NTIA Report 98-350

A Federal Standard on Electronic Media

William Ingram
Evie Gray

U.S. DEPARTMENT OF COMMERCE
William M. Daley, Secretary

Larry Irving, Assistant Secretary
for Communications and Information

August 1998

Contents

page

Figures

page

Tables

page

A Federal Standard on Electronic Media

W. Ingram and E. Gray[1]

Abstract. Federal telecommunications standards have traditionally been distributed in hard copy, *i.e.*, in book form, with mandated 5-year revisions requiring a new book or new change pages for every updated edition. Distributing Federal telecommunications standards on the World Wide Web (Web) or on CD ROM in electronic form presents possibilities for speeding and enhancing user access to specific subsections of the document, providing wider distribution to the intended audience, and promoting conservation of paper resources at the same time. Federal Standard 1037C on the Web and on CD ROM represents successful applications of HTML. In several ways, a large telecommunications glossary was ideally suited for presentation in hypertext format. The most significant advantage to the hypertext format is the rapidity with which users can jump from definition to definition without having to turn a precise (and often large) number of pages to arrive at the next desired definition. Yet, the very great size of the *Glossary* introduced one of the biggest hurdles in its hypertext development. The ITS editors of the *Glossary* surmounted that hurdle by using Perl scripts to generate automatically the many thousands of required hyperlinks in the large glossary. This paper describes those automated techniques. In addition, the paper addresses special considerations (of equipment, software, and image display) for presenting a user-friendly HTML product.

In the months since the electronic version of Federal Standard 1037C first became available, more than 450 copies of the CD ROM version have been distributed and the Web version has been accessed more than 75,000 times. This shows that there is a wide and ready audience for electronic access to the *Glossary*, and, very likely, other Federal standards.

key words: CD-ROM; Federal telecommunications standards; glossary; HTML; hypertext; Internet; NII

1. INTRODUCTION

1.1 Federal Standards

Federal telecommunications standards (FED-STD *or* FS), which are developed by the Federal Telecommunications Standards Committee (FTSC), assist in satisfying the Federal Government's mission of enhancing reliability and ensuring interoperability of the Government's resources. The benefits to the users of strong, accepted standards include the control of product quality, which promotes trade; the interoperability of equipment and services, which promotes communication and productivity; and the support of interdependent national economies, which reduces the cost of doing business by amplifying the benefits of economies of scale while reducing the need for retooling and retrofitting.[1]

[1]W. Ingram and E. Gray are with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, CO 80303-3328.

1.2 Organization of This Report

This report addresses the advantages of electronic media for distribution of Federal telecommunications standards. The paper describes the details of developing Hypertext Markup Language (HTML) code for a large Federal telecommunications standard, and illustrates the presentation of Federal standards on the World Wide Web (WWW *or* Web)[2] and in CD-ROM format, using FS-1037C, *Glossary of Telecommunication Terms,* as an example. Section 2 examines the advantages for distribution and use of Federal standards in electronic form as compared to hard copy. Those advantages are listed first for hypertext in general, and then for Web documents, and finally for CD ROM's. Section 3 discusses the background and history of the FS-1037C series of glossaries. Section 3.5 presents details of developing HTML code for FS-1037C. Section 4 gives details on equipment and software considerations for WWW presentation of FS-1037C, in particular, and other multimedia documents, in general. Section 5 illustrates equipment and software concerns for CD ROM distribution. Section 6 provides conclusions.

1.3 Traditional Distribution

Traditionally, Federal telecommunications standards have been distributed in hard copy,[3] such as in book format, with mandated 5-year revision cycles requiring a new book or new change pages for the book for every updated revision. Historically, Federal standards were distributed by the General Services Administration (GSA) or by NTIS (National Technical Information Service). In today's computer-driven society, however, the preparation and issuance of printed material in hard copy may be neither sufficient nor optimal for making the results of the standards committee's work available to the public. For many people today, the most convenient, most available, and most-often used source of information is the WWW.

1.4 WWW Distribution

The FS-1037C, *Glossary of Telecommunication Terms*, was put on the World Wide Web using HTML coding (ver. 3.0) to extend access to the large and growing audience on the Internet. Accommodations were made to provide compatibility with most computer platforms and with a variety of hypertext browsers. After the document had been on the Web for only three months, the number of accesses to the *Glossary* exceeded 75 per day. Web distribution of the *Glossary* vastly increased the availability of the document. Earlier editions of the *Glossary* (FS-1037A and 1037B) were reportedly best sellers, but the editors were never able to determine the extent of the document's circulation or popularity. The "hit counter" (*i.e.*, the record of how many times the Web page was accessed) on the Web edition has provided a better idea of FS-1037C's use: after seven months on the Web, the hit counter showed that the Standard had been accessed more than 12,000 times. Slightly more than a year later, the hit counter has recorded more than 75,000 visits to the page.

[2]An international, virtual-network-based information service composed of Internet host computers that provide on-line information in a specific hypertext format.

[3]Printed copies of FS-1037C are available from
NCS, Mrs. J. Orndorff
701 S. Court House Rd.
Arlington, VA 22204-2198
Telephone: (703) 607-6204

Advantages of Web distribution of the Standard include:

✔ **Early access to the approved Standard.** Web distribution of the Standard allows potential users early access to the official version of the Standard months before it rolls off the presses in hard copy.

✔ **Reduced cost to the user.** Purchasers of hard copies of the Standard are required to pay for the printed copy. No fee or purchase price is associated with using the Web version (other than the normal access fees paid by the user to their own Internet Service Provider (ISP)).

✔ **Early collection of review comments.** In the past, proposed revisions to the Standard were collected only a few months before the revision process was initiated. With Web availability, reviewers may respond immediately with comments (via e-mail or otherwise). This review mechanism is particularly helpful for those reviewers who are not able to participate in the revision process and who cannot see early drafts of the *Glossary*. Also, easier access to the Web copy (relative to the hard copies) and the organized layout of the Web version, allows readers to provide better, quicker, more thorough, and more organized critique.

✔ **Portability.** Users need not carry around a heavy paper copy of the Standard (which contains 450 pages). Users need not purchase multiple hard copies of the Standard in order to have a copy handy at multiple locations.

✔ **Accessibility.** By making Federal standards more accessible (on the Web) and easier to use (in hypertext), the many benefits of having the standards are realized sooner and more fully.

✔ **Wide distribution.** Increased distribution and use of the Standard emphasizes the advantages of having a standard: interoperability, portability, and uniformity. These advantages are particularly strong in the case of a vocabulary standard, where standardized definitions of components and services help the manufacturers, vendors, and users alike. Without a standardized vocabulary, advances in technology could not be unambiguously communicated. Unless the advances are effectively communicated, those advances are lost, and all their benefits with them.

1.5 CD ROM Distribution

The completed FS-1037C, *Glossary of Telecommunication Terms*, was put onto one CD ROM in three versions:

1) Portable Document Format (PDF) (with a freeware reader included on the CD).

2) HTML format (with a browser included on the CD) viewable on several popular platforms such as Macintosh™, Windows 95®, Windows 3.x®, and DOS®.[4] The CD ROM's are distributed by the National Communications System in Arlington, Virginia. (License fees were paid to the producer of the browser to allow inclusion of a licensed copy of the browser on each copy of the CD ROM.)

[4]Certain commercial equipment, software, services, or materials are identified in this paper to specify adequately the procedures. In no case does such identification imply recommendation or endorsement by the National Telecommunications and Information Administration, nor does it imply that the software, material or equipment identified is necessarily the best available for the purpose.

The HTML version of the Standard differed from the hard copy version in format only; the technical content remained unchanged. Also, the HTML version contained new sections called "help," "welcome," "hints," "files," and "credits," that were not part of the hard copy and that provided additional, unofficial information to the viewer of the HTML version. Additionally, the figures in the HTML version are presented in color. (Details of the development of the HTML version for the CD ROM and for the Web are discussed in Section 4 below.)

3) WordPerfect® 6.1 files (as archival files, readable by any machine having an executable version of WordPerfect® 6.1).

2. ELECTRONIC MEDIA

This section describes hypertext characteristics and the advantages of the hypertext medium for presentation of FS-1037C.

Effective development of electronic versions of Federal standards requires consideration of the exact medium for presentation, which depends, in turn, on the nature of the information contained in the standard. The two types of electronic distribution addressed in this paper are via CD ROM and the WWW.

CD ROM's are an obvious choice for storing very large documents, because the CD ROM can hold 660 Mb of information. If data is compressed (using existing compression algorithms) before putting it on the CD ROM, approximately twice as much data (more than a gigabyte) can be stored in the 660 Mb on the CD ROM. Further, a CD ROM can accommodate multimedia; *i.e.*, text, hypertext, pictures (drawings, digitized photos, and video clips) as well as sound recordings (audio clips). All varieties of these media can be stored and accessed interactively on one CD ROM.

The World Wide Web also provides access to all of the above formats for information retrieval. Web pages, like CD ROM's, can contain still pictures, video and audio clips, and hypertext.

2.1 Advantages

As mentioned in Section 1.3, Federal standards have traditionally been printed and circulated in hard copy, like a book, with mandated 5-year revisions requiring a new book (or new change pages) for the updated edition. Distributing Federal telecommunications standards on electronic media presents possibilities for providing direct access to the subsections of the standards themselves and wider distribution to the intended audience, as well as for conserving paper. On-line versions of standards can be accessed even from a remote field site, using a computer and a modem. The advantages of electronic copies of Federal standards include: speed of access, accurate—and very rapid—text searches, ease of logistics, and—in the case of hypertext versions—external links to related documents and internal links within the subject standard itself.

If a paper copy of the *Glossary* is needed, the user can order one from the Government or use the electronic files, available on-line, to print out a copy on his or her own printer.

The format of this large glossary is particularly suited to presentation in the frames of a

computer browser. The individual definitions can be displayed in their standard dictionary format [2] as if they were individual index cards.

Electronic, dynamic versions of documents have additional, obvious advantages for continuous updating of information. However, the focus of this paper is on glossaries that are only updated every five years. Electronic updating of the documents, as a part of an on-going editing effort, is outside the scope of this paper.

2.2 Hypertext

The advantages of hypertext for non-sequential navigation of a glossary are many. One especially powerful advantage is the ease of use of a hypertext index. The HTML version of the *Glossary* has a hypertext index of the major subject categories of terms defined.

The index, contained in Appendix B of FS-1037C, contains a list of 30 different topics within the field of telecommunications for which telecommunications terms are defined. Under each topic is a list of term names that are defined within the *Glossary* and that are related to that topic. Users of the *Glossary* can use this Appendix to examine whole families of related terms. Viewers of the HTML version not only can view a family of related term names in any category of the index, but can select any of those term names individually to read the appropriate definition immediately with a click of the mouse button. Meanwhile, the list of categories in the index remains on the screen in a separate window.

2.3 Frames

Within a hypertext presentation, the display area can be broken up into smaller display areas. Each of these smaller display areas can be addressed independently, and is called a "frame." A frame can be used to keep certain display items visible to the viewer while changing the information displayed in other frames. For the hypertext version of FS-1037C, four separate frames are used to display information. The four frames are shown in **Figure 1**. More information on frames is presented in Section 3.9.1.

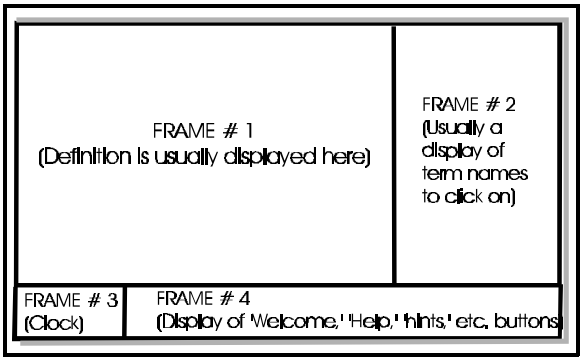


Figure 1. Configuration of the four frames used in the hypertext version of FS-1037C.

2.4 WWW and CD

Having the *Glossary* available on the Web greatly increases its availability and potential for use. Workers at a field site no longer need be concerned that they may not have the latest paper edition of the *Glossary* on hand; they need only look at the *Glossary* on the Web to be assured that they have the latest edition.

Posting the hypertext version of the *Glossary* on the Web also has a significant advantage for classroom use, where it is not necessary to purchase dozens of copies of the *Glossary*, one for each member of the class. Copies of

the *Glossary* can be viewed, via the Internet, on every computer in the classroom, simultaneously.

