June 2001

Digital Archiving: A Work in Progress

Ana Arias Terry

Informed Strategies, ana.terry@informedstrategies.com

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

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

Recommended Citation

Terry, Ana Arias (2001) "Digital Archiving: A Work in Progress," *Against the Grain*: Vol. 13: Iss. 3, Article 9.
DOI: <https://doi.org/10.7771/2380-176X.3489>

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.

Author Care: from page 26

at **Florida State University**, also an author for **MCB** journals, shares some ideas on publisher responsibilities to authors. In his assessment, publishers should be able to ensure that the author: receives knowledgeable and objective reviews of his intellectual property, is provided with timely input on reviews and publication, and is given professional editorial assistance whose suggestions are agreed to by the author. Adequate publicity of the author's work is also, he says, reasonable to expect.¹⁵

But the relationship is a two-way process that fails to function properly if the author simply waits for all to be dropped on his lap. "Publishers should think clearly and completely about what they want to guarantee

authors by way of rights," says **Goldsmith**, "and then state these clearly so that authors understand what their own responsibilities are vis-à-vis the publisher."

The key to a successful author care program that's sustainable and fair must be based on mutual trust that can only develop when both publisher and author keep their respective ends of the collaborative bargain. This is not an on-off process. It's an iterative event that requires commitment and a willingness by the publisher to continually review what's being offered, how it can be improved, and seek author feedback. It equally requires the commitment of the author to take an active part in becoming informed about what's offered, produce high quality scholarship, and provide constructive suggestions to the publisher.

References

1. Smith, Jane. January 1, 2000. "Editorials: giving something back to authors." *British Medical Journal*. Available [Online]: www.bmj.org/cgi/content/full/320/7226/6.
2. Rennie, D. 1999. "Editorial peer review: its development and rationale." In: Goodle, F., Jefferson, T., eds. "Peer review in health sciences." London: *BMJ Books*.
3. See fn 1.
4. Smith, Jane. December 12, 200. Email interview.
5. See fn 1.
6. Kybett, Adrian. November 27, 2000. Email interview.
7. Crabtree, Robert. November 27, 2000. Email interview.
8. MCB University Press. "Authors' charter." 2000. Available [Online]: www.literaticlub.co.uk/writing/charter.html.
9. Tamas, Mick. November 30, 2000. Email interview.
10. See fn 8.
11. Toledano, Kathryn. November 8, 2000. Email interview.
12. See fn 8.
13. See fn 11.
14. See fn 9.
15. Goldsmith, Ronald. November 6, 2000. Email interview.

Digital Archiving: A Work in Progress

by **Ana Arias Terry** (Vice President, Informed Strategies, 633 S. College Ave., Suite F, Fort Collins, CO 80524; Phone: 970-472-5985, Fax: 970-490-5982) <ana.terry@informedstrategies.com> www.informedstrategies.com

The field of digital archiving is new in human history and presents a wealth of challenges exacerbated by the increasing rapidity with which the associated technologies are evolving. To grapple with the issues involved, let's take a stroll along the river of time.

Then to Now (sans digital)

The earliest archives can easily be blamed on happenstance. We are just fortunate that pigments, hidden from sunlight on cave walls and which didn't get washed with water and minerals, tended to endure. And that another early publishing media, chiseled stone, just happened to weather very well—or at least until our modern industrial age invented acid rain.

But in the more formal history of archiving that has accompanied the time of commercial publishing, which covers the last 500 years, the medium has been ink and paper. Archival efforts were driven by a few key goals whose constitution varied from archival body to archival body. These basically came down to preserving material for research, for cultural heritage, for legal reasons, and for consumer interest.

Traditionally there has been a separation between the libraries and archives.¹ In the experience of some there is rarely a library budget item for archiving.² Much of the formal archiving is done as an internal support structure as **Rick Anderson**, Electronic Resources/Serials Coordinator, The University

Libraries, **University of Nevada**, puts it — "a kind of defensive mode" — designed to make sure documents will be available for future needs. Libraries, on the other hand, usually operate under the guise of research and personal study support, and public interest support.³

Selection of what to archive and what not to archive has been a combination of the goals of the institution and its funding. The resulting criteria varies amongst archiving bodies. **Anderson** simplifies it nicely into two core criteria: the beautiful and the useful. How each institution defines those depends upon their motivation to archive.⁴ Who is it that said, "History is written by the archivists?"

