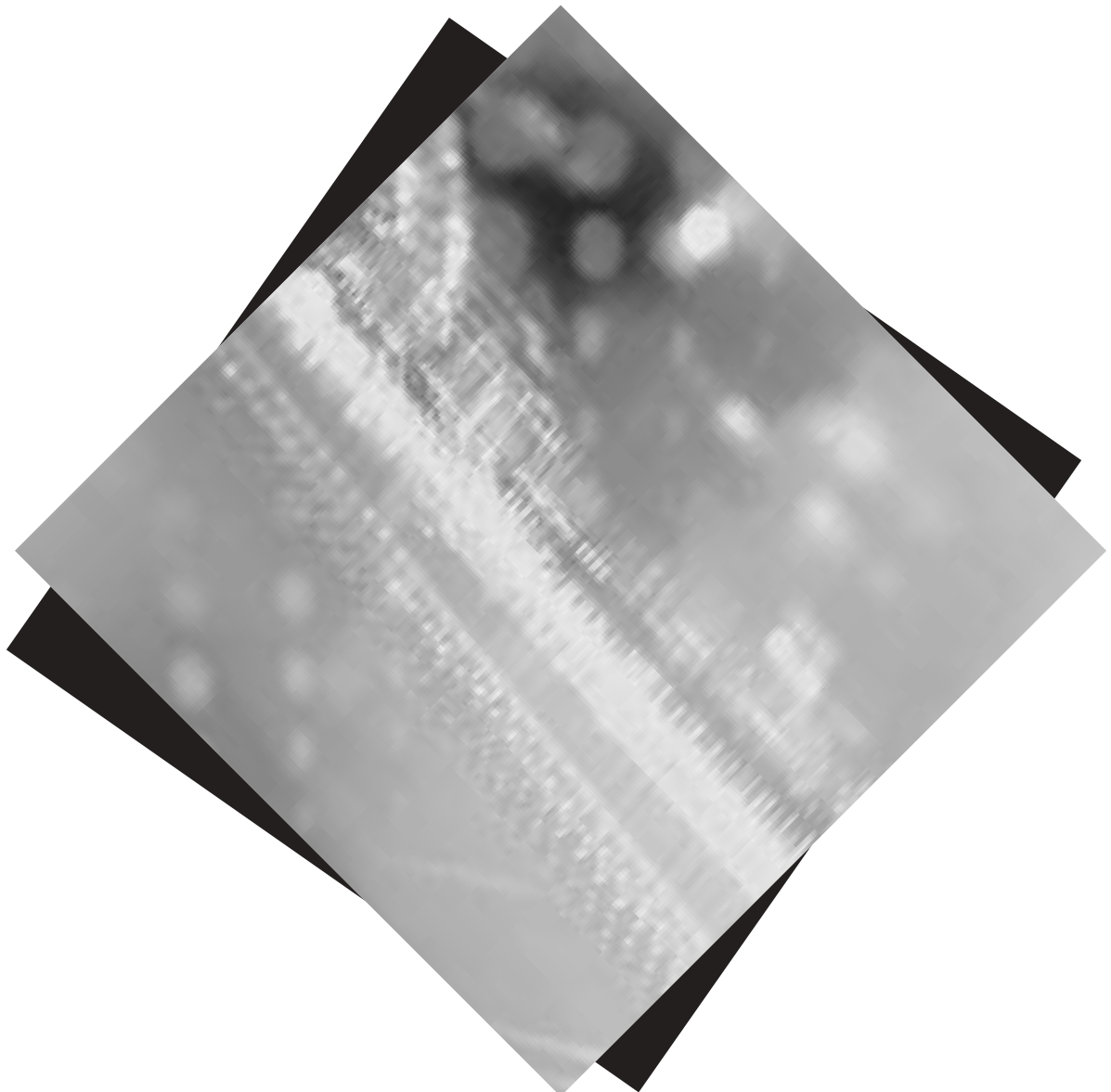


# Understanding the Internet and World Wide Web

Every discipline has its own set of words, terms and acronyms which may seem incomprehensible and sometimes alienating to outsiders.

