

# **Present Change, Past Lessons, and Future Challenges: Dealing with networked information in the '90s**

Andrew Treloar  
School of Computing and Mathematics  
Deakin University, Rusden Campus  
662 Blackburn Road, Clayton, Victoria, 3168  
Ph: +61 3 244 7461 Fax: +61 3 244 7460  
Internet: treloarac@brt.deakin.edu.au

NOTE: I now work at Monash University – current contact details at  
<http://andrew.treloar.net/>

**School of Computing and Mathematics Technical Report C93/27**

## **Abstract**

This paper examines the current situation with respect to networked information in terms of change and its consequences. Some of these dimensions of change are considered, and their impact on networked information and its users are discussed. Particular times of transition in the recent past of information technology are then examined. Some possible lessons are extracted from each episode, and their application to the Internet community noted. Finally, some challenges confronting the management of networked information are considered in the light of the lessons learnt. In addition to the need for richer documents, better management of electronic information, and more sophisticated information management tools, it is suggested that fundamentally new ways of organising information might be necessary. The conclusion is that the real challenge currently facing the international information community may be to move beyond the old ways of doing things and create the fundamentally new ways of working with information that will be required to meet the information challenges of the future.

## **Keywords**

Internet, networked information, information organisation,archie, gopher, WAIS, information technology.

## 1. Living with change in the present

Dealing with networked information in the early 1990's can be described as trying to drink from a firehose. For many of us, the experience is often both bruising and dampening - not only is the information coming out too fast to take in properly, the hose itself is continually flailing around, dragging us with it, and making it even harder to take a sip! This section provides one analysis of why this is so.

It is clear that many factors contribute to this situation, but they can all be viewed as aspects of change. It has become a truism to talk about the phenomenal rate of change in what is sometimes called 'The Information Society', but this makes it no less true. Not only is the rate of change dramatic, but the rate itself appears to be increasing. We often find ourselves wishing things would just slow down a little so we might absorb the implications of, and opportunities in, the current situation, without having to grapple with what is coming around the corner.

All of this is brought into sharp focus in what is developing on the 'Net' (used here loosely as a synonym for AARNET, the larger Internet of which it is a part, BITNET, Usenet, and the like)<sup>1</sup>. Here, the rate of change is affecting the range of information sources coming online, the amount of information available through each source, the expansion of connections to access these sources, and the technologies to support such access.

The range of information sources and the amount of information in each source means that keeping track of what is available is rapidly becoming impossible (if it has not already become so). The expansion of connections and access technologies increases the possibilities for information access, but means that there are more things to keep track of. The time when one 'new renaissance' individual could know about everything available are long gone - now the challenge is to be able to locate the information sources required.

The rate of change has a number of dimensions, and it is perhaps useful to consider some of these in a little more detail. There is an information technology dimension, which consists in turn of the computer technology and communications technology used to work with information. There is a people dimension, relating to the roles of information professionals. Lastly, there is an information dimension, relating to the nature of the information itself.

### 1.1 Changes in computer technology

The hardware that makes our computers work, and the software that enables us to use this hardware is perhaps the most visible dimension of this world of change. Moore's Law predicts a doubling of silicon chip power every 18-24 months, and the last two decades of processing and primary storage technology have been driven by this exponential increase. Similar, if less dramatic improvements have occurred in secondary storage and output technologies. The end result is that most computer hardware is obsolete as soon as it is acquired, and that price per unit performance continues to plummet.

---

<sup>1</sup> This paper will draw on the Internet for most of its examples, but much of the discussion is also relevant to other online information systems.

This increase in hardware performance has driven and made possible great advances in software technology. Some of the increased power has been allocated to improving the interface between computer and user, most notably with the proliferation of graphical user interfaces. Some has been allocated to improving the stability of the underlying operating systems. Much of the rest has gone into doing things on the desktop that either used to require a mainframe, or could not be done before at all. This increased power to process information often involves increased complexity for the user. How many of us use more than 20% of the functionality of our word-processors, or really understand to use Window's OLE or Apple's Publish and Subscribe? 'Fatware' applications overwhelm us with their 'everything and the kitchen sink' approach, and the seeming inevitability of creeping featuritis.

Overall, despite the gains from improvements in user-interface technology, the losses due to increased complexity and increased expectations of what users should be able to do on their desktops mean the user needs to run faster just to stay in place.

These changes in computer technology make it difficult to plan ahead and absorb the implications of what we have now. Like children in a lolly-shop, we are dazzled by the possibilities on offer and the hype associated with this year's 'hot topic', to say nothing of the promised developments just around the corner. As users the pace of change makes it extremely difficult to keep up with training in how to use the latest technology. We can spend so much time mastering upgrades that it erodes the time we might spend on using what we have already.

## **1.2 Changes in communications technology**

At the same time as these changes in computer technology, much the same technical advances have driven an increase in the capabilities of communication technology.

The last decade has seen the Local Area Network (LAN) move from being an esoteric high-tech item to taking its place as a core part of an organisation's information infrastructure. The emphasis at the workgroup level is increasingly on networked PC's rather than stand-alone machines. Standards like Ethernet, 10Base-T, and TCP/IP are simplifying the task of linking together different machine architectures into heterogenous networks. In general, this means that we can focus more on what we want to get rather on how to get it.

