

World Wide Web History Project'

<http://1997.webhistory.org/project/archives.html>

## About the Archives:

### *Building a Digital Library*

When completed, the World Wide Web History Project's **digital library** will serve as a public repository for all source material related to the origins and development of the Web, including historic video, audio, documents, and software.

Visitors will be able to connect to the Project's library over the Internet via a Web client and browse through the various media in the repository, using either a traditional forms-based HTML interface or a graphic timeline that will allow them to visualize the contents of the entire repository as a series of events.

In addition to the [project members'](#) own video and audio tapes of interviews, materials already collected for the archives include tens of thousands of pages of electronic mail, papers, and digital documents, hundreds of photos, and a large portion of the software that played key roles in the Web's development.

### *The Importance of Historical Systems*

Since the birth of the Internet, many different classes of collaborative software have evolved over time: chat systems, news readers, and mailing list servers, to name a few. To date no Internet-based software has been written that allows people to create personal, community and research histories using the full potential of hypermedia.

The ability to view and manipulate histories and see the relationships among events allows one to put things in context, and can help foster a true sense of community and progress. In addition, such systems have the potential to greatly accelerate research when populated with rich content. Very little progress has been made in the development of popular time-based interfaces, even though they offer a natural way to visualize information.

There have been many problems with previous digital library efforts: they have been bulky and difficult to navigate, set up, and administer; they have suffered from a lack of agreement on development standards; they have not been well documented; and they have not been written with personal use and set up in mind, hindering their popularity.

The digital library system developed as part of the Project hopes to alleviate some, if not most of these problems. It is one of the hopes of the Project that this system will be used not just as a historical resource, but also as a valuable research and development tool.

## Footsteps

*...in the Silicon Sands of Time*

The digital library system currently being developed for the Project is called **Footsteps**. The system will allow media to be entered, catalogued, and searched using multiple criteria.

### *Design and Specifications*

Major design goals for Footsteps include:

#### **Cross-platform capability**

The system is designed to run on a multitude of different platforms. This way, anyone can use the software to view, edit, and record histories, regardless of platform. Most of the system will be written in Java, allowing for as much platform independence as possible. The use of Java also meets other goals, such as having an object-oriented design and an internationalizable interface. Where speed is a concern, we will create native (versus interpreted) binaries from the Java source code using popular Java to C translation tools, though such translation will not be necessary to run the system.

#### **Open standards compatibility**

To avoid proprietary solutions which may quickly become obsolete, Footsteps will encode and store media using a number of international and de facto standards endorsed by the ISO and the World Wide Web Consortium. Although few de facto standards exist for the development of digital library systems, the architecture will make use of methods and ideas developed by the U.S. Library of Congress, the OCLC, and various digital library consortia.