The CD ROM version requires a CD ROM reader, but allows the same access to the hypertext version as is available on the Web. This version provides the same advantages of nonlinear navigation, *viz.*, rapid access to definitions and quick access to families of definitions. Availability of the CD ROM is more limited than on the Web, since copies of the CD must be obtained from the National Communications System (NCS), whereas anyone with a computer connection to the Internet can browse the *Glossary* at the URL shown in **Figure 2**.

http://glossary.its.bldrdoc.gov/fs-1037

Figure 2. The URL of the on-line version of the *Glossary*.

The CD ROM and the Web versions of the *Glossary* contain additional PDF files of the entire *Glossary*. *Glossary* pages presented in PDF format resemble a photograph of the hard-copy page of the *Glossary*. Each page of the PDF file begins and ends at the same point in the text where the hard copy begins and ends. The figures in the PDF file are shown in shades of black and white, without the color shown in the hypertext files. However, the PDF reader included on the CD ROM only allows string searches to be made within individual sections of the Standard. Full hypertext linking is not available in PDF.

3. *GLOSSARY* BACKGROUND

This section outlines the history of the development of FS-1037C, its intended audience, its content, and how the format of that content is changed by the HTML presentation.

3.1 Previous Glossaries

For more than two decades—and with the sponsorship of the National Communications System and the U.S. Department of Defense—the Institute for Telecommunication Sciences (ITS), of the U.S. Department of Commerce and National Telecommunications and Information Administration (NTIA), has had a significant role in the development of the state-of-the-art telecommunications glossaries listed in **Table 1**. The chronology of these glossaries is shown in **Figure 3**.

3.2 Federal Standard 1037C

The FS-1037C *Glossary* represents the review by and coordination with a dozen experts in as many agencies, over a period of 18 months.

In 1996, FED-STD-1037C, *Glossary of Telecommunication Terms*, was developed as a revision of the 1991 *Glossary* in the usual hard-copy format, with the addition of 80 drawings, 50 displayed equations, a list of acronyms and abbreviations (Appendix A of FS-1037C), and an index of principal families of term names (Appendix B of FS-1037C).

Table 1. A Listing of the Full Titles for the Federal Telecommunications Glossaries in the FS-1037C Series

Year	Glossary
1976	MIL-STD-188-120, *Glossary of Telecommunication Terms*
1980	FS-1037, *Glossary of Telecommunication Terms*
1979	NTIA-SP 79-14, *Fiber Optics Glossary*
1979 - 1980	NBS Handbook 140, which became EIA-440A, & IEEE 812, *Optical Waveguide Communications Glossary*
1986	FS-1037A, *Glossary of Telecommunication Terms*
1991	FS-1037B, *Glossary of Telecommunication Terms*
1996	FS-1037C, *Glossary of Telecommunication Terms*

The *Glossary* was developed by the Subcommittee to Revise FS-1037B, which is a subcommittee of the Federal Telecommunications Standards Committee, chaired by the NCS. The Institute for Telecommunication Sciences—which is the research and engineering arm of NTIA—in Boulder, Colorado, chaired the Subcommittee, edited the document, developed the hypertext version of the *Glossary*, put it on the Web, and produced the CD ROM.

This latest glossary, which is the August 1996 revision of the 1991 Standard, has been updated to include terminology and definitions from the fields of networking, the NII (National Information Infrastructure, or the Information Superhighway), CD ROM's, video teleconferencing, browsers, and other, new Web terms.

Figure 4 and **Figure 5** show an example of a revised definition (the definition of *D region*). It is a revision and compilation of the definition that appeared in FS-1037B.

The illustrations and definitions were developed after open, interactive committee review, expert staff review, and 120 days of public review and comment, in keeping with requirements for preserving due process in standards development [3]. All definitions and figures were subject to the same rigorous scrutiny and open review process before the *Glossary* was approved by GSA as a Federal standard (according to practices in place in 1996).

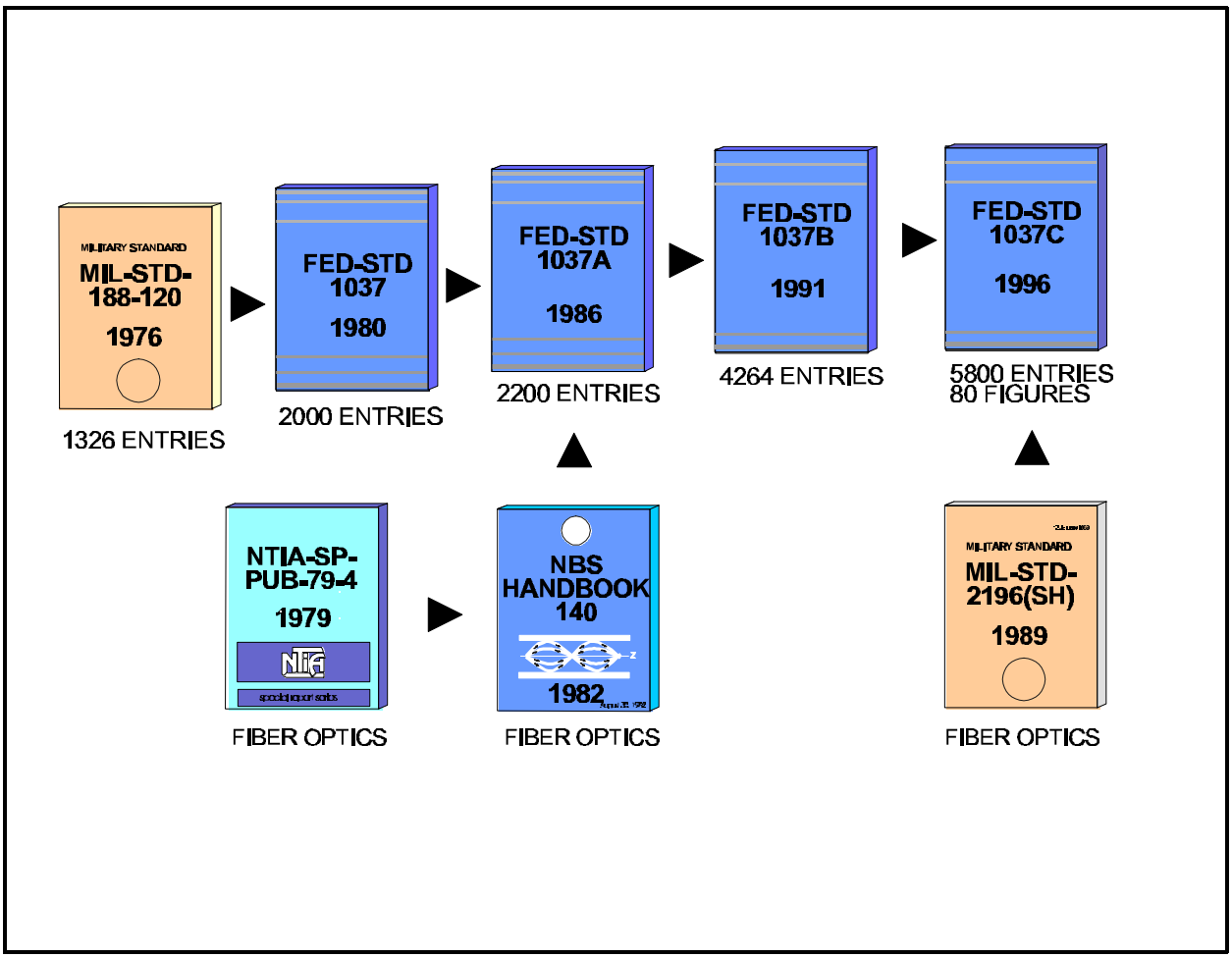


Figure 3. History of the development of FS-1037C, *Glossary of Telecommunication Terms.*

D region: That portion of the ionosphere existing approximately 50 to 95 km above the surface of the Earth. (188) *Note:* Attenuation of radio waves, caused by ionospheric free-electron density generated by solar radiation, is pronounced during daylight hours. Because solar radiation is not present at night, ionization ceases, and hence attenuation of radio waves ceases.

Figure 4. A sample definition from FS-1037C.

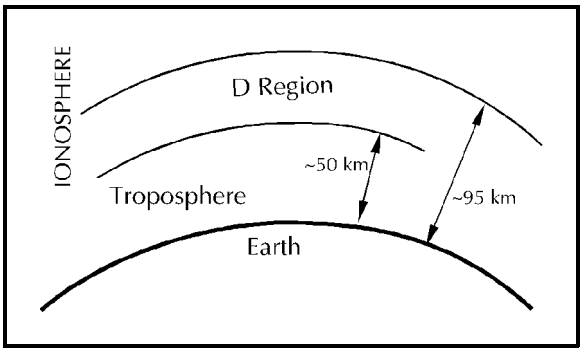


Figure 5. Sample illustration from the definition *D Region*.

The *Glossary* includes definitions in the disciplines of:

✔ Communications Security
✔ Data Communications
✔ Data Processing
✔ Fax
✔ Fiber Optics Communications
✔ Grounding and Bonding
✔ National Security/Emergency Preparedness
✔ National Information Infrastructure
✔ Networks (including Internet, Intelligent Nets, ISDN, Broadband ISDN, Network Management)
✔ Premises Wiring
✔ Radar
✔ Radio Communications
✔ Spectrum Sharing
✔ Telegraphy
✔ Telephony
✔ TV (UHF, VHF, Cable, and High-Definition TV)

The *Glossary* is designed with the following prospective users in mind:

✔ Acquisitions Agents
✔ Government Standards Writers
✔ Government Standards Users
✔ NII Planners
✔ Operations and Maintenance Workers
✔ Research and Development Workers
✔ Technical Writers and Editors
✔ Telecommunications Designers
✔ Telecommunications Instructors
✔ Telecommunications Managers and Planners
✔ Telecommunications Vendors
✔ Test Personnel

3.3 Hypertext *Glossary*

When the draft Standard was complete and approved by GSA as an official, mandatory standard, the Institute for Telecommunication Sciences developed it into hypertext format, using HTML 3.0. At the time of this development, HTML 3.0 had not yet been formally standardized. Many of the features found in HTML 3.0, however, have already been incorporated into the newest generations of browsers, despite this lack of formal standardization. The Institute decided to conform with as many of the features listed in the proposed version of HTML 3.0 as possible. It is likely that any specifications from HTML 3.0 that are widely enough implemented, will become, at least, a *de facto* part of HTML 3.0.

One example of an HTML 3.0 specification that the ITS editors used, before its formal adoption, is HTML frames. In the three years since the editors incorporated frames into the electronic version of FS-1037C, every new version of each of the major Internet browsers has supported the use of frames. It is likely that every subsequent version will support frames as well.

3.4 Example of a Term in Hypertext

The only difference between the hard copy and the electronic copy of the Standard is in format. For example, in hard copy, the definition of *D region* (with its figure) wraps between two columns (and pages), and the figure is shown in black and white. In the PDF version, the same format is preserved.

In the hypertext version, however, each separate definition is displayed on a screen of its own, rather than sharing a page with many

other entries in typical dictionary format. **Figure 6** shows two successive hypertext screens that display first the definition of *D region* (of the ionosphere), and then the multimedia drawing for *D region*. The arrows in **Figure 6** point to the successive screens that appear with each of the two mouse clicks in the definition of *D region*.

3.5 Hypertext Language

This section discusses the development of HTML and identifies some of the coding to convert a document into HTML format for presentation on the Web. This section also describes the techniques used to generate automatically the many thousands of links needed in the hypertext version of FS-1037C. Special cases—such as figures and equations–are also addressed.

Hypertext Markup Language (HTML) is a programming language designed to allow an author to create an interactive, multimedia display for a remote viewer. Some HTML features allow the author to present the user sights and sounds relevant to the document, as well as to download documents containing related information.

3.6 Description of HTML

In the late 1980s, the Internet (which, at that time, was still known as ARPANET—the Advanced Research Projects Agency Network) had evolved into a way for scientists and other researchers quickly and easily to share information with each other. As an example, one scientist could place his or her data file into a specific directory on a particular computer. The scientist could inform colleagues where and how to find that data file. Those colleagues could then download the data file and analyze it. Any conclusions, comments, or additional data from those colleagues could be sent back to the original author electronically by E-mail or as a new data file.

The popularity of this system of data exchange prompted several universities to build menu systems to catalog all of the documents that their scientists were making available to other researchers. These menu systems allowed a user to find a specific document by selecting choices from a series of static menus.

Early menus required the user to type in a number or letter representing the category of interest. New menus would sequentially be presented to the user until the desired document was found and downloaded.

Later menu systems were more interactive and allowed the user to move a cursor around the screen and to select menu items with a computer mouse or with the cursor keys on the keyboard.

Until this point, network site administrators had been building their menu structures in a tree-like hierarchy, with each menu item leading downwards to a submenu. Each submenu contained a number of branches to further submenus or, at the lowest levels, to the documents themselves. Also, all menu items pointed exclusively to documents and submenus within the facility (*e.g.*, a university or a research laboratory).

