A Never-Ending Project for Humanity Called "the Web"

Fabien Gandon
fabien.gandon@inria.fr
Inria, Université Côte d'Azur, I3S, CNRS
Sophia Antipolis, France

Wendy Hall
wh@ecs.soton.ac.uk
University of Southampton
Southampton, UK

ABSTRACT

In this paper we summarized the main historical steps in making the Web, its foundational principles and its evolution. First we mention some of the influences and streams of thought that interacted to bring the Web about. Then we recall that its birthplace, the CERN, had a need for a global hypertext system and at the same time was the perfect microcosm to provide a cradle for the Web. We stress how this invention required to strike a balance between the integration of and the departure from the existing and emerging paradigms of the day. We then review the pillars of the Web architecture and the features that made the Web so viral compared to competitors. Finally we survey the multiple mutations the Web underwent no sooner it was born, evolving in multiple directions. We conclude on the fact the Web is now an architecture, an artefact, a science object and a research and development object, and of which we haven't seen the full potential yet.

CCS CONCEPTS

• **Information systems** → **World Wide Web**; • **Social and professional topics** → **History of software**.

KEYWORDS

history, world wide web

ACM Reference Format:
Fabien Gandon and Wendy Hall. 2022. A Never-Ending Project for Humanity Called "the Web". In *Proceedings of the ACM Web Conference 2022 (WWW '22), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3485447.3514195



1 INTRODUCTION

In April 2017, the British computer scientist Sir Timothy John Berners-Lee was awarded the Turing award for having invented the World Wide Web, the first Web browser, and the fundamental protocols and algorithms allowing the Web to scale. The same month, the estimate is that the Web has more than 3 billion direct users throughout the world and it is indeed difficult to think of a human activity that has not been impacted by the Web. This paper recalls the main historical steps in making the Web, its foundational principles and its evolution.

2 SPIDERS ON THE SHOULDERS OF GIANTS

In this section we highlight the main influences and streams of thought that interacted to bring the Web about and show that Tim Berners-Lee was "standing on the shoulders of giants" when he proposed his project in "Information management: a proposal" [7]

The search for techniques to organize and the means to efficiently access masses of collected information was an omnipresent motivation in the prehistory of the Web. Looking back in time, scholars envisioned worlds of inter-connected information centuries before the Web existed. The ideas of mapping and indexing associations between ideas, facts, and documents long pre-date the existence of computers, and in many ways, they reflect the sophisticated way in which information is indexed by the human brain. When our only means of communicating information was via the written or printed word on paper, inter-connecting pieces of information on different pages was very difficult other than via manually created indexes. With the advent of machines, people were able to imagine a time when the machine could support vast quantities of similar inter-connections.

The beginning of the twentieth centuries witnessed one of the earliest systematic world knowledge collection and management initiative with Paul Otlet's work on collecting, archiving, indexing and categorizing documentation and his Mundaneum project home of a universal bibliographical system using the universal decimal classification. Otlet foresaw the world-wide networked environment [25].

Among the direct influences for the Web is a seminal paper by Vannevar Bush entitled "As We May Think" [8]. Bush foresaw the explosive increase in the production of scientific information and predicted the need for a machine to help scientists follow developments in their discipline. He posed, as a scientific challenge in itself, the problem of improving access to knowledge and proposed "a sort of mechanised private file and library". It would use mnemonic index codes to point to and rapidly access any part of imported documents. These attachment points would also allow links to be created between any two elements, thereby externalizing the points and links of association. Bush called his system Memex for "memory extender" recording trails which users build as they move through the information, so that their paths of discovery can be saved and recalled later or passed on to others externalizing "the association of thoughts, in accordance with some intricate web [sic] of trails carried by the cells of the brain" [8].

Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.
WWW '22, April 25–29, 2022, Virtual Event, Lyon, France
© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9096-5/22/04…$15.00
https://doi.org/10.1145/3485447.3514195

Bush's design was based on mechanical technology, but the principles behind the design are, in many ways, even more valid today because of developments in information technology. To bring about such a system, it was necessary to wait for the appearance of computers. In an article at the 1965 ACM conference, Ted Nelson proposed "a file structure for complex, changing and indeterminate information," [22] where he referenced Bush's article and coined a neologism to refer to this collection of interconnected texts: hypertext, and even hypermedia. The terms are often used interchangeably, although, strictly speaking, hypertext refers to linking text documents, and hypermedia refers more broadly to linking media of any type. In libraries we follow references from one book to another. The idea behind hypertext is that this process can be automated, and that we can harness the power of computers to make the search and query process easier. Nelson argued that it would become possible to store digitally anything that anyone would write, record, photograph, or film, and to produce a system that could connect any piece of information to any other piece of information. Nelson's vision of a universal hypermedia system, Xanadu, is most fully explored in his book Literary Machines [21]. He defines hypertext as non-sequential writing and views hypertext as a literary medium, but the term encapsulates a wider set of meanings that includes both cross-referencing—linking between documents—and the more general associations between ideas.

