

CONTENTS:

[Editorial](#)

[Communicating Off The Press](#)

[Libraries 'Wasting' £2.25m a Year](#)

[Library Adds Research Service, Expands Access](#)

[Digital Preservation: Paradox & Promise](#)

[The 2001 Nobel and IG Nobel Prizes](#)

[Press Release by IFLA: Publishers and Librarians Promote Common Principles on Copyright in the Electronic Environment](#)

EDITORIAL

The conference came and went very successfully and seems to have happened a very long time ago. It is always exciting to read the feedback sheets from the delegates, just as you feel buoyed up by praise from one, someone else sweeps down you into the depths who absolutely hated the very thing that the previous person loved! Some of you were very descriptive, if I ever visit Caesars again I will definitely take note of the 'pulsating carpets'! Everyone enthused about Charles Oppenheim and Martin White which proves that we get more 'bang for your buck' if we bring international speakers over here to address all of us, rather than sponsoring individuals to go overseas.

To follow on with this trend, our next meeting on the 22nd October will feature Mary Ellen Bates, who will speak about "The Super Searcher : with the main focus on the Web". Mary Ellen has been a speaker at a number of recent SIA meetings and her website can be visited at <http://www.Batesinfo.com> for more information.

The chaos created by the World Trade Centre has touched us all and I have included an article on digital preservation. This is something we should all be practicing with our own collections, we take the internet for granted, but as this article shows, the accidental or purposeful push of the 'delete' button, or when a tragedy such as the WTC occurs, it can have enormous ramifications for the world of information.

COMMUNICATING OFF THE PAGE

The web has spawned a whole new generation of library publications, backed up with a do-it-yourself attitude

Marylaine Block

I was looking at old issues of The U*N*A*B*A*S*H*E*D Librarian last week and thinking that, in a sense, it was the original library zine: a personal magazine where the editor/publisher's philosophy and passion for his subject permeated every page. It also served as a filter, reading widely and sharing brief excerpts from every piece the editor thought would be useful or provocative. Which is to say, he was also running one

of the first weblogs (or blog), albeit without the web. As it is, he created a model of self-publishing followed by many who now are on the net.

Now, there are zines by lipstick librarians and laughing librarians, anarchist librarians and progressive ones (though no conservative ones that I can find). There are web sites for librarians who call themselves renegade, cool, ska, marginal, modified, pernicious, and psycho. There are blogs galore, Jessamyn West's Librarian.net, Steven Cohen's LibraryStuff, the shared blog at LISNews. the Internet Scout Weblog, and more.

The speed with which weblogs share links to news stories and commentary about major developments is impressive; by the time a printed magazine could have printed that news, people would have been reading the decision and arguing about it on weblogs and forums for more than a month.

The informality of the net, and the ease of e-mail, leads readers to see blogs and articles on the net not as complete, finished pieces but as the beginning of a conversation, an invitation to add their own voices to the fray. When I surveyed 20 librarians who publish zines and blogs for this article, all the respondents said that the greatest pleasure was getting feedback from all over the world. For Blake Carver, of LISNews, creating an online community was the actual goal 'I wanted a site that had a bunch of people contributing what they thought was important. Slashdot for librarians.' T.J. Sondermann, the 'Library Geek,' says, 'I do it for a sense of belonging. To both the library community and the weblogging community.' When I asked Jessamyn West why she began librarian.net, she answered, 'The best thing for me about the Internet is that if you are a freaky person with fringe interests you can find someone else to share those interests with. Who would have known there was a tattooed librarian subculture?'

Why do they do it? The reasons librarians create zines and blogs are as varied as the librarians. I'm the only one in the lot who hoped to make money doing it, at least indirectly. My zine has served my professional purposes by leading to reprint requests and other writing assignments. For Gary Price, the indefatigable finder of reference sources on the invisible web, his weblog, the Virtual Acquisition Shelf & News Desk, is simply an easier way than e-mail to let his followers know about new sites he's added to his many web pages. For Library Geek Sonderman, it's a way of helping him to stay current with professional issues. 'Because I scour the net so much, looking for links, I'm continually updating my knowledge base," he says.

