February 1997

Technology and its Implications for Serials Acquisition

Clifford Lynch
University of California

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

Part of the [Library and Information Science Commons](#)

Recommended Citation

Lynch, Clifford (1997) "Technology and its Implications for Serials Acquisition," *Against the Grain*: Vol. 9: Iss. 1, Article 23.
DOI: <https://doi.org/10.7771/2380-176X.1996>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Technology and its Implications for Serials Acquisition

by **Clifford Lynch** (University of California, Office of the President,
300 Lakeside Drive, 8th floor, Oakland, CA 94612-3550)
<clifford.lynnh@ucop.edu>

Libraries have been overwhelmed with the financial dimensions of the serials crisis. They are trying to mobilize and apply technology to help solve the crisis, and recently have been giving serious consideration to the negotiation of large-scale agreements with publishers which rely on technology-based delivery systems and electronic content in essential ways as a means of introducing some new flexibility into the financial parameters. I am concerned that we may be making some unrealistic assumptions in trying to create a technical delivery system that will be acceptable to our users as part of these agreements. So I want to explore some facets of this dilemma as well.

In the last eighteen months we've seen the evolution from exploratory prototypes of electronic scholarly publishing to large corpora of electronic materials appearing packaged as actual products in the marketplace. The technical mechanisms underlying that development are very interesting. In the early days of electronic journal delivery prototypes, such as the Tulip Project, there seemed to be an implied notion that publishers would simply be the data providers and that the subscribing libraries or aggregating organizations like OCLC would be responsible for mounting and providing access to this content.

We learned a few things in the pilot projects. First of all we learned that it's pretty hard to mount this electronic material. I think all of the institutions who participated in projects like Tulip would agree it that took longer and required much more work than anybody

expected. I think organizations like OCLC would agree that they've had to dedicate huge resources into producing quality projects in the electronic journal arena. We've had problems with standards — basically every publisher's content seems to be different. We have not seen much support emerging for mounting content in most of the commercial integrated library systems; so you're mounting the content somewhere separate from the system that's running your catalog and managing your acquisitions. Basically, in the absence of extremely detailed and rigorously applied standards (and perhaps even with such standards, since the consensus nature of the standards-making process virtually ensures that any standards that are adopted will include a good deal of flexibility), we can't scale to a system in which each library can mount locally content from all of the publishers in which they're interested, even for serials.

What we're seeing instead is publisher Web sites. The major scientific and technical publishers are now becoming access providers and each of them is setting up a Web site. So the general content access arrangements that are under discussion today are about patron access to journals on such Web sites, rather than scenarios where libraries license and locally mount content for use by their patrons.

Let me define some terms and draw some distinctions between them. When we talk about electronic content today — and I mean primarily scholarly publishing with an emphasis on scientific and technical material, but not exclusively so — there are really two different concepts that are conflated in these discussions. One is delivering paper by electronic means: A product

that was conceptualized and designed for paper is put into digital form and made accessible through a network. Studies indicate that people are choosing what they do and don't read by viewing the digital form on screen. If they decide they have to read the material, they print it — they convert it back to paper locally on demand. These users may not file the paper; they may treat it as kind of a transient user interface by reading it and tossing the paper when they're done, especially if it wasn't a good article. But the issue is really about delivering paper at a distance, displacing storage costs, and making access more convenient because it can be printed on a distributed basis. It's interesting to note that almost no one is talking about offering books within this model. Although it's not implausible to print a fifteen page article on your desktop laser printer, you don't really want to print a two hundred page book, going through toner cartridges and reams of paper, only to have two hundred unbound pages floating around. This just reinforces the notion that we're really talking about decision-making online and printing to read relatively brief items and not entire books.

There's something else going on: a series of explorations, discipline by discipline and journal by journal, of what we can do if we release ourselves from the constraints of paper and design indigenously digital material that can include video clips

or simulation models or three-dimensional objects you can turn and view from different perspectives. Those concepts are still very much in the

How do the Web site administrators know that the users trying to access their material are in fact affiliated with an institution that has licensed rights to some or all of the Web site content?

experimental stage, and I think it's an open question in many fields how much this really contributes to improved scientific or scholarly communication. I think we all want to believe that it contributes a lot, but we're still learning how to construct and author those genres.

So my emphasis here is really on the big-scale system which is, today, still print-on-paper-oriented, and not on the relatively small number of indigenously digital works. Publications on Web sites today that are being offered commercially are basically arranged by publisher and scattered around. When we think of libraries making license agreements with those publishers that allow library patrons to make use of journals on these Web sites, it's important to remember the technical issues involved, because they're really problematic.

The first problem is authentication — site licenses which guarantee rights on behalf of an institution's users for access to materials on some Web site. How do the Web site administrators know that the users trying to access their material are in fact affiliated with an institution that has licensed rights to some or all of the Web site content?

