

Web 3.0 and medicine

Make way for the semantic web, says **Dean Giustini**

Medical blogs and RSS feeds were the hot technology topics in 2003, and we were debating the merits of newer models of scholarly publishing in web 2.0, such as open access and medical wikis.^{w1} Can web 3.0 be here already?

Recently, a neurologist devised an apt medical metaphor for web 3.0. He suggested that, "The development of the graphical web from its early days in 1995 to the social web of late 2007 is comparable to the developing brain." He went on to say that, "Whereas web 1.0 and 2.0 were embryonic, formative technologies, web 3.0 promises to be a more mature web where better 'pathways' for information retrieval will be created, and a greater capacity for cognitive processing of information will be built." (Personal communication, A Wong, 2007.)

So what is web 3.0, and why is it called the semantic web (table)? Although both terms are used interchangeably, they convey slightly different, if complementary, views of the new web. The web 3.0 label is often used as a marketing ploy for "the next big thing." An important feature of web 3.0 is that it enables computers to talk to each other so that they can perform the tasks necessary for us to do our work. However, a primary feature of web 3.0 is that it uses metadata—data about data. This will transform the web into a giant database, and organise it along the lines of PubMed, or one of our trusted medical library catalogues.^{w2}

Somehow, the term semantic web has escaped the reproach of web 3.0, perhaps because it was coined by the respected web expert Sir Tim Berners-Lee in his landmark paper in *Scientific American*.^{w3} His ideas continue to have tremendous salience. Berners-Lee's view is that we need to use semantic annotation to express the meaning latent in web documents, by drawing out inferences in documents deep within the web. As a pioneer in search technology, and director of

the World Wide Web Consortium, Berners-Lee maintains that access to a global "web of data"—what weaves the entire web together into a coherent whole—should help to solve humankind's most complex problems.^{w4}

To understand why we need web 3.0, let's examine the current state of the web. Currently, access to endless reams of unorganised information in web 2.0 shifts the online habits of doctors to searching, not finding. Consequently, medical librarians believe that it is necessary to build better mechanisms for information retrieval.^{w5 w6} As a colleague said to me recently, "we need find engines, not search engines."

In medicine, finding the best evidence has become increasingly difficult, even for librarians. Despite its constant accessibility, Google's search results are emblematic of an approaching crisis with information overload, and this is duplicated by Yahoo and other search engines. Consequently, medical librarians are leading doctors back to trusted sources, such as PubMed, Clinical Evidence, and the Cochrane Library, and even taking them to their library bookshelves instead. Unless better channels of information are created in web 3.0, we can expect the information glut to continue.

Web 3.0 is likely to have a big effect on medicine in 2008. In bioinformatics, it will become more common to process ever larger amounts of data. In fact, experts in bioinformatics already search for data from disparate systems, and they have started to build rich semantic relations into information tools for knowledge discovery. Finally, greater capacity for creating knowledge in medicine will be possible if we have the will to publish clinical data openly and transparently, and subject it to scrutiny.^{w7}

Developing a more personalised health-care system will be an important challenge for doctors in web 3.0. In an era of greater personalisation, treating patients' health problems according to their genetic profiles will depend on using the latest information technologies.^{w8} Even the treatment of new diseases and warning systems for natural disasters will benefit from the merging of epidemiological datasets with virtual, three dimensional tools like Google Earth.

Social software enthusiasts may well find that the new web will be fertile ground for the creation of knowledge. Although already popular, wikis may well serve as platforms for the exploration of web 3.0. One innovative wiki—Wikiproteins—is already using semantic technologies. In contrast to other wikis, Wikiproteins imports data mined from several of the world's leading biomedical databases, such as PubMed, UniProt, and the National Library of Medicine. Its integrated entries are a useful combination of genetic information and scientific literature. Notably,

Glossary

Data mining—a process of knowledge discovery or retrieval of hidden information from data banks and clusters of databases

Mashup—a web application or site that mixes content from multiple sources

Medical wiki—a website or similar online resource that allows users to add and edit medical information collectively

RSS (really simple syndication)—a format for sharing content between different websites

Semantic web—a project that intends to create a universal medium for information exchange from 2008 and beyond by putting documents with computer processable meaning (semantics) on the world wide web

Social tagging—the application of freely chosen labels, or tags, to web documents, web pages, and photo sharing sites, such as www.flickr.com

Web 3.0—a term used to describe the evolution of the web, and our responses to it, in finding and organising new information

the confluence of databases in Wikiproteins yields more than two million factual associations for data mining and over five billion associated pairs.^{w9}

Information retrieval in web 3.0 should be based less on keywords than on intelligent ontological frameworks, such as the National Library of Medicine's Unified Medical Language System, Medline's trusted MeSH vocabulary, or some other tool. The National Library of Medicine is working on automated indexing, which may be part of the solution for searching the biomedical web.^{w10}

The question of whether <http://del.icio.us> and www.connotea.org—two popular social tagging sites—will be useful in web 3.0 remains doubtful.^{w11} Social tagging or "indexing" has limitations because of poor control of synonyms, homonyms, spelling conventions, and other linguistic variations. Think about the myriad ways we describe a heart attack; these variations have enormous implications for searching and require control to optimise retrieval. A smarter medical web is coming. Its two most exciting features will be the better organisation of documents and a deeper use of the knowledge base in medicine. Logically, web 3.0 should bring order to the 21st century web in the same way that Dr John Shaw Billings's *Index Medicus* brought order to medical research back in the 19th century.^{w12}

Dean Giustini UBC biomedical branch librarian, University of British Columbia Biomedical Branch Library, Diamond Healthcare Centre and Vancouver Hospital, BC, Canada V5Z 1M9
dean.giustini@ubc.ca

Competing interests: None declared.

Provenance and peer review: Commissioned; not externally peer reviewed.

This editorial was first published in the *BMJ* (2007;35:1273-4).

At a glance: web 2.0 and 3.0

Web 2.0	Web 3.0
"The document web"	"The data web"
Abundance of information	Control of information
Controversial	No less controversial
"The social web"	"The intelligent web"
The second decade, 2000-9	The third decade, 2010-20
Google as catalyst	Semantic web companies as catalyst
Wisdom of the crowds	Wisdom of the expert
Mashups, fragmentation integration, new tools	
Search, search, search	Why search, when you can find?
Google's Pagerank algorithm	Ontologies, semantic systems
Lawless, anarchic	Standards, protocols, rules
Print and digital	Digital above all else



References w1-w12 are on student.bmj.com.