SLAC-PUB-7757
February 1998

The World Wide Web and High Energy Physics

Bebo White

Stanford Linear Accelerator Center
Stanford University
Stanford, California 94309

(Submitted to Annual Review of Nuclear Particle Sciences)

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

CONTENTS

ABSTRACT

High energy physics and the World Wide Web (WWW) share a rich history. The Web, developed at CERN as a collaboration tool and quickly adopted by the Internet community, has become a communications phenomenon. This article reviews early WWW development and its basic technology. I also summarize some significant applications of Web technology past and present. Prospects for future use are discussed.

1. INTRODUCTION

The conferences on Computing in High Energy Physics (CHEP) offer a regular occasion for people active in the field to meet, to discuss progress and evaluate technologies, and to make projections for the future. The goal of the conference series has been to determine how advanced computing and networking can effectively enable high energy physics. CHEP92 held in September 1992 in the Alpine setting of Annecy, France was no exception. The technical program included papers in areas such as "Triggering and Data Acquisition," "QCD Calculations, Architecture and Experience," "Computer-Aided Detector Design," and "Mass Storage Technology."

Yet, at the conference wrap-up session, Terry Schalk of the Santa Cruz Institute for Particle Physics (SCIPP) indicated that a significant conference highlight for him was the discussion of a networking technology recently developed at CERN auspiciously named the World Wide Web or more simply WWW (1). CHEP92 was held before the introduction of Mosaic by the National Center for Supercomputing Applications (NCSA) at the University of Illinois which took the Internet world by storm in February 1993. It was certainly before WWW became the household word and have the ubiquitous presence that it has today. At the time of CHEP92, the number of Websites internationally was probably less than fifty, many of which were at high energy physics institutions. The first Website in the United States at SLAC (the Stanford Linear Accelerator Center) had only recently "come on-line."

It may be difficult to think of the 1990-1993 time period as "early days," but the phenomena that is the World Wide Web was developed during that so recent period. Since that time, the HEP community has played a vital role in the development and creative usage of the Web and has helped to shape what has become known as the "killer application of the Internet." The demands of increasingly complex HEP experiments will continue to push the capabilities of the Web and provide imaginative applications usable not only by HEP but by the entire Internet community.

2. THE CERN WORLD WIDE WEB PROJECT

The background of WWW in high energy physics illustrates the need for collaboration in HEP. Most of the features of the early Web servers and browsers were developed to address that need. The Web's beginnings were far simpler than the enormous technological and commercial presence into which it has recently evolved. Many of today's dedicated "Web-surfers" would hardly recognize some of the early implementations of the technology.

An evaluation of the needs of the HEP community at CERN and its experimental collaborators and a review of current technological trends such as FTP and Gopher led Tim Berners-Lee and Robert Cailliau of the CERN Computing Division to write a project proposal to their management in March 1989. This proposal was entitled, quite simply, "World Wide Web: Proposal for a Hypertext Project"(2).

In the preface to their proposal, Berners-Lee and Cailliau write,

> "The current incompatibilities of the platforms and tools make it impossible to access existing information through a common interface, leading to waste of time, frustration and obsolete answers to simple data lookup. There is a potential large benefit from the integration of a variety of systems in a way which allows a user to follow links pointing from one piece of information to another one. This forming of a web of information nodes rather than a hierarchical tree or ordered list is the basic concept behind Hypertext.
>
> The project [WWW] will aim:
>
> - 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."

In many ways, the timing of this proposal was ideal. CERN had become the largest Internet site in Europe. In his "A Short History of Internet Protocols at CERN," Ben Segal recalls "...an entire culture had developed at CERN around 'distributed computing', and Tim [Berners-Lee] had himself contributed in the area of Remote Procedure Call (RPC), thereby mastering several of the tools that he had needed to synthesize the Web such as software portability techniques and network and socket programming. " Segal also contends "...the Web could have emerged considerably earlier if CERN had been connected earlier to the Internet" (3).

In October of 1990, Berners-Lee and Cailliau submitted a revised proposal. By November, 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.

March 1991 brought the development of the CERN "linemode" client/browser (Figure 1.), which was executable on numerous platforms but which displayed output only on character-based (VT100) terminals. The WWW 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 WWW to the HEP community.

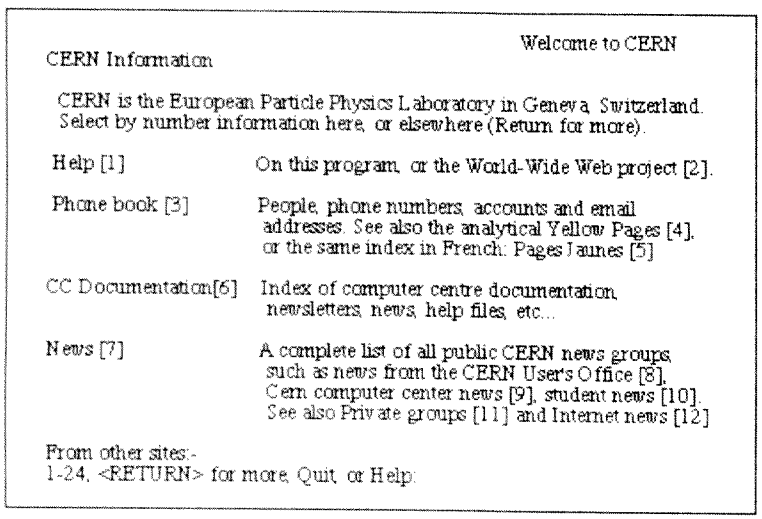


Figure 1. The First CERN Linemode Browser

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 the Midas browser developed by Tony Johnson at the Stanford Linear Accelerator Center (SLAC). Both of these browsers pre-dates the more well known Mosaic browser at NCSA and were freely available to Internet users. Features provided in the Viola and Midas browsers appear to have had an impact on the Mosaic development process.

3. WWW CONCEPTS

The World Wide Web was designed to incorporate two fundamental concepts: hypertext documents and network-based information retrieval.

3.1 *Hypertext*

In 1965, Ted Nelson devised the term "hypertext" to describe non-sequential or non-linear text. It was defined as "a body of written or pictorial material interconnected in a complex way that it could not be conveniently represented on paper" (4). Nelson's intention was to define a mechanism for presenting information which more closely resembled the human thought process which involves free association rather than the linear limitations which he felt that the written word imposes.

While there are classic examples of non-sequential text applications, computer and networking technologies made it possible to realistically implement Nelson's vision of a " document universe" or "docuverse." An essential feature of computer-based hypertext is the concept of computer-supported links (both within and between documents). It is this linking capability which supports the non-sequential organization and presentation of text so fundamental to the hypertext paradigm.

Simple documents become hypertext documents when they contain words or phrases act as links to other documents. Typically hypertext documents are presented to a user with the text that can act as a link highlighted in some manner, and the user is able to access the linked document by selecting the highlighted area. The manner by which the link is selected is dependent upon the Web browser implementation. Some browsers have text interfaces (e.g., the CERN linemode browser and the Lynx browser from the University of Kansas) where links are specified via menus or highlighted text. Other browsers such as Midas, Mosaic, the Netscape Navigator and the Microsoft Internet Explorer support graphical user interfaces (GUIs) which allow a user to point to a link (usually with a mouse) and activate the transfer to the linked document (usually with a click).

HTML Hypertext pages and documents are defined within the WWW environment using the HyperText Markup Language (HTML). HTML is an implementation of SGML, the Standard Generalized Markup Language, a text-formatting system widely used in the HEP community. Quite simply, HTML is a small collection of document markup tags which can be used to define hypertext documents which are portable from one computing platform to another. HTML contains elements which are characteristic of descriptive and referential markup. These markup tags permit a page or document author to specify how information will be presented and which page objects will operate as hypertext links. The portability of HTML marked-up documents permits their independent use by Web browsers (5).

Numerous HTML references, standards documents, and style guides are widely available on the Web. HTML books and guides are available from most well-known and lesser-known technical publishers. The number of these books in print rivals that of books on any other computer programming language or application.

3.2 *Network-Based Information Retrieval*

The World Wide Web accomplishes network-based information retrieval by using a client-server architecture. A WWW user executes a WWW client program (usually referred to as a browser) on their local computer which is connected to a network. The client fetches documents from remote network nodes by connecting to a server on that node, requesting the document to be retrieved, and displaying that document to the WWW user.

3.2.1 THE WWW CLIENT-SERVER MODEL The client-server model offers advantages to both the information provider and the information user. The information provider is able to maintain control of their documents since they reside locally on the provider's computer. Furthermore the documents can be maintained by the information provider in any format, provided that format is recognized by the WWW server and user's client. This capability means that the minutes of a HEP collaboration meeting can be available to the entire collaboration as soon as the secretary makes them available to a WWW server. In addition, these minutes are maintained at a single location eliminating the problem of having multiple copies stored at different collaborator sites.

This client-server model can be naturally extended to allow documents to be dynamically created in response to a request from users. For example, the user's request can query a database and the result of that query returned as a hypertext document. The WWW interface to the SPIRES HEP preprints database is an implementation of this capability.

From the information user's perspective, all the documents on the Web are consistently presented as hypertext. If the information provider has done their job satisfactorily, the user will have access to a usable, consistent, reliable and accurate on-line resource. That user will not be required to know or understand any of the technical details surrounding the acquisition or even the location of the document provided.

3.2.2 HTTP A number of existing protocols and one new protocol are used in the WWW model. The HyperText Transport Protocol (HTTP) is a new protocol for file retrieval (not just hypertext) designed to operate as quickly as is needed in response to a request resulting from a hypertext link selection. HTTP is a very simple Internet protocol, similar in implementation to FTP or NNTP. A WWW client (browser) sends a document identifier with or without search words, and the WWW server responds with hypertext or plain text. The protocol runs over TCP, using typically one connection per document request. The client/browser renders the document as it is being received. The Multipurpose Internet Mail Extensions (MIME) standard is used to specify document types that the client supports, typically a variety of video, audio, and image formats in addition to plain text and HTML. Browser interpretation of these document types is known as "format negotiation."