The Internet (or Net) is no exception. But to make understanding it even more difficult, the Net is not a concrete object that can easily be quantified, weighed or held in your hands to be inspected.

However, having an understanding of the underlying principles behind the Internet and a good grasp of its language will help those managing websites feel more confident about dealing with issues and communicating with colleagues and the web development team.



**In this chapter you will learn:**

- ◆ what the Internet is and the context in which it began
- ◆ what the World Wide Web is
- ◆ who devised it
- ◆ its scope
- ◆ how it works
- ◆ who regulates it
- ◆ some of its specialist terms.

**The Internet – A definition**

The Internet is a worldwide communications system that connects computers and networks of computers to each other. It uses numerous protocols (sets of technical rules) which dictate how information is sent and received within the system. The computers, modems and software that enable people to use the Internet all obey the protocols (whether they be PCs, games machines, TVs or embedded in refrigerator doors).

**Where did the Internet begin?**

The Internet (or Net) had its origins in a communications project called ARPANET (Advanced Research Projects Agency Network) which was established in 1969 by the U.S. Department of Defence. Its aim was to ensure that those organisations engaged in defence-related research could share data with each other securely via a robust communications network. It was devised to ensure that should one or more link in the network go down then the network as a whole would still survive and function.

The fact that the network's more famous offspring, the Internet, does not universally crash despite constant instances of successful hacking and unleashing of debilitating viruses is testament to the foresight and ingenuity of its inventors.

It was not long before researchers and academics outside the defence area began using ARPANET to communicate with each other through email and to transfer computer files from one computer to another. This was a great advance on having to mail or carry disks or computer program cards in order to share programs and data.

The Internet itself emerged in the early 1980s when the set of transmission control protocols (TCP) that had been devised for ARPANET were extended to provide users with more functionality. These were called Internet protocols (IP). The new functionality and the set of rules that governed it was, and still is, labelled TCP/IP. This was a significant development and meant that far more organisations could join the world-wide network of computers that constituted the Internet.

**What is the World Wide Web?**

In 1989 a group of scientists, led by Tim Berners-Lee, at the European Laboratory for Particle Physics in Geneva, Switzerland (the CERN team) began working on a new branch of Internet protocols that offered a new way of linking documents, files and addressing on the Internet. Berners-Lee called the project, World Wide Web (the Web). They developed a HyperText Transfer Protocol which is mostly referred to by its initials, http – the beginning of every website address. The protocol is a set of rules that enable every document on the Internet to be connected to every other document via a hyperlink, a link that the author can place anywhere in the document. The link may enable a user to jump to another document within the same website or to a totally different document or website.

The World Wide Web (WWW or the Web) can therefore be defined as, the collective term given to the vastly distributed universe of documents and data residing on the Internet that can be accessed using the http protocol. The Web is a sub-set of the Internet. Among other sub-sets in widespread use on the Internet are those built around the "ftp" - file transfer protocol and the "irc" - Internet relay chat protocol.

A website is a set of interlinked files containing words, images, video and sound and usually links to other websites.

**[http://www.unitedfocus.com.au/pdf/Choosing\\_Name\\_for\\_Website.pdf](http://www.unitedfocus.com.au/pdf/Choosing_Name_for_Website.pdf)**

The Internet application protocol to be used to access the document.	The domain name providing the address of the computer on which the file is stored. unitedfocus = the owner's name com = it is a commercial organisation au = country is Australia	The document is stored on the computer in a directory or folder called PDF.	The name of the document to be retrieved from that directory or folder.
--	---	---	---

A website is a set of interlinked files containing words, images, video and sound and usually links to other websites, all of which have been coded so that computers using http can find the sites and see their content. People with appropriate software, typically "browsers", can interact with whatever activities are available, such as selecting images to print or selecting and paying for goods and services.

A crucial characteristic of this protocol is that it is "stateless" so that each access, by each user, is entirely independent of prior accesses. As a stateless protocol, http is perfect for viewing static objects, not changed by a user's interaction.

