Data Libraries as Vending Machines; Or, What We Can Learn From Arthur Dent

by Laura Guy [1, 2]
University of Wisconsin-Madison

"...technology causes trouble. As a major agent of change it intrinsically, not accidentally, dislocates and distresses established relationships and forces economic, political or social change[3]."

Abstract
Technological change has had a tremendous impact on how we do our jobs. It not only has affected how we organize and provide access to information, but how our users conduct their research. This change has created new challenges for our profession, not the least of which is wondering if it will make us obsolete by replacing us with knowledge-based systems. The nature of these changes is discussed and we fantasize a bit about the *data library vending machine*. Finally, we look at our users and how we might best continue to provide them with the services that they need.

The Vending Machine Analogy
When I was a little girl we would visit my grandfather where he worked in one of the state office buildings in St. Paul, Minnesota. In the basement of this building there was a little cafeteria--a lunch counter. It was staffed by one or two people, and they sold sandwiches, soup and drinks.

I'm sure that similar lunch counters existed in office buildings all around the country. But if you go to one of these buildings now you will likely see a bank of vending machines.

What happened? Although it may be overly simplistic, it appears that the people lost their jobs to technology. What were the reasons? Vending machines may be cheaper. They take up less space. They are on duty 24 hours a day, seven days a week. They may be more efficient. They don't require vacation time or a health care plan. They don't complain about work conditions.

In the last few years, we have seen vending machine technology advance. Some of them can talk. They can take $1 bills and even $5 bills and make change. They dispense hot soup and cold salads.

The questions that we are dealing with are as follows: Can systems be developed that provide users access to data without the need for data librarians? To what extent can what we do be replaced by vending machines--that is--data vending machines? What are the tasks that can or *should* be automated or eliminated by technology, and after that happens, will there be anything left for *us* to do?

Coping with Change
I doubt that there is one of us who isn't simply breathless at the speed of the technological change that we are experiencing. Today's leading edge technologies quickly become commonplace. It's likely that as professionals our primary task over the next decade will be coping with this change. Much of the transformation is evolutionary in nature rather than fundamentally discontinuous: the change builds on itself. The speed and breadth of the transformation we are experiencing creates interesting challenges as well as opportunities. As the old axiom goes: "God protect me from living in interesting times." These are indeed interesting and exiting times for us, brought in part on the following changes:

1. Change caused by technological advances in hardware:
Client/Server technology: the architecture of the computer systems we use is changing rapidly. Hardware is becoming more "personal" and "portable." Personal devices are connected to powerful servers that are part of a distributed information system.

Storage Media capacity: Multi-gigabyte local storage capacity is becoming commonplace. Advances in disk technology and compression rates facilitate the storage of vast amounts of information on-line and provide interactive access.

Network capacity: The development of Ethernet and Token Ring networks enhances connectivity and provides extremely reliable data throughput. Today's hardwired networks will become tomorrow's *wireless* networks, enabling users to connect to anywhere from anywhere.

Speed capacity: Over the last few decades we have seen the doubling of raw CPU power every 18-24 months. Today's 20 and 50 MIPS machines will become tomorrow's 1,000 MIPS machines.

Memory capacity: We now have 16-megabit memory chips available, and soon we will see 64-megabit chips. Future computers will be able to store hundreds of thousands of typed pages in a computer's main memory and millions of pages on local disk drives.

2. Breakthroughs in Software:
New software architecture: The hardware revolution has enabled the development of easier-to-use software. This software in turn encourages the creation of collaboration tools and work group environments. The interfaces are more powerful, and applications are moving from the stand alone model to intelligent workflow. We see multi-tasking and multi-media capabilities.

The virtual environment: Graphic techniques and animation are transforming the way we visualize information and complex computations. Artificial Intelligence modeling and virtual reality coupled with enhanced visualization capabilities allow users to explore and interact with a virtual environment. If you think this sounds a bit farfetched, watch one of the *America's Cup* programs on ESPN. Software developed by Silicon Graphics incorporates a variety of measures such as wind speed and direction, compass readings, boat location, speed and course information into an astounding real-time graphic display.

Document-based systems: The concept of "document" has become more complex with the advent of hyperlinked data, text, graphics, video, sound, and so forth. The tools that operate on these documents also have become more sophisticated. Information becomes document-based when documents exist separately from the applications that create and operate upon them.