3.2.3 UNIVERSAL RESOURCE LOCATORS (URLS) The power of the WWW model lies not in a complex protocol, but in the Universal Resource Locator (URL). [Note: some references define URL as Uniform Resource Locator or Unique Resource Locator.] The URL, designed to be compact, unambiguous, and printable, contains elements identifying the retrieval protocol to be used, the server to which the request is directed, and the document to be retrieved from that server. The URL is actually only one instance of a larger family of addressing mechanisms named URIs (Universal Resource Indicators). In general, the URI scheme addresses more specific issues such as guaranteeing access to documents in the event they are relocated.

The three elements are expressed in a URL in the following form:

<protocol>://<node>/<location>

The first element specifies the protocol to be used to access the document (e.g., HTTP, FTP, Gopher, NNTP, etc.) The second element specifies the network address of the server from which the document is to be obtained. The third element specifies a file path of the requested document on the WWW server. Therefore, the URL

http://www.slac.stanford.edu/home.html

defines a request for a file named "home.html" from a server with the Internet address "www.slac.stanford.edu." It specifies that this file should be retrieved using the HTTP protocol.

URLs may be transparently provided in a hypertext document using HTML markup or may be explicitly specified by a reader within a WWW browser. Expanded URL formats allow for non-specific document retrieval such as those resulting from the execution of a server-side command (e.g., common gateway interface scripts), a database query and for servers operating on non-standard port addresses.

Figure 2. illustrates how the URL addressing scheme, the HTTP protocol, and format negotiation server to integrate the client-server model used by WWW.

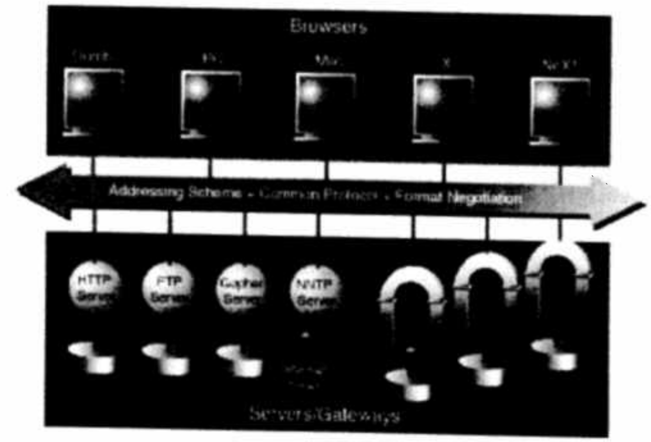


Figure 2. The World Wide Web Client-Server Model

4. EARLY HEP WWW APPLICATIONS

The World Wide Web did not immediately take the CERN physics community by storm. Hansjorg Klein of the Delphi collaboration recalls that a Hypertext Colloquium given by Tim Berners-Lee at CERN in November of 1990 was "overwhelming to many physicists." Early efforts to get Delphi experiment "on the Web" were not extremely successful. The adoption of WWW typically occurred (and still occurs) as a result of a particular need or application recognized by an experimental collaboration.

4.1 *SPIRES*

One of the first examples of a specific application of the World Wide Web to an HEP requirement was its ability to provide a readily-accessible, easy to use interface to the SPIRES databases at 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 HEP including experiments, institutes, publications, and particle data.

The largest of the SPIRES databases is the HEP preprints database, containing over 335,000 entries. Prior to its implementation on WWW, the only way to access the SPIRES databases 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, in order to successfully access information in the database, a rudimentary knowledge of the somewhat esoteric SPIRES query language was required.

The Web, coupled with the widespread adoption of Bulletin Boards as the primary means of distributing computer-readable versions of HEP preprints, has revolutionized the ease of access and usefulness of the information in the SPIRES databases (6).

The SPIRES WWW server was one of the very first WWW servers established outside CERN and certainly one of the first to illustrate the power of interfacing WWW to an existing database, a task greatly simplified by WWW's distributed client-server design. Using this interface it is now possible for researchers to locate papers within the database without any knowledge of the SPIRES query language, by using simple fill-out forms (Figure 3.). Experienced SPIRES users are still able to use the SPIRES query language through this interface.

Access to some of the more advanced features of SPIRES, such as obtaining citation indexes and lists of top-cited papers, are also available via hypertext links. Since the access to the database uses WWW it can be used remotely from site on the Internet.

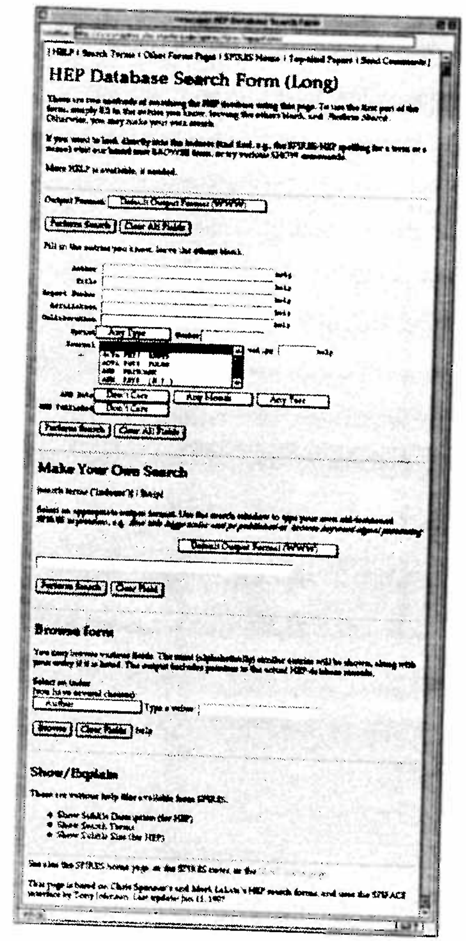


Figure 3. HEP Database Search Form

By linking the entries in the SPIRES databases to the computer-readable papers submitted to electronic Bulletin Boards at Los Alamos and elsewhere, it is also possible to follow hypertext links from the database search results and access either the abstract of a particular paper, or the full text of the paper, which can be viewed on-line and/or sent to a networked printer. A typical day may have over 10,000 accesses to the HEP preprints database from hundreds of institutions.

The WWW interface to SPIRES has now been extended to cover other databases including experiments in HEP, conferences, software, institutions, and information from the Lawrence Berkeley Particle Data Group.

4.2 *Home Pages*

Many HEP experiments and laboratories quickly realized how Web documents could be used to describe both their mission and results. Home pages and welcome pages for these laboratories and experiments were usually provided their Web authoring experience. Experiment "working pages" provide detailed information about an experiment to insure that collaborators are current in matters of data collection, analysis and software changes. Collaboration archives on the Web provide ready access to agendas, minutes, and schedules.

Home pages also became an integral part of a common computing environment. The ability of WWW to integrate with existing on-line systems permitted, for the first time, a common implementation-independent interface. Bjorn Nilsson recalls that in June 1993, ALEPH was able to generally provide information that was previously only available on-line on ALWS. A Web version of ALWHO provided accurate phone list and e-mail data. ALWS HELP and detector status were also provided.

The SLD collaboration incorporated its data monitoring system into a Web application available from its home page (Figure 4). This facility uses WWW forms to provide interactive access to databases containing up-to-date information on the performance of the detector and the event filtering and reconstruction software. Information can be extracted from the databases and used to produce plots of relevant data as well as displays of reconstructed events. Using these tools collaborators at remote institutes can be directly involved in monitoring the performance of the experiment on a day-by-day basis.

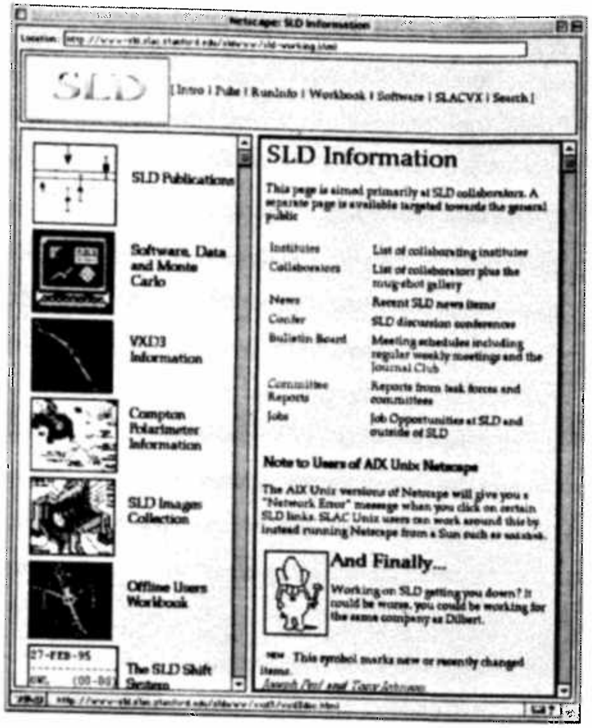


Figure 4. SLD Home Page

Major physics organizations such as the American Institute of Physics (AIP) and The American Physical Society (APS) also quickly recognized the potential of WWW as a communications medium. The APS first established a Website in January, 1994. By November, 1994 a second Web server was necessary to provide the desired services. The AIP debuted on WWW in April, 1994. Visitors to these organizational Websites (Figure 5.) have access to a wide variety of information - meetings, publications (including research journals), career services, fellowship opportunities, etc.

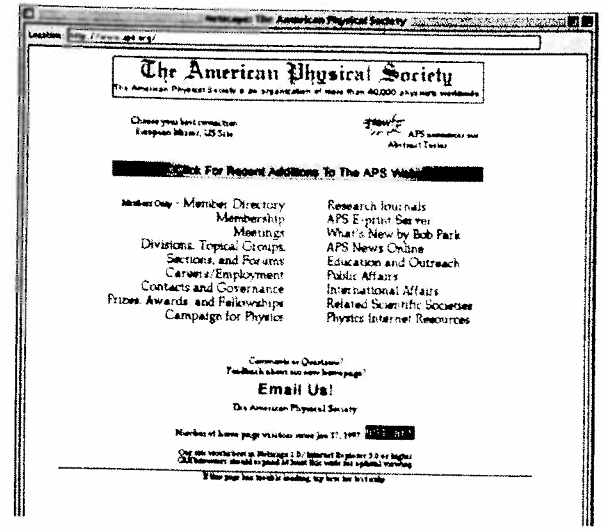


Figure 5. American Physical Society (APS) Home Page

The Particle Data Group (PDG) is an international collaboration that reviews Particle Physics and compiles and analyzes data on particle properties. PDG products are distributed to an estimated 30,000 physicists, teachers, and other interested parties. The Review of Particle Physics is one of the often most cited publications in HEP. The PDG home page, hosted by Lawrence Berkeley Laboratory (LBL), has made this distribution more accessible using WWW.

