The International Children’s Digital Library: A Case Study in Designing for a Multilingual, Multicultural, Multigenerational Audience

Hilary Browne Hutchinson, Anne Rose, Benjamin B. Bederson, Ann Carlson Weeks, and Allison Druin

The challenges encountered in building the International Children’s Digital Library (ICDL), a freely available online library of children’s literature are described. These challenges include selecting and processing books from different countries, handling and presenting multiple languages simultaneously, and addressing cultural differences. Unlike other digital libraries that present content from one or a few languages and cultures, and focus on either adult or child audiences, ICDL must serve a multilingual, multicultural, multigenerational audience. The research is presented as a case study for addressing these design criteria; current solutions and plans for future work are described.

The Internet is a multilingual, multicultural, multigenerational environment. While once the domain of English-speaking, Western, adult males, the demographics of the Internet have changed remarkably over the last decade. As of March 2004, English was the native language of only 35 percent of the total world online population. As of March 2004, Asia, Europe, and North America each make up roughly 30 percent of Internet usage worldwide. In the United States, women and men now use the Internet in approximately equal numbers, and children and teenagers use the Internet more than any other age group.

Creators of online digital libraries have recognized the benefit of making their content available to users around the world, not only for the obvious benefits of broader dissemination of information and cultural awareness, but also as tools for empowerment and strengthening community. Creating digital libraries for children has also become a popular research topic as more children access the Internet. The International Children’s Digital Library (ICDL) project seeks to combine these areas of research to address the needs of both international and intergenerational users.

Background and Related Work

Creating international software is a complex process involving two steps: internationalization, where the core functionality of the software is separated from localized interface details, and localization, where the interface is customized for a particular audience. The localization step is not simply a matter of language translation, but involves technical, national, and cultural aspects of the software. Technical details such as different operating systems, fonts, and file formats must be accommodated. National differences in language, punctuation, number formats, and text direction must be handled properly. Finally, and perhaps most challenging, cultural differences must be addressed.

Hofstede defines culture as “the collective mental programming of the mind which distinguishes the members of one group or category of people from another.” These groups might be defined by national, regional, ethnic, religious, gender, generation, social class, or occupation differences. By age ten, most children have learned the value system of their culture, and it is very difficult to change. Hofstede breaks culture into four components: values, rituals, heroes, and symbols. These components manifest themselves everywhere in software interfaces, from acceptable iconic representations of people, animals, and religious symbols to suitable colors, phrases, jokes, and scientific theories. However, as Hofst notes, culture is like an iceberg; only 10 percent of the characteristics of a culture are visible on the surface. The rest are subjective, unspoken, and unconscious. It is only by evaluating an interface with users from the target culture that designers can understand if their software is acceptable.

Developers of online digital libraries have had to contend with international audiences for many years, and the MARC and OCLC systems have reflected this concern by including capabilities for transliteration and dia-critical characters (accents) in various languages. However, it is only more recently, with the development of international character-set standards and Web browsers that recognize these standards, that truly international digital libraries have emerged. Greenstone, an
open-source software project based in New Zealand, allows people to create online digital libraries in their native language and culture. OCLC recently completed a redesign of FirstSearch, a Web-based bibliographic and full-text retrieval service, to accommodate users with different software, languages, and disabilities. Researchers at Virginia Tech redesigned CITIDEL, an online collection of computer-science technical reports, to create an online community that allows users to translate their interface into different languages.

Researchers have also realized that beyond accessibility, digital libraries have enormous potential for empowerment and building community, especially in developing countries. Witten et al. and Downie describe the importance of community involvement when creating a digital library for a particular culture, both to empower users and to make sure the culture is accurately reflected. Even more than accurately reflecting a culture, a digital library also needs to be understood by the culture. Duncker notes that a digital-library interface metaphor based on a traditional physical library was incomprehensible to the Maori culture in New Zealand, who are not familiar with the conventions of Western libraries.

