

**Testimony of Sir Timothy Berners-Lee
CSAIL Decentralized Information Group
Massachusetts Institute of Technology**

**Before the
United States House of Representatives
Committee on Energy and Commerce
Subcommittee on Telecommunications and the Internet**

**Hearing on the "Digital Future of the United States: Part I -- The Future of
the World Wide Web"**

This document on the Web [<http://dig.csail.mit.edu/2007/03/01-ushouse-future-of-the-web>] [[PDF](#)]

Chairman Markey, Ranking Member Upton, and Members of the Committee. It is my honor to appear before you today to discuss the future of the World Wide Web. I would like to offer some of my experience of having designed the original foundations of the Web, what I've learned from watching it grow, and some of the exciting and challenging developments I see in the future of the Web. Though I was privileged to lead the effort that gave rise to the Web in the mid-1990s, it has long passed the point of being something designed by a single person or even a single organization. It has become a public resource upon which many individuals, communities, companies and governments depend. And, from its beginning, it is a medium that has been created and sustained by the cooperative efforts of people all over the world.

To introduce myself, I should mention that I studied Physics at Oxford, but on graduating discovered the new world of microprocessors and joined the electronics and computer science industry for several years. In 1980, I worked on a contract at CERN, the European Particle Physics Laboratory, and wrote for my own benefit a simple program for tracking the various parts of the project using linked note cards. In 1984 I returned to CERN for ten years, during which time I found the need for a universal information system, and developed the World Wide Web as a side project in 1990. In 1994, the need for coordination of the Web became paramount, and I left to come to MIT, which became the first of now three international host institutes for the [World Wide Consortium](#) (W3C). I have directed W3C since that time. I hold the [3Com Founders chair](#) at MIT where I pursue research on advanced Web technologies with the [MIT Decentralized Information Group](#). The testimony I offer here today is purely my own opinion and does not necessarily reflect the views of the World Wide Web Consortium or any of its Members.

The special care we extend to the World Wide Web comes from a long tradition that democracies have of protecting their vital communications channels. We nurture and protect our information networks because they stand at the core of our economies, our democracies, and our cultural and personal lives. Of course, the imperative to assure the free flow of information has

only grown given the global nature of the Internet and Web. As a Federal judge said in defense of freedom of expression on the Internet:

The Internet is a far more speech-enhancing medium than print, the village green, or the mails.... The Internet may fairly be regarded as a never-ending worldwide conversation.[\[1\]](#)

Therefore it is incumbent on all of us to understand what our role is in fostering continued growth, innovation, and vitality of the World Wide Web. I am gratified that the United States and many other democracies around the world have taken up this challenge. My hope today is to help you to explore the role this committee and this Congress has in building upon the great advances that are in store for the Web.

I. Foundations of the World Wide Web

The success of the World Wide Web, itself built on the open Internet, has depended on three critical factors: 1) unlimited links from any part of the Web to any other; 2) open technical standards as the basis for continued growth of innovation applications; and 3) separation of network layers, enabling independent innovation for network transport, routing and information applications. Today these characteristics of the Web are easily overlooked as obvious, self-maintaining, or just unimportant. All who use the Web to publish or access information take it for granted that any Web page on the planet will be accessible to anyone who has an Internet connection, regardless whether it is over a dialup modem or a high speed multi-megabit per second digital access line. The last decade has seen so many new ecommerce startups, some of which have formed the foundations of the new economy, that we now expect that the next blockbuster Web site or the new homepage for your kid's local soccer team will just appear on the Web without any difficulty.