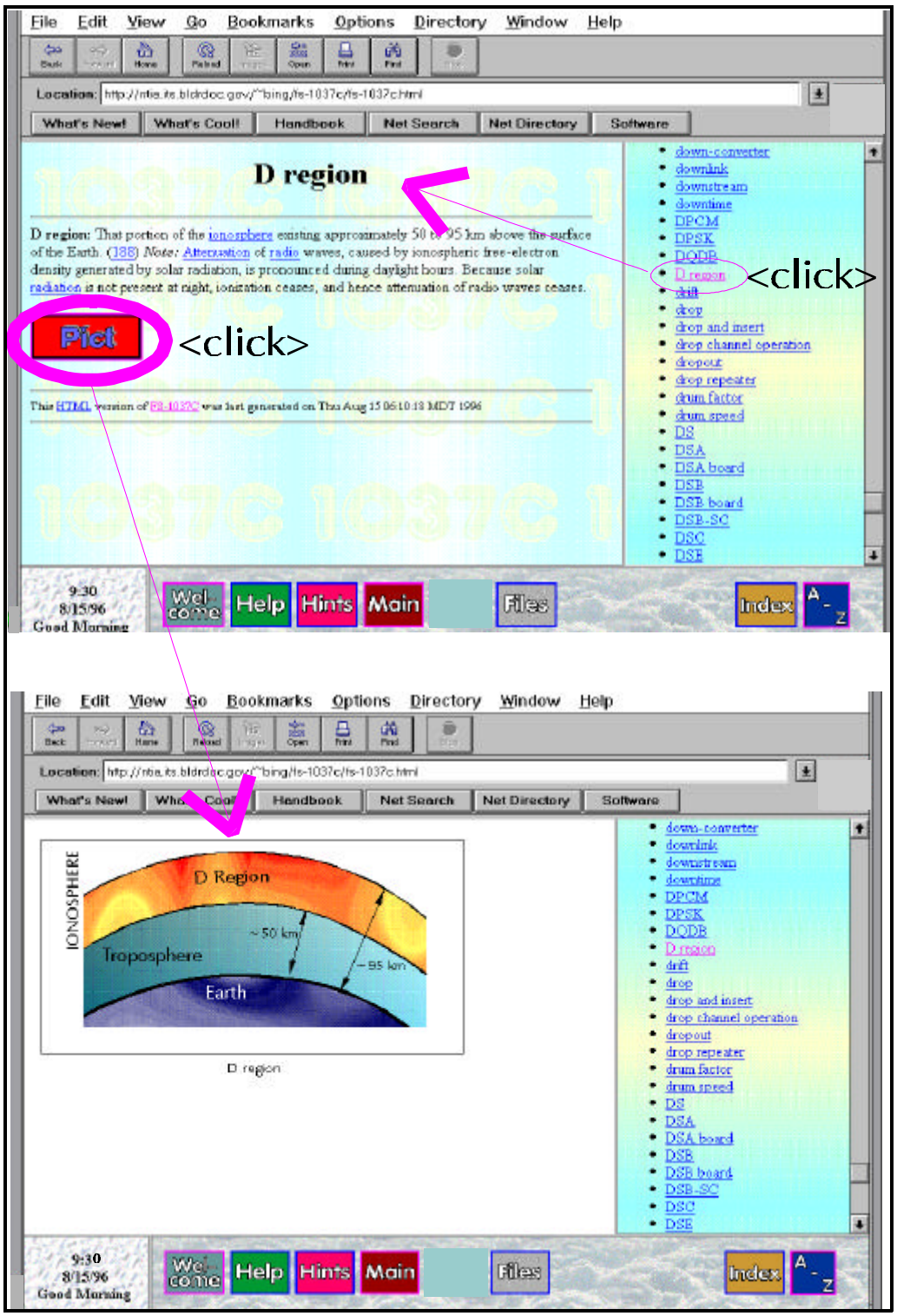


Figure 6. Two successive screens for the definition and illustration of *D region* in the hypertext version of FS-1037C.

As the number and types of different document categories expanded, it became evident that many documents were appropriate to be listed in several different categories. Rather than save a second, identical copy of the document onto the system's hard drive, the administrator would simply build separate menu items, to the single document file, from each appropriate menu. In addition, menu administrators began adding menu items that referenced documents or menus at remote facilities. These procedures began to give the menu systems more of a spider-web structure than a tree hierarchy. Soon thereafter, these menuing systems gained the nickname, "World Wide Web." Similarly, the individual menus became known as "Web Pages."

The next step in this evolution occurred when these Web sites adorned their pages with simple graphics such as University logos and pictures of the authors. Unfortunately, most users were using simple, nongraphical ASCII terminals that could not accept or display such images.

The administrators at some of these Web sites solved this dilemma by writing a simple program that a user could run, on his or her own computer, that would correctly display the included graphics. This graphical front end (any interface program that is installed and run at the user's site is called a "front end") was called a "network browser" (later to be known as a "Web browser") and was simply given to any user who requested it.

Web-page authors could now send special triggers to any user running the author's specific browser. These special triggers caused text to be displayed in various colors or sizes and caused sections of the text to blink.

As more and more users began using these network browsers, the programmers continued to add enhancements and more sophisticated features to the browsers. Some browsers began to include special features that other browsers could not understand. Document authors and users became confused and frustrated, since users would now be required to use specific browsers to view specific pages. The Web authors and users formed a consortium[5] to develop a standardized computer language that would allow authors easily to create and present pages of data that could be correctly displayed on any type of browser.

The result of the consortium's effort was the Hypertext Markup Language. HTML provides a series of simple, ASCII tags that indicate changes in text format, font size, and text positioning. For example, the paired code, "" is often written into a Web page and transmitted to a browser. This pair of codes indicates to the browser that any text between the codes should be displayed as bold text. For example, the following line of HTML code,

```
The fire was so hot that
it <b>melted</b> the
framework
```

would be displayed to the viewer as,

The fire was so hot that it **melted** the framework.

[5]Information on the National Center for Supercomputing Applications (NCSA) can be found at their Web page, http://www.ncsa.edu/

Slightly more complicated HTML codes are used to display in-line graphics, to display data in tabular form, and to create links to other pages of information. For example, the HTML code,

```
<img src=”cover.gif”>
This is the cover of the
report.
```

would display a graphic file (called “COVER.GIF”), followed by the line of text, “This is the cover of the report.”

More complete tutorials for creating and using HTML documents can be found at several different sites on the World Wide Web[6] (and, they are all presented to the viewer using actual HTML encoding).

3.7 Selecting HTML

To distribute the “C” version of FS-1037 electronically, the ITS editing team identified two types of electronic distribution. The first type was on a physical medium, such as floppy disks or CD ROM. Floppy disks were quickly dismissed as an option when the size of the word-processing files for FS-1037C grew to more than 20 Mb of storage space. However, distributing the files on a CD ROM remained a viable option, since a CD can hold as many as 660 Mb of data. Several different versions of the *Glossary* could be included on one single CD (*e.g.*, hypertext, word-processing documents, and Postscript® files).

The other method of electronic distribution selected was on-line distribution over the Internet. The Internet, while having the potential to store an unlimited amount of data, is limited in that data can only be transferred between sites at a finite rate. Currently, this rate is not sufficient to transfer the entire *Glossary*—more than 20 Mb of files—in the time necessary to qualify the *Glossary* as a single, on-demand (*i.e.*, instantly accessible) resource.

However, when the *Glossary* files are separated into small files and catalogued into a database, the small bits of data that any one user would want to read at any one time (*i.e.*, a single definition), can quickly be found and transmitted. With this rapid transfer time, the *Glossary* does qualify as an on-demand document.

The HTML system of Web-like menus was judged, by the editors, to be adequate for putting FS-1037C on the Web in hypertext. Additionally, the growing popularity of the World Wide Web assures that current (and future) browsers, as well as HTML itself, will be supported and upgraded for many years to come.

3.8 Perl Scripts

The editors examined several means of converting the large text files of FS-1037C to HTML files for presentation as Web pages. Some level of automation was required to generate the huge number of links required to make the document fully interconnected within itself. The Perl programming language was selected as the tool for this automation task.

[6]Three such on-line resources are,
“http://WWW.Stars.com/”
and
“http://home.mcom.com/home/how-to-create-web-services.html”
and
“http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html”

"Perl" is the abbreviation for Practical Extraction and Report Language. It was created as a utility language to extract information from large text files and to manipulate that information to create summaries and reports. Before Perl was created, most computer languages were designed primarily to accept and to manipulate large volumes of numerical data. None of those languages could gracefully handle large volumes of textual data.

Federal Standard 1037C contains a very large volume of text. The Standard contains more than 5,800 separate terms and definitions and requires more than 450 pages to print all of them. Early experiments, using word-processing macros to generate Web pages, showed that the large volume of text quickly overwhelmed the capacity of those word-processing programs. Some trials showed that such a macro might take more than a week to convert the 5,800 terms into simple Web Pages that contained no file linking or other advanced HTML features.

By using Perl scripts (also called "macros" or "programs") those same 5,800 terms could be converted into simple Web Pages in less than two hours. Most of this time savings came from the fact that

1) the Perl macros were being run on a (faster) computer workstation rather than on a desktop personal computer; and

2) the Perl language is a very simple language, and its scripts can be executed much faster than word-processor-based macros.

Later, after all the links and advanced HTML features had been added, this Perl-based conversion process required about 4 hours for complete execution. No tests were done to see how long it might take for word-processing macros to accomplish the same conversion, but 8 weeks is an off-the-cuff estimate. Illustrations of Perl script conversions are shown in Sections 3.8.5 and 3.8.6.

One further advantage of using Perl scripts is that the conversion process can be automated. Workstations, such as the one used for this project, are typically on-line 24 hours a day. Timer programs can be created that automatically trigger the conversion process to run at a predetermined time. Normally, the selected time for execution would be outside of normal business hours, when, it is assumed, other users of the workstation would not be inconvenienced by the extra processing load on that computer.

For this project, the conversion process (*i.e.*, the Perl script) was set to run daily at approximately 6:00 p.m. This conversion process would create 5,800 new files in a few hours—long before the next day's scheduled start of business. The editors of the HTML documents could then examine the previous day's *output* Web pages, find and correct any problems in the *input* ASCII files, and then check the generated results the next day.

Also, the daily HTML files were created in a subdirectory that was made available over the Internet. Invited parties could view the 5,800 files with their Web browsers, and could submit electronic comments about linking and format errors or inconsistencies in the evolving draft files. Local examiners could use the

printing feature in their browser to print out any pages with errors, circle the errors in red ink, and hand them to the editors of the HTML document. In this way, any discovered errors could be fixed by the end of the business day, allowing the corrections to be automatically incorporated into the new version of the files, generated later that evening. The local examiners could view the regenerated files the next day to verify that the errors had been corrected. This constant feedback was valuable to the stability of the conversion process and for quality control of the final product.

3.8.1 Converting to ASCII

Since Perl is mainly designed to manipulate text, all information in the *Glossary* that can not be represented by normal text must be converted into ASCII[7] characters, or deleted. Each of the 30 data files (one for each of the 26 letters of the alphabet and 4 files for the introductory sections and appendices) needed to be converted from their native word-processing format into plain ASCII text files.

Any information contained in the non-ASCII codes would be lost in the conversion process. Therefore, all non-ASCII codes were removed separately or changed to ASCII characters that represented the intended information. Methods for converting the various non-ASCII characteristics of the *Glossary* are described in the following sections. Automation assisted in this otherwise time-consuming task.

[7]The ASCII code (CCITT Recommendation V.3) defines 128 allowable characters, most of which are the standard characters found on a typewriter.

3.8.2 Font Characteristics

Many of the *Glossary* definitions contain text that is either bolded, italicized, superscripted, or subscripted. A word processor will keep track of which text is bold, for example, by inserting special codes directly into the text of the document. These word-processor codes are normally invisible to the document editor and are stored, with the document file, as special, non-ASCII codes. These codes are ignored when this document is converted from the word-processing format to plain ASCII text. As a result, all bold, italics, superscript, and subscript information is lost.

To avoid this loss, special word-processing macros were written and executed that would do preemptive conversions. That is, before the *Glossary* files were converted into ASCII files, all of the unconvertible codes would be changed into special codes that could survive the conversion process. One macro, for example, searches a word-processing document, locates the word-processor's special code for "turn the bold on," and replaces it with the string, "", which is the HTML equivalent for "turn the bold on." The macro also finds the matching "turn the bold off" codes and replaces them with, "". Four special macros had to be written (one each for bolded, italicized, superscripted, and subscripted text) and executed for each of the 30 files.

