The World Wide Web and Emerging Internet Resource Discovery Standards for Scholarly Literature

STUART L. WEIBEL

ABSTRACT

THE WORLD WIDE WEB (WWW) HAS BECOME an important medium for the dissemination of scholarly information. This article discusses the technology of the Web and why it is likely to have a lasting impact on the dissemination of scholarship. The role of the display and indexing of structured text is discussed, particularly the relationship of HyperText Markup Language (HTML) and Standard Generalized Markup Language (SGML), as well as problems associated with matching the needs of session-based document retrieval and the stateless architecture of the Web. The relationship of existing bibliographic description standards to emerging standards for the description of networked information resources is described.

INTRODUCTION

Access is the heart of the library. All other functions—selection, acquisition, cataloging, and preservation—derive from the basic objective of matching the information needs of users to the materials that will satisfy those needs.

The rapid development of networking and electronic dissemination of information forces upon us both opportunities and burdens. The opportunity is to provide the greater flexibility and convenience that networked information affords. The burden is to integrate these services with the existing library infrastructure such that users are not confronted with two disjoint information environments.

Stuart L. Weibel, OCLC Office of Research, 6565 Frantz Road, Dublin, OH 43017-3395
LIBRARY TRENDS, Vol. 43, No. 4, Spring 1995, pp. 627-44

Technology is at the heart of all aspects of networked access to scholarly information: economics, protection of intellectual property rights, preserving the record of scholarship, and even the politics of publishing. The choices made—and those forced upon us—comprise the landscape in which the information ecology will evolve.

The present discussion focuses on the technological issues of electronic access to scholarly publishing, and, in particular, the World Wide Web (WWW) as a medium for the delivery of scholarly information. The role of structured text on the Internet is addressed (specifically the relationship between Standard Generalized Markup Language and HyperText Markup Language). Finally, the relationship of MARC standards to the evolving Uniform Resource Identifiers and related standards are discussed.

THE WORLD WIDE WEB: THE FRONT END OF THE INTERNET

World Wide Web browsers, such as NCSA's Mosaic, are the front end to the Internet; in a year's time they catapulted the Internet (and the WWW in particular) to the forefront of the public consciousness and garnered enormous attention in the national press. Will the WWW persist or is it simply the latest technology *du jour*? Is it a suitable vehicle for the delivery of scholarly information or simply a pretty (inter)face? The ability of Web protocols to embrace other Internet protocols (such as ftp and Gopher) augurs well for its extensibility over time. The WWW is likely to be an enduring part of the Internet landscape and should be considered a keystone for the delivery of scholarly information. What makes this technology important, and how will it affect access to scholarly information?

THE PRETTY, EASY ANSWER

WWW browsers provide users with an easy-to-use interface that introduces some of the virtues of print aesthetics to online interactions. Pretty is more than just attractive; good typography improves reading speed and comprehension and makes reading easier on the eyes (Gould et al., 1987). The ability to convey typographic emphasis—bold, italic, font size, and style changes—allows authors to provide visual inflection that is unavailable in the unembellished display of ASCII text.

The ease with which graphics are made available on the Web also accounts for much of its appeal. One can argue whether the information content of the net has been improved in proportion to the additional demands that graphics impose on network bandwidth and workstation performance, but there is no question that users are attracted to graphics, and the old saw that a picture is worth a thousand words still applies (the problem is, it costs ten thousand words, but not to quibble); without graphics, scholarly (or any other) publishing will not prosper on the Internet.

The point and click idiom of hypertext makes it simple for even an inexperienced user to traverse links among documents and collections. The enormous popularity of NASA's Shoemaker-Levy comet images (JPL-a, 1994) illustrates the potential to afford rapid public access to the latest information in the sciences. Reports of this extraordinary event were not limited to grainy newspaper images or the glossy selection of news magazines a week later but were available to millions of Internet users on their desktops within hours of being available to the scientific community. Almost 2 million files were served to nearly 50,000 unique host computers in approximately three weeks in July 1994 from NASA's Jet Propulsion Lab (JPL-b, 1994). The Shoemaker-Levy home page was visited more than 3 million times in the next six months.