Today I will speak primarily about the World Wide Web. I hesitate to point out that the Web is just one of the many applications that run on top of the Internet. As with other Internet applications such as email, instant messaging, and voice over IP, the Web would have been impossible to create without the Internet itself operating as an open platform. [\[2\]](#)

A. Universal linking: Anyone can connect to anyone, any page can link to any page

How did the Web grow from nothing to the scale it is at today? From a technical perspective, the Web is a large collection of Web pages (written in the standard HTML format), linked to other pages (with the linked documents named using the URI standard), and accessed over the Internet (using the HTTP network protocol). In simple terms, the Web has grown because it's easy to write a Web page and easy to link to other pages. The story of the growth of the World Wide Web can be measured by the number of Web pages that are published and the number of links between pages. Starting with one page and one site just about 15 years ago, there are now over 100,000,000 web sites[\[3\]](#) with an estimated over 8 billion publicly accessible pages as of 2005. What makes it easy to create links from one page to another is that there is no limit to the

number of pages or number of links possible on the Web. Adding a Web page requires no coordination with any central authority, and has an extremely low, often zero, additional cost. What's more, the protocol that allows us to follow these links (HTTP) is a non-discriminatory protocol. It allows us to follow any link at all, regardless of content or ownership. So, because it's so easy to write a Web page, link to another page, and follow these links around, people have done a lot of this. Adding a page provides content, but adding a link provides the organization, structure and endorsement to information on the Web which turn the content as a whole into something of great value.

A current example of the low barriers to reading, writing and linking on the Web is the world of blogs. Blogs hardly existed five years ago, but have become an enormously popular means of expression for everything from politics to local news, to art and science. The low barriers to publishing pages and abundance of linking ability have come together, most recently with blogs, to create an open platform for expression and exchange of all kinds.^[4] The promise of being able to reach anyone over a communications system that will carry virtually anything (any type of information) is somewhat like other infrastructures we depend upon: the mail system, the road system, and the telephone system. It stands in contrast to more closed systems such as the broadcast or cable television networks. Those closed systems perform valuable functions as well, but their impact in society is different and less pervasive.

The universality and flexibility of the Web's linking [architecture](#) has a unique capacity to break down boundaries of distance, language, and domains of knowledge. These traditional barriers fall away because the cost and complexity of a link is unaffected by most boundaries that divide other media. It's as easy to link from information about commercial law in the United States to commercial law in China, as it is to make the same link from Massachusetts' Commercial Code to that of Michigan. These links work even though they have to traverse boundaries of distance, network operators, computer operating systems, and a host of other technical details that previously served to divide information. The Web's ability to allow people to forge links is why we refer to it as an abstract information space, rather than simply a network. Other open systems such as the mails, the roads or the telephones come to perform a function in society that transcends their simple technical characteristics. In these systems, phone calls from the wireline networks travel seamlessly to wireless providers. Mails from one country traverse borders with minimal friction, and the cars we buy work on any roads we can find. Open infrastructures become general purpose infrastructure on top of which large scale social systems are built. The Web takes this openness one step further and enables a continually evolving set of new services that combine information at a global scale previously not possible. This universality has been the key enabler of innovation on the Web and will continue to be so in the future.

B. Open Foundation for Information-driven Innovation

The Web has not only been a venue for the free exchange of ideas, but also it has been a platform for the creation of a wide and unanticipated variety of new services. Commercial applications including eBay, Google, Yahoo, and Amazon.com are but a few examples of the extraordinary innovation that is possible because of the open, standards-based, royalty-free

technology that makes up the Web. Whether developing an auction site, a search engine, or a new way of selling consumer goods, e-commerce entrepreneurs have been able develop new services with confidence that they will be available for use by anyone with an Internet connection and a Web browser, regardless of operating system, computer hardware, or the ISP chosen by that user.^[5] Innovation in the non-commercial and government domains has been equally robust. Early Web sites such as Thomas have led the way in efforts to make the legislative process more open and transparent, and non-commercial sites such as the Wikipedia have pioneered new collaborate styles of information sharing. The flexibility and openness inherent in Web standards also make this medium a powerful foundation on which to build services and applications that are truly accessible for people with disabilities, as well as people who need to transform content for purposes other than that for which it was originally intended.