3.8.3 Special Characters

The *Glossary* files also contain a large number of special, non-ASCII characters, such as Θ (Greek characters), ∞ (mathematic symbols), and © (typographic symbols). Without special treatment, none of these characters would

have survived the conversion process and each would be displayed as a blank space.

In HTML, some of these special characters can be displayed by using certain strings of ASCII characters. For example, the "®" symbol can be created by inserting the string "`&reg;`" wherever the "®" character is needed. A line of code containing this substitution,

```
<i>Note:</i> Ada&reg; is
a registered trademark
of the U.S. Government.
```

would be displayed on a browser as,

Note: Ada® is a registered trademark of the U.S. Government.

A word-processing macro was written and executed (on all 30 files) for each special symbol that had an HTML equivalent. **Table 2** shows several special symbols or characters, frequently used in the *Glossary*, with their HTML equivalents.[8]

Unfortunately, not all of the special characters used in the *Glossary* had equivalent HTML character codes—none of the Greek characters and most of the typographical characters did not. Initial reports on future versions of HTML show that these symbols will have character codes created for them, but these codes had not yet been implemented into current versions (at the time of this report) of network browsers. The editors did not use these codes for the hypertext documents.

Table 2. The HTML equivalents for unusual characters used in the *Glossary*.

Special Symbol	HTML Code
® (registered trademark)	`&reg;`
′ (prime)	```
° (degrees)	`&#176;`
ç (c w/cedil)	`&ccedil;`
• (multiplication)	`&#149;`
× (multiplication)	`&#215;`
½ (one half)	`&#189;`
& (ampersand)	`&`
< (less than)	`<`
> (greater than)	`>`
ü (u w/umlaut)	`&uuml;`
± ("plus or minus")	`&#177;`
é (e w/acute)	`&eacute;`
è (e w/grave)	`&egrave;`

The *Glossary* was examined word by word. As each new unconvertible symbol was discovered, a bitmapped image of that symbol was designed with graphical art software. A special line of HTML code was inserted into the word-processing document to replace the unconvertible symbol. This HTML code would

[8]Several lists of special characters and their HTML equivalents can be found at,
http://www.uni-passau.de/~ramsch/iso8859-1.html
or
http://www.w3.org/pub/WWW/MarkUp/html-spec/html-spec_13.html

instruct a Web browser to display the referenced bitmapped image into the line of text where the symbol should be.

For example, in the following line of code,

```
...where <i>M</i> is the
magnitude and <img src =
"../symbols/theta.gif">
is the phase angle.
```

the Greek symbol θ (Theta) is replaced by the HTML code,

```
<img src =
"../symbols/theta.gif">,
```

which tells the browser to retrieve the bitmapped image called "theta.gif" and display it where the θ should be. The result looks just as the definition was originally intended,

```
...where M is the
magnitude and 2 is the
phase angle.
```

Several word-processing macros were created and executed to convert all of the special characters, in the 30 files, into links to bitmapped images. Use of bitmapped images to display equations is addressed in Section 3.8.5 below.

3.8.4 Reserved Characters

HTML coding uses several common ASCII characters to represent very specific things to the browser. The "<" and ">" characters (commonly known as the "less than" and "greater than" symbols) are used by HTML as brackets or field delimiters. The browser interprets any information after a "<" as processing instructions that describe how to present text and graphics. If the document editor includes this line of text,

```
Use an angle x, where x
< 45, to reset the
device.
```

it would, unfortunately, be displayed to the viewer as,

```
Use an angle x, where x
```

because any information after (and including) the symbol "<" would be interpreted as processing commands rather than as text to be displayed to the reader. The browser would not recognize the characters after the "<" as a valid processing command, and would simply ignore them.

To avoid this problem, each mathematical use of the symbols "<" and ">", in the word-processing document, was replaced with the HTML codes, "`<`" ("lt" is an abbreviation for "less than") and "`>`" ("gt" is an abbreviation for "greater than"), respectively. A word-processing macro was written and executed to substitute automatically each instance of the *bona fide* mathematical use of the symbols in each of the 30 files.

Similarly, the symbol "&" (ampersand) is a symbol that HTML uses to indicate to the browser that the characters following it are to be interpreted as a special display code. To avoid problems, each instance of the conjunctive "&" character within the text was replaced by the HTML code, "`&`" (in HTML, "amp" is an abbreviation for "ampersand").

3.8.5 Equations

At the time of the conversion of the *Glossary* from word-processor format to HTML format, only version 2.0 of HTML was available and widely supported. That version did not support the use of an in-line equation creator. Versions 3.0 and 4.0 of HTML are being created and are likely to become available sometime in the future. The new versions may contain codes specifically designed for displaying equations. Although descriptions of many of the proposed specifications for in-line equations in these versions are available, most browsers will not support the equation codes until these new versions of HTML become approved standards. For this reason, the ITS editors used HTML code that was currently available and supported, to construct equations. In the years since design of the electronic version of FS-1037C was started, HTML versions 3.0 or 4.0, and their equation codes, have not been approved or incorporated into any popular browser software.

The definitions in the *Glossary* contain more than 200 equations. Most of the equations are simple equations that could be inserted directly into a line of text using plain ASCII characters and a few Greek and special characters (for example, x + 7 ≤ B * 8). Fifty of the equations, however, were complicated enough to warrant using the word processor's equation editor. The following is a sample equation,

$$f(t) = \frac{1}{2\pi}\frac{d^2\theta(t)}{dt^2} \quad .$$

Version 2.0 of the HTML code does not have any mechanism to support equations built with a word processor's equation editor. Even if the individual equation components could be converted into plain ASCII text, the final structure of the equation could not be guaranteed when the user viewed it with a browser.

Most browsers have internal word-wrapping capabilities that allow the editor of an HTML document to be somewhat cavalier about inserting too many or too few line breaks or spaces between words and paragraphs. The browser will correctly formulate the text into neat paragraphs no matter how many extra spaces and line breaks the editor has added to the raw text and no matter what browser configuration is used. While this word-wrapping feature is quite useful for displaying paragraphs of textual information, it is inadequate for displaying equations, for which the relative placement of variables and symbols on the computer screen is critically important.

Bitmapped images of the 50 complicated equations were created and stored on the hard disk. By using the "image display" HTML code for each equation, the equations were displayed, in-line, at the desired points in the *Glossary*. Some sample code to display one such equation is,

```
<i>Note:</i> Frequency
fluctuation, <i>f</i>
(<i>t</i>), is given by
<p> <img src =
"../equats/37c_16.gif">
<p> where <img src =
"../symbols/theta.gif">
(<i>t</i> ) is the phase
angle of the sinusoidal
wave with respect to
time, <i>t</i>.
```

The result of this line of code is that the GIF file, "37c_16.gif" (the file name for the image of this particular equation) is inserted into display and is shown as,

Note: Frequency fluctuation, $f(t)$, is given by

$$f(t) = \frac{1}{2\pi}\frac{d^2\theta(t)}{dt^2} \quad ,$$

where $\theta(t)$ is the phase angle of the sinusoidal wave with respect to time, t.

During the process of converting the word-processing files into HTML files, each of the 50 equations was manually converted into a bitmapped image, stored on a hard drive, and replaced with a short, coded string of characters that represented the file name of the image. Later, a Perl script converted the coded string into the correct HTML code to display the equation. For example, the Perl script converted the string "#e37c_16#" into the HTML code "<img src = "../equats/37c_16.gif">", which is the proper code for displaying the subject equation.

3.8.6 Figures

As was the case for the in-line equations, there was no way to translate the in-line figures directly from the word-processing document into the HTML format. Each of the 80 figures had to be converted into a bitmapped image and inserted into the HTML document with special codes. Section 3.9.2 describes the methods used for such conversion of the figures.

The HTML code for inserting these bitmapped graphics into the on-line document is very similar to the code used for inserting equations. For example, the following lines of code,

```
<a href ="../gifs/
37c-cov.gif">
<b>Telecommunications:
Glossary of
Telecommunication
Terms</b> </a> <p>
<a href = "../gifs/
37c-cov.gif"> <img src =
"../gifs/ 37c-covs.gif"
align="top"> </a>
<br>Date of Publication:
August 7, 1996<hr>
```

would display on a browser as shown in **Figure 7**.

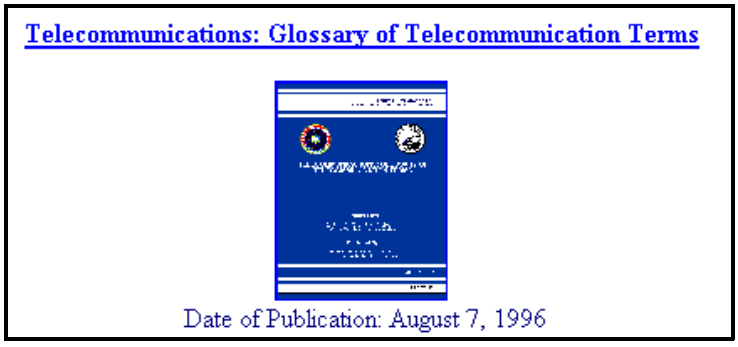


Figure 7. A sample of an in-line image.

The HTML code, <img src=""">, tells the browser to display an image on the user's screen. In this example, the displayed image has the name, 37c-covs.gif, and is a

small, thumbnail version of the front cover of FS-1037C.

A larger, more legible version of the cover is available to the user by clicking the cursor on the small thumbnail image. This was accomplished by using a dynamic link, described in Section 3.8.7.

During the process of converting the word-processing files into HTML files, each figure was manually replaced with a short, coded string of characters that represented the file name of the appropriate figure image. Later, a Perl script converted the coded string into the correct HTML code to indicate to the user that a figure is available for viewing. Unlike the equation images, the figure images could not reasonably be displayed in-line. Displaying these images in-line would be prohibitively slow for users with slow modems or slow Internet connections. To accommodate slow transmission rates, a small "Pict" icon was inserted below the definition to inform the user that a picture was associated with the definition and that this picture could be viewed by clicking on that "Pict" icon.

For example, the Perl script would convert the string "`#idregion`" into the HTML code "`<a href = ″dregion.gif″> <img src = ″$_pict.gif″></a>`". The result of the code is displayed in **Figure 8**. The actual figure that would be displayed (in this example, when the user "clicked" on the "Pict" icon) is shown in the lower half of **Figure 5**.

3.8.7 Anchors and Links

Dynamic links can be inserted into any HTML document. Section 3.9.3 describes why editors would want to use dynamic links in their Web pages.

For the example used in the previous section (**Figure 8**), the terms, "ionosphere," "attenuation," "radio," and "radiation" are terms for which the reader might want to look up individual definitions as they relate to the current definition (they are underlined in the figure).

A Perl script was written to examine the text of each of the 5,800 definitions and to look for any term name, within the text of a definition, that was itself a term defined elsewhere within the *Glossary*. This was the most complex and time-consuming task of the automatic conversion process.

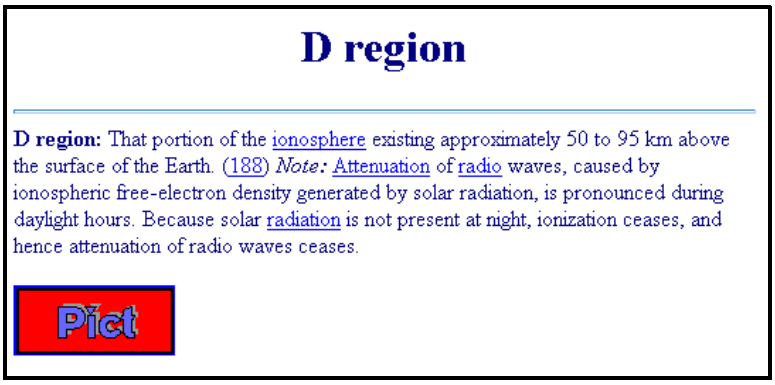


Figure 8. Sample definition with access to an on-line illustration.

The first step in this task was to generate a computer file listing all 5,800 term names. This list of terms (or "anchors list") was used by the Perl script to identify defined term names

embedded within the text of other definitions. Links were built connecting each such term name to its own definition.

Only 4,300 of the 5,800 term names in FS-1037C were judged to be suitable anchors for such links. Some terms, such as those terms designated as "deprecated" (*i.e.*, terms that used to be relevant to the current state of telecommunications, and are now obsolete, but are kept in the *Glossary* for backward compatibility), were manually removed from the anchors list. Also, any term name that was not accompanied by a full definition (*i.e.*, any term name that was an abbreviation only or a synonym only), was manually removed from the anchors list.

