ABSTRACT: The Marine Biological Laboratory has had a WEB site (www.mbl.edu) since January 1994. The usual pointers to information about the laboratory, its research programs, educational programs, and library are all present. The statistics demonstrate that the most heavily accessed document/file is the Marine Resource Catalog that was included as an afterthought. The MRC catalog lists the marine specimens that serve as biological models for research and education. The original file was a simple text document that listed taxonomic and ordering information along with prices. Based on communications with scientific users, both at the institution and over the net, the webmaster was able to determine that web browsers were really looking for information about these animals including images, scientific and biomedical publications, and links to molecular and genetic sequence data. The development of this resource is supported by the Howard Hughes Medical Institute and the National Library of Medicine.

INTRODUCTION:

The MBL/WHOI Library is the research library for the Marine Biological Laboratory, the Woods Hole Oceanographic Institution, the United States Geological Survey, three sites of the National Marine Fisheries Service, and the Sea Education Association. It also supports a substantial summer research and education program that involves faculty and students from over 200 universities in 25 countries. Many of these individuals return to the Library throughout their careers for study and research.

Since 1888, the MBL/WHOI Library has provided both current literature and the history of scientific thought to research faculty and graduate students in the biomedical, marine, and oceanographic sciences. The Library is known internationally for the breadth of its coverage, for its ease of access, and for the service given by its staff. The Library houses a journal collection of 170,000 volumes representing 6,000 titles, with 2,000 current subscriptions, a book collection of 36,000 volumes, and electronic access to journals and databases. Primary areas of the physical and electronic collections are biology, physiology, oceanography, neuroscience, ecology, zoology, marine geology and geophysics, natural history, and expeditions.
The MBL/WHOI Library is in the process of completing a one million dollar grant from the Howard Hughes Medical Institute that helped move the Library towards becoming a conduit for electronic information. Also, the Library, working on a contract with the National Library of Medicine, teaches an outreach program in Medical Informatics. The library through these sources has moved in the direction of placing itself on the path of electronic delivery of information to its scientists and patrons. The Library has become an information provider on the Internet and records over 1500 computer logins daily to its information server from users both locally and around the world. It is our intent to continue to use high speed data networks to provide a wide range of information to scientists throughout the world.

Problem

From the collection of data on the WEB logins we were able to determine the most heavily accessed file we had on-line. Not surprisingly, the Marine Resource Catalog logged in more accesses than any other file (Figure 1). We decided this file would be the model we used to build our knowledge base for the scientists in Woods Hole.

The web page that had the Marine Resource Catalog was incorporated into a new page that dealt with the Marine Resource Center as a whole. The Marine Resources Center (MRC - Figure 2) at the Marine Biological Laboratory is one of the world’s most advanced facilities for maintaining and cultivating marine organisms essential to advanced biological, biomedical, and ecological research. Information needed by the scientists is not only the seasonal availability of animals and price quotes but the scientific and biomedical literature surrounding them. Using the basic 205 animals as a focal point we began to build the knowledge base collecting other information about these animals. These animals have linked the research and education programs to the library’s collection development policies for the past 100 years. As with the physical collection it is now necessary to build the electronic information base concerning these animals.

Methods

Making decisions on what data were to be collected and in what format it should be kept for indexing and retrievability needed to be addressed. The best format for keeping and maintaining data is not necessarily the best format for presentation. This dilemma is especially true for publishing on the INTERNET. Since publishing in this fashion is so new and dynamic there are few tools for transposing data to an HTML page. HTML specifications are evolving very quickly so that such tools would need constant revision and updates in order to stay current. These observations led us to recognize an essential split between the data content and the infrastructure used to present the content to the user.
It was determined that the data would be maintained in a database manager (FileMaker Pro) program and published in HTML. Having a database manager which provides a structure for data that can be easily separated from the content is essential as well as the ability to export the data. This portability of data is essential in a world of rapidly changing hardware and software. HTML is a markup language which embeds text codes directly into the content of a file. The file remains a text file which transmits effectively over the net in standard ASCII and therefore requires no special tools to read it. Learning HTML is relatively easy; however, implementing HTML without some careful planning can create some serious problems. Having accepted this dichotomy between the database and HTML we needed something that would not lock us into HTML - if it changed to something else, i.e. HTML 1.0 has moved to HTML 2.0 with 3.0 already developed and Netscape pushing new HTML features with every release; SGML is on the horizon promising more robust tagging capabilities and XXXX will be the killer application of the future. It is just as critical to have the means to separate the HTML tags from the data as it is to have the means to embed the tags so that if either the content or the infrastructure change, updating can be done efficiently.

This plan allows us to keep our data portable yet take advantage of the latest features in Web publishing. We want to be able to accommodate growth and change in our data and to have those changes quickly reflected in our web site.

