

# Building a Virtual Pacific Neighborhood Cultural Exposition

By David L. Wasley

University of California, Berkeley

At the January 1994 meeting of the Pacific Neighborhood Consortium, it was proposed that the PNC membership use Internet network and information server technology to develop a "Virtual Museum of the Pacific." However, the term "museum" implies perhaps a collection focused on historical information. This paper develops the idea of a virtual "Cultural Exposition" where the widest possible range of information -- both historical and current -- can be displayed and shared with neighbors on the Internet.

I wish to emphasize at the beginning that this is merely a proposal. I hope that it will be interesting to the Pacific Neighborhood Consortium membership and that it will generate further discussion and ideas. No one person or group can collect the information or create all the exhibits envisioned. If realization of this proposal is to be successful, it must be a collaborative effort with active participation from all PNC members.

## **Background on the Technology**

Before discussing the practical details of this proposal, I think that it is useful to have a common understanding of some of the technology that might be used. It is the availability of powerful and inexpensive client and server software for personal computers that makes possible many of the ideas presented below.

In 1989, computer scientists at the European Laboratory for Particle Physics (CERN) began developing a "language" for defining complex documents in terms of a set of components with specific attributes. This language includes definitions for complex formatted text, various kinds of graphical images, and "hot spots" or "buttons" that can be used for feedback from a reader to cause specific actions to occur, such as jumping to another portion of the document. Today the W3 Consortium administered by the Massachusetts Institute of Technology (MIT) in cooperation with CERN continues this work.

A key element of this language is the ability to define links between one document and others. Linked documents need not reside on the same computer but may reside on other servers anywhere on the Internet. Because of this concept of distributed but

interlinked documents and servers, the CERN and MIT system has come to be known as the "World Wide Web" or simply WWW (or W3).

A WWW "document" may be any combination of text, graphical images, sound, moving images, "buttons", and forms into which a user can enter data. All components of a document have a well defined format so that any document created on one computer can be decoded and displayed on a wide variety of other computer and workstation platforms. Already, inexpensive software exists for displaying WWW documents on Macintosh, IBM-type PC and Unix workstations. Of course, the particular capabilities of a given workstation might limit the display of the more complex document types such as sound or moving images.

The language that allows definition of complex WWW documents is known as the "HyperText Markup Language" or HTML. The term "hypertext" refers to the concept of "jumping" from one document or document section to another. The term "markup" refers to the embedded commands within the document that are not seen by the reader but are used by the client "browser" software to understand index information or the way in which a document should be displayed. HTML documents can be created using a variety of common tools on most types of workstation computers.

There are software packages that implement translation from common image and document formats to HTML. For example, many personal computer word processing software programs can produce files in "Rich Text Format" (RTF) which is a commonly understood standard for formatted text. A public domain software package called "RTFtoHTML" converts documents in this form to HTML by using heuristic assumptions about components of the document.

In addition to the language that defines complex documents, a protocol has been defined for transferring such documents between servers and end-user client computers. This protocol is called the "HyperText Transfer Protocol" or HTTP. HTTP is an application layer protocol that runs on top of Internet TCP/IP data transport. Using HTTP, a WWW client sends an HTTP request to a WWW server and the server responds with either the requested document or an appropriate error message.

A third key element in the WWW technology is Uniform Resource Locators or URL's. A URL defines an information resource by type, server location, and specific file identifier. All hypertext links in HTML documents are in the form of URL's. For example, the WWW server for the Museum of Paleontology at the University of California, Berkeley, has a URL of

<http://ucmpl.berkeley.edu/subway.html>

which indicates that it is accessible using the HTTP protocol; it is located at the Internet node named "ucmpl.berkeley.edu" and the document is in a file named "subway.html" (which by convention indicates that it contains a document constructed with the HTML definitions). Other protocols that are supported typically by WWW client software include FTP, gopher and TELNET.

Today thousands of WWW servers exist on the Internet, many of which are excellent examples of the use of computer graphics and creative writing to make interesting and very useful HTML documents. All of the documents in the Appendix to this paper were acquired from WWW servers around the world. Please note several very nice examples of top level documents (or "home pages") that were produced by Chiang Mai University and Chulalongkorn University.

By combining the power of complex document construction, standards for the format of document elements, a consistent way of locating such documents, and a standard protocol for transfer of such documents over the Internet, we have available to us a simple but elegant framework for sharing a wide variety of information among PNC members and with the educational community.

### **A Proposal for a Pacific Neighborhood Cultural Exposition**

One of the founding principles of the Pacific Neighborhood Consortium is the sharing of information in electronic form using our evolving networks. Information can be as specialized as Buddhist texts or as general as cultural and historical databases. One of the results of this exchange can be an increased sense of community - the creation of a spirit of "neighborhood." With this in mind, I propose that the PNC membership organize a project to actively promote and develop a coherent Pacific Neighborhood Cultural Exposition (PNCE) to bring into focus the diverse interests and resources found among our institutions and within our countries.

