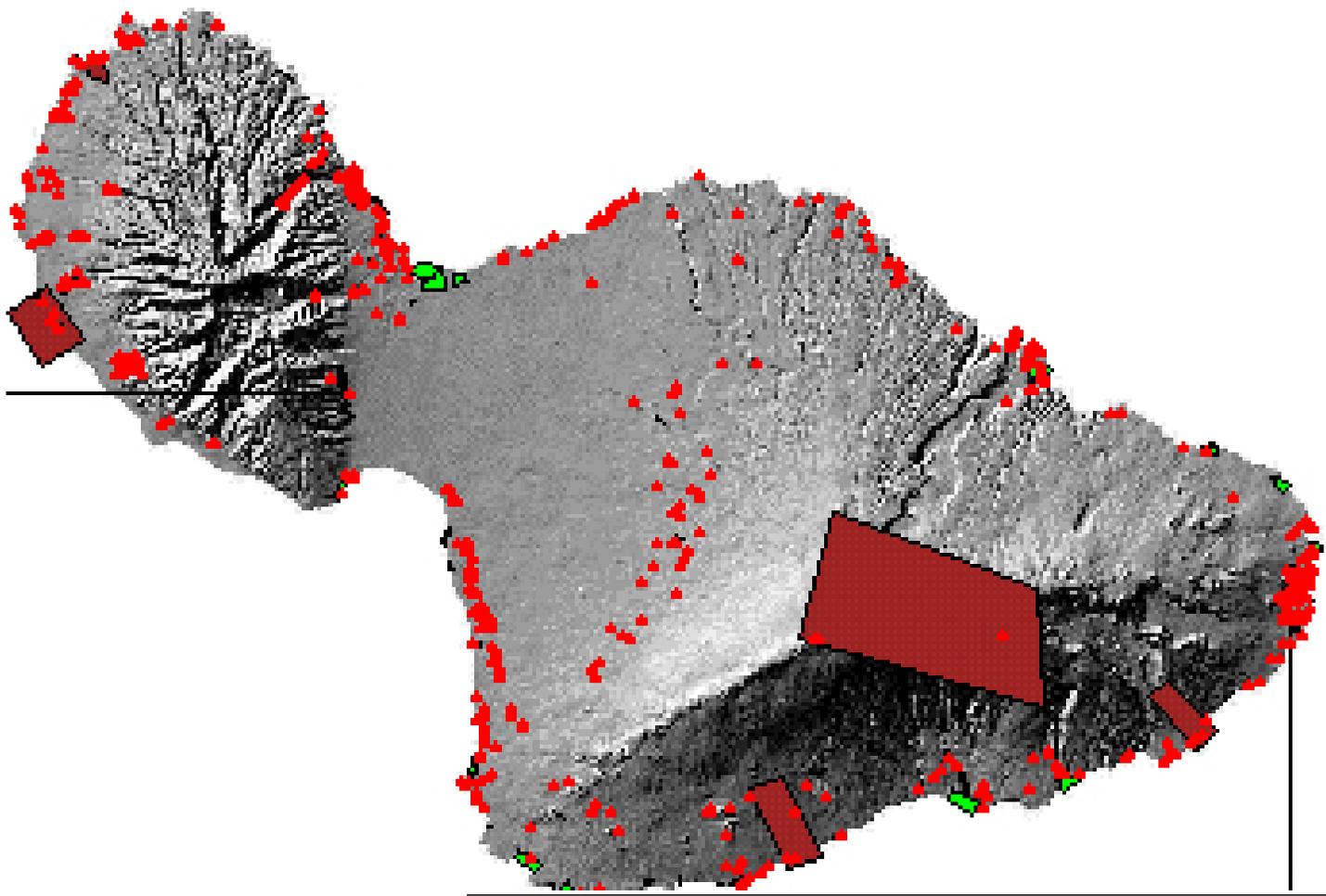# Migration of the Hawaii State Historic Preservation Division Databases to the World Wide Web | 1998-12