Also removed from the anchors list were term names that are heavily used in telecommunications writings, but that were not used in the same sense as the FS-1037C definition. For example, the word "*line*" is defined in FS-1037C specifically as a transmission line or as a scanning line in imaging systems. If the word "*line*" is used in any other sense (a common desk dictionary defines 40 other, valid definitions), a link to the *Glossary* definition was not created, as the FS-1037C definition would have been misleading in that context. **Table 3** lists the most common anchors in the anchors list.

Many of the common term names in the *Glossary* (such as "*second*," "*group*," "*base*," and "*space*") shared the above problem. A telecommunications expert was required to examine each link created by the Perl script and to determine whether or not that particular link was appropriate.

Table 3. The Most Common Anchors in FS-1037C

Term Name (anchor)	Number of Links to the Term
signal	556
data	471
system	454
transmission	452
information	387
communications	369
time	347
frequency	325
network	308
circuit	271
power	246
radio	237
user	230
computer	204
channel	188

The expert was required to add a simple mark next to any words or phases that were initially linked by the automatic link-building process, but that should not have been linked. In cases where approximately 90% or more of the links to a particular term name were incorrect, that term name was manually removed from the anchors list. Following such removal, the link-building Perl script was executed using the shortened anchors list; the new execution

deleted all of the *removed* anchors' links from the Web pages.

When term names were removed from the anchors list to prevent inappropriate linking, the process left term names unlinked in cases where the link *was* appropriate and should have been created (fewer than 10% of the cases). In these cases, the telecommunications expert was required to insert a special code to those term names (within definitions) that would force the creation of a link despite the term name's not being on the anchors list.

For example, the word *branch* is used within 34 different *Glossary* definitions. The definition of *branch*, from FS-1037C, describes its usage in three cases only: as related to a) computer programming, b) network architecture, and c) electrical power distribution. In 31 of the 34 occurrences, the word *branch* was not used in any of the three defined connotations (in most of these non-defined connotations, it was used in relation to telephone equipment, power balance, or as a specific "*branch* of science"). Therefore, *branch* was removed from the anchors list. In the three remaining occurrences where the word was used in its *Glossary*-defined sense, special codes were inserted to force creation of the link regardless of the removal of *branch* from the anchors list.

This examination process and automatic link-building process required several weeks and many iterations to accomplish. A new Perl script was then written specifically to delete links from, and to add links to, term names that had the special reviewer's marks.

When the final hypertext version was completed, several more weeks were required for double-checking all of the links both for functionality and appropriateness.

Some term names appear more than once within a given definition. Sometimes (but rarely) term names appear within their own definitions. Normally, however, a term name is not used within its own definition because such use would be circular and would undermine the defining phrase. For example, it would be circular to use the word *red* within the definition of the color, *red*. The conversion Perl script was designed to identify the occurrences of a term name within a definition and to delete any recursive or duplicitive links.

The HTML code to build a link uses the "`a href`" command and looks like this:

```
That portion of the
<a href="_ionosphere.htm
l">ionosphere </a>
existing approximately
50 to 95 km above the
surface of the Earth.
```

In the above example, the word, *ionosphere* would appear to the user as an underlined word, usually highlighted in blue. The user could then "click" on the underlined word and the definition for *ionosphere* would replace the current definition on the screen. The above example would look like,

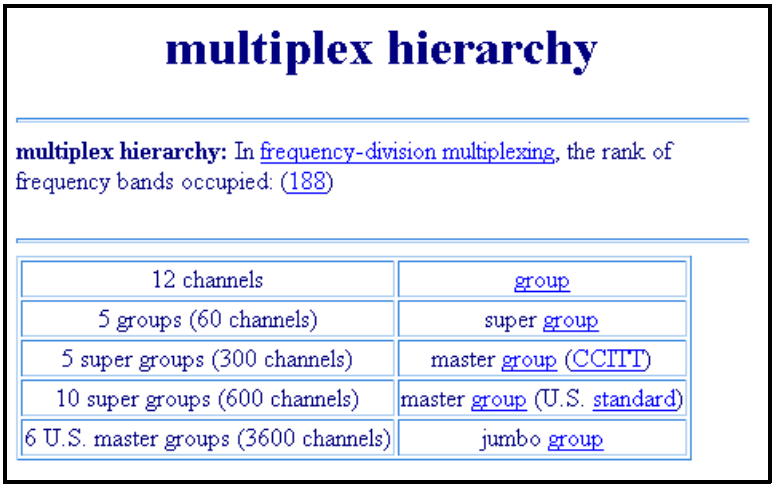
```
That portion of the
ionosphere existing
approximately 50 to 95
km above the surface of
the Earth.
```

Another example is described in Section 3.9.3.

3.8.8 Tables in HTML Format

There are five tables in FS-1037C. These tables represent data whose presentation in tabular form enhances comparisons, delimits concepts, or aids in reader comprehension.

Version 3.0 of HTML contains commands that allow the editor of a Web page to present data in tabular format. Many Web browsers have already incorporated the table-making codes despite the fact that HTML version 3.0 is not yet being adopted as a standard. The editors of FS-1037C decided to use the table features of HTML version 3.0. **Figure 9** shows one of the *Glossary* definitions that contains a table.



multiplex hierarchy

multiplex hierarchy: In frequency-division multiplexing, the rank of frequency bands occupied: (188)

12 channels	group
5 groups (60 channels)	super group
5 super groups (300 channels)	master group (CCITT)
10 super groups (600 channels)	master group (U.S. standard)
6 U.S. master groups (3600 channels)	jumbo group

Figure 9. A definition showing an example of a table.

The HTML code to build a table is fairly complex. Each table was completely deleted from the word-processing file, because all of the table-building commands are non-ASCII codes. Then each table was reconstructed in content and format by retyping all of it into the final HTML files.

This is the code that was used to create the table in **Figure 9**:

```
<table border> <tr align
= center> <td>12
channels</td> <td>
<a href = "_group.html">
group</a> </td> </tr>
<tr align = center>
<td>5 groups (60
channels) </td>
<td>super <a href =
"_group.html"> group</a>
</td> </tr> <tr align =
center> <td>5 super
groups (300 channels)
</td> <td>master <a href
= "_group.html">
group</a> (<a href =
"_ccitt.html">
CCITT</a>)</td> </tr>
<tr align = center>
<td>10 super groups (600
channels)</td>
<td>master <a href =
"_group.html">group</a>
(U.S. <a href =
"_standard.html">
standard</a>) </td>
</tr> <tr align =
center> <td>6 U.S.
master groups (3600
channels)</td> <td>jumbo
<a href = "_group.html">
group</a> </td> </tr>
</table> <p></b>
```

Appendix A of the *Glossary* is an extremely large table of abbreviations and acronyms, many of which are used in the text of the Standard. This large table requires special loading instructions, which are given in the "hints" section. As in the rest of the *Glossary*, hypertext links in the large table of Appendix A allow the viewer to click on defined entries

to view the definition of the linked abbreviation or acronym.

Appendix B of *this* paper contains a Web address that describes, in more detail, the techniques of table creation.

3.9 Display Considerations

3.9.1 Large Files

The FS-1037C editors designed the Web pages to be as user-friendly as possible. One of the ways in which they preserved the user-friendly presentation was by limiting the amount of information and graphics on any one Web page. Further, they accommodated viewers with slow Internet connections by sizing graphics carefully and by providing clickable icons for viewers to view the expanded graphics. The results of these considerations promoted a clean, uncluttered look to the finished Web page.

If all of the information on a single page of the hard copy of the *Glossary* were presented as a single Web page, it could cause intolerable delays when accessing the Standard. Instead, the Web pages for the *Glossary* were designed as if they were a series of index cards. Each Web page is kept small, with just a title (usually the term name), the relevant information (usually the term definition), and links to other information (usually a diagram or other definitions). Various menus allow the viewer to jump to specific definitions and sections of the Standard. By using this format, the viewer is able to navigate 5,800 different terms and definitions with almost no delay at all.

For the hypertext version of the *Glossary*, several menu bars were created that allow the user to "jump" quickly to other parts of the document. The menu bar across the bottom of the Web page (**Figure 10**) contains buttons that allow the user to jump to any other main section of the document.

Figure 10. The main menu bar for the hypertext version of FS-1037C.

The "Help" button of the menu, for example, displays basic information about how to use and navigate through the Web pages. The "Files" button of the menu provides the user with a menu of options that allows the user to download compressed versions of the *Glossary* files and other files related to the *Glossary*. The "Index" button of the menu displays Appendix B (of FS-1037C), which is a list of the major subject headings in the *Glossary*. Under each major heading in Appendix B is a list of terms associated with those subjects. Each term name in Appendix B is cross-linked to its respective definition and can be accessed in hypertext form with a mouse click.

The "A-Z" button of the menu (**Figure 11**) is equivalent to a Table of Contents for the *Glossary*. By clicking on the Foreword, Introduction, Appendix A, or Appendix B menu items, the user causes the selected section to be displayed on the screen.

If any of the letters of the alphabet is selected, the "A-Z" menu is replaced with a menu of term names beginning with that letter, as shown in **Figure 12**.

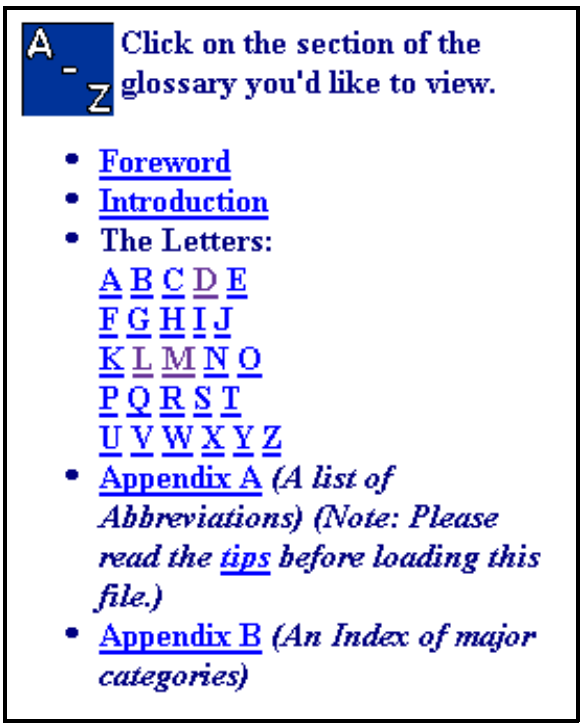


Figure 11. The "Table of Contents" menu bar.

From this list of term names for a particular letter, the user can scroll through the term names and select a particular definition to view. The ITS editors chose to display the hypertext version of the *Glossary* in a "frames" format on the screen. The frames, for purposes of discussion, are numbered as shown in **Figure 1**. When the specific term name is selected, the main viewing frame (Frame #1) is replaced by a display of the definition(s) for that term (**Figure 8**; frames are discussed in detail in Section 4.2).

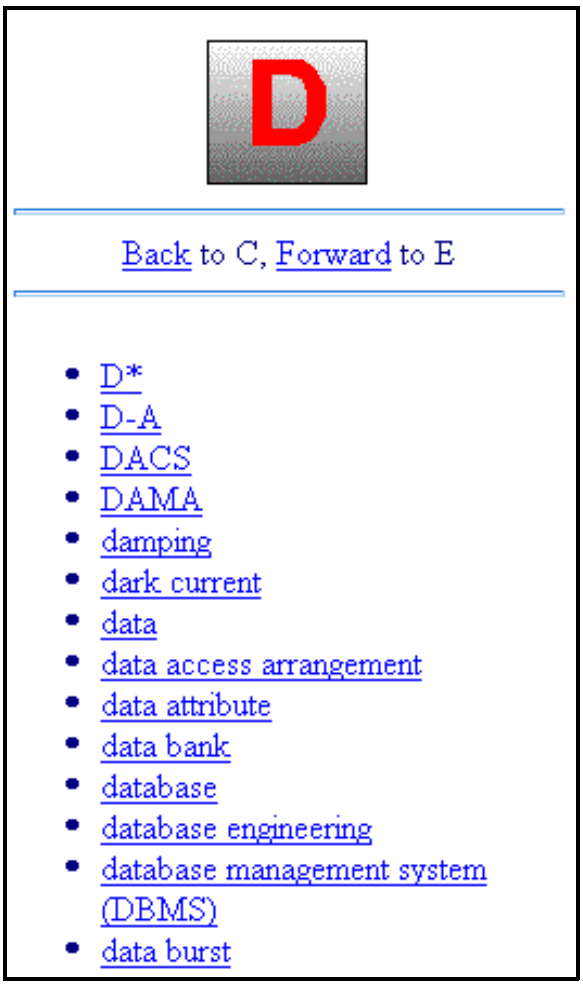


Figure 12. The beginning of the list of term names for the letter "D".

