SLAC-PUB-7815
May 1998

The World-Wide Web and High-Energy Physics

Bebo White

Submitted to *Physics Today*

Stanford Linear Accelerator Center, Stanford University, Stanford, CA 94309

Work supported by Department of Energy contract DE–AC03–76SF00515.

The World-Wide Web and High Energy Physics

(Submitted to *Physics Today*)

In his State of the Union address in January 1998, President Bill Clinton told the U.S. Congress - "We should enable all the world's people to explore the far reaches of cyberspace. Think of this -- the first time I made a State of the Union speech to you, only a handful of physicists used the World-Wide Web. Literally, just a handful of people. Now, in schools, in libraries, homes and businesses, millions and millions of Americans surf the Net every day." [1]

Who were these physicists that President Clinton was referring to, and why did the physics community provide the fertile ground from which the Web was born? While the Web has come to mean many things to many people - electronic commerce, digital libraries, mass advertising, and many others - its roots lie in the simple need for scientists living and working far apart to collaborate with one another on common interests and purposes.

CERN, located on the Franco-Swiss border near Geneva, is probably the most successful high-energy laboratory in the world. Its success can be measured both in terms of the number of Nobel prizes research there has garnered, and in terms of how it has proven that international collaborations can work together conducting sophisticated experiments. Experiments at CERN (and other high-energy physics laboratories) use complex particle detectors which cost millions of dollars. It is not unusual for an experimental collaboration to involve hundreds of physicists from institutions spread throughout the world.

Two major elements of successful collaboration are communication and documentation. International collaborators must be able to quickly share not only experimental data, computer program code, and drafts of scientific papers, but also more mundane administrative information, such as meeting minutes and agendas and telephone and e-mail lists. Computing and networking technologies provide useful tools for collaborations, but can also create roadblocks - collaborating institutions use different computer operating systems and document preparation systems. It was out of this environment and to address these problems that the Web was born at CERN.

The development of the Web at CERN was actually the convergence and syntheses of three separate technologies - networking, document/information management, and interface design. Application of these technologies to the needs of the experimental collaborations led to the solution that became the Web. However, as will be described later in this article, recognition of the Web as the solution was not immediately realized.

"In the beginning was - chaos. In the same way that the theory of high-energy physics interactions was itself in a chaotic state up until the early 1970's, so was the so-called area of 'Data Communications' at CERN. The variety of different techniques, media and protocols used was staggering; open warfare existed between many manufacturer's proprietary systems, various home-made systems, and the then rudimentary efforts at defining open or international standards. There were no general purpose Local Area Networks (LANs); each application used its own approach." [2]

Network development at CERN continued through the late 1970's and the 1980's. CERNET, beginning in 1976, linked a number of mainframes and minicomputers. The STELLA Satellite Communication Project, from 1981-83, used a satellite channel to link CERNET running between CERN and INFN-Pisa and a Cambridge Ring network between CERN and the Rutherford-Appleton Laboratory (RAL) in the United Kingdom. In 1983, Ethernet was first introduced at CERN. DECnet and BITNET provided networking capability for VAX VMS and IBM VM/CMS platforms, respectively. An overall networking strategy was still lacking.

Between 1985 and 1988, CERN local networks moved methodically to TCP/IP. In 1989, CERN opened its first external connections to the Internet and by 1990 it was the largest Internet site in Europe. CERN quickly became a major force in the development and support of the Internet both in Europe and worldwide.

In his "*A Short History of Internet Protocols at CERN*," Ben Segal writes, "A key result of all these happenings was that by 1989 CERN's Internet facility was ready to become the medium within which Tim Berners-Lee would create the World-Wide Web with a truly visionary idea. In fact an entire culture had developed at CERN around 'distributed computing'…It is my belief that the Web could have emerged considerably earlier if CERN had been connected earlier to the Internet." [2]

Network developments at CERN provided the networking infrastructure needed for the development of the Web. Missing, however, was an information infrastructure that could be easily delivered to the user community over these networks using a technology such as the Web.

In 1983-84, a working group was set up to design an electronic document handling system that would facilitate the storage and retrieval of CERN's engineering/computing documents. At that time, document preparation at CERN made heavy use of Waterloo Script (on IBM VM/CMS). In 1985, some groups began experimenting with SGML (the Standard Generalized Markup Language). The first prototype of this documentation system was implemented as a virtual machine on VM/CMS and named CERNDOC. It was one of the first electronic documentation systems to use a client/server model. It used SGML, though it wasn't dependent on a particular DTD (Data Type Definition) - the Web's HTML is an SGML DTD.