At the same time, the use of fibre optics for backbones within user sites, and as data super highways between sites has meant that interconnecting LANs into Wide Area Networks (WANs) is steadily becoming easier and more cost-effective. These high-bandwidth data channels mean that enormous amounts of data of all types can be moved across town, across the country, or across the world. The amount of data that fibre optics can carry<sup>2</sup> means that it is the ultimate data firehose. We are only just starting to come to terms with the implications of this scale of communications bandwidth.

---

<sup>2</sup> Up to 28 Giga ( $\approx 10^9$ ) bits/second, according to some recent work.

### **1.3 Changes in the roles of the professions**

Until recently, the worlds of the computer professional, communications engineer, information specialist, and librarian were relatively discrete. They occasionally drew on each other's expertise, but had their own largely separate specialist concerns. These worlds are now colliding and blurring at the edges as these professionals try to grapple with the special problems of the Net and what it can provide.

There is still a need for narrow specialists in each of these areas, due to the technical complexity and scope of the areas involved. At the same time, there is also a need for 'boundary' professionals who can take a holistic view and coordinate the activities of these domain specialists. Many of us are having to become such boundary professionals by default - there is no one else available or willing to take on the challenge.

### **1.4 Changes in our perceptions of information**

The nature of information itself is also changing. Information is now spoken of as a 'corporate resource' and the new discipline of information economics seeks to determine the value of information. In this environment, information must acquire a dollar value in addition to its existing utility value.

This new view of information as something with a commercial value exists in uneasy tension with the view expressed by the Free Software Foundation as 'Information wants to be free'. Librarians particularly feel this tension, as they have traditionally been associated with the free provision of information services to their clients. In the current economic climate of accountability and user-pays, this may no longer be possible.

Peter Deutsch, co-creator of thearchie Internet resource locator, has expressed a counter view:

"Information has never been 'free' since it costs money to create it, store it and serve it. What we can hope, is that it will be 'freely available' to everyone and I think the Internet holds great promise as a delivery mechanism for this"<sup>3</sup>

In this sort of model, information becomes a utility like gas, water and electricity. The difficulty is coming up with workable charging and payment mechanisms.

This confusion about the nature of information, and its cost/value, can significantly affect our ability to deal with the changing information scene.

## **2. Lessons from the past**

George Santayana said that those who do not study history are doomed to repeat it. What are some of the lessons from the recent history of information technology that we can apply as we grapple with the changes taking place in the information world today? This section selects some developments, extracts from them lessons and

---

<sup>3</sup> Dern [1993], p. 15.

suggests an application of those lessons to the problems of managing networked information.

## 2.1 Arrival of the mainframe

**Scenario.** The development of the mainframe computer and its use in centralised data processing departments at universities and in large corporations was the first large-scale application of computer technology to the management of information.

The early days of the mainframe era have been characterised as a time of high priests<sup>4</sup>, dressed in formal white garments, who performed arcane rituals as they consulted the computer for its prophecies. While this description is somewhat exaggerated, it does reflect the distance between the user and the information technology. Many users felt intimidated and uneasy in their dealings with the mainframe and its staff, and this transferred itself into an unwillingness to use the computer unless absolutely necessary.

**Lesson.** It is easy to exclude people from new technologies, without even trying. Potential users need to be actively encouraged and assisted to use such technologies, often via practical examples of how it might help them. Telling them how to use it and what it does is rarely enough.

**Application.** New Internet access tools need to be designed to be as inclusive as possible, preferably based on ways of working that people already understand. Training sessions for new users should be hands-on wherever possible, with lots of real-world, relevant examples.

## 2.2 Advent of personal computing

**Scenario.** The introduction of the personal computer (PC) heralded a new era in the use of technology to manage information. Users were able to exercise personal control over the technology, without an intervening priesthood. As well, they were able to run applications, like spreadsheets, that were not available on the mainframes. In the early days of the PC, the common criticism was that it was only a 'toy' computer. Nonetheless, people purchased and used these early machines for a wide range of purposes.

**Lesson.** The element of personal control is important, and people are prepared to use machines with less power to retain that control.

**Application.** Internet access tools need to retain that sense of user control, and be based on the user's desktop for this reason, as well as good technological ones.

## 2.3 Rise of the graphical user interface

**Scenario.** The phenomenal success of graphical user interfaces (or GUIs - also called direct manipulation user interfaces ) has been a feature of the late 1980's and the early 1990's. People are now seriously forecasting the 'death of DOS' and large software companies like Microsoft and WordPerfect are talking about the last versions of their DOS products.

---

<sup>4</sup> Almost never priestesses, although they were sometimes permitted to serve as acolytes.

The early criticism of these interfaces was that they offered ease of use but not power, that they were 'too easy for real users'. Now that software manufacturers have improved at graphical interface design, it is clear the power/ease of use dichotomy is a false one. Ease of use and power can, and should, coexist. The success of these interfaces (witness the speed with which Windows has been adopted, and the rash of GUIs layered onto Unix) indicates that people want interfaces that engage them, and that provide increased visual feedback. 'Playing' with the Macintosh Finder is more 'fun' than typing in a DOS command.

