The World Wide Web

Abstract

The World Wide Web (WWW) is a system for creating, organizing, and linking documents so that they may be easily browsed. The Web has transformed the ways in which we communicate, learn, and socialize, and it has changed the ways in which we think about information, information seeking, and interacting with information systems. It is, moreover, one of the principal factors underlying globalization, in the process creating a vast array of connections involving individuals, groups, institutions, and providing a platform that has redefined workflow in many organizations through computer-to-computer data interchanges as well as the creation of collaborative communities. The Web has succeeded because: (1) many relevant conditions were "right" and (2) it has relied from the outset on a simple, derivative architecture, consisting of the Hypertext Markup Language (HTML), the Hypertext Transfer Protocol (HTTP), and the Uniform Resource Locator (URL). The Web's stewards have managed the continuing development of the underlying technologies to ensure its openness and in ways that have lead to gradual changes and subtle transformations, rather than radical shifts. At the same time, the Web's stewards, most notably the World Wide Web Consortium

(or W3C), have fostered important innovations, such as the development of the Extensible Markup Language (XML) and the Cascading Style Sheets (CSS) specification, the proposal to develop the "Semantic Web," and the evolution of HTML leading to the development of HTML5. In the process, the World Wide Web has had profound effects on libraries, librarians, and library users, changing the way in which librarians relate to vendors, clients, bibliographic utilities, and other libraries, and giving rise to new, often highly creative approaches to serving readers.

Introduction

The World Wide Web is a system for creating, organizing, and linking documents so that they may be easily browsed. Created by Tim Berners-Lee, the World Wide Web is also one of the most remarkable developments of the last 25 years, and it is virtually certain that it will continue to be a pervasive influence on both information producers and information consumers for the foreseeable future.

The Web has transformed the ways in which we communicate, learn, and socialize. Perhaps even more to the point, the World Wide Web has changed the ways in which we think about information, information seeking, and interacting with information systems.

The World Wide Web may be an incomplete and imperfect manifestation of the ideas about hypertext that Ted Nelson set forth in the mid–1960s, but it has changed the ways in which we think about the world, and it has changed forever how ideas, information, and knowledge are shared.[1] According to Thomas Friedman, in his *The World Is Flat: A Brief History of the Twenty-First Century*, the World Wide Web is one of the principal factors underlying globalization, in the process creating a vast array of connections involving individuals, groups, institutions, and providing a platform that has redefined workflow in many organizations through computer-to-computer data interchanges as well as the creation of collaborative communities. As Friedman has also noted, it is an environment that seems almost ideally suited to the needs of information seekers with what he calls a high "curiosity quotient" — Friedman believes that when curiosity is combined with passion in the exploration of a subject of interest, an individual of average intellectual endowment may be able to acquire knowledge comparable that of a highly intelligent person, because of the vast amount of information resources available through the Internet — and it clearly appeals to writers in search of new and more expressive modes of communication.[2] For them, documents are, as Lisa Gitelman has observed, "instruments used in the kinds of knowing that are all wrapped up with showing, and showing wrapped up with knowing," and the Web affords both technologies and cultural milieus of greater power and scope than traditional, analog forms of information exchange.[3] The product, from the perspectives articulated by Timothy Morton, are often "hyperobjects," by which Morton means objects so massively distributed in time and space that they transcend "spatiotemporal specificity."[4]

Less flattering are the views of critics like Louis Menand, who has characterized the Web as an imaginary space — he calls it a "spatial imaginary" — in which visual change is often experienced (and confused with) as a physical change. Menand argues that the use of "real estate vocabulary," in the form of terms such as "address," "site," and "domain," reinforces this dislocating illusion and changes how we think about information resources and use them in ways that obscure underlying realities.[5]

The emergence of Web 2.0, a new layer of activities shaped by participatory architectures based on cooperation rather than control, lightweight programming models, enriched user experiences, and a fuller realization of the Internet as a platform for computing, changed yet again the way in which we think about and use the Web and its contents. In its first phases, Web 2.0 allowed users to comment on published articles, participate in social networks, tag items such as digital photographs, images, and documents, and share Web bookmarks.[6] In the second phase of Web 2.0, software as a service came to maturity, through the integration of application programming interfaces (APIs), Ajax programming using JavaScript and the Document Object Model, and cloud-based storage, in the form of Web-based applications such as Google Docs, YouTube, and Microsoft Office 365.

More recently, HTML5, a synthesis of HTML and XHTML that integrates the Document Object Model (DOM) into the markup language and offers new opportunities for the incorporation of audio and video media, has further enhanced what may be conveyed through a Web page. It includes processing models designed to encourage more interoperable implementations, extends and improves the markup available for documents, and introduces markup and APIs for complex web applications.[7]

Looking to the near future, it seems likely that the ideas associated with the Semantic Web will soon begin to have more obvious effects, transforming the Web from a vast file system to an equally vast database capable of supporting various processes, including discovery and search, with perhaps unparalleled precision.

The Semantic Web has long been a controversial subject, marked by high aspirations and serious doubts. The debate began the day Berners-Lee, James Hendler, and Ora Lassila unveiled their proposal, focusing mainly on questions about its feasibility.[8] There were almost no doubts expressed about the desirability of this vision for the future of the Web, but many experts were not optimistic about the success of the initiative, owing to its complexity, its stringent requirements, and, as Clay Shirky observed, because most of the data we use is not amenable to the syllogistic recombination that the Semantic Web presumes.[9] Others have noted, similarly, that the proposal "disregards the fundamental fuzziness and variability of human communication," and that the "rigid formality" which characterizes the Semantic Web cannot be enforced or ensured, resulting in an "interoperative polyglot" akin to, for example, RSS (Rich Site Summary or Really Simple Syndication).[10]

However, the vision of a near future in which semantically oriented technologies that systematically describe the content of the Web are coupled with artificial intelligence to create a new layer within the Web infrastructure has persisted.[11] More important, essential parts of this new infrastructure have been built, and the transformation, in which metadata in standardized forms pervades the network and affords the basis for a wide array of services, ranging from more precise retrieval of information to the automatic generation of documents, is well under way.

But the doubts persist. In 2010, the Pew Internet Research Center surveyed a group of experts on Web technologies in an effort to understand the prospects of the Semantic Web. Some of the experts, 41 percent of the survey's 895 respondents, thought that the concepts on which the Semantic Web is founded would be realized by

2020, while 47% of those surveyed expressed skepticism about its feasibility, agreeing with the notion that "[b]y 2020, the semantic web envisioned by Tim Berners-Lee will not be as fully effective as its creators hoped and average users will not have noticed much of a difference."[12]

Around the same time, Berners-Lee returned to the debate, arguing then and later that efforts to markup and link data sets, but especially data sets derived from scientific research, would lead inexorably to a new version of the Web organized on the basis of semantic information interpreted by both humans and computers.[13]

Another aspect of Berners-Lee's vision for Web is the annotation. It is a feature that Berners-Lee had originally intended to incorporate, but in the effort to retain control over the technology and guarantee its openness in the mid–1990s, it was set aside. But when he wrote *Weaving the Web* in the late 1990s, Berners-Lee noted that "[w]e need the ability to store on one server an annotation about a Web page on another [server]."[14]

In recent years, the idea of creating a standard for annotations and integrating it into the Web infrastructure has been taken up by the W3C and others, in the form of Open Annotation Data Model. The primary aim of the Open Annotation Data Model is to create a "single, consistent model," within "an interoperable framework for creating associations between related resources, annotations, using a methodology that conforms to the [a]rchitecture of the World Wide Web," and in so doing provide "a standard description mechanism for sharing [a]nnotations between systems, with the mechanism facilitating sharing or migration of annotations between devices."[15]

There is considerable interest among developers in the annotation as a mechanism for information enhancement and exchange, manifest in a variety of projects active at this writing. But it is not clear if there is a widespread interest among users. Other projects of similar purpose, such as the W3C's Annotea Project, have met with limited success.[16] Perhaps even more to the point, there is no sufficiently simple mechanism for support of the Open Annotation Data Model that is available for deployment; so, the model and its potential remain untested at this writing.

How Big is the Web?