Eric van Herwijnen (one of the architects of CERNDOC) recollects, "Tim [Berners-Lee] (and Mike Sendall, who later became something like the 'adopted godfather of the Web') was very interested in CERNDOC and spent quite some time talking about possible extensions to CERNDOC. The way that images were handled was particularly unsatisfactory. PostScript was just becoming popular at that time, and the system couldn't automatically translate from IBM's EBCDIC to ASCII for VAX clients. I believe that Tim got the client/server idea from CERNDOC as well as the idea of format negotiation through a special purpose protocol that relied on something looking like SGML as the document format." [15]

With the Internet providing the transfer mechanism, and CERNDOC providing a prototype for content, the only piece of the Web missing was the user interface. FTP (File Transfer Protocol) and Gopher were accepted Internet information retrieval systems, but neither provided a uniform, platform-independent user interface. And perhaps worst of all, both systems required user knowledge of a specific set of commands. The solution to the interface problem was found in pointing and linking features of hypertext!

In his 1989 proposal "Information Management: A Proposal" [3], Tim Berners-Lee wrote, "In 1980, I wrote a program for keeping track of software with which I was involved. Called Enquire, it allowed one to store snippets of information, and to link related pieces together in any way. To find information, one progressed via the links from one sheet to another, rather like in the old computer game 'adventure.' I used this for my personal record of people and modules. It was similar to the application HyperCard produced more recently by Apple for the Macintosh. A difference was that Enquire, although lacking the fancy graphics, ran on a multi-user system, and allowed many people to access the same data."

In his proposal, Berners-Lee offered a solution for the management of general information about accelerators and experiments at CERN and the problems of loss of information as "the integration of a hypertext system with existing data, so as to provide a universal system, and to achieve critical usefulness at an early stage." Figure 1 shows a diagram from the proposal illustrating how the proposed system would integrate such diverse information sources such as VAXNOTES, USENET, IBM GroupTalk (which the author implemented at CERN), and CERNDOC. This proposal never used the term "World-Wide Web."

