WEB-BASED INFORMATION MANAGEMENT SYSTEM FOR THE LONG TERM ECOLOGICAL RESEARCH PROGRAM IN KENTING, TAIWAN

Yang-Chi Chang*, Meng-Tsung Lee **, and Kun-Chi Lai ***

Key words: web-based system, LTER, information technology, coral reef ecosystem.

ABSTRACT

This paper introduces a web-based information management system which has been set up as a collaborative facility to allow key environmental and ecological information access for the Long Term Ecological Research (LTER) in Kenting, Taiwan. The coral reef ecosystem degradation in Kenting has concerned both public and private sectors, for which a LTER team was formed to investigate the problem. A multi-functional web-based system integrates diverse data collections and analyses efforts of team members during the project span. It also provides authorities with effective information and knowledge for sustainable coral reef management. Three information technologies, web Database Management System (DBMS), web geographic information system (GIS), and Data Warehousing (DW), successfully implement the complex system. Web DBMS is to provide the fundamental data transaction and query, while DW is to support advanced data query by aggregating available datasets. Web GIS, on the other hand, facilitates map-based spatial data visualization and query. Detailed discussions regarding system data integrations and operational functions are presented.

I. INTRODUCTION

Kenting, on Hengchun Peninsula, is located at the southern tip of Taiwan, to the west of the Pacific, east of the Taiwan Strait, and bordering the Bashi Strait to its south. The Kuroshio warm current passes through the peninsula, and splits into two streams, thus making Kenting a unique location. One stream flows directly into the Kenting coastal water, bringing abundant marine organisms with it breeding some sixty percent of the world’s coral species. A healthy coral reef system provides a sound habitat for numerous fishes and marine creatures, and is also one of the richest ecosystems in biodiversity. This small coastal area is rich with marine ecology and other scenic resources, making it one of the most popular resorts, attracting million of tourists every year. However several emerging threats, including ever-increasing human activities, have endangered the integrity of local coral reef ecosystem. For examples, over-developed tourism has put pressures on the coral reef. These include improper construction, water pollution, sedimentation, over-fishing, and various recreational water sports. On the environmental impact, Taiwan is not alone, since Bryant et al. in 1998 [2] revealed that more than eighty percent of Southern Asia coral reefs are at risk, primarily owing to human impact.

This research project aims to develop a LTER information management system for the coastal waters at Kenting on the basis of information technology using data integration and sharing. This system could effectively integrate ecological data from all research sub-projects and digitize original data, there by combining spatial information for better storage efficiency and providing an easy-to-use user interface for the decision makers. The modeling capability of this system would also highlight data digitization advantages by extracting useful information for decision support. The users would no longer need to sort through large quantities of written documents, nor spend time looking for suitable information. The LTER project would be greatly improved and make a more significant contribution to ecology conservation. Several advanced information technologies are adopted in this research project, including Management Information System (MIS), web Geographic Information System (GIS), decision support with analysis tools, and Internet as a platform to develop a data integration and information sharing mechanism. The Database Management System (DBMS) component is established with MIS and is responsible for marine ecological and other related database connections. GIS handles geographic database and associate graphic objects with corresponding attributive data. Users or decision makers can access remotely to the system server through Internet, i.e. using user-friendly browsers to inquire or read statistics and information accumulated from long term ecological research. Furthermore, such
findings can be transformed into more useful information or knowledge using decision support and analysis tools.

II. INFORMATION TECHNOLOGY APPLICATIONS

It is a common consensus that each LTER project should systematically provide long term scientific data for decision support in terms of environmental and ecological protection, reasonable resource utilization, and sustainable development. The ILTER (international LTER) network, established by scientists worldwide engaged in LTER, suggests infrastructure development for information exchanges and distributed database management [8] in view of important information technology application. A LTER information committee has been established as part of the United States LTER project. Its purpose is to serve each of the ecological investigation sites and provide effective data management [10]. Therefore it is important that comprehensive data integration and sharing mechanism be established while a LTER project is under way. Fortunately, tremendous amounts of ecological and environmental data with temporal-spatial dimensions can be processed with the help of present day information technologies, and information sharing is possible with the Internet as a communication backbone. Chen [6] presented a DBMS for the Nanjenshan LTER site in Taiwan, and Chen and Yang [7] addressed the issue of how to establish and manage Taiwan LTER network information.

