

The threads of Web 2.0

Based on a presentation given at the UKSG seminar 'Caught up in Web 2.0? Practical implementations and creative solutions for librarians and publishers', London, 22 November 2007

Explanations of 'Web 2.0' often resort to buzzwords and incomprehensible phrases like 'long tail' and 'crowdsourcing', confounding the reader and making the topic inaccessible to users not familiar with the terms.

This article attempts to take a fresh look at Web 2.0, teasing apart the concept into five fundamental trends or 'threads', each of which is explained in isolation, with reference to illustrative examples.



LEIGH DODDS

Chief Technical Officer
Ingenta

In Stanley Kubrick's film, *2001: A Space Odyssey*, there is a classic scene where the astronaut Dave Bowman is confronted by an alien artefact, the black monolith, and discovers that inside it is a whole universe; an entire world of possibilities. It would be nice to think that Tim O'Reilly and his associates went through the same kind of process in 2004, when they went into a brainstorming session, examined the web and discovered that, far from being a wasteland following the smash and grab of the dotcom boom and bust, there was actually a whole world of new possibilities¹.

At the time, there was a clear resurgence in creativity around web application design, with renewed interest and innovation in online services. It also became apparent that, while the dotcom boom had clearly impacted the level of trust that many larger businesses had in the web as a viable platform and an area for investment, this had not impacted people's ability to interact, do real business, work and research in online environments. Indeed, if anything, those methods of interaction were becomingly increasingly diverse. In short, O'Reilly found that the web was full of people.

A critical mass of online users meant that there were many more options for businesses exploring the online arena. The web *was* a sustainable economic platform, and a huge amount of effort had been invested into standardizing a number of key technologies to ensure that the web would be a stable environment.

Attempts to understand and define Web 2.0 are often confounded by a confusing array of buzzwords

and phrases like 'folksonomy', 'digital natives', 'the wisdom of crowds', and 'long tail'. O'Reilly's own definition¹ runs for several pages and resorts to describing eight general design patterns while contrasting exemplars of Web 1.0 and Web 2.0 companies. While the article is a good exposition of the principles, it is far from being a recipe for Web 2.0 success. Web 2.0 is a difficult concept to pin down because of its many facets: it is definitely not a (single) technology and is really more of a mindset. Web 2.0 is not a sudden leap forward: it is more the culmination of a number of related social, business and technical trends. The aim of this article is to take a slightly different approach to defining Web 2.0 and, rather than resort to buzzwords, to try to tease out and highlight the fundamental trends – the individual 'threads' – that make up the concept. Each of the threads has its roots in the early development of the web and, in combination, should illustrate the evolution of the web we see today.

Users and community

The first of these threads is the notion of community.

Communities have always been an aspect of the online environment. Originally, website owners fretted about adding discussion forums and similar features to their sites in order to build a community, in an attempt to make the service 'sticky' and ensure that users would keep returning.

It was this 'Venus fly-trap' model of engaging with users that gave rise to the notion of the online 'portal' that would become your definitive view of the web, the only resource you would ever need. The Web 2.0 mindset stands this view on its head, recognizing that these communities already exist. For a site to be successful, it needs to reach out to this global audience and identify how it can provide information and services to those existing communities in a way that fits in naturally with their existing ways of interacting, their existing tools and workflow.

The other aspect to the community thread is the recognition that a large, global user community is an asset in its own right and one that can enable new business models and features. These users are not an 'asset' in the original, Web 1.0 world-view that aimed to monetize users through advertising revenues. Instead, a user community is an asset in the sense that it can contribute towards improving an application.

The archetypal example of this is the Amazon recommendation engine. Amazon mine the aggregated data of their users' purchasing patterns in order to recommend other products that they might wish to buy. In essence, this is nothing new; all businesses use market and sales analysis to improve their products and services. The difference is that, on the web, the feedback cycle can be much quicker and there is much more fine-grained data available. Within a web application it is possible to collect much more information about what people actually do, as opposed to what they *say* they do!

