

## Rip Van Winkle in the Digital Era: The Future of National Libraries

Deanna B. Marcum, Associate Librarian for Library Services, Library of Congress, Washington, D.C., U.S.A.

Address to the National Library of the Netherlands, Koninklijke Bibliotheek, The Hague, The Netherlands, 25 June 2004

I feel privileged to be asked to participate in this gathering to honor Dr. Vonhoff on the occasion of his retirement from the National Library of the Netherlands. I join you in congratulating him on his distinguished career. And I thank you for giving me an opportunity to contribute to this gathering some thoughts about the future of national libraries.

Before I get to the future, however, I would like to say something about the past. I have discovered that your national library contains a wonderful collection of books about the Netherlands published in America. This, as you well know, is a collection of some three hundred works from the nineteenth and twentieth centuries given to the library by Hendrik Edelman, who was born in the Netherlands but came to live and work in the United States. The collection includes what is probably the most famous American children's book about the Netherlands—Mary Mapes Dodge's story entitled *Hans Brinker, or the Silver Skates*. It is from such books as these that so many of us in America developed stereotypes that all Hollanders wear baggy pants and cute little hats, live in windmills, and, when they aren't ice skating, make cheese and grow tulips.

Well, you do make great cheese and you do grow beautiful tulips, but most of us in America, when we grow up, learn to respect you for many other things as well, not least of all for your international outlook and your accomplishments as linguists. A friend of mine who once lived in the Netherlands said it was hard to learn Dutch because often when he tried to speak it, people would recognize his accent and respond in English.

In keeping with your discussion today of the humanities, I want to remind you of a cultural connection between us that runs far deeper than Hans Brinker and his skates. Some of the first prose fiction of lasting merit that American literature produced arose from the section of our country that was explored and settled by Dutch people in the seventeenth and eighteenth centuries. Formerly New Netherland, this is now primarily New York City and the Hudson River Valley. Today, American school children are less likely to read about Hans Brinker than they are about Rip Van Winkle, Katrina Van Tassel, Derrick Van Bummel, and Nicholas Vedder. These are characters in two short stories written in the early part of the nineteenth century by Washington Irving, a prominent writer of the time in New York City. These stories continue to hold a hallowed place in our national literature.

Some of you will recognize Rip Van Winkle as the lazy farmer who fell asleep for twenty years in the Kaatskill Mountains after drinking Dutch gin, which was given him by the ghosts of the explorer Henry Hudson and the crew of the Dutch ship called *Half Moon*, which, back in 1609, had explored the Hudson River. The other story, called the Legend of Sleepy Hollow, tells of a night-riding headless horseman, who scared an itinerant school teacher named Ichabod Crane away from a region of prosperous farmers of Dutch descent and from a girl named Katrina Van Tassel, whom he had hoped to marry. Irving jokingly attributes these stories to a fictitious scholar whom he named Diedrich Knickerbocker. In Irving's words, this character was—

*an old gentleman of New York who was very curious in the Dutch History of the province, and the manners of the descendants from its settlers. His historical researches, however, did not lie so much among books as among men; for the former are lamentably scanty on his favorite topics; whereas he found the old*

*burghers, and still more, their wives, rich in that legendary lore, so invaluable to true history. Whenever, therefore, he happened upon a genuine Dutch family, snugly shut up in its low-roofed farmhouse, under a spreading sycamore, he looked upon it as a little clasped volume of black-letter, and studied it with the zeal of a bookworm.<sup>1</sup>*

Today, thanks to a project that your library and mine are carrying out together, researchers into the history of New Netherland will have a far easier time than Irving's old Diedrich Knickerbocker did. Books on Dutch history in America are no longer "scanty." The National Library of the Netherlands and the Library of Congress in the United States have cooperatively created an Internet-accessible digital library collection entitled *The Atlantic World: America and the Netherlands*. This is a bilingual, multi-format digital collection that includes books, maps, manuscripts, and photographs, among other materials. The initial stage, completed last fall, focuses on the period between 1609 and 1664, when the Netherlands established the colony of New Netherland, located in parts of the present-day American states of New York, New Jersey, Delaware, and Connecticut. Later stages of the project will cover the role of the Dutch and Dutch-American patriots in the American War of Independence, large-scale migration from the Netherlands to the United States in the nineteenth and early twentieth century, the "Holland-Mania" that resulted from the discovery of Dutch painting and culture by American artists and scholars in the late nineteenth century, and relations between our countries during World War II and the post-war period, including a new wave of immigration after 1945.