The paradigm that guides the organization of information under this proposal is that of an exhibition hall where galleries are created by PNC members interested in participating. I envision a wide range of "visitors" to this virtual exhibition, from students of all ages to those with mere curiosity. The exhibition hall would be "virtual" of course, consisting of one or more WWW servers and a coordinated set of HTML documents.

Each PNC member would be encouraged to create HTML documents and other databases relating to their country, region, or areas of interest. Information could be of any type or format that is desired by the organization creating it but there should be some consistency across at least the top level documents so that visitors to any gallery will find the next gallery familiar and there will be a feeling of cohesiveness about the whole set of exhibits.

The PNC itself would maintain a "virtual PNCE Visitor Center" that would help visitors find exhibits of interest. In technical terms, this would be the "home page" for the PNCE documents. For example, the home page might be a map of the Pacific basin where each country on the map would have a hypertext link to a more detailed map of that country which in turn might have links to yet more detailed maps or other forms of information indices. Text and icons on each "page" would link to HTML documents of interest at that level. Search capabilities could be incorporated to bypass this hierarchy.

Some of the "galleries" that might be created for each country could include:

Visual arts collections with both images and text descriptions.

Ethnographic collections including text, images, audio, and perhaps film clips of regional peoples, both current and historical. The University of Alaska, for example, has an on-line database of this type relating to some of the Eskimo peoples.

Architectural displays showing historical and current monuments and population centers.

Maps and other geophysical data.

Image and text descriptions of the member institutions, the programs it offers, research it is supporting, etc.

Regional or indigenous music collections with accompanying images of instruments and text describing the significance of the work.

Writings and other important cultural documents from the region. For example, some of the great poetry collections of China or Japan might be put in machine readable form (Kanji or other standard format) by a PNC member institution and made available in the Virtual Museum of the Pacific for interested scholars everywhere.

The diagrams below show one way in which the PNCE might be organized. Since any given document can be linked to from any number of places, there could be many different logical views of the total set of documents. For example, an HTML document devoted to economic data might link to specific individual documents to retrieve data from each country. These same specific documents might also be found from the HTML document describing "Industry, Economics and Technology" under the country data document. I believe it would be worthwhile to have consistent organization within a generally useful logical view of these documents, at least down to the types of categories I've shown. Additional categories and views can be created as needed.

Another exciting possibility is to maintain parallel sets of documents in various major languages. For example, the Chiang Mai University home page is available in Thai as well as English. There are technical problems in doing this and it would take a set of people willing to devote time to translating such documents but I believe it would result in a very meaningful resource.

In Figure 1 I have included a "Concession Stand." The concept of this is to provide a place where common documents, software or data might be stored for retrieval by "visitors." Such items might include new versions of public domain WWW clients, sound or image decoding software, and special fonts needed in order to display PNCE documents. It should be possible to offer compressed versions of some of the PNCE documents that could "taken home" by "visitors."

### **A Practical Implementation Strategy**

The major impediments to implementation of this proposal include:

staffing and funding

location of WWW server(s)

identification or creation of information resources

In addition, several technical problems remain that a PNC working group might address. In particular, the representation of non-English text in HTML documents needs to be more standardized. Unfortunately non-English character fonts are not uniformly available with WWW client software. PNC might undertake to make such fonts available and might also identify any additional standardization needed to ensure that non-English HTML documents are constructed in a well defined and compatible way.

Other issues to be considered include the possible need for access control in some cases, intellectual property rights, and scaling problems if the PNCE becomes highly successful.

Staffing might be coordinated through the PNC Secretariat or through a PNC member institution. Initial staffing would include a project leader with library or information management skills and at least one half-time technical assistant. Additional staffing in the development stage might come from PNC member institutions on a voluntary basis. Funding would be needed to cover dedicated staff as well as server platform(s).

Since server locations and the logical topology of the servers would not be visible to "visitors", the actual location of HTML documents is somewhat arbitrary. In order to prototype the PNCE, it should be possible to start with one or at most a few servers containing HTML documents or databases transferred to them by PNC member institutions. As the PNCE evolves, each member institution can decide whether they wish to operate their own server or make use of the centrally managed server(s).

A practical problem will be to find efficient ways to make the PNCE available given the constrained Internet network capacity currently in place between some locations. Transferring HTML documents can require a large amount of network capacity, especially when images or sound are included. If every "visitor" retrieves the same document(s), this can represent a significant load on wide area networks as well as servers.

In order to work around this problem, I propose that an early task for the project should be to develop a method of caching automatically copies of all PNCE documents at strategic locations around the Pacific basin. As part of this caching process, embedded URL's would have to be modified automatically as well. End users in each region could then be guided to use the regional copy of the PNCE Home Page instead of relying on the main central one. In this way, the same set of "exhibits" can be made available to everyone with limited use of constrained parts of the network. It also would spread the processing load across multiple WWW servers.

Clearly many on-line information resources already exist as HTML documents and can be linked into the logical structure proposed above. I leave it to the PNCE project manager to address the issues of locating these resources or creating new information resources.

The Pacific Neighborhood Cultural Exposition envisioned in this proposal will be meaningful and successful only if all PNC members contribute ideas about what resources should be accessible and are willing to help develop them and make them available to all of us.