Community service: For many of the self-publishers, it's a chance to render a service, to fill a hole in the web of information. Jenny Levine was one of the first to do this, back in 1995, with her late lamented Librarians' Site du Jour. 'I began it to bring home to the librarians in my system the power of this new tool,' she says. 'The two biggest complaints I heard

about the net were that people didn't have time for this new stuff, and, even if they did, they didn't know what to do once they got online. So my goal was twofold: 1) to highlight valuable resources, and 2) to give librarians a reason to go on the web every day.'

John Guscott, who publishes an actual print magazine, *Library Futures Quarterly*, with a companion e-mailed newsletter, *Library Futures Express*, began his site because there was no library publication that was trying to consolidate in one place much of the current thinking about the future of libraries. He sees his publication as a current awareness tool, as well as a planning resource, for harried library administrators. Juanita Benedicto and Colleen Bell began *New Breed* librarian because they wanted to provide more of a full-service publication, that combined news items with articles, interviews, a job section, and advice (*Ask Susu*).

It's also true that most of us think self-publishing is just fun. We enjoy what Dodge calls 'unfettered self-expression' and what Eris Weaver, the 'Bellydancing Librarian,' calls 'the ability to put just whatever the hell out there with no editing or censorship.' She loves busting stereotypes and figures there can never be too much humour. The web, as Matthew Wright of *Library Underground* notes, 'offers freedom of speech to a maniacal degree.' Linda Absher originally thought lipstick librarian would just be a chance to 'make a few friends laugh. I had no idea I'd receive so much response and notoriety.' Brian Smith began one of his other web sites, *laughing Librarian*, because he's 'written and drawn parodies and other funny stuff for as long as I can remember' and thinks it's important for librarians to be able to laugh about their profession,

The blogs and zines. are part of an Internet society that values attitude, community, and breezy informality. To some, it makes the formal library literature suddenly look a bit stodgy and behind the curve. To others it becomes more apparent what professional editors are up to all day long.

In time professional publications will take on some of the attributes of web publications, perhaps relegating all news functions to blogs they'll offer on their own web sites. Library publishers could well do what the traditional news media did when the web came along: take advantage of the possibilities it offers for interactivity and reader responses to articles in the magazine, because the new generation of web-based librarians doesn't want to just read an article or discussion, they want to contribute to it. Publishers could use their web sites for lively opinion, discussion groups, and spot surveys. By combining peer-to-peer journalism with traditional library journalism, they could smoothly transition a new generation of web-based librarians to their magazines. If that happens, librarians could have in one package the best of both worlds the reliability of serious and systematic research and problem-solving, from some of the most knowledgeable librarians in the business, with the immediacy, playfulness, diversity of voices, and complete interactivity of the zines and blogs.

LIBRARIES 'WASTING' £2.25M A YEAR

Do you remember to check your own catalogues first ... ?

UK libraries could be wasting as much as £2.25 million a year through ineffective document ordering, according to new research findings. In the UK, some 3 million documents are delivered annually with an average payment of £5 per document. The research claims that around 15 per cent of document delivery requests come from libraries that already have subscriptions to the relevant journals. Meanwhile the recent surge in demand for individual articles does not seem to have affected demand for the journals themselves. While half of articles now looked at come from sources other than the original journal, the study reveals, journal subscriptions seem to be holding up. These are two of the first results to emerge from independent research being conducted on behalf of the Ingenta Institute (www.ingenta.com). The research project -- looking at user behavior when sourcing scholarly information -- was devised to fill a real gap in industry knowledge. David Brown, director of strategy of the Ingenta Institute, said that present business models 'are based on assumption and hearsay' because the publishing industry has no central body that currently organises such research. Brown said that while publishers could take comfort from the fact that document delivery does not seem to be affecting subscriptions rates --'it looks like additional demand' -- libraries need to look at the 'clear inefficiencies' in their document request systems.