Modern information technologies are present in various literatures and widely used in studies related to environmental resource management and planning. Boston and Stockwell [1] described Australia’s World Wide Web (WWW) service called ERIN (Environmental Resources Information Network), which provides key Australian environmental information. They concluded that the WWW is an effective technology for a data center to provide information of various kinds, including attributive, geographic data, and modeling results. Loh et al. [9] described the importance of information management and exchange to sustainable development using WWW information technology. They also developed an information system which provides users with easy access to data. More recently, Chang and Wang [3] carried out a preliminary configuration for a solid waste management system framework in Taiwan. They suggested an approach integrating current hardware and software and combining information systems developed for each type of wastes. They also proposed a geographic data inquiry function and a client-server environment for improved waste management efficiency.

Chang et al. [4] developed a web-based information system for scrap vehicle recycle management. They used Internet as an information sharing mechanism to enhance operational efficiency within the scrap vehicle recycling program, where the participants in the recycling channel did not have easy and clear communications in the past. To reach the system goal, various information technologies were used including web DBMS, web GIS, and web computing. Chang and Chang [5] designed a web-based decision support system for sustainable river aquatic environment management in urban areas. The web system integrates the joint research effort, using data warehousing and web computing techniques, allows decision makers to perform remotely the complex data query and modeling analysis. A more sustainable management scenario can be devised effectively using such web system. These various information systems bring modern information technology to environmental resource management fields and enhance managerial performance.

III. SYSTEM DEVELOPMENT PROCESS

A system development process is divided into four stages as shown in Fig. 1. The first covers fundamental works including data collection, data format development, and database establishment. The second stage begins with system planning and design, and develops a Kenting LTER project web page. A web GIS is established in the third stage to add spatial query. The final stage executes knowledge management using related technologies and decision support analysis tools.

1. Stage One

The first stage collects diverse data from each sub-project. This work is critical because it lays the foundation for effective data integration, which will support analyses in later stage. The current study develops universal data formats to incorporate different kinds of original data, such as fish, coral reef, and algal data. Original data are stored in Microsoft Excel® format, which allows us to clean data, such as replace empty records with zero, for preliminary data quality check using the same software. Then, several Microsoft VBA (Visual Basic for Applications) scripts have been written for automatic data Extraction, Transformation, and Loading (ETL). Finally, data is stored in Microsoft Access®, the database management software, in the desired data format once the ETL process is complete.

The main purpose for building such a database is to support advanced analyses, particularly ECOPATH model development which represents the coral reef ecosystem dynamic. Therefore a strong data integration that collects necessary information is needed. To do so, the current study consults with a model builder and set up a universal data format to cover important information. The database is divided into environmental and ecological data sets representing the two distinct categories of LTER sub-projects. Most common data fields (survey items) in the original data compose the universal data formats. Table 1 shows the universal data format for environmental related sub-projects. The two data set groups contain the same main...
items, including subject, time, location, measurement, and others. Each main item has one or several sub-items, which are the real data fields in the database. For example, research title, researcher, feature, and sampling item are data fields used to describe the main item “Subject”. The ecological data format is similar to that in Table 1, except for three more sub-items, family, genus, and species, placed under the main item “Measurement”.

Each LTER sub-project has a different recording data format in the Microsoft Excel® files, as mentioned earlier. The current study serves as the data hub in the Kenting LTER project to fulfill effective data integration. That is, each sub-project submits original data without considering how to satisfy the universal data formats. The research team then studies all data formats and writes the VBA scripts for efficient ETL performance to handle diverse data sets given by each sub-project. The VBA scripts contain three components, including data field set up, data loading, and data writing. The database is updated three times per year using the VBA scripts, which greatly shorten the database pre-process time. Table 2 shows the complete list of sampling items stored inside the databases, which cover environmental and ecological aspects of information.

The database built using the universal data format, as defined in this stage, can be transferred to a database based on other metadata standards in the future. In case of further data integration, the complete study of the target metadata format is a prerequisite. Afterward, the Kenting LTER databases in the system can be converted, using XML (eXtensible Markup Language), into the specific data format suggested by the required metadata. XML provides a systematic way to describe structured data and thus making it possible as a data storage and interchange format. As the whole world has recognized the importance of this new data storage format, most popular DBMS have supported XML operations. The emergence of XML has made database transfers between different DBMS or metadata formats feasible and easier. Taiwan National Digital Archives Program has successfully applied this approach to