Since the World Wide Web does not operate under any central authority, the question of the Web's size is difficult to answer precisely. Domain Name System (DNS) services, the Internet services that translate domain names into IP (or Internet Protocol) addresses, list the domain names that exist, but not every domain contains a Web site, many domains contain more than one Web site, and DNS registrars are not obliged to report how many domains their databases contain. So, most of what is known about the size of the World Wide Web is based on survey results, which differ substantially, and/or the number of pages indexed by Web search engines, such as Google and Yahoo. However, within the limits of what can be measured, there is evidence that the Web not only continues to grow at a rapid rate, but also that it is taking on increasing complexity and substance in the content it transports.

The Internet has been growing exponentially since at least 1993. Current estimates indicate that slightly more than a billion live Websites have been created since 1991. Today, there are at least 760 million Websites, with approximately 103 million new sites added in 2013 alone. How many of the current Web sites are active? The number depends on how "active" is defined. One source indicates that 67 percent of the current sites are active, while another suggests that about three-quarters of the active sites are "parked," or dormant.[17]

According to the findings of surveys last updated in 2014, there are 2.7 billion Web pages that have been indexed, and approximately 2.6 billion Web users. Current estimates indicate that slightly more than a billion live Websites have been created since 1991, that there are roughly 672 million Websites in existence, with about three-quarters of them dormant (or "parked").[18] (This finding is largely consistent with the results of a series of studies conducted by OCLC between 1997 and 2003, in which investigators discovered that perhaps as many as half of the Web sites on the Web had effectively been abandoned by their respective owners.[19])

Active sites present a total of 14.3 trillion pages, 48 billion of them indexed by Google, and consist of a total of 672 exabytes of accessible data. More than 2 trillion searches were conducted through the Google search engine in 2013, by an estimated 1.45 billion users.[20] (Another source indicates that the number of Web users is much larger, in excess of 2.5 million people.) [21]

In principle, the World Wide Web remains an open and egalitarian enterprise. Anyone can launch a Website. But the vast majority of the top 100 Websites, the most visited sites, are run by corporations, the most important (and almost only) exception being Wikipedia.[22]

The so-called "deep Web," the part of the Web that is not indexed by search engines, which is generally restricted in its access, and which may include non-HTML, unlinked, dynamic and/or scripted content, is thought to be much larger than the "surface Web." Recent estimates suggest that the "deep Web" may make up as much as 90 percent of the Web, but the size and continuing growth of the Web make it impossible to determine precisely how large it is or how much of it is part of the "surface Web" or the "deep Web." However, according to Bright Planet, an organization that specializes in content extraction, the “deep” or “invisible” Web contains nearly 550 billion individual documents, as compared to the one billion documents contained within the "surface Web.”[23]

Almost 60 percent of the Web’s users are employing mobile devices, i.e., smartphones, tablets. The popularity of various Web browsers is a matter of constant contention, but it appears that Internet Explorer remains the most popular Web browser at this writing, with a combined marketshare of 49 percent. Google Chrome has a marketshare of 18 percent, and Mozilla’s Firefox browser and Apple’s Safari each have shares of approximately 11 percent.[24,25]

According to the HTTP Archive, the average page is currently 1890 KB in size, compared to 828 KB in May 2012, with an annual growth rate of about 50 percent. [26,27] Images typically make up about 55–60 percent of the overall payload. The use of CSS and JavaScript is increasing, whereas the number of requested pages including Adobe Flash is slightly below 30 percent and declining markedly.[28]

It is not entirely clear why Web pages are growing in size and complexity. It seems reasonable to assume, however, that there are several highly influential factors. The first is increasing penetration of broadband Internet services, which allow the creators of Web content to create and transfer files of substantial size with ease. Moreover, the widespread availability of broadband services, particularly in the developed world, means that many files with richer content can reach large audiences quickly and without difficulty. A second factor is competition. Competition for the attention of Web users, but particularly for consumers using the Web as a marketplace, is intense, and there is evidence that richness of content is often a necessary condition for success. Finally, there is the availability and use of authoring and site management tools that make it relatively easy to create HTML and/or XML documents of considerable complexity, and to be able to do so without detailed knowledge of the relevant coding languages.

Another, major factor adding to the complexity of the Web is the use of streaming media, which has increased by more than 100% each year since 2000. Audio or video files that are transmitted continuously from a server and can begin playing as they are being downloaded to a computer. This process is now enhanced by the Dynamic Adaptive Streaming over HTTP (DASH) standard, which has made video-on-demand a 'standard' Internet application similar in its impact to email and Web browsing.[29]

While videos account for only a small percentage of the responses from Web servers, the most popular video services, Netflix, YouTube, Amazon Video, Hulu, and iTunes, generate between 50–55% of the bytes transferred via the Web, with Netflix producing slightly more than 31% of the traffic in 2014, and YouTube accounting for about 12% of the overall traffic.[30]

Interest in downloading or streaming videos via the Web is likely to continue to grow in accord with broadband's market penetration and the increases in the speed of data transfer that are now an integral part of the market for Internet services, and with improvements in the DASH protocol and end-to-end congestion controls. So, it seems reasonable to assume that the demand for videos will continue to grow, perhaps even sharply, and that the bandwidth required to fulfill those requests will continue to be a major factor in the use of the Internet, the management of Internet backbone resources, and the politics of the Internet.

Is the Web getting faster? Google claims that it is significantly faster today than it was only a couple of years ago, owing mainly to improvements in the core infrastructure, much faster mobile networks, and improvements in Web browser design.[31] However, there are other sources that dispute such findings, but resolution of the question is not possible at this time, because there is no consensus about the methods of measurement or analysis.[32]

A Brief History of the World Wide Web

The development of the World Wide Web may be divided into three phases. In the first, experimental phase, Berners-Lee developed a collaborative hypertext environment based on an adaptation of the TCP/IP protocol and aspects of SGML. As noted in his original proposal, "the working structure of the organization is a multiply connected "web" whose interconnections evolve with time."[33] Berners-Lee then made the source code available over the Internet, facilitating experiments with and improvements of the basic technologies, a process that continues to this day largely through the World Wide Web Consortium, or W3C.

According to the *Encyclopedia of Computer Science*:

> *Hypertext is both the concept of interrelating information elements (linking pieces of information) and the name used to describe a collection or web of interrelated or linked nodes.*[34]

Vannevar Bush is generally given credit for developing the idea of hypertext, as first articulated in an article entitled "As We May Think," which was published in the July 1945 issue of *The Atlantic Monthly*.[35] However, recent research indicates that the ideas set forth by Bush were not new, and that the idea of a machine that would connect an individual to diverse sources of information and effectively help to synthesize such information had also been expressed previously, perhaps most notably by H.G. Wells in his pre-WWII lectures on a "World Brain," by Paul Otlet, a Belgian bibliographer who proposed in 1934 a plan for a global network of "electric telescopes," which would allow anyone in the world to access to libraries of books, articles, photographs, audio recordings, and films, and by Emmanuel Goldberg, who patented a device called the "Statistical Machine" in 1927 that allowed a user to search and retrieve large volumes of data stored on microfilm by using a so-called search card, and who later proposed a technique that would allow a user to enter a query via telephone.

Bush called his machine the "Memex." Wells talked about an interactive encyclopedia controlled by subject experts that would function as a form of collective intelligence. He wrote:

> "[*W*]*hat I am saying … is this, that without a World Encyclopaedia to hold men's minds together in something like a common interpretation of reality, there is no hope whatever of anything but an accidental and transitory alleviation of any of our world troubles.*"[36]

Otlet called his knowledge network the Mundaneum. (Previously, Otlet and Henri La Fontaine had launched a project called the Universal Bibliography, or *Répertoire Bibliographique Universel*, a plan to catalog of all the world's published information. The project eventually resulted in the creation of more than 15 million entries, stored on index cards and classified under a system called the Universal

Decimal Classification, an adapted version of the Dewey Decimal System.)[37]

In considering the history of hypertext, scholars tend to focus on the various mechanisms that were proposed as ways of bringing greater order to the formal communications of scholars and affording greater precision in the use of such materials, and in the process may not always pay enough attention to Bush's notion that memory is an associative process, and that we should build information storage and retrieval systems whose organizing structures mimic associative memory as closely as possible. So, in Bush's mind, the linking functions that connected chunks of text and articles to one another, what we now call hypertext links, represented the best available way to imitate this aspect of human memory. (This was not a new concept; in fact, the notion can be traced back to Aristotle, whose thoughts about memory and recollection may be viewed as the first articulation of principles of association and order. Bush's greater service was bringing these ideas into discussions about how to make information systems more effective, and by suggesting that "spatializing" ideas, or the chunks of text, data, and imagery, into more discrete units of presentation could enhance memory and recollection.[38])

Similarly, the idea of a "world brain" as set forth by Wells may sometimes be dismissed too quickly, in part because Wells was vaguer than, say, Otlet or Bush, about how his system might work at a mechanical level, but also because he used an ancient and familiar form, the encyclopedia, as the basis for his proposal. What gets lost are Wells' insights into the sociology of knowledge production; specifically, his recognition of science as a highly collaborative enterprise and his sense that technology, in the form of a distributed network, could enhance and extend the process known as "peer review," and thereby bring greater order and clarity to human knowledge.