I am grateful to you at the National Library of the Netherlands for joining with us in this project, which we at the Library of Congress have included in our Global Gateway program, an effort to establish cooperative digital libraries with national libraries from around the world. I am also grateful to the other Dutch institutions that you enlisted to contribute to the digital collection: the Netherlands National Archives, the University of Amsterdam Libraries, and the Municipal Archives of Amsterdam. Together we have created a wonderful research resource for scholars, students, and others in both our countries and around the world.

I am grateful that productive relations between our libraries continue in other areas as well. For example, the National Library of the Netherlands and the Library of Congress are working together in an international Alliance for Bibliographic Standards, under the auspices of the Conference of Directors of National Libraries within the International Federation of Library Associations. We are joined in this alliance by the national libraries of Australia, Germany, Great Britain, and Portugal. Our hope is to improve international coordination of standards and practices for bibliographic and resource control, involving, among other things, standards for metadata, persistent identifiers, and interoperability, which are all important for digital library access..

Additionally, on behalf of the Library of Congress, I have been working with Director-General Van Drimmelen of your library on plans for digital archiving. I congratulate you on the agreements that the National Library of the Netherlands has been reaching with publishers to collaborate with them in preserving scholarly material. Your library has announced agreements "to secure the long-term digital archiving of all research published by BioMed Central," to preserve the continuing output of Kluwer Academic Publishers, whose Web platform contains more than "235,000 articles from 670 journals and more than 600 e-Books covering the areas of science, technology, and medicine,"<sup>2</sup> and to archive the journals in science, technology, and medicine published by Elsevier, whose back issues you have been digitizing.<sup>3</sup> Such arrangements with digital publishers are critical for all of us who care about preserving the output of scholarship and science. On that point, let me begin the remarks you asked me to provide about the future of national libraries.

It is of course important to acknowledge from the outset that national libraries in the more than one hundred countries that have them are not all the same. They differ from country to country in nature and purpose. Peter Johan Lor, former national librarian of the National Library

of South Africa, recently provided a useful and interesting categorization of national libraries into three main types. One, he said, are the “older national libraries, such as those that evolved from European royal libraries,” which put “emphasis on the nation’s documentary heritage, the management and preservation and exploitation of rich collections of old, rare, and valuable materials” as a “service to learned scholars and researchers.” Second, he said, are national libraries that “arose in the nineteenth and twentieth centuries in response to nationalistic and modernizing movements,” which “place emphasis on the development of national infrastructure . . . to support the work of the nation’s libraries and information agencies,” basically serving other libraries. Third, he said, are “national library services,” which “arose in developing countries . . . to offer services to the general population through a network of public libraries, school libraries, and other libraries, much as a metropolitan or county public library service would in a developed country.” Particularly now, he notes, these “orientations are not mutually exclusive.”<sup>4</sup>

To this I should add that my own library is particularly idiosyncratic. Although the Library of Congress is funded primarily by national tax revenue, collects extensively in many areas, makes resources available to the public-at-large, national and international, and receives, through its Office of Copyright, all books for which U.S. copyright is requested, our *primary* client by law is the United States Congress, our national legislature, for which we provide reference services. Moreover, we are not a national deposit library, mandated to collect the nation’s recorded culture, much of which is held instead by research libraries of major universities in the United States along with the National Agricultural Library and the National Library of Medicine.

Nonetheless, we feel, as many national libraries do, a great deal of responsibility for America’s bibliographic treasures; we have contributed substantially to the development of infrastructure for American libraries generally; and now in the digital era, we are becoming a provider of collections that can be accessed through other libraries as well as by Internet-connected individuals throughout our nation and the world. Indeed, digital technology is substantially broadening the views of many national libraries, among others, of what they can and should become.