With the hypertext transfer protocol established, it was then necessary to construct a computer language for creating documents that used the hyperlink function. The language devised to create these web documents with their hyperlinks was called HyperText Markup Language (HTML). The language provides the means by which programmers not only can create hyperlinks to other documents but can specify such things as what text, images, sound and video are displayed on the screen and their layout, the various font types and sizes to be used and colour schemes.

With http and HTML developed, a further challenge was to develop a user-friendly addressing system so that documents and websites could be found on the Internet. The addressing scheme that was devised, and which is at the heart of the Web, is called the Uniform Resource Locator (URL). A unique URL is assigned to every

document on the Web – just as every house in a street is assigned some identifying, unique number or name. The URL contains the name of the protocol required to access the documents on the Web (eg http) and a domain name that identifies the address of the computer on which the document is stored (eg www.unitedfocus.com.au). The URL also contains the location of the directory or folder on the host computer in which the documents to be retrieved are located, and the name of the actual document that is to be displayed.

Actually, a URL is, in its raw state, a series of numbers. A typical one might look something like this: http://203.2.124.97. However, for reasons of human psychology, these URLs are not easily remembered, nor do they reveal anything of the nature or content of the website to which they are assigned. Domain names were devised to overcome this problem. They are plain English words chosen by the website owner to convey meaning and to aid the memory of users.

The Internet is a communications system that is defined by a specific technical protocol (TCP/IP) while the Web conforms to a sub-set of that protocol (HTTP).

The domain name does not replace the URL number but is the “public face” of the website. When a domain name is typed into a browser it is translated into the URL number, usually without us seeing it happen. The translation is done by domain name servers (specialist computers) on the Internet.

So, instead of trying to remember a URL such as `http://203.2.124.97`, we can simply type in the plain English web address, `http://www.unitedfocus.com.au`, and the domain name servers do the translation for us.

In January 1992 the CERN team overcame a major challenge of the newly-created Web: to create a user-friendly means by which the public, using computers, could locate and read HTML documents on the Web and be taken to related documents located at another URL. The software they developed was called a “browser”. The first browser they devised did not use graphics like the modern versions of Netscape and Internet Explorer, just text.

The browser marked a quantum leap in user-friendliness and resulted in a rapid uptake of the Web. It did not take long for innovators to improve on the first model. In 1993 Marc Andreessen and a team at the University of Illinois released a graphic-based browser called Mosaic. The next year Andreessen co-founded the Netscape Communications Corporation and released the first version of Netscape Navigator in December 1994. In 1995 Microsoft Corporation introduced its browser, called Internet Explorer. Although there are other browsers, Internet Explorer and Netscape overwhelmingly dominate the market today.

## Who manages the Internet and the Web?

As explained earlier, the Internet is a communications system that is defined by a specific technical protocol (TCP/IP) while the Web conforms to a sub-set of that protocol (HTTP). The Internet and the Web have been able to evolve to meet many technical challenges and accommodate new protocols and applications since they were devised in the 1980s and 1990s respectively. However, the Internet and Web are not owned by any one person, company or government. So who is managing and regulating them?

Since their creation, a number of not-for-profit bodies have carried the responsibility for ensuring the integrity of the protocols and their application. Today, two recently formed organisations are now recognised internationally as managers of the Net and Web protocols:

- ◆ The Internet Corporation for Assigned Names and Numbers (ICANN)
- ◆ The World Wide Web Consortium (W3C)

