Internet opportunities for distributed construction information

Vanier, D. J.; Turk, Z.

Publisher's version / Version de l'éditeur:

International Journal of Construction Information Technology, 2, 3, pp. 45-65, 1994

NRC Publications Record / Notice d'Archives des publications de CNRC:
https://nrc-publications.canada.ca/eng/view/object/?id=7700cab5-0b2f-4e90-9bf7-5d51e0991fbe
https://publications-cnrc.canada.ca/fra/voir/objet/?id=7700cab5-0b2f-4e90-9bf7-5d51e0991fbe

INTERNET OPPORTUNITIES FOR DISTRIBUTED CONSTRUCTION INFORMATION

(DEC 1994)

SUBMITTED TO INTERNATIONAL JOURNAL OF CONSTRUCTION INFORMATION TECHNOLOGY

Dana J. Vanier, Senior Research Officer
Institute for Research in Construction, National Research Council Canada
Ottawa, Canada K1A 0R6
vanier@IRC.LAN.NRC.CA

Ziga Turk, Asst. Prof.
Faculty of Civil Engineering, University of Ljubljana
Ljubljana, Slovenia
ziga.turk@fagg.uni-lj.si

For more information, please contact:

Dana J. Vanier, Research Officer
Institute for Research in Construction, National Research Council Canada
Ottawa, Canada K1A 0R6
Vanier@IRC.LAN.NRC.CA
(613) 993-9699

Internet opportunities for distributed construction information. *International journal of construction information technology*, 1995, vol. 3, no 2, pg. 45-65, ilustr.

INTERNET OPPORTUNITIES FOR DISTRIBUTED CONSTRUCTION INFORMATION

ABSTRACT: The electronic distribution of construction documents such as building codes and standards, product information, technical literature, and research papers using magnetic and optical media is fast becoming commonplace. Wide Area Networks (WAN), and most notably the Internet, can provide new opportunities for the distribution of construction information. Generally, this technology is maturing rapidly and access to the Internet is now widely available. As a result, it is now possible to provide new services for the distribution of construction information to researchers, architects, engineers and builders. In the paper we describe the underlying technology and review the opportunities for its employment in research and practice in the construction industry.

Keywords: Internet, Wide Area Network, World Wide Web, Building Codes, Technical Information.

INTRODUCTION

Builders and the information society

Local area networks (LAN) are contributing to the use of information technology (IT) as a communication media between individuals and departments within organizations. Not only are LANs used to share expensive devices like printers, plotters, disks, modems, etc., they are becoming the major 'backbones' in modern offices for sharing information, documents, discussions, correspondence and work flow. With the development of high speed, digital telecommunications, these (electronically) closed organizations can open themselves up to the world. Computer-based communication backbones are spreading to encompass whole industries, countries and the entire world. In the developed world the creation of so-called "information highways" is becoming one of the most important national investments in the infrastructure, as reflected in the US National Information Infrastructure project (Table 1, item 1) and a European parallel (Table 1, item 2).

The traditional ways of spreading engineering and architectural knowledge electronically are to use magnetic (diskettes) or optical media (CD-ROMs, Compact Disk Read Only Memory) to distribute software and various construction documents such as building codes and standards, product information, technical literature, and research papers (Bourdeau, 1991; CCB, 1993). However, finding out about which products exist, or how can they be obtained, is still an obstacle to the rapid, effective dissemination of construction information. This is followed by a time lag to have the products delivered to the client. In addition, even stable information like the ones listed

above has to be constantly updated. Therefore, magnetic or optical-based information is not that appropriate for dynamically changing information, like that created within a building project.

Opportunities for information dissemination

The Internet provides new opportunities for the distribution of construction information. The largest wide area network existing on Earth is the Internet (Table 1, item 3). It is an inter-network consisting of over 20,000 networks that connect over 2.2 million computers and is estimated to be used by 20 million people (Table 1, item 4). Originally started as a communication tool between military contractors in the USA, it is fast becoming a communication standard for everyone dealing with computers and computer networks (Krol, 1993). Generally, the technology is maturing rapidly, even exponentially. By relaxing the "acceptable use policy", Internet access is not only limited to government, educational and research organizations, it is now opening up to the commercial users. In fact, using free software, Internet is accessible by anyone with a personal computer or a terminal and a modem. As a result, the Internet can provide the necessary backbone for the distribution of construction technical information to researchers, architects, engineers and builders.

This paper proposes using the Internet to access the abundance of construction information that is distributed around the world. Included in the benefits are ease of access, instantaneous updates, and rapid information exchange. Other opportunities include the availability of multiple editions of documents, friendly standardized user interfaces, and transparent information access. The Internet also provide opportunities to link into other computer programs on distributed servers and to efficiently browse and search large text documents.

The paper describes four types of Internet access: (1) access through a *gateway service*, (2) *remote modem access*, (3) *Point to Point Protocol* (PPP) or *Serial Line Internet Protocol* (SLIP) access, and (4) *direct Internet access*. In addition, the paper describes six different Internet services: (1) electronic mail (e-mail) and List Servers, (2) Usenet, (3) Telnet, (4) FTP (File Transfer Protocol), (5) Gopher, and (6) World Wide Web (WWW). Although the types and services all appear to be completely different in functionality, they have surprising similarities such as quickly sending sparse messages around the world.

TECHNOLOGIES

General

Physically the Internet consists of computers that are connected to the network through a network adaptor. The network consists of the wires and the hardware that modulates and routes digital signals from one computer to another. Logically the computers are grouped into networks

- one network being limited to an organization or a department. Each network has a unique number, as does each computer within the network. The data sent from one computer to another are broken into fairly small packets which are then broadcast into the network. Using these addresses, the routing software sends the packets, perhaps over different routes, to the receiving computer, which then reassembles them into the original data. An even more abstract view would reveal that the packets are catalogued by type and that packets of a particular type (*e.g.* e-mail, files, hypertext) are sent, and expected, by different applications on the sending and the receiving end. It appears as if each application uses a different "channel" of the network.

To understand each other's communication the sender and the receiver must use a standardized protocol. The *lingua franca* of the Internet is the Transmission Control Protocol (TCP), often called the Internet Protocol (IP) and identified as TCP/IP. The basic vocabulary of TCP/IP is mainly preoccupied with addressing, acknowledgements and transmission of raw data. However, there are several application protocols that specialize in the exchange of particular types of data such as files, hypertext information, or e-mail. The following subsections describe the various types of access available and the most useful Internet services. These have many common characteristics:

- A service consists of a client program, server program and an application protocol they both use to communicate. The computer running the server program is often called "host computer" or "remote host".
- The client program's task is the user interface and the presentation of the information. The server program's main task is to provide the information. They have client-server architecture.
- As a rule, client programs work with only one host server at a time, but can appear to be working with many host servers. Host servers can be contacted by more than one client program at a time, and
- Client programs and host servers communicate with each other using a specialized protocol. The names of the protocol end with letter P for protocol (*e.g.* FTP, HTTP)
- The client program is a normal program. However, the host server software is a deamon, that is, it responds automatically to inquiries (program names end with D for demon, *e.g.* FTPD, HTTPD). Therefore, host server machines must be multi-tasking machines, where more than one program can run at a time. Multi-tasking is not necessary for client programs.
- All services described in this paper are interactive applications, but do not work in real time - in theory no connection is too slow to use any of the services.
- Access to the host server is generally restricted: a client program must have permission to log on. But the majority of host servers also offer public access to anyone.