EBRARY ADDS RESEARCH SERVICE, EXPANDS ACCESS

Ebrary, one of the first online research companies, has announced that it is expanding its business beyond the ebrary.com online library to make its books, journals and periodicals available through Internet portals, marketing channel partners, search engines and other Web sites. At the same time, the company unveiled its ebrarian program for channel partners, enabling them to market and securely distribute their works to wide communities of Internet users. The company's business model allows users to browse ebrary's texts for free, paying only for printing or downloading the information they need.

The ebrarian program is built on four points to help publishers become more efficient in serving customers, increase title usage and revenues and allow usage tracking.

- First, the ebrarian Infotools provide users with easy-to-find definitions, translations, biographical information and maps, as well as following links to additional materials available on the Web.
- Second, ebrary increases partners' site "stickiness," the rate at which readers stay at a site after linking to it, by allowing partners to integrate information in ebrary's collection with any content the partner wishes to make available to its customers.
- Third, the ebrarian program provides real-time activity reports on micro-transactions, document viewing, usage and purchases, providing partners with data about which titles are used most often, while protecting user privacy.
- Finally, ebrary's open-access model enables search engines and Web sites with search capabilities to freely index the company's content, and provide links directly to ebrary's collections, increasing site traffic for the portal, the publisher and ebrary.

By charging only for copying or printing information, Warnock said, ebrary ensures copyright protection, while recapturing for publishers and their marketing channel partners revenues that would have been lost to photocopying services. According to Warnock, the free viewing also encourages the sale of print and electronic documents by allowing consumers to "browse before buying." According to company officials, the price to license ebrarian will vary according to organisations' individual requirements, including customisation and hosting services.

DIGITAL PRESERVATION: PARADOX & PROMISE

Suppose one morning the world discovered that, with no advance notice, tens of thousands of pages of popular content had mysteriously disappeared from the web site of a major federal agency. The outcry would be immediate and loud. Senators would demand hearings. Pundits would proclaim depravity. Editorialists would decry the decline of public values.

Yet that is exactly what happened on January 20, 2001, Inauguration Day in the United States. When George Bush took over the presidency, he also took possession of the White House web site, www.whitehouse.gov. All of the previous content of that site, and its companion searchable document archive, www.pub.whitehouse.gov, were completely wiped clean, replaced with a skeleton site for the new administration. The result was a massive example of "link rot" in one of the most popular sites on the web. AltaVista reported 170,000 links to the site--many of them "deep links" (i.e., deep within the hierarchy of a web site)--that were suddenly broken. It is impossible to know how many thousands or millions of personal bookmarks were similarly trashed.

Of course, no one would expect that the new President would want to preserve a complete collection of the speeches and official communications of the Clinton administration. Historians, however, might take a different view; such documents are vital to analysis of a presidency. Citizens, too, might want to be able to look up a presidential document, whether it is a disaster decree from 1993 or the text of the pardon of Marc Rich.

In this case, not all is lost to history. The National Archives and Records Administration (NARA) has undertaken an effort to preserve several renditions of whitehouse.gov across the Clinton years--which happened to correspond to the years of the web revolution. While the NARA archive at clinton.nara.gov lacks much of the functionality of the former whitehouse.gov sites, at least the content itself is preserved--or so we are assured. (Over a month after the transition, NARA's site said, "Please check back with us. We are busy processing records and will post our project schedule as soon as possible.") In fact, not only did NARA take steps to preserve the content of the Clinton White House, it also asked all

agencies in the Executive Branch to take "snapshots" of their web, sites as they existed at the end of the Clinton administration. On January 12, 2001-- just eight days before the inauguration--Lewis J. Bellardo of NARA sent a fax to "agency record officers and information resource administrators" with this request; news reports indicate many agencies had trouble complying. While one appreciates NARA's efforts, one wonders how much more could have been accomplished with more planning and notice.