Technical issues have been minimal for print media. As **Kathryn Toledano**, director of Business Development for **MCB University Press**, succinctly puts it, "We just had to make sure it kept dry and didn't burn."⁵ Certainly there are other issues like acidity and humidity, but they basically amount to the details.

The remaining piece of the archive puzzle is access. An archive serves little benefit if people can't get to that which is being archived. Funds bought shelves in buildings. Librarians provided expert locating of material for those who sought research, scholarly, or professional assistance. And the education many people received early in life provided the basic skills necessary to read the materi-

als. So the primary issues with access, such as education, shelves, and building distribution, tended to be outside the archival process.

Now (media stew)

The infusion of digital media in our information stream hasn't changed the basic goals behind archiving. According to **Regina Reynolds**, Head of the National Serials Data Program at the **Library of Congress**, "Our reasons are the same as those of any civilizations which have recorded and preserved the output of their culture for thousands of years." And, as she points out, **George Santayana** gave us a more compelling reason: "Those who do not learn from the past are condemned to repeat it."⁶

Beyond that similarity with the pure print archiving history, everything is up for grabs.

The boundaries between traditional libraries and archives are blurring in the realm of digital media. "We're really talking about finding a way to preserve our cultural and intellectual heritage at a time when that heritage is increasingly finding its expression in highly ephemeral contexts and formats," says **Anderson**.⁷ With the variety of technologies

continued on page 30

supporting digital materials and “ownership” issues muddled by access versus ownership paradigms, it has become less clear where archival responsibility, and even archival opportunity, lies.

If an archival body such as a library is buying access to an e-journal and then discontinues the subscription, how can they archive something they don’t own, but that the publisher owns? “If the library has paid for this information, should they not be allowed to keep it?” says **Bev Bruce**, director of **iGoli**, an international consultancy specializing in special project implementation and consortia licensing for publishers, intermediaries, document delivery organizations, and corporations. While she thinks librarians are the experts in archiving, publishers can make a significant contribution. “I believe that publishers can help in the process by making their licenses liberal and allowing ownership rather than access of the data, in the same way as you do with hard copy,” says **Bruce**.⁸

This will have to be a collaborative effort, which will probably be spearheaded by more formal, non-commercial archival and library bodies.⁹ Publishers, however, still ought to have a major participatory role.

As with anything archival, credibility means long-term stability and appropriate technical savvy. Many argue that publishers run the risk of merging with others or going out of business. Traditional libraries are still faced with the issue of the invisible budget item for archiving. So large bodies like the **US Library of Congress**, the **British Library**, and their counterparts in other countries are likely the best options for coordinating this collaboration. Such bodies have a demonstrated history of archival responsibility and fiscal stability. As **Toledano** states, “Because they’ve demonstrated that they can do and have been doing it in the print medium, they are seen as more credible repositories.”¹⁰

So as the “who” is being hashed out, we are also faced with a much more complex “what.” **Jerome McDonough**, Digital Library Development Team Leader, Elmer Holmes Bobst Library, **New York University**, feels that “Selection criteria for materials to add to a collection, or to deaccession, should not be significantly affected by the move to digital media, as these decisions should be driven by the current or potential use of the information, not the media carrying it.”¹¹ We’re back to the beautiful and useful. The trouble comes in identifying where the boundary between one item and the next is, and how, therefore, to capture that.

An example of the implications of this capturing issue can be illustrated with **MCB’s Corporate Communication Internet Re-**

search Register. This is a database of current research on corporate communication topics that is continually being updated by the researchers themselves. The records outline the research, what it is, who’s funding it, who the researcher is collaborating with, current status, reference papers given regarding it, and the like. Access to the **Register** is provided as part of the regular journal subscription. What then constitutes an archive of an issue of a journal? Do you provide a snapshot of the complete **Register** at the time the issue is made available? Or do you provide a snapshot of the complete **Register** at the time moments before the succeeding issue is made available? Or do you just capture any **Register** records that relate to articles in the issue? Or are the two treated separately, even if an article references something in the **Register**?¹²