There were a number of experiments with hypertext from the mid–1960s through the end of the 1980s, including Ted Nelson's Xanadu Project, whose goal was the creation of a computer file system based on hypertext concepts, and Douglas Engelbart's NLS/Augment, the first distributed, shared-screen, collaborative hypertext system in 1968. Other prototype and commercial hypertext systems appeared in the 1970s and 1980s, including Document Examiner, gIBIS, Guide, Hypergate, HyperTIES, Intermedia, MacWeb (by LIRMM), Max, Neptune, Note-Cards, PHIDIAS, StorySpace, Writing Environment, and ZOG/KMS. (In June 2014, fifty-four years after first announcing the Xanadu Project, Nelson finally released a prototype of his system, which may be viewed at http://xanadu.com/xanademos/MoeJusteOrigins.html. According to information on Nelson's Website, he believes today that the Xanadu document format could supplant PDFs, but is unlikely to displace the basic architecture of the World Wide Web.)[39]

One hypertext system, Apple's HyperCard, which was based on the idea of virtual "cards" in a stack akin to a Rolodex and a scripting language called HyperTalk, came into widespread use as part of the package of applications developed for and included with Apple's Macintosh computer. But HyperCard was never fully adapted for use in networked environments or recompiled to run under OS X, and Apple ceased work on its development during the transition to the OS X operating system, eventually dropping it altogether in 2004. It has been argued, moreover, that HyperCard was not a genuine hypertext system, because it lacked many of the navigation, annotation, and structural features that characterize "true" hypertext systems. But, true hypertext system or not, by the time the powers that be at Apple had lost interest in HyperCard, another initiative, developed by Tim Berners-Lee and eventually known as the "World Wide Web," had taken hold, changing forever the Internet and how we think about computerized text.

In 1989, Tim Berners-Lee was working as a software engineer at CERN, the European Organization for Nuclear Research. (The name CERN is derived from the acronym for the French "Conseil Européen pour la Recherche Nucléaire", or European Council for Nuclear Research, a provisional body founded in 1952 with the mandate of establishing a world-class fundamental physics research organization in Europe.) Scientists working at CERN used all sorts of operating systems and software on their computers, and as a result, one scientist often could not find or access another scientist's research. But then Berners-Lee noticed:

> "*All these systems looked different but in fact you're reading stuff on a screen and sometimes clicking on bits. So you could imagine a thin layer which would map all these existing systems into one virtual system. Wouldn't that be cool*?"[40]

Berners-Lee's supervisor at CERN thought the idea was vague but worth exploring, so he commissioned Berners-Lee to create a system for collaborative authoring and document sharing that could run over the local area network at CERN and entail a number of different types of computers and operating systems.

To make such a system operational, Berners-Lee need to create a mechanism for the transport of textual data and a document format that could be interpreted by all of the operating systems that would be connected across the CERN network. And, because he was interested incorporating ideas about hypertext into the system, he need to establish a way of creating links within and among documents. According to Berners-Lee:

> *"The idea of the Web was prompted by positive experience of a small "home-brew" personal hypertext system used for keeping track of personal information on a distributed project. The Web was designed so that if it was used independently for two projects, and later relation-*

> *ships were found between the projects, then no major or centralized changes would have to be made, but the information could smoothly re-shape to represent the new state of knowledge."*[41]

Owing to the requirements of his assignment, the limited resources available for its support, and inspired by the success of the "home-brew" hypertext system, Berners-Lee elected to focus his efforts on the use of existing resources available to him under open licenses and/or technical standards. He realized almost from the beginning that the transport layer of his system could be built on top of the Internet protocols — the so-called "TCP/IP" suite of rules for conveying data over the Internet — by adding a compatible layer, which he dubbed the Hypertext Transfer Protocol, or *http*.

Berners-Lee designed the Hypertext Transfer Protocol to function as a request-response protocol under a client-server computing model, and he wrote the programming code for both a client and a server capable of carrying out the specified actions. In accord with this model, a client application, typically what we now know as a Web browser, submits an request message formed on the basis of the Hypertext Transfer Protocol to a server that has been outfitted to recognize and respond to requests formatted in this way.

The key element of the request is the Uniform Resource Locator, or URL. The Uniform Resource Locator is a specific character string that constitutes a reference to a resource. The now-familiar components of a URL are:

- the scheme, which defines how the resource will be obtained;
- the domain name or numeric IP address of the destination location for the URL;
- the port number; and
- the path specifying the location of the resource requested on the server.

The URL may also include:

- a query string containing data to be passed to software running on the server; and
- a fragment identifier, which, if present, specifies a part or a position within the overall resource or document.

The server, which provides resources such as HTML files and other content, or performs other functions on behalf of the client, returns a response message to the client, using another URL to route the response to the client. The response contains completion status information about the request and may also contain requested content in its message body.

In addressing the question of how the documents mounted on the server and fetched by the client would be formatted, Berners-Lee confronted a thornier problem, in part because the process of standardizing the formats for electronic documents was nascent. However, Berners-Lee discovered the Standard Generalized Markup Language (SGML) in the early going and recognized that under the SGML grammar a document is separated into three parts: an SGML declaration, a prologue, and an instance, with the prologue constituting a document type definition, or DTD. He noted further that the SGML declaration determines the lexicon of the grammar, specifies the document character set, and establishes the code positions associated with those characters, binding the abstract syntax of SGML to a concrete syntax expressed through the DTD. The concrete syntax formulated by Berners-Lee on the basis of SGML and the ISO standard defining the 7-bit coded character set for information interchang became known as the Hypertext Markup Language, or HTML.[42,43]

So, the Hypertext Markup Language (HTML) began as a subset, also referred to as a "document type definition," of the Standard Generalized Markup Language (SGML).[44] As noted previously, Berners-Lee had been commissioned to create a system under which documents could be shared, edited, and annotated across a network, and the project required a way of formatting documents using a code base — ASCII (the American Standard Code for Information Interchange), as it happens — that was shared by a variety of computers and operating systems. Early on, Berners-Lee became aware not only of the

Standard Generalized Markup Language, he also realized that SGML allowed for the creation of formatting subsets and related expressions rendered in ASCII.

The DTD Berners-Lee created was a simple expression, conceptually and syntactically. And that simplicity plays no small part in the success of the World Wide Web, because it is clear in retrospect that a more complicated scheme would have been much more difficult to implement or standardize. However, that same simplicity also became a curse of sorts, as it became clear that the wide variety of document types that authors sought to make available via the Web could not be supported adequately by a single DTD.

The system introduced by Berners-Lee provided a text-only interface. Acceptance of this system was slow. The text-only browser was not easy to use and the resources to which it had access were limited. However, that state of affairs changed in 1993, with the creation of a graphical Web browser for UNIX, known as Mosaic, by Marc Andreesen, an undergraduate student working at the National Center for Supercomputing Applications (NCSA), and the distribution via the Internet of versions of Mosaic for Windows and the Mac OS in 1994. Mosaic's introduction was the culmination of the first stage of Web development, and its effect was transformative. From the perspective of end users, the Web became a graphical medium, with Mosaic as the lens, and while it would take another ten years before Web content began to achieve real richness and sophistication in visual terms, Mosaic was a clear indication of what was possible.