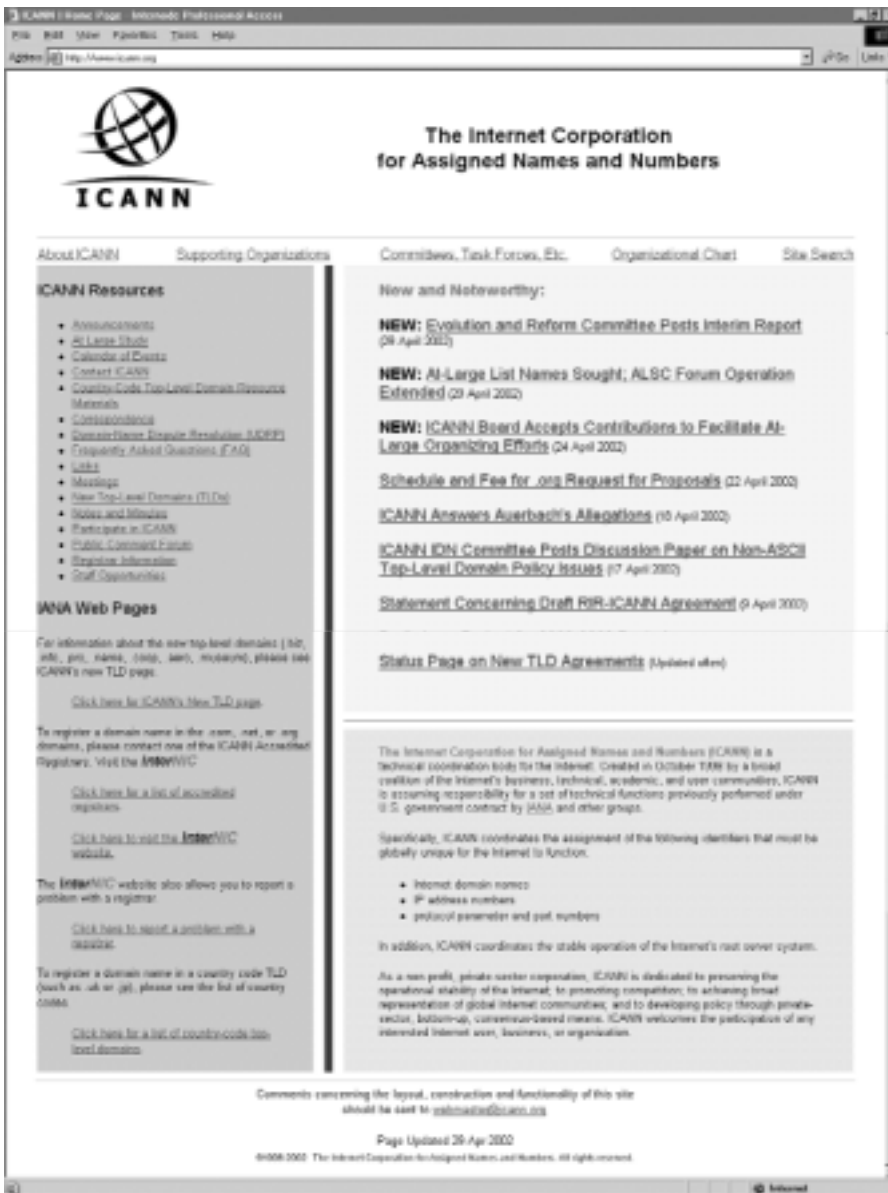


Figure 1.1 www.icann.org

### ICANN - www.icann.org

The Internet Corporation for Assigned Names and Numbers coordinates the technical aspects of the Internet. It was created in October 1998 by a coalition of the Internet's business, technical, academic and user communities. ICANN's most important responsibility is the assignment of three identifiers that must be globally unique for the Internet to function:

- ◆ Internet domain names
- ◆ IP address numbers
- ◆ protocol parameter and port numbers

In addition to this role as postmaster of the Internet, ICANN coordinates the operation of the Internet's root server system. The root servers, of which there are about fifteen located throughout the world and maintained by various organisations, contain the top level domains (TLDs). (see [www.icann.org/general/abouticann.htm](http://www.icann.org/general/abouticann.htm))

There are a number of other related and important organisations responsible for monitoring the Internet and ensuring its advancement. Most have subscription email lists that are open to anyone and which provide updates and news from the organisation. These lists will be of interest to those responsible for the technical maintenance of their organisation's website.

### Internet Society – www.isoc.org

The following explanation of the Internet Society comes from, *The Tao of IETF – A Novice's Guide to the Internet Engineering Task Force*, Susan Harris, June 18, 2001.