The same virtues that make the Web appealing to the neophyte netmaven will improve access to ideas by scholars. Every chemistry student learns the law of mass action in the first week of class: molecules are more likely to combine chemically if they bump into one another more frequently. So it is with ideas; make it easier to bump into them, and they will combine more often in the minds of users to form more complex ideas.

The ease of use of Web browsers is complemented by the ease of publishing information on the net. Any determined user can master the technology of mounting a Web server; never has the technology of mass distribution of information been so accessible. The technology of the Web is the modern equivalent of a printing press, except you could never get a printing press for free. Anyone with an Internet connection can download a Web server and browser for free from places like the National Center for Supercomputing Applications (NCSA) at the University of Illinois (where NCSA Mosaic was developed) (NCSA, 1994) or CERN, the European High Energy Physics Research organization where the Web began (Berners-Lee, 1994).

Early examples of scholarly publishing on the Web are already appearing. Johns Hopkins University Press, for example, has made some of its journals available on an experimental basis under Project Muse (MUSE, 1994). Prototypes of this nature foretell a future where universities and even departments will publish more of their scholars' efforts. Such publications are not likely to replace the peer-reviewed commercially published journal (the added-value of professional publication staffs is underrated), but the ease of network publishing will blur the distinctions between the formal scholarly journals and less formal means of publishing, making the so-called gray literature more important.

Scholarly societies are establishing their presence on the Web in growing numbers, recognizing it as an important new means of maintaining contact with their members. The University of Waterloo Library maintains a master list of such resources that at this writing includes ninety-three scholarly societies (Scholarly Societies Project, 1994).

The Web lowers barriers to networked publishing, making it easier to encroach upon the traditional markets of publishers. But, at the same time, the idiom of hypertext affords publishers a means to protect their position by providing links among their products, thereby enhancing their value as part of a larger whole. Following links is the native idiom of World Wide Web navigation. The opportunity to weave a publication into the context of related scholarship (by embedding explicit links to related articles) will enhance the usefulness of the literature to the scholar and provide a competitive advantage for publishers whose scope of publications is large. Publishers will be driven to add value to their journals through enhanced features, and this competition will help drive the quality and capabilities of electronic publishing to higher levels.

OCLC's Electronic Journals Online supports such linkage today. For example, in the *Online Journal of Current Clinical Trials,* cited references are linked to their corresponding records in the MEDLINE database and can be retrieved by clicking on the highlighted reference. As more and more journals are delivered online, articles will be retrievable in the same manner, creating a web of scholarship that will be accessible with far less effort than with conventional paper literature.

Elsevier's TULIP project (Willis et al., 1994) suggests one model for promoting access to journals electronically without incurring the expense of full electronic markup. Essentially an electronic microfilm, scanned pages are delivered to the workstation to be read on the screen or printed. Although not as desirable as having structured full text, it is a practical and affordable means of making previously published journals accessible electronically, and it is a conceptually simple matter to link these page images to citations to them that will occur in fully electronic journals. But it is structured full text that is the key to the added value of electronically published journals, and these benefits will help to overcome the disadvantages of screen-based display and the less than ideal means we now have for managing online text.

The Importance of Structured Text

HyperText Markup Language, the lingua franca of Web documents, is basically a narrowly constrained subset of Standard Generalized Markup Language (SGML). HTML is a relatively simple (if limited) way of introducing the notion of structured data to users and would-be providers alike (for an eloquent discussion of the importance of structured text and SGML, see Coombs et al., 1987).

SGML is an international standard (ISO 8879) for the description of text in machine-readable form. An SGML document consists of text that is marked up with descriptive tags that specify the function of a given element within the document. As a formal language construct, an SGML document can be parsed against a Document Type Definition (DTD) that unambiguously defines what elements are allowed and where in the document they can (or must) occur.

Procedural versus Descriptive Markup

Conventional computerized typography defines the page layout of documents. Often referred to as procedural markup systems, they are oriented toward page description (that is, specification of where characters are placed on a page) rather than the description of document structure. While such systems are well suited to the production of paper-based journals, they are inadequate to the demands of electronic delivery.

