Important Issues in Preservation and Access of Digital Content for Children

Howard Besser
Moving Image Archiving & Preservation Program, New York University, New York, USA
Department or Division Name, Institution Name (no postal address), City, Country.
http://besser.tsoa.nyu.edu/howard/

Abstract:

Though digital distribution offers new and interesting ways to provide access to works for children, content in digital form poses challenges to our traditional approaches to conservation. This paper will first review some of the more interesting and challenging projects (such as the International Children's Digital Library and Internet Archive Children's Library), and discuss their innovative approaches to distribution and engagement around issues of childrens' literature. It will also discuss the standards for metadata and scanning that are necessary for building interoperable digital childrens' libraries.

The rest of the paper will focus on conservation-related issues. The author will explain why some principles of analog conservation (like focusing on the "original work") must be re-thought for conserving works in digital form. He will discuss international best practices for digital preservation (such as the OAIS reference model and the PREMIS metadata data dictionary). And he will explain why various projects he has worked on (InterPARES, Preserving Digital Public Television, Activist Archivists) have recommended that people involved with digital preservation need to be involved early in the life-cycle of a work in order to improve the chance that it will be preservable. He will emphasize that conservators need to collaborate with authors early in the process of creation.

The paper will discuss the recommendations and practices that have been developed for digital preservation that were created for libraries (and other organizations) that already have strong technological infrastructures (in terms of equipment, technologically trained staff, and monetary resources).

Keywords: Children, Digital Libraries, Preservation, Access

1   International Children's Digital Library

The International Children's Digital Library[1] (ICDL) is a freely accessible online collection of over 4,500 digitized children's books written in more than 60 different languages from more than 200 countries. Its target audience is children between 3 and 13 years old. All books are presented as published in their original language, complete with illustrations. It was originally launched in 2002 by the University of Maryland (USA) as a library science research project to better understand children's online habits[2] as well as to encourage a love of reading. In its early years, the Internet Archive's Childrens' Library[3] was an important collaborator. ICDL is now run by the International Children's Digital Library Foundation (ICDLF).
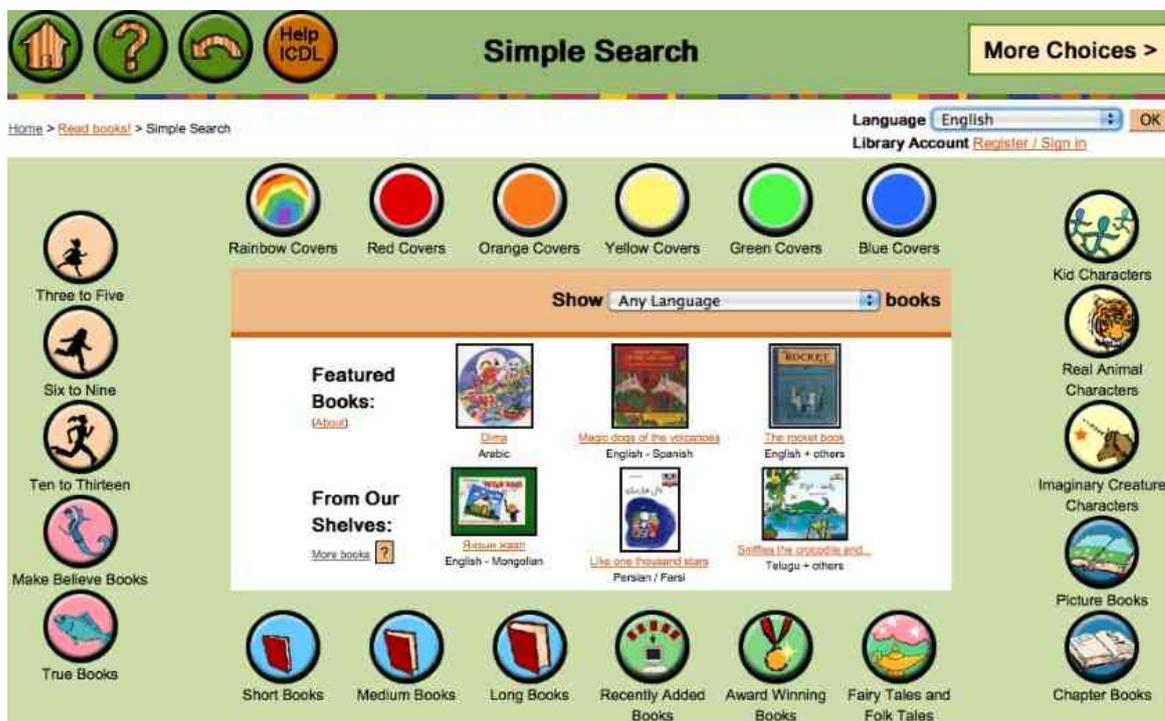

figure #1: ICDL simple search interface in English[4]