The answers right now are basically unacceptable. They are user passwords or restriction by IP source addresses, basically whatever physical network to which these people happen to be connected. Given that many people do their work from home, given that many people are now using commercial Internet services to get access, and given that we are moving into mobile environments, IP address filtering simply isn't going to work.

continued on page 36

Passwords aren't going to work. Assigning individual passwords to the incoming students every fall for each of the publishers with whom you have a license and expecting them to remember these passwords is basically insane. It's unworkable.

We clearly need some really sophisticated authentication infrastructure. There are two basic sorts of proposals under discussion. There's the Kerberos system that's been around for a number of years. But Kerberos was really designed to manage access inside an organization, not to deal with interorganizational issues. There are some extensions that permit what's called cross-domain authentication, but that really wasn't its original focus. Kerberos doesn't deal much with inter-organizational trust models. The other approach is a digital certification hierarchy where you set up trusted machines that will vouch for people. For example, University X might give users digital certificates that are signed by that institution's certification machine. The publisher needs to know only that the licensing institution's certification machine is at some well-known address in order to validate the digital certificate. It's a very structured distributed trust model. I think that until we get those kinds of underpinnings in place, the authentication problem alone is a showstopper.

We need to think through printing, given that most people basically are going to be using this digital information to decide what to print. Printing items by bringing them down to the user's machine and out again as you would do with a Web browser is not necessarily the only process you're going to want. This is particularly true if you are dealing with some form of bit-mapped representation that is going to be painful over a slow line, and which will tie up large amounts of disk space in transit. You're going to want routing both from the workstation and directly from the publisher server to third party printers on the network. That is a more tractable technical problem, I believe, and there's been some good work done on that, such as the CUPID program.

A third problem is content consistency. There's a real problem with how we think about journals in digital form and Web sites. Here's a concrete scenario: You put up a Web site; you get some users in; you show them the journal; you say, "How do you like it now that it's electronic?" The users say, "Oh yeah, this is pretty good." But the content provider has only changed the format of a journal, and that's what the users are reacting to. The publisher hasn't faced the users with the idea that the whole system which they use to access the literature is changing in fundamental ways. This system of access defines how users locate content, navigate within that content, and use it. Different publishers will organize and present content differently. Different publishers may offer different ways of access to content.

So we have a large-scale system coherence problem. There's a lot of coherence and commonality in the print journal, in conventions like the table of contents and the idea of issues, in binding issues together and shelving material in a common fashion in libraries, in business models like subscriptions. All of this is in flux in the electronic environment. We've also addressed the coherence problem within disciplinary literatures in the print world through abstracting and indexing services. We still need to link records in abstracting and indexing services to the content that's out there on those publisher Web sites as this content shifts from printed to electronic formats. Who's going to do that — individual institutions or publishers?

Whoever takes on this role of keeping links up to date as sites move and publishers change and journals shift from one place to

another is going to be carrying a pretty hefty responsibility. Clearly we need better linking technology. One promising tool is the new serial item and contribution identifier standard that's about to be published, which essentially lets you produce keys computationally from an abstracting and indexing record or from content and then potentially match on those keys. You could potentially use a real-time, interactive, computational process rather than hard (precomputed) links that way.

There's another problem, though: Many of the sites that are being set up are not individual-object-addressable. Let me explain: This is a crucial point which seems to have eluded most publishers. As a user, you want to be able to point into an object (like an article) on one of these sites so that when you follow the link you go directly to a specific article — not so that you follow it first to a welcome page that gives you the news of the day, offers you discounts on some new journal, offers a trial issue, and only then presents you with a long list of journals from which to select and a choice of which year and issue. You need to be able to make links directly across sites down to the level of individual pieces of content. I believe that that is an absolutely critical design criterion as we start moving content to the Web in this environment. And obviously, until we get individual-object-addressability (or if we don't get it), making meaningful links in abstracting and indexing databases is going to be a problem. This is also an essential prerequisite for linking from a citation in one article to the text of the cited article, which is something that users really want.

Let's hypothesize that we achieve individual object linking in the abstracting and indexing databases. There is yet another interesting implementation issue here that we need to address. When we first started mounting abstracting and indexing data-

bases, many of us were a little concerned because these databases provided information about materials to which we had subscriptions as well

as to those to which we didn't. Some librarians (with the best of motives, I might add) thought it might be much better if they could produce subsets of such databases that only covered what the library actually possessed, or at least presented that content first and then provided users with the other material as an extra option. To inform users about material without thinking about how we're going to provide them access is a real disservice.

So now we're going to have these abstracting and indexing files with links in them. Do we show the users only the links to which we think they might have access because the library has a license in place? What about cases where the user has an individual subscription? Many professional societies now are offering access to some of the content on their Web sites as a member benefit; all you need is your member number and perhaps a password (yes, yet another stupid password!) to access the information.

