The Pre-Web Internet and Information Management

0

Waiting for the Web's Arrival

It was not long ago, on an Internet that did not yet include the World Wide Web, that finding files and resources or even locating a site on the Internet was considered a challenge. One could not hop quickly and effortlessly from one Internet site to another with the click of a mouse button or view the Internet and interact with it using a colorful, icon-rich graphical interface. Instead, everyone worked hard to gather and maintain his or her own list of Internet sites, kept notes on each site's resources and how to access them, and consulted with friends and colleagues to learn about additional Internet resources. Overall, one's interaction with the Internet was very narrow in focus and primitive in operation, particularly by today's standards. But the Web changed all this, decidedly and permanently.

What most people will never be able to fully appreciate, given how relatively few people traveled the Internet before the Web's arrival in the early 1990s, is the magnitude of the change effected on the Internet by the introduction of information management systems in general and the Web in particular. Early information management systems — Archie, Gopher, WAIS, and the Web most of all — transformed the Internet from a vast, little known repository of independent and isolated information resources into an integrated, searchable, and navigable information space. This transformation was equivalent to taking piles and piles of books, periodicals, photos, illustrations, and other documents — enough to fill any number of football stadiums — and installing that information in a library.

Information items (or objects) that had been accessible across the Internet but that were for the most part hidden, undiscovered, unadvertised, or unrecognizable were found by these information management systems; they were then categorized, labeled, and tagged with a unique address so that they could be easily and quickly located and retrieved. Each system created its own searchable catalog or index, making it possible to browse its collection of information resources and quickly locate specific items. Each system also provided additional tools to further assist individuals in locating the information they needed. In other words, information management systems brought order and organization to an Internet that was as difficult to use and disorganized as it was distributed and decentralized.

The information space created by these information management systems has come to hold a seemingly endless number of resources and a wealth of information the likes and extent of which the world has literally never before known. Computers and networking fundamentally changed the way information could be stored, reused, and shared. The Internet and information management systems made that information findable, accessible, and searchable. The sheer volume and breadth of information now available over the Internet demonstrates that the Internet revolution is in large part an information revolution. Information management systems — especially the World Wide Web — precipitated and fueled this information revolution when they transformed the Internet into an information space.

Before the Web's arrival, the Internet was difficult for experienced computer people to use and nearly impossible for everyone else. If you did not know what you were looking for, precisely where it was to be found, and how to access it, you were out of luck. If you needed help, it was out there, somewhere. But it, too, was not all that easy to find or, for that matter, to follow. Imagine that you had heard of a new Internet site that contained weather forecasts organized by region and city, but you had not been told the site's address. No search engine existed to help you find the site. Moreover, if you did manage to locate the site, there was no browser to help you examine and navigate the site's contents. The information was there, contained in thousands of text files describing the weather conditions for each region and city and thousands of graphics files storing the associated weather

satellite images. But it was up to you to discover the information you wanted, retrieve it, and determine how you were going to work with it.

The tools that existed to travel the Internet and retrieve information consisted of basic commands like TELNET or RLOGIN to access a remote site and FTP to transfer files. It required time, persistence, and a fair amount of experience in the use of these commands to achieve the desired result; and since every Internet site was different, coming up with a method for locating, accessing, and using information from one site did not necessarily simplify matters when it came time to perform the same basic operations on another site. The entire process was manual, time-consuming, and technically challenging. There was little hope that this Internet — no matter the volume or diversity of information it contained — would ever gain widespread popularity or become a household word.