<table>
<thead>
<tr>
<th>Table 1. Universal data format for environmental data sets.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main Item</strong></td>
</tr>
<tr>
<td>Subject</td>
</tr>
<tr>
<td>Feature</td>
</tr>
<tr>
<td>Sampling Item</td>
</tr>
<tr>
<td>Time</td>
</tr>
<tr>
<td>Location</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Measurement</td>
</tr>
<tr>
<td>Others</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Complete data content in the databases.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research Title</strong></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Subject</strong></td>
</tr>
<tr>
<td><strong>Marine alga</strong></td>
</tr>
<tr>
<td><strong>Coral reefs</strong></td>
</tr>
<tr>
<td><strong>Fish</strong></td>
</tr>
<tr>
<td><strong>No. of catch per species</strong></td>
</tr>
<tr>
<td><strong>No. of catch per species</strong></td>
</tr>
<tr>
<td><strong>Density</strong></td>
</tr>
<tr>
<td><strong>No. of catch per species</strong></td>
</tr>
<tr>
<td><strong>No. of catch per species</strong></td>
</tr>
</tbody>
</table>
convert the original databases defined by various metadata into the databases satisfying Dublin Core.

2. Stage Two

The current study developed a web-based client-server communication system using the Internet for efficient data sharing and access. This stage involves system planning, system design, and web page development. The three-tier software architecture is adopted in the initial system-planning phase, including presentation tier (user interface), logic tier (network server and application server), and data tier (database). Users can perform remote data inquiry and model running via Internet with this architecture. The application server handles user requests and conducts the technical jobs, such as accessing geographic objects, modeling process, and database management. Fig. 2 shows the schematic diagram of the system architecture. The web server is the system kernel, with four components. The components in this system are described as follows.

The database management component is responsible for retrieving information from the connected databases, such as attributive database. The model bases management component invokes the modeling process using various model types, such as the statistic model. These two components have strong connections, as the modeling process requires input parameters taken from the database. The geographic display and query component, link with the database management component to support basic GIS functions. The analytic results generation component joins the model bases management component to process model execution results. System users are primary research project team members and decision makers in the public sectors. They can access the system through the easy-to-use interface, i.e. WWW browser, with Internet connection. Therefore the users don’t need to purchase new hardware or consider software configuration.

Four major functions in the system, as shown in Fig. 3, include data management, file management, web GIS, and knowledge management. Only authorized individuals can access all functions due to restricted data release by each sub-project PI. Data management allows users to define their own query conditions, and results are displayed promptly on the web browser. Meanwhile users may choose to save query results in their local PC as an Excel® file for further analyses. The system not only stores numerical data, but also related LITER project documentations. The current study for example, collected project proposals, yearly reports, meeting presentation files, and so on. A full-text searchable electronic file index is available in the system to help users quickly find the needed documents. Web GIS enables users to perform geographic layer display and spatial data query over the Internet. Knowledge management is the last system function, providing decision support functions. Data warehouse and indicator calculations are currently implemented. A more complex simulation model and data mining analyses can be added in the future. Microsoft FrontPage® software is used for web page development, and the system runs on Microsoft Windows 2000 server® with IIS® (Internet Information Services) as the web server.

3. Stage Three

The third system development stage establishes a web GIS using software Autodesk MapGuide®. Web GIS enables users to perform map-based spatial data visualization and analyses without actually purchasing a GIS. This technique greatly promotes the benefits of using GIS for both daily life and scientific study. There are presently two types of web GIS applications, server-side and client-side. Server-side application processes GIS functions on the server side and generates the map as an image. Client-side application shares GIS functions to the client and generates real graphic objects taken from the vector data. MapGuide® uses the client-side technique and usually has better system performance. However, users are required to install MapGuide® plug-ins. The server performs heavy computing tasks due to future modeling processes, so client-side web GIS is more suitable for the current system because it reduces server workloads.

A web GIS development defines system functions according to project demands, and collects relevant data including attributive and spatial data. Map browsing functions (such as zoom in, zoom out, and pan around) and spatial query should be adequate for project demands. The system should also provide base maps, sampling sites, and aerial photos of Kenting. These
geographic data, with different formats, need to be transformed into SDF (Spatial Data File) file format, a binary format used in MapGuide®. Meanwhile, the geographic attributes of the sampling sites are the corresponding databases established at the first stage. The final step integrates geographic layers and attributive data in an internet environment, such that spatial query over the web is possible.

4. Stage Four

The current study attempts to integrate both environmental and ecological data monitored by the Kenting LTER project. Extensive data is expected to be stored in the system as time progresses. Therefore, the current study employs Data Warehouse (DW) to efficiently and effectively manage the ever-increasing database. DW is the new innovative technique in database management system development. It reveals a strong need for advanced data query to support subject-oriented analysis. DW design is not the same as relational database. The greatest difference is that DW aggregates several two-dimensional tables into a multi-dimensional cube or hypercube. Fig. 4 shows a data cube example, which looks just like a cube with three dimensions representing location, time period, and monitoring items. The measured value is the entity or element inside the cube. All information is aggregated in advance, as in building a data warehouse with systematic arrangement, so users can easily find what they are looking for without getting lost in a huge dataset.