For example, your library and mine, through our partnership to create the digital collection I described earlier on relations between the Netherlands and America, have in effect become digital publishers. We have organized a body of material, made digital copies of it, and are providing it to users online—a body of material that did not exist as an aggregated entity before but was scattered among various collections. In the past, some libraries may occasionally have published printed editions of documents, but nothing on the scale of the digital collections we are now making accessible. And the more that national libraries publish digitally, the more they are going to become preferred information access points and less repositories of last resort. The digital libraries we are building take our resources to Internet users far beyond our own walls. Indeed, as digitization proceeds, national libraries will become libraries for the world.

Big questions about digital publishing confront national libraries. First, how much of our vast collections will we digitize for Internet access? Your national library’s collections now contain more than 3.3 million items, including 2.5 million books and 15,000 current periodicals. You also have many manuscripts, medieval and modern, along with a huge collection of newspapers that were published within the Netherlands and within its former colonies around the world. Much of that material is valuable for use by scholars in those lands, and in countries such as mine, as well as in your own. Ideally, national libraries would digitize everything they possess as fast as their finances, logistical capabilities, and technological infrastructures would permit.

But whether the goal is to digitize all or some, a question of priorities arises. Which parts of our collections should we digitize with resources now at hand? So far, I think, most of us have tended to digitize collections in which we could interest financial supporters, or collections that easily lent themselves to use in testing new technologies, or collections so heavily used that digitization might relieve pressure on the originals. Even then, however, we have sometimes

discovered, as the digital repository called JSTOR has learned from digitizing back issues of old journals, that digitization creates use. A lot of digitized items are getting much more use than they ever did within the physical library, and sometimes interest generated by digitization has led to greater use also of the originals.

Additionally, in selecting for digitization, we have used items that were not tied up by copyright, licensing, and other intellectual property restrictions. This is as true of the Meiji Era digital collection offered online by the National Diet Library of Japan as it is of the American Memory collection of mostly pre-twentieth century books that the Library of Congress has put online from its own and other American collections. In the United States, and I think in other countries, copyright is a major obstacle to mass digitization. But even among materials that we are free to digitize, national libraries are now beyond the stage of doing just experimental and demonstration projects. We need more systematic, far-seeing ways of deciding what to digitize next.

However, digitizing collections is not the only major consideration in the future of national libraries. Given what technology is now making possible, National Libraries of the first kind in the analysis I described by Mr. Lor—older libraries with established mandates to collect their nations' intellectual output—will find themselves having to do that in new and challenging ways. They will need to collect materials that have been created digitally by others—materials that exist only in digital form. And such materials are being created in great quantities and multiple formats. I am talking about vast quantities of data from observational research that scientists have electronically aggregated for study now and in perpetuity. I am talking about databases created by scholars working on particular projects individually or in collaboration—*e-scholarship*, as we have taken to calling it. I am talking about educational resources aggregated or created by professors for their classes as aids to study and teaching. I am talking about Web sites on which groups and individuals all over the world are mounting information about, it seems, almost every conceivable subject. Some of these sites are valuable and reliable enough to be used by others, and many of them are evidence of activities in our time that will warrant historical study.

Perhaps it is the function of the libraries of universities to capture and preserve the most valuable *e-scholarship* generated on their own campuses. The Massachusetts Institute of Technology in the United States, for example, has created a program called DSpace to do that at MIT, and is making that program available to some other universities. Perhaps learned societies and scientific laboratories will take responsibility for capturing, in forms that permit long-term access, the databases of most ongoing value for their individual fields of study. And perhaps special entities will arise to evaluate Web sites and save those that have most evidentiary and informational usefulness for the future. An organization called the Internet Archives, founded by Brewster Kahle, is trying to do that in the United States. But the Library of Congress is working with him on a Web site capturing project, and national libraries, whether or not they collect electronic resources themselves, may need to help see that somebody does.

There is, however, an additional problem, which is the need undertake digital collecting at the point of creation. If libraries are to preserve and provide future access to digital documents, they must encourage creators to produce those documents in formats that libraries can most easily preserve and make perpetually accessible, and must help creators do that. Guiding digital document creators is currently difficult, however, because libraries, themselves, have so much yet to learn about digital preservation. This brings us to a major future requirement for national libraries. No one knows better than your librarians here that digitizing material and taking custody of digital disks and tapes from others is the easy part of electronic librarianship. The hard part is keeping digital material in readable form.