(In the early days of the World Wide Web, as programmers worked to develop browsers and other tools, questions about how to format HTML documents for presentation arose, and those questions became particularly acute with the development of graphical Web browsers such as Mosaic. Under SGML, formatting for presentation had been treated as a separate issue, with the focus placed instead on structural tagging. As graphical Web browsers came into more general use, one of the limits of this approach became obvious -- most of the formatting of Web documents was being rendered by the browsers themselves and determined in large measure by the default settings of the specific browser in use. While permitting the browser to format the Web page had advantages, it also placed sharp limits on the extent to which the author of a Web page could control how that page was presented to a user. In the short term, the solution was to incorporate formatting attributes and tags into HTML, which in many ways adulterated the original concept. The longer term solution came in the form of a compatible but distinct language for formatting, which came to be known as the "Cascading Style Sheets," or CSS, specification.[45])

In the second phase, the infrastructure that supports the Web through the present day was established. Jim Clark and Andreesen founded Netscape and released its Web browser in late 1994, ending the experimental phase in the Web's development and initiating its commercialization. Microsoft, after ignoring the development of the Internet for a number of years, responded with the development of a Web browser called Internet Explorer and the subsequent release of other, Web-related technologies, including the Internet Information Server. The Apache Project, dedicated to building secure Web server software in an open source environment, was established in 1995. In 1999, as the Web turned from this "buildout" phase to a period of rapid expansion predicated on increasingly stable server technologies and more and more sophisticated approaches to document rendering, IBM embraced open source software and played a key role in establishing the Apache Foundation, which has become one of the leading organizations in the ongoing development of the software that runs the Web. In 2001, the first podcast was presented (in the form of a Grateful Dead recording), Wikipedia was founded, and Pope John Paul II sent the first papal e-mail from a laptop in his office at the Vatican. In 2003, Apple's iTunes music download service was launched. In 2004, Tim Berners-Lee was knighted, and Google became a public company.[46]

In this "buildout" phase, one of the most critical developments was the introduction of the Cascading Style Sheets (CSS) specification in 1994. The purpose of CSS was two-fold: first, it was intended to separate structural markup and formatting in order to simplify both coding and interpretation; and, second, it was designed to enlarge and enrich opportunities for formatting HTML documents, ranging from the ability to define families of fonts to absolute positioning and the use of so-called “floats” to wrap text around images to the cascade itself, which was defined as “the process of combining several style sheets and resolving conflicts between them.”[47, 48]

How important has CSS been to the development of the Web? Håkon Wium Lie, who wrote and published the first CSS specification in 1994, has asserted that CSS “saved HTML,” because it gave “authors a way to express their designs without adding new HTML tags.” Perhaps more to the point, CSS has afforded authors an effective mechanism separate and apart from structural markup for creating visually richer and more interesting Web pages.[49]

In the third phase, which began in 2004–2005, the Web entered into a more highly interactive phase, characterized by network-resident applications, participatory architectures, the increasing use of XML-based technologies, and metadata interchanges that presaged the Semantic Web.

Under HTML5, yet another phase in the development was initiated. HTML5 provided a more coherent framework for creation of Web pages and the development of Web applications. For example, whereas HTML 4 was imprecise in defining the structure of a document, HTML5 provides an "outline algorithm," under which all content residing with the <body> of a document is also part of a section, and sections are defined explicitly within the <body>, <section>, <article>, <aside>, <footer>, <header>, and <nav> tags.[50] In a related vein, under HTML5 headings, e.g., <h1>, <h2>, etc., are defined and ranked within sections of the <body> of the document. Relative ranking of the headings matters only within a section, with the structure of the sections determining the outline, and not the heading rank of the sections.

HTML5 supports video and audio tracks without plugins, provides programmatic access to a resolution-dependent bitmap canvas that is useful for rendering graphs, graphics, or other visual images, native support for scalable vector graphics (SVG) and math (MathML), and features supporting the development of and access to rich applications. (The HTML5 <canvas> element is used to draw graphics, on the fly, via scripting. The <canvas> element is a container for graphics, with the script, usually a JavaScript, which actually draws the graphics. A canvas is a rectangular area on an HTML page, with incorporated methods for drawing paths, boxes, circles, text, and adding images.)

HTML5 has been in use throughout its development. According to a 2014 survey, 42% of 10,000 developers polled are using the combination of HTML, CSS, and JavaScript for all or part of their mobile applications, and Gartner Research has identified HTML5 as one of the top 10 mobile technologies for 2015–2016, as "an essential technology for organizations delivering applications across multiple platforms."[51, 52]

With the publication of the fifth version of HTML5 in late 2014, the World Wide Web Consortium announced that HTML5 would serve as the core of the Open Web Platform, through which the W3C intends to lower the cost of developing cross-platform applications by focusing on the following issues:

- security and privacy;
- Web design and development;
- device interaction;
- media and real-time communications;
- performance and tuning;
- usability and accessibility; and
- related services, including the social Web, payments, annotations, Web of data.[53, 54]

Why Has the Web Been Successful?

The Web has succeeded not only because many relevant conditions were "right," but also because it has relied from the outset on a simple, derivative architecture, consisting of the Hypertext Markup Language (HTML), the Hypertext Transfer Protocol (HTTP), and the Uniform Resource Locator (URL). A URL is a URI that also specifies the location of an identified resource and the protocol for retrieving it. In popular usage and in many technical documents it is often confused as a synonym for uniform resource identifier.[55] In the beginning, as Berners-Lee has noted, the World Wide Web represented a "basically trivial" expression of ideas, mainly about hypertext systems, that had been in circulation, in some instances for many years. [56]

Moreover, since the advent of the graphical browser in 1995, the technologies of the World Wide Web have been easy for end-users to manipulate, and that aspect coupled with the increasing rich blend of text, graphics, and links, has created a huge global audience. Of equal importance is the fact that the Web's stewards, including Berners-Lee in his role as director of the World Wide Web Consortium (W3C), have managed the continuing development of the underlying technologies in ways that have led to gradual changes and subtle transformations, rather than radical shifts. (Berners-Lee established the World Wide Web Consortium (W3C) in 1994 as a means of ensuring that Web-related standards would remain open. It was founded at the Massachusetts Institute of Technology Laboratory for Computer Science (MIT/LCS), with support from the European

Commission and the Defense Advanced Research Projects Agency (DARPA), and quickly expanded its membership to include many other organizations, including most of the leading companies in the computer industry. By 2014, W3C membership had increased to 385 organizations.) For example, the path from the first version of the HTML to HTML 4.01 and XHTML and HTML5 is one marked by a commitment to interoperability and spirit of accommodation, whereby "[e]ach version of HTML has attempted to reflect greater consensus among industry players so that the investment made by content providers will not be wasted and that their documents will not become unreadable in a short period of time."[57] At the same time, however, the interest in innovation that has driven the development of the Web has been manifest in other aspects of this stewardship, most notably in the development of the XML and proposal to develop what Berners-Lee and others refer to as the "Semantic Web."

In the longer term, the World Wide Web has succeeded because it is networked, but separate from the Internet. Berners-Lee has argued that this separation is of fundamental importance, because it allows the two layers of technology, the Internet and the Web, to work together while advancing independently.[58] Moreover, the value of what the Web provides has grown in proportion to data, services, and users connected through it. Positive network externalities, the so-called network effect, have been reinforced by the extensibility of the Web's underlying technologies and the availability of an increasing large and diverse set of resources, often available to both content creators and users at little or no cost.

Of at least equal importance, the World Wide Web has succeeded and will endure because of a commitment to open technical standards that was made at the outset by Berners-Lee, and which he and others have sustained over the years, through the W3C, the Apache Foundation, and a host of other efforts. The commitment to open standards has thwarted efforts to exert private control over the Web and its technologies, and it has been a key factor in the Web's remarkable growth, allowing other innovators and entrepreneurs to leverage the Web's technologies in often creative ways. In the beginning, Berners-Lee opted for technologies based on open standards, because his sponsor required interoperability, and because he could afford no other choice. In so doing, he set in motion processes that have radically altered the course of the digital document's evolution.