As an aside, Michael Swaine has argued that rather than talking about easy interfaces and power interfaces, we should be discussing the difference between hacker interfaces and luser interfaces<sup>5</sup>. He believes that hackers are interested in whatever the machine is able to do; that is in the machine as an end in itself. Lusers are interested in using their machines to get their jobs done, but have little interest in the machine other than this. A third group is emerging in the personal computer world who are neither hackers or lusers but want to push the machine in creative ways. The only way to do this is to interact with the machine in the most flexible way possible, via some form of language. Currently this is called scripting, but may be something quite different in the future. Notice that many of the high-end applications currently available have their own macro/script language built-in. It is interesting that Apple, the original popularisers of the GUI for the mass market, have now designed Applescript as an extension to their operating system.

This third group of users are the ones currently driving many of the developments in Internet tool design. They are interested in pushing the boundaries of what is possible with current technology, but not as an end in itself. The focus is firmly on producing tools that help solve real problems.

**Lesson.** Ease of use and power can coexist. Users want interfaces that engage them and that are (preferably) built on familiar metaphors.

**Application.** Access tools need to be based on familiar user-interface technology; if they are not there will be considerable resistance to their use from novice users. The tools must not sacrifice power for ease of use.

## 2.4 Advent of multi-media

**Scenario.** Just as GUIs were the flavour of the late 1980's, Multimedia seems to be the flavour of the early 1990's. Everybody seems to want to claim that they are 'doing' multimedia, 'working towards' multimedia, or 'pioneering' multimedia. Driven by increases in desktop computing power, more capable display devices, and the widespread use of CD-ROM as a capacious (if slow) publishing medium, multimedia has well and truly come to the desktop. Even after discounting all the hype, it appears clear that multimedia in some form will constitute a significant part of future desktop interfaces. Witness the recent release of the Apple AV range of desktop computers, sporting integrated video capture, the Plaintalk speech recognition and generation technologies, and built-in CD-ROM<sup>6</sup>. The creation of documents containing more than plain ASCII text has been commonplace for years,

---

<sup>5</sup> Swaine [1993], p. 42.

<sup>6</sup> For more details consult *PC Week*, August 18, 1993, p. 15, among other articles.

and current word-processors contain facilities for adding voice annotations to text. For many users these new developments are opening their eyes to the inherent visual poverty of the monospace, monochrome text that has been so prevalent for so long.

**Lesson.** People are sensory beings, and appreciate having as many senses stimulated as fully as possible, in computer interfaces as in much else.

**Application.** Internet tools need to provide a rich sensory environment, and support multimedia information as fully and transparently as possible.

### 3. Solutions for the future

The online information community worldwide faces a number of challenges as it tries to keep pace with the changes identified earlier. How can we encourage our colleagues to use the wealth of information resources currently available? How can we ourselves keep track of new developments? How can we contribute to the development of new access tools and methodologies? This section will consider current tools and suggest some directions for future work, bearing in mind the lessons from the past just outlined.

#### 3.1 The 'sufficient bandwidth' myth

Before considering some of the possible solutions, it is necessary to dispose of one common myth about online communications, the myth of sufficient bandwidth. There is not now, and will never be, enough bandwidth to support all we want to do. There are upper limits set by physical constants, and the laws of supply and demand ensure that demand will always outstrip supply. Chip Morningstar, in an article discussing an early virtual reality system<sup>7</sup>, points out that what we want to communicate will always be pushing the envelope; compressed sound and high-resolution graphics images now, full-motion HDTV video clips or massive volumetric time-series data in the future. Every time the link between Australia and the US is upgraded it saturates after a matter of months. Peter Deutsch points out that whenarchie was started the link to the outside world was only 9600 bps:

“Of course, we saturated it very quickly and followed its continued expansion with our own exponential growth curve, but I believe that would have happened no matter how much bandwidth we had”<sup>8</sup>

There is a fixed upper limit for the Internet, even for fibre optics, and we need to work within it. As Deutsch argues:

“What we have to realise is the Internet changes what is possible in terms of services. We need to find ways to use its properties in our tools, not just shoehorn existing services onto the net...At the same time, people connecting onto the Internet are finding they can do quite a lot with slow links.”<sup>9</sup>

---

<sup>7</sup> MorningStar and Farmer [1992], p. 279.

<sup>8</sup> Dern [1993], p. 14.

<sup>9</sup> *ibid.*

### 3.2 Current tools

Existing Internet access tools<sup>10</sup> divide loosely into two groups.

What might be characterised as the 'older' tools include text-only email and mainframe<sup>11</sup> implementations of telnet<sup>12</sup> and ftp. These tools were designed to enable users to do at a distance what they could already do on their local machines. Thus telnet provides remote logons, ftp provides remote copying, and so on. These tools generally use a command-line interface and talk to standard online terminals. Rather than being built around a true client-server model, the tools assume similar 'symmetric' programs running on the remote machines. Thus, with an implementation of ftp running on two machines, either end can initiate action. Useful though these tools were and are, they provide nothing that is conceptually different from working with a local machine.