This case exemplifies several of the challenges of preservation in the digital age. Not only is NARA motivated by its very mission to preserve the digital content under its auspices, but the Presidential Records Act of 1978 makes it the law. Alas, not every custodian of digital content operates under such motivation. In many cases, content may be born in digital, form, live online for months or years, and then vanish--without a trace, never to be recovered.

The changing of the political guard is but a single cause of such disappearances. We can identify different modes of digital death:

- **The New Replaces the Old:** Every time an organisation has new information to put on a web site or CD-ROM or other digital format, there is a tendency to publish the new content and simply to overwrite or toss the old.
- **Content Reorganisation:** Particularly in the web sphere, it is popular to reorganise content space periodically: when a new master is assigned to care for the content or when significant change in the content structure has occurred. Many an Error 404 stems from simple reorganisation.
- **Death of a Sponsor:** When the sponsoring organisation of a collection dies, so too may its digital content.
- **Sponsor Loses Interest:** Most traditional print publishers consider the "back file" to be an asset worth protecting. Web-based publishers seem to lack this long-term view. For instance, Internet World, a print publication with a companion web site, has published since 1994. Until recently, a complete archive of back issues appeared on the web site. Now, the archive extends back only to July 1999.
- **Sponsor Fears History:** In many cases, corporations may consciously avoid maintaining historical documents for fear of litigation. Ford, IBM, and other major companies have been sued in recent years for infractions alleged to have taken place over 50 years ago. As corporate knowledge moves to the company intranet, corporations will face an increasing dilemma: to what extent are we protecting ourselves from potential litigation, and to what extent are we destroying corporate knowledge?
- **Lost Functionality:** The Disappearance of www.pub.whitehouse.gov, an MIT-developed site that offered rich search functionality, exemplifies this particular form of digital death. While NARA has preserved the raw content, the NARA search interface pates in

comparison to the MIT product.

- **Media Format Obsolescence:** Anyone who owns data stranded on a 5 1/4" floppy disk knows the impact of media format changes. A newer storage technology supplants your format of choice; unless you take steps to copy old content to newer media, your data become stranded.
- **Content Format Obsolescence:** Data stored in a proprietary format, such as an older version of WordPerfect or an obsolete tool such as PC Write, may be completely unreadable, to current versions of software.
- **Disaster:** Whether a small-scale disaster (e.g. server meltdown) or a disaster that affects a campus (e.g. the 1994 earthquake that destroyed much of Cal State-Northridge), disaster can wipe out digital data whose sponsors have failed to provide for adequate offsite backup.

We live in an era when digital storage costs are at an all-time low. A decade ago, disk space cost almost 400 times more than it does currently. This apparent improvement carries costs as well. Writing in *Library Journal* in 1999, Stewart Brand cited the overarching paradox of digital preservation: people think of digital content as inexpensive--and inexpensive things are not worth preserving. And with greater storage comes the desire for ever larger files--digital feature films were unimaginably large just a few years ago. Still, the new storage capacities seem impressive.

In 1991, the network organisation of the Committee on Institutional Cooperation (Big Ten schools plus the University of Chicago) saw a need to preserve electronic serials and launched the GICNet Journal Archive. This ambitious digital preservation project sought to capture and archive the content of numerous nascent electronic journals. Alas, essential funding never appeared, and GICNet itself ceased operations in 1997. The GICNet Journal Archive vanished with it. Ironic, indeed, to lose not a mere collection but an archive whose purpose was to prevent loss of electronic content.

The raw cost of disk space is merely one component in the costs of offering digital content online. Disks must be connected to servers, server hardware and software also cost money. For a large or popular collection, electricity begins to become a factor. Systems administration requires talents that are in short supply these days, adding significant costs. In some cases, custodians of content may choose to archive it offline. rather than paying all of these costs of online storage. In this case, the custodians must choose an offline digital storage medium that they think will endure for a period of time. Unfortunately, we have not settled on such a single digital storage medium

Capturing Web Sites

There are two basic approaches to archiving a web site for posterity.

