

“Knowledge Biodiversity:” the Precarious Economics of Preserving the World’s News

*A paper produced for the Association of College and Research Libraries Meeting,
Minneapolis, MN
April 2005*

Bernard F. Reilly, Jr.
Center for Research Libraries

For a decade now alarms have been sounded, by voices as diverse as Deanna Marcum of the Council on Library and Information Resources and novelist and bibliophile Nicholson Baker, about the loss of important recorded knowledge and cultural heritage. The situation, it goes, is that the record of human achievement maintained continually by libraries for centuries by preserving books, manuscripts, photographs, newspapers, films and sound recordings, and other types of artifacts is today in jeopardy.

This paper examines the challenges of preserving a small subset of the record of human achievement that is particularly at risk: the news or, more specifically, news from the developing world. In this realm the peril is not, as one might suspect, the inadequacy of technologies for preserving and delivering analog and digital content. Digital media have vastly improved our ability to reproduce, store, and disseminate both printed and born-digital news content. Rather, endangering the survival of such knowledge are the *economic* forces that affect the development and deployment of those technologies. It is true that the rapid change and obsolescence of formats, devices, and practices do present challenges. But the growing consolidation of the publishing and broadcast media and the economics of the knowledge market constitute a far greater threat.

What these economic forces portend is not simply a future scarcity of historical information, nor the loss of massive amounts of knowledge. The overabundance of information today is patently obvious. It appears rather that the first casualty of modern market forces is the multiplicity of viewpoints and perspectives on our world that comes from a diversity – or “biodiversity” -- of knowledge sources. The threat is less information poverty than the homogeneity of knowledge. Advanced scholarly research in the humanities and social sciences requires a rich and varied pool of information and documentation. An understanding the world’s cultures and their respective achievements requires access to the records of those cultures. Certain economic forces now threaten the survival of important documents, communications, and evidence critical to the heritage of people and regions outside the industrialized world.

Two recent studies undertaken by the Center for Research Libraries provide evidence that this threat is real. The first, a survey of major research libraries that hold foreign newspapers, was undertaken by Melissa Roser and James Simon of the International Coalition on Newspapers (ICON). ICON is a preservation project supported by the Center for Research Libraries and the National Endowment for the Humanities.

The second study arose from an investigation, led by the Center for Research libraries and supported by the Andrew W. Mellon Foundation, of prospects for systematically archiving political Web communications. The Political Communications Web Archiving investigation (PCWA) examined the economic, curatorial and technological challenges of preserving the voluminous on-line content produced by political parties, insurgencies, factions, and other non-governmental entities in the developing world. This study too dealt with news content, by examining the proliferation of news sources and alternative news outlets which challenge and confound traditional library collecting methodologies.

In addition to reporting on the findings of these studies, I will here suggest measures that can and should be taken by libraries to stem the loss of critical knowledge and heritage.

Preserving News from the Emerging World

For centuries news has been a primary source mined by humanities and social science scholars in every discipline. Newspapers yield important data on everyday life, politics, government, commerce, and culture. As the “first rough draft of history,” they mirror the values, biases, aspirations, perceptions and (equally important) misperceptions of their communities and of their times.

Major research libraries in the US and other parts of the industrialized world have traditionally undertaken to preserve newspapers and other heritage materials from regions where the infrastructure necessary to maintain and preserve those materials either has not existed or has been suppressed for political or other reasons. The wisdom of such a paternalistic approach, a model inherited from the colonial era, has been affirmed by recent events like the destruction of much of the contents of the Kabul Public Library, Kabul University Library and Afghanistan’s national museum by the Taliban and Afghan warlords during the 1990s, the natural destruction of parts of the national archives of Mozambique, and the emptying of Iraq’s national library during the 2003 U.S. invasion.

