

LA-UR- 02 - 1683

Approved for public release;
distribution is unlimited.

c. 1

Title: A QUANTUM LEAP:
INNOVATION IN THE EVOLVING DIGITAL LIBRARY

Author(s): Richard E. Luce

Submitted to: The Samuel Lazerow Memorial Lecture
San Jose State University
San Jose, California
October 18, 2001



Los Alamos

NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

The Samuel Lazerow Memorial Lecture
San Jose State University, October 18, 2001

“A Quantum Leap: Innovation in the Evolving Digital Library”

Rick Luce

Research Library Director, Los Alamos National Laboratory

It is an honor to give the Lazerow lecture tonight and to discuss digital library developments from the perspective of working at a national laboratory. Tonight I would like to consider what lies ahead given the evolution in scientific research, how that impacts the development of digital libraries, and finally, look at some of the challenges ahead of us. I'm particularly interested in giving this talk tonight because it provides an opportunity to talk to those of you who are students. You represent the next generation of professionals who will to confront some of the challenges I will outline tonight, as well as those of you who are the mentors and teachers of the next generation. The two roles are pivotal in terms of the challenges on the horizon.

Most of you are familiar with the information literacy challenges we face as a nation. As the library director of a national laboratory's science library, I am also acutely aware that we also have a real problem with the lack of scientific literacy within the general population in this country and it has a corresponding impact on decision-making in a technological society. Those of us engaged in supporting scientific research, or just generally interested, should be concerned about this fact because science and technology are at the foundation of our success as a nation in the 20th Century. For our nation to continue to be successful in the 21st Century, we will need to improve on the state of scientific literacy.

LANL Background - Los Alamos National Laboratory is one of the largest multi-disciplinary laboratories in the nation. It is spread out over 43 square miles and employees about 12,000 people. It is where Robert Oppenheimer brought a team of scientists together in World War II and gave birth to nuclear physics. Today the laboratory is one of the world's premiere sites for high performance computing and the birthplace of efforts such as the human genome project. The laboratory is currently constructing a building large enough to house three football fields, to support the newest generation of super computers. Supporting world-class science in such an environment comes with some very high and demanding expectations related to associated information needs and library services. Tonight I will use the Research Library at Los Alamos as a case study for innovation in the digital library.

Research Library Background - The Research Library can be characterized as a hybrid library. It has as a special library's focus on highly developed services and an academic library's orientation on deep collections. We have a staff of ~52 FTE's, which has been roughly constant over the last decade. When I arrived in Los Alamos in mid 1991, the Library had perhaps a half dozen PC's, no local area network, and a library system unconnected to the Laboratory's network. It was quite surprising to arrive at a national laboratory and find that the library was so far behind the curve in terms of

The Samuel Lazerow Memorial Lecture
San Jose State University, October 18, 2001

technology. About 1/3 of the staff were professional librarians and 2/3 of the staff were paraprofessionals, as they were categorized at that time. Today, we are a hybrid organization, composed of a combination of library/information science professionals, a business component, and people with very strong technical skills in IT. I believe this type of a hybrid organization is where future digital libraries will need to move.

The Research Environment Today - How can we characterize the environment in science today? The following quote from The Chronicle describes a fundamental shift in the nature of scientific research:

Instead of spending six months doing an experiment which you can then understand in an afternoon when you're done, you can do an experiment in an afternoon and it takes you six months to figure out what you've got.¹

The vast quantities of information available to the researcher also present new challenges. The requirement to mine the data in vast repositories is summarized as follows:

People of my generation grew up programming computers; I imagine the next generation of graduate students will have to become data-mining experts because that's the technology they'll be using.²

Those quotes represent a real shift, which is also a harbinger of a shift in user requirements from digital library resources. The traditional way of doing science is changing. Scientists are looking at so much data sometimes they don't quite know what it is they are looking for. So, we will need new tools in order to really exploit the information that comes out of this experimental process if we are to continue to meet their needs.