*“The Internet SOCIety (ISOC) is a professional membership society with more than 150 organisational and 6,000 individual members in over 100 countries. It provides leadership in addressing issues that confront the future of the Internet, and is the home for the groups responsible for Internet infrastructure standards, including the Internet Engineering Task Force (IETF) and the Internet Architecture Board (IAB).*

*The Society's individual and organisational members are bound by a common stake in maintaining the viability and global scaling of the Internet.*

## Understanding the Internet and World Wide Web

*They comprise the companies, government agencies and foundations that have created the Internet and its technologies, as well as innovative new entrepreneurial organisations contributing to maintain that dynamic. Visit their home pages to see how Internet innovators are creatively using the network.*

*The Society is governed by its Board of Trustees elected by its membership around the world."*

**Internet Engineering Task Force – [www.ietf.org](http://www.ietf.org)**

*The Internet Engineering Task Force (IETF) is a large, open, international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. It is open to any interested individual.*

*The actual technical work of the IETF is done in its working groups, which are organised by topic into several areas (routing, transport, security, etc.). Much of the work is handled via mailing lists. The IETF holds meetings three times per year." (The Tao of IETF - A Novice's Guide to the Internet Engineering Task Force, Susan Harris, June 18, 2001)*

**World Wide Web Consortium – [www.w3c.org](http://www.w3c.org)**

The World Wide Web Consortium (W3C) was founded in October 1994 by Tim Berners-Lee, inventor of the Web.

The following is an abbreviated form of the information provided on the W3C website.

*By promoting interoperability and encouraging an open forum for discussion, W3C commits to leading the technical evolution of the Web. In just over five years, W3C has developed more than 20 technical specifications for the Web's infrastructure. However, the Web is still young and there is still a lot of*

*work to do, especially as computers, telecommunications, and multimedia technologies converge. To meet the growing expectations of users and the increasing power of machines, W3C is already laying the foundations for the next generation of the Web. W3C's technologies will help make the Web a robust, scalable, and adaptive infrastructure for a world of information. To understand how W3C pursues this mission, it is useful to understand the Consortium's goals and driving principles.*

W3C concentrates its efforts on three principle tasks:

- 1 Vision:** W3C promotes and develops its vision of the future of the World Wide Web. Contributions from several hundred dedicated researchers and engineers working for Member organisations, from the W3C Team led by Tim Berners-Lee and from the entire Web community, enable W3C to identify the technical requirements that must be satisfied if the Web is to be a truly universal information space.
- 2 Design:** W3C designs Web technologies to realise this vision, taking into account existing technologies as well as those of the future.
- 3 Standardisation:** W3C contributes to efforts to standardise Web technologies by producing specifications (called Recommendations) that describe the building blocks of the Web. W3C makes these Recommendations (and other technical reports) freely available to all.

**“Let us not forget that the value of this great system does not lie primarily in its extent or even in its efficiency. Its worth depends on the use that is made of it .... For the first time in human history we have available to us the ability to communicate simultaneously with our fellow men, to furnish entertainment, instruction, widening vision of national problems and national events. An obligation rests on us to see that it is devoted to real service and to develop the material that is transmitted into that which is really worthwhile.”**

Herbert Hoover, US Secretary of Commerce, 1922. (About the radio)

## Design Principles of the Web

The Web is an application built on top of the Internet and, as such, has inherited its fundamental design principles:

- ◆ **Interoperability:** Specifications for the Web's languages and protocols must be compatible with one another and allow (any) hardware and software used to access the Web to work together.
- ◆ **Evolution:** The Web must be able to accommodate future technologies. Design principles such as simplicity, modularity, and extensibility will increase the chances that the Web will work with emerging technologies such as mobile web devices and digital television, as well as others to come.
- ◆ **Decentralisation:** Decentralisation is without a doubt the newest principle and most difficult to apply. To allow the Web to “scale” to worldwide proportions while resisting errors and breakdowns, the architecture (like the Internet) must limit or eliminate dependencies on central registries.

These principles guide the work carried out within W3C Activities.