- Since there may be hundreds of thousands of host servers in the world, there exist specialized programs to find information offered by these host servers (Table 1, item 5).
- For all the listed services, high-quality freeware or shareware client programs exist for the DOS/Windows™, Macintosh™ and UNIX™ workstation (Table 1, item 6). Host server software of good quality is also available for multi-tasking operating systems. An ideal host server machine would run UNIX™ but there are good implementations for most services for DOS/Windows™ (Table 1, item 7) and Macintosh™ (Table 1, item 8), as well.

It is relatively easy to create a host server on the Internet; as a result, a vast quantity of information is being made available. This also spawned the creation of new host servers which specialize in finding the right information in specific domains.

Types of Internet access

Before describing the software part of the Internet, it is necessary to describe the types of Internet access. These are hardware issues and are described in increasing order of speed, functionality, and cost.

Gateway access means that the user is directly connected to a proprietary information service such as Compuserve™; in this type of access the user is only remotely connected to the Internet. Typically, the user is restricted to the services offered by the gateway; for example exchanging e-mail, accessing Usenet News or accessing data originally from the Internet that has been translated to the service. As of today, there is a limited number of *gateway services* that provide access to the more robust Internet services; however, demand is growing for additional functionality and the gateway companies will respond in the near future. Some already provide full *terminal access* as described in the following paragraph.

Terminal access allows the user to access a host server in a "terminal emulation" mode. That is, the user is connected to the host server but the software that connects to the Internet is resident on that host server, NOT the user's computer. *Terminal access* is acceptable for some Internet services such as e-mail, Usenet News, Telnet, and to some extent for Gopher and WWW. *Terminal access* is unacceptable for FTP because of the slow transfer speeds of low speed modems. Typically *terminal access* means that an alphanumeric (*i.e.* text) user interface is used and, as a result, the interaction is less friendly and more confusing to novice users. In addition, only alphanumeric information (text) can be accessed interactively, NOT the associated graphics or multimedia information (although these can be downloaded to the user's computer and displayed using other software applications). Normally, all that is required for *terminal access* is a terminal or computer and a high speed modem (2400 bits per second (bps) or more). Typically,

the initial cost for this type of access is to the order of $10.00 to $50.00 per month; the full monthly cost is dependent on the connect time and amount of data transmitted (Table 1, item 9).

Direct access is the most complete type of Internet access. *Direct access* means that the user's computer becomes one of the tens of millions of computers on the Internet and can (providing it has the proper software) enjoy or offer all the services available for the Internet. Technically there are two ways to give a workstation *direct access* to the Internet: (1) the workstation is connected to a LAN which is part of the Internet or (2) the workstation individually connects and becomes part of the Internet.

Direct access for individual workstations is enabled through Point to Point Protocol (PPP) or Serial Line Internet Protocol (SLIP). PPP or SLIP enables the client program to use a high speed modem (9600 bps or greater is recommended) and standard telephone lines to connect to an Internet node and become part of that network. After the connection has been established the workstation acts as if it is another computer on the same LAN as the dialed host server. Such access type is appropriate for home usage or for smaller companies that do not have a LAN or do not need continuous Internet services.

Direct access through LAN means that the user can continuously, interactively and instantly access the Internet and all its functionality. This is possible because LANs have a permanent connection to the Internet, typically over leased lines, optical or cable networks. These lines are shared by all workstations at the same time. The access speeds range from 56 kbps (kilo or thousands of bps) connections to T1 connections at 1.5 mbps (million bps) to T3 connections at 40 mbps (Carroll and Broadhead, 1994). Initiatives in the USA such as the National Research and Education Network (NREN) say that networking at "Gigabit speeds" is possible in the near future (Carroll and Broadhead, 1994). *Direct access through LAN* is mandatory for all information providers and is recommended for larger organizations with continuous Internet access requirements.

INTERNET SERVICES

E-mail and List Servers

E-mail is rapidly becoming the standard for communication in the 1990's. Not only does it permit effective communication within a LAN; with the proper hardware and software, there is also the possibility of transmitting information to the entire Internet. E-mail systems involves: (1) a user program for reading, writing and archiving mail, (2) a postoffice computer with a mailbox for each user and an out-box for messages, and (3) a network of computers which connect the postoffice computers. Typically a user writes a message and drops it in the postoffice's mailbox; the mailbox locates the destination computer and uses SMTP (Small Machine Transport Protocol)

to forward the message to the receiver's postoffice, and the receiving postoffice then places the message into the receiver's mailbox. Some postoffice software notifies the receiver when new incoming mail has arrived.

There is not much more to say about e-mail except that it is fast (taking minutes to send messages around the world), efficient (less labour-intensive than fax technology), reliable (if a message is not delivered in three days, the message is returned to the sender), and inexpensive (free to most Internet users). One can also keep an audit trail for all correspondence received and sent. What more can you ask of a communication tool? Typically Internet e-mail addresses (also known as a RFC822 address as opposed to a X.400 address) are identified by the "@" (pronounced 'at') sign. The first author's e-mail address is 'Vanier@IRC.LAN.NRC.CA'; an e-mail address is not case-sensitive, so the address can be in upper or lower case characters. The address is in uppercase so there is no confusion between the L's or I's and the number "one". The address actually means that the e-mail address is at a postoffice named "IRC", on a 'LAN' router, and in the 'NRC' gateway in 'CA'nada. Most commercial postoffice software permits sending e-mail messages outside the LAN; this is normally found in a SMTP option for addressing messages.

List servers are a simple extension to e-mail. Behind list server's e-mail address is a robot program that distributes any incoming message to a list of subscribers. It is an extremely efficient method to disseminate information; however, a flood of electronic junk mail is possible, and this flood might draw attention away from important messages. Also when a user is subscribed to similar lists, it is possible to receive multiple broadcasts of the same message.

Usenet News

As opposed to e-mail, which is typically one-to-one communication, a world wide conferencing system known as US Net News offers thousands of discussion groups on all imaginable topics (Table 1, item 10). Civil engineers should have a look at news:sci.engrng.civil. Anyone can post a message to a discussion group and the whole audience of the discussion group is able to read it and comment. Many people only think of Usenet as the Internet, but in the authors' views Usenet is a glorified bulletin board.

Telnet

Telnet enables users to access a remote host server on the network as if it were their own. This means the user "logs onto" a host server and runs a program on that host server; the input-output is redirected to the user's client program. Since Telnet only transfers data that is typed on the keyboard or shown on screen (which are both limited by the speed of human typist or reader) even the slowest communication lines are fast enough for this type of Internet service.

The use of Telnet is universal, from programming or using complex engineering software on a remote supercomputer to queries into information systems provided by travel agents, airlines, libraries and government agencies. Each of those programs offers a different user interface to access the information and limits the options to transfer that information to the user's computer. Typically Telnet is used with the first two types of Internet access, namely access through a *gateway service* or *remote modem* access.

File Transfer Protocol (FTP)

FTP is a service which transfers files between the client program and the host server. FTP can move files, megabytes in size, from the client program to the host server, or retrieve files from the host server. One uses commands like 'put', 'get', or 'cd' to transfer files or to change directories. Typically, password protection is available to prevent unauthorized use of the host server. Anonymous FTP access (Table 1, item 11) is available on many host servers around the world.

