *Rich Internet Applications (RIA)* is often used to describe the form and functionality of web 2.0 websites, many of which offer a user experience similar to desktop applications. Popular examples include sites such as YouTube.com and Flickr.com where users generate and organise content by uploading, sharing, collectively ‘tagging’ and commenting on data. Recently established sites such as Swivel.com and MapTube.com have attempted to do the same with statistical data. In the case of Swivel.com, OECD and Eurostat have signed as source suppliers and have trialled engaging with the Swivel community in their role as official data suppliers.
- 2.5. A key technical enabler of the Web 2.0 toolkit is the *Application Programming Interface (API)*, a set of functions that allow a web-based resource to be remotely accessed by an external client. For example, the auction site eBay.com has published an open, public API that allows external developers (i.e. the wider community, not employed with eBay) to develop their own interfaces for accessing eBay content. Auctionmapper.com is a practical example of this, where an external developer has produced a geographical interface to eBay (allowing users to see auction items on a map centred on their home location). This API-driven model allows eBay to extend its audience (and revenue, shared with the external developer) with minimal investment (just the cost of developing and documenting the API, rather than the cost of developing a multitude of interfaces).

### **3. Web 2.0 and the Public Sector**

- 3.1. Although many of the exemplar instances of Web 2.0 have come from the private sector, it is important to recognise that is not just a commercial vehicle. Indeed the term “Government 2.0” is emerging to describe the application of Web 2.0 approaches to public services. The recent successful presidential election campaign of Barack Obama drew particular emphasis to the potential use of Web 2.0 approaches to generate more open government and greater freedom of information.
- 3.2. Also in the USA, the Library of Congress recently established a pilot project, uploading thousands of old photographs from its image archive to Flickr and invited the wider web community to contribute relevant metadata (in many cases the Library know very little about the photographs, or their subjects) by tagging and linking content. “*This pilot project is a statement about the power of the Web and user communities to help people better acquire information, knowledge and – most importantly – wisdom*” (Raymond, 2008).

- 3.3. In the UK, a key technology and education agency have recently reported on *“the [positive] impact that Web 2.0 can have on the motivation and engagement of pupils. We need to ensure that these benefits are extended to all learners”* (BECTA, 2008). More than half of the teachers questioned in the BECTA survey believed that Web 2.0 technology should be used more often in the classroom. Of particular interest was BECTA’s conclusion that pupils found a sense of ownership and engagement when sharing and publishing their work online, encouraging attention to detail and an overall improved quality of work.
- 3.4. The full impact of Web 2.0 is still emerging and, not surprisingly, has yet to be fully embraced by National Statistics Institutes (NSIs). Garnder (2008) suggests that this may be due to *“the need to be cautious in order to maintain public credibility”*, in addition to long standing resource constraints affecting public sector organisations. But OECD and Eurostat’s engagement with Swivel.com has set a precedent among official statistics suppliers. In experimenting with the Web 2.0 community, both organisations are suggesting that the wider use of their data (and associated commentary) outweighs potential concerns over the credibility of using the (social, chatty) Web 2.0 platform. Furthermore, one of the potential benefits of Web 2.0 is its ability to use the distributed community of the World Wide Web to achieve results that would be difficult within a single organisation – perhaps this is a long term solution to the resource constraints affecting producers of official statistics?
- 3.5. In the UK, there is now a very clear steer from central government that the public sector *must* embrace the Web 2.0 philosophy for disseminating data. Tom Watson MP, the Minister for Transformational Government, noting that *“today the most successful websites are those that bring together content created by the people who use them”* asserts that *“we will be embedding data mash-up into thinking across all of government”* (Watson, 2008). An early product of this assertion has been the ‘Show Us A Better Way’ competition – a £20,000 public prize to be awarded to the best idea for ‘mashing up’ public sector data using Web 2.0 techniques. As part of the competition, several data sources have been made available as web 2.0-style data services, including small area Neighbourhood Statistics from ONS (which has been independently trialling ‘web services’ as a way of feeding data to customers without requiring them to physically download data by visiting the website). By the closing date of the competition (30<sup>th</sup> September 2008), the Showusabetterway.co.uk website had received several hundred entries
- 3.6. ONS’ recent move towards Independence from government has brought with it the release of a new Publication Hub (to act as single source of information on all official statistical releases across government) together with plans for a revitalised NS Online (the official site of ONS). As well as reinforcing the independence remit, there is also a clear intention for ONS to engage with a wider audience through accessible, engaging and relevant content. These goals seem to be in broad agreement with the Web 2.0 philosophy outlined above. But to what extent (if at all) do Web 2.0 approaches, - commonly used for photos, text and video - need to be modified to successfully accommodate official statistics?