SGML is the leading example of descriptive markup systems, so termed because a system uses a formally defined tag set to describe the role of a document component, rather than specifying the procedure to display it. The details of character placement on a screen (or piece of paper, for that matter) are left to the application software on the user's machine.

For example, in a procedural markup scheme, a title might be tagged to specify page centering, a type size four points larger than body text, and bold style. In a descriptive scheme such as SGML, it would simply be marked as a title. Final details of the display would be determined by the application responsible for its presentation (a Web browser, for example).

An added benefit of such a device-neutral encoding of the data is that the text is more readily reused by other applications. For example, it is a simple matter to extract subcomponents of a structured document (citation and abstract information, perhaps) that might be made available in a separate product or service.

The structure of text also provides the underpinnings for fielded searching. A user can search such a corpus for terms occurring only in the title or abstract, or perhaps for all papers that cite a particular paper, or papers written by a particular author. The ability to embed such structure (and use it as the basis for index searches) will be increasingly important as larger online full-text collections become common and users are confronted with the problems of managing ever-larger retrieval sets.

HTML: Less is Less (but Simplicity is a Virtue)

HTML falls far short of the full power of SGML, and it suffers from the intermixing of structural markup (identifying the elements of a document) and display, or procedural markup (specification of how objects should be displayed on a page or screen). Nonetheless, it is a relatively straightforward idiom for imparting useful, if minimal, structure to networked information. While not up to the demands of formal publishing, its simplicity makes it ideal for less formal applications and occasional publishing tasks.

As HTML evolves (its enhancement is mediated by the activities of a Working Group of the Internet Engineering Task Force), it is likely to acquire greater expressive power, but it is not likely to be imbued with the full richness of SGML. If this were to happen, the Web browsers themselves would become far more complex to develop and maintain, and, more importantly, the simplicity of Web publishing as we know it today would be overwhelmed by the complexity of SGML.

A more likely scenario is the promulgation of SGML display applications that act in tandem with Web browsers much as external graphics viewers now support the display of image data without actually being part of the Web browser itself.

The first such SGML display engine has recently been announced by SoftQuad, a Toronto based vendor of SGML software and systems (SoftQuad, 1994). This product, named Panorama, is being made available in a public domain version and a commercial version (Panorama Pro, with somewhat enhanced capabilities).

Formal publishers will thus have a mechanism for distributing typographically complex, SGML-encoded, materials while occasional or less formal publishing will benefit from the simpler idiom that HTML affords.

An Interim Solution: The Translation of SGML to HTML

When SGML viewers are commonly available and widely supported by information providers, many of the representational problems of HTML text will become moot. During the transitional period leading to that state, the delivery of complex scholarly text requires an interim solution involving the translation of more complex markup (such as SGML) to HTML. The translation facility developed in the OCLC Office of Research is being used to provide Web-based access to Electronic Journals Online (Weibel et al., 1994). The first journal to be supported thus is the American Institute of Physics journal, *Applied Physics Letters Online.*

The translator parses an SGML document and decomposes it into a grammar tree. Each SGML entity in the document is translated into either its HTML-specific counterpart or a bitmap of the appropriate font character. Each formula (i.e., equations or mathematical notation) is extracted and translated from 12083 SGML to TeX, a computer-based typography system, and subsequently rendered to generate a corresponding bitmap. The 12083 SGML standard is a recommended style of SGML for books and journal publishing (12083, 1994).

Figures and tables are handled similarly to equations. However, in these cases, a reduced size or thumbnail image is embedded in the running text. It in turn is linked to a corresponding full-size image that can be downloaded at the user's discretion. The thumbnail images reduce initial image-loading burdens and provide a better-proportioned page display (full-resolution figures in electronic documents are typically of awkward proportion when included inline in running text). The full-sized image is displayed by selecting the thumbnail image, thereby invoking the appropriate external viewer.

Retaining SGML Structure for Indexing

The original SGML versions of the documents are used to build an inverted-file database, which is used to search the collection and generate pointers to the HTML version of the document. It is important to note that the original SGML markup is retained in this database, and this markup supports searching in specific fields (for example, limiting the search term to occurrences in the title or abstract). Thus, although the delivery of scholarly journals into the WWW involves some display formatting compromises, it need not result in loss of structured document searching capabilities.

