

Access : Not Just Wires

Karen Coyle
October 8, 1994

I have to admit that I'm really sick and tired of the Information highway. I feel like I've already heard so much about it that it must be come and gone already, yet there is no sign of it. This is truly a piece of federal vaporware.

I am a librarian, and I find it especially strange to have dedicated much of your life to the careful tending of our current information infrastructure, our libraries, only to wake up one morning to find that the entire economy of the nation depends on making information commercially viable. There's an element of Twilight Zone about this because libraries are probably our most underfunded and underappreciated of institutions, with the possible exception of day care centers.

It's clear to me that the information highway isn't much about information. It's about trying to find a new basis for our economy. I'm pretty sure I'm not going to like the way information is treated in that economy. We know what kind of information sells, and what doesn't. So I see our future as being a mix of highly expensive economic reports and cheap online versions of the National Inquirer. Not a pretty picture.

This is a panel on "access". But I am not going to talk about access from the usual point of view of physical or electronic access to the FutureNet. Instead I am going to talk about intellectual access to materials and the quality of our information infrastructure, with the emphasis on "information". Information is a social good and part of our "social responsibility" is that we must take this resource seriously.

From the early days of our being a species with consciousness of its own history, some part of society has had the role of preserving this history : priests, learned scholars, archivists. Information was valued; valued enough to be denied to some members of society; to be part of the ritual of belonging to an elite.

So I find it particularly puzzling that as we move into this new "information age" that our efforts are focused on the machinery of the information system, while the electronic information itself is being treated like just so much more flotsam and jetsam; this is not a democratization of information, but a devaluation of information.

On the Internet, many electronic information sources that we are declaring worthy of "universal access" are administered by part-time volunteers; graduate students who do eventually graduate, or network hobbyists. Resources come and go without notice, or languish after an initial effort and rapidly become out of date. Few network information resources have specific and reliable funding for the future. As a telecommunications system the Internet is both modern and mature; as an information system the Internet is an amateur operation.

Commercial information resources, of course, are only interested in information that provides revenue. This immediately eliminates the entire cultural heritage of poetry, playwriting, and theological thought, among others.

If we value our intellectual heritage, and if we truly believe that access to information (and that broader concept, knowledge) is a valid social goal, we have to take our information resources seriously. Now I know that libraries aren't perfect institutions. They tend to be somewhat slow-moving and conservative in their embrace of new technologies; and some seem more bent on hoarding than disseminating information. But what we call "modern librarianship" has over a century of experience in being the tender of this society's information

resources. And in the process of developing and managing that resource, the library profession has understood its responsibilities in both a social and historical context. Drawing on that experience, I am going to give you a short lesson on social responsibilities in an information society.

Here are some of our social responsibilities in relation to information :

- Collection
- Selection
- Preservation
- Organization
- Dissemination

Collection

It is not enough to passively gather in whatever information comes your way, like a spider waiting on its web. Information collection is an activity, and an intelligent activity. It is important to collect and collocate information units that support, complement and even contradict each other. A collection has a purpose and a context; it says something about the information and it says something about the gatherer of that information. It is not random, because information itself is not random, and humans do not produce information in a random fashion.

Too many Internet sites today are a terrible hodge-podge, with little intellectual purpose behind their holdings. It isn't surprising that visitors to these sites have a hard time seeing the value of the information contained therein. Commercial systems, on the other hand, have no incentive to provide an intellectual balance that might "confuse" its user.

In all of the many papers that have come out of discussion of the National Information Infrastructure, it is interesting that there is no mention of collecting information : there is no Library of Congress or National Archive of the electronic information world. So in the whole elaborate scheme, no one is responsible for the collection of information.

Selection

Not all information is equal. This doesn't mean that some of it should be thrown away, though inevitably there is some waste in the information world. And this is not in support of censorship. But there's a difference between a piece on nuclear physics by a Nobel laureate and a physics diorama entered into a science fair by an 8-year-old. And there's a difference between alpha release .03 and beta 1.2 of a software package. If we can't differentiate between these, our intellectual future looks grim indeed.

Certain sources become known for their general reliability, their timeliness, etc. We have to make these judgments because the sheer quantity of information is too large for us to spend our time with lesser works when we haven't yet encountered the greats.