In fact, these are the very types of issues that the **British Library** is attempting to sort through. The **Library** is exploring what processes and systems need to be developed for digital archiving. It is currently in discussions with publishers about providing the **Library** with voluntary deposit of content to help address what ought to be included and what the technological and process implications are. **Toledano** sees great value toward the body of knowledge and experience in such an effort, and **MCB** has made a commitment to participate in this project.¹³

“It’s very early days, but we are one of the first of the larger UK publishers to actually have agreed in principle with the **British Library** that it will make its e-journal and other e-content available on a voluntary basis,” says **Toledano**. “We’ve agreed to work with them on an experimental basis so that they can understand what processes, systems, innovations they need to do, what kind of relationships they need with a publisher to get this content, and what standards need to emerge.”¹⁴

“My guess is that everything produced electronically will be ‘archived’ in some fashion,” surmises **Bruce Heterick**, director for Library Relations, **JSTOR**. “Whether that archive is dark or light, whether the archive exists for three months or 300 years, and whether that archive has been created to serve the needs of the scholarly research community or simply the public in general, are but a few of the questions we need to ask.”¹⁵

Toledano adds, “So I’m with the **British Library**, whose starting point is to go for archiving as much as possible, at least until we have better research on how best to capture things like dynamic content or on what to base the selection criteria. But that will have implications in terms of cost. So the decision on what to archive digitally may be a wish list and may be that which can be afforded or that which can be funded on a sustainable basis.”¹⁶

Anderson points out why it’s critical to

address the issue of what to archive now. “The creation of information no longer necessarily involves the creation of physical artifacts that can be counted on to sit there on a shelf or in a folder while we figure out how to organize them,” he states.¹⁷

Different archival bodies are trying different approaches to this. Some are archiving reference materials if they are internal to the same corpus. Others will go so far as to get the text of the referenced material if it is external to their corpus.¹⁸ And all of this hinges on the major concern for intellectual property.

This brings us to the piece mentioned earlier that needs resolution. How does one archive something that the archiving body only has access privileges to, yet that material may be important in context to what can be archived? Something more formal needs to be arranged to ensure that the holder of the referenced material archives those pieces. All of this necessitates a large-scale collaborative effort preferably without regard for political borders.¹⁹

Once archival value is assigned, the longevity of that value still needs to be assessed

continued on page 32

References

1. Anderson, Rick. November 28, 2000. Email interview.
2. Heterick, Bruce. December 13, 2000. Email interview.
3. See fn 1.
4. See fn 1.
5. Toledano, Kathryn. December 1, 2000. Personal interview.
6. Regina, Reynolds. December 12, 2000. Email interview.
7. See fn 1.
8. Bruce, Beverly. December 20, 2000. Email interview.
9. See fn 6.
10. See fn 5.
11. McDonough, Jerome. December 12, 2000. Email interview.
12. See fn 5.
13. See fn 5.
14. See fn 5.
15. See fn 2.
16. See fn 5.
17. See fn 1.
18. Hodge, Gail M. “Best practices for digital archiving: an information life cycle approach.” *D-Lib Magazine*. January 2000. Vol 6, No. 1.
19. Terry, William D. December 20, 2000. Personal interview.
20. O’Neill, Jill. December 11, 2000. Email interview.
21. Meyers, Barbara and Beebe, Linda. 1999. “White Paper: Archiving: from a publisher’s point of view.” *The Sheridan Press*.
22. See fn 8.
23. See fn 2.
24. See fn 20.
25. See fn 2.

periodically. **Jill O'Neill**, director of Planning & Communication for **NFAIS**, points out that, "The activities of selection, collection development and management arose, at least in part, from the recognition by information professionals that not all information retains a constant value to a community of users over time."²⁰

The technical issues, while voluminous, aren't nearly as daunting. And they basically boil down to the need for large, continual funding for technology. With today's rate of technical evolution, it's possible that media will survive longer than the technology required to access it. A means of preserving both the content and the means to access it are critical to digital archiving.²¹ At that point, **Bruce** says, "The technological aspects of archiving can be taken care of by the libraries [or archiving bodies], be it preservation—retaining the original technology that the content was accessed by; emulation—emulating the technology that was originally used to access the content; or migration—moving the content to an upgraded piece of technology."²²