The 'newer' tools differ in four important respects from these older implementations. Firstly, they make use of the client-server approach to provide quite different front-end (client) and back-end (server) programs, communicating via standardised protocols. Secondly, they often come in both GUI and terminal-based versions. The GUI versions usually require TCP/IP as the underlying data transfer mechanism, often over Ethernet, and provide the richest functionality and greatest ease of use. Note that the terminal versions, despite being handicapped by limited displays and much lower bandwidths, often do a surprisingly good job. Thirdly, they support facilities that are different to working on a local machine, things that are very different to the normal operations of logging on and copying files. Fourthly, they are based around a global world view; that is, they require the Internet or something like it to be meaningful and truly useful. The best known of these newer tools arearchie, WAIS and Gopher. Less well known are Veronica, WWW and Mosaic. Hytelnet provides a way for users to manage information about these tools and information sources. There are also GUI versions of ftp and telnet.

Archie<sup>13</sup> currently provides a searchable database of files available from anonymous ftp sites worldwide. The database is kept current by periodically 'polling' the ftp sites for changes to their holdings. Thearchie database is invaluable for locating particular files, if the name or part of it is known. Bunyip Information Systems Inc. is currently seeking to extend its usefulness to provide general centralised searching of distributed collections of information.

WAIS<sup>14</sup>, standing for Wide Area Information Servers, is a distributed networked database system, based on an extension of the NISO Z39.50-1988 standard. This

---

<sup>10</sup> For readability's sake, reference to a particular tool will imply the protocols that support that tool. This is to avoid making clumsy distinctions between, for instance, the ftp protocol and the ftp tool. Most users only see the tool, and don't care about the underlying protocols as long as they don't break.

<sup>11</sup> This is used as a convenient, if sloppy, shorthand for all multi-user systems, including Unix boxes and VAXen.

<sup>12</sup> Concise definitions for unfamiliar terms can be found in the glossary at the end of this paper.

<sup>13</sup> See Deutsch [1992] for more details, and Dern [1993] for discussion of Bunyip.

<sup>14</sup> See Kahle [1992], Stanton [1992] and Stein [1991].

describes a model for an “originating application” (client) to query databases at one or more “targets” (servers). Wiggins [1993] summarises its capabilities as:

“WAIS allows a user to do keyword searches for documents, scanning a single server or multiple servers. WAIS responds to a search with a list of documents sorted by a weighted score--the higher the score, the better the match to the search. WAIS is strong in its ability to allow users to sift through a large body of documents stored in servers across the Internet. It is more appropriate for finding a small set of documents among a large set of candidates than for serving as a menuing or browsing tool”<sup>15</sup>

Hundreds of WAIS servers, often specialising in particular subjects areas, are now available, running on various back-end systems. Both GUI and terminal-based WAIS clients exist, although the terminal client swais is somewhat clumsy to use. WAIS also allows retrieval of the documents themselves once found.

Gopher<sup>16</sup> has been attracting a lot of interest recently. It too is based on a true client/server model. Client software running on a user’s personal workstation is preferred due to the better user interface and access to still images, audio files, and other resources, but terminal gopher clients also exist.

Rich Wiggins, one of the original Gopher developers, and now overall Gopher coordinator describes Gopher like this:

“In a nutshell, Gopher offers access to files and interactive systems using a hierarchical menu system. The organisation of the menu is defined by the administrator of each server. The resources that each menu item describes, such as ASCII files, Telnet sessions to other computers, and submenus, are called “documents.” The scope of a Gopher server could be limited to a campus or to a subject area. Or, a Gopher server could include a catalogue of a variety of Internet resources. ... The user connects to one of the thousands of Gopher servers now in production around the world and receives an initial (or “root”) menu. When you use a tool like Archie, you are asking the server a specific question (e.g, “Tell me what FTP sites can provide a copy of the program named PKZIP”). When you connect to a Gopher server, you are asking it “Give me a menu listing the resources you have to offer.”<sup>17</sup>

Once resources have been located, they can often be retrieved to the local workstation, depending on the resource type.

The difficulty with ‘Gopherspace’ is knowing which Gopher server contains the required resources. Enter Veronica<sup>18</sup>, or the “Very Easy Rodent-Oriented Netwide Index to Computerized Archives.” Like archie, Veronica polls its target servers periodically to build its database. A user connecting to Veronica specifies a keyword to search for and is returned a list of document titles from throughout Gopherspace

---

<sup>15</sup> Wiggins [1993], p. 48.

<sup>16</sup> Wiggins [1993], Stanton [1992].

<sup>17</sup> Wiggins [1993], pp. 7-8.

<sup>18</sup>Wiggins [1993], pp. 36 - 38.

that match. Currently Veronica suffers from performance problems and difficulties in interpreting its results, but these problems are being worked on.

In contrast to the hierarchical, menu-based, one hierarchy per server organisation of Gopher, World-Wide Web<sup>19</sup> (also written as WWW or W<sup>3</sup>) is a distributed hypertext system. Wiggins [1993] comments:

“Advocates of WWW claim that this is an essential feature--the only effective way to deliver documents chosen from a vast universe of information is via hypertext. They point out that providing pointers within the text of documents can make it far easier for users to find and peruse materials of interest. The WWW model lets users seamlessly cruise from server to server and from topic to topic. It allows the user to retrace steps as well. Like Gopher, WWW can launch Telnet sessions to connect to online services.