Usage analysis allows a service to adapt to its users without them having to sit down and fill in a lot of preferences and wade through configuration options in order to tailor the site for their own needs. The goal is to analyse the behaviour of a user community and use that to improve the user experience. While many view Web 2.0 as being about users annotating, commenting on, or explicitly contributing towards online resources, implicit interactions are just as important. The user experience should improve through the continual ongoing implicit participation that takes place through the use of a service by the combined user community.

Openness and open data

Another key shift towards Web 2.0 involved letting users more directly participate, more explicitly

contribute. This is the area of Web 2.0 which receives the most attention, and is where the second 'thread' in Web 2.0 comes into play.

The Web 1.0 environment was largely about basic document sharing: publishing documents, and letting users download content and exchange documents through e-mail, and was about the creation of websites that served as content repositories.

The natural extension to this was to move away from the Internet as simply a platform for exchanging information, towards the Internet as a platform for creating and working with information; moving from a distribution system towards a collaborative environment. Collaboration and participation are key ideas in Web 2.0. What possibilities arise when there is not only easy access to information, but also the tools to manipulate and enrich that information in order to use it in new, unforeseen, ways? There are a number of illustrative examples that can be drawn upon to answer that question.

Photo sharing has always been a common online activity. Early photo publishing websites like Webshots² provided users with the means to upload photos and share them with family and friends. But now there is one website that dominates the online photo-sharing space: Flickr³. Flickr is a poster-child for Web 2.0 and as a service has always embodied many of the Web 2.0 principles. But what is the difference between Webshots and Flickr; how did Flickr manage to corner the market?

Without underplaying the clear and well designed user interface, it is safe to say that the key innovation in Flickr was making the public sharing of photos the default option. Photos are not locked away so only friends and family can access them, but open and accessible to all. This simple change made it much easier for photos to be shared, and enabled a whole new set of social activities around the content. There are Flickr communities based around photos of particular locations and themes, there are communities that will critique your photos and tell you how to become a better photographer or suggest ways to better crop and present a photograph. None of that would be possible without open, public sharing of photos being the encouraged norm. As a result, Flickr has amassed a huge database of photos⁴ and related metadata, a massive user base, and now offers much more than simple photo-sharing: the site is an excellent

source of freely available stock photography, and a secondary set of applications are now able to build their own business models around the content, offering, for example, photo printing and editing applications⁵.

Another example of where open, public sharing of information creates new possibilities is in the area of social bookmarking. The ability to bookmark favourite websites was an early browser feature, but sites like del.icio.us⁶ launched, which offered the facility to share bookmarks online. On the surface this might have seemed a fairly uninteresting option unless faced with the need to routinely transfer bookmarks between computers. But, like Flickr, del.icio.us proved to be popular and has quickly gathered a database of not just bookmarks but also keywords that users had associated with the content. From this was born the notion of folksonomies⁷: users cataloguing and categorizing content, enabling new ways to find information. Again, publishing seemingly uninteresting, personal information has led to unforeseen social applications and benefits. The del.icio.us homepage is a useful resource for finding new and popular information, as is browsing through the site itself. Sites like Connotea⁸ and Citeulike⁹ are demonstrating the utility of social bookmarking in the academic area.

It is not only content that is moving online, but also the tools to content authoring. For example, the Google Docs¹⁰ service allows users to create documents and spreadsheets using a browser-based application. Again, on the surface this seems like a step backwards: tools like Word and Excel offer a much richer set of features and a better user interface. But they are inherently single-user applications. Online office suites allow users to collaborate on a document at the same time, rather than having to pass it back and forth via e-mail. And, as the information is already online, there is no separate publishing process involved in distributing the finished version.