The subsequent years witnessed the development of the first hypertext editors using, in particular, another invention from the same decade: the mouse, with the interfaces and interactions that it allows. We owe these inventions in particular to Douglas Engelbart of the Stanford Research Institute (SRI), who himself received the Turing award in 1997 for his pioneering and inspiring vision of man-machine interactions and the key technologies that allowed their development. In particular, his NLS ("oN-Line System") system combined, in the 1960s, hypertext, a text-editing interface, and the mouse. This system would later be renamed Augment, in reference to the search program of Engelbart who, with an eye toward the pioneering work of the period in artificial intelligence, proposed working on augmented intelligence [12].

The development of hypertext systems stayed in the realm of the research lab for most of the 1970s and 1980s. The systems grew in sophistication as the underlying technologies became more powerful. First-generation systems were implemented on mainframes and were largely text-based. Work on the so-called second-generation systems took advantage of the more advanced user interfaces of 1980s workstation technology. These systems supported graphics and animations, as well as fully formatted text documents. They were also able to provide graphical overviews of the hypertext structure and multi-user support [11]. With the advent of the PC in the mid-1980s came the emergence of a new generation of hypertext systems that were much easier for individuals to use—notably Hyperties, OWL's Guide, and Apple's HyperCard. HyperCard was made available for free on every Macintosh computer in 1987, and it did more to popularize hypertext in the late 1980s than any other event. The first ACM Hypertext conference was held in the same year. But such systems could only be used to build relatively small applications.

It is important to understand that, as hypertext systems were emerging, the age of the Internet was also starting. Many research laboratories were already using email as a standard means of communication, and users were anxiously awaiting the ability to share files over the network. Systems such as Gopher and WAIS were being developed to enable files to be downloaded easily, without the need for expert technical knowledge. Although the hypertext concept had then been created, it remained, for years, essentially limited to applications that were locally executed on a computer. The world was hungry to share information over the Internet, and the time was ripe for the emergence of an easy-to-use, Internet-based hypertext system to facilitate these activities. Communication involving packet switching (inter-networking), which was developed between 1972 and 1975 with the work of Louis Pouzin (IRIA and then Inria [23], Vinton Cerf (SRI), and Robert Kahn (DARPA, [9]), culminated in 1978 with the (TCP/IP) standards and the beginnings of the Internet. Network communications applications then multiplied: email (SMTP), mailing lists, file transfers (FTP), remote connections (Telnet), discussion groups, etc. Here as well, we find two Turing award winners from 2004: Vinton Cerf and Robert E. Kahn. The crossing between internet and hypertext research communities led to the development of hypertext systems that handled information on large scale and in distributed environments. Notable examples were the Hyper-G system from the University of Graz [2], the Microcosm system from the University of Southampton [15], and of course, the Web itself [5].

3 WEAVING THE…MESH

Timothy Berners-Lee arrived in 1980 as a consultant for CERN (*Centre Européen de Recherche Nucléaire*), as a young physics graduate and a self-taught computer scientist. Faced with the quantity of information and documentation that he and his colleagues had to deal with, Tim wrote a program (Enquire) to store information and to support links within it. It worked on a multi-user operating system and thus allowed multiple users to access and contribute to the same data [7]. In comparison to the CERNDOC documentation system, which was based on a constrained hierarchical structure, Tim found that the arbitrary and bidirectional links within Enquire were more flexible and scalable.

Tim then returned to the private sector between 1981 and 1983, where he worked on real-time remote procedure calls, and thus in the field of networks and network programming, before returning to CERN in 1984. At that point, he was convinced that there was a need for a global hypertext system at CERN where thousands of people crossed paths from different areas of specialization and worked on different instruments [6]. In March of 1989, to persuade his management, Tim wrote "Information Management: A Proposal" [7] where he described a distributed hypertext system that he would call…"Mesh".

In 1989, CERN was the largest Internet hub in Europe and Tim and his colleagues worked on RPC (Remote Procedure Call) allowing a program running on a given machine to invoke procedures from programs running on other machines [18]. The Unix operating system was very popular and natively integrated functions such as support for Internet protocols and multi-user environments. Tim had the idea of extending the link references to include network

addresses, in order to weave a "mesh" between documents made available on different machines. Tim incorporated the principles of hypertext and integrated TCP and DNS into them in an architecture where clicking a link becomes, conceptually, a remote call for procedures, making the Web less a network of documents and more a network of procedures, as what is called now the REST (REpresentational State Transfer) [14].