- Preserve the content in the format in which it is stored on the server, whether flat-file HTML, Java or JavaScript, database records, etc. This assumes access to original content normally only granted to the webmaster or system administrator.
- Capture the content as experienced by an end user browsing the site, employing a software tool. This can be done by anyone with access to such tools.

Web content archived in internal format assumes that the future reviewer will be able to use all the tools necessary to cope with all the various forms of content on the site. Moreover, if the content is merely dumped to offline media in a native format, it will be necessary to restore the operating environment in which it was served as well as the content. In order to archive a web site externally, software must play the role of an end user with a web browser.

Site-centric archiving of web content, undertaken by a webmaster, tends to be very complete but may require significant work to restore to a usable web experience. Client-centric archiving, done by an external party, tends to yield renditions of the site that can be navigated using any browser--but the rendition may be very incomplete.

No matter what the medium, when digital content is preserved offline, custodians must be prepared to migrate the content to new formats. This may be necessary because the old format becomes obsolete. or it might be because the physical storage medium is in danger of wearing out. Archivists agree the only answer is periodic digital inspection of a sample of the corpus and copying to new media as required.

The Internet Archive

A great deal of the digital content we care about exists on the web. Brewster Kahle's ambitious Internet Archive project attempts to capture all of the content on the publicly accessible web site. The archive now consists of some 40 to 70 terabytes of content online, with data for 1999 through the present on disk. Content dating back to the start of the archive is on tape. Every month ten or so more terabytes of web content are contributed. By contrast, according to estimates by Kahle and by other experts, all the text of all the holdings of the library of Congress would require about 17 terabytes of storage.

While the archive's initial focus is on the textual content of web pages, it is branching into multimedia content as well as nonweb content. For instance, the archive is experimenting with capturing television for on-demand streaming of old content (e.g., news programming) via the web. Thus far, tools for navigating the archive are primitive. Kahle says the goal is to produce an interface whereby a scholar could 'dial back the clock' to a particular date and surf the archive's view of the web on that

day.

Over 93 percent of new information produced is created in digital form-- or "born digital." Berkeley researcher Peter Lyman observes, 'Information produced by institutions (governments, corporations, universities) is far more likely to be preserved because institutions have an urgent need to create and preserve their own archival histories. But private individuals rarely have the motivation to preserve digital documents, and don't always have the resources or expertise to do so if they recognise the need.' Even individuals or small organisations that recognise the need to preserve their digital content may find that the tools they choose tailor them. Driveway.com, a provider of free web-based digital storage, had amassed a user base of some two million customers. On February 20, 2001, the company announced the demise of this service effective March 5. A customer on a two-week vacation would have returned to find precious files no longer online.

Whether for individuals or for institutions, however, the issue for custodians of digital content is which items and collections that are "born digital" ought to be preserved. Not all content is worth the cost of preservation.

The copyright barrier: During the 2000 election campaign, the Internet Archive partnered with the Library of Congress to capture for posterity various campaign-related web sites. Yet the archive's Kahle indicates that questions about intellectual property rights cloud the ability to offer unfettered access to these materials, once freely available on the web. Rights management issues affect every digital preservation project.

Is there hope?

While digital content can be ephemeral, digitisation can also preserve content. Many projects to digitise paper or analogue media incorporate preservation as a goal. For instance, the National Gallery of the Spoken Word (www.ngsw.org) seeks to digitise speech content from the Vincent Voice Library at Michigan State University, among other sources. Much of the Vincent collection is on old reel-to-reel audiotape that is brittle or decomposing. Once converted to digital format, the prospects for preservation actually increase.

Once it is in the digital realm, its custodians must fight the various modes of digital death. To solve the problem of media obsolescence, there is one simple approach: copy the digital content to new media as it evolves. To combat format obsolescence, researchers propose several approaches:

- **Preserve the Software.** This requires keeping the original software program used to view or edit digital content intact. The problem of preserving corresponding operating systems and hardware compounds the issue.
- **Refresh or migrate** as appropriate.