The list of term names remains on the screen in Frame #2, while the selected definition appears in Frame #1. This configuration allows the user to select and view term names that are in alphabetic proximity to the selected one. The list of term names (in Frame #2) is headed by menu items that allow the user to jump ahead to the next letter of the alphabet or to go back to the previous letter. (Of course, the user may also click on the "A-Z" icon to go to any

other, nonadjacent-letter section of the *Glossary*.)

3.9.2 Figures and Illustrations

Most of the 80 figures were created with one specific, commercial drawing package. The chosen drawing package uses vector graphics (**Figure 13**) to draw and to store the diagrams. Many other drawing packages use bitmapped (raster) graphics only; these bitmapped packages are commonly referred to as "paint" programs. The differences between vector graphics and bitmapped graphics are highlighted below.

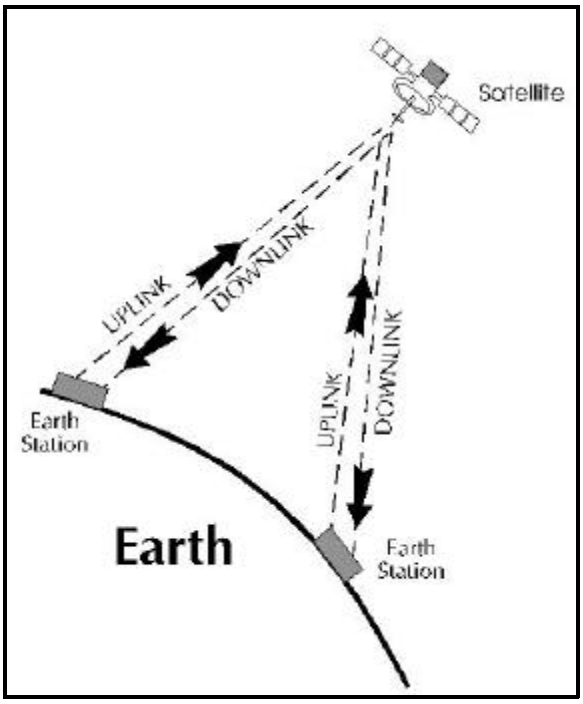


Figure 13. The *Glossary* image for "downlink" (vector graphics).

An image created with a typical paint program is static. A virtual grid is created in the computer's memory. At each intersection of the grid (called a "pixel"), a byte of information exists that indicates to the computer what specific color to display, on the computer screen, at that pixel location. The result is an image that is conceptually similar to a computerized scoreboard at a sporting event, with each pixel in a bitmapped image being equivalent to a single light on the scoreboard.

Unfortunately, a bitmapped image will look "grainy" if it is enlarged, just as if a person were to look at the computerized scoreboard while standing only a few feet in front of it.

The graininess problem can be reduced by creating a denser grid of pixels (See **Figure 14**, **Figure 15**, and **Figure 16**). But, since each pixel requires a fixed amount of computer storage space, the total space required to store a bitmapped image increases geometrically relative to its increasing grid density.

Vector graphic representations, on the other hand, create virtual images. Almost all parts of the image are represented by straight lines (vectors). Each line is stored in memory as a set of coordinates and associated attributes (starting point, ending point, color, width, *etc.*), and each line is drawn on the screen at the time when the user views the image. For moderately complicated images, a drawing stored as vector graphics requires significantly less space than is required for storage of the same image as bitmapped graphics (at usable resolutions). **Table 4** lists file-size requirements for selected bitmapped images and for a comparable vector graphics image.

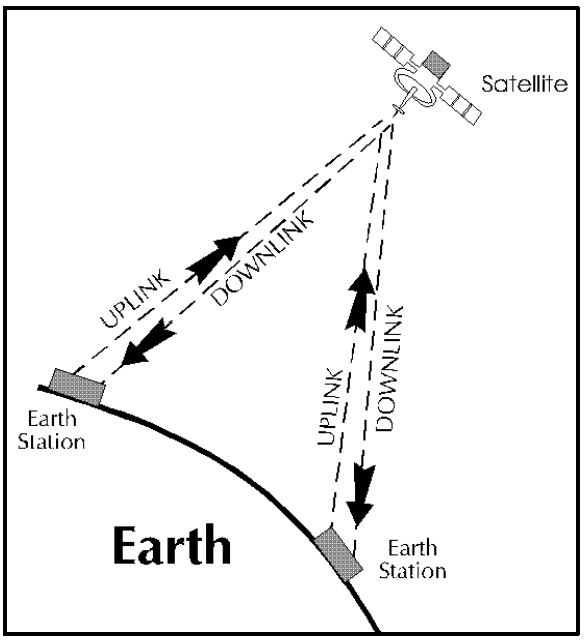


Figure 14. The *Glossary* image for "downlink" (550 × 600 pixel bitmapped graphics).

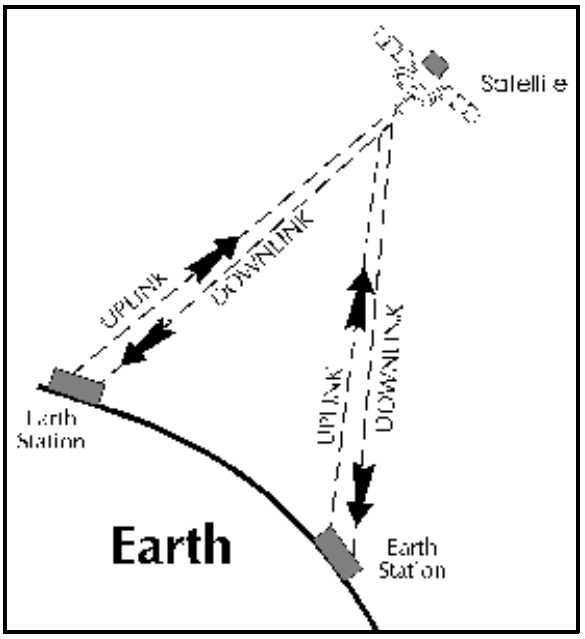


Figure 15. The *Glossary* image for "downlink" (275 × 300 pixel bitmapped graphics).

Vector graphics do not suffer from the graininess problem when expanded. Since all parts of the drawing are stored as a series of lines, the drawing is created "fresh" every time a user views it. It can be viewed at any resolution and the lines will be redrawn to look sharp at that particular resolution.

While all 80 of the images for the paper version of the *Glossary* were created and stored as vector graphics, the HTML and the network browsers had no way to create and display vector graphics for the viewer of a Web page. All images displayed by a network browser must be transmitted and displayed as bitmapped images. Therefore, each of the *Glossary* images was converted into bitmapped graphics.

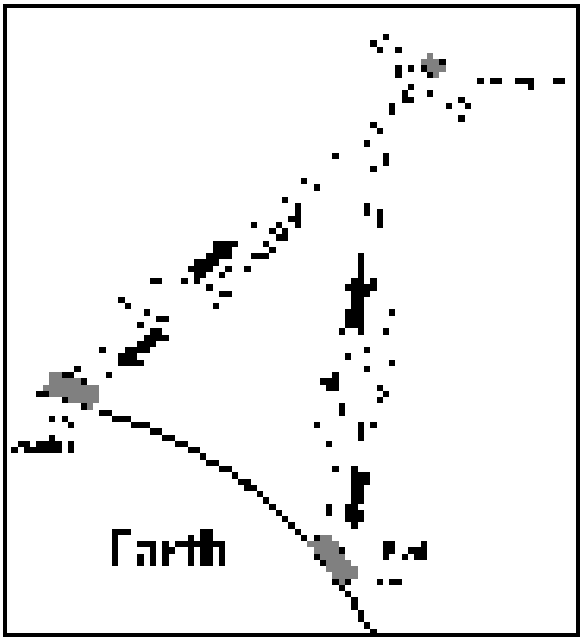


Figure 16. The *Glossary* image for "downlink" (90 × 100 pixel bitmapped graphics).

This conversion process was accomplished by loading the vector images into computer memory and displaying the image on the screen. Then, the image was captured from the screen, cropped, resized to the required image size, and stored as a bitmapped file. The appropriate dimensions for the final, cropped image were the subject of some debate.

A bitmapped image that is 10 pixels by 10 pixels, for example, contains 100 pixels and requires $(100 * t + h)$ amount of time to transmit to the user, where t is the amount of time needed to transmit one pixel, and h is the amount of time needed to establish the transmission (*i.e.*,the overhead). An image that is 20 by 20 pixels contains 400 pixels and would take approximately four times longer to transmit than an image that is 10 by 10 pixels.

Table 4 shows the file sizes for three possible resolutions for one of the images used in the *Glossary*. Note that the file size for **Figure 15** is approximately (allowing for some overhead bits within each image) nine times larger than **Figure 16**, while the file size of **Figure 14** is approximately four times larger than the file size of **Figure 15**.

Many users of the on-line *Glossary* are connected to the Internet by fast communication lines. For these users, the screen size of the images in the *Glossary* is not a major concern, with regard to time spent waiting to receive transmission of the entire image over the communications line. For these users, the time needed to receive a large bitmapped image is only a fraction of a second more than that required to receive a smaller image.

Table 4. Different File-Size Requirement for Storage of Different File Formats

Image File Format (pixels)	File Size; Disk Space Required (in bytes)
90 × 100 bitmapped (**Figure 16**)	10,280
vector graphics (**Figure 13**)	29,416
275 × 300 bitmapped (**Figure 15**)	83,880
550 × 600 bitmapped (**Figure 14**)	332,280

A user who is accessing the *Glossary* with a modem, through a commercial ISP, or through a local firewall, may not be able to receive large amounts of data in an acceptable amount of time. For these users, small images are very much preferred over large images.

The results of simple transmission tests are shown in **Table 5**. The increase in transmission time is not linearly related to the file size because of the overhead time needed to establish and maintain the transmission link.

Table 5. Transmission Times for Different File Resolutions

File Resolution	Time to Transmit via direct connection (seconds)	Time to Transmit via 14.4-kb/s modem (seconds)
90 × 100 bitmapped (**Figure 16**)	0.15	11.4
275 × 300 bitmapped (**Figure 15**)	0.18	46.6
550 × 600 bitmapped (**Figure 14**)	0.55	192.1

For a user connected via a direct communications line, the difference in transmission times is small enough to be almost unnoticeable (only a 0.4-s difference between the graphics displayed in **Figure 16** and **Figure 14**). For a user connected via a 14.4-kb/s modem, the receive-time difference can be as much as 3 min. Such a delay may be unacceptable to that user.

Although the graphics in **Figure 16** transmit much faster than the other two sample figures, their image quality is too poor to be used for the on-line version of the *Glossary*. The resolution of the graphics in **Figure 15** is also too grainy (in the editors' opinions) to be used.

Determining the appropriate pixel density for a presentation is always dependant on the final presentation medium. Amateur photography equipment has an equivalent resolution of approximately 4,000 dots per inch (DPI). The master copy of this report was printed on a printer that prints with a resolution of 600 DPI. Traditional offset printing (as was done with this report) and photocopying or reproduction with dithering cannot preserve resolution rates much more than approximately 400 DPI (which is why the reader of this report probably cannot see much of a difference between **Figure 13** and **Figure 14**). For comparison, it is noted that a 15-inch (38-cm; measured diagonally) computer monitor (set to display 1024 × 768 pixels), has a resolution of approximately 115 DPI.

As each bitmapped image was created for the *Glossary*, the smallest size—in the opinion of the *Glossary* editors—that was still readable on a computer screen, was used. This size was slightly different for each display image, and usually resulted in an image size approximately 200 to 400 pixels wide by 200 to 400 pixels tall. The display-image quality is between the qualities shown in **Figure 14** and **Figure 15**. The actual display-image resolution selected for the image shown in **Figure 13** was 305 × 343 pixels, which displays on a 15-inch (38-cm) monitor as an image approximately 3¼ × 3¾ in. (8.3 × 9.5 cm).

For the hard copy of the *Glossary*, all of the diagrams were created as black and white vector graphics. When converting the images for use with the on-line version, the editors added color to the diagrams, and, very occasionally, they increased the size of the

fonts within the diagrams to ensure readability of the displayed text at the chosen resolution.

3.9.3 Anchors and Links

Browsers support the concept of "linking," which allows the creator of the HTML document to highlight certain parts of the text that are destined to be accessed by a click of a mouse. When the user sees one of these highlighted areas, the user moves the cursor onto that area and presses the cursor button or mouse button. Doing this causes the browser to load a completely different section of the document (or an entirely different document). It is the responsibility of the editor of the HTML document to build the link and to select an appropriate destination for the user.