Unfortunately, funding for American academic and independent research libraries has been in decline, while the cost of acquiring, managing and preserving library holdings has continued to rise. As a result the major Western research libraries today are less able to provide stewardship of heritage materials from other regions than they were in the past. The ICON survey revealed considerable attrition in the extent to which developed-world libraries fulfill their traditional stewardship role. The survey suggests that the major U.S. research universities and national institutions acquire and preserve fewer non-U.S. newspapers than they did two decades ago. Moreover, the study found that in libraries where historical print runs of newspapers are still preserved, holdings are often no longer systematically or comprehensively cataloged.¹

The survey queried the libraries of 40 institutions, most of them academic libraries at Carnegie Doctoral Research-Extensive universities, Doctoral Research-Intensive universities,

¹ The ICON survey report is available on-line at http://www.crl.edu/PDF/ICON_Survey_Results.PDF.

and their Canadian equivalents; public and national research libraries; colleges offering masters and four-year liberal arts programs; and a few specialized institutions.

Survey results indicated that most libraries collect international newspapers only for immediate or current awareness use, and do not retain hard copy for any length of time. At universities where advanced research in the humanities and social sciences is not explicitly supported, hard copies are often discarded within six months to a year. Even the largest North American research universities, the survey found, retain few non-U.S. titles of newspapers for which either microfilm or full-text electronic editions are available.

For “libraries of last resort,” like the Library of Congress, National Library and Archives of Canada, the British Library, and the New York Public Library permanent retention of print versions of foreign newspapers is the exception rather than the norm. The Library of Congress still maintains subscriptions to an estimated 800 titles, but defines the “best edition” of those titles as microfilm, generally disposing of paper copies, except for titles “being of special note,” when they are replaced by microfilm. In 1988 the National Library and Archives of Canada discontinued subscriptions to microfilm or original editions of all but eight major foreign newspapers.

Moreover, academic and public libraries today are acquiring fewer foreign newspapers to begin with. On-line editions of the newspapers and news aggregation services provided by Lexis-Nexis, NewsBank, and Thomson Gale are the preferred means of keeping library users abreast of recent events. News is available online immediately, without the days or even weeks that elapse before the arrival of the paper edition. The problem is that on-line subscriptions to such news do not create back runs of news content under the control of the individual libraries that print subscriptions did. While major Western news organizations like the New York Times and the Tribune Company make “back runs” of their on-line news available by subscription through the aggregators like NewsBank, such is not the case with most newspapers from emerging regions.²

The Mexico City newspaper *La Jornada* is a case in point. An important chronicle of the Zapatista uprising in Chiapas during the 1990s, the paper’s on-line news reports from that period are no longer maintained on line. Their fate illustrates the dynamics of the media free market at work: rather than provide an archive of those reports, cited by scholars as a uniquely valuable record of the insurgency, the editors of *La Jornada* marked the 19th anniversary of the rebellion with a heavily branded video promoting the sale of a book of photographs and texts culled from the newspaper’s files.³

The consumer-driven behaviors of media organizations are creating other preservation challenges as well. Until 2002 the Library of Congress and the Center for Research Libraries regularly microfilmed the principal daily newspaper of Jamaica, *The Gleaner*. Published in Kingston Jamaica since 1834, the newspaper holds a wealth of historical information on the modern nations and peoples of the Caribbean. The Library-CRL effort ensured availability

² Some foreign on-line news is archived and made available through commercial aggregators in products like NewsBank’s *Access World News*, but only for a limited number of years. Information on *Access World News* is available on-line at <http://www.newsbank.com/features/awn.html>.

³ The promotion page for the *La Jornada* book is at <http://www.jornada.unam.mx/linearota/index.html>.

of a back run of over a hundred years of *The Gleaner* – until 2002. In that year the Gleaner Publishing Company executed an agreement with Heritage Microfilm, Inc. giving Heritage, as for-profit firm, exclusive rights to reformat and distribute *The Gleaner*, in both microform and digital formats.

Because Heritage now holds exclusive digital rights the agreement effectively shifts control of back files of that important newspaper to the commercial sector, where it is marketed primarily for genealogists and family history researchers. Henceforth the availability of this important knowledge and heritage asset will most likely be contingent upon the persistence of market demand, rather than scholarly need.