Hawaii Department of Land and Natural Resources

MIGRATION OF THE
HAWAII STATE HISTORIC PRESERVATION DIVISION
DATABASES TO THE WORLD WIDE WEB

HISTORIC PRESERVATION DIVISION
DEPARTMENT OF LAND AND NATURAL RESOURCES
STATE OF HAWAII
AUGUST 1998

ABSTRACT

This report presents the results of a project named "Migration of the Hawaii State Historic Preservation Division Databases to the World Wide Web" The project was undertaken by the Historic Preservation Division (SHPD), Department of Land and Natural Resources, State of Hawaii with support from the National Park Service's National Center for Preservation Technology and Training (NCPTT grant MT-2255-6-NC-032).

The work identified and developed an economical approach for providing World Wide Web access to information about historic resources developed and maintained by the SHPD. Bibliographic information on a wide range of reference material maintained in the SHPD library, and data from the Division's Geographic Information System have been made accessible through the WWW.

Various software and "freeware" combinations were reviewed during the course of the work. The low-cost ($175), platform-independent, mSQL (Hughes Technologies) program was selected for use as the UNIX database server Access to GIS data uses "JShape", a "freeware" package that uses "JAVA" technology to access and display GIS data.

Issues regarding the sensitivity of historic resources to enhanced access were addressed by the project and inappropriate data has been identified and excluded. Database tools for the creation of WWW public notification reports for historic preservation reviews have been developed, however, the process has not been implemented pending receipt of SHPD level edited data for display.

The WWW site for the project is presently:

"http://mano.icsd.hawaii.gov/~ckomoek"

A permanent URL reference for this page will be assigned as part of the State of Hawaii's World Wide Web development in the near future.

INTRODUCTION

The State of Hawaii Historic Preservation Division (SHPD) manages historic resources in Hawaii through a multifaceted program that includes the development of an inventory of historic sites and a library of reports concerning cultural resources in Hawaii. The Division is actively developing computer database and Geographic Information System (GIS) resources that provide enhanced access to information about the historic resources in the islands While these resources are valuable management tools for the Division, public access to these resources has been restricted due to technical problems, costs, and concern about the security and sensitivity of historic sites. This project is intended to promote the public dissemination of these resources.

Recent developments in information systems like the Internet provide an opportunity to develop public access for these kinds of data at reasonable cost, while protecting sensitive materials. This project focuses on the development of a "data rich" World Wide Web site for historic resources in Hawaii using economical computer software and hardware resources. The project also addresses the public's interests and needs for this data and examines technical solutions to policy issues.

Several considerations were addressed in the design of the SHPD WWW site. Cost was a primary factor but software portability was also an important consideration as our implementation was to be developed, tested, and brought online using extant hardware and software platforms. Good software portability also allows others to use the approach with less initial effort .The design was heavily influenced by the nature of the data, which consisted of large relational databases and GIS data sets.

Following this introduction this report is divided into three parts:

Part 1. Bibliographic Database Application

Part 2. GIS Application

Part 3. Data Selection

Part 4. Summary of Implementation

Part 5. References

HISTORIC PRESERVATION BIBLIOGRAPHY

The development of the Historic Preservation Bibliographic database was based on public domain database software that is intended to run in a UNIX environment. This project utilized freely distributed software for several reasons Cost was the primary consideration, but this solution also addressed a concern with portability.. The software needed to be compatible with the host site, a SPARC station running SOLARIS, and with our test site, a 486 running LINUX. Software distributed as "freeware" frequently include the source code .This allows the users to re-compile programs for their target computer environment, and facilitates the transfer of data and programs.

Database Selection

The large number of UNIX and non-commercial software packages available, provide range of software solutions. The primary task here was the selection of a combination that fit existing computer resources and data needs. The degree to which data is structured is important to the selection of a query server. If the data is loosely organized or of varying detail and length, a database that indexes blocks of text, such as "free-WAIS" is better . More conventional data from a flat file or relational database is better accessed using a structured database approach like Structured Query Language (SQL). In addition, the means by which the database is made accessible to the Internet is important. Some packages are far more adaptable in this respect and deserve consideration for their ease of implementation. In most cases, however, a large number of tools are usually available to simplify the "html" user interface.