User interfaces are available in more than a dozen languages, and are designed to be simple for children to search for books by color, book length, intended age group, fiction or non-fiction, type of book, language of book (which may be different than the language of the user interface), etc. (see figure #1). No e-Reader software is needed; the books are not downloaded, but are instead read through a standard Internet browser such as Firefox or Internet Explorer. But this means that the reader needs to maintain an Internet connection throughout the reading experience. And the copyright holders have given the ICDL

---

[1] http://en.childrenslibrary.org/

[2] Allison Druin, Benjamin B. Bederson, Ann Weeks, Allison Farber, Jesse Grosjean, Mona Leigh Guha, Juan Pablo Hourcade, Juhyun Lee, Sabrina Liao, Kara Reuter, Anne Rose, Yoshifumi Takayama, and Lingling Zhang, *The International Children's Digital Library: Description and analysis of first use,* **First Monday 8**:5, May 5, 2003, http://firstmonday.org/ojs/index.php/fm/article/view/1050

[3] http://archive.org/details/iacl

[4] http://www.childrenslibrary.org/icdl/SimpleSearchCategory?ilang=English

permission to scan and present all books that are part of the collection. Books cannot be downloaded, copied, or printed.

The ICDLF's mission was outlined several years ago on the website by their Executive Director Tim Browne:

> "the mission of the International Children's Digital Library Foundation is to prepare children for life in an ethnically and culturally diverse world by building the world's largest online multicultural repository of children's literature. We have set out to change the world, book by book, byte by byte. We offer free access to exemplary works from more than 42 countries. Our online library can be accessed for free in 11 languages with innovative software that was developed by hearing from young people about their needs, interests, and capacities. "[5]

This international cross-cultural approach to childrens' literature is mentioned repeatedly throughout their website: "Ultimately, the Foundation aspires to have every culture and language represented so that every child can know and appreciate the riches of children's literature from the world community."[6] And their stated Goals emphasize this:

> "To create a collection of more than 10,000 books in at least 100 languages that is freely available to children, teachers, librarians, parents, and scholars throughout the world via the Internet. The materials included in the collection reflect similarities and differences in cultures, societies, interests, lifestyles, and priorities of peoples around the world. The collection's focus is on identifying materials that help children to understand the world around them and the global society in which they live. It is hoped that through a greater understanding of one another that tolerance and acceptance can be achieved."[7]

A secondary goal is for the children of immigrants to gain exposure to their cultural heritage by reading the literature from their parents' culture. "A fundamental principle of the Foundation is that children and their families deserve to have access to the books of their culture, as well as the majority culture, regardless of where they live."[8]

The ICDL operates with over a thousand volunteers around the world.[9] They identify books for the collection; they secure rights; they package and send the books either physically or digitally; they tell others about the collection and recruit new users. Many librarians are involved in ICDL collection development activities by selecting quality literature, as ICDL's "goal is to build a collection of books that represents outstanding historical and contemporary books from throughout the world."[10]

Because project co-founder Alison Druin specializes in user interfaces for children, the project was designed to figure out the best types of user interfaces to support children in searching, browsing, reading, and sharing of digitized books. Sets of children become "design partners" in the development of optimum user interfaces.

---

[5] http://en.childrenslibrary.org/about/letter.shtml
[6] http://en.childrenslibrary.org/about/mission.shtml
[7] http://en.childrenslibrary.org/about/goals.shtml
[8] http://en.childrenslibrary.org/about/mission.shtml
[9] http://en.childrenslibrary.org/contribute/ambassadors.shtml
[10] http://en.childrenslibrary.org/about/mission.shtml

The ICDL takes advantage of its aggregation of digital books to offer other types of interesting services. ICDL Exhibitions[11] presents "collections of books from around the world with similar themes. The exhibitions include an overview of the theme, links to the exhibition books, descriptions of why the books were included, and related activities."[12] Exhibition themes include strong women, overcoming difficulties, water as a natural resource, how seasons change in different parts of the world, cultural differences, and several others.