The World Wide Web Consortium endorses the Modern Paradigm for Standards, which is based on "[r]espectful cooperation between standards organizations, whereby each respects the autonomy, integrity, processes, and intellectual property rules of the others." Under the Modern Paradigm, standards are developed on the basis of five principles:

- Due process, under which decisions are made with equity and fairness among participants, where no one party dominates or guides standards development, and standards-making processes are transparent and opportunities exist to appeal decisions, as well as review and update standards;
- Broad consensus, whereby processes allow for all views to be considered and addressed in order to facilitate agreement;
- Transparency, based on easily accessible records of decisions and the materials used in reaching those decisions and public comment periods provided before final standards approval and adoption;
- Balance, ensuring that standards activities are not exclusively dominated by any particular person, company or interest group; and
- Openness, providing relevant information to all participants and interested parties equally.[59]

Impact of the World Wide Web on Libraries

The impact of the Web on libraries has been profound, changing the way libraries relate to vendors, clients, bibliographic utilities, and other libraries. What is more important, the World Wide Web has liberated librarians and engendered new and often highly creative approaches to serving readers. The Web has had similar effects on publishers, ranging from scholarly presses to the U.S. Government Printing Office.[60]

An example of how influential the Web has been may be found in the case of electronic journals. In the early 1990s, the publisher Elsevier ran The University Licensing Program, or TULIP, an experimental project aimed at establishing an infrastructure for the distribution and delivery of e-journals.[61] What is significant about TULIP in retrospect is that engineers at Elsevier assumed that the most effective and efficient means of delivering content would be to create a system under which digital subscriptions were maintained locally (meaning, in this instance, that copies of subscribed content would be maintained on servers situated with the subscriber's Internet domain) and updated regularly via the Internet. What they did not envision was an Internet that would be fast enough to support the real-time distribution of journal articles or the emergence of a hypertext environment capable of linking databases to e-journals, journal articles to other journal articles via hyperlinked citations, articles to datasets, and so on. By the end of the 2000s, research libraries had reached a "tipping" point, whereby electronic journals, almost all of them delivered to libraries by means of the World Wide Web, outnumbered print subscriptions.[62]

Integrated online library systems have been enhanced in a wide variety of ways, incorporating published and personal reviews of materials, support for bookmarking and personal account management, Web-based readers, citation formatting services, virtual reference services, etc. In recent years, access to library resources has been improved markedly, through the ongoing development of increasingly powerful discovery services and link resolvers. (The development of link resolution services is great example of how the commitment to open standards has enhanced the functionality and efficacy of the Web as a medium for information interchange. Such services are based primarily on the OpenURL standard, which was developed by Herbert Van de Sompel and others in the late 1990s and early 2000s and then established as NISO standard in 2004.[63])

Google Scholar is another interesting and important case in point where libraries and the Web are concerned. Google Scholar is designed to access to the scholarly literature through a single portal, supplying a means to find scholarly papers, abstracts, and citations, locate papers via libraries and/or Web sites, and learn about key works in "any area of research."[64] [65] In many research library environments, it is also being used as a discovery system. (How effective is Google Scholar? It has been a matter of controversy, with some studies suggesting that Google Scholar's coverage was substandard, but a recent study found that "as regards strict scientific impact, the analysis of GS data provides very similar results to the results obtained from traditional citation-based databases, with the advantage of being able to retrieve a larger and more varied number of citations, since they come from a wider range of document types, different geographical environments, and languages different to English."[66,67])

In addition, owing to the considerable efforts of archivists, curators, and scholars, there is a large and constantly expanding body of primary source materials available via the Web. Well-established projects like the American Memory Project at The Library of Congress continue to grow in terms of the number of collections available and the availability of collateral services.

The American Memory Project draws upon the collections of The Library of Congress and other institutions to provide "a digital record of American history and creativity."[68] In the process, materials to which access had previously been highly limited have become available to the general public, in most instances accompanied by expert commentary. An early contribution to the American Memory Project concerning poet Walt Whitman and his notebooks is a good case in point. Conceived as a "test bed" for digital preservation and making collections available via the World Wide Web, digital facsimiles of four of Whitman's notebooks, including a notebook that contains early drafts of segments of The Leaves of Grass, were rendered and made accessible via Web in a presentation entitled Poet at Work: Recovered Notebooks from the Thomas Biggs Harned Walt Whitman Collection. [69] The significance of Poet at Work may be expressed on many levels, but it may be sufficient to say that in making available a body of material that has previously been available only to credentialed Whitman scholars, The American Memory Project changed forever the relationship between The Library of Congress and library users, and that the World Wide Web served not only as the medium of presentation, but also as a agent of change in a process that has brought The Library of Congress substantially closer to the goal of "serving the public as a resource for education and lifelong learning."

Many newer projects are distinguished not only by their original content but also by their use of newer, interactive technologies; for example, the Georgia Virtual History Project is dedicated to recording the history of the state in digital forms and making those records available via the Web to "multiple audiences," ranging from middle school students and the general public to college students and scholars.[70] Another example is the eHistory Web site, as developed by the University of Georgia's Center for Virtual History. eHistory was founded in 2011 by two historians in the belief that digital technologies afford new forms of research, in which students, scholars, and members of the public may act as collaborators. Like sites that call on “citizen scientists” and to gather and analyze data, eHistory projects involve “citizen historians” in amassing and analyzing historical data. The creators of eHistory that what has been labeled "citizen history," often better reflects the way knowledge is created and consumed in the digital era.[71]

The Web, or, to be more precise, the audience that the World Wide Web brings, has been the impetus for a series of scanning/content preservation projects, ranging from the controversial Google Books project to JSTOR to the Internet Archive. And the World Wide Web is also changing the form of the journal article, as publishers move from the PDF format to composite document formats based in significant part on HTML5. What Elsevier has dubbed the "article of the future” is a good example of widespread efforts to deploy “better ways to create and deliver the formal published record” by taking advantage of the expressive possibilities manifest in the continuing improvement markup and scripting languages.[72]

General Effects of the World Wide Web

Technological Impact

The technological impact of the World Wide Web begins with the fact that Berners-Lee constructed a hypertext system based on open standards and capable of running in networked environments. It ends, for the time being, with the ongoing development of the Semantic Web and the emergence of compound documents within a multimodal "interaction domain." It encompasses the development of an increasingly sophisticated system for structuring and formatting digital documents within an open technical framework and the creation of languages that define, respectively, concepts and relationships within domains of knowledge—the Web Ontology Language (OWL)—and establish a functional basis for statements, in the form of triples, e.g., (Subject, Predicate, Object) or (Subject, Property, Value), linking data in order to describe both concepts and objects—the Resource Description Framework (RDF). [73]

On another level, the World Wide Web has changed expectations in regard to computing and networking, to the point that with the rise of so-called cloud computing, the notion of the Web as a computing environment has taken on genuine meaning, in the form of myriad services, ranging from cloud storage services such Google Drive to productivity software like Microsoft's Office 365.

Social and Cultural Effects

The social and cultural effects of the World Wide Web are many, but perhaps the most pronounced have been observed in publishing, information retrieval, and collaborative work. Simply put, the Web has changed the meaning of the word "publish," providing new and breathtakingly broad connotations and linking traditional forms of publication, such as book, newspapers, and journals, with blogs and wikis.

According to research conducted by the Pew Research Center, 14 percent of the adult population in the United States used the Internet. In 2014, 81 percent of U.S. adults use laptop and desktop computers somewhere in their lives, 87 percent of the adults in the U.S. use the Internet, and 73 percent use social media, a pattern of growth that Pew's researchers attribute largely to the growth and popularity of the Web and the emergence of more interactive formats within the framework of the Web. In the so-called second generation of the Web, users have taken an increasingly active role. Almost all users are now able to "'create new content, share it, link to it, search for it, tag it, and modify it' – Wikipedia, Facebook, Twitter and YouTube being the most significant, and now classic, examples of this Copernican revolution."[74] The social impact of an enterprise as vast as the World Wide Web is difficult to gauge in the fullest sense, but 90 percent of the Internet users polled by Pew believe that the Internet and the Web have been good for them personally, 76 percent regard the Internet and the Web as positive developments for society in general, and two-thirds of them indicated that "online communication has generally made them socially richer.[75,76]

In the realm of information retrieval, the Web has been a test bed that renewed and enlarged interest in information retrieval as a set of complex conceptual problems and procedural issues. At a practical level, the Web has facilitated the development of powerful tools, such as the Google Search Engine, it has placed those tools at the disposal of hundreds of millions of users, and it has created a competitive environment for IR services that virtually guarantees ongoing, vigorous commitments to basic research and development.

In terms of contemporary culture, the Web has altered how people inform themselves. News comes from the Web sites of CNN, The New York Times, the BBC, and tens of thousands of other outlets on the Web. Stock quotes, television program listings, restaurant menus, airline flight information, up-to-date weather information, satellite maps, the current address of a distant relative or a long-lost friend—these are all types of information that hundreds of millions of people use the Web to locate and retrieve.