The lesson from the proliferation of new applications and services on top of the Web infrastructure is that innovation will happen provided it has a platform of open technical standards, a flexible, scalable architecture, and access to these standards on royalty-free (\$0 fee patent licenses) terms. At the World Wide Web Consortium, we will only standardize technology if it can be implemented on a royalty-free basis. So, all who contribute to the development of technical standards at the W3C are required to agree to provide royalty-free licenses to any patents they may hold if those patents would block compliance with the standard. ^[6] Consider as a comparison the very successful Apple iTunes+iPod music distribution environment. This integration of hardware, software, Web service shows an intriguing mix of proprietary technology and open standards. The iTunes environment consists of two parts: sales of music and videos, and distribution of podcasts. The sale of music is managed by a proprietary platform run by Apple with the aim of preventing copyright infringement. However, because Apple uses closed, non-standard technology for its copy protection (known as Digital Rights Management), the growth is seen as limited. In fact, Apple CEO Steve Jobs recently wrote that the market for online music sales is being limited by the lack of open access to DRM technology.^[7] By contrast, the podcast component of iTunes is growing quite dramatically, providing a means for many small and large audio and video distributors to share or sell their wares on the Web. Unlike the music and video sales, podcasts are based on open standards, assuring that it's easy to create, edit and distribute the podcast content.

C. Separation of Layers

When, seventeen years ago, I designed the Web, I did not have to ask anyone's permission. The Web, as a new application, rolled out over the existing Internet without any changes to the Internet itself. This is the genius of the design of the Internet, for which I take no credit. Applying the age old wisdom of design with interchangeable parts and separation of concerns, each component of the Internet and the applications that run on top of it are able develop and improve independently. This separation of layers allows simultaneous but autonomous innovation to occur at many levels all at once. One team of engineers can concentrate on developing the best possible wireless data service, while another can learn how to squeeze more and more bits through fiber optic cable. At the same time, application developers such as myself can develop new protocols and services such as voice over IP, instant messaging, and peer-to-

peer networks. Because of the open nature of the Internet's design, all of these continue to work well together even as each one is improving itself.

II. Looking forward

Having described how the Web got to where it is, let us shift to the question of where it might go from here. I hope that I've already persuaded you that the evolution of the Web is not in the hands of any one person, me or anyone else. But I'd like to highlight three areas in which I expect exciting developments in the near future. First, the Web will get better and better at helping us to manage, integrate, and analyze data. Today, the Web is quite effective at helping us to publish and discover documents, but the individual information elements within those documents (whether it be the date of any event, the price of a item on a catalog page, or a mathematical formula) cannot be handled directly as data. Today you can see the data with your browser, but can't get other computer programs to manipulate or analyze it without going through a lot of manual effort yourself. As this problem is solved, we can expect that Web as a whole to look more like a large database or spreadsheet, rather than just a set of linked documents. Second, the Web will be accessible from a growing diversity of networks (wireless, wireline, satellite, etc.) and will be available on a ever increasing number of different types of devices. Finally, in a related trend, Web applications will become a more and more ubiquitous throughout our human environment, with walls, automobile dashboards, refrigerator doors all serving as displays giving us a window onto the Web.

A. Data Integration

Digital information about nearly every aspect of our lives is being created at an astonishing rate. Locked within all of this data is the key to knowledge about how to cure diseases, create business value, and govern our world more effectively. The good news is that a number of technical innovations (RDF which is to data what HTML is to documents, and the Web Ontology Language (OWL) which allows us to express how data sources connect together) along with more openness in information sharing practices are moving the World Wide Web toward what we call the Semantic Web. Progress toward better data integration will happen through use of the key piece of technology that made the World Wide Web so successful: the link. The power of the Web today, including the ability to find the pages we're looking for, derives from the fact that documents are put on the Web in standard form, and then linked together. The Semantic Web will enable better data integration by allowing everyone who puts individual items of data on the Web to link them with other pieces of data using standard formats.

To appreciate the need for better data integration, compare the enormous volume of experimental data produced in commercial and academic drug discovery laboratories around the world, as against the stagnant pace of drug discovery. While market and regulatory factors play a role here, life science researchers are coming to the conclusion that in many cases no single lab, no single library, no single genomic data repository contains the information necessary to discover new drugs. Rather, the information necessary to understand the complex interactions

between diseases, biological processes in the human body, and the vast array of chemical agents is spread out across the world in a myriad of databases, spreadsheets, and documents.