But the Web turned this situation around in two critically important ways. First, it gave information providers the tools they needed to make their content findable and accessible. The Web enabled them to organize and manage their information resources, to present their information in a simple, architecture-independent, non-proprietary format, to make their resources more readily known and available to others, and to control access to them from across the Internet. Second, it enabled information providers to interconnect information in a new and powerful way through something called hyperlinks. A hyperlink is an electronic cross-reference that functions like a traditional cross-reference — a notation in one place that directs the reader to related, often more detailed information in another place. More than that, however, a hyperlink also functions to physically interconnect information — a reader clicks on a hyperlink notation and is taken to the referenced information, or destination. Through hyperlinks, information providers became able to interconnect their information with information and resources on other Internet sites, as well as to interconnect information between their own documents and other information objects on their own site. As will become clearer in the following chapters, it is this interconnection of information through hyperlinks and the resulting formation of an ever-growing web of criss-crossing information threads that is responsible for the Web's name. Even more importantly, this interconnection of information

has had the largest and most lasting impact on our use of the Internet.

In addition to enhancing the organization, integration, and accessibility of information on the Internet, the Web also addressed the limitations and frustrations faced by individuals in search of information. It accomplished this through the creation of a new type of simple, fast, easy-to-use computer application called a Web browser. A browser made it possible for anyone to navigate the Internet's information resources. It kept the details of how and where information was stored hidden from the user; and it transparently handled the complex and tedious process of copying information files and determining how to interpret and display them. Browsers made locating and retrieving information as easy as clicking on a highlighted word or typing in a simple address. In effect, browsers and the Web brought the Internet and its ever-increasing store of information to the individual, greatly reducing the effort and skill necessary to discover what the Internet had to offer and how to put those resources to use.

You may not think of the Web as an information management system. You see its product, or effect, in the form of Internet destinations where you can check the local weather forecast, track and trade stocks, look up what's playing at the neighborhood movie theater, or perhaps find your soul mate. You use it as a tool that helps you pay bills or balance your checkbook, maintain and share your calendar, keep informed about current events that are related to your particular interests, or do research for a school project or for writing a book. But the Web serves these functions, and many others, by operating as a relatively simple, but powerful and distributed information management system. It succeeds so well at this task that many people mistakenly equate the Web with the Internet, confusing the Web — a service that provides access to the vast amount of information available across the Internet — with the large, complex infrastructure of host computers, routers, wires, transmission devices, and protocols for managing the safe, fast, and efficient transmission of data that is the Internet.

The Web's beginnings were strikingly different than those of the Internet. As explained in *The Internet Revolution*, the inspiration for the creation and development of the Internet came from the military and its need for a decentralized and redundant communications network. The demand for more economical and

egalitarian use of expensive and remotely located computer resources also contributed significantly to the Internet's design and development. Additionally, the contributions of individual Internet users had an early and lasting impact on the Internet's form and function. The unanticipated introduction and immediate popularity of such services as email, chat, and newsgroups are examples of just such contributions. The Web itself is another example.

The birth and early development of the Web, however, were the result of a far more localized, focused, and controlled effort. Its creation can be attributed to a single person, Tim Berners-Lee, and to a single location, CERN, the European Particle Physics Laboratory in Geneva, Switzerland. Berners-Lee was undoubtedly in the right place at the right time. He worked in an environment filled with the latest computer and networking technology at a time when the Internet and networking standards had spread internationally and commercialization of the Internet was on the horizon. An inventive, smart, and determined individual, Berners-Lee faced the task of organizing and improving access to a large, disparate body of information at CERN. But instead of developing a specific solution that would only meet the information requirements at CERN, he engineered a distributed, all-purpose information management system that could accommodate virtually any type of information, one that we now happily and universally refer to as the Web.

Many people believe that the Web is something utterly new and unique. In some respects, this is true. Certain pieces of the Web's technology, along with its influence on the evolution of the Internet, its impact on business, and its revolutionary effect on the distribution and interconnection of information, are without precedent and represent distinct and singular achievements. But the motivation behind the Web's creation was neither new nor unique. Moreover, some of the core concepts and much of the engineering that contributed to the Web's construction and that were critical to its success, such as hyperlinks, had existed in one form or another for many years, as is described below. Berners-Lee's inspiration for the Web dates back considerably further than the ideas and technology on which he built the Web — to a popular, household book from Victorian England. The following