CERN also already had a programming approach to document writing (e.g., LaTeX, SGML) [18], and this had a direct influence on the idea of turning to a simplified markup language that will become HTML. Starting in 1990, the Web documentation was itself in HTML and the Web began self-documenting, opening up the possibility for everyone to add to this documentation, and to learn to "weave through weaving." This type of reflexivity also made the Web a space designed for humans and machines to co-evolve and collaborate. It supports co-understanding, co-documentation, and the co-conception of all the objects that are part of it, starting with the Web itself.

Another, relatively poorly known characteristic of the Web at the time of its conception, is that the Web was conceptually open to writing: everything in its architecture allowed not only the consultation (HTTP GET [13]) but also the modification of the Web (HTTP PUT, POST, DELETE [13]). The first client prototype imagined by Tim was a WYSIWYG browser and editor called W3 [5]. But for diverse reasons (security, ease of use, etc.) and because of a lack of editors [5], the most popular browsers and servers did not include this possibility of modifying Web pages. It would only be reintroduced, at a certain level, with the invention of Wikis (WikiWikiWeb) in 1994 by Ward Cunningham.

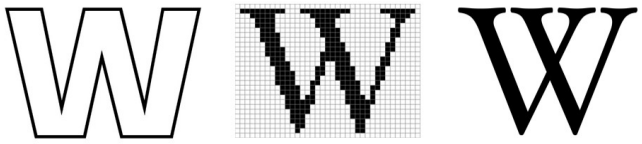
Tim prototyped his first server and browser on a state-of-the-art NeXT Cube workstation. It had a powerful object-based and graphical programming environment that integrated the large-scale networking of computers (TCP/IP and Internet), high level programming languages, progress on human-machine interactions, and notably graphical and multi-window interfaces. With this NeXT and its environment Tim was on the shoulders of the giants we mentioned to start his development. At the end of 1990, the first server and the first browser were tested through an Internet connection. The browser was called World Wide Web, which would become the name of the hypertext that it would give rise to and which would send the name "Mesh" to the dustbin of history.

4 BREAKING THE THREAD

But the Web proposal was not just integrating and hybridizing existing ideas, it was a remarkable exercise in striking a balance between the integration of and the departure from the existing and emerging paradigms of the day. In particular, Tim was convinced that generality and portability were more important than other extensions and he had the idea to "build small, but viral."

The core functionality was to allow random associations between arbitrary objects as well as an incremental and trivial contribution for objects and for links by different contributors, justifying a decentralized approach [4]. The Web always was about moving beyond the silos imposed by the existing applications and searching for independence with respect to the field, task, hardware, operating system [6] and file paths and systems [4]. Tim's proposal liberated hypertext from a central server and the data and links were decentralized on the Internet. As a price to pay, the links became unidirectional and were no longer necessarily maintained. The 404 error was born.

The Web allowed precisely what traditional hypertexts were trying to prevent. However, these breaks were the necessary conditions for the scaling and virality of the Web. Yet this loss of functionality led to new adaptations. For example, the absence of a central index and of bidirectional links would lead to the creation of directories (e.g., Yahoo! in 1994) and search engines (e.g., Altavista, Google) to locate pages and links. The search engines grew to fill the gaps of the Web, the two systems are completely synergistic — one can't exist without the other. Search engines can function because of the way in which the Web encodes hypermedia data, and the Web has overcome some of its inherent disadvantages through the development of search engines. Compared to the state of the art at the time, Tim argued that the network was everything while many other hypertext systems ran on standalone workstations. He showed us that "scruffy works." It didn't have to be perfect, because we aren't perfect, but it had to be good enough. In addition, he argued that the model and architecture had to be decentralized and non-proprietary, that the standards had to be universal, and that it had to be free to use. Ontologically speaking, either everyone would use the *World Wide* Web, or no one would.



5 WEB PILLARS: IDENTIFY, DESCRIBE AND COMMUNICATE

The Web architecture was thus built on three pillars:

- The first, and the most important, is that of **identification**. The major prerequisite for weaving a web is to have anchors. The history of identification on the Web goes through UDIs (Universal Document Identifier), URLs (Uniform Resource Locator), URIs (Universal Resource Identifier), and IRIs (Internationalized Resource Identifier). URLs and URIs are formats for addresses and identifiers allowing the localization, or simply the naming and referencing on the Web, of any type of resource, it is thus a resource-oriented vision of the network [4].
- The second fundamental notion of Web architecture is that of communication and data **transfer**. For this, the HTTP protocol allows a request to be made, starting with a URL address, for a representation of the resource identified and localized by the URL, and then the data corresponding to the representation is either obtained or error codes are produced indicating an encountered problem.