First, and simplest, digital media such as magnetic tapes and optical disks, from which information is obtained by generating electronic impulses on computer servers, are far more fragile than books and journals, from which information is obtained by physically removing items

on a shelf. In relatively short times, magnetic tapes can become brittle, optical disks can become cloudy or scratched, and breakdowns can occur between the surfaces and the backing of these media. Just minor damage can make an entire document unreadable or recoverable only with great effort. Storage conditions, handling practices, and media improvements can make a difference, but experiments by the National Media Laboratory in the United States have indicated that magnetic tapes and optical disks may not last more than 30 years and some become unreliable in fewer than five.

Unfortunately, electronically recorded information is in danger of becoming unreadable even before the digital media deteriorate. This is because of the rapidity with which the computer hardware and software used to create digital files become obsolescent. Few if any of us today use the hardware platforms and software programs on which we once created digital documents. Some of the companies that produced and supported older computer systems no longer exist. And computer languages continue to change. Unmigrated data is now unreadable or can be recaptured only through considerable effort and expense. Three years went into recovering digital data from the 1960 U.S. Census after discovery that machines capable of reading the data were no longer available. And some digital records have been lost forever.

We cannot feasibly preserve digital materials by keeping all the old software and hardware on which digital resources are generated. If we did that, every repository of digital data, such as a library, would have to maintain an ever expanding computer museum, complete with replacement parts and maintenance capabilities. Instead, we seek ways to emulate or mimic old systems on new ones. Or much more commonly we transfer, or “migrate,” as technicians say, digital data to new media, new formats, and new systems.

This solution, alas, is imperfect. Frequently rewinding stored tape, periodically recopying it, separately storing backup copies, and performing other functions needed in digital migration all take time and cost money. Also, migration requires decisions about how much of the “look and feel” of a digital document to try to carry forward along with basic content, and we are far from sure that we can migrate content over long periods without significant loss.

The problem is compounded by the continuing invention of new and increasingly sophisticated formats for digital communication. Multimedia programs that rely on complicated combinations of software and hardware are particularly difficult to preserve. And digital preservation will have to deal with formats yet to be invented. Moreover, we are challenged by the growing volume of digital material to be preserved. Preservation methods must be able to handle increasing quantities of digital information as well as multiple, complex formats.

That’s not all. In preserving digital information we also face problems of authenticity and reliability. In digital files, changes can be made imperceptibly, even unintentionally. Systems in which databases are embedded may replace old data with new data automatically. And a digital document is indistinguishable from any electronic copy of it. Which is the original to be preserved? How do we ensure authenticity in digital documents we preserve?

So far there is no one, perfect solution for the problem of digital preservation. If we make progress, it will likely be through varying approaches developed incrementally. Your library is taking a significant part in this effort. I have already mentioned the agreements announced by the National Library of the Netherlands to secure the long-term digital archiving of research published by Elsevier, BioMed Central, and Kluwer Academic Publishers. And all national libraries are interested in your work with the IBM corporation to develop “a new electronic deposit system that can meet large-scale and high-quality storage requirements and support digital preservation and functionality.”<sup>5</sup>

We at the Library of Congress have received appropriations to create a National Digital Information Infrastructure and Preservation Program in our country that we hope will also be useful elsewhere. This program has developed an architectural model for federated digital preservation and a plan for preservation research and development, which we expect soon to implement by funding research proposals from institutional partners. And all of us in national

libraries are keeping track of preservation progress through a clearinghouse maintained by the National Library of Australia in an initiative called PADI, an acronym for Preserving Access to Digital Information.

Of necessity, digital preservation will be a major function of national libraries in the future. Who else can be depended on to keep the digital heritage that we are developing safe and accessible for future, ongoing access and study?

For many of us in national libraries, the largest question is, how can we afford these new roles and responsibilities? Books continue to be published in print—we receive some 800,000 a year through the U.S. Copyright Office. Other analog materials such as manuscripts, maps, photographs, and audiovisual materials continue also to be added to our collections. How can we finance the addition of digital libraries and services on top of our ongoing traditional ones? One of the major sources of impetus for digitizing and archiving scholarly journals has been the increasing cost both of buying and shelving printed versions.