WWW can serve as a browsing tool much as Gopher does. Under Gopher, it is common for a document to include pointers such as “Look in the 'Campus Eateries' folder for more information.” To follow that advice, the user must leave the current document and open the folder in question. Under WWW, the pointer text could highlight “Campus Eateries” within every document where it would be helpful to mention it; a click on the embedded title would open the referenced document. Such multiple links are unobtrusive and do not require the user to hunt through other folders. It is hard to deny that embedded links are easier for the user to navigate.”<sup>20</sup>

With WWW, the location of the target of these links could be another directory on the current machine, another machine on the same campus, or a machine on the other side of the world. The ‘Web’ is truly international. There are a number of WWW servers and both GUI and terminal clients.

Mosaic and Hytelnet are two very different ways of managing the diversity of tools, protocols and information sources.

Mosaic is a sort of ‘super-client’ written by the National Centre for Supercomputing Applications. It provides connections to World-Wide Web, Gopher, USENET News, FTP, and supports numerous document types including various image formats, audio, and PostScript. Mosaic aspires to:

“provide the user with a customizable hypertext view of the networked universe. ...Mosaic also provides helpful navigation aids such as a complete history of where a user has been in her wanderings through the Internet; it even uses color to highlight menu items selected earlier in a session”<sup>21</sup>.

Mosaic currently only runs under XWindows, but is being ported to Macintosh and perhaps Microsoft Windows.

---

<sup>19</sup> Berners-Lee [1992]

<sup>20</sup> Wiggins [1993], p. 49.

<sup>21</sup> Wiggins [1993], p. 46 - 47.

Unlike Mosaic, HYTELNET is not itself a client tool but rather a program that helps users identify needed Internet systems. Created by Peter Scott<sup>22</sup> of the University of Saskatchewan it is essentially a hypertext directory of Internet sources and resources including a wide range of OPACs and CWISes. Versions of HYTELNET exist for UNIX and VMS (with direct telnet access), IBM-PC (as a TSR or memory-resident utility), and Macintosh (as a Hypercard stack). Frequent updates to the data files are provided through an electronic mailing list. HYTELNET is essentially only a resource discovery tool. It does not retrieve files or perform index searches like Gopher - other programs are necessary for this.

### **3.3 New challenges and associated issues**

Challenges abound wherever one looks on the Internet at present. There are issues that need to be resolved at every level from the format of documents to the very ways in which information itself is organised. How can our 'lessons from the past' inform what we do in these areas in the future?

#### **3.3.1 Data and documents**

Internet data comes in two basic forms - ASCII data, and binary data<sup>23</sup>. ASCII data is largely used for email and text documents. Binary data covers everything else.

ASCII data is the 'lowest common denominator'. While it can be read and written by all current computers, it has a number of significant limitations. It provides no support for languages that use any additions to the 26 letter Roman alphabet. It certainly provides no support for languages that do not use the Roman alphabet at all. It severely restricts the range of formatting options, and provides no way to represent different fonts and other typographic information. ASCII is so limiting that John Warnock, co-inventor of Postscript, has referred to it as the 'ASCII jail'. The proposed new Unicode double-byte standard would accommodate all current scripts, but would not solve the typographic problems. These will have to be addressed at the level of the binary document.

Binary data consists of program files, documents in a bewildering variety of formats, and encoded data files. Most users want to be able to work with documents created by other people. These documents are increasingly being created on a range of hardware and software platforms, and include much more than simple text. They already can contain multiple fonts, layout instructions, graphics and scanned images. Voice and video annotation is being provided by some vendors. There is no current standard that allows for the creation of such a document on one machine and its viewing on a range of others. Adobe is proposing its Acrobat<sup>24</sup> technology, based on Postscript and Multiple Master fonts, as a generic solution to the interchange of documents containing text, graphics and images. At present this technology has difficulties in supporting editing of received documents.

---

<sup>22</sup> For more details, refer Scott [1992].

<sup>23</sup> These are in fact the only two types of data that ftp understands.

<sup>24</sup> Refer Warnock [1992] for details.

In the domain of what is currently text-only email, the proposed new MIME standard is a way of providing richer email documents by building on top of existing email protocols. MIME is also being investigated as another generic document interchange solution.

It may be that all of these technologies need to succeed. What is certain that 7-bit ASCII is now clearly inadequate as the lowest common denominator. As multi-sensory computer users we demand much more.

### **3.3.2 *Managing electronic information***

Many of our techniques for managing information are also conditioned by our past, in this case our pre-electronic past. Library cataloguing rules are designed to describe physical items owned by and residing in libraries, and do not cope well with the fluid nature of electronic information. Electronic documents can be changed much more quickly than can a book. Is each version of such a document equivalent to a new edition? Electronic documents can be copied to multiple locations very quickly. Should one try to monitor this spread? Electronic documents can easily be moved from server to server. How can one keep track of where a document now resides?