In addition to international libraries, a number of researchers have focused on creating digital libraries for children. Recognizing that children have difficulty with spelling, reading, and typing, as well as traditional categorization methods such as the Dewey Decimal System, a number of researchers have created more child-friendly digital libraries. Pejtersen created the BookHouse interface with a metaphor of rooms in a house to support different types of searching. Külper et al. designed the Bücherschatz interface for children who are eight to ten years old using a treasure-hunt metaphor. Druin et al. designed the QueryKids interface for young children to find information about animals. Theng et al. used the Greensite software to create an environment for older children to write and share stories.

The ICDL project seeks to build on and combine research in both international and children’s digital libraries. As a result, ICDL is more ambitious than other digital library projects in a number of respects. First, it is designed for a broader audience. While the digital libraries already described target one or a few cultures or languages, ICDL’s audience includes potentially every culture and language in the world. Second, the content is not localized. Part of the library’s goal is to expose users to books from different cultures, so it would be counterproductive to present books only in a user’s native language. As a result, the interface not only supports multiple languages and cultures, but it also supports them simultaneously, frequently on the same screen. Third, ICDL’s audience not only includes a broad group of adults from around the world, but also children from three to thirteen years of age.

To address these challenges, a multidisciplinary, multilingual, multicultural, and multigenerational team was created, and the development was divided into several stages. In the first stage, completed in November 2002, a Java-based, English-only version of the library was created that addressed the searching and reading needs of children. In the second stage, completed in May 2003, an HTML version of the software was developed that addressed the needs of users with minimal technology. In the third stage, completed in May 2004, the metadata for the books in the library were translated into their native languages, allowing users to view these metadata in the language of their choice. The final stage, currently in progress, involves translating the interface to different languages and adjusting some of the visual design of the interface according to the cultural norms of the associated language being presented. In this paper, the research is presented as a case study, describing the solutions implemented to address some of these challenges and plans for addressing ongoing ones.

ICDL Project Description

The ICDL project was initiated in 2002 by the University of Maryland and the Internet Archive with funding from the National Science Foundation (NSF) and the Institute for Museum and Library Services (IMLS). Today, the projects continues at the University of Maryland. The goals of the project include:

- creating a collection of ten thousand children’s books in one hundred languages;
- collaborating with children as design partners to develop new interfaces for searching, browsing, reading, and sharing books in the library; and
- evaluating the impact of access to multicultural materials on children, schools, and libraries.

The project has two main audiences: children three to thirteen years of age and the adults who work with them, as well as international scholars who study children’s literature. The project draws together a multidisciplinary team of researchers from computer science, library science, education, and art backgrounds. The research team is also multigenerational—team members include children seven to eleven years of age, who work with the adult members of the team twice a week during the school year and for two weeks during the summer to help design and evaluate software. Using the methods of cooperative inquiry, including brainstorming, low-tech prototyping, and observational note taking, the team has researched, designed, and built the library’s category structure, collection goals, and searching and reading interfaces.
The research team is also multilingual and multicultural. Adult team members are native or fluent speakers of a number of languages besides English, and are working with school children and their teachers and librarians in the United States, New Zealand, Honduras, and Germany to study how different cultures use both physical and digital libraries. The team is also working with children and their teachers in the United States, Hungary, and Argentina to understand how children who speak different languages can communicate and learn about each other’s cultures through sharing books. Finally, an advisory board of librarians from around the world advises the team on curatorial and cultural issues, and numerous volunteers translate book and Web-site information.

ICDL Interface Description

ICDL has four search tools for accessing the current collection of approximately five hundred books in thirty languages: Simple, Advanced, Location, and Keyword. All are implemented with Java Servlet technology, use only HTML and JavaScript on the client side, and can run on a 56K modem. These interfaces were created during the first two development phases. The team visited physical libraries to observe children looking for books, developed a category hierarchy of kid-friendly terms based on these findings, and designed different tools for reading books.