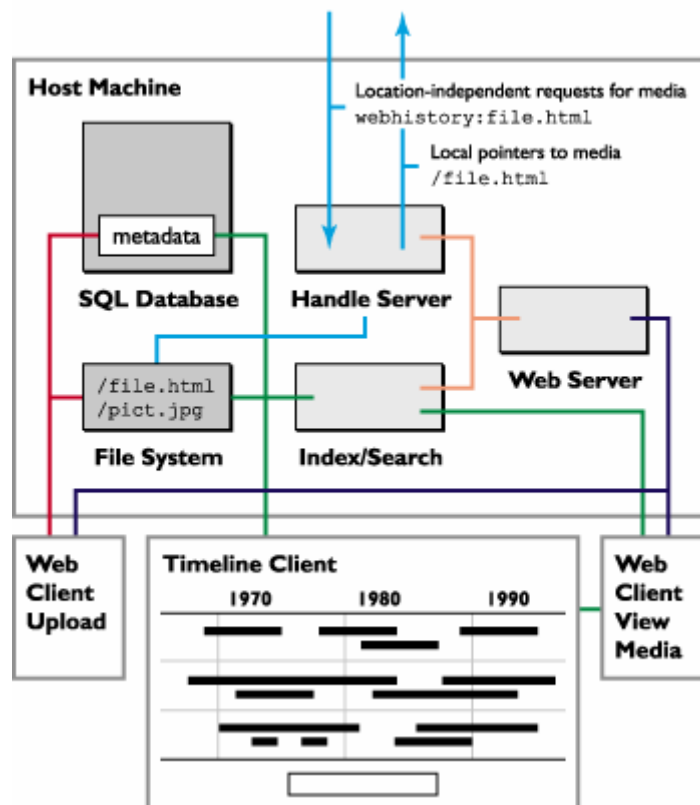
#### **Ease of use and ease of access**

Footsteps will be able to present a hyperlinked, chronological view of events stored within the system. Media will be linked to these events and cross-referenced with other media, enabling one to better understand and discover relationships between items in the repository. Both standard HTML and graphic interfaces will be provided so that even the simplest Web clients can view the timeline.

#### **An event-centric model**

Unlike many digital library interfaces which focus on collections of media, the Footsteps interface is centered around the presentation and selective filtering of events sorted by time. Although more traditional interfaces to the repository's media will be implemented (such as media-based catalog searches), the system is intended to take advantage of collections of events and the media associated with those events, and not just collections of media.

## THE FOOTSTEPS SYSTEM ARCHITECTURE



### *The General Architecture*

First, we've designed a way to **store media in the native filesystem**. The system may make use of ISO CD-ROM naming and hierarchical standards to some extent, so the entire filesystem can be placed on a DVD or other removable media in a platform-independent manner. Media is organized using methods developed at the U.S. Library of Congress. For each image in the system we expect to offer at least three versions - one suitable for online browsing (lossy, 72dpi), one for midrange use (lossy, at most 300 dpi), and one for archival purposes (lossless, at least 2500 dpi).

Second, **metadata storage and manipulation** is provided through a SQL database, which can be replaced with other SQL-compliant databases as needed. Database communication will be implemented with JDBC. The metadata format follows that of the Dublin Core, which specifies a simple metadata set that in one sense is a subset of USMARC, the U.S. machine-readable bibliographic specification. Each piece of media in the system is classified as an event, person, place, or thing. The schema and relationship structure makes use of ideas embodied in Taligent's CommonPoint operating system. Annotations are provided for in the schema and can be attached to any media, including other annotations.

Additional software to **index and search** media in the filesystem is used in conjunction with the database's capabilities to provide the visitor with the ability to cross-reference and browse repository media according to a number of criteria.

A **handle server** translates external requests for media to local mappings. This allows the media in the repository to be referred to in a platform and location independent

manner. This also allows the entire system to be replicated and mirrored, and still look consistent to the outside world no matter where it exists on the Internet.

An **interface to submit new media** is implemented with a browser interface and will likely allow one to use an HTML-based form and the HTTP POST method to submit files.

An **interface to browse media** is also implemented with a browser interface and will allow one to search, cross-reference, and browse through the media stored in the system using a number of criteria.

The **timeline interface** is used by visitors in tandem with the browsing interface. The timeline provides a chronological view of events in the system; each event is linked to media concerning that event. Events can be annotated collaboratively and grouped into different topic rows. Event text can be searched using this interface. In addition, it may be possible to view trails of influence and/or development among events. Research is being done into using distributed objects for client/server communications via CORBA and other methods so that timeline objects can be shared and manipulated easily by a variety of clients.

### *Development Schedule*

The core architecture and schema has been written, and the interface and backend programming is slowly progressing. There is no set date for initial public launch.

### *For More Information...*

Please contact Kevin Hughes ([kevinh@webhistory.org](mailto:kevinh@webhistory.org)), the History Project's digital library architect.

In addition, you may wish to go to these sources dealing with various related standards and technology:

A [Digital Libraries Resource Page](#)

[The USMARC Home Page](#) at the [Library of Congress](#)

[The Dublin Core Home Page](#) at the [OCLC](#)

The LOC's [American Memory Project White Papers](#)

The [Lifestreams Project Page](#)

[UI for the LOC National Digital Library](#) by the University of Maryland's Human-Computer Interaction Laboratory