Institution and experiment resources on WWW (and other Internet services) have made research in high energy physics more accessible to the general public. Many HEP Websites now specifically contain information and resources targeted at non-HEP audiences such as the educational community (Figure 6.) and the press. A future commitment to provide such resources could play a major role in the design of Internet-based physics curricula.

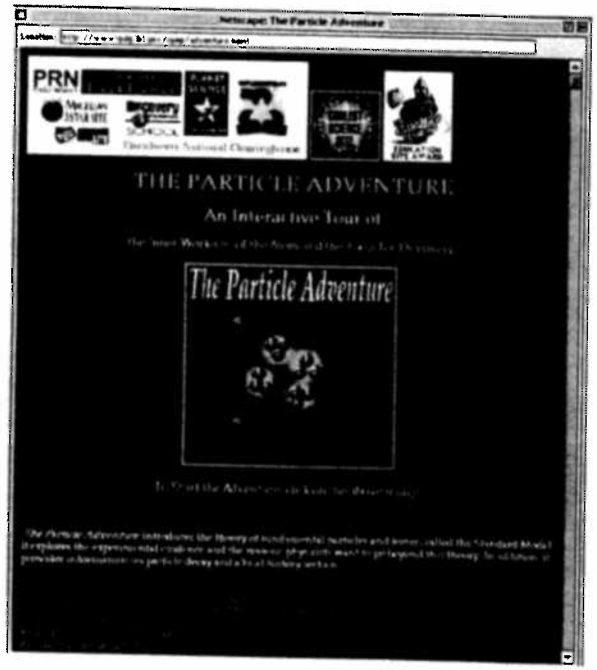


Figure 6. The Particle Adventure

Web pages provide an excellent mechanism for exceptional public announcements. Just hours after Prof. Martin Perl of SLAC was awarded the 1995 Nobel Prize, Web pages containing background information on the tau discovery, and Perl's first post-Nobel news conference (including photographs taken with a digital camera) were available to the world via WWW. Web pages on the Fermilab server chronicle the discovery of the top quark and the generation of anti-hydrogen (Figure 7.).

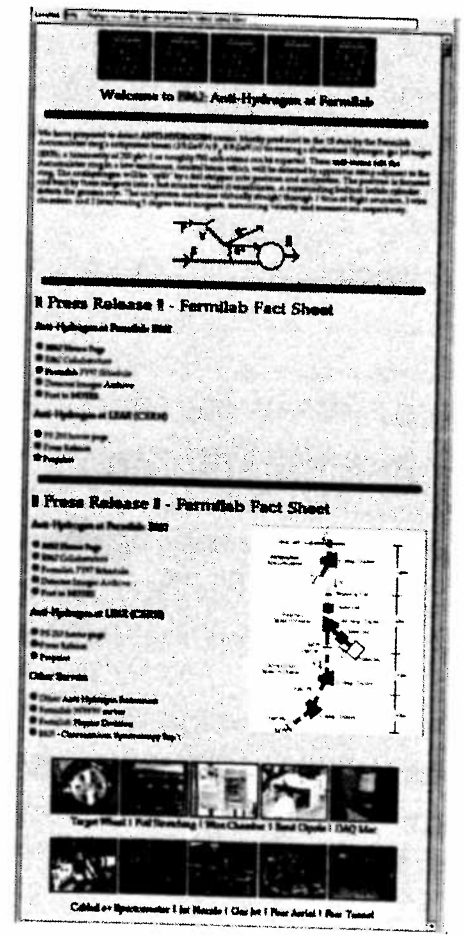


Figure 7. Welcome to E862 - Anti-Hydrogen at Fermilab

4.3 *Software Distribution*

WWW has played an important role in the success that HEP 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 and associated databases to remote sites. A decade ago, HEP researchers relied extensively on community wide software tools in addition to software developed for an individual experiment or research project. Today these same researchers are making increasing use of software developed outside of the HEP community including CAD/CAM systems, visualization software, symbolic mathematics packages, databases, software engineering packages, software for coarse and fine grained parallelism, C++ class libraries and many others. The Web provides a mechanism for keeping track of important software developments around the world in the many relevant fields or even within HEP itself.

FreeHEP consists of a collection of software and information about software which is useful in HEP and related fields (7). It was first proposed at the 1991 HEPLIB meeting to address the distribution of HEP and non-HEP software within the high energy physics community. FreeHEP consists of:

1. a global compilation of software useful in HEP from within HEP, other fields, and commercially;
2. tutorials on common subjects and reviews on the subject areas covered;
3. FTP and WWW access to information about software packages, documentation, source code, tutorials, and reviews;
4. a mechanism for authors of software to distribute their packages;
5. a mechanism for users to communicate with software authors and/or with other users.

When first implemented, the FreeHEP compilation was available primarily via FTP. The compilation is now also imported into the SPIRES system which provides software database operations via hypertext.

5. CURRENT HEP WWW APPLICATIONS

5.1 *Conferencing Systems*

HyperNews is a Web-based application developed by the National Center for Supercomputing Applications (NCSA) that allows large groups of people to participate in electronic conferencing. It allows a user to read and contribute commentary, and supports various types of notification as activity in the area of discussion progresses. Each HyperNews conference begins with a "base article," which is followed by a nested list of postings by conference participants (8).

The BaBar experiment has adopted HyperNews as a mechanism for intra-collaboration communication. Figure 8. illustrates the wide variety of "conferences" in which BaBar collaborators are able to participate.

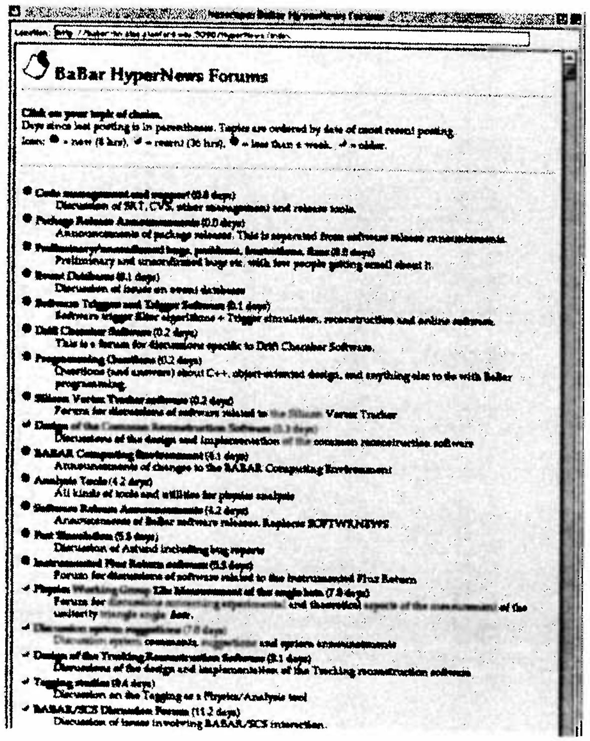


Figure 8. BaBar HyperNews Forums

While HyperNews is primarily text-based, the postings can contain multimedia elements. However, experimental Web-based conferencing systems available in the near future will support real-time audio and video, shared white boards, and other features now found in videoconferencing systems.

5.2 *Virtual Reality*

One of the most exciting research areas associated with WWW is the development of the Virtual Reality Modeling Language (VRML). The first discussions of incorporating virtual reality into the Web began at the First International WWW Conference at CERN in May, 1994. The VRML concept, like that of HTML, is to specify a very simple language capable of defining virtual reality scenarios. VRML documents can be translated into viewable objects on almost any computing platform, from small personal computers to very high-end graphic workstations. While the amount of graphical detail may vary between this range of platforms, the essential elements of the virtual world will be the same (5).

In 1994 the VENUS (Virtual Environment Navigation in the Underground Sites) project at CERN began an effort to develop a high performance 3D Web browser, particularly equipped to deal with large territorial and engineering models. This tool was named i3D and is intended to be used for LHC Engineering Design Integration. i3D will allow the entire HEP community to access and evaluate LHC's virtual prototypes on the Web (9, 16).

Components of the LHC detectors will be designed at many institutions located around the world. These conditions make design validation and integration an extremely difficult task. By its ability to load geometries via hyperlinks, i3D provides an ideal integration stage. Its data models, since i3D uses the WWW client-server model, can be maintained at the sites where they originate. Each object can be loaded on-line from wherever it is being designed, and a virtual world created in a users VRML-enabled browser.

The advantages of using i3D for LHC design integration are enormous. By using a database server, it is possible to manage centrally the hyperlinks and virtual worlds structure, in order to keep control of the overall project. For instance, the experiment's management controls naming of the components and associates URL's linking to geometry, drawings, home pages, etc. i3D then queries this database to obtain the appropriate information whenever a user selects a graphic hyperlink or "hyperdoor." Therefore navigating through an LHC detector with i3D allows not only an overall view (Figure 9.), but also the ability to obtain technical or organizational information on each component.