- **Emulate.** Provide software tools to emulate essential viewing applications or operating environments.
- **Encapsulate.** When storing digital content, store not only the text, images, audio and video, etc., that end users consume, but also store the necessary information to interpret the content using new systems across time.

These strategies become progressively more theoretical. It is one thing to copy from a floppy disk to a CD-R; it is quite another to build a software emulator for WordStar under DOS or to store a Microsoft Word 2000 document in an encapsulated form understandable to some future software system.

These approaches address some of the technological aspects of digital preservation. The economic, social, and legal barriers may be the more daunting, however. Nevertheless, libraries and librarians have long had a mandate for preservation as a part of their mission. Jeanne Drewes, assistant director for access and preservation at Michigan State University Libraries, who consults with libraries that have suffered disasters, notes, 'Digital is the preservation issue now and for the foreseeable future. After many years of dealing with paper-based materials, libraries and archives have an understanding of the issues and the process for preserving materials.'

This article originally appeared in the Spring 2001 issue of netconnect.

THE 2001 NOBEL AND IG NOBEL PRIZES 2001

Nobel Prize Winners
<http://www.nobel.se>

2001 Ig Nobel Prize Winners
<http://www.improbable.com/ig/ig-pastwinners.html#ig2001>

This week, the Nobel Foundation announced the winners of its six awards for 2001, the 100th anniversary of the Nobel Prize. The United Nations and its Secretary General Kofi Annan were awarded the Nobel Prize for Peace. The Prize in Literature went to V.S. Naipaul 'for having united perceptive narrative and incorruptible scrutiny in works that compel us to see the presence of suppressed histories.' Three Physics laureates were awarded, and the Nobel prize for Chemistry went to an international team of researchers. (The Nobel Foundation's online "e-museum" was reviewed in the August 8, 2000 Scout Report for Social Sciences & Humanities.) At the official Nobel Website, press releases for these categories, as well as Economics and Physiology/ Medicine, can be read in English, French, German, or Swedish, and links to the Curriculum Vitae and publication lists of the laureates are given. Back in the United States, another set of awards were passed out this week: the 19 Nobels. Awarded by humor rag The Annals of Improbable Research, the Ig Nobels honour people whose achievements 'cannot or should not be reproduced.' This year's illustrious Ig Nobel laureates include, for Medicine, the publisher of "Injuries Due to Failing Coconuts," in the Journal of Trauma, the founder of the Apostrophe Protection Society (Literature), and the Peace Prize goes to the

Lithuanian who built an amusement park known colloquially as 'Stalin World.'

(From the Scout Report: Oct 12, 2001)

PRESS RELEASE BY IFLA. PUBLISHERS AND LIBRARIANS PROMOTE COMMON PRINCIPLES ON COPYRIGHT IN THE ELECTRONIC ENVIRONMENT

12 October 2001

The joint Steering Group of the International Federation of Library Associations and Institutions (IFLA) and the International Publishers Association (IPA) is working to promote common principles on copyright in the electronic environment. They are pleased to announce further progress on the following principles:

While the fundamental principles underlying copyright protection in the print environment remain the same in the electronic environment (in this sense, "digital is not different"), the Group recognises that the advent of new technologies has fundamentally changed methods of publication and dissemination as well as rights management (in this sense, "digital is different"). Bridging the digital divide is best achieved by government programmes increasing funding for the provision of books and other publications in libraries as well as for connecting end-users to the Internet, especially in developing countries and disadvantaged groups in developed nations. Exceptions and limitations to copyright in the public interest remain necessary in the electronic environment, in order to maintain an equitable balance between the rights of creators and distributors and the needs of users but the nature and extent of exceptions and limitations must be assessed by applying the three step test. Libraries are key players in ensuring long-term preservation archiving of electronic information, through appropriate arrangements with publishers. However, the conditions of access and other technical and policy issues require further discussion among stake holders.

For the full text of this press release, go to <http://www.ipa-ue.org>.

Southern African Online User Group
<http://www.saoug.org.za>
Copyright © 2001
Update: December 2001