FINAL REPORT
Submitted 5/21/03

Project Title: **Integrating Bibliographic Information into Geographic Information Systems**
Project No: **R/HBT-060PD** Period: **4/15/01-2/01/03**
Index: **NA050B** Activity Code: **RDB9**
Project Award: **$8,775**
Principal Investigator: **Janet Webster**

Project Technical Summary
Accomplishments to Date
We carefully examined the challenges of creating a GIS bibliographic layer and found that it was technically complex. While experts thought it was possible, the time and effort involved outweighed the benefit. So, we successfully created a map interface to the Yaquina Bay Bibliography using simpler tools than full GIS. As the full GIS capability was elusive, we were not able to link to existing GIS of the Yaquina Bay. We will alert the Mid Coast Watersheds Council and the EPA Newport Lab of the availability of the map interface and share our project metadata and description of the challenges of creating a true GIS layer.

Links to full text have been inserted where currently available. Scanning equipment has been purchased through another grant and initial digitizing of key documents will start this summer.

Benefits Achieved or Expected
The map interface expands the access to the Yaquina Bay Bibliography by allowing people to explore the published research visually rather than through a keyword search. The maps and aerial photos provide an intuitive approach that many prefer. Additionally, my renewed focus on the Yaquina Bay Bibliography lead to revisions and additions including citations to recent research as well as additional historical material.

Project Report Discussion:
Interest in Oregon's estuaries and watersheds continues to grow, and the need for easier access to previous work remains high. Development of assessments and management plans often requires background information. In 1999, Oregon Sea Grant and the Environmental Protection Agency responded to this need by funding the creation of the Yaquina Bay Bibliography. This comprehensive bibliography allows researchers, managers and the general public to easily identify and locate over 1,160 citations of research papers, technical reports, local documents, maps, and datasets. This information provides background for proposed studies and context for management decisions.

We expanded access to the Yaquina Bay Bibliography by developing a map interface. This geographic tool allows users to query the database by geographic location rather than searching keywords. Users click on a location and can zoom in or out, linking to citations to work done at the various locations. This visual, geographic access enhances

people's use of the literature on Yaquina Bay and simplifies querying of the bibliographic database. Ease of access will save researchers' time and will possibility make managers more aware of past work and benchmark studies. The geographic display also helps researchers, students and managers identify under-studied areas as well as those of obvious high interest.

Technically, this tool is a simple solution to linking bibliographic records to geographic locations. While not elegant, it is inexpensive and uses readily available methods rather than expensive software solutions. Our approach has potential for use in other estuaries where information exists but is poorly archived and not readily accessible. Our initial objectives were to create a bibliographic layer that could be imported into ArcView or ArcInfo, the standard geographic information system. While simple in concept, this proved to be difficult in practice for two main reasons. First, ProCite does not structure the bibliographic data in an open, relational database. This made it difficult to construct the geographic link and export it into a simple GIS data layer. Secondly, we did not have access to an ArcView IMS server and did not have the intensive time and funding resources needed to set one up and maintain it. Consequently, we could not exploit ArcView's ability to read DBFs and convert them to a shape file. Instead, we created static images maps with multiple scales and JPEGS that are linked to the bibliographic database. This solution is appropriate when an organization, such as a watershed council, does not have an intensive GIS infrastructure and still wants to use geographic information creatively. (See Supplemental Information for the project's metadata.) This map is on the Worldwide Web at http://osulibrary.orst.edu/guin/yaqbib.htm.

Another objective was to link to the full text of certain documents of high interest. This has not yet been accomplished due to limitations of equipment and staff. Equipment has been purchased with funds from another source. We anticipate experimenting with this in the near future. Documents will be scanned as appropriate and posted to the Web with the URL embedded in the citation. Consequently, when a user clicks on a geographic location, the corresponding citations will appear, and users can then link to the full text where available.

The linkage of bibliographic information to geographic location continues to be a challenge for the library community. The approach taken in this project adds another tool for the community to consider. We intend to compare our approach to others and share our missteps and successes. Ms Hiveley presented the concept to the MidCoast Watersheds Council Technical Team and the MidCoast Watersheds General Council. A follow-up announcement of the availability of the tool has been made. Ms Hiveley is currently finishing her master these and the development this tool helped inform her research on the use of information gathering on a watershed as an outreach tool. The project was shared with library colleagues in the marine sciences through a quarterly newsletter and will also be more fully explained at the 2003 IAMSLIC Conference in Mystic, Connecticut. Mr. Justin Mills, the GIS consultant, plans to present the project at a regional GIS conference, GIS in Action 2003, and will submit it as a poster to the annual ESRI sponsored GIS conference in the summer of 2003.