An attempt to address some of these issues is being made. USMARC now includes a general field called Electronic Location and Access for this purpose. The rationale is as follows:

“In effect, it was decided that FTP sites, list servers, and the like constituted electronic locations that conceptually parallel physical ones. The paper form of a document might be on a shelf in a library, while a bitmapped form might be available from a file server on the Internet. A new field was invented for ‘Electronic Location and Access’ (856), including data elements for type of access (e.g., e-mail, FTP, and Telnet), host name, path name, file name, and similar information necessary to access or retrieve a data resource over the network.”<sup>25</sup>

The difficulty with this scheme is that the location and access information can change very quickly. The solution lies in another scheme being developed by the Internet Engineering Task Force. Under this scheme, a Universal Resource Identifier (URI)--much like the ISBN--would be assigned to each object by the originating agency. A Universal Resource Locator (URL), similar in concept to the Electronic Location and Access field, would identify a location. According to Caplan [1993]:

“Only URIs would be imbedded in the bibliographic description, and computers would associate the URI with one or more URLs in much the same way an Internet host name (HARVARDA.HARVARD.EDU) is associated with its IP address (128.103.60.11) by the name server system. However, someone needs to do all this; an infrastructure needs to be developed and responsible agencies in agreement on responsibilities and procedures. Once this mechanism is in place, we can decide what to do next with the Electronic Location and Access field. Meanwhile, it allows us to begin building records and testing the feasibility of catalog access to electronic data resources.”

---

<sup>25</sup> Caplan [1993], p. 64

Finally, there is the problem of what is an electronic information source anyway. Is it reasonable to distinguish between documents and systems/services? Is there a distinction between a database and the retrieval system required to access it?

There is clearly a good deal of work to be done before management of electronic information approaches the state of the art with respect to traditional materials. This work needs to be done in such a way as to support the development of efficient, intuitive and transparent ways for users to access the information.

### **3.3.3 New information tools**

Most of the latest generation of Internet access tools are governed by the Windows-Icons-Menus-Pointing (WIMP) desktop metaphor that has not changed much since the days of the Alto at Xerox Parc. The use of the desktop as the main structuring device for the interface may not always be the most appropriate for all application domains. What we need are completely new tools to support new ways of working with information. A number of researchers are developing such new ways of working with large amounts of structured information.

One of the more promising developments is the work being done by Robertson et. al. [1993] at Xerox Parc. They have developed an experimental system called the Information Visualizer, which exploits the capabilities of the next generation of desktop workstations to provide a range of new information tools. 3D visualisations including animation, use of colour and shadows allow the user to dynamically explore hierarchical organisations, linear sequences, spatial structures, continuous data, and unstructured information. The properties of these visualisations are tuned to the pattern detection properties of the human perceptual system.

Developers of new Internet access tools must keep abreast of these latest developments in user interface techniques. The information available on the Net is vast in scope, loosely structured, constantly changing, and wildly heterogenous. Only the best techniques will serve.

### **3.3.4 Information organisation**

Finally, we need to consider the very ways in which we organise information itself. The great majority of information on the Net is organised in hierarchies of one form or another. Hierarchies have a long and proud tradition in the management of networked information. The library of King Assurbanipal of Assyria (667-628 BC) was organised along hierarchical lines by subject matter. We are all familiar with hierarchical forms of organisation and they have become widespread in all organisations.

Storage hierarchies lie at the root of most file systems, particularly those underlying DOS, Unix and the Macintosh. These file system hierarchies then influence the ways in which information on such systems is structured. Most ftp archives use a subject division which maps directly onto the underlying directory structure. Most gopher servers organise their information into nested hierarchies of menus and submenus.

For the most part, hierarchical structures work well, but they have some real limitations for networked information. These limitations are particularly obvious if the hierarchy is inherent in the structures used to store the information (as opposed to the hierarchy only being reflected in some sort of additional index).

Firstly, it is necessary to understand the hierarchy to be able to efficiently locate an item. The path taken at each branching point reduces the number of items left that can be located. This can be a strength if the correct path is chosen by speeding the location process. If not, the item can not be found without backtracking. Anyone who has tried to find a file on someone else's hard disk will recognise the problem.

Secondly, most hierarchies provide only one possible location for any item. Similar items cannot be stored in more than one place unless inconsistently placed. This can be a problem if one wants to store a document under Articles/Current/VALA-93 and Documents/Internet/Discussion.

There have been some partial solutions to these problems. One solution to the 'one location' problem has been to implement 'cross-links' within file system hierarchies. This is done via aliases on the Macintosh and link files under Unix. Another attempted solution is to overlay the original hierarchy with another organisational structure altogether. Windows program groups are a form of this.

Most of these solutions, however, are patches and do not address the real problems with hierarchies. As the Net grows in size and complexity, hierarchical forms of organisation will prove increasingly inadequate.

A more challenging alternative is to abandon hierarchies altogether and try for another organisation altogether. This is the approach taken by advocates of hyperlink schemes, such as the WWW. Here there is no implied hierarchy, and users can follow links to go where they will. But hyper solutions have their problems, and users complain of not knowing 'where' they are, whatever that means in the context of organised information.

What may be needed is a completely new way altogether to organise and access information. If so, it will need to draw upon user's experiences and intuitions, while providing the necessary extra power. Perhaps some modification of facet analysis might prove a useful starting point. Only further research will tell.