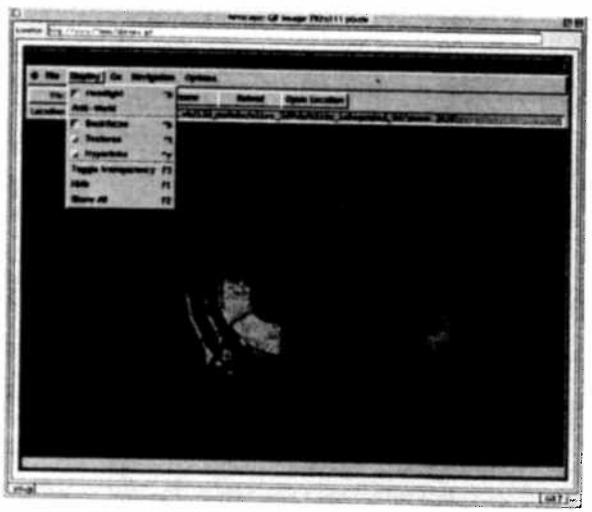


Figure 9. An i3D View of an LHC Detector

5.3 *Software Development and Maintenance*

Programmers who develop, use, maintain, and modify software are faced with the problem of scanning and understanding large quantities of documents, ranging from source code to requirements, analysis and design diagrams, user and reference manuals, etc. Their task is non-trivial and time consuming, because of the number and size of supporting documents, and the many implicit cross-references contained. In large distributed development teams, such as HEP collaborations where software and related documents are produced at various sites, the problem can be extreme. The fundamental concept of the LIGHT (LIfe cycle Global HyperText) system, developed by the Programming Techniques Group, ECP Division at CERN and the ALEPH experiment, is to provide access to all documentation associated with a software application or system, including source, via WWW with all cross-references automatically established (10). For example, links in a listing of a subroutine call written in FORTRAN refer to the subroutine's documentation; a link in a data element leads to the corresponding data definition, etc.

This concept and the LIGHT system has been applied to the JULIA reconstruction program of the ALEPH experiment at CERN. With JULIA/LIGHT, ALEPH programmers, documentation writers, maintainers of the data model and end user physicists are able to view through the Web the entire FORTRAN source code, data definition, and data design diagrams, as well as the JULIA, ALEPHLIB, CERNLIB, ADAMO, and BOS manuals (Figure 10.). All these documents are connected with hypertext links making it possible to inspect them in the widest variety of ways. The final production version of JULIA/LIGHT is expected to consist of approximately 6,000 HTML pages with 150,000 hypertext links.

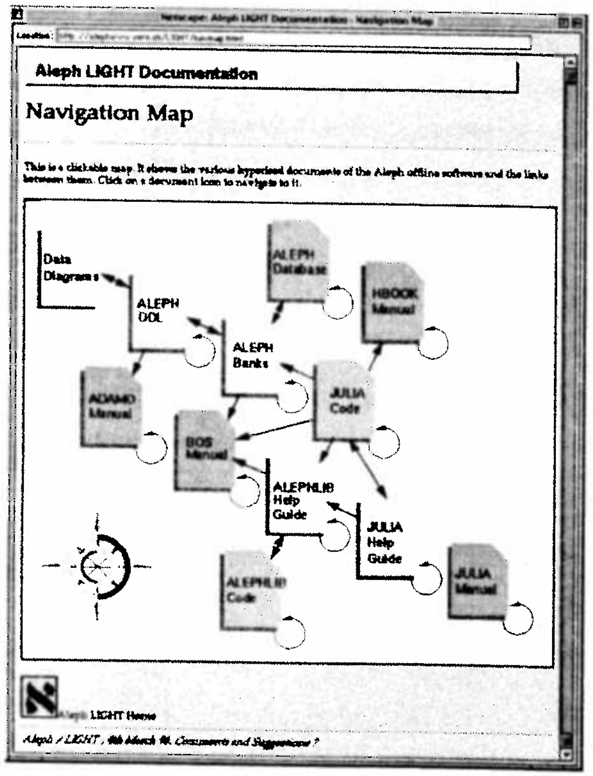


Figure 10. ALEPH LIGHT Documentation - Navigation Map

6. FUTURE HEP WWW APPLICATIONS

6.1 *Electronic Journals*

WWW has long been recognized as an effective medium for the distribution of documentation. In the near future it is likely to provide the basis for electronic journals in all disciplines not just HEP. The American Physical Society (APS) has a vision of the future of physics publishing, circa 2020. According to Burton Richter, former APS President and Director of the Stanford Linear Accelerator Center (SLAC), "any physicist, any place in the country, can turn on his computer and for free browse through the table of contents of any APS journal." He adds that this browser "can select those things about which he wants to see an abstract, and then, after deciding what he might read, ask for the article itself and eventually pay for it like you pay your telephone bill" (11).

At first glance, Web-based publishing of scientific journals should be a relatively simple matter. HTML provides most of the required typesetting capability. Those capabilities absent (e.g., equation layout) can generally be accomplished with use of graphical elements.

The lack of a widespread availability of journals on the Web appears to be due to the fact that most parties involved are uncertain as to what "electronic publishing" actually means. Authors are primarily interested in electronic manuscript preparation and submission. Readers are interested in facilities provided for document/journal retrieval. Librarians focus on the delivery of information to users, but they often overlook archiving. Publishers concern themselves with the handling of electronic manuscripts, copyediting, formatting, typography and the production of versions suitable for delivery to the end users. They also worry about collecting the revenues needed to keep their operations financially viable. Authors and publishers are both concerned about the resolution of copyright and intellectual property rights associated with Web-published content (12).

Until technology advancements are able to address these issues, Web-based distribution of journals is unlikely to become commonplace. In the meantime, Web-based publishing and archiving of conference proceedings may provide a prototype. Figure 11. illustrates how the proceedings of the 1995 CHEP conference have been made available to Web users.

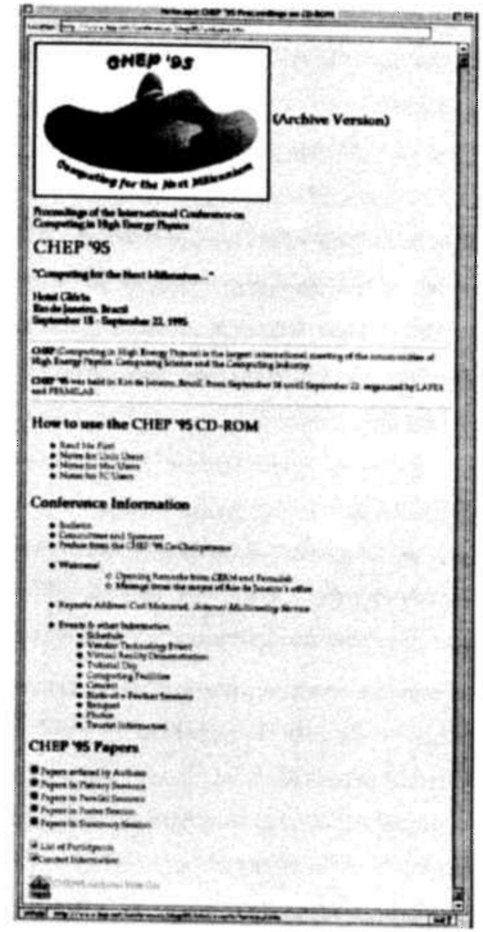


Figure 11. CHEP95 Proceedings

6.2 *Active Objects*

In addition to the virtual worlds demonstrated by the i3D LHC application, future Web-based HEP applications are likely to be enhanced by the inclusion of active objects. This technology, already demonstrated in a number of prototypes, will enable active objects such as spreadsheets or data plots to be embedded into objects such as HTML or VRML documents. While older browsers might only display these objects statically, the newer browsers will allow a user to interact with the object, perhaps by rotating a three dimensional plot, or expanding and rebinning a particular area of a data plot.

A current effort to provide similar functionality is EDWIN (Event Display WWW INterface) developed by DELPHI (13). EDWIN was designed to provide a generic method of interfacing event displays to the Web. Using an HTML form, a physicist can request a certain event and view (angle, zoom factor). The server returns the requested view as a bitmapped graphic. The requester iterates by resubmitting the form until satisfactory results are achieved. The EDWIN model is effective, but is very slow. Active object technologies will provide similar data interactivity but with faster response times.

6.3 *Collaboratories*

Scientific collaborations such as those in HEP currently rely heavily on face-to-face interactions, group meetings, individual action, and hands-on experimentation. Group size and attendance can vary widely. Technologies such as videoconferencing have begun to address some of these issues. A future solution may be found in "collaboratories."

The concept of a "collaboratory" came from William Wulf while working at the US National Science Foundation (NSF). Wulf merged the words "collaboration" and "laboratory" and defined a collaboratory as a ' ...center without walls, in which the nation's researchers can perform their research without regard to geographical location - interacting with colleagues, accessing instrumentation, sharing data and computational resource, and accessing information in digital libraries' (14)

A collaboratory facilitates scientific interaction within a group by creating a new, artificial environment in which group members can interact. This new environment must be socially acceptable to those who participate and improve their ability to work. Many computing tools must be brought together and integrated to allow seamless interaction. Some of these tools are already in wide use, such as WWW and e-mail, while others, like telepresence - the immersive electronic simulation of "being there" - are still being developed.

To facilitate scientific work, collaboratory systems must support the sharing of secure data, analysis, instruments, and interaction spaces. Videoconferencing improves communication and helps reduce the likelihood of major misunderstandings between collaboration members. Drawings, budgets, schedules and status reports can be shared in a collaboratory system insuring that engineering and manufacturing efforts are to specification, within budget and on schedule.

Because it is impractical to have all the critical personnel permanently located at the experiment site, collaboratory technology could even be used to decentralize the traditional central control room. Control of most HEP detectors is now done by networked computers, and permitting control over the Web from remote sites is a logical and easy extension of this trend.

The US Department of Energy has initiated a series of five major testbed projects known as the Distributed Collaboratory Experiments Environments (DCEE) Program. One of these testbeds is the Advanced Light Source (ALS) collaboratory at Lawrence Berkeley Laboratory. The model for this collaboratory is a loosely integrated set of Internet capabilities that appear as extensions to the Web. The ALS collaboratory has established a set of requirements that are similar to those of the other virtual laboratory DCEE projects. The requirements define a wide range of cross-platform functions that would allow researchers to interact with remote colleagues in a rich, in-process style. These requirements include:

* audio/video conferencing;
* chatting;
* shared computer display and whiteboard;
* shared on-line electronic notebook;
* file sharing with access security, safety, and data confidentiality;
* on-line instruments, remote experiment monitoring, computation and visualization;
* Web browser synchronization.

It is easy to recognize how the implementation of a collaboratory integrates many of the Web technologies discussed earlier in this chapter. The collaborative working environment can be presented as a virtual world. Web-based library technologies provide prototypes for electronic notebooks. Even the efforts supporting commerce on the Web are likely to make valuable contributions in areas such as security and privacy. Developing Web technologies such as Java will likely play a major role in the implementation of collaboratory requirements.