We wish to acknowledge two people in particular for their expert assistance on this project. Pat Clinton of Newport's U.S. EPA Lab provided the base EPA data as well as practical assistance with the project. Ralph Garrano of Earth Design Consultants consulted on the general concept assisting with our initial planning.

Supplemental Information

Outreach activities:

Formal presentation to MidCoast Watersheds Council on the project, Fall 2001
Information presentation to MidCoast Watersheds Council Tech Team, Fall 2001
IAMSLIC Newsletter, February 2003
Announcements to HMSC community

Students:

Heather Hiveley, Masters student in Fisheries & Wildlife, Oregon State University, Database technician
Justin Mills, RARE Intern, OSU Extension, GIS Consultant and technician

Project Metadata:

See Appendix.

Lay Outcomes Summary

Discovering what others already know is key to doing good science and management. The Yaquina Bay Bibliography helps scientists, students, managers and citizens find out what research and studies have been done on and around Yaquina Bay, Oregon. Over 1160 citations to material are accessible through the Worldwide Web using either a keyword search of the database or a map interface. The map interface helps people locate an area of particular interest and then discover what has been published on that area. Starting with a base map of the basin, a user can zoom in through aerial photos to more exact locations. Coverage ranges from the Yaquina Bar to the upper reaches of the Yaquina River. Emphasis is put on scientific studies, but many management and historical reports are included. This information resource is a great tool for all interested in what has been studied in the Yaquina Bay and what still needs to be researched.

Appendix

Yaquina Bay Bibliography GIS Project Metadata

Phase 1: Proof of Concept
Initial work was done to identify the project's complexity, key challenges and approach. One key goal was to test the feasibility of using static HTML imagemaps for the search interface.

- Used ArcView 8.1
- Gave a generic Yaquina label to a couple of sixth fields in MCWC assessment map.
- Exported to JPEG and opened into Adobe Image Ready to draw imagemaps of selected 6th fields.
- Created HTML file with the imagemap defined.
- Linked HTML file to WebPoster search engine
    - Accessed the WebPoster search engine via the HMSC web page.
    - Modified the WebPoster HTML text to allow the Yaquina Bay geographic query to define the search term rather then the WebPoster query, and then successfully direct it to a search page.
        - Stripped out all the text that was specific to the WebPoster searches and not standard HTML.
        - Replaced the WebPoster query box with the Yaquina Bay geographic terms.
        - Made searching WebPoster transparent to the user by storing search term, database choice, and other WebPoster options in hidden inputs.

Phase 2: Geographic Terms Data Layer Development
Of the 119 original geographic terms in the Yaquina Bay Bibliography, 71 had latitude/longitude coordinates derived from the USGS Geographic Names Information Server. Coordinate information was added for an additional 28 terms, leaving 20 terms without coordinates.

- Imported geographic terms list into Microsoft Excel to convert from HTML file to a spreadsheet.
- Fixed any incorrect latitude/longitude coordinates.
- Added a leading zero in front of all coordinates as later conversion required three digits.
- Converted list to a dBase IV (.dbf) file for use in ArcView.
- Downloaded an Avenue (ArcView scripting language) script and used ArcView 3.2 to convert location from degrees/minutes/seconds to decimal degrees, as required by ArcView.
- Imported .dbf file into ArcView 8.2 and created an event them from the converted latitude/longitude locations.
- Added 1995 digital orthoquads (DOQs) as an aerial photography base for refining locations.

- Corrected latitude and longitude for approximately 75% of location to match locations on the DOQs.
    - Refinements were made with the assistance of USGS maps, US Forest Service data and other data sources.
- Set definition query for the layer to exclude all geographic terms still missing location data specified and exported remaining points to an ArcView shapefile.

Phase 3: Base Map Production

A variety of maps were available as a potential base map for the display. We sought a balance between detail and screen readability. This meant finding a base map that could provide the most visual information in the simplest form. We decided to start with a coarser overview and then add the finer scale of the aerial photographs. We used the following data sets:

Yaquina Bay Geographic Terms produced above (Guin Library, 2002)
1995 DOQs (Oregon Geospatial Data Clearinghouse, 2002)
City and County Outlines (Oregon Geospatial Data Clearinghouse, 2002)
Mid-Coast Densified Streams (U.S. Forest Service, 2001)
Yaquina Bay Outline (EPA, 2001)

- Used ArcView 8.2
- Projected data in the Oregon Custom Lambert projection as defined at the Oregon Geospatial Data Clearinghouse.
- Designed a 3"x4" layout to be exported at 96 dpi containing 2 overview maps, a scale bar and a brief description of the overview map.
    - Designed small overview map at 1:700,000 scale showing major streams, cities and the current extent of the detail map.
    - Designed the detail map at an initial scale of 1:500,000.
        - City, county and Yaquina Bay outlines and the streams layer were set to be visible only at scales smaller than 1:20,001.
        - DOQs were set to be visible only at scales 1:20,000 and larger.
        - Geographic Terms were set to be visible at all scales.
- Created maps used for navigation in the web application.
    - Exported layout to a JPEG image with the detail map set to various scales. Scales were chosen to present easy to navigate groups of terms or single terms where appropriate. The Yaquina River groups were exported at 1:150,000 scale.
    - The exported JPEG images were compressed with IrfanView to reduce the overall space required for the search application.