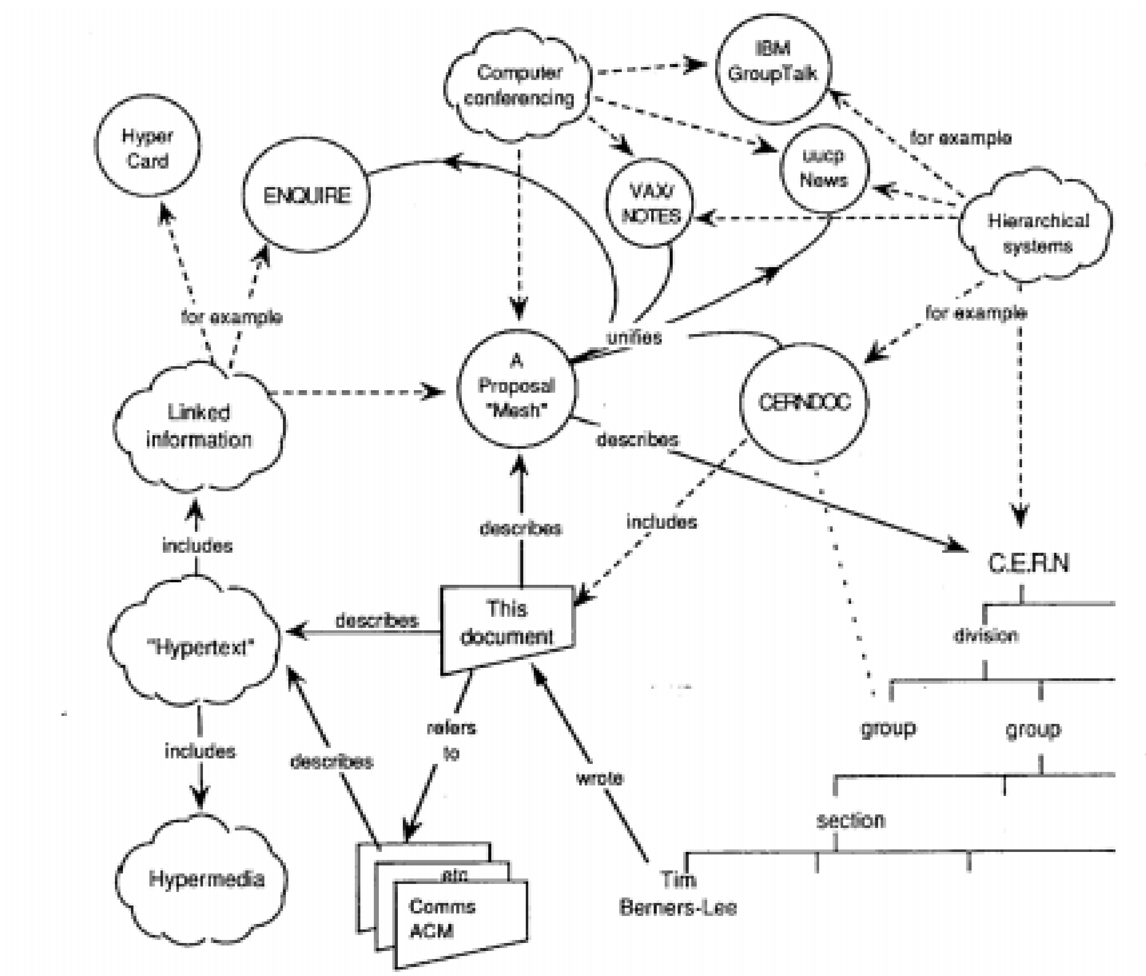


Figure 1. A CERN Distributed Hypertext System

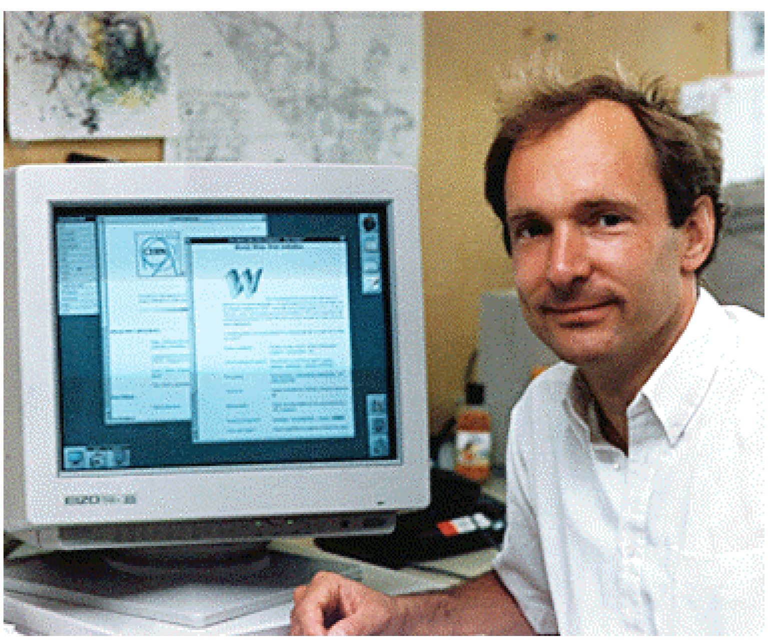
Independently of Berners-Lee, Robert Cailliau was also interested in use of hypertext as a solution to CERN's information retrieval problems. He was especially interested in platform-dependent hypertext systems such as HyperCard and hypertext scripting languages such as HyperTalk. In 1990, Berners-Lee and Cailliau decided collaborate on a revision of the 1989 proposal to CERN management. The new proposal was entitled "World-Wide Web: Proposal for a Hypertext Project" [4].

Cailliau recalls the discussion with Berners-Lee (held at the CERN cafeteria) over what to name the proposed project: "We agreed that Greek names were out - too old-fashioned. Then one of us had the idea: 'What about the World-Wide Web?' 'That'll do,' said the other. 'Let's put that down, and think of something better later.' " [15] Later never came.

In October of 1990, Berners-Lee and Cailliau submitted the revised proposal. It defined the key goals of the project as,

- “ to provide a common (simple) protocol for requesting human readable information stored at a remote system, using networks;
- to provide a protocol within which information can automatically be exchanged in a format common to the supplier and the consumer;
- to provide some method of reading at least text (if not graphics) using a large proportion of the computer screens in use at CERN;
- to provide and maintain at least one collection of documents, into which users may (but are not bound to) put their documents. This collection will include much existing data;
- to provide a keyword search option, in addition to navigation by following references, using any new or existing indexes. The result of a keyword search is simply a hypertext document consisting of a list of references to nodes which match the keywords;
- to allow private individually managed collections of documents to be linked to those in other collections;
- to use public domain software wherever possible, or interface to proprietary systems which already exist;
- to provide the software for the above free of charge to anyone.” [4]