Web Browsers as Remotely Programmable User Interfaces

The World Wide Web is a prominent example of client-server computing. A client is a program that issues a request for service to another, largely independent, piece of software that may reside on the same machine or, more often, on another machine. The two are linked to the degree that they share a communications protocol—a formally specified language for communicating with each other. The protocol in the Web that supports this client-server communication is called the HyperText Transport Protocol (HTTP).

An HTTP server is a fairly simple piece of software that "listens" at an Internet port for requests issued by a Web browser. In its simplest terms, this request is a string of characters known as a Uniform Resource Locator (URL) that specifies a scheme (ftp, Gopher, or http, for example), a host name (a machine on whose file system the resource resides), and finally, what is interpreted as a path to the location of a file on the host machine (this is not strictly the case but illustrates the basic workings of the protocol).

The server, having received a request for a document under its control, sends that file (typically HTML-encoded text) to the client that issued the request. Links may be embedded in the text, allowing users to jump to a different part of the document or another document entirely. In effect, these links become navigational controls embedded in the document, allowing the information provider to program the user interface to a limited degree.

HTML Forms: Getting Information from the User

A capability known as HTML Forms allows the document provider to interact with the user by putting up a template of text entry fields and several varieties of check boxes. The content and structure of the forms can be tailored to the specific task at hand, in effect a remote programming of the Web browser's capability that is as easily modified as any HTML file. Thus the user has what appoaches a universal client application—familiar in its appearance and behavior but adaptable to a wide variety of search and retrieval situations.

From the provider's point of view, changes in search capabilities no longer require redistributing software to an entire customer base but rather a relatively straightforward change in a data file.

In the future, it is likely that the HTML standard will support a persistent toolbar capability (not unlike the toolbars currently found in Microsoft Windows and Apple Macintosh application software). When this is available, controls will not scroll away as the user moves through a document as they do now.

One can imagine a future in which scholarly documents contain dynamic data objects that, through the click of a button, will launch virtual experiments or demonstrations based on the published data and the models, all mediated through a universal browser that serves as a familiar entry point to all the user's information resources.

Current Limitations of the World Wide Web

The problems that currently exist with the Web as a document delivery medium fall out into two broad categories: (1) representation and rendering of complex text, and (2) the stateless model of Web transactions.

Representation and Rendering Problems

There are three major impediments to the representation of complex scholarly material in current Web browsers:

1. *Mathematical Representation:* Mathematical notation is among the most challenging aspects of any typographic system. HTML does not currently support mathematical notation at all; all such objects must be represented as rendered bit images (simple graphics).
2. *Formal Models for Tables:* Tables are notoriously complex as typographic objects. In an electronic delivery medium, the normal problems of expressing tabular material in a coherent notation are compounded by the need to express the underlying logical structure of tables in such a manner that they can be parsed by software that can read tables for the visually impaired (or convert them to braille). Currently, HTML supports neither objective, though work is underway to substantially improve the table model so that more complex information will be represented effectively, both logically and for display.

   The recent submission of the HTML 2.0 standard for formal standardization under the Internet Engineering Task Force includes markup that will support the International Committee for Accessible Document Design (ICADD) recommendations for marking up electronic text for the visually handicapped. ICADD includes representatives from standards bodies, disabilities organizations, governments, and vendors of software and hardware. The ICADD text transformation process description forms an informational annex to ISO 12083, the SGML application for books and journals (ICADD, 1994). This is a first and important step toward making the Web accessible to those with visual impairments.
3. *Character Sets:* Currently there are no generalizable methods for displaying alternative character sets in HTML documents, thereby severely limiting the expression of any but a standard character set. At this time, the only characters that implementers can be sure will be present on all platforms for all browsers is the lower 128 characters of standard ASCII. Characters not found in this basic set must be rendered as small bitmapped images. This state of affairs will be rectified in future versions of the HTML standard, but until then, implementors must resort to the contorted work-arounds such as those described earlier in OCLC's SGML to HTML conversion.

