Voces del Caribe – A Digital Repository

An Interactive Qualifying Project Report

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements of the

Degree of Bachelor of Science

by

_____

Michael Fakharzadeh

_____

Artur A. Janc

_____

Katarzyna Koscielska

Date: March 2, 2007

_____

Professor H. J. Manzari, Project Advisor

1. digital repositories

2. the Caribbean

Abstract

Voces del Caribe is a collaboratively developed digital repository of cultural resources from the Caribbean. For this project we researched and evaluated choices for creating digital archives, and developed the design framework for the *Voces del Caribe* on-line system. The result is a bilingual website presenting information about Caribbean culture grouped into sections devoted to arts, film, language, history and music, and containing related information of use to scholars and students of Caribbean cultural heritage.

Table of contents

1. Introduction

Voces del Caribe, or Caribbean Voices, is an interactive digital archive of images, texts and music from the Caribbean. *Voces* is a scholarly tool that aims to preserve and make accessible the images, texts and music that identify the Hispanic Caribbean. *Voces* is designed to provide an interactive resource for students, scholars, and the general public worldwide. The digital archive, when finished, will educate even the most informed of the music, arts, language, history, and maps of the Hispanic Caribbean.

Libraries, museums, and archives are currently lacking funding. Buffalo and Erie County, New York closed 16 of their 52 libraries due to a 6 million dollar deficit in late 2005 (Buffalo). These public resources are increasingly unable to maintain and protect the artifacts that they possess. In Canada, school libraries are spending an average of only $2,000 on new library materials per year (Weiss). Digital libraries and archives are an excellent and cost-effective way of preserving and presenting the artifacts traditional institutions contain. Since the majority of digital libraries and archives can be accessed over the Internet, one simply requires a computer to have a wealth of information at his/her fingertips.

Libraries, archives, and museums are also having difficulty attracting patrons. In Oregon City in 2006, the libraries were open only 23 hours per week due to a lack of funding and interest; in January 2007 the city increased its library budget slightly, allowing the facilities to be open 35 hours per week (Mayes). The world is moving into a digital age and all of these resources must follow. In 1996, libraries only spent 2% of their acquisition budget on electronic resources; today, that percentage has jumped to 30% (ACLS). This movement is one of the motivations for the creation of digital repositories that contain a variety of resources from libraries, museums, and archives. The digital age that we live in was one of the major incentives

for the *Voces del Caribe* project.

A major goal of this project is cultural preservation. The *Voces del Caribe* digital archive will protect cultural artifacts that may be lost otherwise. By converting artifacts into digital information, we are accomplishing two tasks: as mentioned before, we are preserving them; but these artifacts, that would otherwise be unseen by the public, are also going to be readily available for anyone. Thus the *Voces del Caribe* digital archive will provide central access to cultural artifacts that would otherwise be spread out among distant countries, archives, libraries, museums, and websites.

Academic collaboration is another significant aspect of the *Voces del Caribe* project. The website will become a venue for academics, as well as public, to share their knowledge about the Hispanic Caribbean. Students will have the ability to use the website for certain coursework or even essays. Others will have the ability to converse one-on-one with experts about any questions they may have. The future of the *Voces del Caribe* project will bring an immeasurable amount of knowledge that will be shared with the rest of the world.

The *Voces del Caribe* digital archive will provide samples of art, organized by country, for the public to discover, including film, music, and paintings. All of this art will come from a diverse variety of sources and will, in turn, be preserved digitally. Films are listed with relevant links that provide in-depth information. The digital archive will also contain maps that can be enlarged and thoroughly examined. We will embrace articles written by scholars about the Hispanic Caribbean by posting them on the website, when completed. These resources will immerse academics, as well as the public, into the culture of the Hispanic Caribbean.

As the *Voces del Caribe* project expands, scholarly communication and collaboration will be a crucial facet of the website. Once the project's website is complete, there will be a section

for academic communication so that any questions or comments can be answered just by posting them on the site. Eventually, we will allow the general public access to this forum and will be able to have their questions or comments answered by members of the academic community. This feature is one of the key interactive portions of the *Voces del Caribe* website.

The *Voces del Caribe* interactive website is designed to be easy to use and a valuable resource to anyone who utilizes it. The layout (as shown in Figure 1) places the site's navigation at the top of the screen which allows fast and easy access to everything that the site offers. We set up the website so that the user can select between the English or Spanish version of the website, allowing Spanish and English speaking users easy access to the website's content. The website's digital archive will include an encyclopedic search engine that will allow access to any resource that the user may need. The site uses various colors to catch the user's eye, but not distract or annoy him/her. These simple features make the *Voces del Caribe* digital archive easy to access and explore.
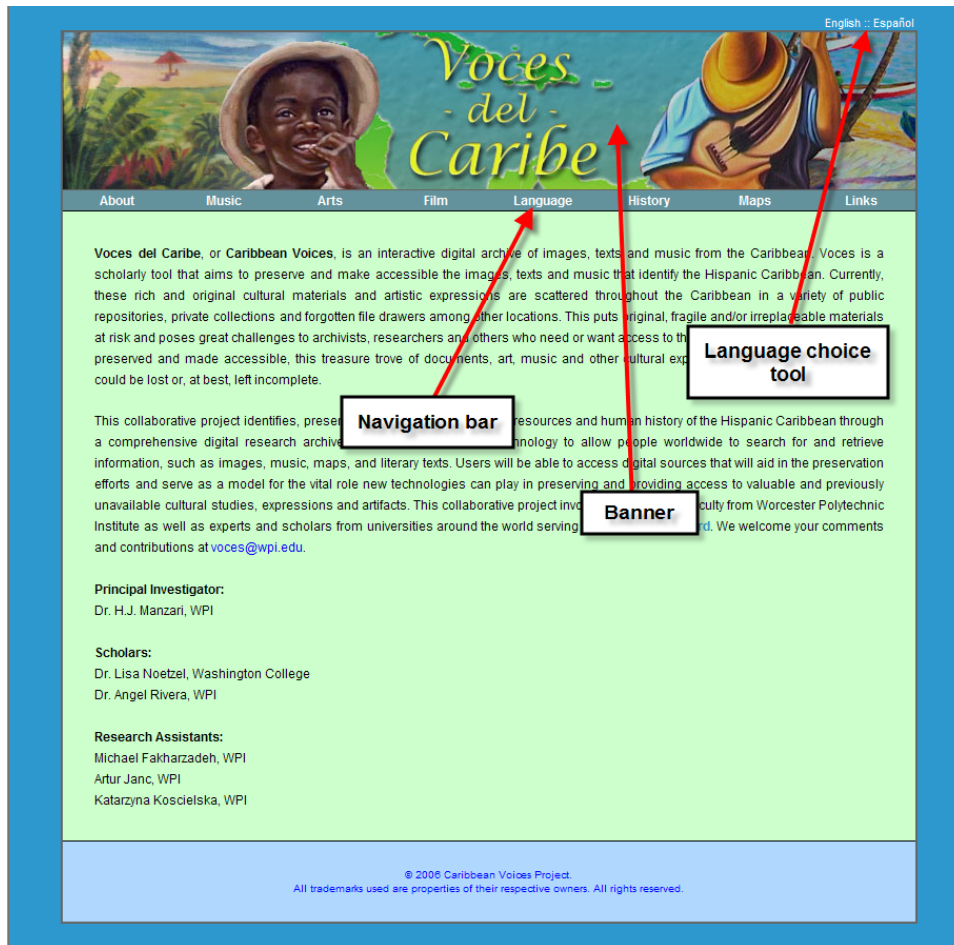
Figure 1. Voces del Caribe.org main page.