- The third fundamental notion is that of the **representation** of exchanged data. As the Web was initially motivated by the representation and exchange of documents, HTML was the first language proposed to represent, store, and communicate Web pages. After HTML, other language will be proposed to exchange other kinds of data than Web pages. In fact, Tim revisited the Web architecture in 1996 [4] putting an accent on addressing, protocols, and content negotiation, a native mechanism of HTTP, offering the possibility to serve, for a given URI, different versions of the same resource, and was directly inspired by the format negotiation mechanism of "System 33" [24] from Steve Putz at Xerox PARC.

Even more than each pillar individually, the way in which they are combined in a simple and elegant architecture is the core of the Web invention and greatly contributed to the virality of the Web [6].

6 MAKING THE WEB WORLD WIDE

At the end of 1990, the first server and the first browser were tested through an Internet connection and at the start of the 2000s there were more than 26 million Web servers running in the world. What made the Web so viral compared to competitors?

Some of the reasons were technical, starting with Tim's decision to stress the importance of generality, portability, and extensibility, and his position that they were more important than the satisfaction derived from taking advantage of the latest capabilities of computers (such as graphics at the time). If the Web today has made a major, sustainable technical contribution to the computer science community, it is in particular due to its simplicity, elegance, and extensibility [1].

Tim recognized that simplicity was a necessary condition for the generalized adoption of the Web. His simplification of protocols, including his insistence on the absence of states in the HTTP protocol, made the design easy to implement. In addition, the use of human-readable scripts made the system easy to debug, as well as virally reusable by copy-paste-adapt [6]. We all made our first Web page by copying-pasting-adapting the source of an existing page.

From the beginning, Tim also stressed that the result had to be sufficiently attractive in use to allow the information contained to surpass a critical threshold, and that this mass of information would in turn encourage ever more use and contributions. In that, he recognized the power of network effects as a positive feedback of Metcalfe's law, which states that the value of a network is proportional to the square of the number of its users/members. As value increases, more agents join the network to get the benefits, including information that they own into the network, and thus further increasing its value.

To achieve this, Tim proposed to support links between existing and new databases [7], and a systematic compatibility with what was already in existence. From its creation, the Web integrated gateway servers to import legacy resources and to allow access to other applications using techniques such as CGI (Common Gateway Interface). In this way, the Web would integrate, inter-operate, and ultimately absorb existing systems, most notably WAIS and Gopher. This approach facilitated the complete migration of communities of users of older systems to the Web. Tim also designed retro-compatibility or descending compatibility with earlier protocols (FTP, Gopher, WAIS, etc.) as an inter-operability constraint, a show of flexibility, and above all as an assurance of its evolutive nature for the future [4, 5]. Further, with approaches like the CGIs, the automatic generation of dynamic pages and the possibility of referencing and reading them immediately played a key role in the weaving of a Web that was sufficiently resourced (numerous nodes), connected, and dense (numerous links) [5].

Beyond these technical choices, other choices made for economic, legal or social reasons also made a difference in the acceptance of the Web. First of all, in 1993, the CERN announced that the architecture and the technological foundations of the Web would be made open source, free of rights, and without any fees [10]. Another decisive initiative was the establishment, in 1994, of the World Wide Web Consortium (W3C). Before the W3C, Web standards had been published in the form of RFCs (Request For Comments). They would thereafter be published as recommendations from the W3C. The consortium played a primordial role in the normalization of the evolution of the Web architecture, allowing it to grow without losing the standard inter-operability that gave it its universality.

Finally, the Web also benefited from important external positive factors, in particular the democratization of broadband access. It wasn't until the late 1990s that we saw broadband became readily available, thus enabling the growth of the home computing industry. With large numbers of users possessing high-speed access to the Internet in their homes, companies trying to sell things on the Web had a market for their products.

7 MUTATIONS OF THE WEB

No sooner, the Web emerged from its initial chrysalis of a distributed hypertext documentary system that it started evolving in multiple directions revealing and reaching for its full potential as universal hypermedia software architecture for the internet and every connected thing.

7.1 Mobile Access to the Web

As of 1996, companies were interested in providing Web access to mobile phones. In 1999, the QR code was made open source, which would contribute to its distribution and its being made available in image recognition applications on mobile phones, in particular for gathering URLs displayed around us. In the same period, WAP (Wireless Application Protocol) and WML (Wireless Markup Language) were proposed for adapting Web access and content to the constraints of mobile phones and their connectivity issues. These first attempts, as well as those made with PDAs with wireless network cards, would have to wait until the mid-2000s and the arrival of smartphones to have platforms upon which they could reach their full potential. The problem was first to overcome the limitations of mobile connections (screen, bandwidth, limited interactions, limited computational power, connection costs, etc.). Then, either by resolving these problems or through their disappearance with increases in the power of the terminals and networks, the

work moved to the question of the deeper adaption of usage under mobile conditions: geo-localization, adaptation of interactions and interfaces, access to personal data, contextualization, audio and vocal interaction, augmented reality, coupling of several terminals, etc. Today, mobile Web applications are common, and many native mobile applications have in fact largely been developed using the languages and standards of the Web. In addition, lower entry-level prices and the democratization of smartphones mean that in certain areas of the world, mobile phones are now the primary means that people have of accessing the Web and are the most widespread way of having an initial contact with the Web.