Even language has been altered. In the English-speaking world, "Google" has become a verb. A neologism arising from the popularity of the eponymous search engine, the American Dialect Society chose it as the "most useful word of 2002," it was officially added to the Oxford English Dictionary on June 15, 2006, and to the 11th edition of the Merriam-Webster Collegiate Dictionary in July 2006.[77] (An interesting side note: Google has actively discouraged the use of the word as a verb, compelling some lexicographers to use a lowercase version of the word in order to avoid legal conflict with Google.)

The Web has also provided an operational as well as social environment consistent with the rapid development of systems for collaborative work, encompassing content management systems, collaborative authoring tools, learning management systems for asynchronous learning, Internet telephony, video conferencing, and resource sharing.

Scientists use the World Wide Web as an environment for providing remote access to and control of scientific instruments, they use the Web as the medium for a grid-based architectures that combine entail parallel distributed computation, distributed data management and archiving, and interactive integrated visualization tools in support of their research.[78]

Today, as "cyberinfrastructures" designed for research and development emerge, the focus of development in systems for computer-supported cooperative work is shifting to knowledge sharing within collaborative frameworks and building shared work environments that incorporate the ontological schema and services of the Semantic

Web. However, to date, the most significant collaborative project to be hosted by the Web is Wikipedia.

Wikipedia has created a remarkable amount of controversy, but it also stands as witness to the equally remarkable social forces that the Web has unleashed. It may not have been Jimmy Wales's aim to stand H.G. Wells notion of the "world brain" on its head, but Wikipedia and its many allied projects have demonstrated the reach of digital volunteerism and have shattered, perhaps forever, the myth that valuable knowledge resides mainly within universities.[79] The Wikipedia model will undoubtedly undergo many modifications in the years ahead, some of them intended to bring collaborative models for building knowledge resources closer to the technocratic control that Wells envisioned, but no matter what happens in the future, the Wikipedia of the early twenty-first century will be remembered as revolutionary in nature and effect, because it has changed the way in which we think about encyclopedias and how they are made.

The Future of the World Wide Web

When contemplating the future of the World Wide Web, it is necessary to consider what we actually know and understand about the system in its many dimensions and manifestations. Not long ago, it was argued that "[d]espite the Web's great success as a technology and the significant amount of computing infrastructure on which it is built, it remains, as an entity, surprisingly unstudied."[80] That state of affairs has changed markedly in recent years, to the extent that many studies have been conducted and published, resulting in a plethora of data. However, what we know or understand about the Web, and particularly in reference to the Web as a social machine, remains limited and ultimately inadequate. In *Weaving the Web*, Berners-Lee "hypothesized that the architectural design of the Web would allow developers, and thus end users, to use computer technology to help provide the management function for social systems as they were realized online."[81] In view of the fact that the success or failure of Web technologies often depends more on social factors than it does on technological issues, it may be argued that the ability to design and deploy successful applications requires a significantly better understanding of the features and functions of the social aspects of the systems.

At a technical level, the future of the World Wide Web is well defined, at least in the near term, by HTML5 and the Semantic Web. Each is an evolutionary extension of the World Wide Web.

The Semantic Web is an initiative and a broad area of work in which the semantics of information and services on the Web are defined in order to increase the precision of the results that are delivered to users in response to their queries. [82] The core idea of the Semantic Web is to create the metadata describing data, which will enable computers to process the meaning of things. Once computers are equipped with semantics, they will be capable of solving complex optimization problems.

The Semantic Web is based on Tim Berners-Lee's vision of the Web as a universal medium for data, information, and knowledge exchange and his belief that the existing Web may be so transformed through schemes based on the XML, focusing specifically on shifting the underlying structure of the Web from a vast file system into a huge, integrated database by marking up data and documents by content. From another perspective, the Semantic Web comprises a set of design principles, collaborative working groups, and a variety of enabling technologies. Some of the key elements of the Semantic Web are expressed as future possibilities that are yet to be implemented or realized. Other elements of the Semantic Web are expressed in formal specifications. These elements include Resource Description Framework (RDF), a variety of data interchange formats (e.g., RDF/XML, N3, Turtle, N-Triples), and notations such as RDF Schema (RDFS) and the Web Ontology Language (OWL), all of which are intended to provide a formal description of concepts, terms, and relationships within a given knowledge domain.

Is the Semantic Web inevitable? The answer is yes, insofar as it is the latest expression of humanity's desire to create better tools, because it is the logical next step in the development of information processing and distributed computing services, because its realization will play a major role in driving the global economy of the twenty-first century, and because it will also play a critical role in the development of the systems that are needed to ensure continuing progress in science, biomedical research, and other matters vital to humanity.[83] When will the Semantic Web arrive? The key building blocks—RDF, OWL, etc.—are here already, but it will take some time, perhaps a long time, to annotate and capture the world's information in the appropriate ways.

One of the factors that may speed or delay the arrival of the Semantic Web is the speed with which intelligent applications, software implementing concepts from artificial intelligence that facilitate machine-to-machine communications, grow in operational sophistication. Realizing Berners-Lee's vision depends on an ability to generate and assimilate metadata on a scale and at a speed beyond collective human capabilities, and that process will depend, in turn, on the skill with which software agents (or "bots") are designed and deployed.[84]

There are issues associated with HTML5, too. Owing to the rapid pace of device change and the attendant platform fragmentation – each browser and device vendor chooses what HTML5 features to implement – the deployment of HTML5 is uneven. Software tooling is regarded as inadequate, and many developers are not yet proficient in its uses.[85] Even bigger problems stem from the differences in purpose that troubled the development of HTML5. Web developers want HTML to become a better application platform, and, even though HTML5 has been approved by W3C, there is continuing concern among developers that the W3C, the organization that made a long and probably ill-advised commitment to XHTML, is neither flexible enough nor future-oriented enough to provide the necessary stewardship.[86]

But the future of the World Wide Web will also be defined by a series of collateral developments and factors. Mobile computing is clearly one of the major factors. A rapidly growing array of smartphones and tablet computers already provides access to a large majority of the Internet's users in the United States, and because the devices themselves are growing more powerful and more sophisticated, this trend is not only expected to continue but to reshape more general patterns of computer (and Web) usage. Gaming systems will continue to play an important role in the future of the Web, as will online video, already a major force on the Web in the form of YouTube and NetFlix. Software as a service will serve as a medium for more and more end-user computing, in many instances supplanting the personal computer, and evolutionary changes in the design of operating systems and applications will produce growing amounts of structured information, thus paving the way for more intelligent computing.

In education, the Semantic Web's greatest impact will be in the ways in which data integration influences how knowledge is aggregated, organized, and presented to students. As noted in a related essay:

> One vision of a well-developed semantic web includes a search feature that would return a multimedia report rather than a list of hits. The report would draw from many sources, including websites, articles from scientific repositories, chapters in textbooks, blog dialogue, speeches posted on YouTube, information stored on cell phones, gaming scenarios played out in virtual realities—anything appropriate that is accessible by the rules of Web 3.0. The report would consist of short sections that coalesce around knowledge areas that emerged naturally from your research, with keywords identified and listed conveniently off to one side as links.
>
> The information in the report would be compared, contrasted, and collated in a basic way, presenting points of agreement and disagreement, and perhaps associating these with political positions or contrasting research. Because the web knows something about you, it also alerts you to local lectures on related topics, books you might want to read, TV programs available through your cable service, blog discussions you might find relevant, and even local groups you can contact that are also focused on this issue. Unlike a standard report, what you receive changes as the available information changes, and you might have wiki-like access to add to or edit it. And be-

> cause you told your agent that this topic is a high priority, your cell phone will beep when a significant development occurs. After all, the semantic web will be highly inclusive, providing a common language for many kinds of media and technologies, including cell phones. The net result, ideally, is that you spend less time searching and sifting and more time absorbing, thinking, and participating.[87]

Education will also be affected, perhaps profoundly, by the rise of "just in time" learning. Unlike traditional educational models, which are essentially supplier-driven systems that work efficiently for instructors, "just-in time" learning is predicated on consumer-driven systems that are designed to work effectively for students, making learners active participants in the educational process. The new model focuses on learning rather than on teaching. More to the point, "just-in-time" models achieve their goals by moving from standardized to customized content, from discrete time and place to anytime and anyplace delivery, and from passive lecture models to interactive and applied learning, all of which can and will be supported by the technologies of the Semantic Web.