The *Voces del Caribe* project will not conclude when this IQP is completed. The content created from the arduous research done by our group during the span of this IQP is far from everything that the Hispanic Caribbean has to offer. Academic collaboration is the most important constituent to this project's potential success. For example, currently the Spanish portion of the website is incomplete. This is largely due to the lack of a native translator that has the time to translate all of the English content. With the help of the academic community, the *Voces del Caribe* website will become bilingual. The degree of academic collaboration will dictate the efficacy of the *Voces del Caribe* project since a majority of Hispanic Caribbean

culture can be manifested best by experts and natives.

The potential concepts that can be applied to this project are endless. When future academics continue work on this project, new features will inevitably be shaped. This project's purpose was to create a framework and some initial content that can be further developed. This goal was met through careful planning and thinking. Prospective academics will have the opportunity to combine their creative abilities and ambitions, and expand the *Voces del Caribe* project into an internationally renowned resource of Hispanic Caribbean culture.

2. Background

In order to create an online archive of cultural material it is crucial to understand the role of digital archives in the process of preserving educational content, as well as the challenges faced by such projects. Throughout the past 15 years, the importance of digital repositories has increased significantly, both due to the ease of accessing digital content and the decreasing cost of migration. We have therefore seen a great variety of projects aiming to migrate, digitize and present cultural, artistic and intellectual content. Such initiatives can be divided into four main categories: digital extensions of existing libraries, domain-specific and general on-line repositories, on-line museums, and standards projects aiming to ensure long-term preservation and accessibility of resources.

As the most common goal of digital migration is to allow a large audience to gain access to digitized resources, we have focused on analyzing digital collections and institutions with a strong on-line presence. We evaluated both freely available databases, as well as paid sites gathering resources of educational and scholarly use. To analyze the process of migrating traditional library resources we reviewed information about the status of status of public libraries in the United States. The on-line repositories chosen for review were ones often utilized in a college environment for educational purposes. The digital museum collections we studied belong to the most famous museums worldwide, as those institutions possess the most resources to undertake digitization efforts.

When analyzing the situation of libraries we have found that digital additions to traditional libraries have become almost indispensable in the modern world. The last decade has seen libraries struggle to attract and retain patrons; we found that only through the modernization of the infrastructure and adding computing resources can libraries remain valuable educational

facilities. We discuss the challenges facing existing libraries, as well as discuss the implications of digitization of library resources as a strategy for coping with them, in Section 2.1; we also present digital libraries – collections of publications, such as journal articles, usually designed to provide domain-specific information to scholars.

Section 2.2 presents the concept of an on-line repository – a usually searchable collection of digital resources, either devoted to a particular topic, or a general database, often created in effort to digitize an existing collection. One example of a domain-specific repository is the Connecticut History Online website, resulting from collaboration between multiple institutions, and providing on-line access to photographs and historical materials related to the history of Connecticut. Another increasingly popular example of such a repository is a campus-wide system collecting scholarly and education-related materials submitted by faculty and students in order to preserve the project created as part of coursework or research.

Section 2.3 discusses how on-line museums approach digitization. Whereas digital archives and repositories usually contain mostly written resources, on-line museums often present audiovisual and interactive content. Most on-line museums were created to establish a digital presence for an existing museum and enable potential visitors to view a small part of the museum's most valuable resources on-line. However, some museums present their entire collections on-line. Another type of an on-line museum is one that exists in a purely virtual form – usually being a collection of resources from a particular domain or the works of a certain artist; we analyze an interesting example of the Diego Rivera on-line museum.

A substantial part of migrating existing cultural and intellectual resources is choosing the right digitization strategy, computing platforms, file formats and presentation format. There have been many industry-wide, as well as government efforts to create methodologies of converting

content into a digital form, while ensuring that the data will remain accessible. In section 2.4 we discuss the most important postulates of theoretical approaches to digitization, as well as practical conclusions from institutions which undertook digitization efforts.

2.1. Digital libraries

In recent years, the rapid development of communication technologies and ability to find relevant information online has decreased the world's dependence on traditional databases of written information. This has lead to the fact that libraries are not being funded sufficiently in many communities due to decreased public interest.

2.1.1. Digital efforts of traditional libraries

Many towns around the country are being forced to raise taxes in order to fund their libraries. In Buffalo and Erie County, New York, 16 of their 52 libraries are being closed because of a $6 million dollar deficit; even after these libraries are closed both Counties will have to reduce their expenditures in order to reduce the deficit (Buffalo). The article states that these libraries will be closed "to reduce the deficit." In Oregon City and Clackamas County, Oregon, libraries are experiencing a decline in funding. Last year, Oregon City libraries were only open for 23 hours per week due to budget cuts; in January 2007 funding was slightly increased and they are now open for 35 hours a week (Mayes). In Lebanon, PA the Lebanon Community Library lost one-sixth of its budget, beginning in 2004. The book budget was cut from $125,000 to $65,000, the library was forced to close on Sundays, and two full-time employees were forced to leave (Rhen). In Canada, 93.3% of schools have libraries. The mean expenditure on new library materials in those libraries was a mere $2,000 (Weiss). With rising costs and diminishing funding, the outlook for libraries is bleak.