A prime use of anonymous FTP is to access shareware, graphical images, and updates to software applications. A few host servers in the world contain directories of files from many of the FTP servers ("FTP sites"). These host servers may be queried as to where a particular file can be obtained: the service is called Archie (Table 1, item 12).

Gopher

Users of FTP must spend considerable time changing directories and viewing their contents (using respective commands such as 'cd' and 'dir'). So in 1986, at the University of Minnesota, they developed a host server that automatically creates a menu from the file names in a directory. The menu is displayed by the client program with the file names as the menu options. If an option is selected, then the host server either sends the contents of the selected directory or the selected file. Because of the simplicity to organize information for Gopher (named after a University of Minnesota mascot) there are now thousands of Gopher servers around the world (Table 1, item 13).

Later on Gopher servers provided access to Telnet sessions (a menu entry would start a Telnet application), to FTP archives, to Archie and WAIS (see below) search engines and to Gopher servers on other computers. This last feature created the so-called "gopher space" which consists of all Gophers servers in world and enables the user to access any information on any Gopher server by moving through the menu structures. The more advanced Gopher+ protocol includes multimedia capabilities and multiple representations of the same file so that client programs with only alphanumeric capabilities show only the text portion of documents while more sophisticated client programs can have a full Graphical User Interface (GUI).

Finding information in Gopher space is eased by a service called Veronica (Table 1, item 14). Similar to Archie, a number of Veronica servers can index information on all registered Gopher menus and files, and can allow Boolean searches of that information - the results are returned as a Gopher menu.

WAIS is a full-text database system produced by Thinking Machines Corp, Apple Computer and Dow Jones and placed in the public domain (Table 1, item 15). Full-text databases allow retrieval of documents or document parts by specifying any of the words which occur in them. On the host server the information is indexed using full-text inverted index techniques. The client program queries the host server giving a list of words as input. WAIS returns pointers to the appropriate information and ranks the documents found, positioning what it considers the best match at the top.

A typical use of WAIS is rapid searching of vast quantities of poorly-structured or heterogeneous information. WAIS can search texts, DXF files, e-mail archives or FTP sites at the same time, and is often used to search *all* the information on a particular host server.

World Wide Web (WWW)

WWW was initiated at the European Laboratory for Particle Physics (Table 1, item 16) in 1989/90. It serves hypertext and hypermedia documents and interfaces to other services on the Internet (Krol, 1993). Hypertext is fast access connection to electronic documents. This technology permits users to browse quickly over text and graphics, while permitting sophisticated keyword searches and other administrative functions such as history trails. Combined with the Internet, hypertext because a powerful browsing and searching tool.

The native format of the WWW documents is the Hypertext Markup Language (HTML). It is a markup language based on the Standard Generalized Markup Language (SGML). Automatic converters from various word processing formats such as Rich Text Format (RTF) to HTML already exist and are free of charge (Table 1, item 17). The protocol used between the client program and host server is the HyperText Transport Protocol (HTTP). Similar to Gopher, the HTTP server can automatically generate a HTML document from file/directory structure and serve it to the client program. A graphical client program is recommended for a comfortable use of the WWW. Perhaps the best known client program is NCSA Mosaic (Table 1, item 18), developed by the National Center for Supercomputing Applications (NCSA). It supports both graphics and text in a hypertext environment (NCSA, 1994), and is available for many different platforms. Other high quality graphical client programs include Air Mosaic™ from Spry Corporation (Table 1, item 19) and Netscape™ from Mosaic Communications Corporation (Table 1, item 20). Alphanumeric client programs such as Lynx provide acceptable browsing capabilities.

Browsing the WWW employs standard hypertext functions like moving through a hyperlink to another location, going back to a previous location, or reviewing the history of locations visited. The hyperlinks may be internal, ones that refer to a different point in the same document, or external and point to some document located virtually anywhere on the Internet. The links are referenced by the Uniform Resource Locator (URL) and Uniform Resource Name (URN) which give a unique name to any information on the Internet. They also include the service type, host server machine and directory/file name on that machine. Examples of URLs can be seen in the Table 1.

1 US Project	gopher://is.internic.net/11/infoguide/about-internet/nii/
2 Europe	http://www.earn.net/EC/bangemann.html
3 Internet	http://www.internic.net/infoguide/gopher/about-internet.html
4 History	gopher://is.internic.net:70/00/infoguide/about-internet/history/timeline
5 Searching	Perhaps the most complete is: http://cui_www.unige.ch/meta-index.html
6 Mosaic	http://www.ncsa.uiuc.edu/SDG/Software
7 Servers	http://www.rpi.edu/Internet/Guides/decemj/internet-tools.html
8 MacHTTP	http://www.uth.tmc.edu/mac_info/machttp_info.html
9 Gateway services	Compuserve™, Mindlink™, or Hookup™ are current examples of commercial services.
10 Usenet	http://gopher.ictp.trieste.it/news/News.html
11 FTP Anonymous	Anonymous FTP means that the users log on with "anonymous" as the user identification and their Internet Protocol number as the password.
12 Archie	gopher://istge.ist.unige.it
13 Statistics	Annual rate of growth for Gopher traffic for 1993 is 997%. This information was obtained from treese@crl.dec.com on the Internet.
14 Veronica	Very easy rodent-oriented net-wide index of computerized archives. gopher://veronica.scs.unr.edu:70/00/veronica/how-to-query-veronica
15 WAIS	Located at http://www.vuw.ac.nz/who/Nathan.Torkington/ideas/www-primer.html
16 WWW at CERN	The World Wide Web Initiative, European Laboratory for Particle Physics, Geneva, Switzerland. http://info.cern.ch/hypertext/WWW
17 rtf2html	ftp://ftp.cray.com/src/WWWstuff/RTF/rtftohtml_overview.html
18 NCSA Mosaic	http://www.ncsa.uiuc.edu/SDG/Software/MacMosaic/MacMosaicHome.html
19 Spry	http://www.spry.com/
20 Netscape	http://home.mcom.com/
21 dates.ce	http://audrey.fagg.uni-lj.si:80/cgi-bin/madame_d/ICARIS/dates.ce
22 HTML	http://www.pcweek.ziff.com/~eamonn/crash_course.html
23 WinWord macro	http://audrey.fagg.uni-lj.si:80/ICARIS/tvdmi93/hpaper.htm. A list of tools is available at http://info.cern.ch/hypertext/WWW/Tools/ Word_proc_filters.html
24 GREP	Grep is a UNIX utility to find lines or paragraphs containing matching words or regular expressions.
25 ICARIS	http://audrey.fagg.uni-lj.si/toc-fagg.htm
26 IADS	Host: sgml1.ex.ac.uk, Login: anonymous ftp, Location: /iads.
27 FastTag	They are available from Avalanche Development Company and parse most common word processed documents to create SGML output.
28 Omnimrk	Available from Exoterica, Ottawa, Canada
29 Tagwrite	Available from Zandar Corporation
30 Rainbow	Rainbow is a SGML-like neutral format from Electronic Book Technologies, in conjunction with several key SGML vendors/promoters
31 rtf2html	ftp://ftp.cray.com/src/WWWstuff/RTF/rtftohtml_overview.html
32 Plinth	http://www.aiai.ed.ac.uk/~andrewc/plinth/datasheet/contents.html
33 STEP	file://ftp.cme.nist.gov/pub/step
34 Express	http://www6.informatik.uni-erlangen.de:1200/Express/exp2html.html
35 dates.ce	http://www.fagg.uni-lj.si/ICARIS/dates.ce/
36 ICARIS	http://www.fagg.uni-lj.si/ICARIS/
37 J. Brown	gopher://una.hh.lib.umich.edu:70/00/inetdirsstacks/archi:brown
38 ICARIS	http://www.fagg.uni-lj.si/ICARIS/home-ptr.html
39 NICE-L	http://www.fagg.uni-lj.si:80/ICARIS/NICE/