For much of the last half-century libraries' preservation efforts were augmented by an army of small microform publishers like Norman Ross, Inc., Custom Microfilm Services, and Scholarly Resources. With the consolidation of scholarly publishing in recent years these small players have not fared well. Custom Microfilm Services, an important source of newspapers like the Cuban Communist Party organ *Granma* and Chilean newspapers of the Salvador Allende period, dissolved in the late 1990s and much of its stock was dispersed.⁴ Scholarly Resources and Norman Ross Publishing were acquired in recent years by Thomson Gale and ProQuest respectively, rendering uncertain the future availability of their less widely subscribed titles in microform.

Prospects for preserving foreign broadcast news are not much brighter. For over fifty years the Foreign Broadcast Information Service (FBIS), a program of the federal government under the auspices of the Central Intelligence Agency, has recorded and disseminated to U.S. government agencies and academic libraries news broadcasts and political reporting from conflict zones, developing countries, and other parts of the world of interest to U.S. diplomatic, policy and intelligence communities. The *FBIS Daily Reports* consist of transcriptions and translations of radio and television broadcasts, news agency transmissions, newspapers, journals, technical reports, and bulletins from government and independent news outlets in hundreds of countries. The reports have long been a critical source for graduate research in international studies, business, law, trade, and political science.

Aside from being an important service to academic and policy researchers in the developed world, the *Daily Reports* also serve a secondary preservation function. Because the media organizations and even the national libraries in many of the regions monitored by the FBIS are ill-equipped to preserve the products of their country's news outlets, in many instances the *Daily Reports* are the only surviving record of broadcast news -- and news culture -- from emerging parts of the world.⁵

⁴ The microfilm negatives and positives disposed of by Custom Microfilm when they folded were salvaged from a jobber by Professor Henry Snyder of the University of California at Riverside. The Center for Research Libraries was able to take possession of several hundred reels of Custom stock of Latin American, Eastern European, and Middle Eastern materials.

⁵ The precarious situation of the audio archive of three decades of *Radio Afghanistan*, narrowly rescued from destruction at the hands of the Taliban, was documented in a 2002 British Broadcasting Company report. See William Reeve, "Kabul's Radio Treasure," *BBC World: From Our Own Correspondent*, May

Today the *Reports* are available online through the *World News Connection*® (WNC), a subscription-based product of the for-profit Thomson Gale Corporation. Firms like Thomson Gale serve as the de facto archives of these materials: they store and maintain them in electronic form and make them available to a large population of users. On the surface this looks like a winning proposition: Thomson Gale can provide indexing, storage, access, and delivery of news content more efficiently and less expensively than libraries.

The problem is that their products are most often aimed at the largest possible markets, to ensure the kinds of margins that profit-making organizations require. Such organizations tailor their products and services to lucrative markets like K-12 educators, the policy research community, and government agencies. The organizations cannot be relied upon to preserve heritage materials for smaller communities of stakeholders, such as those doing advanced humanities research, future historians, and the residents of emerging regions. These parties constitute small, even negligible markets for commercial publishers and aggregators.

The largest markets favor the most recent news content, a different set of materials than historians. Fortunately, for the time being at least, the *FBIS Daily Reports* are also available on microfiche for the smaller, library market by NewsBank. In the book publishing world, however, history has shown that content that is not heavily used, or heavily subscribed, quickly goes “out of print.”

News and Political Communications

The challenge of adequately archiving even the most important news from emerging regions is compounded by the rapid proliferation of news sources. The second investigation, funded by the Andrew W. Mellon Foundation and coordinated by the Center for Research Libraries, analyzed materials mounted on the World Wide Web by non-government political organizations and actors.⁶

The Web has emerged within the last decade as a vital medium of political communication. It now serves political activists, parties, popular fronts, and other non-governmental organizations (NGOs) as a global message board through which to communicate with constituents and the world community. The Web provides a widely accessible and relatively unrestricted medium for the rapid broadcast of information and public posting of critical documents such as manifestos, statements, constitutions, declarations, and treaties. The use of information and communications technologies

16, 2002. The report is available in the on-line BBC archive, at http://news.bbc.co.uk/1/hi/programmes/from_our_own_correspondent/1979482.stm

⁶ The report of the Political Communications Web Archiving project is available at <http://www.crl.edu/content/PolitWebReport.htm>. Project participants were drawn from New York University, the Latin American Network Information Center at the University of Texas at Austin, Cornell University, Stanford University, and the Internet Archive.