On-Line Analytical Processing (OLAP) is required for providing fast, consistent, and interactive access to multi-dimensional views of the pre-built data cube or data warehouse, supporting the decision analysis by DW. Therefore even a decision maker without extensive mathematical and statistical training can achieve the complex data inquiries, such as identifying intrinsic relationships of some factors using long term historical data. The current study uses the Microsoft SQL Server 2000® enterprise edition to develop the DW and OLAP service. There are three steps in building a DW and providing an OLAP service as shown in Fig. 5. The first step transforms various database sources into Microsoft SQL Server® using the Data Transformation Service (DTS) provided by the software. This step typically spends a long time on the “extract, transform, and load” (ETL) process. Once all the data sets are in the system, the next step is to build the data cube by aggregating the two dimensional tables. A data cube usually consists of several dimension tables (such as: location, time period, and monitoring items in Fig. 4) and one fact table (the measured value in Fig. 4). Universal data formats design in the first stage already considers how to build the data cube. Therefore it is easier to integrate the complex ecological and environmental data as a whole data cube at this stage. A cube browser is then needed to connect to the cube for OLAP service after creating the cube. Cube browsers usually provide user-friendly interfaces that allow end users to easily interrogate the cube. There are three basic types of cube browsers. The first is Office Web Components (OWC), such as the Pivot Table analysis provided by Excel®. The other two applications are third-party tools and custom-built applications. In the current system, OWC is used to insert a pivot table and a pivot chart in the web page. The pivot table data sources connect with the data cube using the OLE DB (Object Linking & Embedding Database) for OLAP.

The future work primarily relates to knowledge management for decision support based on the established database. For example, model simulation, using ECOPATH, and data mining calculations notify user of refined information for policy planning support. Integrating a web computing program (or model) requires both parameter-inquiring and model-processing procedures. The parameter-inquiring procedure collects scenario setup information by the web page user. Such information sends model input parameters to the server when the model-processing procedure is invoked. Several techniques are available to implement server remote computing, such as Component Object Model (COM), Java® and .NET®. The appropriate one will be chosen to meet future project demands.

In summary, the complete software configuration is shown as Fig. 6. The web server IIS® associated with ASP® (Active Server Pages) dynamically generates web pages as requested by user. The system combines HTML (Hypertext Markup Language) and ASP files for querying databases, displaying geographic information, performing OLAP and indicator calcula-
tion, and even future modeling analyses. To support these complex services, two application servers are in use, which are MapGuide® Server and Microsoft SQL Server®. Original raw data are collected and transformed into the databases and data warehouse by the ETL and DTS process. Afterward all the data can be accessed via Microsoft SQL Server® using the data and knowledge management functions provided by the system. In terms of the spatial information display, the digitized geographic layers were first transferred to the SDF files. MapGuide Author® was then utilized to design the Map Window File (MWF), which defines in advance how the geographic data will be displayed. When users make a request to the MapGuide® Server through the IIS® server to send and download MWF, the MapGuide Plug-In program for the browser will be activated to display the vector data. The graphic objects displayed on the browser are also linked to the attribute databases, which are manipulated by Microsoft SQL Server® and can be accessed via ASP. Therefore, both spatial information and the corresponding attribute data are provided to users when making a spatial query.

IV. SYSTEM DEMONSTRATIONS

The Kenting LTER information management system is developed and in use based on the system design as described earlier. Several screen captures of the web system show current study results. The system interface is in Chinese, but detailed English notations are added to explain system functionalities. The fundamental system function allows the users to access the server side database. Data query functions are divided into simple and advanced queries to accommodate various users’ needs. Advance data query allows users to issue more complex query conditions, while the simple query uses only sampling “site” and “time” for data inquiry. Advanced query interface allows users to aggregate several query conditions, such as family or genus name, amount, site, and time, using logical “and” and “or” as shown in Fig. 7.

If users don’t know the family or genus of species, they can press the magnifier icon, and a complete list in this particular database shows up in a popup window. The query results are shown as a table in the bottom of the web page, which can be saved as an Excel® file for further analyses. Meanwhile, the LTER database also links with remote databases for better referencing. Linking fish sampling data in Kenting with the “FishBase,” providing fish taxonomy according to Academia Sinica in Taiwan is shown in Fig. 8. Based on family, genus, or species name of a record, the corresponding fish is shown in the popup window. The user can therefore perform efficient cross-reference in their data query.