The decline of library funding could be the result of a move into a digital age that will leave libraries, at least in the traditional sense, behind. In the 1980s the ACLS surveyed 4,000 scholars in order to obtain "what they think about a wide range of issues of greatest concern to

their careers, their disciplines, and higher education in general." The first response was "the rapid increase in computer use." (ACLS). In 1980, it was reported that only 2% of all respondents owned a computer. By 1985, that number increased to 45% (ACLS). Email, Web pages, search engines, and digital media did not exist just a generation ago. Now all of these items are neither exotic nor new (ACLS).

According to the ACLS Report, "librarians, curators, archivists, and the private sector are aligning around the building of a 'global digital library'." In order for a library to be modern, it is required to have expensive technological equipment like desktop computers, projectors, and laptops. Most modern libraries also have digital repositories of articles and other resources. These repositories usually consist of resources from all around the world. Our project aims to contain resources from all around the world and support the 'global digital library' movement.

Books are no longer the most important aspect of a library. People look for online access and technological readiness when searching for a library. Libraries today spend millions on digital resources; meaning they too are moving towards a unified digital infrastructure. In 1996, libraries spent 2% of their acquisition budget of electronic subscriptions and resources. Currently, the average portion of the library's acquisition budget which is spent in digital materials is 30%. Some institutions even spend 50% of their budgets on digital materials (ACLS). Subscriptions to Internet databases like Lexis Nexis (enabling its customers to access and search 5 billion documents pertaining to law and business) allow libraries to provide massive amounts of information for anyone willing to use them.

2.2.2. Digital library databases

Digital libraries are the key to a widespread, easy-to-access, and low-cost solution for

distributing knowledge that is difficult to gather. The public, students, teachers, and scholars desire online access to primary source materials which are currently held in museums, historical societies, libraries, collections, and archives. This includes books, journals, newspapers, magazines, government documents, manuscripts, maps, photographs, satellite images, census data, recorded sound, film, and broadcast television (ACLS). Digital resources favor openness and collaboration among the world. They also offer ways to reunite separated collections, like The International Dunhuang Project. Information about any certain subject is usually spread out over many sources (from libraries to archives). Digital libraries have the ability to educate people about cultures which are unknown to them, thereby potentially eliminating stereotypes.

A digital library can also be used as an "e-textbook" for certain courses and projects within courses. Marc Levoy of the University of California states that, "once humanities faculty begin using the laboratory in their research, they would also find creative ways to fold its technology into their teaching, for example through project-based assignments in upper-level courses (ACLS)." Provost James O'Donnell of Georgetown University further explains this concept,

"[humanities computer labs] are zones of experimentation and innovation for humanists, within and without traditional institutions. These zones should be part and parcel of the formal academic structure. Ghettos are not the answer. We need instead the creation of privileged but open communities, where the very best young people are challenged to invent, experiment, break things, and succeed (ACLS)."

When students wish to travel to other countries they will benefit greatly from the cultural background a digital library will provide, thus reducing culture shock.

In a recent study it was found that there were 55% more Internet searches conducted in 2005 than in 2004, while only 3% rise in the number of Internet users was recorded (Surge). This

suggests that Internet users are more likely than before to look for the information they need online. The American Civil Liberties Union currently provides all of the American public the ability to view and critique the Patriot Act by posting it on their website. By doing this, the ACLU has helped bridge the gap between what happens in Washington, D.C. and what happens at home, wherever that may be.

The Internet Movie Database (IMDb; example page shown in Figure 2) is also an outstanding example of how easy-to-access and useful a digital database can be. IMDb allows anyone who has access to the Internet to look up any information about nearly any movie, TV show, or unfinished work from anywhere in the world. It is therefore an extensive database of movie facts and an Internet phenomenon of its own. Created in 1990, it started as a "a hobby project by an international group of movie fans" (IMDb Help), to be gradually developed into a huge repository of information about around 900,000 movies and TV series (IMDb Help), mainly thanks to the contributions from filmmakers and movie fans all over the globe. IMDb lets its users add and correct the content directly through their website, pending the approval of administrators. It is also fully searchable using a variety of criteria, which coupled with a simple and user-friendly interface make it a powerful tool for searching information about the topic. Needless to say, IMDb is an important inspiration to us and authors of other database-repository-type projects, as it relies on community interaction for accumulating content, collects and showcases vast amounts of facts in a compact, easily searchable form and finally, is accessible to anyone, anytime, and free of charge. Another important aspect of IMDb is the ability of users to rate, discuss, and view previews of nearly every movie in the database. This concept, if applied to a digital library, has a tremendous amount of potential.

Figure 2. Example movie page from IMDb website

2.2. Digital repositories

To discuss the role of on-line digital repositories we first present the reasons which prompted many organizations to digitize their existing content. We also introduce the concept of a purely virtual digital repository, an on-line collection of resources gathered from different sources, without an existing physical collection. We present case studies and discuss the outcomes of several important projects of this nature.

2.2.1. Rationale

The recent increase in the popularity of the Internet and the benefits of allowing a world-wide audience to view on-line content has prompted many institutions to consider adopting digitization strategies. Large collections of artistic and historical resources have been converted to a digital format, mostly in the form of texts and images. Many organizations have seen several benefits of creating on-line repositories of digitized cultural resources.

The most apparent benefit of creating a digital repository is that the content is now available to all Internet users, allowing potential patrons to find out information about certain collections. The possibility of quickly searching for relevant information based on any number of different criteria makes browsing through a collection much more effective.

In addition to easier access to information about a collection, digitizing resources enables a new type of collaboration between institutions. Where creating shared collections has often been cumbersome due to location and ownership constraints, a cross-institution digital repository could now present materials from different museums or archives thematically, adding value to both collections. As many organizations have noticed, the collaboration can be extended to

scholars and patrons, allowing them to comment on the resources, adding valuable information.

The third immediate advantage of creating a digital repository is risk mitigation – natural disasters which have the potential of destroying or damaging physical resources will not affect digital versions of documents or images.

Whereas most on-line repositories of cultural materials were started by institutions to migrate existing content into the digital format, there are several examples of purely virtual repositories. Such sites usually contain digital representations of artwork created by a certain artist, or devoted to a particular topic. This is an interesting concept, as it allows for creation of scholarly and cultural projects which, while having large educational and cultural value, can be carried out with very limited resources.