What implication does this have for digital libraries? When you talk to scientists, they often speak about a race against time. It's a particularly acute issue today. Think about people working on new forms of anthrax detection. They're racing against time with lives at stake. People working in the biomedical field are racing against disease. We are living in an environment where we need to anticipate and supply research information quickly. The PI's grant is going to end, or they are trying to solve a significant problem and they are focused on how they can do that quickly. It follows that the library's focus then, in a supporting role, is concerned with how can we efficiently and effectively support that process. Accordingly, efficiency in our process is a big driver in our environment, and I think it ought to be a driver in every library environment.

We also see a shift in the value equation in supporting scientific research today, and I think beyond science as well. The value used to be exclusively the content – if you could just get the user full text delivered to the desktop, he would be happy. That is no longer sufficient because the value equation today is no longer just the full text of whatever the journal article or book. The real value is in coupling that material with the relationships related to the material, dialog about the article, citations backwards and

The Samuel Lazerow Memorial Lecture
San Jose State University, October 18, 2001

forwards, experimental results, the raw data that one can, as a colleague, take apart and look at also. That larger context now is becoming just as important as the article itself.

Collaboration is also a clear driver in science and it has implications for how our systems need to work. Our laboratory is known for being able to very quickly set up project teams that work with people across the globe. These teams come together very quickly and often they are disbanded very quickly as the members move on to other projects. A good example of this was the outbreak of hantavirus in New Mexico. Experts from a variety of agencies came together to puzzle through what was going on. Why were people dying in rural New Mexico? What's going on? So that was a case of a team coming together very quickly, having to solve a problem very quickly and then going back and doing other things afterwards. In that kind of an environment the ability to collaborate and the tool enabling effective collaboration are absolutely crucial. One can't take three or four days when you're fighting a disease that can kill you in a matter of hours. You can't afford lost time trying to figure out how to do something more sophisticated than just exchange e-mail. So we need to configure environments that allow people to quickly come together, plug into that environment, and then move where they need to move.

Stages of Digital Library Development - I'm going to propose a model of evolutionary stages of digital library development, which applies in our environment. The first phase is a fundamental and strategic question about make or buy. Do you make or buy systems and content? What is it that the digital library you're building really aspires to do? At Los Alamos, due to the nature of our work, it would not be prudent to have external organizations or people analyzing the research patterns in the Laboratory. So we are very committed to bringing digital content into the laboratory, making it available on Laboratory networks and avoiding sites where that information could be mined, harvested, and sold. We can see the same parallel thinking in the pharmaceutical industries, and for good reasons. We're beginning to see it more in other industries as well. Thus, our perspective is to protect the usage data related to the work of our research community, although many publishers have a hard time understanding this. We bring the information inside where confidentiality can be assured. The statistical information that is collected by the Research Library is used it to understand better what the needs of our own organization are.

If you begin with the premise that you are going to collect digital content, both that created internally, as well as things created outside that you bring in to study, analyze and use, the next stage is to begin to aggregate it together. Much like the physical collection, i.e., where one needs to aggregate things in a building, in the digital environment we need to aggregate a variety of content together as well. This includes primary and secondary content in a variety of different forms. As the digital content is aggregated, it must be put together in an integrated system. That integration then leads to a driver for applications that interoperate with one another. If I'm looking at a chemical encyclopedia, I would like to be able to use it integrated with chemistry articles that I am reading. Acquiring and locally storing electronic provides the ability to control and manipulate content within a framework for the end-user.

The Samuel Lazerow Memorial Lecture
San Jose State University, October 18, 2001

We also need to create such mechanisms for our technical reports, which we have the responsibility for in the Laboratory to tie together not just the reports that are digitized, but also links to the raw data. Rather than looking at a table, let's go in and look at the raw data and experiment with it or change it. It becomes an interesting problem in terms of then tracking what it is that we're really trying to keep track of, because it's mutating all the time in front of us.