7.2 Metadata and Linked Data on a Semantic Web

The PICS (Platform for Internet Content Selection) standardization, which was one of the first recommendations of the W3C in 1996, allowed the filtering of inappropriate content, in particular for children. It was also an early example of the type of decentralization that was desired for the Web in general: the filters capturing user preferences were created and stored in the browsers; the descriptors were stored in the consulted sites, but were generated by third-party authorities. The idea of labelling by, and in reference to, a third-party reintroduces in the Web the idea of complex and rich links and represented an opening towards the general notion of metadata on the Web.

This idea was in line with another evolution towards a Web of documents and structured data. CSS (Cascading Style Sheets) was an important step that marked the beginning of the separation on the Web of content (HTML) from the way in which it is presented (CSS) allowing the formatting of a document to be made independent of and separated from its structure, and also to use the same formatting for multiple documents, or inversely to vary the formatting of a single document. With Web content now separated from its format, the content was able to evolve through the creation and management of its own data and document structures. This was achieved with the standardization of XML in 1998 and, in its wake, of several languages allowing its validation (DTD, XML Schemas), interrogation (XPATH, XQuery), transformation (XSLT), management (XProc), etc.

The same year, Tim published the roadmap for the Semantic Web [3] extending on his idea from an article in 1994 pushing the Web towards resources with semantics oriented towards machines to allow for more automated processing [5]. The 1998 roadmap would open the door to all of the work that was carried out on the Web of Data and the Semantic Web (RDF, RDFS, SPARQL, OWL, etc.). The Semantic Web and Linked Data are reasserting the importance of the links and triples are essentially typed links. The aim is to shift the emphasis of associative linking from documents to data, facilitating a more comprehensive form of automated reasoning. This shift is desirable for at least three reasons. First, it facilitates data reuse, often in new and unexpected contexts. Second, it helps reduce the amount of relatively expensive human information processing. Third, it releases the large quantity of information, not currently accessible, that is stored in other data management systems and applications.

7.3 Web Programming and a Web of Applications

At the birth of the Web a page would already be generated on the fly on request following a click on a simple link, or in response to a form being submitted. On the server, a code would use the CGI method (Common Gateway Interface) to process requests and generate pages in return. With the introduction of frames in around 1996, it became possible to reload only part of a page. The same year, JavaScript appeared and, in Web applications, it became possible to use, even conjointly, programming on the side of the server as well as on the side of the browser. This marked the beginning of a genealogy of programming languages and techniques for the Web, some of them more server-oriented (ASP, PHP, C#, Python, Ruby, Perl, JSP, etc.), others more client-oriented (e.g., Plugins, ActiveX, CSS+HTML), and some that could be used by both (Java Servlet and Applet, JavaScript). Finally, with the XMLHttpRequest component added to JavaScript, it became possible to exchange data between a page and its server, and also, thanks to its DOM (Document Object Model), to modify a displayed page without necessarily reloading it. This technique would be named AJAX (Asynchronous JavaScript And XML) in 2005 and was massively adopted in Web applications to conjugate code being executed on the client with code being executed on the server, allowing fluid interactions with the user. In parallel, the architecture of the Web was being increasingly studied and formalized (e.g. REST architecture [14]). In the early 2000s, the proposition of evolving towards a Web of services (SOAP, Simple Object Access Protocol, and WSDL, Web Services Description Language, standards) opened up a new direction for work, to use the Web as a programming platform, which is currently being realized through APIs and associated languages such as JSON. Some go so far as to talk of the Web as an operating system that goes beyond the global collection of services that are offered on the Internet, and that is independent of the computers and individual objects that are connected to it. They see the Web as the programming and execution environment par excellence for Internet applications.

7.4 Social Medias and Networks on the Web

When speaking of the high points in the evolution of the Web, we see references now to what is called Web 1.0, Web 2.0, Web 3.0, etc., terms that give the false impression of major software evolutions in the Web when they are changes in practice or even in the understanding of the Web. Web 1.0 corresponds essentially to the initial document-based and distributed hypermedia vision of the Web. Web 2.0, which is also called the Social Web, reflects both the opening up of the Web to writing, including the AJAX approach for interactions, and the enormous social exchanges that it permitted, with Wikis, blogs, forums, social media, etc., . We are all familiar with the impact of that evolution. Web 3.0 generally covers the integration of Semantic Web practices and the Web of Data into Web 2.0, such as with RDFa (RDF in Attributes) in the Facebook OGP (Open Graph Protocol) or the use of Schema.org in numerous sites integrating social functionalities (for example, votes, ratings, etc.).