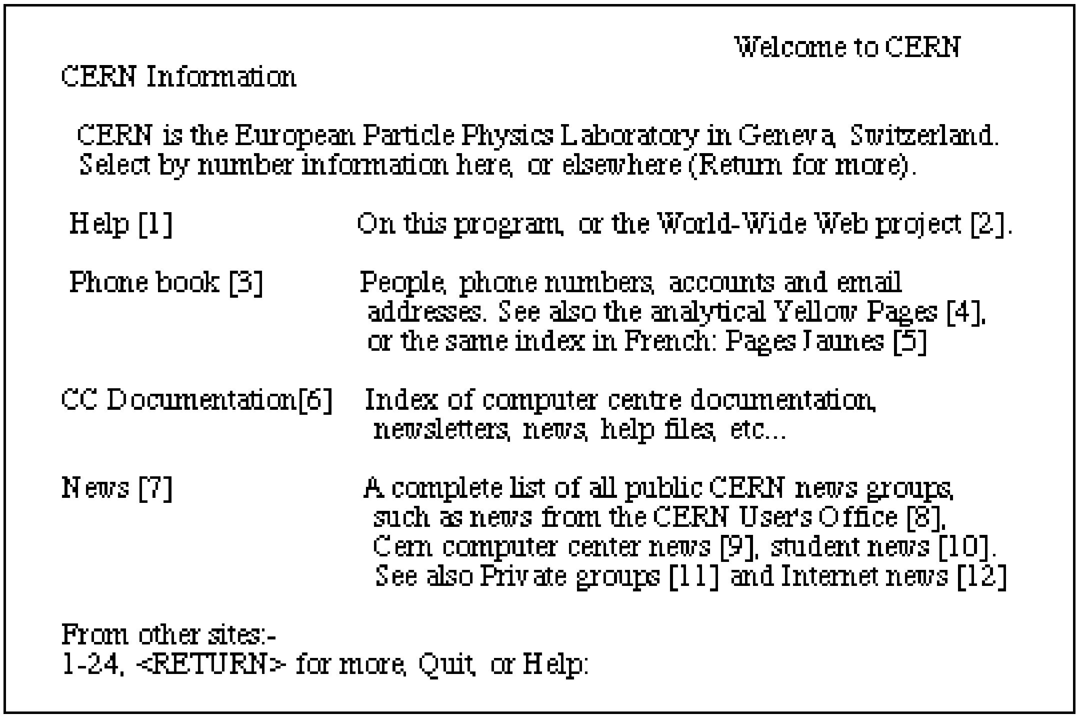
With the Berners-Lee/Cailliau proposal, all the pieces contributing to the development of the Web at CERN were complete. By November 1990, the initial implementation of a Web client for the NeXT platform had been developed. This first browser was able to display documents using multiple fonts and styles and was even able to edit documents (capabilities only now becoming widespread in many browsers) , but access was limited to NeXT users.



Tim Berners-Lee

The NeXT "Cube" Which Ran the Original WWW Server and Browser

March 1991 brought the development of the CERN "linemode" browser (Figure 2.), which could run on numerous platforms but displayed output only on character-based (e.g., VT100) terminals. The Web server code was subsequently ported to the CERN VAX/VMS platform and the IBM VM/CMS mainframe. In December 1991, the CERN Computer Newsletter announced and described the Web to the high-energy physics community.



Welcome to CERN

CERN Information

CERN is the European Particle Physics Laboratory in Geneva, Switzerland.
Select by number information here, or elsewhere (Return for more).

Help [1]	On this program, or the World-Wide Web project [2].
Phone book [3]	People, phone numbers, accounts and email addresses. See also the analytical Yellow Pages [4], or the same index in French: Pages Jaunes [5]
CC Documentation[6]	Index of computer centre documentation, newsletters, news, help files, etc...
News [7]	A complete list of all public CERN news groups, such as news from the CERN User's Office [8], Cern computer center news [9], student news [10]. See also Private groups [11] and Internet news [12]

From other sites:-
1-24, <RETURN> for more, Quit, or Help:

Figure 2. The First CERN Linemode Browser

Contrary to popular belief, the Web did not initially take the CERN and high-energy physics community by storm. Hansjorg Klein of the Delphi collaboration recalls that a Hypertext Colloquium given by Berners-Lee at CERN in November of 1990 was "overwhelming to many physicists." [15] Early efforts to get the Delphi experiment "on the Web" were not extremely successful.