Wide Area Information Service (WAIS) Evaluation

A number of databases were considered in the implementation. The first was "free WAIS". WATS is primarily for use with unstructured text and is part of the ANSI Z39.50 draft Standard. WAIS is able to quickly search large bodies of text for a given string, and to return the

best matches ranked in the order of relevancy Its speed is obtained by constructing an index file for use in processing queries, and search performance is very high. The single greatest advantage in using a WAIS server is that other servers can reference its indexes. In this respect it represents a more general solution for document access.

Many WAIS packages with CGI scripts for "html" interfaces are available, and are recommended for a WWW site which provides access to large amounts of loosely structured text. However, it is ill-suited for structured information which is better handled by a database that maintains data fields. For example, a WAIS search for information about an archaeological site identification number that contains the text string "1993", might include every site report published in "1993".

Another problem with the WAIS systems is that it must generate extensive index files in order to execute rapid searches. While these index files are efficient for searches, they do create some data management problems. The index files that are generated can be as large as the documents they reference, so storage space may become a problem. The WAIS indexing scheme also causes the search engine to rank the relevancy of a query against all entries in the data base and an entirely new index file must be generated each time the database is updated. This is a processor intensive task and may be a problem for databases with frequent updates.

A functionally similar alternative to WAIS, is a PERL package called SWISH. SWISH (Simple Web Interface for Humans) is probably the easiest approach for the initial development of a WAIS-type search engine. While the ease of implementation is attractive, its limitations are similar and is not practical for the SHPD data sets.

Structured Database Evaluation

The structured databases that were evaluated for the project include PERL programs, mSQL, and POSTGRESS.

PERL Software

Several PERL libraries provide tools to parse data into tables and columns, database queries based on a subset of SQL. One of the software that functioned well was Sprite. The

advantages of working with PERL scripts are the extensive libraries available for public distribution, that include CGI utilities for developing internet interfaces. The language handles different data types well and is able to extract information from data files. Well suited to the Unix operating system, it supports pipes, interprocess communications, and the programmer is able to access most shell processes. PERL is similar to C in syntax and although it is simpler, there is a substantial learning curve involved in using it.

PERL is a very flexible and useful tool, but there are some important concerns. Its tight integration with the operating system makes security a concern. Several hacks have been identified (subsequently patched) that allowed users to access system files. Aside from the holes that have been patched, there are also problems inherent to the common gateway interface . Certain characters are handled differently by the CGI and the PERL interpreter. A user can append a character string on a file request that will translate into a shell command in a PERL script. The potential for abuse is obvious, and while careful filtering of incoming file requests can reduce risk, it is a major problem . A large number of Internet service providers have disallowed the use of PERL scripts primary because of security concerns.

Another limitation with PERL is that it is an interpreted language. A database that uses a PERL script will be slower than one that uses a complied binary . The problem becomes compounded with the addition of an "http" interface . There is the potential for a large number of queries to be made more or less simultaneously and the additional overhead adds up as each request spawns a separate process. A "pure" PERL database solution is practical only on low volume sites.

In cases where performance concerns rule out a simple PERL database, the next step is to provide a dedicated query server. PERL may still be used as the scripting language for the CGI interface, but the query server is a complied binary usually running as a background process. The primary advantage of running a dedicated query server is that it lowers the amount of system resources needed to process each query. This translates into faster response times and is a much more robust solution in cases where the server must sustain a large volume of queries.

In this regard, the advantage of using an open ended language like PERL for the user interface is that the designer maintains a greater degree of flexibility as they can still incorporate

there own code or borrow from an extensive library.

mSQL Software

The largest draw back to using a dedicated query server is that most of the solutions being offered are costly, both in terms of hardware and software. There are a number of large companies competing in this area, offering an extensive array of query engines, "html" interfaces, and "middleware" to bridge existing applications and databases. Large software vendors have traditionally been able to license their products based on the number of end users. Now, a single server may be accessible by a very large number of users through the Internet and corporate "intranets". One of the results is that the price of a license for this kind of software is usually high.