Table 1: List of referenced Internet servers

Unlike most other services the WWW is flexible enough to allow more than simple fetching or searching of information. The HTML+ specification enables the creation of forms through which the client program can send large volumes of textual information to the host server (See Fig 1). The HTML forms can be used as front ends to databases, e-mail or conferencing systems, allowing any client program to submit changes and additions to information on a host server, when applicable. This technology promises to provide a universal easy-to-use, platform-independent interface to all the information on the Internet. A good example of the use of HTML forms is a listing of civil engineering conferences provided on ICARIS (Table 1, item 21). The conference listing is searchable on dates provided by the user, as demonstrated in Fig 1. In addition, the 'dates.ce' service e-mails the search results to various List Servers at the beginning of each month; thereby notifying civil engineers of upcoming conference and workshop events and deadlines.

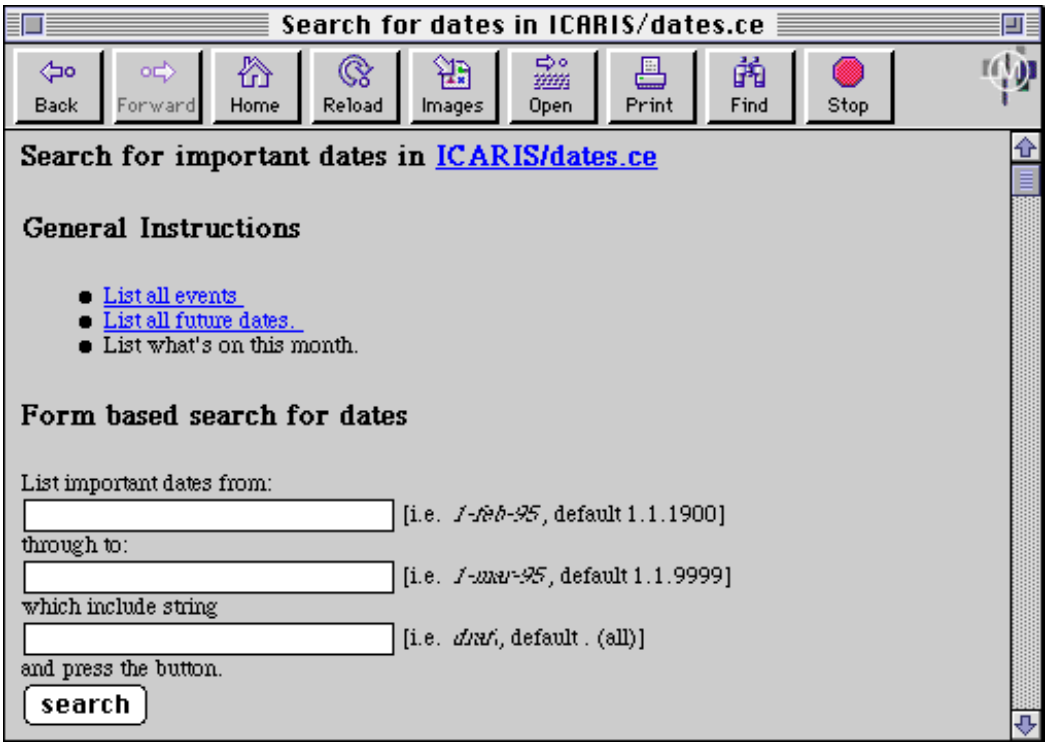

Figure 1: HTML form to add information to on-line databases

Indexed searching is also an important feature in WWW. There are two main approaches to the difficult task of creating search engines for hyperspace. The spider-based approach exhaustively navigates the WWW, following the links it finds and indexing the resulting information in one large central index. Whereas, the list-based engines rely on lists of services that are reported by the information providers to a central authority where they are indexed.

Formats of information for Internet distribution

The example in Fig 2 provides some of the markup for HTML for the ICARIS Home Page shown in Fig 9. The angle brackets delimit the beginning <xxx> and the end </xxx> of the elements, and H, B, and HREF respectively mean header, bold and hypertext reference. The remainder is self-explanatory:

<HEADER>	Header information
<TITLE>ICARIS – CIC research network	Title Page Name
</TITLE>	End of Title
</HEADER>	End of Header
<BODY>	Start of Body
<IMG SRC=/icon/icar_483.gif>	Image link
<HR>	Divider line
<A HREF=w78/><IMG SRC=/ICARIS/cib.gif></A>	One image link
...	Other image links
<A HREF=w78/>CIB-W78</A>	Text links for non-
...	graphical user
<B>ICARIS</B> is an experimental research	Bold ICARIS
network related to integrated CAD in civil	
engineering and architecture. It is	
described in the <B> ICARIS Initiative </B>	Bold
<A HREF=file://audrey.fagg.uni-lj.si/pub/	Hypertext Link
icaris.zip> paper </A>.<P>	

Figure 2: The HTML source for the page rendered in Fig 9

The rule of a thumb for making information available on Internet is: if it is in the computer form and if it uses an industry standard format it can be used by others who have access to the same tools. The services themselves do not care about the content of the data. Some of the more popular formats include:

- ASCII text: The most reliable format, but also the most primitive.
- Post Script: It can be viewed by viewers that are available free of charge; it can be printed by most printers. It is a good format to distribute documents intended only for printing or viewing. It is difficult to create valid PostScript™ files on DOS/Windows™. PostScript™ files are large and must be compressed.
- GIF: This is the best format for raster images; it is compact and unambiguously understood on different platforms.

- JPEG: This format for raster images incorporates compression that deliberately loses some information (that human eye does not notice it) and thus results in even smaller files.
- HTML: This is the formatting language for hypertext on WWW. It is easy to learn (Table 1, item 22) and convertors from Windows™ Help format, TeX™, ASCII, Wordperfect™, Word for Windows™ are available.. The second author wrote a WinWord macro that automatically converts well-written papers to html (Table 1, item 23).
- DXF, HPGL, *.*: Any format is suitable so long as the receiving end is expected to have a program that can view that format. It is important to stress that the host server does not restrict the format to be displayed, it is normally the type of client program software that limits the input/output.

INTERNET PERSPECTIVES FOR CONSTRUCTION

This section outlines the existing and prospective uses of the Internet for the distribution of information concerning construction research and practice.

IRC Audience

IRC has been distributing technical information for the construction industry for the past 50 years. The large majority of this information is in the print medium. The information supplied can be generally categorized into the following areas: marketing, practitioner technology diffusion, and research information. Marketing, in this delimitation, can be viewed as self-promotion for future revenues; practitioner technology diffusion deals with the distribution of construction information and knowledge to the industry, and the research information is that published for the research community. A limited practitioner base use the scientific publication network. This paper concentrates on the last two categories of technical information, recognizing that marketing has special conditions that are best left to experts in that field.