7. CONCLUSIONS

The World Wide Web was originally developed to address the needs of collaborative efforts in high energy physics. Since that time, the Web has undergone many changes, become a major force in the growth of the Internet, and has been used for dissemination of every conceivable form of information. Its widespread acceptance and modifiability are proof that its underlying design based on hypertext and client-server technology was a sound one.

The high energy physics community has adopted WWW as a critical component of its research effort. Web-based applications now support HEP in ways which the original WWW designers could never have imagined. Ongoing HEP developments using the Web closely parallel advances being made in overall WWW and Internet technologies. It is clear that WWW will continue to shape the way that HEP does physics and that HEP will continue to shape the development of WWW.

ACKNOWLEDGMENTS

I wish to thank Robert Cailliau, Hansjorg Klein, Paolo Palazzi, and Bjorn Nilsson for helpful discussions. Kathryn Henniss and Terry Anderson were invaluable in helping me to prepare this manuscript. This work was supported by the US Department of Energy Contract No. DE-AC03-76F00515.

Literature Cited

1. Verkerk, C, Wojcik, W. (edit), *Proc. Int. Conf. on Computing in High Energy Physics, Annecy, France, 1992*
2. Berners-Lee T, Cailliau R., *Proc. Software Engineering, Artificial Intelligence, and Expert Systems for High Energy and Nuclear Physics, La Londe-les-Maures, France, 1992*
3. Segal, B. *A Short History of Internet Protocols at CERN*, CERN PDP-NS, April 1995
4. Nelson, T. *Literary Machines*. Sausalito: Mindful Press (1992)
5. White, B. *HTML and the Art of Authoring for the World Wide Web*. Boston: Kluwer Academic Publishing (1996)
6. Johnson, T. 1994. *SLAC Beamline*. 24 (3):2-9
7. Johnson, T et al., [WWW document]. URL http://heplibw3.slac.stanford.edu/FIND/FHMAIN.HTML
8. LaLiberte, D et al., [WWW document]. URL http://union.ncsa.uiuc.edu/HyperNews/get/hypernews.html
9. de Gennaro, S. *Proc. Int. Conf. on Computing in High Energy Physics, Rio de Janeiro, Brazil, 1995* (1996)

10. Aimar, A et al., *Proc. Int. Conf. on Computing in High Energy Physics, Rio de Janeiro, Brazil, 1995* (1996)

11. Taubes, G. 1993. *Science*. 259:1246

12. Boyce P, Dalterio H. 1996. *Physics Today*. 49(1):42

13. Dönszelmann, M et al., *Proc. Int. Conf. on Computing in High Energy Physics, Rio de Janeiro, Brazil, 1995* (1996)

14. Kouzes, R, Myers, J, Wulf, W. 1996. Computer. 29(8):40-46

15. Agarwal, D et al.,*Proc. Int. Conf. on Computing in High Energy Physics, Rio de Janeiro, Brazil, 1995* (1996)

16. Balaguer, J-F. *Proc. 3rd Eurographics Workshop on Virtual Environments, Monte Carlo, 1996* (1996)

THE WORLD WIDE WEB AND HIGH-ENERGY PHYSICS

The Web—what you may now be using to buy an airline ticket or look up Miles Davis's discography—began life as a tool for improving the flow of information at CERN.

Bebo White

In his 1998 State of the Union address, President Bill Clinton told 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.

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

The nursery

CERN, located on the Franco–Swiss border near Geneva, is one of the most successful high-energy laboratories in the world. Its success can be measured in terms of both the number of Nobel prizes its staff have garnered and its proven record of carrying out huge projects. Experiments at CERN and other high-energy physics laboratories use complex particle detectors, which cost millions of dollars and typically involve hundreds of physicists from institutions spread around the world.

Two major elements of successful collaboration are communication and documentation. Researchers working together in different countries must have easy access not only to experimental data, computer program code and drafts of scientific papers, but also to more mundane administrative information, such as meeting minutes and agendas and telephone and e-mail lists. In general, computing and networking technologies provide useful tools for collaborations, but the implementation of specific technologies can also create roadblocks, since collaborating institutions tend to use different computer operating systems and document preparation systems. It was in this environment and to address these kinds of problems that the Web was conceived at CERN.

BEBO WHITE *is a member of the research staff and the Webmaster at the Stanford Linear Accelerator Center in Stanford, California.*

The early years

The development of the Web at CERN was actually the convergence and synthesis of three separate technologies—computer networking, document/information management and software user interface design. Applying these technologies to the needs of the experimental collaborations led to the solution that became the Web. However, as described below, the Web was not immediately recognized as the solution. CERN's Ben Segal remembers:

> 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 1970s, 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 manufacturers' 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.[1]

Network experimentation at CERN continued through the late 1970s and the 1980s. CERNET, which was started in 1976, linked a number of mainframes and minicomputers. From 1981 to 1983, the STELLA Satellite Communication Project used a satellite channel to link CERNET, the Istituto Nazionale di Fisica Nucleare in Pisa, Italy, and the Rutherford–Appleton Laboratory in Didcot, England. Despite the introduction of Ethernet to the CERN community in 1983, most mainframe and minicomputers still used proprietary networking systems. And although the chaos noted by Segal had been somewhat alleviated, an overall networking strategy was still missing.

While CERN was busy seeking solutions to its networking problems, the global phenomenon known as the Internet was maturing in the US. At the time, no one foresaw that these two efforts would collide and revolutionize the world of communications.

The birth of the Internet

Work on the Internet was initiated in the late 1960s by the US Department of Defense, which wanted to develop a communications network that could survive a nuclear war. In particular, messages had to be able to reach their destinations, even if significant communication routes were destroyed. The solution, the "packet switching"

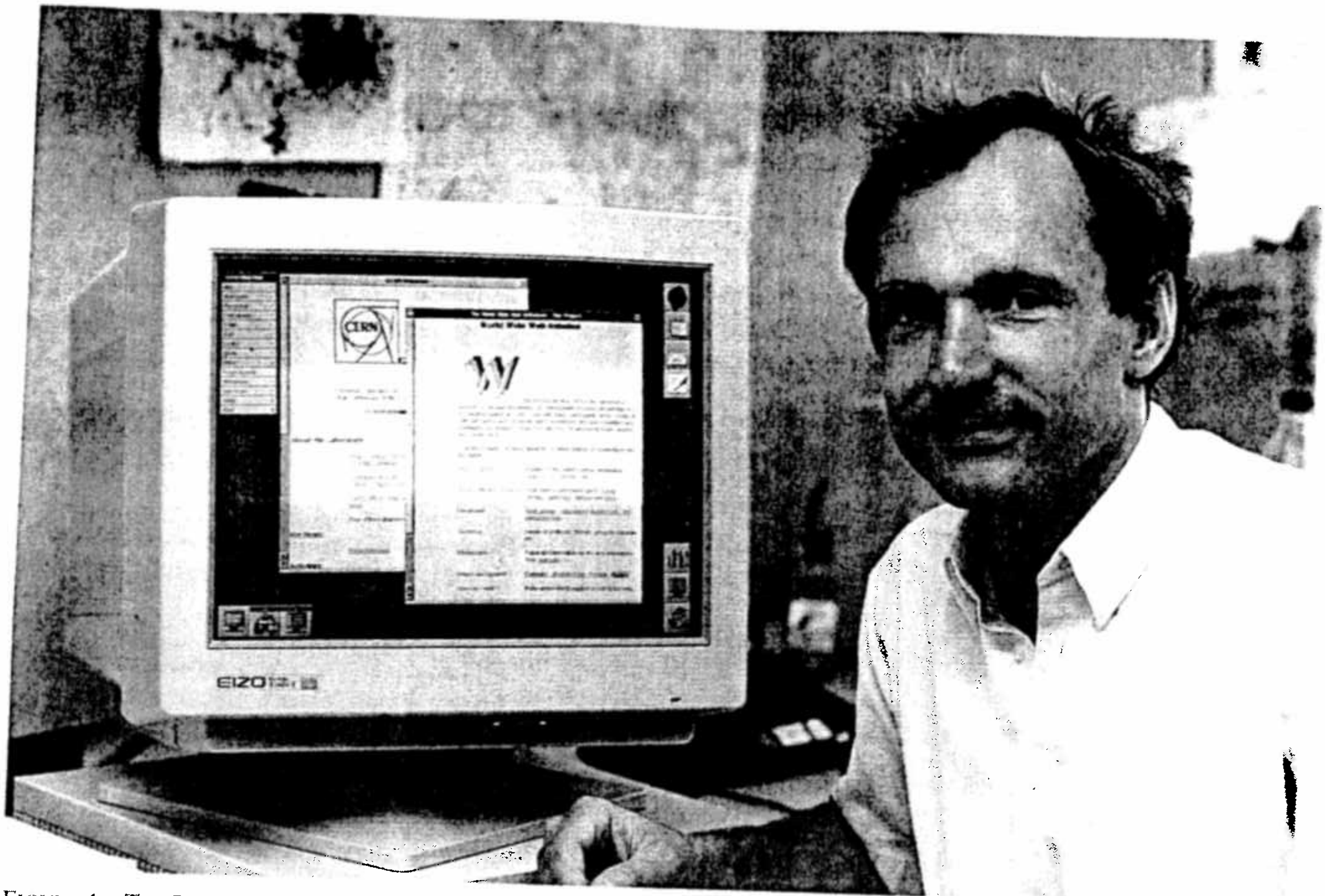


FIGURE 1. TIM BERNERS-LEE, the Oxford-educated physicist-turned-CERN programmer who invented the World Wide Web.

network, was designed so that the particular route a message takes to reach its destination is unimportant. This ability of a message to travel by other routes assures the sender that it will eventually be received. The original Internet was named Arpanet, after its sponsor, the Advanced Research Projects Agency.

Arpanet led to the development of numerous important networking technologies. Among them was TCP/IP (Transmission Control Protocol/Internet Protocol), which is a method for converting messages to data streams, directing them to a particular address on the network and reassembling the message at the destination. Arpanet (and later the Internet) was designed to be a network of networks that contain different types of computers and different networking technologies. To be able to communicate, all the computers and networks simply had to "speak" TCP/IP.