Object-oriented systems: Suited to modeling complex problems and processes, these advanced systems have the ability to self-update and communicate output in a variety of manners (voice, visual, etc.). Software tools are more modular and applications more flexible and powerful.

3. We see a paradigm shift from the *data* processing model to the *information* processing model as described by Ronald Weissman of NeXT[4]. In the new model information becomes content enriched, existing in an environment of "creativity" and "ambiguity" (an example might be *data* in a spreadsheet as opposed to *information* on the World Wide Web; the first is static, one-dimensional, unambiguous and incomplex, the second is dynamic, multi-media, possibly ambiguous and capable of presenting complex subject matter). In more concrete terms, we see documents linked with abstract, index, and bibliographic information, and numeric data linked with meta-data. Such documents may be more accessible and, for the general public, perhaps more captivating.

4. We've seen changes in our users: scholarly research methods have evolved into what Michelson/Rothenburg call "network-mediated scholarship[5]. Scholarly communication and collaboration, as well as the broader research process, have undergone significant transformation. Many of the changes our users are experiencing are in large part technology driven.

5. We are experiencing changes in levels of connectivity. For those of us who are "wired" there is an enhanced ability to access, analyze, disseminate and communicate information instantaneously and without regard for distance.

6. There is an increasing amount of information published in electronic form (for example, the growth of government information) and a growing number of formats for electronic records and information such as e-mail, CD-ROM, magnetic tape, word processor publication, dial-up services, on-line services, G.I.S., spreadsheets, relational databases, floppy disks, and bulletin board systems.

7. Recently there has been a hypermedia revolution and accompanying it the concept of the nonlinear document (and the thought process behind it--which is not that new[6]!). We see multidimensional data that integrate diverse formats of information. Recall the *information* processing and the *data* processing models mentioned above. The new paradigm includes a growing complexity of systems, increasingly sophisticated applications, and a plethora of document types.

It will be essential for us as librarians and archivists to project and assess the importance of these changes. Clifford Lynch warns that in the 1980's "many research libraries...thought that users' needs for access to online database searching were substantially overstated[7]." By not meeting the challenge of a transforming environment we risk making our libraries irrelevant and ourselves obsolete. And while we must remain agile, at the same time we need to be wary of developing systems and new services that may be poorly matched to the needs of our users, poorly designed or poorly implemented.

Decentralization of Information Resources

In the last several decades we have seen the transformation of traditional, paper-based, largely manual information systems into automated electronic systems. The obvious example is the library card catalog, which was transformed in the early 1980's by the development of online public access catalogs (OPACs).

We have experienced a more recent evolutionary change into a distributed or decentralized information êenvironment. For example, by the late 1980's OPACs were beginning to be made available through the Internet. Later, to obviate the need to learn new user interfaces to access the various OPACs, the Z39.50 standard was êdeveloped. The standard is based on client/server technology and greatly facilitates network information access.

This evolution towards distributed resources closely follows the evolution of the Internet itself. The development of the early **ARPANET** in the late 1960's and early 1970's had a strong economic basis because of êthe great expense of computers; it enabled resource sharing. As technology became cheaper, the need to centralize (cost share) decreased.

In the 1980's the *financial* need for groups to come together to share computing resources decreased. The death of centralized mainframe computing followed closely the advent of minicomputers and micros. Increased networking capabilities allowed individuals to link to other individuals for reasons other than cost-sharing and without any regard for physical proximity. In the 1990's the dumb terminal attached to a mainframe computer has been replaced by the stand alone "scholar's workstation," which has powerful CPU, lots of disk space, a CD-ROM drive and an Internet connection.

A familiar manifestation of this decentralization has been the proliferation of information resources on the êInternet. The sharing of information has been greatly facilitated by the development of tools such as anonymous FTP, the Wide Area Information Server (WAIS), and the World Wide Web. Research projects that collect data such as the National Survey of Families and Households (NSFH) now have the technological ability to cheaply and easily become their own access providers. Individuals now have the capability to become resource centers. Bill Goeffe's *Resources for Economists on the Internet*[8] , is a good example of what one person can accomplish using technology no more sophisticated than what can fit on a desktop.