The potential audience for technical information can therefore be divided into practitioners and researchers. Currently, researchers generally have *direct access through LAN* to the Internet through their research centres or universities, whereas practitioners are becoming "connected" using *gateway access* or PPP and SLIP. So both groups have to potential to benefit from Internet distributed information.

IRC On-line

Browsing the current WWW servers on the Internet provides a good view of what information can be readily made available. Organizational data, personnel lists, curriculum vitae information, and project descriptions are typical examples. This is information that is readily available and easy to maintain. In the case of a research laboratory the following data are of interest to both

practitioners and researchers: researcher names, location coordinates, research fields and recent publications, as well as organization structure, laboratory description, equipment resources and current projects. Information of interest to practitioners includes symposia descriptions, topics of lectures, titles of publications and descriptions of products.

The home page of the IRC WWW server illustrates the potential of accessing this type of general technical information:

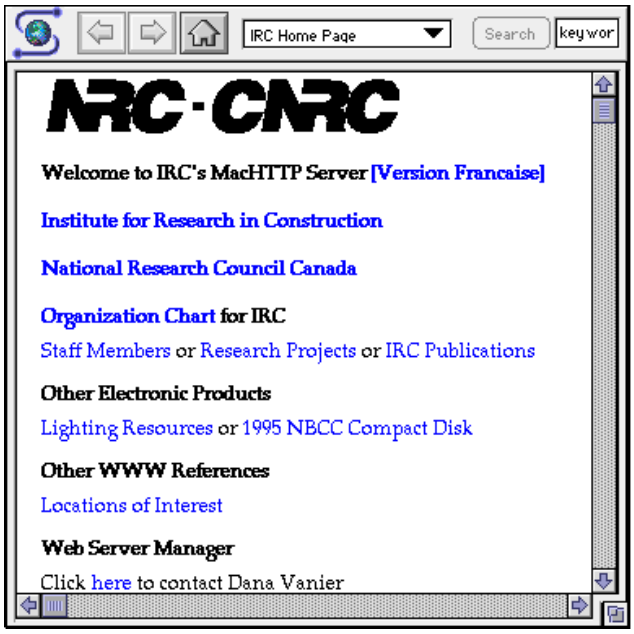

Figure 3: IRC Home Page and associated documents

As can be seen from Fig 3, multilingual editions of these documents are also possible just by adding a pointer to the "Version Française". Pointing at the Organization Chart in Fig 3 produces the screen in Fig 4.

Included with these major functional elements are details about these units as well as listings about the staff members, equipment resources, and current projects. In turn, the staff references include pointers to the functional units, projects and publications. In Fig 5, information about a staff member in the Technology Diffusion Service of the Industry Liaison Branch is requested. In Fig 5, the first reference to the Turk/Vanier paper (TVDMI93) points to the ICARIS server mentioned earlier. This demonstrates the full power of distributed information.

This design follows the structure of an information base developed at IRC for a CD-ROM application (Worling et al, 1992). As can be seen in Figs. 4 and 5, this is not only valuable to a research organization, but any corporate structure would benefit from the same detailed information about their operations.

Currently the information on the IRC WWW server can not be searched using the indexed techniques described earlier, but we are investigating two approaches for search engines to search for information in all the numerous documents on the host server. The brute force "grep" search (Table 1, item 24) relies on a list of files that it searches for specific word occurrences. The result returns a list of files or a list of paragraphs that match the search string. This approach can be a considerable load on the host server since each search must always browse through all the files. The other possibility is to use WAIS to build a full text index of all the files on the host server including the ones not in HTML, but in formats supported by WAIS. In theory the second solution should work much faster and have less load to the host server. In both cases, the final effect on the end user could be similar since the real bottlenecks are the communication lines. Such searches are currently possible on the ICARIS HTTP server (Table 1, item 25).

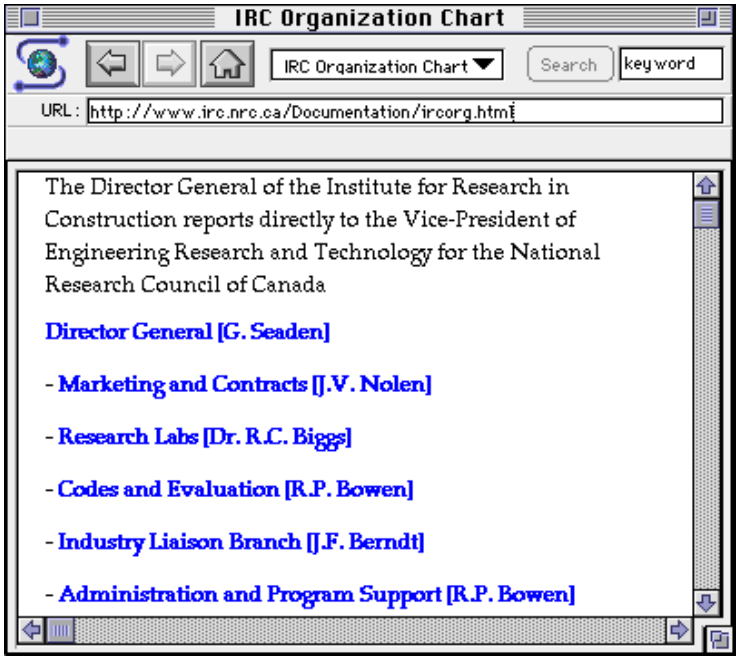

Figure 4: IRC Organization Page

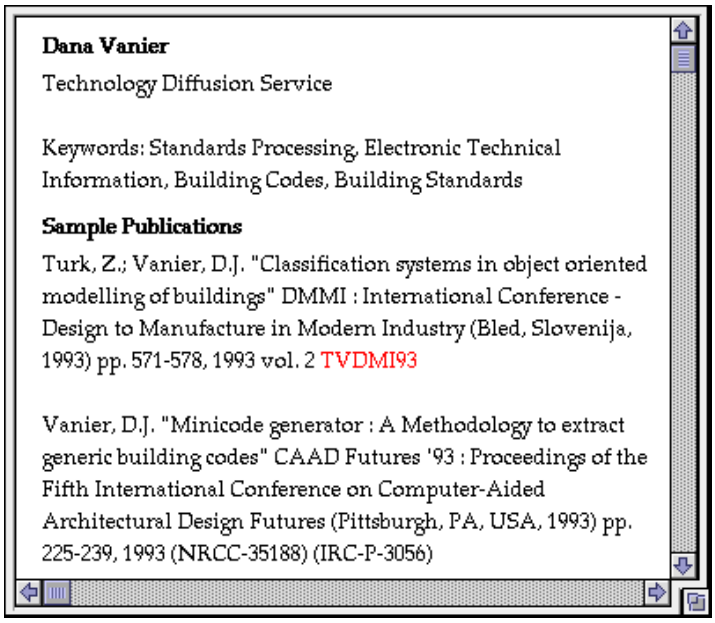

Figure 5: IRC information browsing showing information on staff members

Lighting Resources

In addition to the organization information described above, there are other sources of IRC information that could be of interest to researchers and practitioners alike. Typically, a good application for an information base is one that already exists in electronic format. One example shown in Fig 6 is Lighting Resources, an electronic prototype developed at IRC.