What the Social Web also made visible is that the Web is a massive network of networks, and it has shown us quite dramatically, and in so many different ways, both how networks shape our lives

and the amazing potential of a global hypertext system. The Web is a network designed in such way you can easily overlay other networks on top of it and that is what happened with social networks.

7.5 Web of Things

One upcoming evolution that is currently under study is to make the Web a universal application platform for connected objects, called the Web of Things. Today, we literally envisage making a web of everything. But we can also remember the initial influence that RPCs had on the conception of distributed hypertext and see this evolution as a form of returning to the source. In addition, in representing the Web as a mesh of potential calls for procedures that we invoke with each link that we follow, we understand better why ambitions such as that of crawling, indexing, or archiving the Web are complicated or even paradoxical.



8 THIS IS FOR EVERYTHING

"This is for everyone" is the message with which Tim presented the Web during the opening ceremony of the 2012 Summer Olympic Games in London. The Web is for and affects everyone and is also becoming part of everything around us, and is being deployed in all corners of the world. The Web, for everyone, everywhere, and for everything. This Web, with its worldwide reach, a name which at the beginning could have been seen as immodest, revealed itself in just a few years to be a self-fulfilling prophecy, where its being designed to be universal allowed it in fact to evolve towards the universal.

With the Web of Data, Web of Applications, Web of services, Mobile Web, and a Web that is accessible, international, etc., the Web therefore initiated, very early, its transformation towards a generic hypermedia architecture for programming and interaction, and especially the generalization of a Web potentially linking all types of resources, whether computational or not. The Web can reach anything, as everything can be identified with a URI. As the principles of the Web are extensible and generic, they have allowed us to pass from a documentary vision of a global library to a network of protean resources. One of the major strengths of the Web is in its universality, but we will also see that it requires constant vigilance to be preserved.

9 AN ARTIFACT, A SCIENCE OBJECT AND A RESEARCH OBJECT

Even today, it is striking to see to what point the Web is both well-known and at the same time poorly understood, as we can see for example with the tenacious and all-too-often seen confusion between the terms Web and Internet. The Internet allows for the inter-connection between computer networks and connected objects in general. It provides a communication infrastructure that supports numerous applications such as: email, telephoning, and videoconferencing. The Web is the distributed hypermedia that has become the primary software architecture for applications on the Internet.

Beyond this confusion, the term Web is also often used in a way that fails to differentiate between the founding principles of the software architecture and the object that emerged from it, i.e. the web weaved by its billions of users. The architecture and the object that emerged from it have two histories that are linked but that cover different aspects. Each of these two aspects exhibits, in different ways, complexity that calls for both research and development. The architecture of the Web is based on protocols, models, languages and algorithms that need to be specified, designed, characterized, and validated, as well as on, systematically, constraints such as those related to upscaling, efficiency in time and memory, interoperability, and internationalization. To achieve this, the architecture of the Web and its extensions are the subjects of research, in particular in the digital sciences. Within computer science and the digital sciences – and this is true in other fields as well – the Web has expanded both as a new tool for work but also as a new subject bringing with it both solutions and new problems and needs.

With respect to the object that has emerged from the use of that architecture, the complexity of its usages, the heterogeneity and volume of its content, services, and data, the dynamic nature of some of its traffic, and the life cycles of its resources and communities, are equally sources of complexity that also call for a scientific approach and for theoretical, applied, experimental, and multi-disciplinary research.

The different examples of the evolutions of and the facets of the Web lead us to a new need: that of developing the means for studying the Web and its changes. The more the Web grows in its architectural complexity and in the resources linked to it, the more it calls for transdisciplinary research and development [16]. Tim would even go so far as to say that, ultimately, the Web is more of a social creation than a technical creation [6].

Historically, the Web quickly became a subject of research. Again under the impetus of Robert Cailliau and Tim Berners-Lee, in 1994 the WWW conference series now called "The Web Conference" was started at CERN and became the annual meeting place for the research, development, companies, and major actors of the Web. This research community is vivid [19] and has grown and diversified, for example with resolutely multidisciplinary initiatives such as "Web Science" or more specialized conferences such as ISWC (International Semantic Web Conference).