The third stage that follows is the concept of linking. How do we create dynamic relationships between the kind of data that we're looking at and managing, and data in the outside world that perhaps we can't bring in, or it's not desirable to house locally? We want to seamlessly characterize the relationships between our documents.

After achieving progress with those stages, you come to the point where we have the base to begin to do something that stretches beyond what we were able to do in the paper paradigm. The early stages are a duplication of what we've had in paper for centuries. Perhaps it is faster and more efficient and perhaps not as pleasant to read on the screen as opposed to paper. But once you have those pieces in place, we are then able to move to a position where you begin to add new value in terms of services, electronic publishing, supporting publishing from members of the community, personalized services that begin to suggest things of interest to the user based on patterns of usage and new digital tools.

I'm going to briefly demonstrate some of the applications that we use today, as a place marker. We have a product called *FlashPoint*, for cross-database searching, which followed our earlier developments of individual web databases, which were locally loaded and accessed, via a web interface. Today, we manage about 56 million metadata records. Much like the general development of the web, our users went from being excited about access to the realization that there's too much out there and one may not know which database to use. So, now we support parallel searching across our discreet databases with one command. Search results are returned and identified by database, allowing a user the option of linking directly into the appropriate results in the desired, or linking to the full text. This maximizes the first phase, integrating our resources together. We have nearly a dozen different local web databases and what one wants to do is end up with an environment where the user doesn't have to understand which database is needed, rather here is where I can launch a search and this is where I get the results. As a matter of fact, that closely parallels the behavior from users on the Web today.

MyLibrary @LANL – is our highly customized and personal view of information resources and links. We decided that we would skip over the concept of creating a portal to the information that we have. Because the laboratory's research is so broad, we couldn't find a way to effectively bring everything together in one environment that was not overwhelming to the user. We desired to use the idea of personalized capabilities much like you see in My Yahoo. What I'm demonstrating here is a personalized user interface, and this one happens to be mine. Within *MyLibrary@ LANL*

The Samuel Lazerow Memorial Lecture
San Jose State University, October 18, 2001

I'm able to create services based on what I'm interested in, both with our Research Library and to links externally anywhere on the web.

So, the definition of this library is no longer what collections and authorized links we have digitally or physically. It's really what one would want to connect to, or derive information from, and seamlessly integrating both our local resources and external resources together. For example, I'm reading something on the Web and I happen to see this article. I decide I want to save it and read it later by activating a *bookmarklet* on my toolbar, which I've been able, from this application, to just drag up and drop on the toolbar. I am asked where I want to file this and what name it should be given. I've created these categories, thus I select a category, and that's all that is required in terms of saving the link.

Recently we expanded the "*MyLibrary*" environment by adding collaborative capabilities so it can be used to share resources with others within the Laboratory. We allow teams to share resources or selected tasks within this environment, and still keep other tasks private and personal. So, in this case, I'm adding a library tab and on the third line, there's a "shared library" for read-only users. I can define who those shared users are, as well as shared read/write users. This is the kind of thing that we use in a team environment, where I've got things as a member of the team and I want to put it in a space that the other team members can take advantage of, and they can do likewise. As the user creates new tabs, we are trying to understand where this user is coming from in terms of discipline, and in terms of the kinds of connections they make. Having that information, are we able to make suggestions things that would be of interest?

Creating a new tab will automatically create a list of relevant links to journal titles and databases in that subject domain, e.g., physics or grant resources. We are working on methods to make these suggestions adaptive, so their relative weight begins to rise or fall based on community usage. The idea is to save user time populating new links, as well as exposing the user to associations that might not have been known.

Active Recommendations - I mentioned the idea of personalization, which we're using with *MyLibrary* and taking that further into the area of active recommendations. An active system will make recommendations to a user, based on the user's behavior, and also the collective behavior of the community with which he or she has frequent interactions. We want to push recommendations to the user, while allowing the user to either adapt the recommendations to better target his interest or ignore the recommendations. What are some of the challenges involved in this work?