2.2.2. Case studies

An important example of a digital repository created as a result of inter-institutional co-operation is the Connecticut History Online website (see Figure 3). The project was created in 2000 by the Connecticut Historical Society, Connecticut State Library, Thomas J. Dodd Research Center at the University of Connecticut, Mystic Seaport, and the New Haven Colony Historical Society. The site's goal is to reflect "Connecticut's social, educational, political, civic, and cultural life" in the years 1800-1950. By gathering resources from all its member institutions, the repository is able to present 14,000 photographs, drawings and prints. The entire collection is searchable; several types of education resources are provided, including digital thematic 'journeys' and lesson plans.

Figure 3. Connecticut History Online website

A similar project has been created by the California Digital Library, a part of the University of California. The mission of the Online Archive of California is to "provide access to materials such as manuscripts, photographs, and works of art held in libraries, museums, archives, and other institutions" from the entire state. The OAC contains over 120,000 images; 50,000 pages of documents, letters, and oral histories; and 8,000 guides to collections, all of which are from California museums, historical societies, and archives. It resembles an online catalog, of other institutions' (mainly universities) digital collections, created to allow scholars and research to access relevant cultural materials without having to travel to a particular collection site.

Another related initiative is the Collaborative Digitization Program (see Figure 4), initially created as The Colorado Digitization Program. It aims to provide digital access to cultural heritage collections including human culture, science, and art. The CDP acquires its information through partnerships with various museums, libraries, and archives in Arizona, Kansas, Montana, Nebraska, Nevada, New Mexico, Texas, and Utah. While being similar in

nature to the OAC, it provides a smaller collection of content, but presents various other materials useful to educators, including lesson plans and professional development resources.

Figure 4. Collaborative Digitization Program website

An example of a digital repository which isn't based on any existing physical collection, but is rather a compilation of digitized resources gathered from various existing collections is the on-line Diego Rivera museum (as shown in Figure 5). It contains biographical information, a gallery of paintings and murals, as well as short related movie clips. Overall, the on-line collection provides a wealth of information about the artist, focusing especially on providing digital versions of artwork, making the repository a possibly interesting addition to classroom study of Diego Rivera. The example of this repository shows that efforts to create on-line collections by grouping digital resources is a viable alternative to making on-line repositories by digitization of existing collections.

Figure 5. Diego Rivera Museum website

2.2.3. General-purpose digital repositories

In addition to on-line repositories created by organizations to present their own or shared collections of cultural resources, there have been several efforts to create repositories of materials the scope of which goes beyond the cultural context. An increasingly popular example of such a repository is a campus-wide system collecting scholarly and education-related materials submitted by faculty and students in order to preserve the project created as part of coursework or research (Gibbons). Such systems usually focus on collecting resources from and for a more narrow audience, providing it with easy access to stored resources, but not necessarily presenting those resources in an appealing fashion.

To aid institutions in creating their own repositories, several useful platforms and tools have been created. An example of such a platform is MIT's DSpace, an open digital repository system which "captures, stores, indexes, preserves, and distributes digital research material". The system is used by organizations as an institutional or learning objects repository, as well as for

records management and digital preservation. The system has built-in cataloging capabilities to facilitate searching for specific content. Being open source software, it can be edited by each institution and tailored to its needs. It is primarily used by academic institutions, serving several hundred universities worldwide (DspaceInstances).

NINES is an excellent example of a new movement called organized scholarship. NINES is a group which was founded to develop an online publishing environment for integrated peer-review scholarship. Project MUSE is an online database of full-text journals in the humanities, arts, and social studies. Project MUSE contains 250 different journals which can be thoroughly searched. JSTOR is another online database of over 600 "backfiles of scholarly journals" (JSTOR), some of them a hundred years old. The Google Book Search project is an excellent example of a vast digital library that is readily accessible to anyone who has access to the Internet. While Google's project is quite ambitious, it is important to note that it is estimated that only one-third of the books held in research libraries are being represented by Google Books. This does not include the other forms the books may be in or the fact that many books are not held in research libraries (Lavoie). Google has already catalogued more than eight billion web pages and one billion images.

An exceptional example of a repository of similar nature which also provides useful information on cultural topics is Wikimedia Foundation's Wikipedia. While it is a general-purpose encyclopedia, its open nature stimulates adding new content, making it a useful research aid for many topics. The site allows grouping information about a topic or group of related topics into portals, such as the Caribbean Portal (see Figure 6). While opening a resource to editing by any Internet user has its drawbacks, Wikipedia has become a landmark system showing the possibilities brought about by open collaboration platforms.

Figure 6. Wikipedia's Caribbean Portal

2.3. On-line websites of museums

Until the advent of the Internet, museums were the primary sources of "interactive" cultural information. Many museums, through years of accumulating resources, have been able to create permanent collections of sculptures, paintings, and other artwork. Emerging technologies have brought about many challenges, but also opportunities for museums. In this section we will present some examples of museums and their approaches to digitization.

A positive example of utilizing the capabilities of the Internet is the Louvre Museum webpage (shown in Figure 7). It contains databases and catalogs which allow the visitor to peruse and search for items within the Museum's 35,000-item art collection and 140,000-item prints and drawings collection. Each item has a photo and a description attached, to replicate the impression of being in the museum and reading labels next to each item. There is also professional commentary and analysis by the museum's staff available for each item.
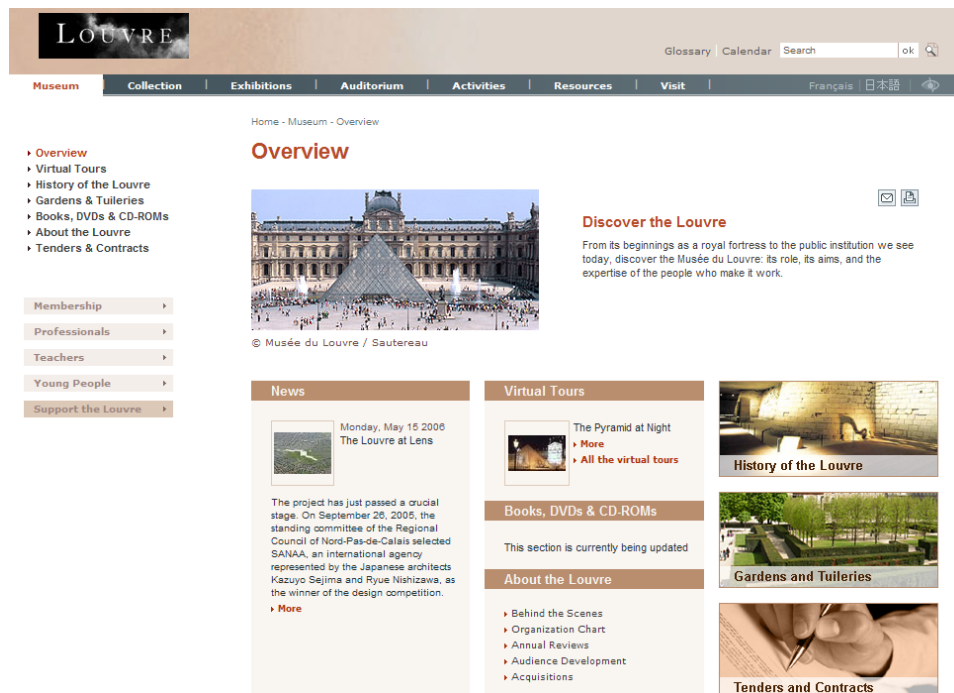
Figure 7. Louvre Museum website

