Introduction to the Semantic Web—Part 1

—by Eric Wolf

This edition of DIY GIScience doesn't emphasize the DIY part as much. Instead, I'm presenting a primer in two parts on what some are calling Web 3.0: the Semantic Web. Let's begin with some historical context and then I will present two of the key concepts that underlie this compelling technology.

In 1989, Sir Tim Berners-Lee had an idea: What if you could make a simple method that would have one document point to another? This wasn't a new idea. In 1945, Vannaver Bush (no relation to George W. Bush) proposed a system which used indexed microfilm that would mimic the "intricate web of trails carried by the cells of the brain." Bush wanted his Memex system to store all of an individual's books, records, and communications in a way which would allow for their instant retrieval. In 1965, Ted Nelson coined the term "hypertext" and later founded a project called Xanadu, a database in which all human knowledge could be stored and retrieved. Apple Computer created the Hypercard software for the Macintosh in 1987 which combined linked information with a graphical user interface.

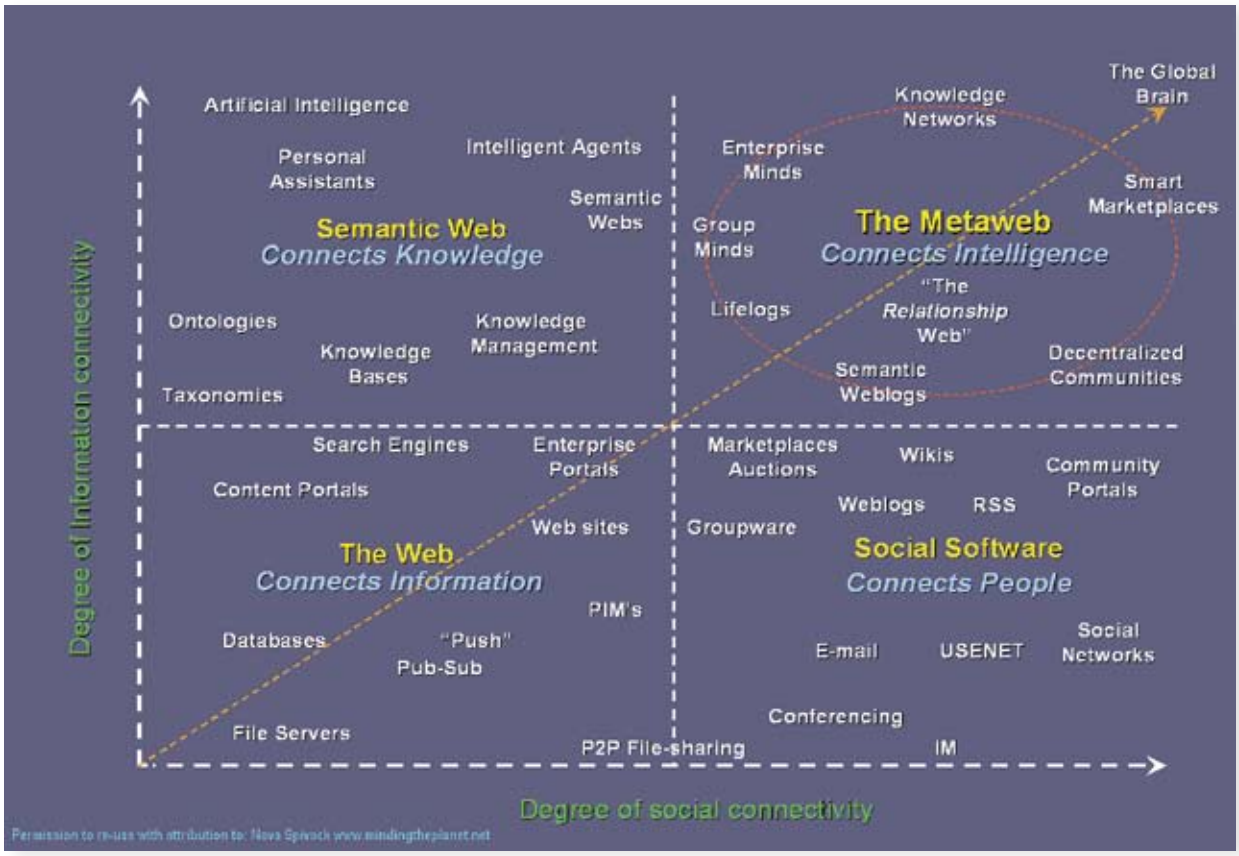
What made Berners-Lee's idea unique was the network. Instead of building a single piece of software to manipulate a database or files stuck on one machine, the future knight (Tim Berners-Lee was named Knight of the British Empire in 2004) created two mechanisms: one which would deliver files on demand and another which would display the files and let the user request others. The requests would be made over a network, in particular, over the Internet, in the form of a Uniform Resource Locator (URL).

At the time, the Internet was an ecosystem of servers and clients with specific, narrow purposes like File Transfer (FTP), remote access (telnet), news feeds (NNTP) and games (MUD). The nascent World Wide Web and the proposed hypertext transfer protocol (HTTP) which made it work were just one of hundreds of different means of exchanging information. But this particular method caught on in a big way. So big that today we equate "the Internet" with "the World Wide Web" and ignore the rest of that ecosystem.

The World Wide Web in many ways is Ted Nelson's Xanadu and Vannaver Bush's Memex. I can find the answer to the most trivial questions with a quick Bing search. I can listen to my favorite music and even find esoteric recordings with web services like Pandora. My written correspondences live in my web-based e-mail account (archive, don't delete!). Even my voice-mail has become absorbed into the web via Google Voice.