Statelessness and Document Retrieval

The Web is an example of a stateless protocol. Each time a browser requests a document from a server, it makes and breaks a network connection, incurring a certain amount of overhead. This keeps the HTTP protocol simple, but when a server and client carry on extended transactions, it is inefficient (imagine hanging up and redialing the phone each time you completed a sentence in a phone call, and the person on the other end forgetting what you said in the previous call!). This is particularly problematic for an information service that benefits from (or requires) a session-based interaction, as is generally the case with reference databases and document retrieval systems. Users may modify a search strategy successively, reducing the size of a retrieval set until it is manageable. Such interactions require maintaining information about the state of a transaction.

Document search and retrieval works better in a session-based model—the server should retain session context for the user (reusable result sets, for example). OCLC has developed a hybrid HTTP-Z39.50 server to bridge the stateless world of HTTP and the session-based Z39.50 world (Weibel et al., 1994). The WebZ server provides Web access to Electronic Journals Online and soon will provide an entry point to the reference databases available under OCLC's FirstSearch system.

The WebZ hybrid server is acting as a Z39.50 client for the HTTP client, maintaining session information for any sequence of interactions from the same client. This is accomplished by putting a session ID in all URLs produced by the gateway and returned by the client. Authentication of the user is required only at the beginning of the session. Thereafter, the session ID in the URL is matched against active sessions maintained by the server, validating the request as legitimate. The sessions are aged by the server; if no subsequent requests are received within an arbitrary time interval, the session is closed and any subequent request by the same client would require re-authentication.

Finding What You Want

No discussion of access to networked information can fail to note the major problem facing the net—i.e., finding something you want. Web browsers are just that—browsers. As browsers, they excel, and only the jaded (or perhaps a too-busy reference librarian) can resist the appeal of clicking from site to site around the world, stumbling upon gems and chestnuts that delight the inquisitive mind. But along with the gems are plenty of dead ends, and to find something from a dead start—either a known-item search or a keyword or fielded search—is quite another challenge, a challenge which the current information infrastructure does not adequately support.

The evolving information infrastructure that has served libraries for many decades has yet to be transferred to the digital world. Some aspects of the two worlds require different solutions, but the problems of networked resource discovery are more like those of the conventional library than they are different. The library community can contribute valuable experience to these solutions. The long-term investment of the library community in MARC records and the Anglo-American Cataloguing Rules (AACR2) (Gorman & Winkler, 1988) represents a working model of how object description can be formalized to support resource discovery and retrieval. Should MARC therefore be the basis for similar systems on the Internet?

MARC AND THE INTERNET

The MARC record must certainly be accorded a hallowed place in the history of library automation, and it is unlikely that it will be supplanted in the foreseeable future as the currency of bibliographic record exchange. It must be seen for what it is, however—a carrier syntax. MARC is a container with well specified capacities and a set of rules (AACR2) that have been developed over a long period of time to guide us in packing these containers.

MARC and AACR2 have been a success partly because the world of library technical services has enjoyed a hegemony over them that allows closely regulated control. The culture of the networked information environment is unsympathetic to this monopoly of form and structure, however, and it is unlikely that this world will stand quietly while the keepers of MARC and AACR2 adapt these standards to networked information. That is the bad news. The good news is that it need not matter. The library community need not force the MARC record onto the rest of the world, nor must it forsake this enduring and useful legacy (not to say, the enormous financial investment in MARC systems).

As a carrier syntax, MARC serves libraries well, but the needs of cataloging, retrieving, and managing electronic objects are sufficiently different that departure from existing MARC fields is inevitable. The critical issue for libraries is that the description standards for networked information objects and services not be made incompatible with existing standards.

The library community must assure that new standards map gracefully into and out of MARC. To the extent that they evolve to meet the new demands of networked objects, libraries will adapt and find ways to incorporate new structures and fields into MARC. The challenge is to assure that the systems that evolve in the electronic communities do not break existing practice.

MARC 856 FIELDS

Acknowledgment of the changes necessary to accommodate networked resources is already evident in the development of the 856 field in MARC—a field given to codification of the information salient to the description of (and retrieval of) electronic objects on the network (USMARC Proposal 94-9, 1994).