As we seek to understand the origins of the Web, appreciate its current state, and anticipate possible futures, there is an increasing need to address the critical questions that will determine how the Web evolves as both a social and a technical network. The study of

the Web—its evolution and its impact on society, on business, and on government—is referred to as Web Science. The Web, and the ways in which people interact with it, change faster than we can observe. This rapid evolution calls for new approaches and new methodologies to enable us to define and implement studies in Web Science. Economic, social, political, cultural, and economic agendas all have a role to play in shaping the new networks, and we need to understand the inter-play between those different perspectives to be able to anticipate the nature of the Web worlds of the future. Web Science must, by definition, be very multi-disciplinary, bringing together researchers from many different disciplines to help us shed light on what is actually going on in this very complex sociotechnical system. In order to monitor and analyze such developments, we need to be collecting evidence on a global scale of the changes brought about by the application of Web technology, and we need to be sharing the data we collect to be re-used in increasingly sophisticated analysis. This is a major role for the Web Scientists of the future.

10 WE HAVEN'T SEEN THE FULL POTENTIAL OF THE WEB YET

The Web has modified our relationship with time and space by giving us the possibility of interacting with distant people and things, by giving us access to information that is not locally available, by recording traces of our actions, by documenting a variety of activities, and by allowing us to revisit them and re-experience them a posteriori and remotely. We no longer note down the address of the dentist, we can find it on the Web. We no longer program our VCR, we look for a recording on-line. We no longer go to the train station to buy tickets, but rather download an e-ticket. The omnipresence and hypermnesia of the Web are so established now that we can't stand it when it's not there to respond to us, and the "no results found" page of search engines is virtually never seen anymore.

At the historical scale of computer science, the Web has not only uniquely proven itself as an architecture that has passed the test of time, it has defined a new era: there is an era before and an era after the Web. An era where the problem was to have information about or access to a service, and an era where the challenge was to find one's self within the overwhelming quantity of information. An era of fragmentation, and then an era of hyper-integration, even over-integration, with the risks that that entails.

But the defense of the Web remains necessary. It is universally useful and used, but it remains fragile, and its initial ideals could prove to be just a passing phase if we do not constantly keep watch on their preservation. Tim Berners-Lee is still fighting today against all forms of re-centralization. The stakes (neutrality, decentralization, democratization, etc.), the dangers (re-centralization, levels of access, a Web at different speeds, etc.), limitations (infrastructure needs, energy, costs, etc.) make the Web a never-ending project more than a final achievement. The architecture of the Web is and must remain robust, even in a hostile environment, neutral, even if its underlying layers are compromised, and resilient, even if the infrastructure is lacking or limited, etc. To protect it, we will have to design the spiders that weave the Web architecture to be extremophile animals.

Inversely, the Web can itself be perceived as a danger. In 1996, Tim wrote about how the diversifying force provided by geography could be weakened by the Web [4]. He also drew attention in a general way to the ethical and societal stakes of the Web: he recalled the impact that the choices made concerning Web architecture had on the forms of society in which we were living; the necessity of revisiting the notion of copyright in a space where copying could take on so many forms; the privacy-related problems introduced by the multiple opportunities to capture data; the impact that Web-based information has on a voting population; the necessity of working hand-in-hand with legislative systems; etc. [4]. These subjects raised in 1996 foreshadowed needs that would be even more urgent two decades later, and the need to actively search for interdisciplinarity in the study of the Web and of its evolution. It is important to understand the Web holistically, in all its complexity, and to encourage the "Web Science" movement towards transdisciplinarity. The three W's of the World Wide Web call for the three M's of a Massive Multidisciplinary Method [16].

Further, the Web must also absolutely become a subject for education and training in itself. Its use (basic principles of browsing, searching, etc.), best practices (critical reading, cross-referenced validation, active contributions, etc.), prevention (protection of privacy, protection of children, etc.), are all subjects that every generation should be taught in school as an important element for equal opportunity.

Whereas at the beginning, before the Web arrived, the problem was to imagine a world with the Web in it, we are now in the inverse situation where people have forgotten or can no longer imagine that there was a world without the Web [26]. Our choices in Web architecture do have an impact on our societies and this only increases with every evolution of the Web. We are witnessing profound changes in society, and we need to be constantly vigilant to ensure that the technology that is such an important driver in these changes is being developed as a force for social good, rather than as one that can be used to harm or control or manipulate.

"The Web as I envisaged it, we have not seen it yet.
The future is still so much bigger than the past."
– Tim Berners-Lee, 2009

Note: This paper is based on two previous publications by the authors "The Ever Evolving Web: The Power of Networks"[20] reflecting on the Web and its evolution and "For Everything" [17] providing an historical perspective on the Web and its mutations.