## **4.0 Conclusion**

This paper has tried to examine the current situation with respect to networked information in terms of change and its consequences. Some of these dimensions of change have been considered, and their impact discussed. Particular times of transition in the recent past of information technology have been examined. Some possible lessons have been identified, and their application to the Net community noted. Finally, some challenges confronting the management of networked information have been considered in the light of the lessons learnt.

There is a wealth of networked information becoming available on the Net. We are mostly trying to manage this information with old techniques and tools based on outdated methodologies. The real challenge facing the international information community may well be to move beyond these old ways of doing things and create the fundamentally new ways of working with information that will be required to meet the information challenges of the future.

## References

- Caplan, Priscilla [1993] "Cataloging Internet Resources." *The Public- Access Computer Systems Review* vol. 4, no. 2, pp. 61-66. To retrieve this file, send the following e-mail message to [LISTSERV@UHUPVM1](mailto:LISTSERV@UHUPVM1) or [LISTSERV@UHUPVM1.UH.EDU](mailto:LISTSERV@UHUPVM1.UH.EDU): GET CAPLAN PRV4N2 F=MAIL.
- Clarkson, M. [1991], "An Easier Interface", *Byte*, February 1991, pp. 277 - 282.
- Dern, Daniel P. [1993]. "Peter Deutsch, 'archie', and Bunyip Systems", *Internet World*, Vol. 4, No. 2 pp. 11 - 16.
- Deutsch, Peter [1992]. "Resource Discovery in an Internet Environment-- The Archie Approach." *Electronic Networking: Research, Applications and Policy* 2, no. 1, pp. 45-51.
- Berners-Lee, Tim et al. [1992] "World-Wide Web: The Information Universe." *Electronic Networking: Research, Applications and Policy* vol. 2, no. 1, pp. 52-58.
- Kahle, Brewster et al. [1992] "Wide Area Information Servers: An Executive Information System for Unstructured Files." *Electronic Networking: Research, Applications and Policy* vol. 2, no. 1, pp. 59-68.
- Lynch, Clifford A. [1990] "Information Retrieval as a Network Application." *Library Hi Tech* vol. 8, no. 4, pp. 57-72.
- Morningstar, C, and Farmer, R. F., [1992] "The Lessons of Lucasfilm's Habitat" in *Cyberspace: the first steps*, edited by Michael Benedikt, MIT Press, Cambridge, Massachusetts, 1991.
- Robertson, G. G., Card, S. K., and MacKinlay, J. D. [1993], "Information Visualization using 3D Interactive Animation", *Communications of the ACM*, Vol. 36, No. 4, pp. 57 - 71.
- Scott, Peter. [1992] "Using HYTELNET to Access Internet Resources." *The Public- Access Computer Systems Review* vol. 3, no. 4, pp. 15-21. To retrieve this file, send the following e-mail message to [LISTSERV@UHUPVM1](mailto:LISTSERV@UHUPVM1).[UH.EDU](mailto:LISTSERV@UHUPVM1.UH.EDU): GET SCOTT PRV3N4 F=MAIL.
- Stanton, Deidre E. [1992]. *Tools For Finding Information on the Internet: HYTELNET, WAIS and Gopher*. Murdoch University Library, Murdoch, WA. [Available for anonymous ftp from host [infolib.murdoch.edu.au](ftp://infolib.murdoch.edu.au), directory pub/soft/nirtools; filenames nirtools.txt, nirtools.ps, nirtools.wp]
- Stain, Richard M. [1991], "Browsing through Terabytes", *Byte*, May 1991, pp. 157 - 164.
- Swaine, M., [1993] "Users and Lusers", *MacUser*, April 1993, 41-43.
- Warnock, J. [1992], "The new age of documents", *Byte*, June 1992, pp. 257 - 260.
- Wiggins, Rich. [1993] "The University of Minnesota's Internet Gopher System: A Tool for Accessing Network-Based Electronic Information." *The Public-Access Computer Systems Review* vol. 4, no. 2, pp. 4-60. To retrieve this file, send the

following e-mail messages to `LISTSERV@UHUPVM1` or `LISTSERV@UHUPVM1.UH.EDU`: `GET WIGGINS1 PRV4N2 F=MAIL` and `GET WIGGINS2 PRV4N2 F=MAIL`. -

## Glossary<sup>26</sup>

**Anonymous FTP.** Anonymous FTP servers allow users to retrieve files without the need for assigned user IDs (literally the word “anonymous” is used as the login ID). This service is offered by many sites on the Internet. The Australian ftp site, `archie.au`, provides ‘shadows’ of a number of significant US archives.

**Archie.** A network service that allows users to discover which anonymous FTP sites house particular files of interest. Developed at McGill University (and now Bunyip Information Systems), future versions of Archie may allow searching for resources by abstract, author name, and other criteria. To access the Australian archie site, telnet (q.v.) to `archie.au` and login as `archie`.

**ASCII file.** A file encoded in the 128-character ASCII (American Standard Code for Information Interchange) character set. The term “flat ASCII file” is often used to refer to a simple text file, with no embedded special formatting codes or binary data. In FTP transfers, “text” and “ASCII” are synonymous.