Arpanet's first "killer application" (a software program whose popularity surpasses its competitors to the point of all but killing them off) was e-mail, thanks to which Arpanet's potential was quickly realized by many people outside the US Defense agencies.

Formally decommissioned in 1989, Arpanet became, more generically, the Internet. Its users scarcely noticed, for Arpanet's functions not only continued, but also steadily improved. The use of TCP/IP standards for computer networking is now global.

Between 1985 and 1988, CERN local networks converged on 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 account entitled "A Short History of Internet Protocols at CERN," Segal writes:

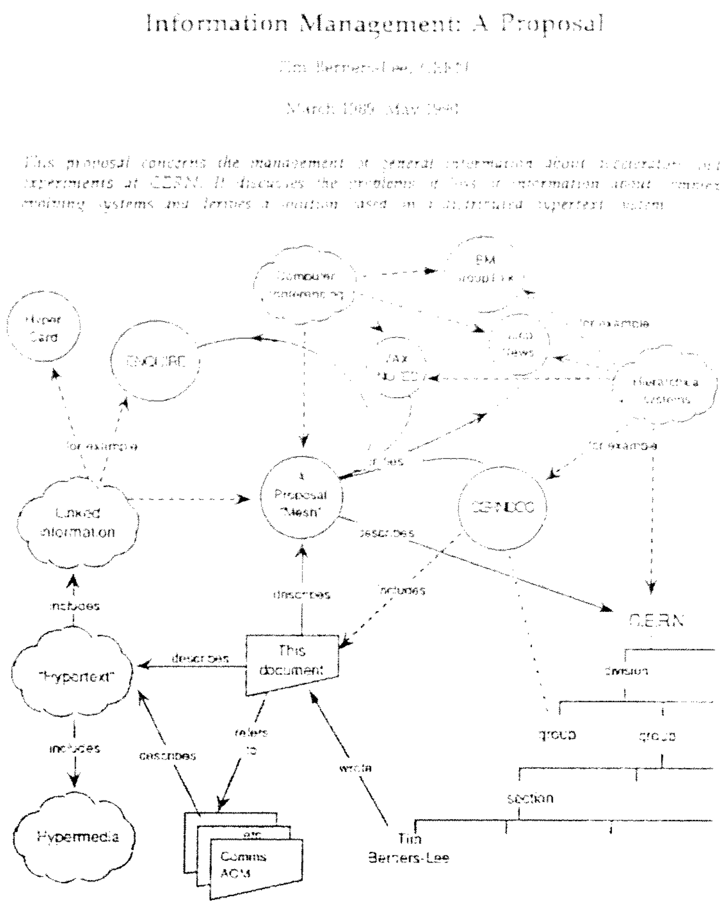
> A key result of all these happenings was that by 1989 CERN's Internet facility was ready to become the medium within which Tim Berners-Lee would create the World-Wide Web with a truly visionary idea. In fact an entire culture had developed at CERN around 'distributed computing. . . .' It is my belief that the Web could have emerged considerably earlier if CERN had been connected earlier to the Internet.[1]

The networking experience of CERN provided a foundation for the development of the Web. Missing, however, was a way to describe information that could easily be delivered to the user community over the nascent Internet.

Hyperdocumentation

Back 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 and computing documents. At that time, document preparation at CERN was centered on the use of Waterloo Script on the IBM central mainframe computer. In 1985, some groups began experimenting with SGML (the Standard Generalized Markup Language). The first prototype of this documentation system was introduced on the IBM mainframe and named CERNDOC. It was one of the first electronic documentation systems to use a client–server model. As Eric van Herwijnen (one of the architects of CERNDOC) told me:

> Tim [Berners-Lee] was very interested in CERN-



[illegible] 2. Tim Berners-Lee's [illegible]

[illegible] not [illegible] computer [illegible] personal [illegible] realities. It was [illegible] HyperCard [illegible] by Apple for [illegible] Macintosh. A difference was that [illegible] running the [illegible] multiuser system, [illegible] many people to access the same data.

In his proposal, Berners-Lee addressed the management of general information about accelerators and experiments at CERN and the problems of inconvenient access to current and archived information. The solution, as he described it, was "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 2 shows a diagram from the proposal, which illustrates how the proposed system would integrate such diverse information sources as VAXNOTES, USENET, IBM GroupTalk (which I implemented at CERN) and CERNDOC.

Independently of Berners-Lee, Robert Cailliau was also looking at hypertext as a solution to CERN's information retrieval problems. He was especially interested in platform-dependent hypertext systems such as HyperCard and in hypertext scripting languages such as HyperTalk.

DOC 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.

With the Internet providing the transfer mechanism and CERNDOC providing a prototype for content, the Web lacked only a 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 from the users' point of view, both systems required knowledge of a specific set of commands. The solution to the interface problem was found in the pointing and linking features of hypertext.

Proposing the Web

Tim Berners-Lee (see figure 1) was a programmer working on networking in CERN's computer division. In a 1989 paper entitled "Information Management: A Proposal," he recalled:

> 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

In 1990, Berners-Lee and Cailliau decided to collaborate on a revision of Berners-Lee's 1989 proposal to CERN management. Cailliau recalls his 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."

The new proposal was entitled "World-Wide Web: Proposal for a Hypertext Project." Later never came.

In October 1990, the two programmers submitted the revised proposal, whose key goals were stated as being

> 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.

With the Berners-Lee–Cailliau proposal, which got the go-ahead in September 1990, all the pieces contributing to the development of the Web at CERN were complete. However, implementing the described model was not simple. It required the development of server software (to provide content) and client software (to receive and interpret content). In addition, the method that allows the client and server to communicate with one another—that is, the Hypertext Transport Protocol (HTTP)—had to be invented, as did the Hypertext Markup Language (HTML), which enables content to be displayed effectively on a client (or browser). HTML is based on SGML, revealing again the Web's CERNDOC roots.

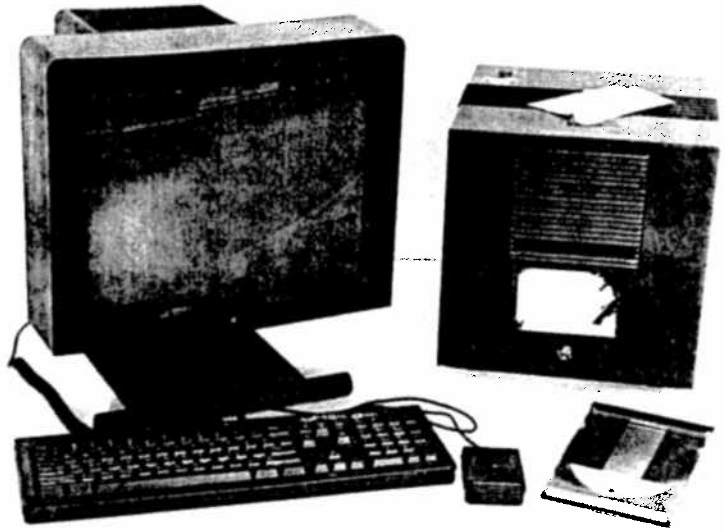


FIGURE 3. THE ACTUAL NeXT COMPUTER that ran the original World Wide Web server and browser at CERN.

By November 1990, the initial implementation of a Web client for the NeXT computer (see figure 3) had been developed. This first client was able to display documents using multiple fonts and styles and could even be used to edit documents (capabilities only now becoming widespread in many browsers), but access was limited to NeXT users.

March 1991 brought the development of the CERN "linemode" browser, which is shown in figure 4. It could run on numerous platforms, but displayed output only on character-based terminals, such as the VT100. Soon afterward, the Web server software was successfully installed on the CERN VAX/VMS platforms and the central IBM mainframe.

In December 1991, the *CERN Computer Newsletter* first introduced the Web to the high-energy physics community. At that point, Cailliau estimates, there were 10 Web servers worldwide.

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 1990 before the Web's formal announcement was overwhelming to many physicists, who viewed it as an esoteric computer science exercise and did not immediately recognize it as an improvement over existing methods for helping physicists to do physics. Early efforts to get the Delphi experiment on the Web were not successful. 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 [Berners-Lee] 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.

And Paul Kunz, a mainstay of the high-energy physics computing community, who was instrumental in establishing the first US Web site, at the Stanford Linear

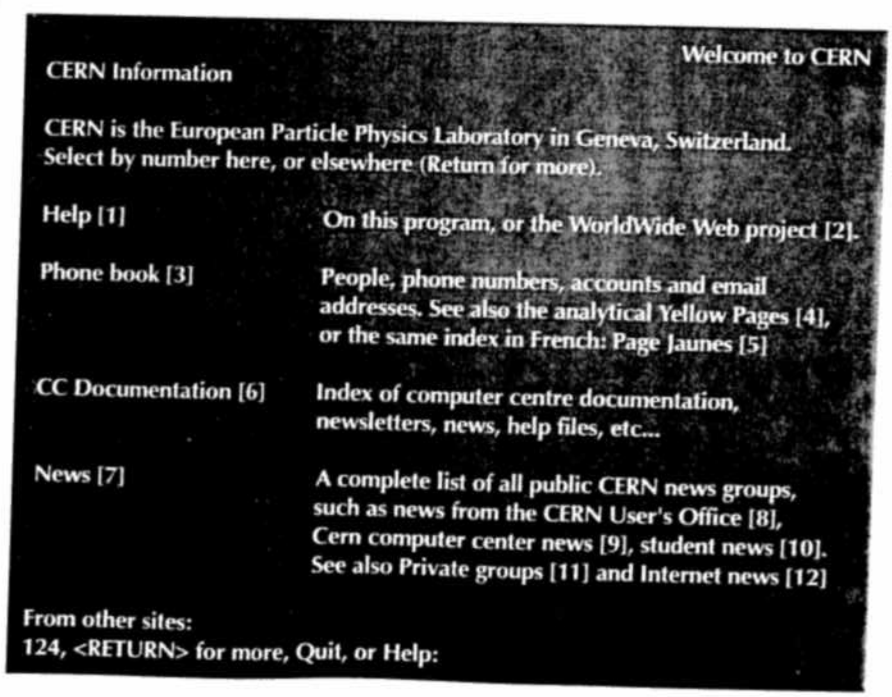


FIGURE 4. A SAMPLE SCREEN from the first World Wide Web browser.