There is currently no simple solution to translating data records into HTML, so we developed some custom tools which enable us to keep the content in a manageable resource. A computer science summer student was given specifications designed by the research team, and he created a set of programs (using PERL) that translates data files into interlinked HTML pages. These programs allow us to quickly change the layout of our data and enable us to update the files and accommodate the changes in HTML, or any other standard, painlessly. If SGML became the standard, we would just write a new SGML template and run the data through it to create the presentation for Web Browsers. This flexibility allows us to present the same data in three different formats - one optimized for Netscape 1.1, one for HTML 1.0 specifications, one for text view (Figure 3).

The key to the system is in the use of one or more template files. The series of PERL programs create HTML files from a text file exported from a database program. FileMaker Pro database is used at MBI, but any regularly delimited file can be processed. A good database manager is essential as the source of information. These tools will serve as the authority source from which Web pages will be generated. In other words, HTML itself would serve as a poor means to manage data. A business might use a database manager to maintain client lists or a product catalog, information subject to regular change. The MBL uses FileMaker Pro for Purchasing, Order Entry at MRC, Admissions, etc., for managing information at the desktop. The Marine
Models data are also managed in FileMaker Pro and subject to regular changes, some more frequent than others.

Publishing this information on the web holds some real problems since you would most likely create a web page for each record as well as an index of names linking to each of the pages. Generating over 1000 files in HTML is a daunting task! Most databases don’t yet have the flexibility to export to the rapidly changing HTML environment. The solution to this problem was to create a program that would allow the contents of a database to be “dumped” into a template or layout and have the program automatically generate the 1000 HTML files. Creating the index means fashioning a new “template” which lists the names and pointers (HTML anchors) to the files previously created.

Ideally, one would create a template for what a typical page would look like and then move the data into it and generate the corresponding 100 or 1000 HTML pages. If one needs to change the look of the pages, just change the template once and run the data through it again. If the data change, just export the data again and run it through the template. By achieving a split between the content and the infrastructure you can change one or the other independently and quickly re-integrate the two into new HTML files.

The tools that we developed give us this flexibility. We can export our marine specimens data to a single file. That file can be moved through several templates. We have one template which is tagged with the latest Netscape tags, another follows the HTML 1.0 specs for other browsers, another template generates a text only version of the data. Each of these templates generates 200 HTML files in a matter of seconds. More importantly, if the database grows to 400 or 1000 records there is no proportional development time to get those files into HTML.

The system is made for maximum flexibility. One template might output 200 files while another outputs one file with all 200 files listed for an easy linked index. Built-in variables allow one record to point to the next or previous record in the database. Programmers can add custom code to do specialized tasks such as conditionals to make decisions based on a field’s contents.

Using this model, data can be maintained in the database of choice, exported as often as desired, and put through the template. Furthermore, as HTML evolves and new features are made available, one needn’t be locked into a particular HTML design. Re-creating 1000 HTML files by hand is totally impractical. The solution here is simply to create a new template and move the exported data into it. Databases all have simple export mechanisms by which data records can be exported to a text file.
Knowledge Base

Overview of Knowledge Base on Marine Animals at the MBL (Figure 4)

Each animal has its own central page that includes the following information.
- Image file (Fig. 4f)
- Common and Scientific Names (Fig. 4a)
- Systematic Taxonomic Information from the Phylum through the Family level (Fig. 4a)
- Links to Taxonomic Keys (Fig. 4a)
- Links to Literature Citations with on-line ordering forms (Fig. 4b)
- Links to Genbank at the National Library of Medicine (Fig. 4c)
- Links to the MRC Catalog information for scientific requisitioning (Fig. 4e)
- Image Library database (Fig. 4g)
- Place holder for links to other resources specific to an organism i.e.
- BBMER (Fig. 4d)
- Navigation bar which gives a hierarchical access to the MBLWeb system (Figure B)
- Built in Help (Fig. 5b)

Because of the limited bandwidth that most people using the net have available to them, one of the first things we noticed was that the small image files were accessed at a high rate. The files on this knowledge base are very image oriented with links to the text at a lower level. Having all the data on one page as opposed to scrolling through seems to satisfy the attention span of the web browsers (Figure 6). Of course one can always turn off the images and just use text for faster retrieval.

The scientific community uses marine models for biomedical research. Searching the literature for the latest published information on the species is time consuming. This resource automates the process so that the program presents the citations to the researcher using pre-established search strategies that are tailored to the research done on the animals. In Figure 5 you will notice that for the animal Limulus the pre-established search strategies take into account the work done on this animal in relation to vision research, lysate research, and basic physiology. Limulus is used heavily in vision and lyocyte research. These categories for searching would not be appropriate in the squid. More appropriate would be physiology and neurology. For other species, like the Leidy's Comb Jelly, general information is all that is available since that animal does not have a large research front.