There are several database engines, however, that are available in the public domain. mSQL is an example of a powerful software package that may be used by students and educational institutions free of charge. The site license for commercial and government use is reasonable ($175 at the time of this writing).

Some of the features that make mSQL suitable are its portability (source code is available), the relatively small amount of system resources needed, and the inclusion of software to query the database on different web servers. In addition, there is a community of developers working with mSQL who develop additional code and patches for the product.

One of the limitations of mSQL is that it does not support hierarchical search criteria and a series of searches much be conducted to select multiple criteria in multiple fields. The next version of mSQL is will support hierarchical search criteria and allow more complex searches to be performed.

POSTGRESS Software

POSTGRESS, a derivative of University of California's (Berkeley) INGRESS, is publicly available and has a number of advanced features. An object oriented program, it was originally developed as a test platform for database experimentation and is able to provide database access through the WWW. POSTGRESS has more features than mSQL, but is also more complex, has

more system resource overhead, and more software "bugs". Support for standard SQL queries is a relatively new development, since POSTGRESS has its own query language . There are advanced features that may make POSTGRESS attractive for some applications. Particularly interesting is the capability to perform spatial searches. The complexity of the software, because many of the features are in the "beta" development stage are major problems in its use.

BIBLIOGRAPHY DATABASE IMPLEMENTATION

The core data set for the database implementation is the SHPD library bibliography. This bibliography references almost all of the archaeological site reports available for Hawaii (presently about 5000) . The database contains a number of fields which where determined to be useful in constructing searches. The fields selected were author, title, year, and TMK (tax map key). The SHPD library number, sequentially assigned as the reports are entered in the database, was also included as a searchable field, since it serves as a key field in the relational structure of the bibliography database.

The database software selected was mSQL, with the CGI interface provided by a program named "PHP /fi" PHP. /fi is a freely distributed package that embeds commands (including mSQL queries) into html documents. Some functions are written specifically for mSQL and provide powerful tools and capabilities for interface development including:

1) Queries can be designed to conditionally return results based on other criteria.

2) A function that converts a string into a format recognized by mSQL as case insensitive.

3) The ability to modify characters in a search string that would not be recognized in an mSQL query. An example is that single quotes are used to define values for database fields, and in a search for "Hawai'i", the text must be converted to "Hawai/i" before querying the database.

PHP /fi is also used to pass variables between WWW pages. This feature is used in the SHPD site to give users the ability to limit their queries to a geographic region through a graphical interface . Administrative tax division maps are used to begin development of bibliographic text searches through the PHP. /fi interface. Another advantage of these persistent variables is that it allows users to modify their queries and then resubmit them.

PART 2

GEOGRAPHIC INFORMATION SYSTEM

The software selected to display Geographic Information System is the freely distributed JShape program (Jshape Software). Environmental System Research Institute's "Internet Map Server" was investigated as an alternative to Jshape. however, this technology is very expensive in terms of both software licensing and hardware requirements.

JShape uses JAVA appiets to display spatially referenced data stored in ARCVIEW "shape" file format, and "tiff" image format. Although source code is not distributed, the program is designed to be compatible with Java enabled browsers such as Netscape (Netscape Communications Corporation) and Internet Explorer (Microsoft).

The program is not client/server based and data must be downloaded from the WWW server side to the client side each time it is used. Limits on the size of the data sets provided are therefore related to the bandwidth available to the user. Unlike most Client/Server programs, however, JShape performs all calculation and drawing functions on the WWW clients' side, and performance is not limited by the server. As in most applications faster computers equate with better performance, however, any Pentium based computer will generally be capable of reasonable performance.

PART 3

DATA SELECTION

The data that were originally selected for access on the WWW site include the SHPD library bibliography, and GIS data on Hawaiian ahupua'a (traditional land division), historic coastlines and fishponds. This list was expanded during the project to include general representations of archaeological site distributions. The following table shows the data that is presently available (x = data online, o data under review) . Other data will be added as they become available.