The extreme end-point of moving more content creation online is to open up the entire authoring and review process, resulting in the creative anarchy exemplified by Wikipedia: anyone can publish anything, anywhere. Whether Wikipedia is a reliable resource or not, it is undeniably a good example of a successful social application¹¹. The debate surrounding Wikipedia, particularly issues such as quality¹², trust¹³ and identity¹⁴, clearly illustrates the potential pitfalls present in this aspect of Web 2.0.

Networked services

The third thread of Web 2.0 is the move towards networked services.

The software industry goes through a cycle on a regular basis where at any one time, either the server or the client is deemed to be the key component in an application. In the beginning was the mainframe, and then came Windows and the rise of the desktop application. And then, later, Sun and Oracle announced that "the network is the computer" and that all users needed were simple networked terminals that would download applications on demand. This cycle continues anew with applications moving off the desktop and onto the web. However, there may be an end in sight, as a kind of equilibrium has been reached.

On one side of this equilibrium there are services like Flickr and del.icio.us that offer users the means to share and annotate information and, crucially, also expose web services or APIs that let that data be accessed from other applications. On the other side there is the evolution of the browser from a simple document viewer into a fully-fledged application platform capable of combining data from multiple sources in order to create new applications.

This ability to mix together multiple services has been dubbed a 'mash-up'. These applications typically run entirely in the browser, on the client, but draw data and functionality from a number of different networked applications. Programmable Web¹⁵ provides a useful resource for exploring the huge variety of applications that are now being built in this way, as well as the growing range of data sources available for remixing.

The potential for networked services to drive usage has been clearly demonstrated by Amazon, who recently reported that web service traffic now accounts for more traffic than all of the global websites combined¹⁶.

New dimensions

The fourth thread of Web 2.0, and a spark for a lot of the creativity that surrounds mash-ups, is the exploration of new forms of information visualization and presentation.

Originally, the web was one-dimensional: people read through documents from start to finish. Documents could be linked together into a

hypertext¹⁷, but the web was still essentially a one-dimensional environment. With the advent of Google Maps, there was suddenly the means to present information in two dimensions, by plotting it on a map. Many kinds of information are better presented in a geographical context, by overlaying onto a map, e.g. statistical data, house prices, etc. It is not surprising that the majority of early mash-ups were about taking an existing data set and re-presenting it on a map. Previously, access to this kind of functionality would have involved a large investment in technology and licensing, as well as specialist knowledge of the mapping domain. But now, geographical visualization is extremely easy, and there are numerous tools that allow users to create custom maps, without any software engineering, in a matter of minutes.

The next obvious step is moving into three dimensions. Services like Google Earth¹⁸ and NASA World Wind¹⁹ provide a three-dimensional interface for presenting geographical data sets. Furthermore, these applications are network enabled so are not just a form of document viewer: data can be streamed directly into the user interface, allowing a dynamic presentation of information.

Services like Second Life²⁰ offer a full three-dimensional, virtual environment within which users can interact and build their own spaces. Both Nature and EduServ have been doing some interesting experiments with the possibilities that a virtual world can open up, ranging from data visualization through to new forms of academic conferences²¹.

Web of data

Mash-ups and networked applications like Google Earth are built on the fifth and final strand that makes up the Web 2.0 tapestry: the web of data.

The original web was about publishing information for humans, but to enable easy integration of networked sources, information needs to be exchangeable in a machine-readable format. In order to allow that integration to happen without a huge amount of co-ordination between the parties involved, those data formats need to be standardized. A successful aspect of Web 2.0 has been the adoption and promotion of key technology standards like XML and RSS that enable lightweight integration with a minimum of technology investment.

The next natural extension to this is a move from machine-readable data to machine-interpretable data. By publishing information about data element means, e.g. is it a stock quote, a document citation, or a weather report, then new kinds of application functionality can be enabled. A machine-interpretable, or semantic web, has always been part of the overall vision for the web and steady progress is being made towards achieving that goal²².

Conclusion

Setting hype and buzzwords aside, and looking at the underlying trends, it is easy to see that Web 2.0 is a natural progression in the development of the Internet. Businesses looking to embrace Web 2.0 need to understand these trends in order to invest effort in the right areas and right directions.