Not all museums have embraced the idea of digitized their entire collections. For example the Metropolitan Museum has only about 6,500 objects from its larger than two million permanent collection accessible online in the form of pictures and short descriptions. The New York Guggenheim Museum offers just a small selection (including works of 169 artists) from its permanent collection available for viewing online.

All the presented museums allow searching their digitized collections, usually presenting advanced search interfaces, helpful to scholars, researchers, as well as regular users.

2.4. Technical standards and policies

For the success of digital migration efforts, the resulting digital resources must be readily available and accessible. Because of the rapid growth of the computer industry, both software programs and hardware resources become obsolete very quickly. In this case, institutions carrying out digitization initiatives must put extra care into choosing platforms which will remain available for a long period of time.

Many standards of representing visual and multimedia content are available, and very often they are mutually conflicting. It also often happens that a particular digital resource looks differently on two computers, because of system-dependent configuration settings. This is can pose a significant threat to the preservation, as parts of digital resources might become lost or unintentionally modified. In section 2.4.1 we present several standards efforts aiming to alleviate this problem.

Institutions have begun to realize that the process of carrying out migration is a non-trivial task and requires careful planning. Strategies for this process have been created and standardized. We discuss several of them in section 2.4.2.

2.4.1. Digital formats

The success of a digital file format can be measured by the scale of its adoption and percentage of computer users able to access information contained in the file. For distributing content on-line an important standard are the Multipurpose Internet Mail Extensions (Freed). MIME extensions allow computer programs to interpret multimedia data, exchange and read text in multiple encodings and understand multi-part message bodies. They are especially useful for

Web browsers and email programs for interpretation of different kinds of multimedia data.

One solution to the problem of digital files depending on system-dependent information was attempted by creating a PDF document standard specially suited for archival applications. The PDF/A-1 format (PDF/A JWG) "aims to preserve the static visual appearance of electronic documents over time and also aims to support future access and future migration needs". It tries to achieve this goal by embedding the necessary fonts, metadata and device-independent color information in the PDF/A file itself. The main goals for this archival standard are for the documents to be device independent, self-contained self-documenting, unfettered, available and easily adopted.

Another approach to the problem is the Bitstream Syntax Description Language (Amielh), which utilizes XML schema to enable multiple devices access the same document in a manner understandable to each of them (so that a document can contain the same relevant information, but look differently on a PDA and a PC). The paper presents an XML framework to describe the structure of a bitstream, so that the document can be transformed to dynamically adapt multimedia data to the network and terminal capabilities.

Even when one file format is widely adopted, data represented in other formats must still be accessible. Without the description of the format representation, a file is *merely a collection of undifferentiated bits*. The Global Registry for Digital Format Representation (GDFR) created by the Harvard University Library is a digital archive of file formats. The registry allows information created using in different file formats to be to understood and deciphered.

2.4.2 Management of preservation efforts

The process of digitizing cultural material must take into account many non-technical factors, including budgeting, the timeframe of the project and the overall goal to be accomplished through migrating materials to digital formats. In order to aid organizations in carrying out this process several standard approaches have been proposed.

An official standard is the ANSI X3.285 (ANSI) describing a metamodel for the management of shareable data. The standard describes the structure of a data registry: a place to keep facts about characteristics of data that are necessary to clearly describe, inventory, analyze, and classify data.

Since the needs of organizations vary widely, databases of standard approaches have been created, so that an organization can choose the most appropriate process for their needs. The DIFFUSE (Dissemination of InFormal and Formal Useful Specifications and Experiences) project is a European Union initiative to collect standards relevant to the information society. The objective of the DIFFUSE project is to "provide a single, value-added, entry point to up-to-date reference and guidance information on available and emerging standards and specifications that facilitate the electronic exchange of information."

In the recent years, two new approaches have emerged, both postulating that digital preservation is a process rather than a specific goal. A document titled *Migration: Context and Current Status* (Netherlands) is part of a governmental initiative in the Netherlands to preserve important information in a digital format. The document describes migration defined as "a set of organized tasks designed to achieve the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation" (Netherlands). The point of migration is to preserve important digital

data utilizing a practical approach, rather than discuss the theory of digital data archiving. The publication outlines methods to ensure the data will be efficiently preserved for future generations.

According to Stewart Granger, it is very difficult to ensure long-term survival of data stored in digital formats unless a complex emulation system is designed and employed. He notes that the idea behind emulation is "to be able to access or run original data/software on a new/current platform by running software on the new/current platform that emulates the original platform". The three most popular emulation options are options are to: emulate applications, emulate operating systems, or emulate hardware platforms.

3. Methodology

3.1. Active Research

One of our group's goals was to make the *Voces del Caribe* website a valuable asset to the Internet community. We exhaustively researched and brainstormed on how to generate a resource that the academic and non-academic community will consider useful. Our group decided that it needed to divide the website into several sections so that it would be easier to navigate. We initially decided to split the website into two versions: Spanish and English. We then proceeded to divide all of the content into the following categories: arts, language, history, maps, and links. Within each of those sections, we further divided the content based on the country that it represented. The result of this structure is an easily accessible digital repository.

Our group did not solely focus on navigation when considering the Internet community. We spent a substantial amount of time encouraging academic collaboration with the *Voces* project. We received a paper from Dr. Lisa Noetzel of Washington College and decided to introduce the language section of the website with her piece. By posting her article on a prominent section of the site, we are encouraging other academics to submit their work, so that the *Voces* repository will include more content written by experts.