A good example of the output of the W3C is their document, *Web Content Accessibility Guidelines 2.0*. These are widely recognised as the benchmark guidelines for developing websites that accommodate people with disabilities.

## Characteristics of websites

We have defined what the Web is in architectural terms but what are its characteristics and features? What can owners and users do with websites? Here are some of the defining features of the Web and websites that make it such a powerful tool:

- ◆ The Web is a distributed publishing and narrowcasting environment that can be used to inform, promote, market, entertain, educate and train and buy and sell goods and services.
- ◆ A website can contain text, images, animations, diagrams, illustrations, maps, video and sound, all of which can be manipulated, edited, copied, stored and audited at any time.
- ◆ A website is capable of being accessed 24 hours a day, 365 days a year, or when users are interested in ‘visiting’ it.
- ◆ A website knows no borders – it is global.
- ◆ Although a website owner does not usually monitor who is viewing the site at any one moment, the total number of times a website is viewed over any period can be calculated and a number of other statistics can be obtained – eg where users are from, popular times of the day for access, popular parts of the site.
- ◆ A website can be highly interactive, allowing the user not merely to see or hear information but to interact with it and other users via email, discussion groups and various types of forums.
- ◆ A website can be constructed to allow users to personalise it so that they see only those aspects of the site that they want to see.
- ◆ A website does not know a person's race, colour, creed, socio-economic status or cultural background.

HTML is a universally recognised set of codes and rules that are employed to format the text and images on a web page.

## Understanding the Internet and World Wide Web

- ◆ A website can provide links to other websites at the click of a mouse button, thereby creating a reservoir of inter-connected information.
- ◆ A website can be used by organisations and consumers to order and distribute goods and services.
- ◆ A website can enable the secure exchange of money for goods and services.
- ◆ Developing and maintaining an appropriate level of understanding of Internet technology and advances in web design and functionality is time-consuming and requires constant attention.

## Languages on the Web

At the heart of every website is the coding or language used to transform the words and graphic images into something that browsers can interpret and display. This transformation is referred to as “marking-up”.

## HTML

The first and most fundamental language used to mark-up content so it can be displayed by browsers is HyperText Markup Language – HTML. HTML is a universally recognised set of codes and rules that are employed to format the text and images on a web page. The codes are very simple and easy to use. There is no license fee or cost for purchasing HTML. Indeed, marking-up a page of text you have created with the word processor, MS Word, can be as easy as saving it as an HTML document using the “Save As HTML” command in the File menu.

However, with simplicity comes limitations. A website built using documents created in MS Word and saved as HTML, will not have the level of interactivity and dynamism that users now take for granted, such as a comprehensive search facility, online forms, roll-overs (where things happen on the screen when the mouse pointer passes over an object) and animated diagrams, maps or pictures.

These are the positive aspects of the Web. But what are its main deficiencies and pitfalls for individuals and organisations?

- ◆ Web users are notoriously impatient and will exit a site if it is not quick and easy to navigate and search – yet the better the quality of the graphics on your site, the longer it will take for it to appear in the user’s browser.
- ◆ The Internet can be frustratingly slow at times due to a number of factors, some of which are beyond the control of the website owner – so no matter how fast your site, it will be too slow for some users at some time.
- ◆ The Web is now so vast in terms of content and number of web pages that it is often difficult for users to find precisely what they are looking for. The challenge is to help them find your site amongst the vast array of competing sites out there.
- ◆ The two main browsers, Internet Explorer and Netscape, not only have multiple versions but they do not necessarily support the set of features you might want for your website. This increases the chance that it will not be accessible to everyone.
- ◆ The features and functionality that are possible to incorporate in a website are developing and changing so rapidly that it is difficult to maintain a website that keeps up its appeal to users, who are an increasingly demanding and savvy audience.

## Enhancing HTML

In order to overcome the limitations of HTML, and to deliver interactive elements, various enhanced languages and codes were developed and continue to be developed. Two of the most widely used are CGI and JavaScript.