But this Xanadu isn't a panacea, and our knight hasn't slain his last dragon. Sir Tim Berners-Lee has been on a crusade to do for your raw data what he did for your documents. The problem with the World Wide Web, as it exists today, is that all of the information is hidden inside chunks of text or blobs of media. Berners-Lee has been working on what is called the Semantic Web as a means to expose these hidden meanings.

Semantic web, p. 26



Like the World Wide Web, the Semantic Web starts off with two basic concepts: all data is to be broken into "triples" and each triple can be accessed via a Uniform Resource Indicator (URI).

A triple consists of a subject, predicate, and an object. The subject is the thing the triple is about. The object is something that modifies or describes the subject. And the predicate determines how the object modifies or describes the subject. For example, say you have a database table that looks like this:

Name	Phone Number	Birthday
George	303-404-5555	5/13/1960
Fred	202-555-1212	7/31/1955

You could form triples like this:
George hasPhoneNumber 303-404-5555

"George" is the subject. "hasPhoneNumber" is the predicate and "303-404-5555" is the object. Each triple of information can be accessed via a URI. URIs happen to look like the URL from the World Wide Web. In fact, a URL is considered a special case of a URI. So a Semantic Web application could use a URI such as "http://friends.org/contacts/george/12345" to access that value. On the Semantic Web, instead of addressing pages where information is embedded in context, the raw data are exposed.

Perhaps the most powerful part of the Semantic Web is that the objects need not be values. Instead, a triple could look like this:

Fred hasFriend http://friends.org/contacts/george/12345

The object could be a URI connecting two pieces of data. In relational database systems, connections between two kinds of data are mapped through relations. When a relation is established, every similar piece of information (other records in the table) has the same kind of relations. In the Semantic Web, the relations can be more fine-grained. And since the "database" is now a "web," two pieces of information need not be in the same system.

If you use Facebook and LinkedIn, you have a great deal of duplicated information. Your basic contact information, your friends, even your picture, may all be duplicated information. If you get a new cell phone, you have to change your contact information in both systems separately. If Facebook and LinkedIn used Semantic Web technology, there could be one set of contact information that each site would access.

Have you ever tried to use a geospatial data clearinghouse—e.g., the Geospatial One-Stop (http://gos.geodata.gov)—to find a dataset? Frequently, clearinghouses don't give raw data. Instead they may simply index the metadata records. The metadata may include contact information for someone who

Semantic web, p. 28, 2nd col.

Semantic web, from p. 26

"Technomanifestos"

The "web" is called the "web" due to its non-linear nature, with hypertext and links taking our thoughts sideways or horizontally (like a web), rather than in a straight, linear fashion (Brate, 2002).

Brate (2002) states that the web works the same way as the human mind, as our thoughts constantly jump and switch subjects. We do not think in the linear way that a book or manual might read, our thoughts are more random and abstract.

Recently I have been concerned that I cannot seem to maintain linear conversations. Brate's theory has comforted me in that I now know that my seemingly out of order thought processing is in fact, quite normal.

The web, like our minds, contains a huge amount of information. The good, the bad and the ugly. Every creation is either a reflection of reality, or perhaps the embodiment of a fantasy, an imagination of the human mind.

Kevin Kelley wrote an article for "*Wired*" online magazine entitled "We are the Web." Kelley states that "we are the web" because we create, update, contribute, alter, debate and discuss what is there. Without us, there would be no web.

I think the web continues to grow and develop as we (humanity) keep growing and adding to its content. The web, like the world, does not discriminate against who or what is there it just is, as we just are. Politics comes into play on the web as it does in real life, because its contents are simply an extension of ourselves (with an IP address).—Excerpts from a blog on http://anothermediastudent.wordpress.com.

Bibliography

Adam Brate. 2002. Everything is deeply intertwingled. In: Ted Nelson and Tim Berners-Lee (eds), *Technomanifestos: Visions from the information revolutionaries*, New York and London: Texere, 2002.

Kevin Kelly, We are the We. *Wired*, August 2005.

The British-born software engineer brought order to cyberspace by creating the World Wide Web (http://www.time.com). He is also behind the semantic web concept.

is supposed to be able to provide the data but that information may be out of date. You get a disconnected number or they don't work for the data provider anymore. If these clearinghouses used the Semantic Web, even if you didn't get access to the raw data, the metadata would likely point to a dataset maintained by the data provider. When the contact person for a set of data changed, the record would be updated and you would be more successful tracking down needed information.

The Semantic Web arises out of a need to break data out of silos of disconnected databases and the contextual shells of web pages. But this only scratches the surface of the capability and complexity of the Semantic Web. In part two of this article, I will explain how the Semantic Web enables more intelligent searching. How, if Sir Tim Berners-Lee's vision holds out, instead of feeding bread crumbs into a search engine and sorting through piles of meaningless links, intelligent agents will piece together the raw data from the right places to provide solutions.

The future of government services—Gov 2.0

In the latest SpatialRoundtable.com discussion on May 12th, ESRI industry solutions manager Christopher Thomas addressed the emerging trend of governments using Web 2.0 technology to improve service—Gov 2.0. Executives from all levels of government, as well as media and geographic information system (GIS) thought leaders, were engaged in a dynamic online conversation to respond to the question: "Can the GIS community provide a platform for engagement that empowers citizens?"

"Government needs to meet high expectation levels," says Thomas. "Citizens want to interact with their government online, and GIS is playing an important role in delivering these valuable Web services."

There are two kinds of Gov 2.0 enthusiasts. One group is engaged in studying emerging technology, including cloud computing and crowd sourcing, while the other primarily sees Gov 2.0 as a movement to improve government. But a separate group—the largest group—comprises citizens who are generally unaware of Gov 2.0. In this environment, GIS is the key platform for delivering transparency and accountability.