The on-line version of the *Glossary* relies on this kind of hypertext linking. The *Glossary* contains more than 5,800 terms that are all related to Government telecommunications efforts. Inevitably, many of the words used to define a particular term are, themselves, defined elsewhere in the *Glossary*. For example, the following term and its definitions, from the *Glossary*, contains four terms (that are underlined) that are also defined elsewhere in the *Glossary*,

> **graded-index profile:** In the core of an optical fiber, a plot of the variation of refractive index such that the refractive index decreases with increasing radial distance from the fiber axis.

In the HTML version of this definition, the user may click on any of the highlighted, underlined terms, which will cause the definition for that term to replace the current one on the screen in Frame #1. The user may, in turn, select any highlighted, underlined term within that new definition so as to view an almost endless series of related terms and definitions.

Each link in a hypertext document must have a corresponding anchor point. The anchor point is considered the destination point for a particular link. When a user selects the link, the Web system searches for the appropriate document, as determined by the link code, and, if that document is found, the browser displays that new Web page on the screen. The example in **Figure 17** illustrates the fundamental concept of linking and the technique for achieving it. In the example, when a user clicks on the "beam" hot spot within the upper left box, the "defs.html" file is retrieved and displayed, starting from the anchor point whose name is "beam."

Each link on a Web page points toward a single target anchor. An anchor, on the other hand, may have a limitless number of links pointing to it.

For a Web page to function correctly, the page editor must ensure that every link points to a valid anchor. Excess anchors, however, will not cause the Web page to function incorrectly and will be unnoticed by the user. The editors of FS-1037C tested the functioning of all links in the hypertext version of the *Glossary*.

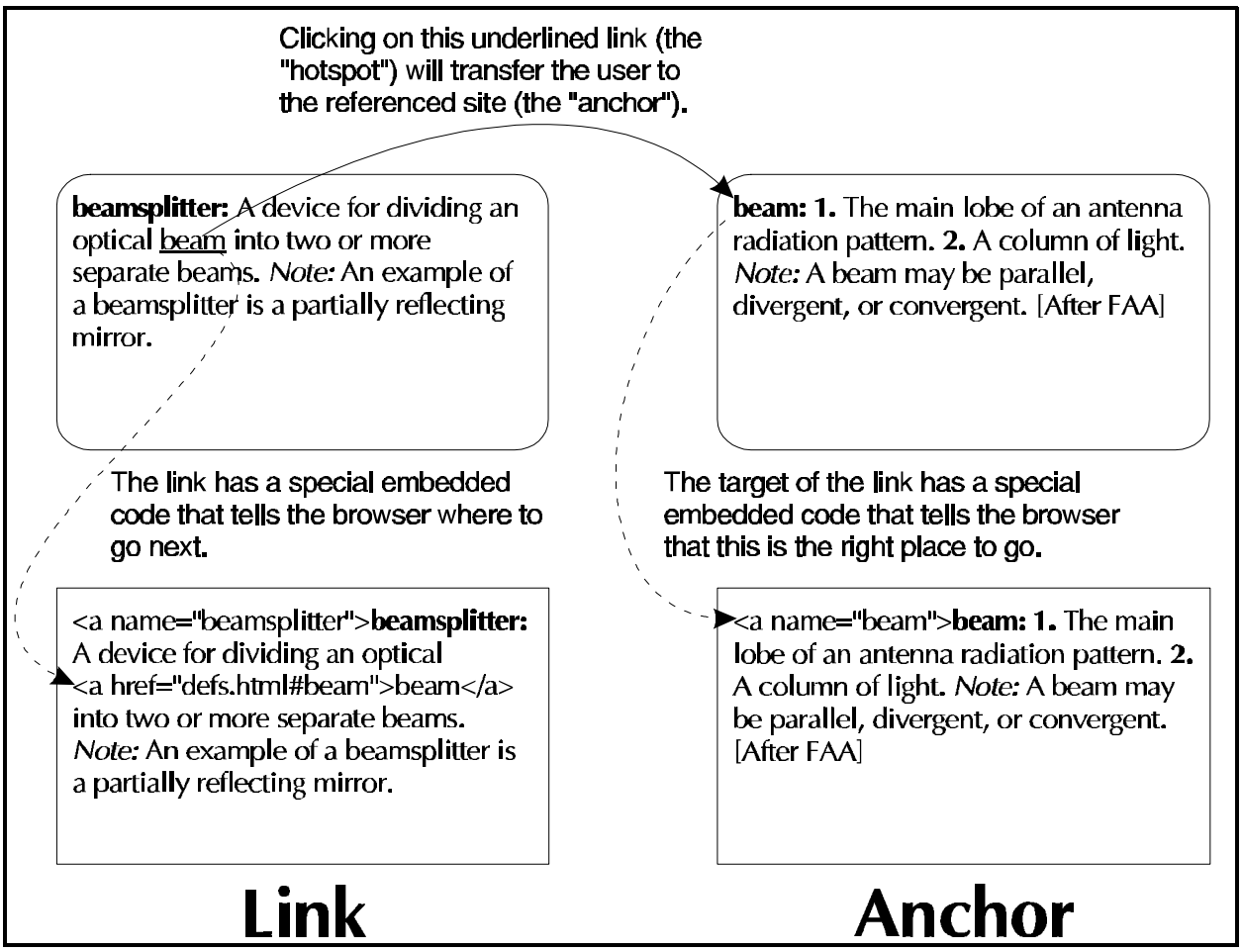


Figure 17. A conceptual diagram of links and anchors invoked when a user activates a hypertext definition.

encompassing 360° (2π radians). *Note 3:* Phase may be represented (a) in polar coordinates by $M \angle \theta$, where M is the magnitude and θ is the phase angle, and (b) in Cartesian coordinates, *i.e.* , an Argand diagram, as $(a + jb)$, where a is a real component and b is an imaginary component such that $\tan \theta = (b / a)$, where θ is the phase angle, and the magnitude, M, is $(a^2 + b^2)^{1/2}$ **2.** A distinguishable

Figure 18. Part of a sample definition showing the use of symbols.

3.9.4 Equations and Symbols

The equations, symbols, and special characters used for this *Glossary* are all displayed "in-line" (as described in Sections 3.8.5 and 3.8.6). When the user views a page with such graphics, the graphics are displayed in the correct position on the screen.

Each equation is displayed on its own line immediately following its mention in the text of the definition. The remainder of the definition follows the displayed equation.

The images of the symbols and special characters are displayed within the sentence or word where they are used. In **Figure 18**, the symbols θ, π, and ∠ are all displayed as in-line images. The symbols ° and ½ are displayed as HTML characters (see Section 3.8.3).

The size of the equations and symbol images was selected to match the displayed size of normal text on a typical browser. The results, however, were not always satisfactory. For example, the title of each term name is displayed in a font size twice as large as the basic font, and any in-line symbols included in a title will, therefore, be displayed half as large as the surrounding text.

4. WWW DISTRIBUTION

This section provides information on sources for FS-1037C, as well as a discussion of hardware and software considerations for the hypertext version of the *Glossary*.

The hypertext version of the *Glossary* can be stored in slightly more than 10 Mb of storage space. Additional features, such as compressed files (the 20 Mb of word-processing files compressed into approximately 7 Mb) and PDF files, increase the storage space needed to more than 30 Mb. Relative to many other Web sites, this is still a rather small set of Web pages. The pages and subpages of information need to be stored on a fast and reliable computer with abundant available disk space and high-speed connections to the Internet. Many commercial ISP's can supply storage space that meets these requirements. The hypertext files for the *Glossary* are stored on and are available from a computer at the NTIA Institute for Telecommunication Sciences, U.S. Department of Commerce in Boulder, Colorado.

4.1 Equipment Considerations

To view the hypertext files on the Web, the reader must use a browser on a computer that has reliable access to the Internet. The preferable connection method is a physical attachment to a local area network that is, in turn, connected via a fast communications line to the Internet. This kind of connection can provide very fast access to the *Glossary* hypertext files.

The other common method of connection is via a modem. Typically, the modem user connects his or her computer, via conventional telephone lines, to a computer host that is connected to the Internet. Transmission speeds for these modem connections are usually 14.4, 28.8 or 56 kb/s.

4.2 Software Considerations

In designing the Web pages for the *Glossary*, special considerations were given to the software that the user might use to view the pages. The whole development of the World Wide Web is still relatively new, and the

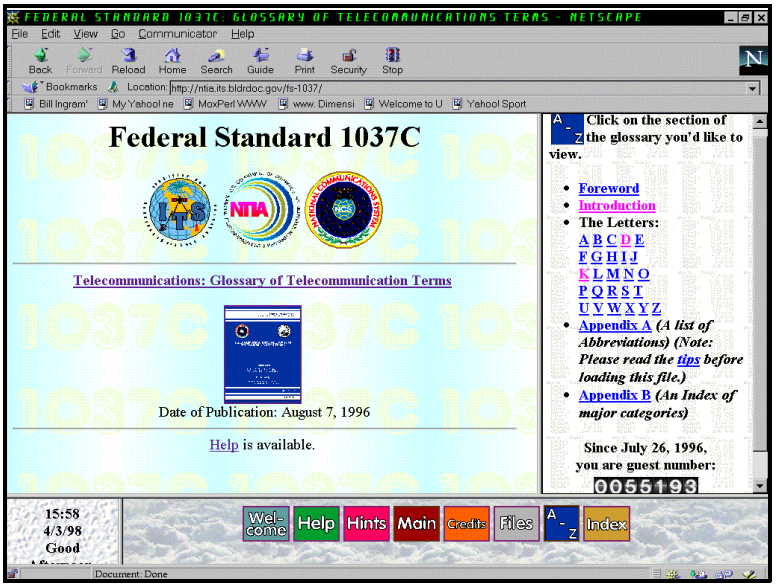


Figure 19. The main *Glossary* page (viewed with Frames).

coding used to create and display Web pages is still in the process of stabilizing.

As a result, many different versions of Web browsers are currently in use. Older versions of browsers cannot recognize some features that are standard on later versions. For example, in 1995, the designers of HTML added a new feature called Frames. The Frames feature allows the designer of a Web site to display several different Web pages on the screen all at the same time, each in a separate portion of the screen, called a Frame. The four Web pages, each in its own Frame, are shown in **Figure 19**.

In Frame #1 is the Main page (shown by itself as **Figure 20**). In Frame #2 is the "A-Z" page (also known as the Table of Contents, see **Figure 11**). The Clock page is shown in Frame #3, and the Menu Bar (see **Figure 10**) is shown in Frame #4. As the user selects various items from the menus, the contents of different frames are replaced (in frames 1 or 2) by the selected pages.

Most browsers released before 1995 will not be able to recognize the Frames features and will only display the information that is destined to be in Frame #1 (**Figure 20**).

The hypertext version of the *Glossary* has been designed to be usable by viewers with older browsers. Viewers with older browsers will be required to navigate the Web pages one at a time. The helpful menus (such as those in **Figure 10**) will not remain on the screen while the viewer reads other information. Demonstrations with such older browsers show that the progression through the *Glossary* pages is relatively slow, but is still straightforward and intuitive.

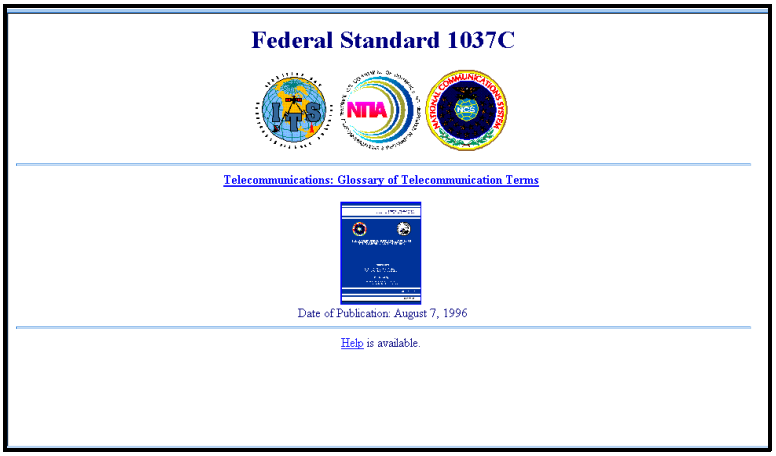


Figure 20. The main page from the *Glossary* (viewed without Frames).

Some users turn off all graphics from their display (to speed transmission time), and some other users are limited to using browsers that can display text only. While these browsers are quick and efficient, they cannot display any graphical information transmitted by Web pages. Web-page authors can include short descriptive text, for each graphic, that will be displayed when a browser cannot display the graphic itself. The *Glossary* does not contain such descriptive text, since it was determined that the graphics were too complicated to be represented by a short line of text.

Instead, viewers of the *Glossary* who are using a text-only browser will see an empty box where a graphic image would have been displayed. Most text-only browsers will, however, allow the user to download and save the graphic onto their hard drive. The user can then view the saved graphics file with any number of graphics-viewing programs.