First of all, we need to know who the individual is, which requires an authentication system to establish a unique user identity. We do this today with crypto cards, or smart cards. If the card is stolen, it will generate numbers for anybody who picks it up but if you don't enter the correct sequence, it doesn't generate the correct code. A user may wish to remain anonymous but then only generic services are available.

The Samuel Lazerow Memorial Lecture
San Jose State University, October 18, 2001

Secondly, to create customized services, knowledge of the user's personal preferences and behavior is required. From that behavior we would like to derive communities of interest, which we can use that to start to make connections, suggestions, and determine patterns, e.g., where is the research moving and where do we need to move to respond to that.

Coupled with user information, we characterize the information resources available, and for that we need a substantial understanding of the corpus of digital resources available. The problem can be broken down into the following components: (1) the keyword structure; (2) calculations of the semantic proximity of terms; (3) citation structure; and (4) user traversal paths which analyzes the linking path(s) a user follows during a session. With a large repository, the algorithms required to calculate these relationships take significant computational processing capabilities, so that is done off-line.

I am describing four different ways of trying to organize and map the relationships between things. We don't believe any one of these four is best, rather overlaying all four together leads to a far more sophisticated and on-target system. Out of that, recommendations are then calculated and presented to the user. Of course there should be a feedback loop, and we analyze what happens to the recommendations sent to the user – are they actually activated and used or not? That would show up in the traversal path when it is analyzed. Finally, the collective user patterns with the recommendations are analyzed and fed back into the system, which causes the system to adapt itself over time. We call this our Active Recommendation System and it is a challenge to scale this up to a large and dynamic environment.

Visualization - We are also seeing a growing need to begin to create visual spaces to look at this mass of data that comprise our digital collections. Today we have roughly six terabytes of information on our local library servers, and that is growing at roughly a terabyte per year. With 56 million citation records in the physical sciences, we are reaching the point where text systems start to bog down, in terms of trying to look at data.

Visual tools are relevant in part because we are visual creatures. If you look at brain research related to how the brain functions, what parts of the brain do what kinds of things, somewhere between 60 to 64 percent of the brain is related in some way to visual functions. That's how humans take in what's going on in the environment around us. It's a very, very quick capability that allows us to scan things, make instant judgments and move on. We have the same kind of problem when we're looking at a mass of experimental data, which is now represented by text. We try to represent the bits in different ways, and so we're looking at how we can use some of the visual functions, in terms of the kinds of things that we'd like to look at, to try to provide insight to our knowledge in masses of data.

Organizational Structure – That is a very quick snapshot of a few of the things that we're working on in our digital library today. How do you structure, organize and manage an organization to accomplish this type of work? We need to create an

The Samuel Lazerow Memorial Lecture
San Jose State University, October 18, 2001

organization able to innovate and rapidly readapt itself to be able to develop and support new digital library capabilities as they evolve over time There are three things that we focus the organization on: customers, operational excellence, and digital library product innovation.

I don't know of a library that will not claim they are customer responsive and most likely individuals in the respective institutions are. But for an organization to be customer focused, a systematic process is required. We use a variety of methods to obtain customer feedback, which is one dimension of a customer-focused organization. We survey our users quarterly, pulsing one fourth of our laboratory user population quarterly for the past 27 quarters. The surveys are coupled with a system for capturing unsolicited feedback. That unsolicited feedback goes into a database that's categorized, and then sent out to appropriate parts of the organization to resolve or follow up on. A Voice of the Customer process is used to obtain feedback regarding future needs, which is used as input into our software development activities. We take new products through extensive testing and focus groups. And, finally, an outreach process is used to both communicate what we are doing as well as listening to our customers. It is essential that our librarians get out from behind the desk and into the field, so that there's a feedback loop going on all the time.