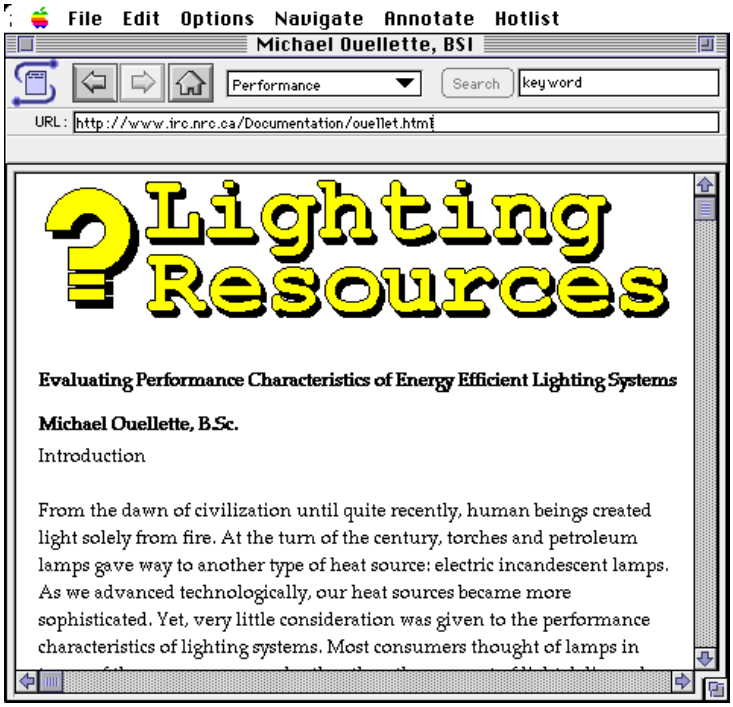

Figure 6: IRC Lighting Resources

Lighting Resources is a prototype electronic publication containing information on lighting research at IRC. It was produced in conjunction with the lighting group and other individuals involved in a lighting symposium (Lighting, 1994). It is currently available as a prototype on diskette and runs under Windows™ 3.1. Lighting Resources is IRC's second prototype investigating the use of information technologies to disseminate technical information. The CD-ROM entitled *Construction Resources* (Worling et al, 1992) employed HyperCard™ and is only available on the Macintosh™ platform. In contrast, the diskette version of Lighting Resources uses SGML as the markup language and currently resides in the Interactive Authoring and Display System Application (IADS) software (Table 1, item 26). The features of the IADS implementation that are of interest to this WWW application are hypertext links to other documents, full colour graphics, graphical maps with cross-references fields, and full-text searching (See Fig 7).

Moving the contents of Lighting Resources to the MacHTTP server was relatively easy. That is, modifications were made to the SGML source document to HTML, graphics were converted to GIF or JPEG from ".bmp", and cross-reference links were reconstructed. As mentioned earlier, HTML is similar to SGML, so brute-force semi-automated conversion (global search and replace) was used; however, tools such as FastTag/Hammer (Table 1, item 27), Omnimark (Table 1, item 28), TagWrite (Table 1, item 29) and the Rainbow (Table 1, item 30) format can convert documents from standardized formats such as RTF (Table 1, item 31).

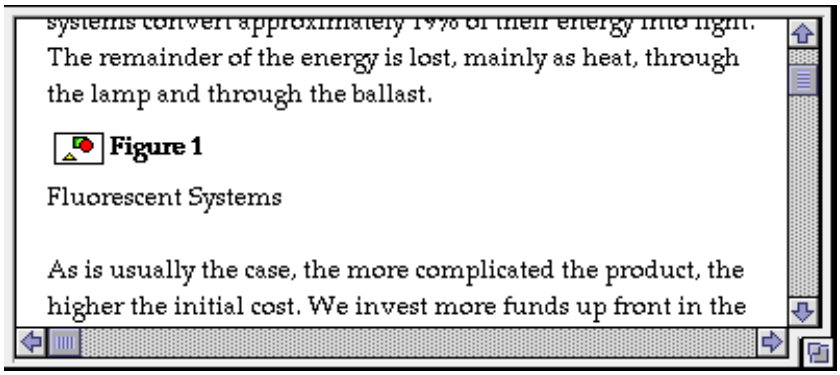

Figure 7: Lighting Resources with links to graphics

As can be seen in Figs. 6 and 7, products such as Lighting Resources can be naturally integrated with the organization information through the authors' names or any of their affiliations.

Exchange of design information

Internet's FTP and e-mail services are the natural infrastructure for the lowest layers of electronic and product data interchange (EDI/PDI). Companies can send drafts and documentation using multimedia e-mail or simply transfer files using FTP. The size of

engineering documents can pose a problem which can be solved either by leasing faster lines, by compression techniques, or by sending data on optical disks or magnetic media. TCP/IP can also be used to interconnect LANs in such a way that the remote networks and local LAN, and the corresponding file systems, merge into one seamless local area network (Turk, 1994).

This distributed construction information source could provide building practitioners and researchers with access to the tomes of construction information available today. These hypertext information bases could not only incorporate technical information and building codes (Watson and Ward, 1992), but could also contain controlled vocabularies, national standards, international standards, national specifications, contract documents, as-built and working drawings and manufacturers drawing, specifications and instructions.

This distributed information can be an invaluable asset to members of the design team, allowing them to communicate directly to their colleagues, consultants, contractors or manufacturers without numerous telephone calls or expensive meetings. The technology can automatically provide an electronic record of all the discussions, decisions and project notes. It also allows members of the design and construction team to transfer or share word processed documents, specifications, change orders, CADD drawings, database records, and hypertext documents in a quick and efficient manner.

The Internet could be visualized as the backbone of communications for a design network. Building designers transmit drawings, sketches and specifications via FTP, and the Internet allows use of services such as EDI, Gopher and WWW for flexible, robust information requests. It could also assist the transfer of information about building codes as suggested by Williams (1993).

Generally, we feel that the enumerated Internet services do not provide the necessary security, locking and version control capabilities to be directly used in a framework for collaborative design environment. However, these are host server features that can be developed in the future to meet these demands from the information suppliers.

Building codes and standards

Hypertext is regarded as a key technology in the passive representation of building codes and regulations (Vanier, 1989; Vanier, 1990; CD-REEF, 1992; Thomas and Worling, 1992; CCB, 1993). WWW, backed with a WAIS full text search engine, makes it possible to convert paper documents into the WWW format quickly and efficiently. For example, there are plans to create a WWW interface into a more advanced forms of hypertext like the PLINTH system (Casson, 1993; Table 1, item 32), which includes typed links and graph browsers not provided by the WWW.

A current project at the University in Ljubljana is converting the latest drafts of Eurocode 1 and Eurocode 8 (EN 1991, EN 1998) to WWW. Eurocode 8 is being converted to Windows™

Help format; from there it can be translated to HTML automatically. The conversion procedure for Eurocode 1 is as follows:

- Using OCR software, the paper version of Eurocode 1 is scanned, recognized, spell-checked and split (according to contents) into documents up to 5 kbytes in size (to achieve transfer times below 5 seconds on slow communication lines).
- A 'text to html' filter is applied to create about 85% of the HTML markup, the rest is done manually.
- The hypertext links for the table of contents and the explicit cross-references are done semi-automatically with human intervention.
- The figures and tables are scanned and inserted as bitmaps.
- The next and previous links, and the link to the table of contents, are added to each page automatically.
- The document is searched for words that are listed in the Canadian Thesaurus of Construction Terms (TC/CS, 1978); each word is substituted with a pointer to an index page of this word. The index page contains pointers to all words in the documents accompanied by some characteristic text, (e.g., nearest section number and title). The index page also contains a pointers to related words ('See Also' references).
- The documents are indexed using WAIS.