Eric van Herwijnen remembers, "As for initial user reactions, I remember that no one believed in the Web at CERN in the beginning. Many people thought that Tim was not right in inventing yet another protocol (HTTP) - now they take part of the claim of having inventing it. The Web was only taken seriously here after Mosaic came along." [15]

Paul Kunz, a mainstay of the high-energy physics computing community who was instrumental in establishing the first U.S. Website at the Stanford Linear Accelerator Center (SLAC), recalls " The first I heard about the WWW was an article in a comp.sys.next newsgroup. Tim announced to NeXTStep users the availability of his application. My thoughts were something like: 'what are these CERN people up to now?' I had absolutely no interest as the application was pushed as a way to distribute documentation, and what developer finds documentation fun?" [15]

Others were quick to see the Web's potential: In early 1992, Ruth Pordes and Jonathan Streets of the FermiLab Computing Division's Online Systems Department (OLS) were considering the problem of providing information about online data acquisition systems to high-energy physics experimenters. Seeing the Web presentation to Artificial Intelligence in High Energy Physics (AIHEP'92) at La Londe, France in February 1992, Streets recommended the Web as being "the best thing around," and OLS decided to adopt it. [5] Paul Kunz attended the same presentation and recalls that the Web demonstration included a connection to the SLAC server, which was already in production at that time.

At the wrap-up session of the Computing in High Energy Physics Conference (CHEP) held at Annecy, France in September, 1992, Terry Schalk of the Santa Cruz Institute for Particle Physics (SCIPP) announced to attendees that the conference highlight for him had been the short presentation on the Web. Since then, Web applications have played a major role at the CHEP conferences. (CHEP97 in Berlin held joint sessions via videoconferencing with the Sixth International WWW Conference in Santa Clara, CA.)

Figure 3. is a photograph of a T-shirt made for CHEP92. The shirt bears the names of the 5 high-energy physics laboratories on the Web at the time - CERN, SLAC, DESY, NIKHEF, and FNAL. The joke amongst the designers of the shirt was that "maybe some day we'll have enough sites such that the names will go around the shirt wearer's chest."