Using the Simple interface (figure 1), users can search for books using colorful buttons representing the most popular search categories. The Advanced interface (figure 2), allows users to search for books in a compact, text-link-based interface that contains the entire library-category hierarchy. By selecting the Location interface (figure 3), users can search for books by spinning a globe to select a continent. Finally, with the Keyword interface, users search for books by typing in a keyword. Younger children seem to prefer the simplicity and fun of the Location interface, while older children enjoy browsing the kid-friendly categories, such as Colors, Feelings, and Shapes.

All of these methods search the library for books with matching metadata. Users can then read the book using a variety of book readers, including standard HTML pages and more elaborate Java-based tools developed by the ICDL team that present book pages in comic or spiral layouts (figures 4–6). In addition to the public interface, ICDL also includes a private Web site that was developed for book contributors to enter bibliographic metadata about the books they provide to the library (figures 7 and 8). Using the metadata interface, contributors can enter information about their books in the native language of the book, and optionally translate or transliterate this information into English or Latin-based characters.

The design of ICDL is driven by its audience, which includes users, contributors, and volunteers of all ages from around the world—more than six hundred thousand unique visitors from more than two hundred countries (at last count). As a result, books written in many different languages for users of different ages and cultural backgrounds must be collected, processed, stored, and presented. The rest of this paper will describe some of the challenges encountered and that are still being encountered in the development process, including selecting and processing a more diverse collection of books, handling different character sets and fonts, and addressing differences in cultural, religious, social, and political interpretation.
Book Selection and Processing

The first challenge in the ICDL project is obtaining and managing content. Collecting books from around the world is a challenge because national libraries, publishers, and creators (authors and illustrators) all have different rules regarding copyrights. The goal is to identify and obtain award-winning children’s books from around the world, for example, books on the White Ravens list, which are also made available to ICDL users (www.icdlbooks.org/servlet/WhiteRavens). However, unsolicited books are received, frequently in languages the team cannot read. As a result, members of the advisory board and various children’s literature organizations in different countries are relied on to review these books. These groups help determine whether books are relevant and acceptable in the culture they are from, and whether they are appropriate for the three-to-thirteen age group. These groups are eager to help; including them in the process is an effective way to build the project and the community surrounding it.

In addition to collecting and scanning books, bibliographical metadata in the native language of the book (title, creator[s], publisher, abstract) are also collected via the Web-based metadata form filled out by the book contributors. It was decided to base the ICDL metadata specification on the Dublin Core because of its international background, ability to be understood by nonspecialists, and the possibilities to extend its basic elements to meet ICDL’s specific needs (see www.icdlbooks.org/metadata/specification for more details). Contributors who provide metadata have the option of translating them to English; they also can transliterate them to Latin characters, if necessary. Regardless of what language or
languages they provide, they are asked to provide information that they create themselves, such as the abstract, in a format that is easily understandable by children. Simple, short sentences make the information easy for children to read, and easier to translate to other languages.

The metadata provided allow the team to catalog the books for browsing according to the various categories and to index the books for keyword searching. Even though translation to English is optional, the English-speaking metadata team needs the metadata in English in order to catalog the books. Since many contributors do not have the time or ability to provide all of this information, volunteers who speak different languages are relied on to check the metadata that get submitted, and translate or transliterate them as necessary. This method allows information to be collected from contributors without overwhelming them, and also helps build and maintain the volunteer community.

### Handling Different Character Sets

The metadata form allows contributors to provide information from the comfort of an operating system and keyboard in their native language, but this flexibility requires software that can handle many different character sets. For example, English uses a Latin character set; Russian uses a Cyrillic character set; and an Arabic character set is used for Persian/Farsi. Fortunately, there exists a single character set called Unicode, an international, cross-platform standard that contains a unique encoding for nearly every character in every language. Unfortunately, not all software supports Unicode as yet. In the first stage of implementation in ICDL, metadata information was collected only in English, so Unicode compliance was not a problem. However, in the next phase of development, which included collecting and presenting metadata in the native language of all of the books, the software had to be adjusted to use Unicode because ICDL supports potentially every language in the world.