For libraries and archives, the changes are likely to be many and great, encompassing many, if not all of the trends noted above. The decline of the importance of the library as a place is inevitable. Competition for the attention and loyalties of digital information consumers will grow only keener, and librarians who wish to survive and flourish will necessarily adopt new approaches to serving clients. However, perhaps the most profound change before librarians and archivists may well be changes in the form of documents, as new technologies and the continuing pressure to improve formal communications in science, technology, and medicine combine to alter both the form and content of the scholarly (or professional) paper. The result will be the compound digital document, basically a framework for integrating text, multimedia, datasets, and hyperlinks, and those documents will present new, serious challenges across the spectrum of bibiothecal functions. (Librarians are certain to face major challenges in the area of client privacy. As the personalization of Web-based services increases, libraries will be challenged by consumer expectations and forced to confront the inevitable conflicts between the demand for more personalized services and the traditional guarantees of privacy that libraries and librarians in many countries have maintained zealously.)

Finally, the dominant role of the United States in the use and development of the World Wide Web will be somewhat diminished, as continuing growth in China, India, and parts of Africa changes the demographics of the Web's user population and broadens the base of developers and providers. Internet governance issues have become more important and more difficult. A new and potentially more balanced order for Internet governance is emerging, but there are myriad challenges to that new order, the most significant of them being government censorship and surveillance. In the midst of these changes, the stability of the Web will depend, as it has for last 20 years, on the World Wide Web Consortium, the commitments that Tim Berners-Lee has fostered with such great success, and the ability of its leadership, in the midst of change and contentiousness, to maintain a unifying sense of purpose.

References

---

[1] Nelson envisioned a "docuverse," in which all data is stored once and never deleted, and where all information is accessible by links from anywhere in the system, supported by non-linear navigation. The World Wide Web realizes important aspects of Nelson's vision, but there are also significant differences, including the fact that there are deletions and information is often stored in more than one place. See *Ted Nelson Discovers Hypertext.* http://www.livinginternet.com/w/wi_nelson.htm.

[2] Friedman, Thomas. The world is flat : a brief history of the twenty-first century (1st Picador ed., Further updated and expanded.). New York: Picador/Farrar, Straus and Giroux. p. 314.

[3] Owens, Trevor. The PDF's Place in a History of Paper Knowledge: An Interview with Lisa Gitelman. The Signal: Digital Preservation. June 16, 2014. http://blogs.loc.gov/digitalpreservation/2014/06/the-pdfs-place-in-a-history-of-paper-knowledge-an-interview-with-lisa-gitelman/.

[4] See http://www.worldwidewebsize.com/.

[5] Menand, Louis. Crooner in Rights Spat. *The New Yorker*, October 20, 2014, p. 84.

[6] O'Reilly, T. What is Web 2.0: Design patterns and business models for the next generation of software. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html (accessed September, 30, 2005).

[7] "HTML5 Differences from HTML4". World Wide Web Consortium. 19 October 2010. Retrieved 2 November 2014. http://www.w3.org/TR/html5-diff/.
[8] Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic web." *Scientific American* 284, no. 5 (2001): 28-37.
[9] Shirky, Clay. The Semantic Web, Syllogism, and Worldview
First published November 7, 2003 on the "Networks, Economics, and Culture" mailing list.
http://www.shirky.com/writings/herecomeseverybody/semantic_syllogism.html
[10] Anderson, Janna and Lee Rainie. Prospects for the Semantic Web: Overview of responses. Pew Internet Research Project, 2010.
http://www.pewinternet.org/2010/05/04/prospects-for-the-semantic-web/.
[11] Markoff, John. Entrepreneurs See a Web Guided by Common Sense. *The New York Times*, November 12, 2006.
http://www.nytimes.com/2006/11/12/business/12web.html?_r=0 (Accessed October 2, 2014.)
[12] Anderson, Janna and Lee Rainie. The Fate of the Semantic Web. Pew Internet Research Project, 2010. http://www.pewinternet.org/2010/05/04/the-fate-of-the-semantic-web/.
[13] Berners-Lee, Tim and O'Hara, Kieron (2013) The read-write Linked Data Web. [in special issue: Discussion Meeting Issue 'Web Science: a New Frontier'] Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences, 371, (1987), 20120513-[5pp]. (doi:10.1098/rsta.2012.0513).
[14] Berners-Lee, Tim and Mark Fischetti. 1999. *Weaving the web: The original design and ultimate destiny of the world wide web by its inventor*. San Francisco: HarperSanFrancisco. p. 163.
[15] World Wide Web Consortium. Open Annotation Data Model Community Draft, 2013. http://www.openannotation.org/spec/core/.
[16] World Wide Web Consortium. Annotea Project. http://www.w3.org/2001/Annotea/.
[17] See http://www.worldwidewebsize.com/.
[18] See http://www.worldwidewebsize.com/.

[19] O'Neill, E.T.; Brian, L.; Rick, B. Trends in the evolution of the public web: 1998–2002 . D-Lib Mag. 2003 , 9 (4) http://www.dlib.org/dlib/april03/lavoie/04lavoie.html.

[20] http://www.factshunt.com/2014/01/total-number-of-websites-size-of.html..

[21] See http://www.internetlivestats.com/total-number-of-websites/#trend.

[22] See http://afrodigit.com/visited-websites-world/.

[23] Boswell, Wendy. How big is the Invisible Web? About Technology, http://websearch.about.com/od/invisibleweb/f/What-Is-The-Size-Of-The-Hidden-Web.htm. Accessed on October 18, 2014.

[24] See http://www.internetlivestats.com/total-number-of-websites/#trend. In this context, "combined marketshare" refers to an aggregation of data for desktop computer and mobile device usage.

[25] Chrome surges, Windows 8.x falls in September. Ars Technica. October 2, 2014. http://arstechnica.com/information-technology/2014/10/chrome-surges-windows–8-x-falls-in-september/.

[26] Trends. HTTP Archive. October 4, 2014. http://httparchive.org/trends.php.

[27] The HTTP Archive – http://httparchive.org/index.php– is a permanent repository of performance information, including the size and construction of pages. Twice a month, the Archive gathers this data from the top 1,000,000 Alexa-ranked websites.

[28] The use of images on the Web constitute a persistent problem. Many images are rendered in the wrong format, many are poorly sized and/or uncompressed, few employ progressive rendering, and many are hosted in multiple locations, thus increasing the risk of additional latency and possible outages.

[29] ISO/IEC 23009–1:2014 Information technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats. ISO, 2014; and T. Stockhammer, "Dynamic adaptive streaming over http –: standards and design principles," in Proceedings of the second annual ACM conference on Multimedia systems, ser. MMSys '11. New York, NY, USA: ACM, 2011, pp. 133–144. [Online]. Available: [http://doi.acm.org/10.1145/1943552.1943572].

[30] Global Internet Phenomena Report. Sandvine, 2014. https://www.sandvine.com/trends/global-internet-phenomena/; and The Broadband Report, Website Optimization, LLC. http://www.websiteoptimization.com/bw/.

[31] Is the Web getting faster? Google Analytics. April 15, 2013. http://analytics.blogspot.ca/2013/04/is-web-getting-faster.html.

[32] Everts, Tammy. The average web page has almost doubled in size since 2010. Radware, June 5, 2013. http://www.webperformancetoday.com/2013/06/05/web-page-growth–2010–2013/.

[33] Berners-Lee, Tom. Information Management: a Proposal. CERN, March 1989-May 1990. http://www.w3.org/History /1989/proposal.html.

[34] "Hypertext." In Encyclopedia of Computer Science. Hoboken: Wiley, 2003. http://search.credoreference.com/content/entry/encyccs/hypertext/0 (accessed June 10, 2014.)

[35] Bush, Vannevar. "As we may think." Bush, Vannevar. "As we may think." The atlantic monthly 176, no. 1 (1945): 101–108. 176, no. 1 (1945): 101–108.

[36] Wells, H.G. The World Brain. London: Methuen, 1938. pp. 34–5.

[37] Wright, Alex, "The Secret Life of Hypertext," *The Atlantic*, May 22, 2014.