Some results of this work are shown Fig 8. Due to unclear copyright policies of international standardization bodies it is still not known whether such versions of their documents will (or can) be offered publicly on the WWW.

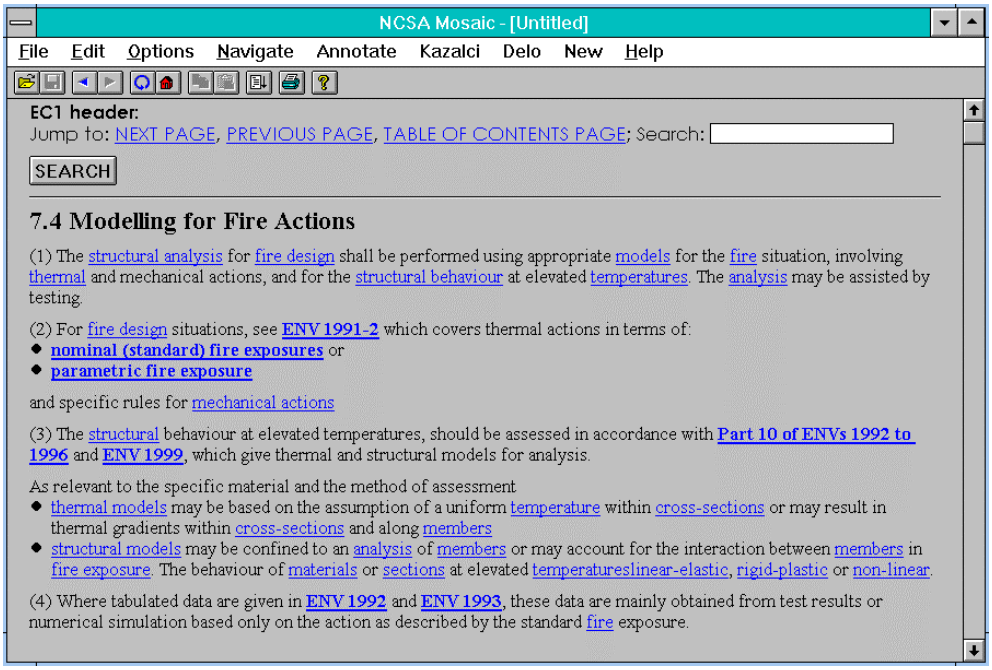

Figure 8: Eurocode 1 as a HTML document displayed by Mosaic.

Joint development of STEP models

International standardization committees are ideal customers for Internet services, and the WAN distributed information seems to provide an ideal collaborative infrastructure. A move in that direction is the Step On-Line Information System SOLIS (Table 1, item 33), which includes dial-up and FTP access to STEP documents and some e-mail discussion lists. Another interesting development is the EXPRESS to HTML converter (Table 1, item 34) which translates EXPRESS models into hypertext; that is, so that one can quickly move from entity's use to entities definition.

The authors can envision a global WWW-based EXPRESS authoring tool. The components would include:

- A central organization would provide administrative tasks, central search engines, participants lists etc., but would delegate the responsibilities for the definition of certain concepts or schemas to an organization somewhere on the Internet. This organization would fix the URL of *the* definition of the concept.
- Developers of a schema would provide a hypertext definition of their schema. All external references would be hypertext links as well and lead to the outside world such as EXPRESS definitions, word dictionaries, thesauri, programming language implementations, or Mosaic annotations. Hypertext links could be connected to textual and graphical comments (as with EXPRESS-G) and discussion threads could be sent over e-mail or entered in a form. These would complete the modelling framework.

Industry information systems

In 1994, Slovenia started a national development project TIGRA (technical information system for the building industry) sponsored by the Ministry of Science and Technology and the Building Centre of Slovenia. It is planning to use TCP/IP connectivity to create an information system for the Building Centre, connecting suppliers, building contractors and government authorities. Multimedia systems based on WWW and Windows™ will be used to gather and disseminate information. Emphasis will be placed on a system design that simplifies the supply and structuring of the information and yet does not hamper efficient access.

Systems to support R&D collaboration

The World Wide Web can be viewed as a huge collaborative environment with its own "white" and "yellow" pages, calls for papers(Table 1, item 35), publishing opportunities, and discussion systems. Since the spring of 1994 the ICARIS (Table 1, item 36) system (Turk, 1993) operates as a WWW server as well. The home page is shown in Fig 9. ICARIS includes pointers to Internet resources, calendar of events, discussion group on education of IT for civil

engineering, on-line membership databases and a distributed bibliography of papers and reports including some with full text and graphics.

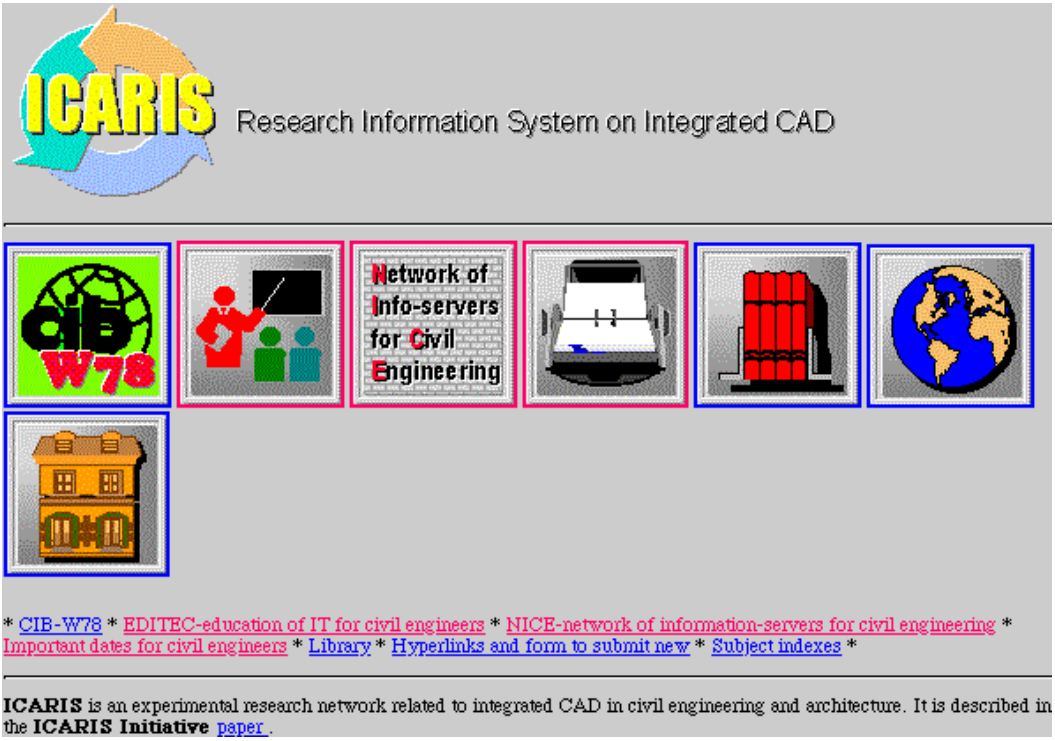

Figure 9: The ICARIS Home Page.