## 4. Risks of Web 2.0

- 4.1. We have already considered one risk of Web 2.0 – the possibility that as an informal, unmanaged environment, it has the capacity to undermine authority in official statistics. But are there other risks and issues which need to be considered?
- 4.2. Web 2.0 has its sceptics. Keen (2007) argues that it is philosophically weak, based on “*superficial observations of the world around us rather than deep analysis, shrill opinion rather than considered judgement*”. Such characteristics are potentially alarming to key stakeholders in official statistics where deep analysis and considered judgment represent desired outcomes.
- 4.3. Godwin et al (2008) have considered ‘perceived and real barriers’ to successful use of social media by government: “*Many agencies view the use of social media as a technology issue, instead of a communications tool, and management decisions are often based solely on technology considerations.*” They also cite privacy concerns, particularly in relation to government surveys as well as the lack of a coherent strategy for using the new tools.

## 5. Rich Internet Applications and Data Visualisation

- 5.1. There are now established guidelines for presenting information on the World Wide Web, many of which are mandated standards for the UK public sector - for example, the *e-GIF* interoperability framework and the W3C (the World Wide Web’s governing body) *Web Accessibility Initiative (WAI)*. WAI provides guidance on making content available to people with disabilities (including visual, auditory, physical, speech, cognitive, and neurological disabilities). Tools and techniques such as making content accessible to screen readers, using appropriate colour and styling rules and enabling keyboard navigation are often used in this respect.
- 5.2. But statistical data is often verbose, complex and difficult to interpret - the presentation and publication of it online presents unique problems. Producing content that is ‘web accessible’ does not necessarily meet all of ONS’ user requirements – there is a distinct need to be mindful of ‘*statistical accessibility*’, that is the ability of the wider audience to understand and interpret statistical data relevant to their interests and decision-making requirements.
- 5.3. The Skills for Life Survey estimated that in England, 6.8 million people lacked Entry Level 3 numeracy skills (Letich, 2006), the level identified by the Moser Report as the minimum standard required to function at work and society in general (for example, an ability to add or subtract decimal figures, or work with fractions). Similarly, a recent survey in Wales estimated that 53% of the adult population of Wales do not have Level 1 numeracy skills, defined as an understanding of “*straightforward mathematical information used for different purposes*” (Welsh Assembly Government, 2006).
- 5.4. So, simply presenting tables of data as tables on web pages, downloads, or Web 2.0-style RSS feeds may not allow ONS to reach its full audience. There is a need to look for devices that can help ONS (and the wider web community) produce content that helps users interpret and reason with data – supporting a wide range of users with differing standards of numeracy.

- 5.5. The web browser has quickly established itself as one of the most ubiquitous and frequently-used software tools in the world today. Virtually every desktop PC sold today and an increasing number of mobile phones and other portable devices now have some form of web browser installed as standard.
- 5.6. The web browser has also matured technically in recent years to cope with the sophisticated demands of Web 2.0 applications – they have changed from simple devices used for viewing static content into powerful tools designed for the efficient loading of (big, complex) data, often expressed in open standards such as XML (eXtensible Markup Language). Moreover, open web browser standards such as Scalable Vector Graphics (SVG, a form of XML), along with proprietary browser ‘plug-in’ formats such as Adobe’s Flash and Microsoft’s Silverlight now allow the production of interactive, animated, vector content within the web browser.
- 5.7. The combination of efficient data loading and the ability to deal with interactive vector graphics means that web browsers are potentially effective vehicles for the visual presentation of statistical data to wider audiences, catering for graphs, maps and other engaging content. Ridgway, Nicholson, and McCusker (2004), having developed tactile, interactive interfaces using Adobe Flash for schoolchildren, provided evidence that “*students can deal with greater complexity in data if they have the appropriate tools and visualization support*”. There are parallels here with the BECTA findings, in that making content appealing and tactile can potentially gain greater engagement with the audience, encouraging reasoning with the data.
- 5.8. ONS have recently established a dedicated Data Visualisation Centre within its Methodology Directorate to explore, amongst other things, browser-based presentation of statistics. One recent product of this centre – *CommuterView* – is a good example of the power of web browsers to deliver verbose, complex statistical data in a meaningful way. Taking the full 2001 Census LSOA-level origin and destination commuting data for the UK (including sub-populations), *CommuterView* allows the user to investigate flow patterns in commuting data interactively (by passing the mouse pointer over an interactive SVG map). The symbology used in the product – flow lines with varying saturation and line thickness to denote quantity of flow – is not new, and is indeed based on earlier work, for example Tobler’s *Flow Mapper* product of 1987. What is new is that, by delivering this product inside a standard web browser, the potential audience (and power of the data to inform) is much greater. No special software licenses are required – in contrast to many statistical packages which are often expensive and unnecessarily complicated for ‘casual’ users.
- 5.9. Feedback from *CommuterView* has been extremely positive, mainly from Local Authorities who have struggled to load, analyse and interpret traditional census origin and destination matrices. It is intended that the Data Visualisation Centre will continue to produce browser-based interactive templates for the customised display of ONS data. The unique possibilities offered by the combination of animation and user interactivity offers real potential for transforming ONS’ outputs.