### **References to WWW Documents and Resources**

There are, of course, many books on the World Wide Web and related information server technologies. All references included here are HTML documents describing WWW technology and are available over the Internet. I have attached paper copies of most of the top level documents mentioned here as Appendices but of course the many embedded HyperText links that are not expanded. The software for WWW clients mentioned below is available using standard FTP clients with anonymous user access.

The WWW server at CERN contains a wealth of information on the history, technology and use of WWW. Some overview documents include:

The CERN WWW project

URL <http://info.cern.ch/hypertext/WWW/TheProject.html> [click here](#)

WorldWideWeb - Summary

URL <http://jinfo.cern.ch/hypertext/WWW/Summary.html> [click here](#)

Kevin Hughes at Enterprise Integration Technologies (EINet) has authored a set of illustrated, brief documents on WWW technology. I have included the following

excerpts:

Guide to CyberSpace

URL <http://www.eit.com:80/web/www.guide/> [click here](#)

What is WWW?

URL <http://www.eit.com:80/web/www.guide/guide.01.html> [click here](#)

What is hypertext?

URL <http://www.eit.com:80/web/www.guide/guide.02.html> [click here](#)

How does WWW work?

URL <http://www.eit.com:80/web/www.guide/guide.10.html> [click here](#)

What software is available?

URL <http://www.eit.com:80/web/www.guide/guide.11.html> [click here](#)

The Workstation Support Services group at the University of California, Berkeley maintains a set of helpful information on WWW, including hints on setting up servers and creating documents. You can visit the UCB Web Corner at:

URL <http://wss-www.berkeley.edu/Webcorner/webcorner.html> [click here](#)

**Software:** Some of the more popular WWW client software packages for personal computers are available freely over the Internet. The NCSA distribution includes many "helper applications" that decode non-text documents so might be a good starting point.

**Mosaic**, developed by the National Center for Supercomputer Applications (NCSA) at the University of Illinois in the US, was one of the first full featured, graphical clients for personal computers.

URL

<http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/NCSAMosaicHome.html> [click here](#)

The FTP source for NCSA Mosaic is <ftp.ncsa.uiuc.edu>

Directory locations for specific platforms are Web/Mosaic/Mac

Web/Mosaic/Windows

Web/Mosaic/Unix

NetScape is the first product from Netscape Communications Corporation but is free to academic institutions. It includes easy access to WWW document indices.

URL <http://home.mcom.com/home/welcome.html> [click here](#)

The FTP source for NetScape is ftp.mcom.com

The software is located in the directory netscape

EINet is a commercial provider of network communication and information services. EINet is a trademark of Microelectronics and Computer Technology Corporation (MCC). EINet has made WWW clients available for Macintosh and IBM-type PC platforms.

URL <http://www.einet.net/EINet/MacWeb/MacWebHome.html> [click here](#)

URL <http://www.einet.net/EINeT/WinWeb/WinWebHome.html> [click here](#)

The FTP source for EINet clients is ftp.einet.net

Software is located in directories inet/mac/macweb

einet/pc/winweb

A selection of WWW server software is described in the document "**W3 Server Software**" maintained by the W3 Consortium.

URL <http://www10.w3.org/hypertext/WWW/Daemon/Overview.html> [click here](#)

## **POINTERS TO APPENDICES**

### **The World Wide Web**

URL <http://info.cern.ch/hypertext/WWW/TheProject.html> [click here](#)

### **WorldWideWeb - Summary**

URL <http://info.cern.ch/hypertext/WWW/Summary.html> [click here](#)

### **Entering theWorld-Wide Web: a Guide to Cyberspace**

URL [http://www.eit.com:80/web/www\\_guide/](http://www.eit.com:80/web/www_guide/) [click here](#)

### **What is hypertext and hypermedia?**

URL <http://www.eit.com:80/web/www.guide/guide.02.html> [click here](#)

### **How does the Web work?**

URL <http://www.eit.com:80/web/www.guide/guide.10.html> [click here](#)

### **What software is available?**

URL <http://www.eit.com:80/web/www.guide/guide.11.html> [click here](#)

### **The Web Corner**

URL <http://wss-www.berkeley.edu/Webcorner/webcorner.html> [click here](#)

### **NCSA Mosaic**

URL <http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/NCSA/MosaicHome.html>  
[click here](#)

### **Welcome to Netscape**

URL <http://home.mcom.com/home/welcome.html> [click here](#)

### **MacWeb Version 1.00**

#### **ALPHA3 07Dec94**

URL <http://www.einet.net/EINet/MacWeb/MacWebHome.html> [click here](#)

### **WinWeb Version 1.0**

#### **ALPHA2.1 11Oct94**

URL <http://www.einet.net/EINet/WinWeb/WinWebHome.html> [click here](#)

### **W3 Server Software**

URL <http://info.cern.ch/hypertext/WWW/Daemon/Overview.html> [click here](#)

**CMV (WorldWideWeb Home Page of Chiang Mai University, Thailand)**

URL <http://www.chiangmai.ac.th/> [click here](#)

**Chulalongkorn University**

URL <http://www.netserv.chula.ac.th/> [click here](#)