The evolution of such description will take place in parallel with other communities that are working to support access to electronic resources. To the extent librarians involve themselves in these ongoing discussions, it will be possible to influence these standards so that they will work together. What are these other standards, and who is developing them?

SELF CATALOGING DOCUMENTS

One of the dreams of library automation is the automation of the cataloging process itself. This goal can be partially realized by the encoding of descriptive information in the document. The Text Encoding Initiative (TEI) represents an important step toward the realization of this goal. The Text Encoding Initiative is a program aimed at developing standards for the markup of scholarly information in SGML; the recently published guidelines (Sperberg-McQueen & Burnard, 1994) represent the culmination of this effort in a formal model for marking up scholarly information. This standard includes the specification of the TEI header, which includes descriptive information normally considered part of conventional bibliographic description. TEI headers are a prime candidate for the expression of cataloging information that will make a document self-cataloging. The developers of this encoding scheme were mindful of this possibility, and considerable attention was paid to the relationship of this encoding standard to existing library standards. Thus, the TEI headers are a prime candidate for the expression of cataloging data for electronic resources. See Giordano (1994) for a discussion of the TEI headers and Gaynor (1994) for a case study of their application for this purpose.

Uniform Resource Identifiers (URIs)

URIs are the trinity of resource location on the net. Comprised of three distinct, but closely related, standards (only one of which is now in common practice), the codification of these identifiers is central to the future workings of the networked information environment (see Duranceau et al., 1994, for an in-depth discussion of URIs).

Uniform Resource Locators (URLs) are the spine labels of networked information objects. They encode the location of an object or resource and, since the rules for specifying a URL are widely recognized and accepted as a standard, decoding the location specified by a URL has become a trivial capability implemented in a wide array of network applications. Decoding of such URLs is at the heart of any Web browser.

Resource location should not, however, be tied to physical location of an object; when an information object is moved, the URLs that pointed to that object are no longer valid. What is needed is a persistent unique identifier that does not change when the location of the object changes; this is the proposed function of Uniform Resource Names (URNs).

Uniform Resource Names

The notion of a Uniform Resource Name is closely related to an ISBN or LC card number. The two fundamental requirements of a URN are that it be unique and persistent into the forseeable future. It involves a naming authority—an agency with the authority to assign these unique identifiers, and with the commitment to maintain a resolution service to map logical object names (the URN) to physical location pointers (the URL). Without this resolution service, the URN is useless. URNs do not now exist except in testbed applications built by those attempting to bring these standards to fruition. The technical and administrative details of their use are difficult and as yet unresolved. Readers of this journal are likely to recognize the similarity of this problem to one of the primary functions of a library catalog—i.e., resolving a title or cataloging number to a library (in the case of an interlibrary loan request) or a shelf location (in the case of a local transaction).

Uniform Resource Characteristics (URCs)

Catalogs do far more than provide for resolution of persistent names to item location. Bibliographic description is the basis for search in the library world, and so it will be in the electronic library. URCs are the least well understood component of the trinity and will evolve slowly and probably painfully as the ongoing dialogue between the computing community and librarians proceeds. The nature of resource description is more complex in a networked environment, but the problems are largely those that librarians have dealt with successfully for decades. The lessons learned in managing libraries must be brought to bear on networked libraries, lest we discover that the so-called libraries without walls are roofless as well.

The Internet Engineering Task Force

These protocols are developed and evolve under the auspices of the Internet Engineering Task Force (IETF). The IETF is a collegial anarchy open to participation by any willing to expose their ideas and proposals to the harsh spotlight of scrutiny (and egos) either at the meetings (held three times a year), or on the working group listservs. The principles of adoption of a standard are basically that a "rough consensus" be achieved, and that protocols be backed by two or more independent implementations. Once considered the realm of "Geek of the Week" types only, sightings of technically oriented representatives of the library community are increasingly common. For an introduction to the culture and ethos of the IETF, see *The Tao of IETF—A Guide for New Attendees of the Internet Engineering Task Force* (Malkin, 1994).