What policies are we going to use for exhibiting these links? It would be immensely irritating to get a result from searching an abstracting and indexing database only to be stopped by messages from various publisher servers either indicating blocked access to the content due to licensing restrictions or requesting your subscriber password or credit card number before letting you click your way through the database to look at the articles. In the world of paper, one merely has to find a book on the shelf or request that it be borrowed from another library. Licensing to borrow and read paper information is a nonissue. We have a real problem between knowing content exists and actually accessing it in this licensed digital world. It is another issue that we're going to have to deal with. Libraries are going to have to think hard about the principles that will guide them in designing a

continued on page 37

Technology and its Implications
from page 36

solution to this problem.

Those issues of accessibility, linkage, individual item addressability, authentication, and printing must be resolved in the world of distributed Web sites, or we may see many users choosing paper over electronic content.

The promised convenience of being able to find articles and print them locally, im-

mediately, at any hour of the day, wherever you might be, is, in fact, not satisfactory given all of the present inherent inconveniences.

We need to look not only at the various financial and licensing ramifications of this technology, but to assure ourselves that the technological infrastructure is going to support useful access. We need to be sure that migrating to electronic content is going to create a world in which our readers want to live and read.

How many of you are familiar with the set of technologies that are variously described as secure containers, digital lockboxes, or cryptalopes? This technology wraps raw contents in some protective software, so that content providers don't have to provide (and lose control over) raw content that users can manipulate directly with their viewers or print directly on their printers. Instead, access to and use of the content is mediated through this protective software which defines what the user can do with it. It may not let you print it. It may only let you view it five times. It may give the user a 60-day license and just lock the user out after that time period.

This is both a troublesome and attractive technology for many reasons. It can perform many interesting functions. For example, it could enable an arrangement between a library and a publisher that dictates and guarantees that the enveloped digital content will be used by no more than two people at a time. This technology could, in a sense, emulate the old doctrine of first sale: A library can buy a book and circulate it, but can't copy it; so the publisher can price the book based on the impact of serialized use by one person at a time. Not only can this software wrapper tell the user how he or she can use the content locally (such as whether it can be printed locally or not); it can also interact with other functions on the Net. For example, this software can enforce a policy that lets you print the content as many times as you want, but will debit you ten cents for every page you print. It can call home.

There are also troubling, perhaps invasive functions: It can tell the content provider how often you're using this information. From a preservation point of view, I can't imagine a worse nightmare than some piece of software that tries to call home when I use it after it's been collecting dust for twenty years, and discovers the whole world has changed, nobody uses that protocol anymore, and the organization which published it no longer exists.

And how does this technology play with fair use? Fair use, let me remind you, is basically a defense. If a publisher accuses someone of infringement of copyright, an acceptable defense is that there was a fair use under its various definitions. But that type of discussion presupposes that you were able to make use of the content in the first place. If you can't make a copy, if you are technically prevented from clipping out a piece of an image in some multimedia work that's inside a secure container, there really isn't much question about fair use. So I believe there is a real issue about how this container technology interacts with our assumptions about fair use.

These kinds of secure containers are likely to show up wrapped around some of the content that comes both from libraries distributing content to their user communities and from publishers to libraries. Already, the technology fits nicely with some of the

things that are being done with the Web now. You can bring applets down along with content, for example, using JAVA; and MIME types are being registered, so that when you get one of these secure lock boxes it's handed off to the right helper application by your Web browser.

This new technology, while it gives us a lot more negotiating room for lower volume uses and other functions which may be

very appealing from a financial point of view, raises fundamental questions about user acceptance. Certainly, we can use

this technology to get away from current all or nothing financial arrangements. Today, material that you used to get only in print is also increasingly offered electronically. But it's often site licensed and very expensive, so you might opt to stay with the print, which is less expensive but serialized for only one user at a time. There are constraints either way, and the purchaser has to make a decision based on both financial and practical realities. I think one of the things we are going to see in the next two years are field trials of enough scope to get some sense as to when these options will and won't be acceptable for users.

The theme I want to underscore here is that we need to be very careful about whether we have technology that can deliver this electronic content for which we are busy negotiating financial arrangements in acceptable ways on a broad systemic basis. It's clear that we can do this for individual bits of content; the key question in my mind is whether we can construct a system that users will find comfortable in making broad use of electronic journals as a substitute for printed ones.

(This is a lightly edited transcript of my remarks at the Charleston Conference, Saturday, November 9, 1996. My thanks to Katina for the transcription.) — CL

SEARCH OUR 1.2 MILLION TITLE DATABASE,
CHECK INVENTORY, PLACE ORDERS, PROCESS
CLAIMS AND REVIEW YOUR ACCOUNT ON
BOOKLINK™

EASTERN BOOK COMPANY'S ON-LINE DATABASE

- STANDING ORDERS
- FIRM ORDERS
- RUSH SERVICE
- CUSTOM INVOICING
- ON-LINE INQUIRY & ORDER

131 Middle Street voice: 1-800-937-0331
P.O. Box 4540 bbs: 1-800-456-5201
Portland, Maine 04112 fax: 1-800-214-3895
e mail: info@ebc.com