FIGURE 5. THE FIRST WORLD WIDE WEB T-SHIRT. Dating from 1992, it bears the names of the five high-energy physics laboratories on the Web at the time—the Stanford Linear Accelerator Center (SLAC), the National Institute for Nuclear Physics and High-Energy Physics (NIKHEF) in Amsterdam, the German Electron Synchrotron Laboratory (DESY) in Hamburg, Fermilab (FNAL) and CERN. The designers of the shirt joked: "Maybe some day we'll have enough sites to go around the wearer's chest."

FIGURE 6. TONY JOHNSON and the Midas browser that he devised.

Accelerator Center (SLAC), recalls:

> The first I heard about the WWW was an article in a comp.sys.next newsgroup. Tim announced to NeXTStep (the NeXT operating system) 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?

Others, however, were quick to see the Web's potential. In early 1992, Ruth Pordes and Jonathan Streets of the Fermilab computing division's on-line systems department (OLS) were considering the problem of providing high-energy physics experimenters with information about on-line data acquisition systems. Seeing the Web presentation to the Artificial Intelligence in High Energy Physics conference (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.[4] 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 (CHEP92) held at Annecy, France, in September 1992, Terry Schalk of the Santa Cruz Institute for Particle Physics told the attendees that the conference highlight for him had been the short presentation about the Web. (The commemorative t-shirt is shown in figure 5.) Ever 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, California.)

According to Cailliau, the number of Web servers worldwide at the end of 1992 was approximately 50.

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

The first browsers

The Web is often described as the killer application of the Internet. However, in those early days, it needed its own killer app 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 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 da-

The Future Impact of the Web on the Physics Community

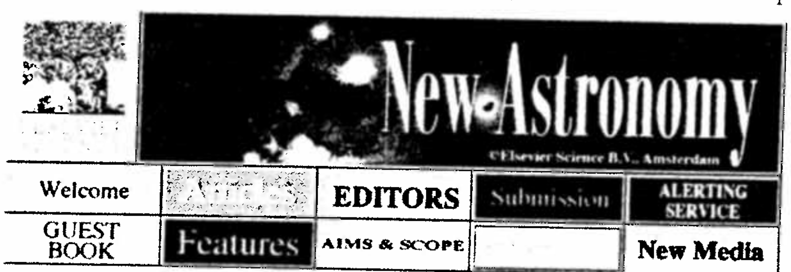
The usefulness of the World Wide Web to the high-energy physics community is no longer questioned, and the operation of a major experiment or laboratory without the Web is almost unimaginable. Not just a tool, the Web has also made research in high-energy physics more accessible to the general public through offering pages written for the educational community and the press. And the Web provides institutions with an excellent mechanism for making exceptional public announcements about such critical events as Nobel Prizes and important discoveries.

Although, for many, the Web has become an advertising and commercial medium whose credibility should not be taken for granted, the high-energy physics and other scientific communities continue to use it in many of the same ways that prompted its conception. There are, however, at least three major areas in the scientific community that could be dramatically changed by some of the new Web technologies—namely, advanced forms of software management, collaborative tools and electronic publishing.

Software management

High-energy physics experiments depend on computer software that is written, maintained and run by large numbers of people. The Web has played an important role, helping experimental groups to manage central databases of program source code while allowing programmers at different locations to develop the source code. It has also significantly eased 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 and the CERN LIGHT (LIfe cycle Global HyperText) system.

But in the near future, Web-based software systems will be quite different and exciting. For example, think of a physicist writing platform-independent analysis software in Java on a local personal computer. The software would be compiled, tested and debuggged locally, but 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 data sets 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. Where local program and data libraries are still necessary, Web "push" technology could provide version control by installing the latest software fixes. Current, or especially significant data sets could also be pushed to experiment collaborators for analysis.



Now three years old, the journal "New Astronomy" was launched simultaneously in Web and print versions.

Collaborative tools

Physicists must collaborate. The first Web-based conferencing system, WIT, was developed at CERN in 1994 as a quick hack. The BaBar experiment makes extensive use of HyperNews, a conferencing system developed by the National Center for Supercomputing Applications. Various other commercial and home-grown tools are also being used in other high-energy physics collaborations.

But the current text-based conferencing systems—and even videoconferencing—provide only tantalizing pieces of the collaborative systems of the near future. DOE2000 is a new initiative[7] that will fundamentally change the way in which 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," which promotes itself as 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."[8]

Electronic publishing

In a PHYSICS TODAY article (January 1996, page 42), Peter Boyce and Heather Dalterio analyzed the electronic publishing of scientific journals and concluded that, ultimately, "it 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." However, their enthusiasm was tempered by their concern that Web-based electronic publishing suffered from at least three major shortcomings: (1) Hypertext Markup Language (HTML) is not sufficiently structured or versatile for scientific literature; (2) papers are useless unless they can be promptly found; (3) electronic journals must collect enough revenue to support the infrastructure, but not in a way that inhibits the use of electronic tools.

Recent advances in Web technology have addressed these three issues head on. The eXtensible Markup Language (XML) presently being defined by the World Wide Web Consortium (W3C) promises to combine the power of SGML with the compactness of HTML.[9] With XML, Web authors will be able to 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 supplementary style sheets (for mathematics, say), XML will provide high-energy physics authors with the tools and functionality previously available only with TeX/LaTeX or SGML.

Continuing advances in search engine technology will make finding material on the Web easier, but will only partly address the frustrating resemblance Web searching bears to indiscriminate drift-net fishing. Document metadata will be used to ensure document applicability, quality and stability. Efforts such as the Dublin Core initiative seek to identify a core set of metadata elements—that is, indexing keywords—for finding things more effectively on the Web. Originally conceived for author-generated descriptions, this initiative has also attracted the attention of archivists at museums and libraries.

Document subscriptions work well in the paper world, but are unlikely to work on the Web. Micropayment technology[10] addresses the revenue generation issues of electronic publishing. Under a micropayment scheme, readers of Web documents will pay cents (or fractions of cents) for access. According to Jakob Nielson, "Micropayments lower the threshold and do not require a big decision before users get their initial benefits."[11]

tabase, to the world's high-energy physics community. Before its implementation on the Web, the only way to access SPIRES-HEP was to log in to the IBM system at SLAC, where the database resides, or to use the QSPIRES interface, which operated only from remote nodes using BITNET. With either method, a rudimentary knowledge of the somewhat esoteric SPIRES query language was required. According to Berners-Lee, the mounting of SPIRES-HEP on the Web was a vitally important factor in the rapid acceptance and utilization of the Web by the international high-energy physics community.[5]

Although ready access to SPIRES may have facilitated the acceptance of the Web, "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 Fermilab.[6] Mosaic brought forth the Web from physics and computer science laboratories and set the path for the role it plays today. The first version of Mosaic emerged in February 1993, but before its release, several other significant Web browsers had already been developed.

The early browsers developed at CERN in 1991–92 had been quickly followed by the first browsers designed for the X Windows computer environment—namely, the Viola browser developed by Pei Wei of the University of California, Berkeley (and later O'Reilly and Associates) and the Midas browser developed by Tony Johnson at SLAC. (See figure 6.)

Like the Midas and Viola browsers before it, Mosaic was freely available on the Internet. Features provided by Midas and Viola undoubtedly influenced the course of Mosaic's development. Fermilab's Pordes 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.[6]

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) for Mosaic and developed the NCSA server. Meanwhile, the small but dedicated Web group at CERN continued its efforts on the CERN server. More important, however, the CERN group developed and supported Libwww, the general software library—the nuts and bolts—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.

Cailliau estimates that the number of Web servers worldwide went from 2500 at the end of 1994 to 50 000 at the end of 1995. This growth was due largely to the availability of server software developed by CERN and NCSA.

Berners-Lee had long argued that the standardization of the Web be open and nonproprietary. His model was that of the X Consortium, which had been formed in 1988 to further the development of the X Window system and whose major goal was 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).

Ultimate irony

Ironically, 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 unanimously approved the construction of the Large Hadron Collider, the world's most ambitious high-energy physics accelerator to date. Because the LHC would cost so much, the council decided not to fund further Web development, and, in negotiations with the European Commission and INRIA (the Institut National pour la Recherche en Informatique and Automatique in France), transferred the CERN Web project to INRIA. It was also agreed that INRIA, along with MIT/LCS, would host the W3C. (Keio University in Japan became the third host of the W3C in 1996.)

By the end of 1997, Cailliau estimates the number of Web sites worldwide had reached 650 000—on every continent and, possibly, in every country. The number of Web users is now certainly in the millions. Web addresses have become almost as widespread as telephone numbers, while Web sites provide every conceivable form of content or service. And, as described in the box on page 35, the Web promises to dramatically change how scientists work and interact.

In reflecting on the current size and scope of the Web, it is important to remember that this revolutionary technology grew from the needs and imagination of the high-energy physics community and that scientific research—even on the ultimate constituents of matter—often leads to significant technical spin-offs.

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

References

1. B. Segal, "A Short History of Internet Protocols at CERN," in-house CERN document (1995). Also available at http://wwwcn.cern.ch/pdp/ns/ben/TCPHIST.html.
2. T. Berners-Lee, "Information Management: A Proposal," in-house CERN document (1989; rev. ed. 1990). Also available at http://www.w3.org/History/1989/proposal.html.
3. T. Berners-Lee, R. Cailliau, "World-Wide Web: Proposal for a HyperText Project," CERN in-house document (1990).
4. P. Garrett, D. Ritchie, *Collaborating over the Web: Libraries and Laboratories or the Librarian and the Webmaster,* Fermilab, Batavia, Illinois (1995).
5. T. Berners-Lee, keynote address, History/Developers' Day, at Sixth International WWW Conference, Santa Clara, Calif., 11 April 1997. Also available at http://www.w3.org/Talks/9704WWW6-tbl.
6. E. Berger, FermiNews, vol. 19, no. 16, 16 August 1996.
7. Available at http://www.mcs.anl.gov/DOE2000/.
8. R. Kouzes, J. Myers, W. Wulf, Computer **29** (8), 40 (1996).
9. D. Connolly, ed., "XML: Principles, Tools, and Techniques, World-Web Web Journal," vol. 2 (4), O'Reilly & Assocs., Cambridge, Mass. (1997).
10. A. Herzberg, "Safeguarding Digital Library Contents: Charging for Online Content." D-Lib Magazine, January 1998.
11. J. Neilsen, "The Case for Micropayments," available at http://www.useit.com/alertbox/980125.html.