Quarterly, we run the searches against CD-ROMS that we purchase and create a text file for each search. A program was created which strips out the citation information and creates two
interlinked HTML pages for each text file and automatically places those files in the appropriate place on the web server. One file has the abbreviated citation for general distribution and the other includes the abstract along with the ILL order information for in-house and Woods Hole affiliated users. The longer version is limited to Woods Hole use in order to comply with copyright restrictions.

The link to the National Library of Medicine is extremely important to the Woods Hole community since so many scientists access the molecular information that resides in the NCBI databases. Again, we have created templates that allow the researcher to go directly to Entrez or Genbank from within the MRC Knowledge base. The search includes the animal’s name and a choice of a nucleotide, protein, NCBI’s Taxonomy database, or the Molecular Biology subset of Medline search.

Future

This knowledge base provides us with a framework to include outside systems that have information we determine will be pertinent and important to scientific research in Woods Hole. This Knowledge Base could not have been built nor will it be maintained without the help of the scientific community that adds comments and pertinent information to the effort. The scientific data that are attached to an animal is mostly the result of scientists asking for that information in the area. This research front is more than one hundred years old in the Woods Hole community—the only thing that is changing is the way scientists are getting the information they need to do their research and how the library is now collecting and that information.
Aquatic Resources Division

**Marine Specimen Database**
This database covers 205 species currently available through our Marine Resources catalog. Other resources such as GenBank and taxonomic information have been integrated into this system. Other items such as marine algae and aquarium sets are available through our catalog as well.

**The Marine Resource Center**
The Marine Resources Center (MRC) at the Marine Biological Laboratory is one of the world's most advanced facilities for maintaining and culturing marine organisms essential to advanced biological, biomedical, and ecological research.

This link is under development (Oct. 3)

Copyright © 1996 by the Marine Biological Laboratory, Woods Hole, MA

---

**Figure 2**

---


<table>
<thead>
<tr>
<th>Phylum</th>
<th>Genus</th>
<th>Common Name</th>
<th>Page Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porifera</td>
<td>Leucosolinia</td>
<td>Organ-Pipe Sponge</td>
<td>Netscape HTML 1.0 Text</td>
</tr>
<tr>
<td>Porifera</td>
<td>Scypha</td>
<td>Vase Sponge</td>
<td>Netscape HTML 1.0 Text</td>
</tr>
<tr>
<td>Porifera</td>
<td>Cliona</td>
<td>Sponge, Boring or Sulphur, yellow</td>
<td>Netscape HTML 1.0 Text</td>
</tr>
<tr>
<td>Porifera</td>
<td>Halichondria</td>
<td>Bread crumb, Yellow Sponge</td>
<td>Netscape HTML 1.0 Text</td>
</tr>
<tr>
<td>Porifera</td>
<td>Halicosa</td>
<td>Finger or Eyed Sponge</td>
<td>Netscape HTML 1.0 Text</td>
</tr>
<tr>
<td>Porifera</td>
<td>Lissodendoryx</td>
<td>Non-Calcareous Sponge</td>
<td>Netscape HTML 1.0 Text</td>
</tr>
<tr>
<td>Porifera</td>
<td>Microciona</td>
<td>Red Beard Sponge</td>
<td>Netscape HTML 1.0 Text</td>
</tr>
<tr>
<td>Porifera</td>
<td>Mycale</td>
<td>Sponge, Mycale</td>
<td>Netscape HTML 1.0 Text</td>
</tr>
<tr>
<td>Porifera</td>
<td>Suberites fusc</td>
<td>Bright Yellow Sponge</td>
<td>Netscape HTML 1.0 Text</td>
</tr>
</tbody>
</table>

**Figure 3**
Figure 5
### Phylum Index

<table>
<thead>
<tr>
<th>Porifera (sponges)</th>
<th>Cnidaria (polyps and medusae)</th>
<th>Ctenophora (comb jellies)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platyhelminthes (flatworms)</td>
<td>Nemertea (ribbon worms)</td>
<td>Chaetognatha (arrow worms)</td>
</tr>
<tr>
<td>Bryozoa (bryozoans)</td>
<td>Mollusca (mollusks)</td>
<td>Annelida (segmented worms)</td>
</tr>
<tr>
<td>Sipunculida (peanut worms)</td>
<td>Arthropoda (jointed-leg animals)</td>
<td>Echinodermata (spiny skinned animals)</td>
</tr>
<tr>
<td>Hemichordata (acorn worms)</td>
<td>Chordata (tunicates and vertebrates)</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 6*