The open-source MySQL database, recently upgraded to allow storage of Unicode data, was already in use for storing metadata. ICDL’s Web applications run on Apache HTTP and Tomcat Web servers, both of which are freely available and Unicode-compliant. However, both the Web site and the database had to be internationalized and localized to separate the template for metadata presentation from the content in different languages. A Unicode-compliant database driver was necessary for passing information between the database and the Web site. Both the public and metadata Web-site applications are written using freely available Java Servlet technology. The Java language is Unicode-compliant, but some adjustments had to be made to ICDL’s servlet code to force it to handle data using Unicode.

To allow users to conduct keyword searches for books in the public interface, Apache’s freely available Lucene search engine is used to create indices of book metadata, which can then be searched. Lucene is Unicode-compliant, but a separate index for each language had to be created, requiring users to select a search language. This requirement was necessary for two reasons: (1) to avoid confusion over the same words with different meanings (bra means good in Swedish); and (2) different languages have different rules for stopwords to ignore (the, of, a in English), truncation of similar words (cats has the same root as cat in English), and separation of characters (Chinese does not put white space between symbols). Lucene has text analyzers for a variety of languages that support these different conventions. For languages that
Lucene does not support, ICDL volunteers translated English stopwords, and simple text analyzers were created by the team.

Finally, HTML headers created by the Java servlets had to be modified to indicate that the content being delivered to users was in Unicode. Most current browsers and operating systems recognize and handle Web pages properly delivered in Unicode. For those that do not, help pages were created that explain how to configure common browsers to use Unicode, and how to upgrade older browsers that do not support Unicode.

By making the ICDL systems fully Unicode-compliant, contributors from all over the world can enter metadata about books in an easily accessible HTML form using their native languages, and the characters are properly transmitted and stored in the ICDL database. Volunteers can then use the same form to translate or transliterate the metadata as necessary. Finally, this information can be presented to our users when they look at books. For example the book *Where’s the Bear?* (Harris, 1997) is written in six different languages. The original metadata came in English, but ICDL volunteers translated them to Italian, Japanese, French, Spanish, and German. Users looking at the preview page for this book in the library have the opportunity to change the display language of the book to any one of these languages using a pull-down menu (figures 9 and 10).

Currently, only the book metadata language can be changed, but in the next stage of development, all of the surrounding interface text (navigation, labels) will be translated to different languages as well. The plan for doing this is to take a similar approach to the CITIDEL and Greenstone projects by creating a Web site where volunteers can translate words and phrases from the ICDL interface into their native language. Like the creators of CITIDEL, the team believes that machine-based translation would not provide good enough results. Unfortunately, the resources do not exist for the team to do the translating themselves. Encouraging volunteers to translate the site will help enlarge and enrich the ICDL community. For languages that do not receive volunteer translation, translation services are an affordable alternative.

### Character-Set Complications

Several issues have arisen as a result of collecting multilingual metadata in many character sets. First, different countries use different formats for dates and times, so contributors are allowed to specify the calendar used when they enter date information (Muslim or Julian). Second, not only do different countries use different formats for numbers, the numbers themselves are also different. For example, the Arabic numbers for 1, 2, 3 are ١, ٢, ٣. Even though Java is Unicode-compliant, it treats numbers as Latin characters, necessitating the storing of Latin versions of any non-Latin numbers used internally by the software for calculations, such as bookpage count.

A third issue is that some of the metadata, such as author and illustrator names, need to be transliterated so their values can be displayed when the metadata are shown in a Latin-based language. Ideally, the transliteration standards used for a language need to be consistent so that the same values are always transliterated the same way. Unfortunately, the team has found no practical way to enforce this, except to state the standard to be used in ICDL metadata specification. When different standards are used, it makes comparison of equal items much more difficult. For example, the same Persian/Farsi creator has been
transliterated as both “Hormoz Riyaahi” and “Hormoz Riahi.” It cannot be assumed that a person is the same just because the name is the same (John Smith), and when a name is in a character set that the team cannot understand, this problem becomes more challenging.

