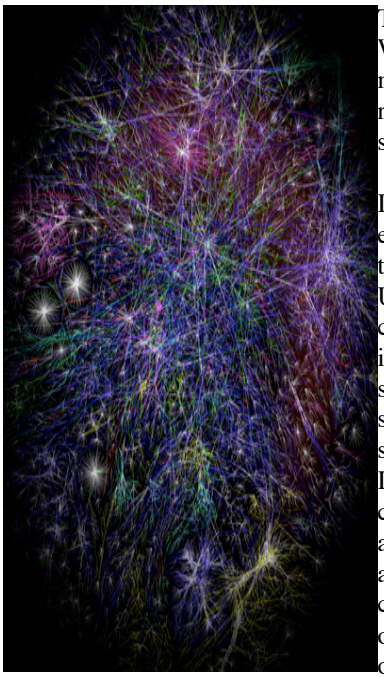
What is the Semantic Web and Why Should You Care?

By Damian Gessler, iPlant Semantic Web Architect, dgessler@iplantcollaborative.org.



The Semantic Web is the linking of concepts and statements on the World Wide Web in a manner discernable for contextualized decision-making by machines. To understand the Semantic Web today and its relevancy to iPlant and science, we need to go back and look for the signatures and commonalities of its intellectual antecedents.

In the early 1990s, computing was still a rather traditional, structured endeavor. Yet transparent to the user, underneath lay a creative and non-traditional foundation. A quarter century earlier the research arm of the U.S. Department of Defense led an effort to construct an informatics and communications network that could withstand unpredictable service interruptions while maintaining connectivity across a broad swath of sub-networks, protocols, and operating systems. The designers were sensitive to the well-known issues with circuit-based communications systems—the telephone system being the canonical example. Instead of addressing this problem with hardware, such as adding more circuits or increasing circuit redundancy and reliability, information architects such as Lawrence Roberts, Leonard Kleinrock, Robert Kahn, and Vinton Cerf struck a bold path: they diminished the role of the circuit to being simply an infrastructural artifact and moved the "smarts" of routing and marshalling data from the circuit system level to the level of operations on metadata riding with the data itself. Under this new model, data was broken into "packets" each containing a copy of its destination address. Nodes in the network would look at incoming packets and determine the best routes and hops based on local conditions. We know this system today as the Internet—brilliant invention, no longer completely understood by any one person, truly bigger than the sum of its parts. The decision to move the routing away from a top-down systems level to a myopic, information-centric model was by no means obvious and had its share of detractors. Yet this move allowed a true network to evolve and scale.

Meanwhile, hypertext—the linking of content from one document to that of another—was maturing under the thinking of Vannevar Bush, Ted Nelson, Douglas Engelbert, and others. But it was not until the early 1990s that these two inventions—the Internet and hypertext—came together in what we now call the World Wide Web. The World Wide Web is a document-centric, hypertext-linked network of content riding on the Internet. The fabulously successful World Wide Web is the work of many, but like the Internet, the brainchild of a few (Sir Tim Berners-Lee along with Robert Cailliau). It is architecturally based on the perspective that a network of components (e.g., documents) creates a value greater than the sum of its parts. Both the Internet and the World Wide Web were paradigm-changing because they enabled a larger usage: they solved a problem in a way that enabled the very use of the system to generate a greater, synergistic value. Building on this intellectual foundation, the Semantic Web takes this approach to its next logical extension.

Connecting computers, like dumb machines exchanging bits, gets one only "so far." Interoperability is not integration. Nor does mere aggregation of data generate synergistic value in knowledge. Designers of the Semantic Web realize that knowledge comes not from just words but from connecting contextualized information. Humans do this every day, but for computers, discerning meaning so as to determine suitability-for-purpose requires computable semantics. Some might think this necessitates some artificial intelligence more appropriate for science fiction than science practice, but the history of modern informatics—the Internet and the World Wide Web—hints that this is not so. Like the Internet and the World Wide Web, the way this is being done in the Semantic Web community is non-obvious, highly infrastructural, and potentially game-changing. The Semantic Web has application to iPlant because, in building a cyberinfrastructure for plant scientists, iPlant has a constituency deep in distributed, interdependent, yet un-integrated data and knowledge. The Semantic Web enables an integration of those data and services driven by the decentralized creative contribution of others; its intellectual antecedents hint that the Semantic Web has the right 'pedigree' for the challenges we face.