This kind of selection needs to be done with an understanding of a discipline and understanding of the users of a body of knowledge. The process of selection overlaps with our concept of education, where members of our society are directed to a particular body of knowledge that we hold to be key to our understanding of the world.

Preservation

How much of what is on the Net today will exist in any form ten years from now ? And can we put any measure to what we lose if we do not preserve things systematically ? If we can't preserve it all, at least in one safely archived copy, are we going to make decisions about preservation, or will we leave it up to a kind of information Darwinianism ? As we know, the true value of some information may not be immediately known, and some ideas gain in value over time.

The commercial world, of course, will preserve only that which sells best.

Organization

This is an area where the current Net has some of its most visible problems, as we have all struggled through myriad gopher menus, ftp sites, and web pages looking for something that we know is there but cannot find.

There is no ideal organization of information, but no organization is no ideal either. The organization that exists today in terms of finding tools is an attempt to impose order over an unorganized body. The human mind in its information seeking behavior is a much more complex question than can be answered with a keyword search in an unorganized information universe. When we were limited to card catalogs and the placement of physical items on shelves, we essentially had to choose only one way to organize our information. Computer systems should allow us to create a multiplicity of organization schemes for the same information, from traditional classification, that relies on hierarchies and categories, to faceted schemes, relevance ranking and feedback, etc.

Unfortunately, documents do not define themselves. The idea of doing WAIS-type keyword searching on the vast store of textual documents on the Internet is a folly. Years of study of term frequency, co-occurrence and other statistical techniques have proven that keyword searching is a passable solution for some disciplines with highly specific vocabularies and nearly useless in all others. And, of course, the real trick is to match the vocabulary of the seeker of information with that of the information resource. Keyword searching not only doesn't take into account different terms for the same concepts, it doesn't take into account materials in other languages or different user levels (i.e. searching for children will probably need to be different than searching done by adults, and libraries actually use different subject access schemes for childrens' materials). And non-textual items (software, graphics, sound) do not respond at all to keyword searching.

There is no magical, effortless way to create an organization for information; at least today the best tools are a clearly defined classification scheme and a human indexer. At least a classification scheme or indexing scheme gives the searcher a chance to develop a rational strategy for searching.

The importance of organizational tools cannot be overstated. What it all comes down to is that if we can't find the information we need, it doesn't matter if it exists or not. If we don't find it, we don't encounter it, then it isn't information. There are undoubtedly millions of bytes of files on the Net that for all practical purposes are non-existent.

My biggest fear in relation to the information highway is that intellectual organization and access will be provided by the commercial world as a value-added service. So the materials will exist, even at an affordable price, but it will cost real money to make use of the tools that will make it possible for you to find the information you need. If we don't provide these finding tools as part of the public resource, then we aren't providing the information to the public.

Dissemination

There's a lot of talk about the "electronic library". Actually, there's a lot written about the electronic library, and probably much of it ends up on paper. Most of us agree that for anything longer than a one-screen email message, we'd much rather read documents off a paper page than off a screen. While we can hope that screen technologies will eventually produce something that truly substitutes for paper, this isn't true today. So what happens with all of those electronic works that we're so eager to store and make available ? Do we reverse the industrial revolution and return printing of documents to a cottage industry taking place in homes, offices and libraries ?

Many people talk about their concerns for the "last mile" - for the delivery of information into every home. I'm concerned about the last yard. We can easily move information from one computer to another, but how do we get it from the computer to the human being in the proper format ? Not all information is suited to electronic use. Think of the auto repair manuals that you drag under the car and drip oil on. Think of children's books, with their drool-proof pages.

Even the Library of Congress has announced that they are undertaking a huge project to digitize 5 million items from their collection. Then what ? How do they think we are going to make use of those materials ?

There are times when I can only conclude that we have been gripped by some strange madness. I have fantasies of kidnapping the entire membership of the administration's IITF committees and tying them down in front of 14" screens with really bad flicker and forcing them to read the whole of Project Gutenberg's electronic copy of Moby Dick. Maybe then we'd get some concern about the last yard.

In conclusion :

- No amount of wiring will give us universal access.
- Just adding more files and computers to gopherspace, webspace and FTPspace will not give us better access.
- And commercial information systems can be expected to be... commercial.