The website also promotes activities that can be organized by librarians or teachers in conjunction with reading books from the site.[13] Activities include scavenger hunts, creative writing, digital storytime, and learning languages. They also have produced a training manual outlining the sorts of activities that teachers might use in conjunction with ICDL.[14]

ICDL also supports a Community Forum[15] where educators, librarians, parents, and others can discuss common issues. Postings have included: suggestions of activities to accompany the reading of a particular book, discussion of resources on bibliotherapy, and requests for identification of a book when someone only remembers fragments of the plot. The Community Forum also serves as a place for discussion of common problems with using the application.

The ICDL also has developed innovative tools such as StoryKit,[16] a free iPhone/iPad App for children to create and illustrate their own electronic storybooks, and share these stories with others. Kindergarten and first grade teachers have used this to help their students learn to compose.

2   ICDL and the Internet Archive's Children's Library

Though they were close collaborators in the early part of the millennium, ICDL and the Internet Archive's Children's Library (IACL) have diverged in both their scope and purpose. In the early years IACL acted as a repository for high-quality digitized scans, and took a major responsibility for metadata development, while ICDL primarily focused on developing a children-friendly user interface and presenting the digitized works within a normal web browser.

IACL, as part of the Internet Archive, has always been concerned with providing to users the highest quality scans. This has led them to prefer downloadable PDFs or high-quality web-based readers that allow for zooming in at very high resolutions. ICDL, on the other hand, has been most interested in building interfaces for children's navigation through content, and on building as large a corpus of digitized material as possible.

This writer speculates that most of the holders of copyright for children's literature were unwilling to permit ICDL to deliver very high quality scans of their works for free on the Internet, so in order to build a large corpus of material (almost half of which is still within copyright), ICDL needed to agree that they would not allow downloads, nor would they permit zooming into pages at very high qualities.

---

[11] http://www.childrenslibrary.org/icdl/ExhibitionList?ilang=English
[12] ibid.
[13] http://en.childrenslibrary.org/books/activities/
[14] http://en.childrenslibrary.org/books/activities/ICDL%20Teacher%20Training%20Manual.pdf
[15] https://getsatisfaction.com/internationalchildrensdigitallibrary
[16] https://itunes.apple.com/us/app/storykit/id329374595

As a result, the scope covered by ICDL is broader than that of IACL because ICDL includes many works still within copyright. And the user interface for ICDL is optimized for the use of children. On the other hand, ICDL requires users to maintain a constant online connection in order to read a book, while IACL permits downloads, so a reader can continue to read a book off-line. IACL has a user interface more designed for adults (see figure #2), and because its corpus does not contain so many of the more contemporary works (because of copyright issues), it is likely used much more by scholars researching older children's literature (while ICDL is likely primarily used by children).

figure 2: Internet Archive's Children's Library[17]

Another major difference of the two approaches is in the implications for preservation. ICDL sees itself as an access service, and doesn't assume any responsibility for long-term archiving or preservation. It likely presumes that the paper copies of all the material in its corpus will be collectively saved by analog libraries across the world. IACL, on the other hand, has the word "Archive" in its name, and assumes serious responsibility for long-term stewardship of its digital content. And, from the standpoint of digital preservation, the ICDL collection would pose serious problems both because the scanned images (at least the ones publicly available) are not high resolution, and because the user interface would not allow a library's archiving web crawler to easily ingest an entire book because viewing a book requires stepping though it one or two pages at a time. In the next section we will turn to issues in digital preservation.

3   Digital Preservation Issues

Digital information is fragile, and requires particular care and handling to preserve it over time. Though novices often characterize the problem as the deterioration of storage media (such as hard disks), professionals working in digital preservation tend to view the problem more as one involving file format changes, appropriate metadata, and the

---

[17] http://archive.org/details/iacl

implementation of policy and procedures to support an environment that will manage digital works over a prolonged period of time. For a more complete explanation of the problems of digital longevity, see (Besser 2000).

In the analog world, conservation and preservation focused on keeping an original artifact in good shape through temperature and humidity controls, and through repair of the object. In the digital world, we need to break away from the notion of the "original object" because we can make a digital copy that is absolutely identical to the digital work that is being copied. And because our file formats become obsolete quite rapidly (think of trying to read a word processing file you created just 15 years ago), we need to manage format changes for the digital files we want to preserve through the process of *reformatting*. This creates a conceptual shift: from preserving a physical object, to managing the informational content that can be completely disembodied from any physical object.