Scientists are not the only ones who need better data integration. Consider the investment and finance sector, a marketplace in which profit is generated, in large part, from having the right information, at the right time, and reaching correct conclusions based on analysis and insight drawn from that information. Successful investment strategies are based on finding patterns and trends in an increasingly diverse set of information sources (news, market data, historical trends, commodity prices, etc.). Leading edge financial information providers are now developing services that allow users to easily integrate the data they have, about their own portfolios or internal market models, with the information delivered by the information service. The unique value creation is in the integration services, not in the raw data itself or even in the software tools, most of which will be built on open source components.

New data integration capabilities, when directed at personal information, pose substantial privacy challenges which are hardly addressed by today's privacy laws. The technology of today's Web already helps reveal far more about individuals, their behavior, their reading interest, political views, personal associations, group affiliations, and even health and financial status. In some cases, this personal information is revealed by clever integration of individual pieces of data on the Web that provide clues to otherwise unavailable information. In other cases, people actually reveal a lot about themselves, but with the intent that it only used in certain contexts by certain people. These shifts in the way we relate to personal information require serious consideration in many aspects of our social and legal lives. While we are only just beginning to see these shifts, now is the time to examine a range of legal and technical options that will preserve our fundamental privacy values for the future without unduly stifling beneficial new information processing and sharing capabilities. Our research group at MIT is investigating new technologies to make the most of the Semantic Web, as well as both technical and public policy models that will help bring increased transparency and accountability to the World Wide Web and other large scale information systems.[\[8\]](#) Our belief is that in order to protect privacy and other public policy values, we need to research and develop new technical mechanisms that provide great transparency into the ways in which information in the system is used, and provide accountability for those uses with respect to what ever are the prevailing rules.

B. Network Diversity and Device Independence

The Web has always been accessible from a variety of devices over a variety of networks. From early on, one could browse the Web from a Macintosh, a Windows PC or a Linux-based computer. However, for a long time the dominant mode of using the Web was from some desktop or laptop computer with a reasonably large display. Increasingly, people will use non-PC devices that have either much smaller or much larger displays, and will reach the Internet through a growing diversity of networks. At one end of this spectrum, the devices will seem more like cell phones. At the other end, they will seem more like large screen TVs. There are, of course, technical challenges associated with squeezing a Web page designed for a 17 inch screen into the two to four inch display available on a mobile phone or PDA. Some of these will be

solved through common standards and some through innovative new interface techniques. All of this means more convenience for users and more opportunity for new Web services that are tailored to people who are somewhere other than their desks.

Growth in access networks and Web-enabled applications presents a number of important opportunities. For example, more robust, redundant network services together with innovative uses of community-based social networks on the Web are coming to play an increasing role in areas such as emergency planning and notification.^[9] Reports about ad hoc communication networks supporting disaster relief efforts are just one illustration of the benefit of the openness, flexibility and accessible of the Internet and Web. This one area is a microcosm of many of the issues that we are discussing today, because in order to work well it requires seamless integration of diverse types of data; repurposing that data instantly into valid formats for a myriad of different Web devices; and including appropriate captions, descriptions, and other necessary accessibility information. I would encourage all web sites designers to ensure that their material conforms not only to W3C standards, but also to guidelines for accessibility for people with disabilities, and for mobile access.

C. Ubiquitous Web Applications

In the future, the Web will seem like it's everywhere, not just on our desktop or mobile device. As LCD technology becomes cheaper, walls of rooms, and even walls of buildings, will become display surfaces for information from the Web. Much of the information that we receive today through a specialized application such as a database or a spreadsheet will come directly from the Web. Pervasive and ubiquitous web applications hold much opportunity for innovation and social enrichment. They also pose significant public policy challenges. Nearly all of the information displayed is speech but is being done in public, possibly in a manner accessible to children. Some of this information is bound to be personal, raising privacy questions. Finally, inasmuch as this new ubiquitous face of the Web is public, it will shape the nature of the public spaces we work, shop, do politics, and socialize in.

D. The Web is Not Complete

Progress in the evolution of the Web to date has been quite gratifying to me. But the Web is by no means finished.