[38] Barnet, Belinda. "Hypertext and Association Space, Time and Hypomnesis." *Convergence: The International Journal of Research into New Media Technologies* 6, no. 3 (2000): 76–100.

[39] Barnes, Susan B. "Computer-user Interface." In Encyclopedia of 20th Century Technology. London: Routledge, 2005. http://search.credoreference.com/content/entry/routt/computer_user_interface/0 (accessed June 10, 2014.)

[40] See: http://www.businessinsider.com/how-steve-jobs-helped-create-the-web-2014-6#ixzz345et5Sga.

[41] Berners-Lee, Tim, et al. "The world-wide web." *Communications of the ACM* Aug. 1994: 76+.

[42] See RFC 1866, Hypertext Markup Language - 2.0, 1995. The URL of the online version is http://tools.ietf.org/html/rfc1866.)

[43] see ISO ISO/IEC 646:1991; for additional information about this standard, see http://www.terena.org/activities/multiling/euroml/section04.html

[44] SGML was developed by Charles Goldfarb, as part of a project at IBM, and was adopted as a technical standard by the International Standards Organization in 1986.

[45] The cascade in "Cascading Style Sheets" refers to the fact that under CSS elements of more than one style sheet may be invoked in order to format a document.

[46] Fifteen Years of the Web, BBC, August 5, 2006. http://news.bbc.co.uk/2/hi/technology/5243862.stm.

[47] Lie, Hakon W.. Cascading HTML style sheets – a proposal. 1994. http://www.w3.org/People/howcome/p/cascade.html. For information about more recent work on CSS and its current status, see the CSS home page at the World Wide Web Consortium, http://www.w3.org/Style/CSS/Overview.en.html.

[48] Lie, Håkon W. Cascading Style Sheets. Thesis submitted for the degree of Doctor of Philosophy, Faculty of Mathematics and Natural Sciences, University of Oslo, 2005.
http://people.opera.com/howcome/2006/phd/#h–357.

[49] CSS: It was twenty years ago today — an interview with Hakon Wium Lie. Dev.Opera, October 10, 2014. https://dev.opera.com/articles/css-twenty-years-hakon/.

[50] Mozilla Developer Network. Sections and Outlines of an HTML5 Document. https://developer.mozilla.org/en-US/docs/Web/Guide/HTML/Sections_and_Outlines_of_an_HTML5_document.

[51] Vision Mobile. Developer Economics Q3 2014: State of the Developer Nation. 2014. http://www.visionmobile.com/product/developer-economics-q3-2014/.

[52] Gartner Research. Gartner Identifies Top 10 Mobile Technologies and Capabilities for 2015 and 2016. February 24, 2014. http://www.gartner.com/newsroom/id/2669915.

[53] The Open Web Platform refers collectively to the standards and specifications developed by the World Wide Web Consortium and its partners, especially as those standards and specifications support the development of new applications and services. For additional information, see http://www.w3.org/standards/.

[54] World Wide Web Consortium. Open Web Platform Milestone Achieved with HTML5 Recommendation Next Generation Web Technologies Build on Stable Foundation. October 28, 2014. http://www.w3.org/2014/10/html5-rec.html.en.

[55] RFC1738—Uniform Resource Locators (URL) . http://www.faqs.org/rfcs/rfc1738.html .

[56] Berners-Lee, T. The World Wide Web—past, present and future . J. Dig. Inform. 1996,(1) http://jodi.tamu.edu/Articles/v01/i01/BernersLee/.

[57] Introduction to HTML 4, from HTML 4.01 Specification, W3C Recommendation December 24, 1999.
[58] Berners-Lee, Tim. Long Live the Web. *Scientific American*, December 2010, p. 83–84.
[59] See http://open-stand.org/principles/.
[60] Yambura, B.S.; Alan, B.A.; David, H.G. Understanding the Web: Social, Political, and Economic Dimensions of the Internet; Blackwell Publishing: Ames, IA, 2003, 36–37.
[61] TULIP, the University Licensing Program. http://www.elsevier.com/wps/find/librariansinfo.librarians/tulip.
[62] Prabha, C. Shifting from print to electronic journals in ARL University Libraries . Ser. Rev. 2007 , 33 (1) 4 – 13 . ISSN 0098–7913, doi: 10.1016/j.serrev.2006.12.001 . [Taylor & Francis Online], [Web of Science ®] .
[63] Van de Sompel, Herbert and Oren Beit-Arie. Open Linking in the Scholarly Information Environment: Using the OpenURL Framework. D-Lib Magazine 7 (3) (March 2001). http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html.
[64] 25. "About Google Scholar," http://scholar.google.com.pitt.idm.oclc.org/intl/en/scholar/about.html.
[65] Helms-Park, R.; Radia, P.; Stapleton, P. A preliminary assessment of Google Scholar as a source of EAP students' research materials . Internet High. Educ. 2007, 10 (1) 65 – 76 . [CrossRef]
[66] Martín-Martín, A.; Orduña-Malea, E.; Ayllón, J.M.; Delgado López-Cózar, E. (2014). Does Google Scholar contain all highly cited documents (1950-2013)? Granada: EC3 Working Papers, 19: November 3, 2014.
[67] Mejer, J.; Thomas, C. Google Scholar's coverage of the engineering literature: An empirical study . J. Acad. Libr. 2008 , 34 (3) 196-201; Jacso, P. Amazon, Google Book Search, and Google Scholar . Online 2008 , 32 (2) 53 – 54 ; Jacso, P. Google Scholar: The pros and cons . Online Inform. Rev. 2005 , 29 (2) 208 – 214 .
[68] About the American Memory Project,". http://memory.loc.gov/ammem/about/index.html.
[69] Poet at Work: Recovered Notebooks from the Thomas Biggs Harned Walt Whitman Collection, http://memory.loc.gov/ammem/collections/whitman/index.html.
[70] See http://www.virtualgeorgia.org/.

[71] See http://www.ehistory.org/.
[72] See http://www.articleofthefuture.com/.
[73] Compound Document by Reference Framework 1.0, W3C Candidate Recommendation July 2007. http://www.w3.org/TR/CDR/.
[74] Levene, M. (2010), An introduction to search engines and web navigation . New York: John Wiley & Sons; Gatto, Maristella. Corpus and Discourse : Web As Corpus : Theory and Practice. London, GBR: Bloomsbury Academic, 2014. p. 207.)
[75] Faris, Robert, and Rebekah Heacock. Measuring Internet Activity: A (Selective) Review of Methods and Metrics. *Berkman Center Research Publication* 22 (2013). p. 27; Fox, Susannah, and Lee Rainie. The Web at 25 in the US. *Pew Research Center's Internet & American Life Project* (2014): 5, 25.
[76] Duggan, Maeve, and Aaron Smith. "Social media update 2013." *Pew Internet and American Life Project* (2013).
[77] "Google (verb)," Wikipedia . http://en.wikipedia.org/wiki/Googling .
[78] Ellisman, M. Cyberinfrastructure and the future of collaborative work . Issues Sci. Technol. October 1 2005 , 43 – 50 .
[79] Wells, H.G. World Brain; Meuthuen : London , 1938..
[80] Hendler, J.; Nigel, S.; Wendy, H.; Tim, B.-L.; Daniel, W. Web science: An interdisciplinary approach to understanding the Web . Commun. ACM 2008 , 51 (7) 60 .
[81] Berners-Lee, T. Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web ; Collins : New York ,2000 .
[82] Berners-Lee, T.; James, H.; Ora, L. The Semantic Web . Scientific American Magazine May 2001http://www.sciam.com/article.cfm?id=the-semantic-web.
[83] Michael, D. The Unfinished Revolution: Human-Centered Computers and What They Can Do for Us ; Harper-Collins : New York , 2001.
[84] Internet Bots. http://en.wikipedia.org/wiki/Internet_bot.
[85] Hilwa, A. The Evolving State of HTML5. IDC, 2014.
[86] "HTML5 Specification Finalized, Squabbling over Specs Continues | Ars Technica." 2014. Accessed November 30. http://arstechnica.com/information-technology/2014/10/html5-specification-finalized-squabbling-over-who-writes-the-specs-continues/.
[87] Ohler, J. The Semantic Web in education . EDUCAUSE Quarterly 2008 , 31 (4) 8 – 948