5. CD ROM DISTRIBUTION

As part of the sponsored effort, FS-1037C , in hypertext, was also made available to computer users on CD ROM. While supplies last, copies of the CD ROM can be obtained (for domestic use only [this is due to restrictive licensing agreements for the browser software]) from

National Communications System
Attn: J. Orndorff
701 South Court House Road
Arlington, VA 22204-2198
phone: (703) 607-6204

After examining several different formats, the *Glossary* editors developed the CD to include both PDF files and hypertext files.

PDF files are simple variants of Postscript® files. Postscript® files are self-contained files that can be transferred to another user who will then be able to copy the file to any Postscript® printer to produce an exact copy of the original printed document.

PDF files can be printed to any kind of printer and can be viewed on any computer (with specialized reading software called Acrobat™). The Acrobat™ program is available, to any user, as freeware (*i.e.*, free of charge). To actually create the PDF files, the ITS editors used other specialized software (for which a fee was charged).

5.1 Software Considerations

The PDF and hypertext files on the CD ROM can be read by any user who has a properly installed PDF reader or hypertext browser, respectively. For this project, the editors judged that the user would be predominately responsible for providing and configuring his or her own viewing software.

Some consideration was given, however, to those users who might not have complete access to browsers and PDF readers. The CD ROM is preconfigured with the Windows® version of the freeware reader called Acrobat™. If the user is using a PC-based

computer and the Windows® 3.1 (or later) operating system, a batch file is available in the root directory that will begin running the Acrobat™ reader with the Table of Contents PDF already loaded.

Additionally, for users of the Windows 95® operating system, the CD ROM is configured to run the Acrobat™ program automatically when the CD ROM is inserted into the drive. A small "auto run" file (shown in **Figure 21**) was created to trigger this automatic loading.

```
[autorun]
OPEN=ACROREAD.EXE
CONTENTS.PDF
ICON=MISC\1037C2.ICO
```

Figure 21. The AUTORUN.INF file on the CD ROM.

Whenever any CD is placed in a Windows 95® CD drive, the operating system looks for a file in the root directory called "AUTORUN.INF". If this file is found, the system will begin to execute the command listed in the "OPEN" line. In this case, the file to be run is ACROREAD.EXE, which is the Acrobat™ reader program. Also, the icon listed in the ICON line (**Figure 22**) will be displayed whenever this CD is referenced by the operating system.

5.2 Hardware Considerations

As mentioned earlier, a single CD ROM can store as much as 660 Mb of data, and the files on a CD ROM can be read by many different brands and models of personal computers.

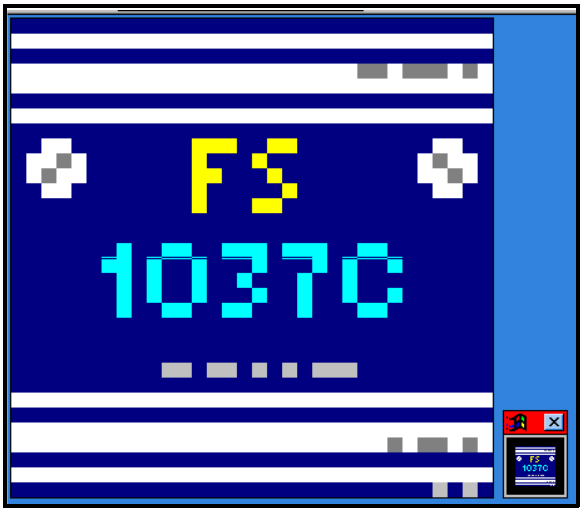


Figure 22. The icon displayed when the CD ROM is running.

In the early development and use of CD ROM's, a CD created by one computer operating system was generally unreadable by any other system. Over the last five years, however, most operating systems have been adopting standards that allow the reading of CD ROM's created by other systems.

Informal tests were conducted by ITS to ensure that the CD ROM used for this project could be read by Macintosh-, DOS-, and UNIX-based systems. The tests showed that each of these systems required up-to-date hardware and software to be able to read the CD. Several older configurations were not able to read some of the files properly.

The recipient of the CD product is expected to ensure that his or her computer hardware and software is up-to-date, allowing him or her to properly read the *Glossary* CD. Some copies of up-to-date, freely distributable software are included on the CD ROM for the convenience

of the user. It is not expected that such software can always be useful for all users.

6. CONCLUSIONS

Federal Standard 1037C on the Web and on CD ROM represents successful applications of HTML. In several ways, the large telecommunications glossary was ideally suited for presentation in hypertext format. The most significant advantage to the hypertext format is the rapidity with which users can jump from definition to definition without having to turn a precise (and often large) number of pages to arrive at the next desired definition. Yet, the very great size of the *Glossary* introduced one of the biggest hurdles in its hypertext development. The ITS editors of the *Glossary* surmounted that hurdle by using Perl scripts to generate automatically the many thousands of required hyperlinks in the large glossary.

Developing user-friendly HTML files for large documents and for wide distribution (via the Web and CD ROM) requires careful processing and special considerations. Among the issues required in the development of HTML files for FS-1037C, *Glossary of Telecommunication Terms,* were the following:

✔ ease of use (*e.g.*, ability to navigate forward, backward, to the index, to the table of contents, and to any selected paragraph/frame/definition, all with a minimum of steps and with a clear context map maintained in one of the frames);
✔ ability to display equations, symbols, and Greek letters;
✔ software for the automatic generation of the anchors list of defined terms;
✔ software for the automatic generation of links to defined terms;
✔ appropriateness of all automatically generated links;
✔ tests of the functionality of all links;
✔ picture size with ranges of monitor quality (granularity considerations);
✔ picture size with various speeds of modems (transmission-time concerns);
✔ adding color to the 80 illustrations to enhance information transfer; and
✔ development of sections providing a welcome, help, hints, credits, and addresses for hard copies

The HTML version of FS-1037C was placed on the Web for the first time on July 26, 1996. As of May 1998, the page has been accessed more than 75,000 times. The access rate has remained steady at about 75 accesses per day. While these numbers are not remarkable compared to other, commercial Web pages (that can have thousands of accesses per hour), what makes these numbers interesting is that many of these users are repeat users, indicating that the *Glossary* is serving them as a pragmatic resource.

Initial feedback shows that viewers of the site find it to be useful and intuitive. Some rough analysis of the first 3,400 initial accesses shows that at least 90 different users have accessed the *Glossary* at least 5 times each (some individual users have accessed the *Glossary* as many as 30 times each). This history shows that the *Glossary* is becoming a resource that viewers will use as a repeat on-line reference, looking up terms and definitions as they are needed.

This *Glossary* supplements advances in technology by providing a standardized, agreed-upon vocabulary with which to record and communicate these advances. Unless these advances can be communicated unambiguously, the advances will be lost, and all the benefits associated with them. FS-1037C, which contains standardized vocabulary for telecommunications terminology that is mandatory for all Federal Government departments and agencies, is a hallmark against that loss. The *Glossary* extends the benefits of common vocabulary to the designer, manufacturers, and to the world marketplace.

7. REFERENCES

[1] Garcia, "A new role for government in standard setting?" *StandardView*, vol. 1, no. 1, 1993.

[2] E. Gray, W. Ingram, and D. Bodson, "The ABCs of writing a technical glossary," *Technical Communication*, vol. 45, no. 1, pp. 23-32, Feb. 1998.

[3] E. Gray, and D. Bodson, "Preserving due process in standards work," *StandardView*, vol. 3, no. 4, pp. 130-139, Dec. 1995.

Abbreviations & Definitions

ASCII	American Standard Code for Information Interchange
CD ROM	compact disk read-only memory
DPI	dots per inch
FED-STD	Federal Standard
FS	Federal Standard
FTSC	Federal Telecommunications Standards Committee
GIF	Graphics interchange format
GSA	General Services Administration
GII	Global Information Infrastructure
HTML	Hypertext Markup Language
ISP	Internet service provider
ITS	Institute for Telecommunication Sciences
NCS	National Communications System
NCSA	National Center for Supercomputing Applications
NII	National Information Infrastructure
NTIA	National Telecommunications and Information Administration
NTIS	National Technical Information Service
PDF	portable document format
Perl	Practical Extraction and Report Language
URL	universal resource locator
Web	World Wide Web
WWW	World Wide Web

Appendix A: Vocabulary from FS-1037C[9]

browser: Any computer software program for reading hypertext. Note 1: Browsers are usually associated with the Internet and the World Wide Web (WWW). Note 2: A browser may be able to access information in many formats, and through different services including HTTP, FTP, Gopher, and Archie.

client: In networking, a software application that allows the user to access a service from a server computer, *e.g.*, a server computer on the Internet.

hypertext: The system of coding that is used to create or navigate hypermedia in a consequential manner.

HTML: Abbreviation for Hypertext Markup Language. An application of SGML (Standard Generalized Markup Language) implemented in conjunction with the World Wide Web to facilitate the electronic exchange and display of simple documents using the Internet.

[The] Internet: A worldwide interconnection of individual networks operated by Government, industry, academia, and private parties. Note: The Internet originally served to interconnect laboratories engaged in government research, and has now been expanded to serve millions of users and a multitude of purposes.

protocol: 1. A formal set of conventions governing the format and control of interaction among communicating functional units. Note: Protocols may govern portions of a network, types of service, or administrative procedures. For example, a data link protocol is the specification of methods whereby data communications over a data link are performed in terms of the particular transmission mode, control procedures, and recovery procedures. 2. In layered communications system architecture, a formal set of procedures that are adopted to facilitate functional interoperation within the layered hierarchy.

server: A network device that provides service to the network users by managing shared resources. Note 1: The term is often used in the context of a client-server architecture for a local area network (LAN). Note 2: Examples are a printer server and a file server.

World Wide Web (WWW): An international, virtual-network-based information service composed of Internet host computers that provide on-line information in a specific hypertext format. Note 1: WWW servers provide hypertext markup language (HTML) formatted documents using the hypertext transfer protocol (HTTP). Note 2: Information on the WWW is accessed with a hypertext browser such as Mosaic, Viola, or Lynx. Note 3: No hierarchy exists in the WWW, and the same information may be found by many different approaches.

[9]These definitions were taken directly from FS-1037C.

Appendix B: Selected Web Addresses

This table contains several WWW addresses (URL's) that are mentioned in this report or that are related to the subjects in this report (*Note:* Each URL must be typed exactly as it appears, including all capitalization and punctuation). All of these Web addresses were verified to be correct as of January 18, 2000.

Table B-1. Selected Web Addresses

Description of the Site	Relevant Section in this Report	URL
Federal Standard 1037C: *Telecommunications: Glossary of Federal Telecommunication Terms*	Section 1, page 1	http://glossary.its.bldrdoc.goc/fs-1037c
The National Communications System	Section 3, page 6	http://164.117.147.223/
The Institute for Telecommunication Sciences	Section 3, page 6	http://www.its.bldrdoc.gov/
The National Telecommunications & Information Administration	Section 3, page 6	http://www.ntia.doc.gov/
U.S. Department of Commerce	Section 3, page 6	http://www.doc.gov/
A Beginner's Guide to HTML	Section 3.5, page 10	http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html
The Web Developer's Virtual Library	Section 3.5, page 10	http://WWW.Stars.com/
Creating Net Sites	Section 3.5, page 10	http://home.mcom.com/home/how-to-create-web-services.html
A Short History of Internet Protocols at CERN	Section 3.6, page 10	http://wwwcn.cern.ch/pdp/ns/ben/TCPHIST.html

National Center for Supercomputing Applications	Section 3.6, page 10	http://www.ncsa.edu/
The Perl Language Home Page	Section 3.8, page 13	http://mox.perl.com/perl/index.html
The HTML Coded Character Set	Section 3.8.3, page 15	http://www.w3.org/pub/WWW/MarkUp/html-spec/html-spec_13.html
iso8859-1 table	Section 3.8.3, page 15	http://www.uni-passau.de/~ramsch/iso8859-1.html
Introducing HTML 3.2	Section 3.8.5, page 18	http://www.w3.org/pub/WWW/MarkUp/Wilbur/
How to Build Tables in HTML	Section 3.8.8, page 23	http://home.mcom.com/assist/net_sites/tables.html
The Netscape Frames Tutorial	Section 4.2, page 32	http://www.newbie.net/frames/
Adobe Systems Inc.	Section 5, page 34	http://www.adobe.com/
Traversing Cyberspace	Appendix A, page 38	http://www.ncsa.uiuc.edu/Indices/Discover/TraversingCyber.html
European Laboratory for Particle Physics	Appendix A, page 38	http://delinfo.cern.ch/