- ◆ CGI (Common Gateway Interface) – is a standard set of rules governing how a website is to interact with programs running on the web server. This interface allows the web server (and thus the user's browser) to run "backend" programs, and to pass data to and from those programs. Examples of exchanging data in this way include performing database searches, doing calculations such as adding up the cost of items purchased in the online shop, generating new images and processing forms, such as application forms. Probably the most popular language used by web developers that complies with the CGI rules is Perl, as it will run on all types of web servers.
- ◆ JavaScript – is a plug-in to Netscape and Internet Explorer that was developed by Netscape and Sun Microsystems Inc. It has wide application to websites, is compatible with most web servers and comes built-in to Netscape version 3 and above and Internet Explorer version 4 and above. It is one of the most commonly used tools by web developers for creating movement, sound and special effects on websites. It is also employed widely for linking the website to other programs and backend databases such as an image library.

## XML

XML (Extensible Markup Language) is a formal subset of the publishing language SGML (Standard Generalised Mark-up Language) developed by the W3C. It allows the people responsible for the content of a website to categorise and index text, headings and other content in the website in a highly organised and flexible manner. Tagging or marking-up words, phrases, headings or data in a consistent way using an XML rule set means that they can be managed easily and new content added and displayed in the same manner as other types of content. Communities of users can agree on usages, called Document Type Definitions (DTDs) that make it possible for them to fully understand exchanged content and manipulate it.

XML is a very powerful, flexible language that imposes discipline on the way content is categorised and stored within a website. Because it provides a rigorous structure for content of a website, it provides an ideal interface between a website and other structured data sets, such as databases and spreadsheets. XML also facilitates a cost-effective means of publishing information from the website in other formats, such as word files or PDFs.

Organisations of a similar type often have a general XML structure for conducting organisation-to-organisation transactions. The standards for these structures are currently being formulated by industry working groups within the W3C. When exploring XML as a potential solution for managing web-based information it will be important to ensure that the design conforms to the standard for your particular type of organisation. This will ensure the long-term viability of your data in a globally-interactive web environment.

If you would like to read more about XML, visit [www.w3.org/xml/](http://www.w3.org/xml/) .

# 1

A plug-in is a software program that supplements a browser (or “plugs in” to it). Once the plug-in has been installed on your computer, it is automatically activated every time the browser is opened.

## Understanding the Internet and World Wide Web

### Browsers and plug-ins

A plug-in is a software program that supplements a browser (or “plugs in” to it). Once the plug-in has been installed on your computer, it is automatically activated every time the browser is opened. Websites that utilise the functionality that plug-ins provide, “look for” the plug-in on your computer and use its codes to, say, start a video, play back some audio or make a logo revolve.

◆ **Acrobat (by Adobe)** – this plug-in, from Adobe Systems Incorporated ([www.adobe.com](http://www.adobe.com)), enables users to read files (eg reports or spreadsheets) that have been converted by their author into Adobe’s Portable Document Format (PDF). PDFs are useful because they preserve the formatting of the original document no matter what computer system or word processor or spreadsheet program the user has. Users who select to download a PDF will need Adobe’s Acrobat Reader program on their computer. It is a free, widely-used program, so many users will have it and know how to use it. (For more on PDFs see chapter 4)

◆ **Shockwave (by Macromedia)** – a plug-in for displaying such things as animated cartoons and interactive games that often include rich graphics and sound effects. ShockWave programs may invite interaction from users, providing such things as buttons to click and objects to rotate. The web developer will have used the programming languages Flash or Director to create effects using the Shockwave plug-in.

◆ **Quicktime (by Apple Computer)** – this is one of the most widely used plug-ins for displaying standard and panoramic images such as a virtual tour of an art gallery and playing pre-recorded audio and movie clips through the user’s browser.

◆ **RealPlayer (by RealNetworks)** – for displaying live streaming video and sound as well as prerecorded sound files.

### Resources

Visit <http://www.unitedfocus.com.au/livingwebsites> for a hotlink to take you directly to:

- ◆ the latest version of the websites referred to in this chapter
- ◆ template documents to download and use
- ◆ online glossaries for explanations of terms used
- ◆ sites of related interest.