Information decentralization causes librarians quake in their shoes. And rightfully so: as the anarchistic nature of the Internet intrudes into information systems there is a recognized lack of standardization, centralized authority, and access control. There is no institutional control over individuals like Bill Goeffe, who may in the twinkling of an eye forsake his resource and allow it to lapse. Similarly with the NSFH, one might ask what happens to the project's data when the funding ends or the research group disbands?

In a decentralized or distributed environment *access* becomes independent of *location*. The physical location of a resource has little meaning. In fact, something that virtually appears as a single resource might physically exist in separate parts in disparate locations. From the user's viewpoint it is not really important *where* the resource lives[9]. In this new information model our old notions of control require reformation. Who is responsible for a resource that consists of multiple copies dispersed among multiple locations? New forms of control may be required for insuring the continuation of important information resources.

Data Libraries as Vending Machines

As librarians and archivists we exist to serve users and preserve information. If we are to be replaced by vending machines it will be because they serve users and preserve information better than we do. They would be *cheaper*, replacing staff and facilities with computer hardware and software; they would be *easier to use*, enhancing the scholar's workstation and providing access to needed information from the desktop; they would be *faster*, allowing for the instantaneous access to any information at any time; they would be *decentralized*, maximizing connectivity through resource sharing; and, they will be *wihout* boundaries, providing access to users no matter where they may be.

Most data users go through a process that consists of four separate steps:

1. identification of potential data sources

2. determination of the usefulness of the data

3. obtaining access to the data

4. obtaining analyses from the data

Any data vending machine environment would have to play a role in each of these steps. It would have to assist users possessing varying levels of experience as they make their way through a research sequence that is not necessarily linear; for example, it may not be until the user reaches step 4 that they realize that the data do not êmeet their needs. The system would have to deal with the lack of standardized terminology to describe data and the lack of standards for formatting and storing data.

The data library vending machine would have to be *smart*. It would need to be capable of a multitude of sophisticated activities such as conducting reference interviews, answering questions, conducting searches, and assisting with problems. It would need to be capable of serving a diverse group of users that spans a wide range of abilities, including computer knowledge, communication skills, and research expertise. It would need to be able to instruct as well as assist, and preserve as well as make accessible.

Who Are Our Users and What Do They Want?
Its an obvious assertion that we need to know who our users are and how we can best meet their information needs. To be sure, they are a diverse group, but in general we can talk about them in terms of what they are usually involved in: the research process. This process can be divided into five parts:

1. identification of sources

2. communication with colleagues

3. interpretation and analysis of data

4. dissemination of research findings

5. curriculum development/instruction of next generation

The research process has undergone a tremendous amount of transformation. These changes in turn have an êimpact on how we, and others associated with the research process, do our jobs. The research paradigm includes a diverse group of agents: researchers, publishers, computer specialists, vendors, librarians, archivists, and professional organizations. Information technology has influenced changes in all of them.

In the last several decades end-user computing has become more convenient, cheaper, faster, more powerful, and êeasier to use with sophisticated interfaces and the advent of interactive environments. Technology has provided increased connectivity that enhances the research process through the expansion of access and the facilitation of communication.

As librarians we are more commonly serving a "remote clientele." Users frequently do not need or wish to "come through the door" to receive assistance. This change, although gradual, impacts directly on the way we provide êour services. Much of the communication is electronic in nature. The need for remote access to documentation as êwell as data is a part of our users' growing demands and expectations for timely and adequate service.

We've seen a transformation in the capabilities of our users. Some are very technologically sophisticated and have access to the most powerful computing resources. These people, typically faculty in my environment, prefer to have minimal interaction with the library. They want *what* they want *when* they want it, and prefer to require little human assistance.

I can imagine this group working well in a "data library vending machine" environment of the type I believe is practical within the constraints of today's technologies: an environment that includes FTP or WWW accessible data and meta-data, intelligent searching, extraction, and analysis capabilities.

Others users, often but not always students, have varying degrees of computer sophistication and varying access to resources. Many have very little experience with FTP or the World Wide Web. They do not have the computer resources required to, for example, download large files off the Internet. Many lack a basic understanding of the nature of numeric data.

I have difficulty thinking of this second group as a potential vending machine clientele. The front end that would be needed to teach these people what they need to know in order to use any given data set is beyond my imagination.