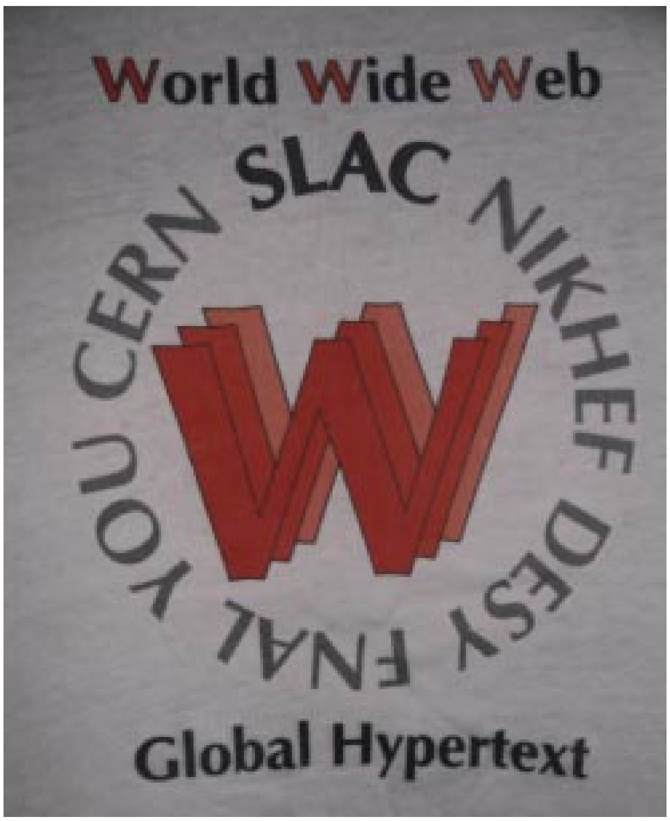


Figure 3. The First WWW T-shirt

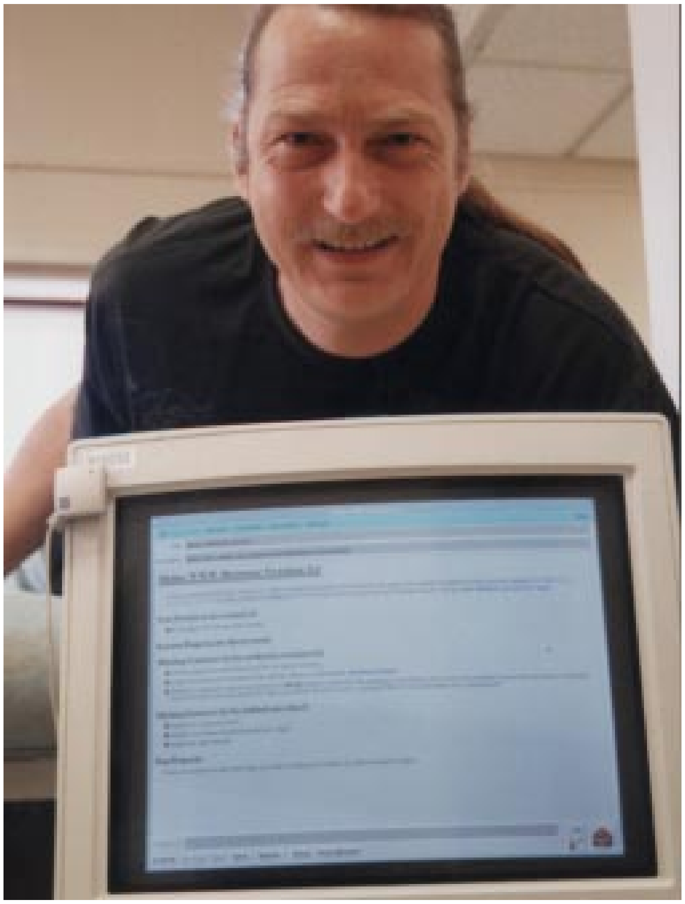
Bjorn Nilsson recalls that in June 1993, the ALEPH experiment at CERN was able for the first time to generally provide information previously limited to ALWS (the Aleph offline system) users. A Web version of ALWHO provided accurate phone list and e-mail data. ALWS HELP and detector status were also provided. [15]

The Web is often described as the “killer application” of the Internet. However, in these early days of its implementation, it needed its own “killer app” in order to “catch on” within the high-energy physics community. That application may well have been the Web’s readily accessible, easy to use interface to the SPIRES databases at the Stanford Linear Accelerator Center (SLAC). SPIRES (the Stanford Public Information REtrieval System) is used to support a set of databases covering a wide range of topics of relevance to high-energy physics including experiments, institutes, publications, and particle data.

Since 1974, the SLAC library has participated in providing SPIRES-HEP, a 300,000 record bibliographic database, to the world’s high-energy physics community. Prior to its implementation on the Web, the only way to access SPIRES-HEP was to log in to the IBM VM/CMS system at SLAC where the database resides, or to use the QSPIRES interface which operated only from remote BITNET nodes. With either method, a rudimentary knowledge of the somewhat esoteric SPIRES query language was required. According to Tim Berners-Lee, the mounting of SPIRES-HEP on the Web was a vitally important factor in the rapid acceptance and utilization of the Web in the international high-energy physics community. [6]

While ready access to SPIRES may have aided in the acceptance of the Web in the high-energy physics community, “It wasn’t until NCSA (the National Center for Supercomputing Applications) made Mosaic that the average person could see why this was really a good thing,” recalls David Ritchie, who chaired the working group that designed the first home page for Fermi National Laboratory. [7] Mosaic brought the Web from physics and computer science laboratories and set the path for the role it plays today. The first version of Mosaic came in February 1993, but prior to its release, Web browser development had hardly been stagnant.

The early browsers developed at CERN were quickly followed by the first browsers designed for X Windows - the Viola browser developed by Pei Wei at the University of California, Berkeley (and later O’Reilly and Associates), and the Midas browser developed by Tony Johnson at the Stanford Linear Accelerator Center (SLAC).



Tony Johnson and the Midas browser

Like Mosaic, both the Midas and Viola browsers were freely available on the Internet. The features they provided undoubtedly had an impact on Mosaic's development. Ruth Pordes of FermiLab remembers, "In 1993 I invited Tim [Berners-Lee] to FermiLab. I contacted NCSA and suggested that we go down and meet them. It was clear they were really going to market Mosaic. That was the first time we realized that he [Berners-Lee] was really going to have a big success." [7]

Through 1993 and 1994, CERN and NCSA continued to be the leaders in Web development. NCSA provided a rapid succession of improvements (and new features) to Mosaic and developed the NCSA server. The small but dedicated Web group at CERN continued their efforts on the CERN server, but more importantly, developed and supported Libwww, the general software library for building Web browsers and servers. Pioneering work on critical Web features such as access authorization resulted from the Libwww effort. In May 1994, the First International WWW Conference was held at CERN. Described by one press pundit as "the Woodstock of the Web," it was rapidly oversubscribed.