These formal listening strategies are run in a coordinated fashion. They feed a set of user satisfaction metrics and process activities that various teams in the library have responsibility for. The output of all this activity, both customer data and our internal processes supporting our products are fed into a quarterly assessment process. These sets of activities are run like a business. A for-profit business is typically focused on quarterly financial results and we are focused on the quarterly customer satisfaction metrics and process behavior metrics related to the systems that we use. Thus this focus is very much like a business, with a slightly different output focus, and we do this rigorously on a quarterly basis. The quarterly assessment feeds a quarterly review of our business plan and results in a quarterly update to that plan. Rather than having an annual business plan, which typically is filed away on a shelf, we continually update this quarterly business plan, which is reviewed biweekly, and often changed about once a month or so.

Over the past eight to ten years I've talked to a large number of libraries about measuring customer satisfaction. Far less than one-half of the organizations I have talked with informally have some sort of general happiness metric which is collected on a regular bases, which I'll call level one. How do you feel about the library? Are you satisfied or not satisfied with the library? That's what I call a happiness meter and typically it is only collected every year or two at best. But what does it tell us when XX percent of some population is happy – happy relative to what? Happy because their expectations are quite modest or they have no other place to go? So, it's *relatively* happy and often without any gradations. So we're not that interested in happiness meters, because we don't really think it tells us anything that we can productively use to improve ourselves.

The Samuel Lazerow Memorial Lecture
San Jose State University, October 18, 2001

The next level of sophistication on an evolutionary curve of customer satisfaction incorporates a ranking of what's important to the end user or customer. This is coupled with an assessment of the satisfaction level related to what is important? To the extent that an organization begins to understand the tradeoffs here, this knowledge allows the organization to become more focused in terms of prioritizing services and level of effort.

The third level of sophistication, (and it is quite unusual to find a library at this stage), incorporates metrics which provide comparative data related to how the organization rate against the best in industry. Now we have a much better idea of what an 88% highly satisfied with some service really means. It is apparent that this requires a great deal of work and a repeatable to manage this systematically. The notion here is to figure out who is the best in the industry at whatever you are trying to do, study what they do, emulate what they do, and use that information to make your organization better. When you formally start to track that and measure how you're doing over time, then you're getting fairly sophisticated in terms of the customer focus.

I am perplexed at the number of library managers who run organizations that can't answer the following questions. *Who are your key customers?* (and sometimes they can answer this one)? Well, I guess it's the folks who come in the door or maybe the mayor. *What are their top three or four needs?* Response: I don't know, I've never asked them. *How well are you doing at meeting those needs?* Response: Well, I don't know, we have twenty different services. *How do you prioritize them?* Response: We don't, we don't have any basis to do that.

I've worked in every type of library except for school libraries. It seems that all libraries, regardless of size or type, have a phenomenal capacity to take on new things but never can let go of low priority things. This results in an appetite for a big smorgasbord of activity but no focus on what we are really good at, or should be good at. I think that is problematic when you have to make difficult decisions.

So, what's the bottom line? I've tried to very generally describe how we've got a formalized process for using customer feedback and customer satisfaction data. We have other processes related to how we run our organization. What are some of the keys for success in the digital library environment? I strongly believe that if you don't measure customer needs, you don't know how you are doing. You can say the water looks hot or it might even be boiling. Wouldn't it make sense to test it out before you stick your whole body in?. The same goes for the customer needs. Before we go off and spend significant dollars, in terms of expensive people developing applications, we'd better be sure we are filling a collective user need and be on the mark with it. We better be sure we're doing something that's not available out in the industry that we could have bought for perhaps a fraction of the cost. We evaluate what we do from the customer perspective, not from our perspective. So, it doesn't matter if we're the first, or if we think it's the best, it matters what our customers think.