We are tied inextricably to the information needs of our constituency and must always monitor these needs êcarefully. It is critical that we work with our users and the other agents involved in meeting their needs (for example information providers and computer scientists) to continue to push forward information management.

The "Expert Systems" Fear

Expert systems, sometimes called "knowledge-based systems" are a type of computer program that uses the knowledge-based techniques of Artificial Intelligence. Simon Hayward[10] described them as computer programs that represent knowledge and apply expertise to manipulate that knowledge and to achieve solutions. Over the last decade or more the fear of being replaced by expert systems has been sounded in many places. By-in-large, this has not come to happen. Certainly we will see, coming out of Artificial Intelligence, the development of intelligent agent technology that will provide aids for locating, evaluating, analyzing and interpreting information. But, for data librarians the fear of being replaced by sophisticated expert systems that interface data and meta-data with users is a concern that may, for the foreseeable future, be unwarranted. The omniscient, omnipotent and omnipresent data library vending machine described above will probably not exist for a long time.

Nevertheless, we must consider: To what extent can what we do be replaced by expert systems? Alternatively, what is it that we do would we *like* to see eliminated by technology? What might we do with the extra time that will facilitate our users and enhance our role in the research process? And, can this process be viewed positively rather than as a threat?

An Informal Users' Survey:

During the months of March and April, 1995, I conducted an informal survey of users of the DPLS. I kept track of the complexity of their needs, their level of sophistication, and their access to resources. The question I wanted to answer was the following: which of these users would operate well in a data vending machine environment?

I found that 30% of our users would be well-served by the vending machine data library. They are experienced and familiar enough with numeric data that they need very little help to use it; these people just need to know where the data are. As more data are made available publicly on the Internet (for example **LABSTAT**, **the PSID, and the Penn Tables**) we will have less frequent contact with these users.

The other 70% are the users who don't know what data sets are available or which data they want to use. They don't have the experience working with data to understand what they need to do to access it or how statistical software works. There are many who simply like to come in and talk about their ideas and appreciate whatever type of feedback they might get. Some users who have little experience using computers are scared and need comforting and hand-holding. Some of our users prefer to look at paper-based resources. There are those who don't have a clue as to how to do secondary analysis.

The continued growth of CD-ROM publishing, and the development of increasingly sophisticated interfaces will change the above percentages. For example, compare an Internet-based interface that merely provides a raw data set (like **FTP**) with an interface that allows people to do **SAS** or **SPSS** runs interactively in real-time on a remotely stored data set (such as can be built today on the World Wide Web). Or compare the latter with an on-line data center system that provides all meta-data associated with a multitude of data sets, including variable-level information, instruments, methodology discussions, bibliographies, and so on.

The percentages will also be changed by the users themselves. Over the years we've seen an increased recognition among scholars of the importance of quantitative analysis skills, and a spreading of this recognition to less traditional fields such as history and education. But as the shear numbers of users doing quantitative research increases, they are coming to us more technologically sophisticated because they are exposed to computers at an ever earlier age.

The new technologies described above promise to eliminate some of the more tedious parts of our jobs such as tape

rollovers and extracting data for users. This will free us up to deal with important issues such as proper levels of documentation, the structure of the information systems providing data and meta-data, and bibliographic, abstracting and indexing problems. It allows us to devote more time to assisting users, teaching, and developing standards and policies. Its likely that in the process we will redefine what it is that we do and what it means to be a "librarian."

We need to ask ourselves what does the word "library" mean? is a library just a physical place or might it become more? An environment? A mind-set? A virtual world? Or, could it become less? A computer attached to the Internet? We must also ask ourselves as user demands and expectations grow, can we provide?

So What Does Arthur Dent Have To Do With All This?
When I first started to think about this panel I remembered something that I had read:

> *"After a fairly shaky start to the day, Arthur's mind was beginning to reassemble itself from the shell-shocked fragments the previous day had left him with. He had found a Nutri-Matic machine which had provided him with a plastic cup filled with a liquid that was almost, but not quite, entirely unlike tea. The way it functioned was very interesting. When the Drink button was pressed it made an instant but highly detailed examination of the subject's taste buds, a spectroscopic analysis of the subject's metabolism and then sent tiny experimental signals down the neural pathways to the taste centers of the subject's brain to see what was likely to go down well. However, no one knew quite why it did this because it invariably delivered a cupful of liquid that was almost, but not quite, entirely unlike tea."* **The Hitchhiker's Guide to the Galaxy**[11].