Web GIS provides basic map browsing functions like a normal GIS, where users interactively view, create, and print their own maps with MapGuide® plug-ins. System built-in functions enable users to operate zoom in and out, zoom to whole layer, pan, and return to previous view. The system also

![Fig. 6. Software configuration of the web-based system.](image1)

![Fig. 7. Advanced data query.](image2)

![Fig. 8. Remote database link with LTER data.](image3)
allows users to quickly find a sampling site. When a user selects the sampling “item” and “site” in the middle frame of the web page and presses OK, the web GIS zooms in to the particular map site, as shown in Fig. 9. Once users find the specific sampling site, they can double click on this geographic object, and the system pops up the sampling data associated with this site. Such spatial data query provides strong linkage between spatial and temporal information, facilitating the LTER study.

There are four basic operations in an OLAP service, including slice, dice, roll-up, drill-down and pivoting. Interactive operation of these actions provides different data perspectives. The “slice” operation selects only one dimension, while the “dice” operation focuses on a sub-cube by choosing two or more dimensions. “Roll-up”, also called aggregation or dimension reduction, allows users to move to a higher aggregation level. The “drill-down” operation is the reverse of a “roll-up”, drilling down to lower data cube levels and retrieving more detailed data. Pivoting, or rotation, changes the data perspective presented to users.

OLAP service uses the pivot table analysis provided by Excel® to display multi-dimensional view of the data cube. The users construct their own pivot table by dragging and dropping preferred dimensions from the data field list to column and row headings, and the pivot table heading. An OLAP service scenario is shown in Fig. 10. If a user wants to compare sampling amounts between fish, juvenile fish, alga, spiral shell, and crustacean, then she (he) must drag all these species from the data field list on the right and drop them in the pivot table column heading as one dimension. The user can further drill-down the species to show the sampling item. Meanwhile, the user can drag the “time” field, in particular 2002 and 2003, to the row heading and “site” field to the table heading to form the other two dimensions. Finally, by placing measurement value in the pivot table center, the user can set up a complete OLAP service configuration. Shortly thereafter the system will show various ecological data set aggregations in the table. Once the pivot table is ready, the user can perform other OLAP operations. For example, the “time” field consists of “year” and “month” hierarchical levels, with the higher “year” level currently displayed in the table. The user may operate a “drill-down” by clicking on the plus sign beside the year “2002” and “2003”, and more detailed data in the lower “month” level will show. A similar operation can be applied to expand the “site” field. Numerical information appearing in the pivot table can be displayed as a bar chart figure concurrently as shown in the lower half of Fig. 10. Therefore more powerful visual analyses of integrated LTER data are possible. Users can effectively and efficiently acquire desired information to support better analysis using the OLAP service.

The current study implements a simple computing process, which calculates abundant fish species in a sampling site. A user can click on a sampling site in the map, and the system carries out the indicator calculation. The result is shown as two pop-up windows, displayed in Fig. 11. One window provides detailed indicator information, while the other shows the indicator bar chart in all sampling sites. More advanced computing processes can be incorporated enhancing data integration benefits in the Kenting LTER study.

V. CONCLUSIONS

Collecting and establishing the ecological information over a long-term is one of the major LTER project contributions. Lack of data sharing in the past however, has greatly diminished research efforts and information acquired from earlier studies. Some gaps may therefore exist in valuable experience and baseline data accumulation. Consequently, it became difficult for researchers to conduct a thorough study when comparing current and past data. An effective solution must not only overcome the problem, but also take the full advantage of modern information technologies. Such an information technique has been widely applied worldwide in many researches by providing successful data management and supporting more
data-oriented analyses.

Modern information technology has tremendous human life impact and affects our professions in many innovative ways. The current study adopts an information technology approach by building a web-based system for the Kenting LTER project. The system uses Internet as a communication platform to facilitate information sharing for each sub-project. Therefore all system functions must be based on WWW techniques, such as web DBMS, web GIS, and web computing. The information system is based on the desired system design, which integrates not only diverse ecological and environmental data but also spatial and temporal data. Advance information techniques, such as data warehouse and decision support by the modeling process, are creative attempts to add to the ecological information system. More analytical tools can be implemented in the future under the system framework, such as data mining for acquiring implicit knowledge hidden behind the data. The benefits of Kenting LTER program can be realized with the help of such information systems.

ACKNOWLEDGEMENTS
The authors acknowledge the support of the National Science Council, Republic of China, under research grants NSC92-2621-B-110-003 and NSC93-2621-B-110-002.

REFERENCES