"If an archive is to be trusted, it will have to have a reasonable plan for generating the

resources necessary to cover the costs of regular and continuing migration of software, systems, and data that is going to be necessary to insure long-term accessibility," says **Heterick**.²³

With digital archives, the access infrastructure is well defined and growing on its own. However, the digital archives have to grapple with how to provide the means for people to access the content of the archive on a continuing basis even as the technologies change. And there is the pervasive funding issue lurking underneath this. "The practice cannot be considered cost-effective," says **O'Neill**, "if there is no balance achieved between the expense of retaining the information and the value derived by the use of the information. Not surprisingly, the value and use of an archive will be closely tied to the ease of extracting information from that archive."²⁴ **Heterick's** thoughts follow a similar path as he points out that, "If we're going to promote trusted archives, then we must also have trusted access points that will take responsibility for maintaining the interfaces that will drive the use of the archives or repositories."²⁵

Now to When (imagination is a terrible thing to waste)

Where does this leave us? With a lot of work to do on what is probably one of

the biggest collaborative efforts in the information community to date.

Individual organizations or even small groups can't tackle this solo. Information is too cross-related and what one entity does affects others. **Toledano** agrees. "It's going to take serious collaboration amongst publishers, librarians, and other repository bodies to actually develop practical and effective options."

One of the benefits of this upheaval will be the creation of a process to deal with the "new." The history of archiving has seen some challenges. The digital part of that isn't an insurmountable obstacle. But it is a major hurdle in the archiving river of time.

The next occasion on which something completely new presents itself, we'll be able to look at the lessons learned addressing digital archiving issues and be able to apply them to clearing that next hurdle. It's also conceivable that the foundations laid with addressing digital archiving will be able to support all new archiving hurdles.

Part of what's being addressed is not just the specifics of digital archiving, but what should be explored when faced with a completely new archiving dilemma.

Biz of Acq — Alibris and 21 North Main: Breaking New Ground in the Search for O.P. Books

Column Editor: **Rob Richards** (Technical Services Librarian, University of Colorado Law Library, Fleming Law Building, Rm 190, 2405 Kittredge Loop Dr., Boulder, CO 80309-0402; Phone: 303-492-2706; Fax: 303-492-2707) <rrichard@stripe.colorado.edu> www.Colorado.EDU/Law/lawlib/rr.htm

by **Dennis K. Lambert** (Head, Collection Development and Management, Villanova University - Falvey Memorial Library, 800 Lancaster Ave., Villanova, PA 19085-1683) <dennis.lambert@villanova.edu>

Column Editor's Note: In recent years, the out-of-print market has consistently borne happy tidings to acquisitions librarians. Fulfillment rates have lept with the rise of Web-based mega-catalogs that aggregate thousands of book dealers' inventories. Now innovative vendors have further improved out-of-print acquisitions, by adopting the practices of Internet retailers. In this month's column, **Dennis Lambert**, Head of Collection Development and Management at **Villanova University Libraries**, profiles two new out-of-print vendors. Lambert finds that these used booksellers raise the bar of customer service standards, and give acquisitions librarians promising choices for out of print purchasing. — **RR**

If you had suggested to me a few years ago that I could now use either of two easy-to-search Internet sites for out-of-print (o.p.) books, I would not have believed you. While I was aware of many improvements in o.p. searching, I still felt it was a highly problematic field. Now, with the arrival of **Alibris** and **21 North Main**, I am quite optimistic.

In addition to being able to locate many more academic books for my library, I have been able to solve all three personal requests made to me in the last six months. I was challenged to find a Spanish-language reader authored by a professor from Cleveland, an o.p. library science text, and a 1972 book of teaching materials for kindergartners. Using **21 North Main** and **Alibris**, I found not only all three of these books, but found them at

reasonable prices.

One was astonishingly cheap at \$7.50; the other two were \$24 and \$26, including shipping charges.

John Zubal calls the embracing of the Internet by o.p. dealers "The Revolution," noting that it began in 1994 with **Interloc**.¹ I was not aware until 2000 of the significant improvements in searchability, convenience, and service introduced by new competitors in a market that has known many years of uncertainty, frustration, and high prices.

The Continuing O.P. Problem

In the past, an o.p. report meant a library could entertain only modest expectations of acquiring that title. O.p. searching has been

continued on page 34