(ICTs) by political actors, particularly in the developing world, is now a significant field of study in the social sciences.

Sadly, the Web sites and documents are disappearing at an alarming rate. The political organizations and actors that produce them, unlike conventional news organizations, have no vested interest in maintaining them after their immediate purpose is achieved. Yet their Web messages and documents are the artifacts of political discourse and struggle.

Again, the Zapatistas provide a case in point. University of Texas Professor of Economics Harry Cleaver studied in depth how the Zapatistas used the Web to broadcast and gain worldwide support for their insurgency during the mid and late 1990s. Cleaver's Web site, *Zapatistas in Cyberspace*, <http://www.eco.utexas.edu/Homepages/Faculty/Cleaver/zapsincyber.html>, preserves many important electronic texts relating to the conflict. Cleaver's site archives postings from supporters' listservs and on-line communiqués from the Ejército Zapatista de Liberación Nacional (EZLN), and provides links to sites where similar texts are archived elsewhere.

Given the role that ICTs played in the Zapatistas' struggle, such messages, Web sites and documents will be critical raw materials for future historians of this important moment in modern Mexican history. Unfortunately, many of the documents have disappeared. On a page of his site poetically titled, "Broken Links, pages that were, but no longer seem to be," Cleaver lists twenty sites that have disappeared, or at least can no longer be resolved from their original URLs.

The PCWA study examined the work of a second researcher gathering similar materials from an entirely different part of the world. As a student at Georgetown University's Edmund A. Walsh School of Foreign Service, Sean McLaughlin analyzed the tactical use of the Internet by several Arab insurgency groups. McLaughlin's "The Use of the Internet for Political Action by Non-state Dissident Actors in the Middle East," published in *First Monday* in 2003, is a lengthy and revealing case study of Political Web production.⁷ McLaughlin studied more than twenty Web sites maintained by three dissident groups: the Muslim Brotherhood in Jordan, Muslim Brotherhood in Egypt, and the Movement for Islamic Reform in Arabia.

His research involved monitoring changes in the sites produced by the subject organizations by accessing those sites, at weekly intervals, throughout 2001-2002. McLaughlin supplemented his real-time monitoring by mining sites from the Internet Archive's *Wayback Machine*. The *Wayback Machine* had archived past instances of some of the subject organizations' sites from as early as 1996.

McLaughlin's study provides valuable information about the behaviors of Political Web producers, specifically about the activities of dissident groups in a region where

⁷ McLaughlin, Sean, "The Use of the Internet for Political Action by Non-state Dissident Actors in the Middle East," *First Monday*, VIII, 11 (November 2003) http://www.firstmonday.org/issues/issue8_11/

ensorship and other state-imposed constraints often suppress the free flow -- and archiving -- of news. He showed how groups like the Movement for Islamic Reform in Arabia (MIRA), founded in 1996 to promote Islamic reform within the Saudi kingdom, provided alternative sources of information and alternative perspectives on events in the region.

Based on his knowledge of these groups McLaughlin was able to document the losses of important evidence. While the materials archived in the *Wayback Machine* were useful to his study, McLaughlin observed that a great deal of original site content, such as images, captions, and sound and multimedia files, had vanished entirely. McLaughlin noted the absence in the archived sites in the Wayback Machine of audio recordings that had originally been available on some sites. This was a serious loss as the Saudi dissident groups posted recorded messages and videos to the sites on a regular, sometimes weekly basis.

In both the Zapatistas and the Arab insurgencies studies, entities other than libraries took on the role traditionally played by libraries and archives, of creating and making available to others important source materials. One may ask, however, why the other systems for knowledge preservation failed here. There are three reasons. The technological challenge is only the most obvious. Until recently most libraries did not have systems in place to enable the methodical selection and storage of Web materials. But today a variety of software and methods are available to harvest from the Web.