CONCLUSION

The Internet is designing us as much as we are designing it; the requirements of providing distributed access to networked information are complex and challenge us economically, socially, politically, and technically. One astute observer at a recent IETF meeting invoked the well-known image of blind men describing the elephant, each describing only as much as he could touch. The problems of providing access are too large for any to fully grasp. We recognize the parts of the problem near and familiar, but the beast is larger than our experience. Bringing the elephant under service will require foresight as well as insight, and the problems are daunting. But the promise is great, and there are no alternatives. Electronic networks are here and their phenomenal growth will not admit inaction. If the library community fails to help describe the elephant, none of us should be terribly surprised to be stepped on by it.

Appendix

On-Line IETF Information

The Internet Engineering Task Force maintains up-to-date, on-line information on all of its activities.

FTP Access

The IETF information described above is available by anonymous FTP from several sites.

- Africa
  ftp.is.co.za (196.4.160.2)
  The Internet-Drafts on this machine are stored in GNU compressed form (i.e., the .gz file extension).
- Europe
  nic.nordu.net (192.36.148.17)
- Pacific Rim
  munnari.oz.au (128.250.1.21)
  The Internet-Drafts on this machine are stored in UNIX compressed form (i.e., the .Z file extension).
- US East Coast
  ds.internic.net (198.49.45.10)
- US West Coast
  ftp.isi.edu (128.9.0.32)

To retrieve this information, FTP to one of the above sites, log in with username anonymous and your e-mail address as the password. When logged in, change to the desired directory (using the cd command), and retrieve the desired files (using the get command).

E-mail Access

Internet-Drafts and other IETF material are available by mail server from ds.internic.net. To retrieve a file, mail a request to mailserv@ds.internic.net with a subject of anything you want. In the body, put one or more commands of the form:

FILE /ietf/1wg-summary.txt
FILE /internet-drafts/1id-abstracts.txt
FILE /iesg/iesg.92-11-10
PATH jdoe@somedomain.edu

where PATH lists the e-mail address where the response should be sent. If you have the mpack utility or a MIME-compliant mail reader, you may want to use the additional command:

ENCODING mime

This command results in the information being returned in a MIME message.

Other Access Methods

IETF-related information is also available via the World Wide Web and Gopher. Both of these services are constantly evolving over time, so a description of their contents will not be given.

- Gopher: gopher.ietf.cnri.reston.va.us
- WorldWide Web: <http://www.ietf.cnri.reston.va.us/home.html>

Mailing Lists

Much of the daily work of the IETF is conducted on electronic mailing lists. There are mailing lists for each of the working groups, as well as an IETF general discussion list and an IETF announcement list. Mail on the working group mailing lists is expected to be technically relevant to the working groups supported by that list.

The IETF announcement list is a "moderated" mailing list that receives the following types of messages:

- meeting logistics,
- agendas for working group and BOF sessions at IETF meetings,
- working group actions,
- Internet-Draft announcements,
- IESG last calls,
- IESG protocol and document actions, and
- RFC announcements.

To join the announcement list, send a request to:

ietf-announce-request@cnri.reston.va.us

The IETF discussion list is open and therefore has a wide range of topics. To join the IETF general discussion list, send a request to:

ietf-request@cnri.reston.va.us

To join most other Internet mailing lists, send a request to the associated "-request" address (e.g., to join the list@listhost list, send a message to: list-request@listhost). Never send a subscription message to the list itself. General inquiries about the IETF should be sent to:

ietf-info@cnri.reston.va.us

REFERENCES

12083. (1994). *Electronic manuscript preparation and markup proposed American National Standard* (ANSI/NISO/ISO 12083). Oxon Hill, MD: NISO Press.

12083. (1994). *International Committee for Accessible Document Design Facilities for Braille, Large Print and Computer Voice* ANSI/NISO/ISO 12083 ANNEX A.8 [Informative]. Electronic Manuscript Preparation and Markup Proposed American National Standard ANSI/NISO/ISO 12083. Oxon Hill, MD: NISO Press.

Berners-Lee, T. (1994). *World Wide Web Initiative,* December 19. Available from: *http://info.cern.ch/hypertext/WWW/The Project*

Coombs, J. H. ; Renear, A. H.; & DeRose, S. J. (1987). Markup systems and the future of scholarly text processing. *Communications of the Association for Computing Machinery, 30*(11), 933-947.