Phase 4: Search Page Programming

Search pages for each geographic term needed to be accessible from the detail maps, the quick search drop-down list and the Yaquina Bay geographic terms list. We wrote a Visual Basic for Applications (VBA) program that took approximately six hours to create and debug. Most of the time was spent on simple text manipulation, but the automatic generation of the geographic place images was a large component. Additional "bells and whistles" could be added if we wanted a more complex web page.

- Used ArcView 8.2 to generate the web pages for the intermediate link between the map and WebPoster. A layout similar to the detail map described above included the aerial photographs and the overview maps.
- Designed an HTML search page template using percent-delimited placeholders where image filenames, the search term, and other information would be placed (e.g. %term%).
- Wrote a VBA program inside ArcView to generate the web pages, the quick search drop-down list, and the small detail images for each geographic term.
    - The program loads the geographic terms layer, the HTML search page template, and prepares the data for processing.
    - For each record in the geographic terms layer, the program uses the same process. It begins with the first record, runs the appropriate processes, and then moves to the next record. This repeats until all records have been processed.
    - Once a record is selected, the program extracts the geographic term, latitude, and longitude for the record.
    - The program generates a file name for the search page and images by stripping out the commas, slashes, spaces, and single quotes from the current geographic term, then appending the appropriate file extension (e.g. .htm or .jpg).
    - After replacing the placeholders in the HTML search page template with the geographic term, location information, and image filenames, the program saves the updated HTML to a file using the filename generated in the previous step.
        - The WebPoster search engine requires search phrases to be in quotes (e.g. “Yaquina Bay”).
    - To generate the air photo and locater maps for the search page, the program uses ArcView to display only the selected geographic term, then exports images (using filenames generated as described) of the maps at 1:20,000 scale and 1:500,000 scale.
        - The filenames for these images were previously generated and inserted in the HTML template, so they will automatically load when the search page is opened.
    - Before moving to the next record in the geographic terms layer, the program appends the geographic term and appropriate codes to a text file containing the JavaScript-controlled drop-down “Quick Search”
    - After the current geographic term is processed, the program moves on to the next term in the layer. When all terms have been processed, the program exits.

Phase 5: Building Imagemaps and Finalizing the HTML

One goal of the project was to make it relatively simple to maintain the tool. As it was created with static, relatively linked HTML and JPEG images, it can be easily posted to a website simply by uploading the entire geoterms folder.

- Created web pages with HTML imagemaps for each of the detail maps described in Phase 3 above so users could search via the maps.

- Used these bare HTML pages as a basic template with tables to contain the various navigation elements.
    - This template was created as text file in Microsoft's Notepad.
- Defined the imagemaps using the GNU Image Manipulation Program (GIMP) and the imagemap plug-in.
    - This is free, open-source software available for many platforms. The imagemap plug-in is somewhat unstable, but saved many hours of coding by hand.
- Copied and pasted the resulting imagemaps as text into the template file.
- Following the completion of the imagemaps used BK ReplaceEM 2.0 (a free text search and replace utility) to insert the updated JavaScript-controlled drop-down "Quick Search" menu into each file.
- Tested all pages using Internet Explorer 6, Netscape Navigator 4.79 and Opera versions 6 and 7. All HTML, JavaScript and imagemap components behaved as expected.

Unresolved Issues:

- Anytime the Yaquina Bay geographic terms list is changed, the application must be rerun and the imagemaps updated by hand. This could be automated if we anticipated updating the terms frequently. If changes are rare, it is easier to write a new page rather than a new application as new terms come up.
- The single-image design of the navigation map images creates substantially larger image files than using several smaller files for the same purpose. The export process could theoretically be modified to produce multiple, smaller files for the navigation maps and overview maps, thus reducing load time and server space required.
- The WebPoster search page would not function properly when accessed with Opera while other browsers worked fine. This difficulty was replicated on several computers and is unresolved at this time.
- Citations using geographic terms without locations need to be reviewed, as they are not currently linked to the map interface. If necessary, searchable geographic terms should be added the citations or coordinates added to terms.
- We continue to discover minor problems caused by using the JavaScript in the Quick Search mode.

Credits and Acknowledgements:

Applications development: Justin Mills
Bibliographic development: Heather Hiveley
Metadata documentation: Justin Mills & Heather Hiveley
EPA Base data: Patrick Clinton