Tim Berners-Lee had long argued that standardization of the Web be open and non-proprietary. His model was that of the X Consortium, formed in 1988 to further the development of the X Window System and having as its major goal the promotion of cooperation within the computer industry in the creation of standard software interfaces at all layers in the X Window System environment. MIT had for many years provided the vendor-neutral architectural and administrative leadership required to make that organization work. In December 1994, after much negotiation, Berners-Lee left CERN for the MIT Laboratory for Computer Science (LCS) to become director of the newly-formed World-Wide Web Consortium (W3C) (http://www.w3.org/).

Perhaps it is the ultimate irony that it was high-energy physics that brought the end to Web development at CERN (and effectively at any other high-energy physics laboratories). In late December 1994 and early January 1995, the CERN Council approved unanimously the construction of the Large Hadron Collider (LHC), the world's most ambitious high-energy physics accelerator to date. Due to stringent budget conditions resulting from the LHC decision, the Council decided not to continue Web development and, in negotiations with the European Commission and INRIA (the Institut National pour la Recherche en Informatique and Automatique, France), transferred the CERN Web Project to INRIA. It was further agreed that INRIA along with MIT/LCS would host the W3C.

There is no longer any question regarding the acceptance or usefulness of the Web to the high-energy physics community. It would be difficult to imagine the operation of a major experiment or laboratory without the Web as an essential tool. Institutional and experimental resources on the Web have made research in high-energy physics more accessible to the general public. Many laboratory Websites now contain information and resources specifically targeted to general audiences, such as the educational community and the press. Web pages have also provided institutions with an excellent mechanism for exceptional public announcements such as Nobel Prizes (e.g., http://www.slac.stanford.edu/slac/hottopic/mperl95/mperl95.html) and important discoveries (e.g., http://www.fnal.gov/pub/top95/top_news_release.html).

While the Web is viewed by many as having evolved into an advertising and commercial medium where the credibility of much of its content should be in question, the high-energy physics community (and perhaps most scientific communities) continues to use it in many of the same ways which prompted its conception. There are, however, at least three major areas in the scientific community that are likely to be dramatically changed by some of the new Web technologies - software development, management, and distribution; collaborative tools; and Web/Electronic publishing.

Software Development, Management, and Distribution

High-energy physics experiments depend on computer software. The Web has played an important role in the success that experimental groups have had in maintaining central program source code bases while allowing distributed software development. It has also made a significant contribution to the process of distributing software, documentation and associated databases to remote sites. Good examples of current applications in this area are the Web interface to the FreeHEP collection (http://slacvm.slac.stanford.edu:5080/FIND/FHMAIN.html) and the CERN LIGHT (LIfe cycle Global HyperText) system (http://alephwww.cern.ch/LIGHT/).

A Web-based software system (development, management, and distribution) for high-energy physicists in the near future will be quite different and exciting. A physicist might write platform-independent analysis software in Java on a local PC or workstation. This software could be compiled, tested, and debugged locally and then uploaded to a remote server. Software users would run the application in their Web browsers after downloading the program code from the server. The server could also provide datasets of the most current experimental data. Such a system could potentially eliminate the need for local program libraries and guarantee access to the latest data. In those instances where local program and data libraries are necessary, Web push technology could provide version control by installing the latest software fixes and patches. Current or especially significant datasets could also be pushed to experiment collaborators for analysis.

Collaborative Tools

Physicists must collaborate. The first Web-based conferencing system, WIT, was developed at CERN in 1994 as "a quick hack." [8]. The BaBar experiment makes extensive use of HyperNews, a conferencing system developed by NCSA. (http://babar-hn.slac.stanford.edu:5090/HyperNews/index). Various other commercial and "home-grown" tools are also being used in other high-energy physics collaborations.

Text-based conferencing systems and even videoconferencing only provide tantalizing pieces of the collaborative systems of the near future. DOE2000 is a new initiative [9] that will fundamentally change the way that scientists work together and how they will address the major challenges of scientific computation. Web technology plays an important role in the implementation of the initiative's concept of the *collaboratory*, a '...center without walls, in which researchers can perform their research without regard to geographical location - interacting with colleagues, accessing instrumentation, sharing data and computational resources, and accessing information in digital libraries.' [10]

Web/Electronic Publishing

Peter Boyce and Heather Dalterio [11] in their analysis of electronic publishing of scientific journals reach the conclusion "In the end it [electronic publishing] creates new capabilities that extend far beyond what paper journals can provide. It is precisely these new capabilities that will make electronic publishing such a powerful tool for scientists." This enthusiastic vision was, however, tempered by their concern that at the time of their work (1996), Web-based electronic publishing suffered from at least three major shortcomings: 1) "HTML is not sufficiently structured or versatile;" 2) "Even the best papers are of no use if they cannot be found when you need them. The research literature is of most value when you can search and recover articles of interest from a large fraction of the entire collection;" 3) "Who will pay…how to charge the users. They must collect enough revenue to support the infrastructure…revenue collection should not inhibit the use of electronic tools."

