

# K-12 Web Archiving Project

## Project Description

There is a growing awareness among libraries, archives, and other cultural institutions that much of the intellectual and cultural output of the nation, and indeed the world, now resides only on the Web. Perhaps more than any other type of content today, Web materials comprise an incredible variety of viewpoints and cultural currents. For more than a decade, a small number of individuals and institutions have recognized the importance of archiving and preserving the often transitory digital cultural artifacts that are distributed over the Web. But so far, the vast majority of decisions about what Web sites will live into the future have been made by adults, and reflect adults' sensibilities about what constitutes the important stuff of history.

Internet Archive, the Library of Congress and California Digital Library (CDL) are collaborating to launch a project that explores archiving the Web from the perspective of adolescents. The project will involve using the Archive-It service from the Internet Archive (See Appendix 1) to create "time capsules" of born digital content available via the Web chosen by students to represent their world. By allowing students to identify sites that will be preserved for the long-term, the pilot will give them a chance to identify and document *their* cultural history and the world that's important to them. Unlike time capsules of tangible objects, which usually remain hidden for decades or centuries, the resulting Web collections will be immediately visible and publicly accessible, with full text search for study and analysis.

Stimulating students to think about history, in context of their own lives, will enrich their study of history, provide an opportunity to actively engage in selecting the matter of history in the future, and help them begin to grasp the tremendous challenges presented by a world in which information can be both generated and removed with a key stroke.

## Participants

- Students from ten high schools, each from a different geographic region of the country. (Middle schools may also be considered.)
- The group of students at each school may represent a class, student club, or other selected group.
- A teacher, who will oversee the students' work, will coordinate each group. The teacher will also select a student to act as the lead on the project: attending training sessions, monitoring crawls, participating in the online wiki discussions, etc.

## Selection Criteria

- Teachers and students must be comfortable using the Web and Web-based tools.
- Teachers and students must be able to dedicate several hours per week to the project initially, and several hours per month thereafter (see timeline below).

- Teachers and students must have an interest in history and technology.
- Teachers and students must have a strong understanding of primary sources and the role they play in documenting cultural heritage.

## **Timeline**

- September 30: Deadline to apply to the program
- Week of October 1: Selected schools are notified (up to 10)
- Week of October 8 through October 15: Archive-It training and Primary Source training (1.5 hours)
- Week of October 20: kick off meeting via web or teleconference with 10 schools (teacher and a student representative), Internet Archive, California Digital Library and Library of Congress (2 hours)
- Week of October 26: participants select websites, write relevant text descriptions, and begin crawling process (2-3 hours)
- November through April: check in conference calls and e-mails; continue crawling; participate in online forum (2-3 hours per month)
- April 30: last day of crawling
- May: evaluation process and completion (2-3 hours)
- End of May: final meeting via web or teleconference for all participants (2 hours)

## **Collection Parameters**

- Collections to consist of Web sites that are primary source sites that reflect the students' lives and interests. (See examples in Appendix 2)
- Selection criteria are purposely left broad.
- Schools will be responsible for selecting their websites and curating their collections. The users can create as many collections as they would like, with three (3) collections active for crawling at any one time.
- Users can archive as many websites as they would like, with 300 websites being crawled at any one time.
- Each school's account can archive up to 10 million documents and one (1) terabyte.
- It is the responsibility of the teacher/student representative not to reach the limit before the end of the project. (Internet Archive partner specialists will advise the users when they have reached 80% of their limit. This information is also readily available in the web application.)

## **Student Responsibilities**

- Identify primary source sites for the "time capsule."
- For each site chosen, provide a one sentence description of the site and a one sentence description of why the site is reflective of the students' interests and lives.
- Work with teacher to provide an overall description of the "time capsule/collection."

- Collaborate and participate in discussions with other groups using online wiki.
- Fill out a questionnaire at the end of the project.

### **Teacher Responsibilities**

- Facilitate the student collection, assisting students in establishing a process for deciding on the sites to be included, and ensuring the content selected does not include pornography, discriminatory materials, or gratuitous violence.
- Work with students to prepare an overall description of the “time capsule” or collection.
- Enter URLs for selected sites into Archive-It.
- Collaborate and participate in discussions with other groups using online wiki.
- Fill out a questionnaire at the end of the project.