Has that approached worked? We began digital library development in 1994 with our first web database. Our scientists had collaborators come in from some of the other

The Samuel Lazerow Memorial Lecture
San Jose State University, October 18, 2001

laboratories, they looked over what we were doing and they went back and said to their librarians, can't we get what they're doing in Los Alamos? That's grown today to a business where we support 29 prominent scientific organizations around the US, such as national laboratories and universities. This makes it hard for many outside the organization to categorize just what kind of an organization we are. Business process efficiency and entrepreneurial expertise is sorely lacking in most libraries. It is absolutely vital, in my opinion, in this new era that we're in.

Lessons – Successful digital libraries are certainly not just about the application of new technologies. It also requires sustaining organizational change. The composition of our organization today is roughly evenly divided between IT staff, librarians supporting both digital and paper collections, and a segment with various business skills. It takes awhile to create enough momentum in the direction you're trying to move to get something flowing. A leader must have a compelling vision and strategy. A leader also must be obsessively focused on the customer. It doesn't do any good to have a vision about where all the troops want to go. It's got to be where you need to go, in terms of delivering something to the customer. You've got to be very, very urgent about pushing change and implementing change. Complacency really kills things. You can think about external change caused by at least three factors: a crisis, a shift in the market, or a technological development. We are living at this interesting point in time where we have all three operating.

I'm going to end up with a shopping list of type of people that we try to attract to the Research Library. I look for people with a passion for service excellence, people who have initiative, enthusiasm, people skills, collaboration skills, technically facile, and lifelong learners. They act as self-empowered agents, they manage their own morale, and they're really here to make a difference. You work in libraries; you work with information because you really believe it will make a difference, all the way from the school kid who's trying to learn to read up to the scientist who's trying to solve a critical problem. So, we all have a role somewhere in that process and those roles are important, that activity is extremely important. You've got to get out on the edge in order to be successful. Coupled with all that, you need to be able to thrive in chaos.

Thanks for the opportunity to talk with you this evening.

Question from audience: So what are the 3 things that library schools should be doing?

Generally, the U.S. educational system here is noted for turning out creative students. So, library schools here ought to be really focused on how we stimulate the creativity process in our graduates, so that when they walk into the workplace they're able to think very creatively, number one.

Secondly, a fundamental requirement today is to facile with technology. And the problem with that is that it's always a moving target, and so to be facile becomes more somebody who is a continual learner. Very small children are incredible learning machines. Those of you who have had children or who can remember young brothers and sisters when they're this size and they're learning how to speak and they're learning how

The Samuel Lazerow Memorial Lecture
San Jose State University, October 18, 2001

to walk, what do they do? They make mistakes, they fall over, they get up, and they absorb. They try and fail and they continue to try. They learn incredibly rapidly. And as we get older, we do things that are very, very safe. So when it comes to technology we need people not with just the latest hot technology skill, because that will be history tomorrow, but we need people who can move through the generations as fast as the technology is moving. I don't know how to teach that. That's probably why I'm not an educator.

Thirdly, one of the things really missing in the library profession is people who are really able to collaborate, to work in a team, to function well as a team member, to succeed as a team member. I see this in the educational system with my kids, more and more activity related to group activities, project work. I ask my children "how does anyone know how much you're contributing"? And then I realize that's not the point. They're learning how to work in a group. They're learning how to get the group to succeed toward some objective. We need some application of that in our library environment as well. It can be difficult hiring staff and getting them up to speed quickly in our environment when they're expected to work as a member of a team and they've not had that experience before.

Comment from audience: I think I'm going to quote you because we have a lot of resistance from students who say – I don't want to work in a group because not everybody in that group is going to work as hard as I'm going to work. And therefore, I'm not going to do the work for them. And we've tried to convince them that working in a team is what they're going to find when they get out in the world. So you may be quoted heavily.

Comment from audience: I think it's pretty good coming up with those 3 things....

-
1. Chronicle: *Learning to Swim in the Rising Tide of Scientific Data* - 6/29/01
 2. Chronicle: *Learning to Swim in the Rising Tide of Scientific Data* - 6/29/01