Duranceau, E. F. (Ed.). (1994). Naming and describing networked electronic resources: The role of Uniform Resource Identifiers. *Serials Review, 20*(4), 31-44.

Gaynor, E. (1994). Cataloging electronic texts: The University of Virginia Library experience. *Library Resources & Technical Services, 38*(4) 403-413.

Giordano, R. (1994). The documentation of electronic texts using Text Encoding Initiative headers: An introduction. *Library Resources & Technical Services, 38*(4) 389-401.

Gorman, M., & Winkler, P. W. (Eds.). (1988). *Anglo-American Cataloguing Rules,* 2d ed. Chicago, IL: American Library Association.

Gould, J. D.; Alfaro, L.; Finn, R.; Haupt, B.; & Minuto, A. (1987). Reading from CRT displays can be as fast as reading from paper. *Human Factors, 29*(5), 497-517.

IETF. (1994). *Internet Engineering Task Force Home Page* (accessed 12-19-1994). http://www.ietf.cnri.reston.va.us/home.html

JPL-a. (1994). *NASA Jet Propulsion Lab Comet Shoemaker-Levy Collision with Jupiter Home Page* (accessed 12-19-1994). Available from: *newsdesk@jpl.nasa.gov http://newproducts.jpl.nasa.gov/sl9/sl9.html*

JPL-b. (accessed 12-19-1994). *NASA Jet Propulsion Lab, World-Wide Web Access Statistics for JPL Comet Shoemaker-Levy Home Page.* Available from: *http://newproducts.jpl.nasa.gov/sl9/access9407_tot.html*

Malkin, G. (1994). *The Tao of IETF—A Guide for new attendees of the Internet Engineering Task Force* (RFC 1718) (Informational). Available from: *http://www.ietf.cnri.reston.va.us/tao.html*

NCSA Mosaic Home Page. (1994). (accessed 12-19-1994). Available from: *http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/NCSAMosaicHome.html*

Project MUSE. (1994). Johns Hopkins University Press, (accessed 12-19-1994). Available from: *http://muse.mse.jhu.edu* Co-Directors: Todd Kelley *<kelley@jhunix.hcf.jhu.edu>* & Susan Lewis *<suelewis@jhuvm.hcf.jhu.edu>*.

Scholarly Societies Project. (accessed 12-19-1994). University of Waterloo Library (accessed 12-19-1994). Available from: *http://www.lib.uwaterloo.ca/society/overview.html* Contacts: Christine Jewell *<cjewell@library.uwaterloo.ca>* & Jim Parrott *<jrparrot@library.uwaterloo.ca>*.

SoftQuad, Inc. (1994). (accessed 12-19-1994). Available from: *http://www.sq.com* (Contact: David Rowan *<daver@sq.com>*)

Sperberg-McQueen, C.M., & Burnard, L. (Eds.). (1994). Guidelines for Electronic Text Encoding and Interchange, Vols. I and II. Text Encoding Initiative (P3). Chicago, IL: Oxford University Press. Also available from: *http://etext.virginia.edu/TEI.html*

USMARC. (1994). Changes to the USMARC bibliographic format to accommodate online systems and services: Proposal 94-9 Network Development and MARC Standards Office
Via e-mail:
send e-mail to *<listproc@loc.gov>*
text of message: get usmarc 94-9.doc
get usmarc 94-9.cov
via telnet: telnet to marvel.loc.gov
login: marvel
select: Libraries and Publishers (Technical Services)
select: USMARC Standards
select: MARBI Proposals and Discussion Papers
(select appropriate proposal)

Weibel, S. L.; Miller, E.; Godby, J.; & Levan, R. (1994). An architecture for scholarly publishing on the World Wide Web (unpublished proceedings of the Second International World Wide Web Conference, Chicago, IL, October).

Willis, K.; Alexander, K.; Gosling, W. A.; Peters, G. R., Jr.; Schwarzwalder, R.; & Warner, B. F. (1994). TULIP—The University Licensing Program: Experiences at the University of Michigan. *Serials Review, 20*(3) 39-47.