REFERENCES

[1] ACM. 2016. Citation Sir Tim Berners-Lee, Turing Award, For inventing the World Wide Web, the first web browser, and the fundamental protocols and algorithms allowing the Web to scale. http://amturing.acm.org/award_winners/berners-lee_8087960.cfm

[2] Keith Andrews, Frank Kappe, and Hermann Maurer. 1996. *The Hyper-G Network Information System.* Springer Berlin Heidelberg, Berlin, Heidelberg, 206–220. https://doi.org/10.1007/978-3-642-80350-5_20

[3] Berners-Lee. 1998. Semantic web road map. https://www.w3.org/DesignIssues/Semantic.html

[4] Tim Berners-Lee. 1996. WWW: Past, present, and future. *Computer* 29, 10 (1996), 69–77.

[5] Tim Berners-Lee, Robert Cailliau, Ari Luotonen, Henrik Frystyk Nielsen, and Arthur Secret. 1994. The World-Wide Web. *Commun. ACM* 37, 8 (Aug. 1994), 76–82. https://doi.org/10.1145/179606.179671

[6] Tim Berners-Lee and Mark Fischetti. 2001. *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor.* DIANE Publishing Company, Collingdale, PA 19023, USA.

[7] Timothy J Berners-Lee. 1989. *Information management: A proposal.* Technical Report. CERN.

[8] Vannevar Bush et al. 1945. As we may think. *The atlantic monthly* 176, 1 (1945), 101–108.

[9] Vinton Cerf and Robert Kahn. 1974. A protocol for packet network intercommunication. *IEEE Transactions on communications* 22, 5 (1974), 637–648.

[10] CERN. 2018. *The birth of the World Wide Web.* CERN. https://timeline.web.cern.ch/timelines/The-birth-of-the-World-Wide-Web

[11] Jeff Conklin. 1987. Hypertext: An Introduction and SurvevJ. *IEEE computer* 20, 9 (1987), 17–41.

[12] Douglas C Engelbart and William K English. 1968. A research center for augmenting human intellect. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I.* ACM, New York, NY, USA, 395–410.

[13] Roy Fielding, Jim Gettys, Jeffrey Mogul, Henrik Frystyk, Larry Masinter, Paul Leach, and Tim Berners-Lee. 1999. *Hypertext transfer protocol–HTTP/1.1, RFC7231.* Technical Report. IETF.

[14] Roy T. Fielding and Richard N. Taylor. 2002. Principled Design of the Modern Web Architecture. *ACM Trans. Internet Technol.* 2, 2 (may 2002), 115–150. https://doi.org/10.1145/514183.514185

[15] Andrew M Fountain, Wendy Hall, Ian Heath, and Hugh C Davis. 1990. MICROCOSM: An Open Model for Hypermedia with Dynamic Linking. , 298–311 pages.

[16] Fabien Gandon. 2014. The three 'W' of the World Wide Web call for the three 'M' of a Massively Multidisciplinary Methodology. In *WEBIST 2014 - 10th International Conference (Web Information Systems and Technologies, Vol. 226)*, Valérie Monfort and Karl-Heinz Krempels (Eds.). Springer International Publishing, Barcelona, Spain, 3–15. https://doi.org/10.1007/978-3-319-27030-2

[17] Fabien Gandon. 2017. For everything: Tim Berners-Lee, winner of the 2016 Turing award for having invented… the Web. *1024 : Bulletin de la Société Informatique de France* 11 (Sept. 2017), 21. https://hal.inria.fr/hal-01843967

[18] James M Gillies, James Gillies, R Cailliau, et al. 2000. *How the Web was born: The story of the World Wide Web.* Oxford University Press, USA, 198 Madison Avenue New York, NY 10016.

[19] Damien Graux and Fabrizio Orlandi. 2022. Through the Lens of the Web Conference Series: A Look Into the History of "the Web". In *Proceedings of The Web Conference 2020.* Association for Computing Machinery, New York, NY, USA, 7 pages.

[20] Wendy Hall. 2011. Network Theory| The Ever Evolving Web: The Power of Networks. *International Journal of Communication* 5 (2011), 14.

[21] Ted Nelson. 1981. *Literary Machines* (3rd ed.). Mindful Press, Sausalito, California.

[22] Theodor H Nelson. 1965. Complex information processing: a file structure for the complex, the changing and the indeterminate. In *Proceedings of the 1965 20th national conference.* ACM, New York, NY, USA, 84–100.

[23] Louis Pouzin. 1973. Presentation and major design aspects of the CYCLADES computer network. In *Proceedings of the third ACM symposium on Data communications and Data networks: Analysis and design.* ACM, New York, NY, USA, 80–87.

[24] Steve Putz. 1993. Design and implementation of the system 33 document service.

[25] Isabelle Rieusset-Lemarié. 1997. P. Otlet's Mundaneum and the international perspective in the history of documentation and information science. *Journal of the American Society for information science* 48, 4 (1997), 301–309.

[26] Neil Savage. 2017. Weaving the Web. *Commun. ACM* 60, 6 (May 2017), 20–22. https://doi.org/10.1145/3077334