Non-academic collaboration will also facilitate the utility of the *Voces* project. Non-academics have the ability to make remarks on the website (via email) and help our group, as well as future ones, make any necessary changes. This feature is quite important because it allows the *Voces* project to be constantly scrutinized and therefore improved.

The main source of information for the Dominican film page (Figure 8) was the Internet Movie Database and the website of Dominican *Dirección Nacional del Cine* (DINAC; translates to National Institute of Cinema). Both provided movie poster images and short descriptions of

the features, which were used in preparing our own synopses of each film. In some cases, online news articles about real-life dramas depicted in the movies were helpful in writing more detailed descriptions (*Código 666, Un Pasaje de Ida,* documentaries of René Fortunato). Overall, all the information about film was collected online and carefully compiled by the authors of this project.
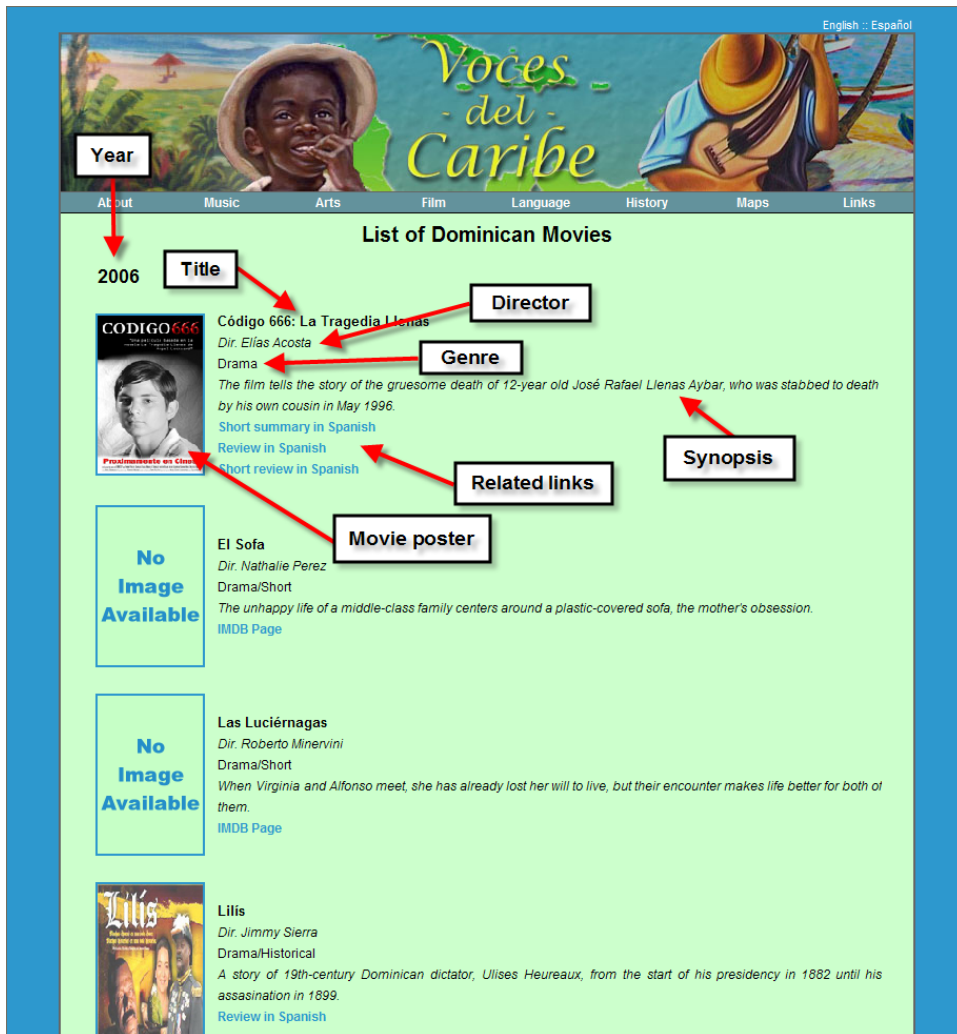

Figure 8. Dominican film page at Voces del Caribe.org

The Language section of *Voces del Caribe* demonstrates another important goal of our project: scholarly collaboration. The author of this section is Dr. Lisa Noetzel from Washington College, who kindly expressed interest in *Voces del Caribe* and decided to write about her own

research in order to contribute content to the project. We are hoping that, in time, *Voces* will become a learning portal, a repository of factual knowledge that people interested in the topic will go to for resources, but also for publication of their own findings. Collaboration and communication unobstructed by distance are the basic goal of the type of project *Voces* can be classified as: an Internet-based learning tool perpetuated and perfected by the academic community.

3.2. Organization

As the topic of our project is the Caribbean culture, we decided to come up with a way to divide the body of knowledge into several major categories. They are ordered according to their subjective importance to the authors of the project: **Music**, **Arts**, **Film**, **Language**, **History** and **Maps**. **Music** comes first, because to us it represents what is most characteristic of the Caribbean: energy, passion, and soul. **Arts** is the broadest category, encompassing literature, visual arts and performing arts with the exception of music and film, which have been given their own categories. **Film** has been separated out, because it plays a major role in popularizing the Caribbean culture outside of the region, and also because its multimedia nature enables it to include and showcase all other categories within it. **Language** category contains information and research articles about the "melting pot of linguistic diversity" (Noetzel) – the fascinating variety of languages and dialects functioning in the Caribbean. **History** section will include factual knowledge about the rich and complicated history of the region. Finally, **Maps** is a collection of various maps of the Caribbean area, used in place of a written geography overview.

3.2.1. Focus: Cultural Heritage

It is important to remember that our project focuses on the cultural aspect of the Caribbean heritage; this is why history and geography of the region are mentioned only briefly, and all the attention is focused on the artistic and linguistic content.

To increase clarity of organization, each category is subdivided by country. This division will also be useful when database search function will be enabled.

3.3. Implementation

For any on-line repository or digital archive system there are important implementation considerations, which affect the outcome of the project itself. The choice of the computing platform, method of presentation (on-line or off-line), search capabilities and database representation of data all have a direct impact on the functionality and accessibility of every digital system.

Our team decided that focusing on on-line collaboration and accessibility are the primary objectives for the *Voces del Caribe* website. Fostering communication between scholars, students and artists is an important goal for the project; it is also of utmost importance that the worldwide community has access to valuable cultural resources related to Caribbean culture.