I believe that part of the solution—and many other benefits—can come through collaboration among libraries within our countries and among national libraries across our borders.

In the United States, we have begun taking a step in that direction through an organization called the Digital Library Federation. Last fall, members of the federation committed themselves to creating jointly something first envisioned in 1995 at the time of the federation's founding—a collaborative digital library that will provide global electronic access to collections in multiple institutions. This Distributed Open Digital Library—or DODL, as it is called—will provide users with one point of entry to multiple, digitized collections. Moreover, the federation hopes to provide services using this distributed library that may include a deep finding system and specialized portals to help scholars locate material for their individual fields and specific studies in multiple libraries. The DODL's collection will begin with materials in the humanities and social sciences, including in special collections, and will transcend regions to concentrate on topics, themes, genres, and formats.

The Digital Library Federation is supported primarily by its members, including the Library of Congress. We are in the process of soliciting additional financing for the DODL, of appointing a coordinator for it, of forming a collections-development working group to plan development of the DODL's content, and of forming a technical working group that will create an enabling infrastructure for the collaborative library. In developing the DODL, we hope to build on a range of achievements in digital library development by federation members and others internationally over the past several years. And the DODL may become an international collection. The Digital Library Federation, which contains 34 major university research libraries and related organizations in America, recently announced that the British Library will join the federation in a strategic partnership. And possible relationships are contemplated with other libraries around the world.

Other international relationships may be possible as well. The digital resources we develop in our individual national libraries can be accessible to people throughout our countries if we work together on standards and technologies for interoperability between our digital library infrastructures. Also, within organizations such as the Council of National Library Directors, we are searching for ways to extend access to electronic information technologies and digital library resources across what has been called “the digital divide”—that is, to countries and peoples of the world who have few computers, limited infrastructure for electronic information use, and little money for its development. Until they, too, can participate, the great benefits of the new digital information technologies for human betterment around the world cannot be fully realized. And extending our digital resources around the world seems to me impossible without expanding the cooperation of the world's libraries, particularly national libraries.

Well, that is quite an agenda I have laid out for the future of national libraries. And I fear it leaves a lot of questions open. Can we successfully collect all the kinds of cultural materials of

research value that are being created in digital forms? Can we learn to preserve digital materials for study by scholars, students, and others indefinitely into the future? Can we extend access to electronic information to have-nots as well as haves in all of our countries? And can we adequately finance efforts to accomplish these extremely important endeavors? I believe that, somehow, we must do so, if we are to realize the full promise of the digital era.

In this era, I feel a bit as Henry Hudson and his Dutch crew must have felt nearly 300 years ago when their little ship, the *Half Moon*, brought them to behold great wonders in what was then the New World. In today's new world for libraries, I also feel a bit as Rip Van Winkle felt when, in Washington Irving's story, he awoke from his twenty-year sleep to find everything around him bewilderingly changed. Technology is changing libraries rapidly now. It may be hard for those of us who have treasured physical books and manuscripts containing the wisdom of the ages and the fruits of scholarship to realize that humanity's cultural and scientific achievements in the future may be recorded in electronic bits and bytes. But so they will, increasingly. And we national libraries have a particular responsibility to preserve the part of the world's intellectual heritage that is being created in that form, too. I look forward to continuing to work with you on that great challenge to our national libraries for the future.

Thank you very much.

---

<sup>1</sup> Washington Irving, "Rip Van Winkle," *The Sketch Book* [there are many editions].

<sup>2</sup> <http://www.kb.nl/kb/pr/pers/pers2003/kb-kap-en.html>; <http://www.kb.nl/kb/pr/pers/pers2003/kb-biomed-en.html>, as of 27 May 2004.

<sup>3</sup> Miriam A. Drake, "Elsevier Teams with Dutch National Library for Digitization Project," <http://www.infoday.com/newsbreaks/nb020903-2.htm>, available as of 6/8/04.

<sup>4</sup> Peter Johan Lor, "National Libraries and the Digital Divide," paper presented to the Conference of Directors of National Libraries in 2003, available as of 6/1/04 at <http://nla.gov.au/initiatives/meetings/cdnl/2003>.

<sup>5</sup> Drake, "Elsevier Teams with Dutch National Library."