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."

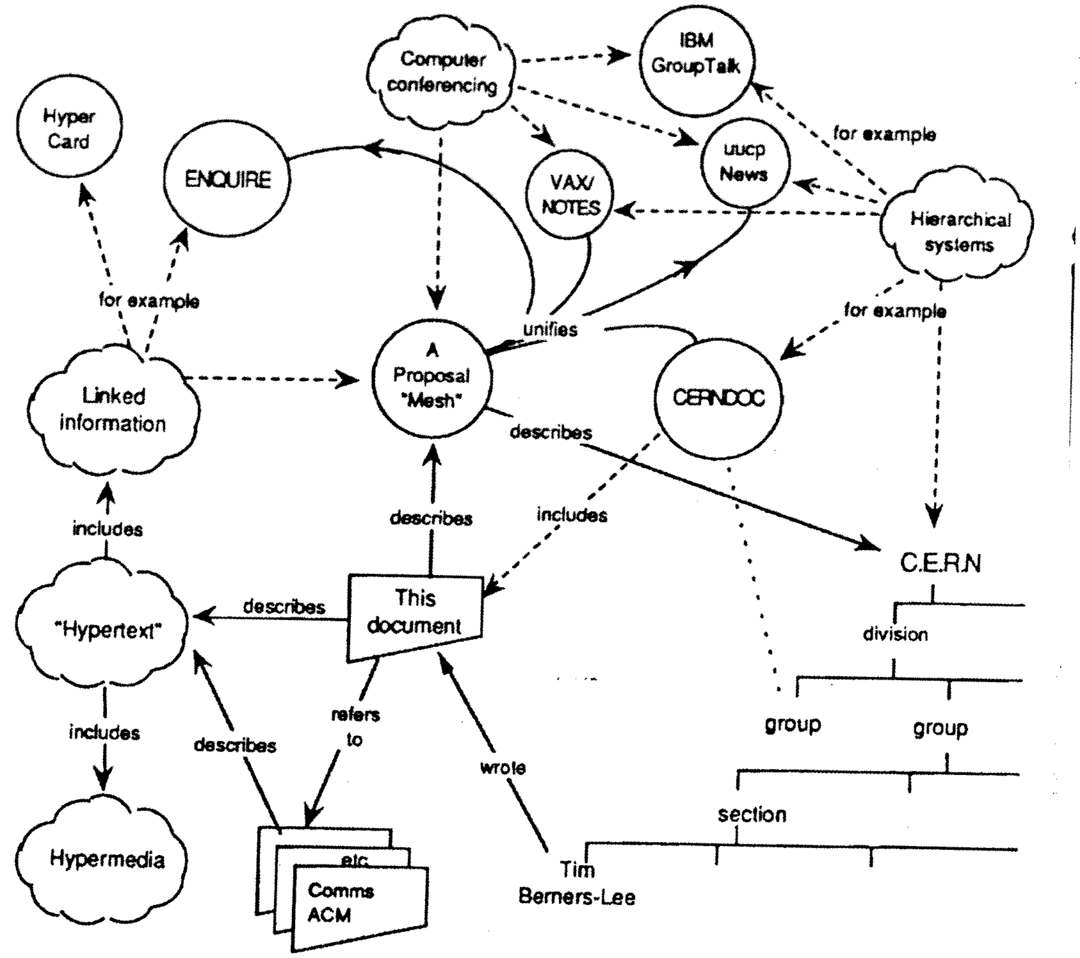


Figure 1. A CERN Distributed Hypertext System

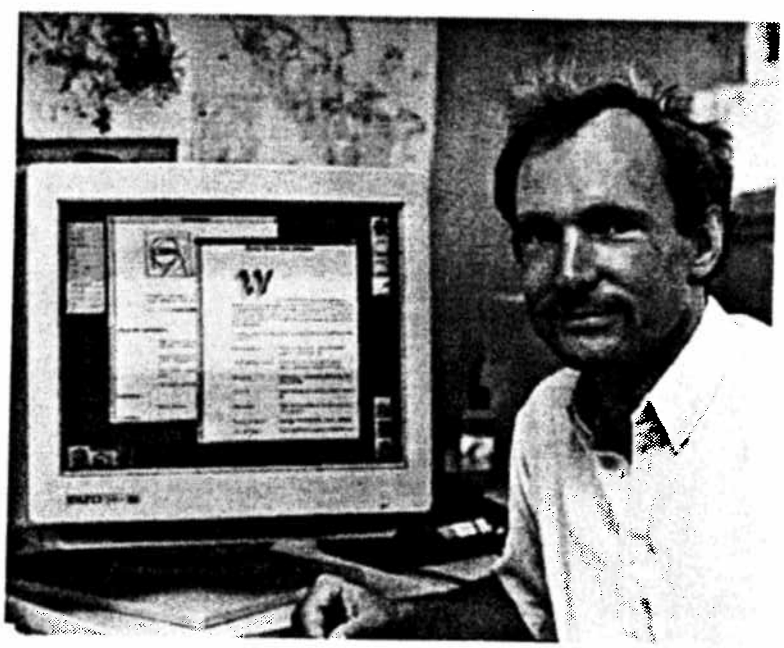
Independently of Berners-Lee, Robert Cailliau was also interested in use of hypertext as a solution to CERN's information retrieval problems. He was especially interested in platform-dependent hypertext systems such as HyperCard and hypertext scripting languages such as HyperTalk. In 1990, Berners-Lee 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."

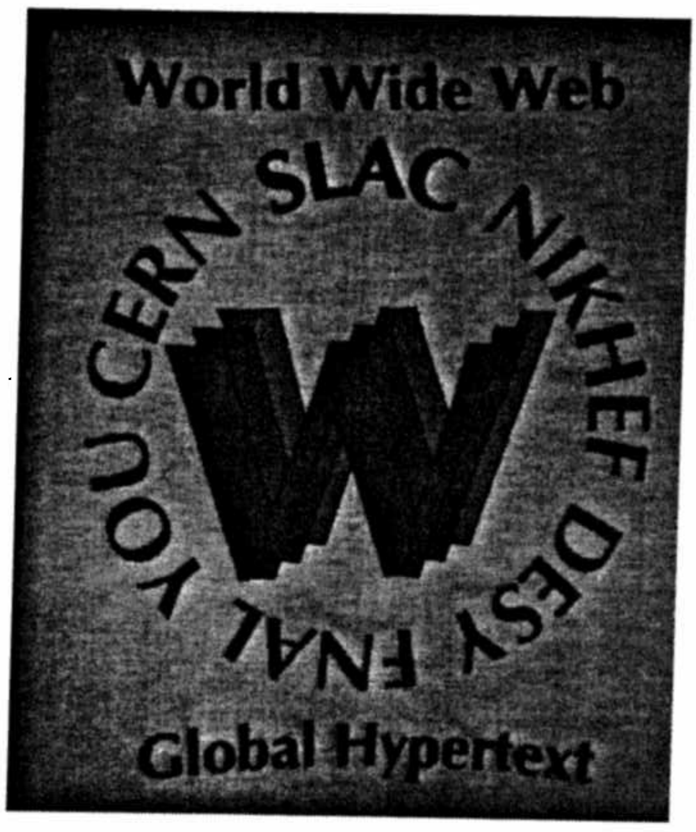


Figure 3. The First WWW T-shirt

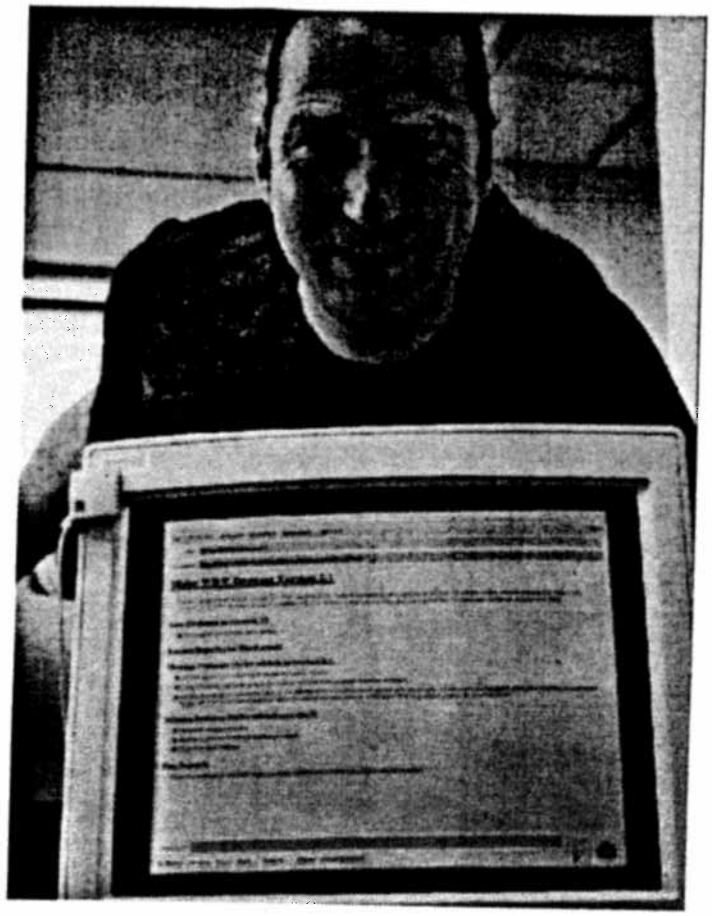
Bjorn Nilsson recalls that in June 1993, the ALEPH experiment at CERN was able for the first time to generally provide information previously limited to ALWS (the Aleph offline system) users. A Web version of ALWHO provided accurate phone list and e-mail data. ALWS HELP and detector status were also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