The second reason is political. Clearly it was not in the interest of the national or university libraries, the libraries of record of Mexico or Saudi Arabia, to preserve the materials at risk. The Zapatistas and the Arab insurgencies were advancing agendas hostile to the local regimes.

The most important reason, however, is economic. There was simply no compelling financial reason for electronic publishers to invest even the minimal capital required to archive the Zapatista and radical Muslim materials. Significant audiences for these materials did not exist at the time.

Furthermore, the CRL studies illustrate that another development undermines our ability to capture and preserve news content: the “atomization” of news. As news sources have proliferated, so have the forms in which information on current events is received multiplied. Web sites, blogs, RSS notification services, and electronic aggregations and indexes like *Access World News* and *Ethnic News Watch* all provide in fragmentary form the record of current events which used to come in a single package, the newspaper. Electronic distribution methods make it difficult to even establish the “artifacts” of news in the digital age. In the paper age: solid, durable editions were the “papers of record.” Consumers’ need for immediate and specialized news had driven publishers and broadcasters to reshape their news content and delivery mechanisms, infinitely complicating the libraries’ task of identifying and gathering news artifacts.

Promising Strategies

As news organizations shift their activities from print and broadcast to the Web arena, libraries must adopt new tactics if they are to retain the capacity to preserve the primary records of history and heritage. Two approaches to sustaining the “biodiversity” of knowledge show promise:

1) *Collective Action*: It is clear that as the media sector changes and adopts new and transformative practices, libraries will have to renegotiate their longstanding relationship with publishers and producers. Publishers and aggregators like ProQuest, Thomson Gale, and others can bring important capabilities and assets to the work of preserving and delivering knowledge resources. They have brought digital conversion, indexing, distribution, rights management, and other such activities to immense scale, achieving valuable efficiencies and economies. Libraries too possess valuable assets, not the least of which is content accumulated from centuries of discerning and informed collecting. Libraries serve different markets than commercial publishers: they look after the interests of scholars, who often produce new knowledge by using little-known or used materials.

Libraries can best deploy these assets collectively rather than individually. *Consolidation in the publishing industry must be countered with collectivism in the library sector*. This must go beyond consortium licensing, and invoke the exchange of a wider array of assets, including rights. By acting in concert through projects like ICON and the ARL-AAU Global Resources Network, libraries can negotiate with producers from a position of strength for the exchange of content, rights, and other assets. Libraries can obtain terms that stretch the boundaries of fair use and first-sale rights, and secure rights to use content that while limited may suffice for purposes of scholarship and heritage preservation. Such terms would favor the communities that libraries represent: researchers present and future and the other stakeholders in the cultural heritage of the developing world.

2. In the future libraries will have to work more closely with scholars/researchers than they have in the past. Researchers like Harry Cleaver and Sean McLaughlin will lead the way to the critical sources on the Web and elsewhere, which libraries can archive, ensuring the integrity and authenticity of important knowledge and heritage materials and access to those sources over time.

Researchers are already leading the way. Models of cooperative management of knowledge resources that have emerged in recent decades in scientific scholarly communities can be adopted by the humanities and by the heritage community. A good example are initiatives undertaken by a number of American and international academic cooperatives to develop astronomical observatories. Cooperatives like the ones that developed the Southern Observatory for Astrophysical Research (SOAR) and the Cero Tololo Inter-American Observatory, joint efforts of multiple universities, federal agencies, and other partners, have successfully used knowledge assets to barter for valuable tangible assets. Such cooperatives have been able to obtain from the governments of Chile, South Africa, and other nations building sites in geographically

remote sites that are well-positioned for viewing outer space, in return for time shares in the observatories developed on those sites.

As mediators between the knowledge market and the research community, libraries have traditionally been relied upon to behave differently than publishers, to look beyond immediate demand to future “markets.” In this role they have ensured the persistence of the record of human achievement for future historians and future consumers of heritage information and knowledge. This role for libraries is one certain, fixed point in a world of constantly changing technology and economic realities.