ISLAND	BIBLIOGRAPHY	SITE	AHUPUA'A	HISTORIC COAST	FISHPOND
Kaua'i	x	x			
O'ahu	x	x	x	x	x
Moloka i	x	x	x	o	x
Maui	x	x		o	
Lana'i	x	o			
Hawai'i	x	o			

Two comments were received with regard to concern about the kinds of information displayed on the WWW GIS. The first, from the SHPD Burial Program, was a request that traditional Hawaiian Burials not be identified in the WWW GIS. The information displayed about a "site" is restricted to an identification number and bibliographic reference. As such, there are no reference to burials in the WWW GIS.

The other concern was expressed at a presentation of the SHPD GIS to the trustees of the Office of Hawaiian Affairs (an elected body that represents Hawaiian interests) . The concern expressed was that the site information displayed in the GIS could be used with Global Positioning System (GPS) devices to loot sites. The SHPD WWW GIS implementation uses an arbitrary coordinate system with limited resolution to prevent this kind of system abuse.

PART 4

SUMMARY

The mSQL database package was software selected to develop access to the SHPD bibliographic database. This multi-platform software supports SQL queries and is available from Hughes enterprises at a very reasonable fee.

PHIP/FI is a "freeware" cgi program with its own scripting language. It is used to develop an interface to mSQL, generate "html" documents "on the fly", and provides functions that are used to pass variables between "html" pages. It also is used to provide automated access logging, but the final implementation will probably use logging methods at a higher system level.

The JShape program's JAVA applets are used to provide users an interactive environment for viewing GIS data files (ARC VIEW "shape" files). The software runs well on current versions of Netscape and Internet Explorer browsers . JShape is freeware and is still being actively developed and new features to increase its functionality are regularly added.

Site Map

Main Page- Basic disclaimer, links for further inquiries, acknowledgment to the agencies and individuals involved.

Bibliographic Search- Provides users with access to the SHPD's report library. Users may choose to limit their search geographically using client-side image maps, or they may go directly to a text form to enter their search criteria. The searchable fields in the bibliography include report number, year, title, author, and tax map key. Results are returned 10 hits at a time in no particular order. It is possible to enter multiple criteria (in multiple fields) but the query engine doesn't rank the hits by relevance, nor is it possible to enter multiple criteria in a single field.

Viewing GIS data Users who have a current Java enabled browser can select various data by island and view them interactively. Data for all the islands is not yet available, but existing data include information on fishponds, districts, traditional land boundaries (ahupua'a), sites (heiau theme, and "all sites"), and modern features such as roads and a terrain map for

reference. Viewers can use a selection screen to limit the data retrieved, reducing the download time considerably.

Site Layout And Resource Requirements

This table shows the resource requirements for the site These numbers don't include the source code for Msql and PHP/FI, nor do they include the text file backup for the bibliographic database . These aren't necessary for the site to run, but they will be temporarily needed if the site is moved to a different server. Overall, 10 megabytes should be more than adequate if everything is moved to a single server, if not, I would recommend allocating *5* megabytes for the front end and *5* megabytes for the remote MsqJ server.

Html files and images associated with the bibliography	50 Kbytes
Data files associated with the bibliography *	3 Mbytes
Html files and images associated with the GIS coverages (includes java le applet).	240 Kbytes
Data files associated with the GIS coverages.	3 Mbyte
PHP/FI Binary (in cgi-bin directory)	300 Kbytes
mSQL Binaries*	780 Kbytes

*The mSQL files and databases can be stored and accessed through a remote server if desired.

Site Maintenance and Updates

The site is checked periodically to assure that the Msql server and the php cgi script are functioning properly. The mSQL daemon occasionally need to be restarted (typically after the server re-boots). This is done using the username specified when mSQL was compiled. The mSQL program is run as a background task . There have been occasions where the cgi-bin directory has become inaccessible or when PHP/CGI wasdamaged and needed to be restored.