There is a real danger that the vending machine data library will almost, but not quite, entirely be inadequate.

The Human Factor
In the vending machine data library might the human factor be missed? it is important to keep in mind that in electronic systems the human factor plays a very important role. In the words of one science fiction author, it is the human factor which gives these systems their "heart." Thus, it will be humans who must deal with pertinent issues such as standards development, information integrity, accountability, and responsibility; it will be humans who realize the importance of meta-data as an essential supplement to standard bibliographic approaches; it will be humans who design the systems, who implement policies and develop the tools and criteria by which the systems will operate; it will be humans who will develop new descriptive systems, finding aids, navigational aids êand informational hooks that are suited to the constantly changing electronic environment and user demands.

There are equations where the human factor hasn't been missed, for example, in bowling alleys there once were "pin boys." It is a cold hard fact that if an entity can be adequately replaced by technology under the current êsystem it probably doesn't deserve to survive. Interestingly, at the University of Wisconsin, human-staffed delicatessens are proliferating. Why? The profit motive? The need for Human Interaction? Could it be that we still do need to have that perfect cup of tea?

Conclusion
We live in a time when concepts like "unique" and "multiple" are becoming obscure, as are "library" and "archive." Our survival and transformation as a profession depends on how we respond to the changes we are experiencing and the new paradigm in which we find ourselves. It is important that these changes not be perceived as threats but as opportunities, and that we work to turn what might be weaknesses into strengths. As the information infrastructure becomes stable and established, focus will shift to areas that are open to contributions that we can play an important role in making. For example:

Meta-data engineering for better methods and tools for describing information.

Development of new descriptive systems and finding aids.

Development of access tools to facilitate navigation and information retrieval.

Development of improved user interfaces.

Development of new governance and control mechanisms over information.

Standards for document and data management dealing with diverse areas such as scanning, text encoding, and storage, and retrieval issues.

Insuring the continuation and widening of the information infrastructure.

Teaching colleagues and users.

Copyright, intellectual property,privacy, and public use issues.

Promotion of the archival mandate and the protection and preservation of electronic information.

We must temper "carpe diem" in an environment of decreasing funds, private sector competition, and the need for us to develop advanced skills and expertise. In light of the changes we are experiencing, it is clear that there are many challenges ahead if we are to remain a viable and useful profession. These challenges will require êinnovation, agility and deep understanding of our environment and the needs of our users. Defining these challenges may help us to continue and grow as a profession; meeting them will definitely lead us to live in interesting times.

1 Paper presented at IASSIST 21st Annual Conference May 9-12, 1995, Quebec City, Canada.

2. I would like to thank Derek Zahn for his substantial contributions to this paper.

3. Coates, Joesph F. "Science, Technology, and Human Rights," Technological Forecasting and Social Change 40 (1991): 389-391.

4. Weissman, Ronald E. "Archives and the New Information

Architecture of the Late 1990's," American Archivist 57 (Winter 1994): 20-45.

5. Michelson, Avra and Jeff Rothenberg. "Scholarly Communication and Information Technology: Exploring the Impact of Changes in the Research Process on Archives, " American Archivist 55 (Spring 1992): 236-315.

6. Bush, Vannevar. "As We May Think," Atlantic Monthly 176 (July 1945): 101-108.

7. Lynch, Clifford. "Archiving the Promise: A Proposed Startegic Agenda for Libraries and Networked Information Resources in the 1990's." In Netwoks for Networkers II, edited by Barbara Evans Markuson, 52-84. New York: Neal-Schuman, 1983.

8. Goeffe, Bill. "Resources for Economists on the Internet." Version 8.02 <URL:http://www.shusu.edu/ftp/economics/EconData/.www/EconFAQ/EconFAQ.html>

9. That is, until it disappears.

10. Hayward, Simon A. "Is a Decision Tree an Expert System?" In Proceedings of the British Computer Society Specialist Group on Expert Systems 4th Conference, 185-192. Cambridge, England: Cambridge University Press, 1985.

11. Adams, Douglas. "The Hitchhiker's Guide to the Galaxy." New York: Harmony Books, 1980