3.3.1. Computing Platform

A natural choice of the platform on which to build our system is the Internet, which not only allows a worldwide audience to access content available in the *Voces* database, but also makes it possible for everyone interested to contribute resources on Caribbean art and culture.

Therefore the interface to the *Voces del Caribe* repository is a website located at <http://vocesdelcaribe.org>.

The *Voces* system, including the database with all articles and images, is hosted on the main Web server at Worcester Polytechnic Institute, maintained by the Computing & Communication Center (CCC). This Unix server provides adequate storage and backup capabilities to ensure that the *Voces* website remains accessible and can grow by incorporating new articles and multimedia resources.

Our website is designed using the Common Gateway Interface (CGI) technology, which allows scripts written in any programming language to be run within the Web server. The scripting language of choice for our group was Python, due to its clear syntax and ability of the scripts to be maintained and extended by other groups continuing the *Voces* project. Each resource on the website exists on the server as a script, which provides layout information (adding the header and footer, so that each sub-page is consistent with the look of the website and has the same banner and navigation menu). Thus adding of a new page only requires including this header information, as well as the real content of the page. An example script from the website is attached in Appendix B.

The presentation layer of the website utilizes the standard HTML 4.0 markup language, along with Cascading Style Sheets (CSS) used to format the pages and dictate the colors. The layout has been tested in several web browsers (including Microsoft Internet Explorer, Mozilla Firefox and Opera) to ensure that the website is universally accessible.

3.3.2. Visual/graphic design

Our group created the *Voces del Caribe* banner using Adobe Photoshop 7.0. We created

34

the background by cropping various parts of Caribbean-themed paintings (clockwise from top left in Figure 9: unknown author's "Guitar Player", "Fisherman's Bay" by Bob Thompson, Sue Snow's little boy painting and "The Beach" by Charles Godwin) and enhancing a basic map of the Caribbean, and organizing them in a coherent, artistic manner. The area above the boy's head is filled with a gradient which allows a seamless transition from the beach to the map. The '*Voces del Caribe*' text was written using two fonts, Lucida Calligraphy and Lucida Handwriting. The colors used in the website's graphic design were inspired by various Caribbean-themed paintings, which are dominated by blue seas and skies, rich green foliage of tropical plants and amazingly colorful flowers. For examples see Appendix A.

Figure 9. Banner composition

4. Appendices

Appendix A. Caribbean Paintings.

Figure 10. Barry Launius "Four Ashore"

Figure 11. Barry Launius "Akumal Sunrise"

Figure 12. Bob Thompson "Haitian Landscape"

Figure 13. Celia Lacayo "Flores del Exilio"

Appendix B. Example scripted page on the *http://vocesdelcaribe.org* website

```python
#!/usr/bin/python

# Required for the browser to understand the script
print "Content-type: text/html\n\n"

# Print out the banner and navigation menu
header = open("include/header.html")
print header.read()

content = """
<div align="justify">
        <p><strong>Caribbean Music</strong></p>
        <a href="http://www.vocesdelcaribe.org/dominican_music.shtml">
                Dominican Music
        </a>
</div>
"""

print content

# Print footer information
footer = open("include/footer.html")
print footer.read()
```

5. References

ACLS Commission on Cyberinfrastructure for Humanities and Social Sciences. <u>The Draft Report of the American Council of Learned Societies' Commission on Cyberinfrastructure for Humanities and Social Sciences</u>. Nov 5, 2005. Accessed on Feb 16, 2007 <http://www.acls.org/cyberinfrastructure/acls-ci-public.pdf>.

<u>American Civil Liberties Union</u>. 2007. ACLU. Accessed on Feb 20, 2007 <http://www.aclu.org>.

Amielh, Myriam and Sylvain Devillers. "Bitstream Syntax Description Language: Application of XML-Schema to Multimedia Content Adaptation." 7-11 May 2002. <u>Eleventh International World Wide Web Conference Proceedings</u>. Accessed on Feb 22, 2007 <http://www2002.org/CDROM/alternate/334>.

ANSI – American National Standards Institute, Inc. <u>American National Standard for Information Technology – Metamodel for the Management of Shareable Data (Draft)</u>. Feb 20, 1998. Accessed on Feb 22, 2007 <http://metadata-standards.org/Document-library/Draft-standards/X3-285-Mgmt-of-Sharable-Data/X3-285.PDF>.

"Buffalo plans to close 20 of its 52 branches." <u>American Libraries</u>. 36.8 (Sep 2005): 13(1). *Student Edition*. <u>Thomson Gale</u>. Accessed on Feb 16, 2007 <http://find.galegroup.com>.

<u>Caribbean Portal – Wikipedia</u>. Nov 26, 2006. Wikimedia Foundation, Inc. Accessed on Feb 20, 2007 <http://en.wikipedia.org/wiki/Portal:Caribbean>.

<u>Collaborative Digitization Program</u>. 2007. Collaborative Digitization Program. Accessed on Feb 20, 2007 <http://www.cdpheritage.org/index.cfm>.

Connecticut History Online. 2001. Connecticut History Online. Accessed on Feb 20, 2007 <http://www.cthistoryonline.org>.

Diego Rivera Web Museum. 2006. Javier A. Rivera. Accessed on Feb 20, 2007 <http://www.diegorivera.com/index.php>.

DIFFUSE Standards. Feb 7, 2007. Digital Curation Centre. Accessed on Feb 21, 2007 <http://www.dcc.ac.uk/diffuse>.

DINAC – Dirección Nacional de Cine de la República Dominicana. 2006. DINAC. Accessed on Feb 21, 2007 <http://dinac.gov.do>.

DSpace Federation. 2007. MIT Libraries & Hewlett-Packard Company. Accessed on Feb 20, 2007 <http://www.dspace.org>.

DspaceInstances. Feb 19, 2007. MIT Libraries. Accessed on Feb 20, 2007 <http://wiki.dspace.org/index.php/DspaceInstances>.

Freed, Ned and Nathaniel S. Borenstein. "Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types." The Internet Engineering Task Force. Nov 1996. Accessed on Feb 22, 2007 <http://www.ietf.org/rfc/rfc2046.txt>.

GDFR - Global Digital Format Registry. Oct 7, 2006. The President and Fellows of Harvard College. Accessed on Feb 22, 2007 <http://hul.harvard.edu/formatregistry>.