Finally, there was the question of how to handle differences in character-set length and direction in the interface. Different languages use different numbers of characters to present the same text. ICDL screens had to be designed in such a way that the metadata in languages with longer or shorter representations than the English version would still fit. The team anticipates having to make additional interface changes to accommodate longer labels and navigational aids when the remainder of the interface is translated.

The fact also had to be considered that, while most languages are read left to right, a few (Arabic and Hebrew) are read right to left. As a result, screens were designed so that book metadata were reasonably presented in either direction. Currently, only the text is displayed right to left, but eventually the goal is to mirror the entire interface to be oriented right to left when content is shown in right-to-left languages. For the problem of how to handle the arrows for turning pages in right-to-left languages—since these arrows could be interpreted as either “previous” and “next” or “left” and “right”—“previous” and “next” were chosen for consistency, so they work the same way in left-to-right books and right-to-left books.

Font Complications

While most current browsers and operating systems recognize Unicode characters, whether or not the characters are displayed properly depends on whether users have appropriate fonts installed on their computers. For instance, a user looking at Where’s the Bear? and choosing to display the metadata in Japanese will see the Japanese metadata only if the computer has a font installed that includes Japanese characters. Otherwise, depending on the browser and operating system, he may see question marks, square boxes, or nothing at all instead of the Japanese characters.

The good news is that many users will never face this problem. The interface for ICDL is presented in English (until it is translated to other languages). Since most operating systems come with fonts that can display English characters, the team has metadata in English (always presented first by default) for nearly all the books. Users who choose to display book metadata in another language are likely to do so because they actually can read that language, and therefore are likely to have fonts installed for displaying that language. Furthermore, many commonly used software packages, such as Microsoft Office, come with fonts for many languages. As a result, many users will have fonts installed for more languages than just those required for the native language of their operating system.

Of course, fonts will still be a problem for other users, such as those with new computers that have not yet been configured with different fonts or those using a public machine at a library. These users will need to install fonts so they can view book metadata, and eventually the entire interface, in other languages. To assist these users, help pages have been created to assist users with the process of installing a font on various operating systems.

Issues of Interpretation

While technical issues have been a major challenge for ICDL, a number of nontechnical issues relating to interpretation have also been encountered. First, until the interface has been translated into different languages, visual icons are crucial for communicating information to young children who cannot read, and to users who do not speak English. However, certain pictorial representations may not be understood by all cultures, or worse, may offend some cultures. For example, one icon showing a boy sticking out his tongue had to be redesigned when it was learned this was offensive in the Chinese culture. The team has also redesigned other icons, such as those using stars as the rating system for popular books. The original icons used five-sided stars, which are religiously significant, so they were changed to more neutral seven- or eight-sided stars.

As the team continues to internationalize the interface, there will likely be a need to change other icons that are difficult to represent in a culturally neutral way when the interface is displayed in different languages. For instance, it is a real challenge to create icons for categories such as Mythology or Super Heroes, since the symbols and stories for these concepts differ by culture. Icons for such categories as Funny, Happy, and Sad are also complicated because certain common American facial and hand representations have different, sometimes offensive, meanings in different cultures. What is considered funny in one culture (a clown) may not be understood well by another culture. Different versions of such icons may have to be created, depending on the language and cultural preferences of users. The team relies on its multicultural members, volunteers, and advisory board to highlight these concerns.

Religious, social, and political problems of interpretation have also been encountered. ICDL’s collection develops unevenly as relationships are built with various publishers and libraries. As a result, there are currently many Arabic books and only a few Hebrew
books; this has generated multiple e-mails from users concerned that ICDL is taking a political stance on the Arab-Israeli conflict. To address this concern, the team is currently working to develop a more balanced collection. Many books published in Hong Kong are received from contributors in either Hong Kong or China who want their own country to be credited with publication. To address this concern, it was decided to credit the publication country as “Hong Kong/China” to avoid offending either party.