## **6. Towards a progressive dissemination strategy**

- 6.1. ONS' Executive Management Group has recently agreed a web strategy which will embrace the Web 2.0 philosophy – by supporting the development of a Web 2.0-style API that will potentially serve two purposes.
- 6.2. Firstly, it would allow ONS to produce web outputs more flexibly. The development of an API would allow multiple ONS databases to be treated as single conceptual database. The API would allow ONS web applications to seamlessly query multiple data sources in small, efficient transactions. In addition to standard ONS interfaces, customised content (such as the visualisations produced by the Data Visualisation Centre) could source data directly via the API. Developing an API purely for internal purposes has many advantages centred on efficiency and reusability. But the API could also be used more widely.
- 6.3. The API could allow *external* users to access statistical content directly. In a similar way to the eBay/Auctionmapper example discussed earlier, ONS could offer external access to the API. This would potentially allow external developers to develop integrated (i.e. Web 2.0-style mashup) services that accessed ONS data without physically visiting the ONS website. In doing so, it would encourage innovation in the use of ONS data and increase the audience for ONS data beyond those who were prepared to enter [www.statistics.gov.uk](http://www.statistics.gov.uk) as an address in their web browser. An early example of the potential of this approach is the recent 'mashup' produced by the BBC using ONS' Personal Inflation Calculator algorithm and data. Users can enter their personal expenditure in a browser and obtain a personal inflation rate. By entering their postcode or clicking on a map they can then share their results geographically with all other users and add comments based on their own perception of inflation. Most of this functionality (excluding the map) has been available on the ONS website for 18 months. Typically, ONS' inflation calculator (hosted at [www.statistics.gov.uk](http://www.statistics.gov.uk)) gets around 3,000 visits in a month. The BBC version obtained over 250,000 visits on launch day alone.
- 6.4. Of course, the development of an API will not mean that ONS will bypass serving data from its own website (in fact it will make that process more efficient). But the potential to reach a much wider audience using Web 2.0 technology would seem to be endorsed by the example of the Personal Inflation Calculator.

## 7. The road ahead

- 7.1. Tapscott and Williams (ibid.) sound a cautionary note for those who seek to avoid embracing Web 2.0: "*Companies that don't share are finding themselves ever more isolated – bypassed by the networks that are sharing, adapting, and updating knowledge to create value.*" Given the emerging steer (both in the US and the UK) for the public sector to embrace Web 2.0 approaches, it seems unlikely that a policy of avoiding collaborative web-based services is viable in the longer term.
- 7.2. Instead, the challenge will be to overcome some of the issues - real and perceived. For example, issues relating to authority, privacy, transparency, quality, insight and copyright. However, one issue is likely to remain troublesome – how can the success of a Web 2.0-centric strategy be measured? In previous generations, publishers of official statistics could

measure success (at least to a superficial extent) in terms of the number of publications sold, or more recently the number of visits to a single website. The move to open, collaborative environments where data is more fluid and pervasive makes the problem even more difficult. But in a sense the problem is not new – rather than measuring simple flows of data, the real focus lies in measuring the effectiveness of decision-making informed by that data.

- 7.3. The opportunity provided by Web 2.0 – to push official statistics into the heart of open, accountable decision-making – is irresistible. The potential to use Web 2.0 as a vehicle for turning statistics into knowledge is perhaps limited only by the ability and willingness of statistics producers to engage with users and vice versa.

## References

- British Educational Communications and Technology Agency [BECTA] (2008) *Web 2.0 technologies for learning at Key Stages 3 and 4*
- Gardner, J. (2008) *Blogs, wikis and official statistics: new perspectives on the use of Web 2.0 by statistical offices*, IMAODBC conference, Finland
- Godwin B., Campbell S., Levy J., Bounds J. (2008) *Social Media and the Federal Government: Perceived and Real Barriers and Potential Solutions*, Federal Web Managers Council
- Keen, A. (2007) *The Cult of the Amateur: How Today's Internet Is Killing Our Culture and Assaulting Our Economy*, Nicholas Brealey Publishing
- Leitch, S. (2006) *Prosperity for all in the global economy – world class skills* HMSO
- O'Reilly, T. (2004) *O'Reilly Media Web 2.0 conference*
- Office for National Statistics [ONS] (2008). *Personal Inflation Calculator* at <http://news.bbc.co.uk/1/hi/business/7669072.stm>
- Raymond, M. (2008) personal blog at <http://www.loc.gov/blog/?p=233>
- Ridgway J., McCusker S., Nicholson J. (2004) *Uncovering and Developing Student Statistical Competences via New Interfaces*, Curricular Development in Statistics Education, Sweden
- Tapscott D., Williams A. (2007) *Wikinomics: How Mass Collaboration Changes Everything*, Atlantic Books
- Watson, T. (2008) *Power of Information Speech*, 31<sup>st</sup> March 2008, transcript at [http://www.cabinetoffice.gov.uk/about\\_the\\_cabinet\\_office/speeches/watson/080331watson.aspx](http://www.cabinetoffice.gov.uk/about_the_cabinet_office/speeches/watson/080331watson.aspx)
- Welsh Assembly Government [WAG] (2006) *Words Count, Numbers Talk*