### **Continuing (after May 2009)**

- Teachers are encouraged to have students revisit the sites that they have archived to see which have disappeared, and how the sites have changed over time.
- Teachers may be called upon for input into developing a broader range student-based Web capture program.
- Teachers and students may be called upon for testimonials of their experiences.

### **About the Partners**

It is at the very heart of the missions of the Internet Archive, CDL, and the Library of Congress to serve as stewards of content that has current and future value to the nation. In addition, all three institutions are committed to serving the K-12 community, particularly as it involves the use of primary source material in the understanding of history.

#### **Internet Archive**

Archive-It is a subscription service developed in 2005 by the Internet Archive. The Internet Archive (<http://www.archive.org>) is a 501(c)(3) non-profit that was founded in 1996 to build an 'Internet library,' with the purpose of offering permanent access for researchers, historians, and scholars to historical collections that exist in digital format.

Internet Archive has the largest public web archive, comprising 95 billion pages, and over 65 million websites in 37 languages for a total of 3 petabytes.

In late 1999, the organization started to expand to include non born digital material. Now the Internet Archive includes over 200 terabytes of texts, audio, moving images, and software along with our web archive.

In 2003, The Internet Archive co-founded the IIPC, the International Internet Preservation Consortium, (<http://www.netpreserve.org/about/index.php>) along with 12 International Libraries, recognizing the importance of international collaboration, standards and best practices for preserving Internet content for future generations.

## **Library of Congress**

For almost fifteen years, through its public Web site and its digitization and outreach programs, the Library of Congress has been a leader in the campaign to integrate more primary source material into school curricula. The Library's site offers a multiplicity of resources for teachers and students through its Learning Page, in addition to its flagship American Memory collections. Beginning in 2001, the Library has also been involved in archiving Web sites. This pilot synthesizes elements of its current Teaching with Primary Sources and Web site preservation programs.

## **California Digital Library (CDL)**

CDL is the University of California's systemwide digital library that supports all 10 UC campuses in creating, managing, sharing and preserving their collections. In addition, through the Calisphere website (<http://www.calisphere.universityofcalifornia.edu/>), CDL provides free access to primary source materials designed around the needs of K-12 educators. Calisphere offers more than 170,000 primary sources are from the libraries and museums of the UC campuses and cultural heritage organizations across California.

## **Contact Information**

- Technical support: Archive-It
  - Molly Bragg [mbragg@archive.org](mailto:mbragg@archive.org) 415-561-6799 x6
  - Kristine Hanna [kristine@archive.org](mailto:kristine@archive.org) 415-561-6799 x5
- Program support: Library of Congress
  - Cheryl Lederle-Ensign [cled@loc.gov](mailto:cled@loc.gov) 202-707-3550
- Program support: California Digital Library
  - Rosalie Lack [Rosalie.Lack@ucop.edu](mailto:Rosalie.Lack@ucop.edu) 510-987-0414

## **Appendix 1**

Internet Archive's subscription service, Archive-It, enables institutions to harvest, manage, search and preserve born digital web based content.

The Web based application features a user friendly interface allowing subscribers to easily set up and catalog collections, monitor crawling and make adjustments post crawl. Collections are hosted at the Internet Archive data center, accessible to the public with full text search.

Archive-It is designed to fit the needs of many types of organizations and individuals, including state archives, libraries, academic institutions, non profits, museums, historians, and independent researchers. The over 50 partners include: state archives, university libraries, federal institutions, state libraries, non-government non-profits, museums, historians, and independent researchers.

For more information regarding the Archive-It service, please visit [www.archive-it.org/onesheet.pdf](http://www.archive-it.org/onesheet.pdf)

## **Appendix 2**

Examples of the types of sites that may be appropriate for this project:

<http://inprogress.typepad.com/kidsforobama/> - child created and maintained blog encouraging other young people to voice their thoughts in the upcoming elections

<http://www.myspace.com/tom> - "Tom" - the creator of myspace.com - is one of that website's most ubiquitous profiles with 220,192,046 friends. MySpace users (many 18 and under) leave thousands of comments each day and check for updates to the website.

Examples of the types of sites that show an adult perspective on what kids want, or are compilations of sources and are NOT the target for this project:

<http://clerkkids.house.gov/laws/InspectLaw/index.html>

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

<http://encarta.msn.com/>