The Web, and everything which happens on it, rest on two things: technological protocols, and social conventions. The technological protocols, like HTTP and HTML, determine how computers interact. Social conventions, such as the incentive to make links to valuable resources, or the rules of engagement in a social networking web site, are about how people like to, and are allowed to, interact.

As the Web passes through its first decade of widespread use, we still know surprisingly little about these complex technical and social mechanisms. We have only scratched the surface of what could be realized with deeper scientific investigation into its design, operation and impact

on society. Robust technical design, innovative business decisions, and sound public policy judgment all require that we are aware of the complex interactions between technology and society. We call this awareness Web Science: the science and engineering of this massive system for the common good.^[10] In order to galvanize Web Science research and education efforts, MIT and the University of Southampton in the United Kingdom have created the [Web Science Research Initiative](#). In concert with an international Scientific Advisory Council of distinguished computer scientists, social scientists, and legal scholars, WSRI will help create an intellectual foundation, educational atmosphere, and resource base to allow researchers to take the Web seriously as an object of scientific enquiry and engineering innovation.

III. Conclusion

So how do we plan for a better future, better for society?

We ensure that that both technological protocols and social conventions respect basic values. That Web remains a universal platform: independent of any specific hardware device, software platform, language, culture, or disability. That the Web does not become controlled by a single company -- or a single country.

By adherence to these principles we can ensure that Web technology, like the Internet, continues to serve as a foundation for bigger things to come. It is my hope, Chairman Markey, members of the committee, that an understanding of the nature of the Web will guide you in your future work, and that the public at large can count on you to hold these values to the best of your ability. I am grateful for the opportunity to appear before you and am ready to help your efforts in future.

Notes

- [1] American Civil Liberties Union v. Reno, 929 F. Supp. 824, 844 (E.D. Pa. 1996) (Dalzell, J.)
- [2] Kapor, M. and Weitzner, D. "Social and Industrial Policy for Public Networks: Visions for the Future". Harasim and Walls, eds. Global Networks: Computers and International Communication. Oxford University Press. Oxford. (1994)
- [3] Netcraft February 2007 Web Server Survey.
http://news.netcraft.com/archives/web_server_survey.html
- [4] Weinberger, D., Small Pieces Loosely Joined: A Unified Theory of the Web. Perseus Books. (2002)
- [5] Note that due to failure by some browser vendors to comply fully with standards, web site developers sometimes have to go to extra trouble to make it so that their sites actually work properly on all browsers.
- [6] Overview and Summary of the W3C Patent Policy, <http://www.w3.org/2004/02/05-patentsummary.html>. W3C Patent Policy. D. Weitzner, Standards, Patents and the Dynamics of Innovation on the World Wide Web. <http://www.w3.org/2004/10/patents-standards-innovation.html>
- [7] Jobs wrote on the Apple Web site: "Imagine a world where every online store sells DRM-free music encoded in open licensable formats. In such a world, any player can play music purchased from any store, and any store can sell music which is playable on all players. This is clearly the best alternative for consumers, and Apple would embrace it in a heartbeat. If the big four music companies would license Apple their music without the requirement that it be protected with a DRM, we would switch to selling only DRM-free music on our iTunes store. Every iPod ever made will play this DRM-free music," Steve Jobs, Thoughts on Music (February 6, 2007), <http://www.apple.com/hotnews/thoughtsonmusic/>
- [8] Weitzner, Abelson, Berners-Lee, Hanson, Hendler, Kagal, McGuinness, Sussman, Waterman, Transparent Accountable Data Mining: New Strategies for Privacy Protection,; [MIT CSAIL Technical Report MIT-CSAIL-TR-2006-007](http://www.mit.edu/~csail/technicalreports/2006-007/MIT-CSAIL-TR-2006-007) (27 January 2006).
- [9] B. Shneiderman, and J. Preece, PUBLIC HEALTH: 911.gov, Science 315 (5814), 944 (16 February 2007)
- [10] "Creating a Science of the Web" Tim Berners-Lee, Wendy Hall, James Hendler, Nigel Shadbolt, Daniel J. Weitzner. Science 313, 11 August 2006. And see the Web Science Research Initiative, <http://www.webscience.org/>