The bibliographic database can be updated by replacing the entire database or appending to existing records. In either case, the data should be extracted and formatted into a text file, that is redirected as input into the Msql database . (i e ~/Minerva/bin/Msql sites < filename msq). The

file that issues the input commands resembles the following:

```
drop table sites \p\g
create table sites rep num char(l0) primary key, year char(6), tittle
char(260), auth char(260),tmk char(260) ) \p\g
insert into sites (rep num,year, tittle, auth, tmk) values ('0—
00001','1970','Artifacts from Mission House Excavations in 1970 [seven
papers) ', 'Crouch, Richard; Munroe, Michael; Failey, Christopher; Chang,
Alicia; Sakai, Steven; Reilly, Hattie F.; Tanaka, Paulette',' 1—2—1-32:002')
\p\g
insert into sites (rep num,year, tittle, auth, tmk) values ('0— 00002',
'1973', 'Analysis of the Moanalua Historic Landmark Application Significance
Statement','Newman, T. Stell; Pyle, Dorothy R.; Takemoto, Anne; Kenn, Charles
W.',' 1—1—') \p\g
```

The proceeding example is typical except that the return characters "\p\g" are included
for legibility. To append to an existing database, delete the first two instructions (drop and
create). When adding records one should also be careful to escape any parenthesis that may have been
in the title (replace ' with /') A database script or a series of global replace commands are
typically used to properly format the data.

The GIS coverages are updated by replacing the shape and text files from ARCVIEW. A
separate directory has been created for each island to facilitate updates. As data is added
or changed, the "html" forms and the PHP/FI scripts are modified to reflect the changes.

12

PART 5

USEFUL LINKS AND REFERENCES

freeWAIS Program
"http://www.wsc.com/freeWAIS-sf/fwsf_1 html"

J-SHAPE Program
Jshape Software
"http://shiuhlin.mil.org"

SWISH Program
"http://www.eit.com/goodies/software/swish/swish html"

PERL and Sprite (a PERL script to generate flat file data bases):
"ftp://uiarchive.uiucedu/pub/lang/perl/CPAN/'
"ftp:/uiarchive.uiuc.edu/pub/lang/PERL/CPAN/authors/Shishir_Gundavaram/Sprite
3.1 readme"
mSQL
"http://hughes.com.au/"
Hughes Technologies Pty. Ltd.
P0 Box 432
Main Beach
Gold Coast
Qld 4217
Australia

PostgressSQL
"http://www.postgresql.org/"

PHP./fi (CGI gateway for mSQL and Postgress):
"http://www.vex.net/php/"

General Reference
"http://cscsun.1.larc.nasa gov/~beowulf/db/web_ access html"

The free software foundation:
http://www.gnu.ai.mit.edu/fsf/fsf html

Hawaii State Historic Preservation Division Experimental Home Page:

DEPARTMENT OF LAND AND NATURAL RESOURCES

STATE OF HAWAII

WELCOME TO THE HISTORIC PRESERVATION DIVISION'S EXPERIMENTAL DATA BASE SERVER
Information on Historic Preservation in Hawaii

This page has been developed with the support of a grant from the:

National Center for Preservation Technology and Training

You can do a text based <u>SEARCH OF MORE THAN 4800 TITLES IN THE DIVISION'S LIBRARY</u>

or with a graphical browser you can

<u>SELECT A GEOGRAPHIC AREA TO NARROW YOUR SEARCH</u>

If you have a JAVA enabled browser like Microsoft's Internet Explorer 4 0, try these links to

Geographic Information System (GIS) views of the. . . .

<u>Island of Maui</u>

<u>Island of Kaua'i</u>

<u>Island of O'ahu</u>

<u>Island of Moloka'i</u>

For a picklist of GIS coverages for O ahu and Moloka'i:

<u>Picklist for O'ahu and Moloka'i</u>

THIS DATA IS IN DRAFT FORM AND IS FOR GENERAL INFORMATION ONLY