

The established system of journals for communicating the results of scientific research is already being challenged by the existence of the web. But we are only in the early days of a new Internet revolution, one which will have a deeper and more disruptive impact on scientific, and other, web publishing, and have profound implications for the web itself. An emerging successor to the web, the [Semantic Web](#), will likely profoundly change the very nature of how scientific knowledge is produced and shared, in ways that we can now barely imagine.

The Web was designed as an information space, with the goal that it should be useful not only for human-human communication, but also that machines would be able to participate and help users communicate with each other. A major obstacle to this goal is the fact that most information on the Web is designed solely for human consumption. Computers are better at handling carefully structured and well-designed data, yet even where information is derived from a database with well-defined meanings, the implications of those data are not evident to a robot browsing the web. More information on the web needs to be in a form that machines can 'understand' rather than simply display.

The concept of machine-understandable documents does not imply some magical artificial intelligence allowing machines to comprehend human mumblings. It relies solely on a machine's ability to solve well-defined problems by performing well-defined operations on well-defined data. So, instead of asking machines to understand people's language, the new technology, like the old, involves asking people to make some extra effort, in repayment for which they will get substantial new functionality -- just as the extra effort of producing HTML markup ([HyperText Markup Language](#)) is outweighed by the benefit of having content searchable on the web.

A new set of languages is now being developed to make more web content accessible to machines. The [Semantic Web Activity](#) run by the World Wide Web consortium is defining new web technologies that will enable successively better tools that make it easier for people to create machine-readable content and make it widely available.

What impact might this have on scientific publishing? In the next few years, we expect that tools for publishing papers on the web will automatically help users to include more of this machine-readable markup in the papers they produce. Whereas current tools using XML (Extensible Markup Language) can allow a user to assert that some part of a document is about an 'experiment', the new languages will let the scientist express that the experiment uses certain chemicals and reagents; that the system used involved some particular organic matter; that the experiment produced gels with certain DNA information on them (and that the images of these gels are located in particular places on the web); and so on.

Papers that include this new markup language will be found by new and better search engines, and users will thus be able to issue significantly more precise queries. More importantly, experimental results will themselves be published on the web, outside of the context of a research paper. So a scientist could design and run an experiment, and create an emerging web page containing the information that he or she wants to share with trusted colleagues (see [Figure](#)). Finding out about experiments and studies in progress will be easy, and work will be able to be modified as a result of interaction with peers, with less need to wait for formal publication. Just as preprints challenge established journals' online versions, these new 'papers in progress' will be a significant challenge to online scientific publishers.

In the long run, the effects on publishing may be far more profound. There is an eternal conflict between operating rapidly as a small group and taking the time to communicate more widely. The former is more efficient but produces a subculture whose concepts and results are not understood by others. The latter can be painfully slow. The world works as a spectrum between these extremes, with a tendency to start small - from the personal idea - and filter over time towards a wider commonality of concept. The joining together of subcultures when there is a need for a wider common language is an essential process in the development of human communication.

The semantic web will facilitate the development of automated methods for helping users to understand the content produced by those in other scientific disciplines. On the semantic web, one will be able to produce machine-readable content that will provide, say, automated translation between the output of a scientific device and the input of a datamining package used in some other discipline, or a self-evolving translator that allows one group of scientists to directly interact with the technical data produced by another.

These new products will allow users to create relationships that allow communication when the commonality of concept has not (yet) led to a commonality of terms. The semantic web will provide unifying underlying technologies to allow these concepts to be progressively linked into a universal web of knowledge, and will therefore help to break down the walls erected by lack of communication, and allow researchers to find and understand products from other scientific disciplines. The very notion of a journal of medicine separate from a journal of bioinformatics, separate from the writings of physicists, chemists, psychologists and even kindergarten teachers, will someday become as out of date as the print journal is becoming to our graduate students.

Does this sound like a crazy science-fiction dream? A decade ago, who would have believed a web of text, conveyed by computer, would challenge a 200-year-old tradition of academic publishing?

[click here to see figure and legend](#)

[Tim Berners-Lee](#) is the inventor of the World Wide Web and is the Director of the [World Wide Web Consortium](#) at the Massachusetts Institute of Technology (MIT), 545 Technology Square Room NE43-356 Cambridge MA 02139 USA. [James Hendler](#) is a Professor in the Computer Science Department at the University of Maryland, College Park, MD 20853 USA, and responsible for research on agent-based computing at the US [Defense Advanced Research Projects Agency](#).