Gibbons, Susan. "Seeking a System for Community-Driven Digital Collections at the University of Rochester." SPARC E-News. Feb-Mar 2002. Accessed on Feb 16, 2007 <http://www.arl.org/sparc/pubs/enews/feb02.html#5>.

Godwin, Charles. The Beach. CaribbeanPaintings.com. Accessed on Feb 26, 2007 <http://www.caribbeanpaintings.com/CharlesGodwin/TheBeachthum.jpg>.

Google Book Search. 2007. Google. Accessed on Feb 20, 2007

<http://books.google.com>.

Granger, Stewart. "Emulation as a Digital Preservation Strategy." D-Lib Magazine 6

(2000). D-Lib Magazine. Oct 2000. Corporation for National Research Initiatives. Accessed on

Feb 23, 2007 <http://www.dlib.org/dlib/october00/granger/10granger.html>.

Guggenheim Museum - The Collection. 2007. The Solomon R. Guggenheim

Foundation. Accessed on Feb 21, 2007 <http://www.guggenheimcollection.org/index.html>.

Guitar Player. Accessed on Feb 26, 2007 <http://cuba.romanvirdi.com/art-musician.jpg>.

IMDb Help. 2007. Internet Movie Database Inc. Accessed on Feb 16, 2007

<http://imdb.com/help/?general>.

The International Dunhuang Project. 2007. British Library. Accessed on Feb 20, 2007

<http://idp.bl.uk>.

The Internet Movie Database. 2007. Internet Movie Database Inc. Accessed on Feb 20,

2007. <http://www.imdb.com>.

JSTOR - The Scholarly Journal Archive. 2007. JSTOR. Accessed on Feb 20, 2007

<http://www.jstor.org>.

Lacayo, Celia. Flores del Exilio. Atenea Gallery. Accessed on Feb 26, 2007

<http://www.ateneagallery.com.ni/artistas/primitivistas/lacayo/img/mas/flores%20del%20exilio.j

pg>.

Launius, Barry. Akumal Sunrise. CaribbeanColors.com. Accessed on Feb 26, 2007

<http://www.caribbeancolors.com/caribbeancolors015001.jpg>.

Launius, Barry. Four Ashore. CaribbeanColors.com. Accessed on Feb 26, 2007

<http://www.caribbeancolors.com/caribbeancolors013002.jpg>.

42

Lavoie, Brian, Lynn S. Connaway and Lorcan Dempsey. "Anatomy of Aggregate Collections." <u>D-Lib Magazine</u> 11 (2005). <u>D-Lib Magazine</u>. Sep 2005. Corporation for National Research Initiatives. Accessed on Feb 16, 2007 <http://www.dlib.org/dlib/september05/lavoie/09lavoie.html>.

<u>Lexis Nexis – Innovative Solutions for Knowledge-Driven Professionals</u>. 2007. Reed Elsevier Inc. Accessed on Feb 20, 2007 <http://www.lexisnexis.com>.

<u>Louvre Museum Official Website</u>. 2007. Musée du Louvre. Accessed on Feb 20, 2007 <http://www.louvre.fr/llv/commun/home_flash.jsp?bmLocale=en>.

Mayes, Steve. "Failed library funding measure had pockets of strong support" <u>The Oregonian</u>. Jan 18, 2007. Accessed on Feb 16, 2007 <http://search.oregonlive.com>.

<u>The Metropolitan Museum of Art, New York</u>. 2007. The Metropolitan Museum of Art. Accessed on Feb 21, 2007 <http://www.metmuseum.org/home.asp>.

The Netherlands. The National Archives and the Ministry of the Interior and Kingdom Relations. <u>Migration: Context and Current Status</u>. By the Digital Preservation Testbed. Dec 5, 2001. Accessed on Feb 22, 2007 <http://www.digitaleduurzaamheid.nl/bibliotheek/docs/Migration.pdf>.

<u>NINES</u>. 2007. NINES. Accessed on Feb 20, 2007 <http://www.nines.org/index.html>.

Noetzel, Lisa. "Introduction." <u>Voces del Caribe</u>. 2006. Accessed on Feb 16, 2007 <http://vocesdelcaribe.org/language.shtml>.

<u>The Online Archive of California</u>. 2007. The Regents of The University of California. Accessed on Feb 20, 2007 <http://www.oac.cdlib.org>.

PDF/A Joint Working Group. "ISO 19005-1:2005 PDF/A-1". Jul 10, 2006. PDF/A Joint

Working Group. Accessed on Feb 22, 2007 <http://aiim.org/documents/standards/19005-

1_FAQ.pdf>.

Project MUSE. 2007. The Johns Hopkins University Press & The Milton S. Eisenhower

Library. Accessed on Feb 20, 2007 <http://muse.jhu.edu>.

Python Programming Language – Official Website. 2006. Python Software Foundation.

Accessed on Feb 21, 2007 <http://python.org>.

Rhen, Brad. "Libraries Cope With Budget Cuts." The Lebanon Daily News. Apr 7, 2006.

Accessed on Feb 16, 2007 <http://www.ldnews.com/fastsearchresults>.

Snow, Sue. 2006. Unknown title little boy painting. CaribbeanPaintings.com. Accessed

on Feb 26, 2007 <http://www.caribbeanpaintings.com/2006/06Snow.jpg>.

"A Surge in Searches (Brief article)." PC Magazine 25.6 (Apr 11, 2006): 21 (1). *Student

Edition*. Thomson Gale. Accessed on Feb 23, 2007 <http://find.galegroup.com>.

Thompson, Bob. Church Day. CaribbeanPaintings.com. Accessed on Feb 26, 2007

<www.caribbeanpaintings.com/BThompson/pages/ChurchDay.html>.

Thompson, Bob. Fisherman's Bay. CaribbeanPaintings.com. Accessed on Feb 26, 2007

<www.caribbeanpaintings.com/BThompson/pages/FishermansBay.html>.

Thompson, Bob. Haitian Landscape. CaribbeanPaintings.com. Accessed on Feb 26, 2007

<www.caribbeanpaintings.com/BThompson/pages/Haitian.html>.

Weiss, Laura B. "It's bleak, north of the border: almost every Canadian school has a

library, yet the average budget is a mere $2,000." School Library Journal 51.6 (Jun

2005): 19(1). *Professional Collection*. Thomson Gale. Accessed on Feb 16, 2007.

<http://find.galegroup.com>.