To embrace Web 2.0, publishers should consider:

- identifying and enabling their user communities; ensuring that published documents can easily be used, reused, and fit neatly into their users' workflow
- encouraging mash-ups and public sharing of data, particularly publication metadata
- exploring the potential for richer user interfaces and options for displaying and presenting information
- understanding that they are part of a global network.

Publishing on the web is no longer about simply distributing documents to users; it involves understanding that the information becomes part of a wider networked environment, and the consequences and options that this entails. For information to be found and used it needs to fit into this network, and be seen to be adding value. Conversely, that network is also able to add value to content being published.

References

1. What is Web 2.0
<http://tim.oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> (Accessed 5 February 2008)
2. Webshots
<http://www.webshots.com/> (Accessed 5 February 2008)

3. Flickr
<http://flickr.com> (Accessed 5 February 2008)
4. Flickr: Two billion photos served
<http://blog.wired.com/monkeybites/2007/11/flickr-two-bill.html> (Accessed 5 February 2008)
5. Flickr to get Photo Editing and Upgraded Printing Features
<http://www.profy.com/2007/10/20/flickr-to-get-photo-editing-and-upgraded-printing-features/> (Accessed 5 February 2008)
6. del.icio.us
<http://del.icio.us/> (Accessed 5 February 2008)
7. Folksonomy
<http://en.wikipedia.org/wiki/Folksonomy> (Accessed 5 February 2008)
8. Connotea
<http://www.connotea.org/> (Accessed 5 February 2008)
9. Citeulike
<http://www.citeulike.org/> (Accessed 5 February 2008)
10. Google Docs
<http://docs.google.com/> (Accessed 5 February 2008)
11. Wikipedia: When in Doubt, Multitudes Seek It Out
<http://pewresearch.org/pubs/460/wikipedia> (Accessed 5 February 2008)
12. Larry Sanger, Citizendium, and the Problem of Expertise
http://many.corante.com/archives/2006/09/18/larry_sanger_citizendium_and_the_problem_of_expertise.php (Accessed 5 February 2008)
13. The law of the wiki
http://www.rougthype.com/archives/2005/10/the_law_of_the.php (Accessed 5 February 2008)
14. Accountable Wikipedia, Citizendium moves into beta today
<http://www.centernetworks.com/accountable-wikipedia-citizendium-moves-into-beta-today> (Accessed 5 February 2008)
15. Programmable Web
<http://programmableweb.com/> (Accessed 5 February 2008)
16. Amazon Web Services: Bigger Than Amazon
http://www.readwriteweb.com/archives/amazon_web_services_bigger_than_amazon.php (Accessed 5 February 2008)
17. Hypertext
<http://en.wikipedia.org/wiki/Hypertext> (Accessed 5 February 2008)
18. Google Earth
<http://earth.google.com/> (Accessed 5 February 2008)
19. NASA World Wind
<http://worldwind.arc.nasa.gov/> (Accessed 5 February 2008)
20. Second Life
<http://www.secondlife.com/> (Accessed 5 February 2008)
21. Eduserv in Second Life
<http://www.eduserv.org.uk/foundation/sl> (Accessed 5 February 2008)
22. The Semantic Web in Action
<http://www.sciam.com/article.cfm?id=the-semantic-web-in-action> (Accessed 5 February 2008)

Article © Leigh Dodds

Leigh Dodds
Chief Technology Officer
Ingenta, a division of Publishing Technology
Tel: +44 (0)1225 361044
Fax: +44 (0)1225 361155
E-mail: ldodds@ingenta.com
<http://www.ingenta.com>
<http://allmyeye.blogspot.com>

To view the original copy of this article, published in *Serials*, the journal of the UKSG, click here:

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=21&issue=1&page=4>

For a link to the table of contents for the issue of *Serials* in which this article first appeared, click here:

<http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=21&issue=1>