Other existing Internet resources

There are many interesting services in existence now, and new ones are being added so quickly that giving a listing in a document distributed on paper would not stand the test of time. These listings include on-line galleries and show rooms of architectural projects, full text publications, STEP documents, research projects summaries and reports, facts on building departments and the staff at universities and research institutions. There are also documents on the Internet that contain up-to-date directories of such services. Perhaps the most exhaustive list of pointers to Internet resources for architecture, engineering and construction industry (AEC) is Jeanne Brown's list, which is updated regularly (Table 1, item 37). Another good starting point for building professionals with WWW access is the ICARIS WWW server (Table 1, item 38). It contains pointers to most AEC home pages, including a list of the latest additions. It is also a good place to register a new resource. An initiative for collaboration among civil engineering information providers on the Internet is called NICE (Network of Information servers for Civil Engineering- Table 1, item 39).

DISCUSSION AND CONCLUSION

This is a preliminary investigation of Internet opportunities for distributed construction information. As can be seen there are numerous potential applications for this technology in the construction industry, quite a few have been attempted by the authors of this paper. This research, in combination with knowledge in practice, can provide the missing ingredients: high quality, edited documents leaning towards practice-oriented information, but supplied on a commercial basis. Almost all the information sources investigated in this paper have been placed on the Internet free of charge, which often reflects a marginal quality of information, poor data structure and primitive presentation format. However, this can be remedied quickly.

The use of the Internet is growing exponentially. In many areas the information flow and volume are doubling in less than a year. The ICARIS information system, for example, delivered 82 kilobytes of information per day and was accessed by 234 client programs (at least as many different humans) in June 1994. In November 1994, only 5 months later, 1307 people accessed ICARIS and downloaded an equivalent of 300 technical papers (similar in length and complexity to this paper) in a month, averaging to 450 kilobytes per day. In contrast, the total usage of the WWW server at Faculty of Civil Engineering in Ljubljana grew ten-fold in a similar period of time. On weekdays this host server delivers approximately 6000 "pages" of information, or about 20 megabytes to the construction community: - an equivalent of twenty books each 500 pages long. Comparing the five-fold increase in ICARIS to the ten-fold change in other areas illustrates that the construction industry and its research must still catch up with the silicon (and optical) curve - curves which mean doubling the computing power every 18 months.

The construction industry will be affected by the development of information highways: but how?. For example, information-intensive tasks particularly related to project planning and construction management such as estimating, planning, benefit-cost analysis, CAD, tendering and contract management can realize immediate benefit from the rapid transfer of documentation provided by the Internet. The geographical location of companies offering "knowledge worker" services such as design, surveying, or consulting will cease to be an important factor in securing new contracts. Niche markets will open for small companies that use the Internet as part of their work strategy: for example, CAD drawing services that deliver copies of drawings to the nearest plotter, in any city or country. In addition, construction companies will be downsizing into lean, specialized and flexible organizations that span large geographical distances; these companies will understand that the networked planet is their market.

How can the industry reach this state of electronic efficiency?

The Internet is proliferating on a "grassroots", or bottom-up approach; individuals and their requirements are establishing the demand and supply. The same can be true for a distributed

construction information network. We should publish all our work on the Internet using the appropriate technology; we should encourage our organizations to do the same; we should ensure data integration through format standardization; we should publish freely and openly; and we should cooperate. This grassroots approach can also encourage others in the profession to do the same: manufacturers, suppliers, consultants, builders and owners can supply their information to close the net, so to speak. Soon we will have all this information at our fingertips.

REFERENCES

Bourdeau, Marc (1991), The CD-REEF: The French Building Technical Rules on CD-ROM, Computers and Building Regulations, VTT Symposium Series, No. 125, Espoo, Finland, 27-29 May, pp. 117-133.

Brown, Jean (1994), Internet Sources In Architecture And Building, University of Nevada, Las Vegas. (gopher://una.hh.lib.umich.edu:70/00/inetdirsstacks/
archi:brown)

Carroll, Jim and Rick Broadhead (1994), Canadian Internet Handbook, Prentice Hall Canada, Scarborough, Ontario, 415 p.

Casson, Andrew (1993), PLINTH: Integrating Hypertext, Semantic Nets and Rule-Based Systems in an Expertext Shell for Authors and Readers of Regulatory Documents, Mayfiled J. and Nicholas C. (eds), Proceedings Workshop on Intelligent Hypertext, CIKM 93, Arlington, VA, USA, 1993.

CCB (1993), CCB Update, CCB Bulletin, Publication of the National Institute for Building Sciences, Washington, DC, Vol. 5, No. 2, (n. pag.).

EN 1991 Eurocode 1 Basis of design and actions on structures, CEN Technical Committee CEN/TC 250.

EN 1998 Eurocode 8 Design of structures for earthquake resistance, CEN Technical Committee CEN/TC 250.

Krol, Ed (1993), The Whole Internet: User's Guide & Catalog, 3rd ed., O'Reilly & Associates, Inc., Sebastapol, CA, 376 p.

Lighting (1994), Effective and Efficient Lighting, Building Science Insight Proceedings, No. 19, Institute for Research in Construction, Ottawa, Canada, 64p.

TC/CS (1978), Canadian Thesaurus of Construction Science and Technology, Industry, Science and Technology Canada (formerly Department of Industry, Trade and Commerce, Government of Canada), Ottawa, Ontario, 2 Vols., (n. pag.).

Thomas, J. Russell, and Jamie L. Worling (1992), SGML and ODMS: Tools for Building Code Writers, Proceedings Joint CIB Workshops on Computers and Information in Construction, Montreal, PQ, 12-15 May, pp. 411-419.

Turk, Ziga (1993), ICARIS Initiative, technical report presented at CIB W78 meeting in Singapore, Aug. (file://audrey.fagg.uni-lj.si/pub/icaris.zip)

Turk, Ziga (1994), Three Prototype CDDM Systems, CIB W78 Workshop on Computer Integrated Construction, VTT, Espoo, Finland, 22-24 Aug.

Vanier, Dana J. (1989), Computerization of Building Regulations, Proceedings of the International Conference on Municipal Code Administration - Building Safety - and the Computer, Winnipeg, Manitoba, 24-28 Sep., pp. 43-62.

Vanier, Dana J. (1990), Hypertext - A Computer Tool to Assist Building Design, The Electronic Design Studio: Architectural Knowledge and Media in the Computer Era, The MIT Press, Cambridge, MA, pp. 283-300.

Watson, A.S., and M.A. Ward (1992), Building Hypertext Representations of Regulatory Material in the Construction Industry: A Case Study, The International Journal of Construction Information Technology, Vol. 1, No. 1, pp. 43-58.

Williams, Fred (1993), A Design Standards Document Server, Master's Thesis, Department of Engineering, Carleton University, Ottawa, Ontario. (http://www.civeng.carleton.ca/Research/Thesis/Fred_Williams)

Worling, Jamie L., B. Scott Mellon, and Dana J. Vanier (1992), Electronic Technical Information-An Access and Distribution System, *Proceedings Joint CIB Workshops on Computers and Information in Construction*, Montreal, PQ, 12-15 May, pp. 572-582.