**Binary file.** Binary files consist of streams of bytes whose meaning is defined by some external format standard. For example, executable computer programs are binary files. In FTP transfers, a binary file is specified by “bin” or “image” settings.

**Client/Server.** A model for distributing system functions between client software, residing on a user's workstation, and server software, residing on a host. The host could be a UNIX workstation, a mainframe, or another type of computer. The client handles most of the information presentation functions, and the server handles most of the database functions. A protocol specifies how the two should communicate. The client/server model is growing in popularity as workstations and networks grow in power.

**CWIS (Campus-Wide Information System).** A university or college information system that offers integrated access to documents (e.g., course schedules, lists of current events, and academic job openings) and systems (e.g., online catalog). CWISes began on mainframes and are now commonly implemented under the client/server model. Gopher (q.v.) and WWW (q.v.) are commonly used for CWISes; other CWISes exist (for instance, the TechInfo software from MIT).

**FTP (File Transfer Protocol).** A standard protocol for sending files from one computer to another on TCP/IP networks, such as the Internet. This is also the command the user usually types to start an FTP session.

**Hypertext.** A scheme for supporting embedded links within documents. While browsing a document with hypertext links, a user can select one of those links and quickly move to the document it points to. Popularized by the HyperCard Macintosh program.

---

<sup>26</sup> Modified from that included as part of Wiggins[1993].

HYTELNET. A hypertext database of Internet systems, such as online catalogs and CWISes. The PC version of the program operates as a memory-resident program and can be accessed while other programs are running. The Macintosh version of the program is implemented as a Hypercard stack and can pass commands to a terminal program.

MIME (Multipurpose Internet Mail Extensions). Initially, an extension of Internet mail standards to allow binary data to be embedded in SMTP mail. Since its introduction in 1992, MIME has been implemented on several computer platforms (often under the name "Metamail"), and it is increasingly viewed as the appropriate standard for handling multimedia information to be moved across dissimilar computing environments, whether sent via e-mail or otherwise.

Mosaic. An integrated client program developed by the National Center for Supercomputing Applications. Mosaic acts as a client for multiple Internet services, including Gopher, WAIS, WWW, USENET News, and FTP. Currently implemented on UNIX systems supporting X Windows and Motif. Versions of Mosaic for the Macintosh and Microsoft Windows are expected.

PostScript. A printer description language from Adobe. PostScript is the dominant format used for desktop publishing. Documents in PostScript format are commonly shared across the Internet and printed on laser printers after retrieval from a remote archive.

SGML (Standard Generalized Markup Language). SGML is a scheme (and an ISO standard) for embedding structural information within a document. SGML is popular in scholarly and electronic publishing as a way to support multiple views of a document. An SGML-compliant set of structures called HTML is used by World-Wide Web.

SMTP (Simple Mail Transfer Protocol). A protocol for sending e-mail messages between computers on TCP/IP networks, such as the Internet. The user does not run a program named SMTP; instead, various e-mail packages know how to utilize this protocol.

TCP/IP (Transmission Control Protocol/Internet Protocol). Technically, TCP and IP are separate protocols; together they allow computers on the Internet to communicate by providing a reliable way for bytes to be delivered in order over a network connection. Connections are made to TCP "ports," allowing multiple connections per machine. A port is described by a number (e.g., Gopher servers typically use port 70).

Telnet. The TCP/IP protocol for remote terminal sessions. Usually implemented as a command of the same name.

Uniform Resource Identifier. An umbrella term for standards that describe Internet resources in a uniform way. The IETF is considering a Uniform Resource Locator, which will be a standard way to name a particular document on a particular network server. Another proposed standard, the Uniform Resource Number, will be a unique number (analogous to an ISBN) assigned to a document or resource regardless of its location and other resource descriptors.

USENET News. A distributed discussion list system, much of whose traffic is carried over the Internet. USENET News services consist of news “feeds” typically provided by one or more servers on a campus network, and news “readers” (i.e., client software that communicates with the server over a news delivery protocol).

Veronica. A service that provides a Internet-wide index of Gopher document titles. Developed at the University of Nevada, Veronica servers periodically poll all the Gopher servers they can connect to and build a title index that can, in turn, be pointed to as a standard Gopher index.

VT 100. The dominant communications protocol for full-screen terminal sessions. The VT 100 standard was defined by the Digital Equipment Corporation in the 70s. Most terminal emulation software packages (e.g., Kermit and PROCOMM) implement VT 100 or its descendants.

WAIS (Wide Area Information Servers). Based on an extension of the Z39.50-1988 standard, WAIS allows searches of documents on one or more WAIS servers. Originally promulgated by Thinking Machines Corporation, WAIS is now offered in a commercial version (by WAIS, Inc.) and a public domain version (by the Clearinghouse for Networked Information Discovery and Retrieval). The WAIS search engine is commonly used by Gopher servers. An index document under the UNIX Gopher server can point to a stand-alone WAIS server.

World-Wide Web. A network-based hypertext document delivery system. Developed at CERN in Switzerland, WWW is growing in popularity. WWW supports embedded links to documents.

Z39.50. A NISO standard defining a client/server model for searching bibliographic and other databases.

Z39.58. A NISO standard describing a Common Command Language for online catalogs.