The digital preservation world has defined two concepts that we use in reformatting. *Refreshing* is designed to respond to the problem of physical deterioration, and *Migrating* is designed to respond to the problem of obsolete formats. *Refreshing* means copying to a new support without changing anything; *Migrating* means changing the actual format when you copy the work. So, in a digital text world, *refreshing* would mean taking a file on a 5.25 inch floppy disk and first moving it to a 3.5 inch diskette, then later to a CD-ROM, then later to a DVD. *Migrating* would mean reformatting a file in Word for Window '95 and re-encoding it for Word 2003, then later re-encoding it for Word 2011.

In order to manage reformatting and other processes necessary to preserve a digital work, most digital libraries have adopted the OAIS conceptual model as a way of understanding how to manage a digital work over time. The OAIS model helps to separate into three discussions: what preparatory work needs to happen in order to ingest a set of digital works into a repository (SIPs), what kinds of transformation can be done within the repository (AIPs), and what needs to happen when getting a digital work out of the repository (DIPs).

Metadata is crucial to managing a digital work over time. We need to know more than merely the standard descriptive metadata that tells us what the work is and who authored it. We also need to know technical and administrative information about how the work was originally created (what file format and encoding schemes were used, whether the work is composed of multiple files and in what order, and for scanned works: what kind of scanner and at what resolution). And we need to track issues such as fixity and integrity in a work over time to make sure that the file has not been altered or corrupted. The most widely adopted metadata standards for preserving digital content are both administered by the Library of Congress: The Metdata Encoding and Transmission Standard (METS)[18] and the PREservation Metadata Implementation Strategies (PREMIS).[19]

To maximize our chances of preserving the work over time, we need to ingest the highest quality of content in the most standard forms possible. For born-digital textual works, this usually means saving original XML coding as part of the text. For scanned works this usually means digitizing at the highest resolution possible and storing in standard PDF files (ideally PDF-A). (This is why the approach of IACL is more preservation-friendly than the

---

[18] http://www.loc.gov/standards/mets/
[19] http://www.loc.gov/standards/premis/

approach of ICDL.) And this always means trying to obtain the highest quality metadata in the most standard format.

Key digital preservation projects have recommended that librarians get involved when digital content is being created in order to help assure that the content creators will be able to provide our digital repositories with high enough content quality and with standardized metadata. The international project on born-digital archives (InterPARES 2) has stressed that archivists need to be involved early in the life-cycle of a digital object instead of just receiving it near the end of its life (Duranti and Preston 2008). And the Preserving Digital Public Television project (USA) has show that preservation is made easier by getting librarians involved in designing the creation phase of a news-gathering program (Besser and van Malssen 2007). This implies that librarians should get involved with the authors of born-digital children's books when they are first starting to write, and should push the author to create the work using standard formats and metadata.

4   Summary

The ICDL offers a wonderful set of tools that allow children to have easy access to a large number of books from different language groups, different countries, and different cultures. It is a very useful system designed for access, but it was not designed for preservation. The IACL, on the other hand, was designed for preservation, but lacks the kind of children-friendly user interface that is an integral part of the ICDL. It is likely that copyright issues have led to these two different approaches. Copyright issues may indeed force future children's digital library projects to make choices between favoring preservation versus favoring access. Proper digital preservation is a difficult task, but digital works can be made more preservable if librarians or archivists get involved with creators early in the creation process.

5   References

Besser, Howard. 2000. "Digital Longevity." In ***Handbook for Digital Projects: A Management Tool for Preservation and Access****,* pages 155-166. edited by M. K. Sitts. Andover Mass: Northeast Document Conservation Center (http://www.nedcc.org/assets/media/documents/dman.pdf)

Besser, Howard and Kara van Malssen. 2007. "Pushing Metadata Capture Upstream into the Content Production Process: Preliminary Studies of Public Television". DigCCurr 2007: An International Symposium in Digital Curation, 18-20 April. Chapel Hill, NC (http://www.ils.unc.edu/digccurr2007/program.html)

Duranti, Luciana, and Randy Preston, eds. 2008. International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records. Padova, Italy: Associazione Nazionale Archivistica Italiana (http://www.interpares.org/ip2/book.cfm)