Finally, some books have been received with potentially objectionable content. Some of these are historical books involving presentation of content that is now considered derogatory. Some include subject matter that may be deemed appropriate by some cultures but not by others. Some include information that may be too sophisticated for children three to thirteen years of age in any culture. While careful not to include books that are inappropriate for children in this age group, the team does not want to censor books whose content is subjectively offensive. Instead, such contributors are consulted to make sure they were aware of ICDL collection-development guidelines. If they believe that a book is historically or culturally appropriate, the book is included. A statement is also provided at the bottom of all the book pages indicating that the books in the library come from diverse cultures and historical periods and may not be appropriate for all users of the library.

Conclusions and Lessons Learned

Designing a digital library for an international, intergenerational audience is a challenging process, but it is hugely rewarding. The team is continually amazed with feedback from users all over the world expressing thanks that books are made available from their countries, from teachers who use the library as a resource for lesson planning, from parents who have discovered a new way to read with their children, and from children who are thrilled to discover new favorite books that they cannot get in their local library. Thus, the first recommendation the team can make based on experience is that creating international digital-library resources for children is a rich and rewarding area of research that others should continue to explore.

A second important lesson learned is that an international, intergenerational team is an absolute necessity. Simply having users and testers from other countries is not enough; their input is valuable, but it comes too late in the design process to influence major design changes. Team members from different cultural backgrounds offer perspectives that an American-only team simply would not think to consider. Similarly, team members who are children understand how children like to look for and read books, and what interface tools are difficult or easy, and fun or not fun. Enthusiastic advisors and volunteers are also a crucial resource. The ICDL team does not have the time, money, or resources to address all of the issues that surface, and advisors and volunteers are key resources in the development process. Bringing together a diverse team as possible is highly recommended. The goals of educational enrichment and international understanding in an international library make it an attractive resource for people to want to help, so assembling such a team is not as difficult as it sounds.

Beyond the human resources, the technical resources involved in making ICDL an international environment necessitate the examination and adjustment of software and interfaces at every level. Unlike many digital libraries that only focus on one or a few languages, ICDL must be simultaneously multilingual, multicultural, and multigenerational. As a result, a third lesson is that freely available and open-source technologies are now available for making the necessary infrastructure meet these criteria. With varying degrees of complexity, the team was able to get all the pieces to work together properly. The more difficult challenge, unfortunately, falls on ICDL’s users, who may need to install new fonts to view metadata in different languages. However, as computer and browser technologies advance to reflect more global applications, this problem is expected to lessen and eventually disappear. Having technical staff capable of searching for and integrating open-source tools with international support to handle these technical issues is highly recommended, as well as usability staff versed in the nuances of different operating systems and browsers.

Finally, the more subjective issue of cultural interpretation has proven to be the most interesting challenge. It is one that will likely not disappear as ICDL’s collection grows and the next stage of development is embarked on for translating the interface to support other languages and cultures. The fourth lesson learned is that culture pervades every aspect of both the visual design and the content of the interface, and that it is necessary to examine one’s own biased cultural assumptions to ensure respect of others. However, with the enthusiasm that continues to be seen in the ICDL team members, advisors, volunteers, and users, future design challenges will be able to be addressed with their help. The final recommendation is to actively seek feedback from team members, volunteers, and users from different backgrounds about the cultural appropriateness of all aspects of your software. It may not be possible to address all cultures in your audience right away, but it is important to have a framework in place so that these issues are addressed eventually.
Acknowledgments

ICDL is a large project with many people who make it the wonderful resource that it has become. We thank them all for their continued hard work, as well as our many volunteers and our generous contributors. We would especially like to thank NSF for our Information Technology Research grant, and IMLS for our National Leadership grant. Without this generous funding, our research would not be possible.

References