Recent advancements in Web technology have addressed these three issues raised by Boyce and Dalterio "head-on." The limitations of HTML for scientific papers are well known. The eXtensible Markup Language (XML) presently being defined by the W3C promises to change this situation. XML has the power of SGML but with the compactness of HTML [12]. With XML, Web authors can create sets of data element tags and structures that define and describe the information contained in a document. XML emphasizes the separation of document structure and content. In conjunction with technologies supporting style sheets (CSS, XSL, and DSSSL) and mathematical markup

(MathML), XML will provide high-energy physics authors with the tools and functionality previously available only with tools such as TeX/LaTeX or SGML.

Continuing advancements in search engine technology will make finding material on the Web easier, but only address a part of the problem faced by researchers. Document metadata will be used to assure document applicability, quality, and stability. Efforts such as the Dublin Core initiative (http://purl.org/metadata/dublin_core) are seeking to identify a core set of metadata elements for Web/Internet resource discovery. Originally conceived for author-generated descriptions, it has also attracted the attention of formal resource description communities such as museums and libraries. The metadata elements of the Dublin Core fall into three groups which roughly indicate the class or scope of the information stored in them: 1) elements related mainly to the content of the resource (e.g., subject or source); 2) elements related mainly to the resource when viewed as intellectual property (e.g., creator or rights); 3) elements related mainly to the instantiation of the resource (e.g., date or format).

Micropayment technology [13] addresses the revenue generation issues of electronic publishing. Document subscriptions work well in "the paper world," but are unlikely to work on the Web. Under a micropayment scheme, readers of Web documents will pay cents (or fractions of cents) for access. "Micropayments lower the threshold and do not require a big decision before users get their initial benefits." [14]

Acknowledgements

I wish to thank Robert Cailliau, Tony Johnson, Hansjorg Klein, Paul Kunz, Jennifer Masek, Bjorn Nilsson, and Eric van Herwijnen for their help and recollections. This work was supported by the US Department of Energy Contract No. DE-AC03-76F00515.

References

1. < http://www.whitehouse.gov/WH/SOTU98/address.html>
2. Segal, B. *A Short History of Internet Protocols at CERN*. CERN PDP-NS, April 1995; also at <http://wwwcn.cern.ch/pdp/ns/ben/TCPHIST.html>
3. Berners-Lee, T. *Information Management: A Proposal*. CERN DD/OC, March 1989, May 1990; also at <http://www.w3.org/History/1989/proposal.html>
4. Berners-Lee, T. and Cailliau, R. *World-Wide Web: Proposal for a HyperText Project*. CERN, October 1990.
5. Garrett, P. and Ritchie, D. *Collaborating Over the Web: Libraries and Laboratories or The Librarian and the Webmaster*. FERMILAB-Conf-95/056, March 1995.
6. Berners-Lee, T. Keynote Address, April 11, 1997, History/Developers' Day, Sixth International WWW Conference, Santa Clara, CA; also at <http://www.w3.org/Talks/9704WWW6-tbl/>
7. Berger, E. *High-Energy Physics: Birthplace of the Web*. FermiNews, Vol.19 Number 16, August 16, 1996.

8. Luotonen, A. and Berners-Lee, T. *W3 Interactive Talk*; also at <http://wwwlab.cern.ch/WebOffice/wit/Overview.html>
9. <http://www.mcs.anl.gov/DOE2000/>
10. Kouzes, R., Myers, J. and Wulf W. *Computer* 29(8):40-46 (1996).
11. Boyce, P., and Dalterio, H. *Physics Today* 49(1):42 (1996).
12. Connolly, D. (edit). *XML: Principles, Tools, and Techniques*. World-Web Web Journal, Vol. 2, Issue 4, O'Reilly (1997).
13. Herzberg, A., *Safeguarding Digital Library Contents: Charging for Online Content*. D-Lib Magazine, January 1998; also at < http://www.dlib.org/dlib/january98/ibm/01herzberg.html>
14. Neilsen, J., *The Case for Micropayments*, <http://www.useit.com/alertbox/980125.html>
15. Private communication with the author