ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

APPLICATION OF WEB SERVICES FOR AQUATIC DATA WAREHOUSE SYSTEM

Hui Cham, SorayyaMalek, Aishah Salleh, Sarinder Kaur, Pozi Milow and Mogeeb Mosleh
Faculty of Science, University of Malaya, Kuala Lumpur,Malaysia
E-Mail: sorayya@um.edu.my

ABSTRACT

Maintenance and monitoring of aquatic systems such lakes, reservoirs and river involves properly documented, valid, and comprehensible data archives. However, aquatic data are collected and kept separately, creating difficulties in data integration. For effective aquatic data management it is important to have databases metadata that have been validated. This study aims to discuss framework for aquatic data warehouse system using web services for sharing database components using standard format and common data exchange method to foster easier data integration and exchange. The key features of the data warehouse comprises of graphical user interface (GUI) developed using ASP.Net. XML to represent metadata for data exchange and transfer, Darwin Core for formatting ecological and biological data management for data exchange protocol in this study.

Keywords: web services, application, aquatic data warehouse.

INTRODUCTION

Database system allows researchers to query and extract data for effective environmental management plan. However researchers faced problems in aquatic data management with data integration for analysis and exploration because of spatio-temporal differences of the data. Aquatic ecosystem around the globe faced problems such as changes in water level, eutrophication, pollution, acidification, and siltation [1][2][3][4][5][6][7]. Furthermore the aquatic databases are designed in heterogeneous formats and using different platforms. Some of the databases are maintained by autonomous bodies and administrations with absence of national policy, implementation, guidelines, and legislation on the hydrological data management. The data collected from various hydrological sites are not standardized, outdated, distributed and not available for sharing. Global level aquatic databases such as World Lake Database [8], Ramsar Site Database [9], GIS WORLDLAKE database [10], and LakeNet[11] contains data that are not up-to-date, important data and details are incomplete or blank and data is stored in heterogeneous form. This leads to difficulties in data sharing due to incompatibility problem. Synchronization between research agencies becomes tougher. This leads to poor dissemination of finding and synchronization in research to develop an effective environmental management plan.

XML are used to represent metadata in ecological informatics due to its inter-operability for data exchange and transfer between dispersed research groups [12]. Web services uses Extensible Markup Language (XML) for communicating between systems and are being used for environmental models [13], for sharing data within remote sensor networks [14], and for developing virtual databases that data are dispersed through various platforms [15]. European Bioinformatics Institute (EBI) deploys web services to enable researchers to use their database [16].

Web services are important for aquatic ecosystem due to increased availability of data and the time needed to access and integrate data from heterogeneous sources. Currently there is limited literature reported on usage of web services in aquatic database. The usage of standard protocol to implement data sharing with web services would increase the interoperability of data and better utilization of the datasets in environmental research or planning.

Issues such as incomplete data, data standards and sharing of aquatic ecosystem data could be solved in this study by providing a standard data representation and interface for data exchange. The aim of this study is to recommend a framework for aquatic data warehouse using web services for data archiving and retrieval for seamless data transfer for heterogeneous database independent of the database platform in aquatic ecosystem. XML document based on Darwin Core data standard format will be used in this study used together with SOAP (Simple Object Access Protocol) messaging protocol to enable heterogeneous data transfer.

MATERIAL AND METHODS

The aquatic data warehouse architecture in Figure-1 contains graphical user interface (GUI) developed using ASP.Net, web services, data standards and data mining tools. Figure-2 illustrates database structure using entity relationship diagram. The partial data structure reflects the properties of hydrological and biological data. The table structure is designed dynamically to allow users to manipulate the variables by using existing defined variables or creating a new variable. The biological data represents population of organism in the aquatic ecosystem is defined using Taxonomic Databases Working Group based on Biodiversity Information Standards. This is common standard used for data exchange using web services for biodiversity data among data warehouse systems that uses similar standards. The application of standard

VOL. 10, NO. 15, AUGUST 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

terminology for data exchange ensures data integrity and uniformity. Figure-3 illustrated the web services process using SOAP messaging protocol in the aquatic data warehouse system for data transfer and exchange. SOAP is defined as a messaging protocol which enables programs that run on different operating systems to communicate using Extensible Markup Language (XML). In order for a successful data transfer to the developed aquatic data warehouse in this study from another aquatic data warehouse or system that can be in heterogeneous format .A SOAP messaging protocol sent by the initiating aquatic data warehouse needs to be verified by the receiving data warehouse. The SOAP message contains information regarding the data format of the initiating aquatic data warehouse which will then be processed by the receiving data warehouse system. The receiving data warehouse will then format the requested data in XML format based on the content in the SOAP messaging protocol. The output for a successful data transfer would be indicated as a feedback message indicating successful data retrieval or failure to the sender or the initiating data warehouse. SOAP Figure-4 illustrates SOAP message transfer structure used in data retrieval and response. Darwin Core is used as the data standard format in this study for the developed aquatic data warehouse system. Darwin Core is one of the most comprehensive current formats for ecological data. It used for managing and developing biodiversity informatics data in Global Biodiversity Information Facility [17]. The aquatic data warehouse in this study deploys web services to process the Darwin Core form of communication message. Ramsar Classification System produced by wetland International Body contains standard identification information of wetlands has been adopted in this study as well. It is an important tool for identifying wetlands in order to inspire their usage and suitable long term management. The classification also serves as a broad framework to aid the rapid identification of the main wetland habitats represented at each site, to provide units for mapping, and to encourage uniformity of concepts and terms in national or regional wetland inventories. The aquatic data warehouse developed in this study allows data transformation from aquatic ecosystem data into Extensible Markup Language (XML) format for data exchange and transfer independent of the database platforms. Imports of data into the following format of Excel (.xsl), Access (.mdb) or Text file (.txt) is incorporated in the developed aquatic data warehouse system as well.

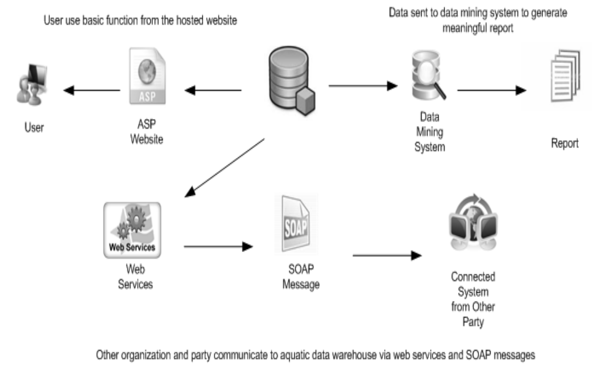

Figure-1. Architecture of the aquatic data warehouse framework system.

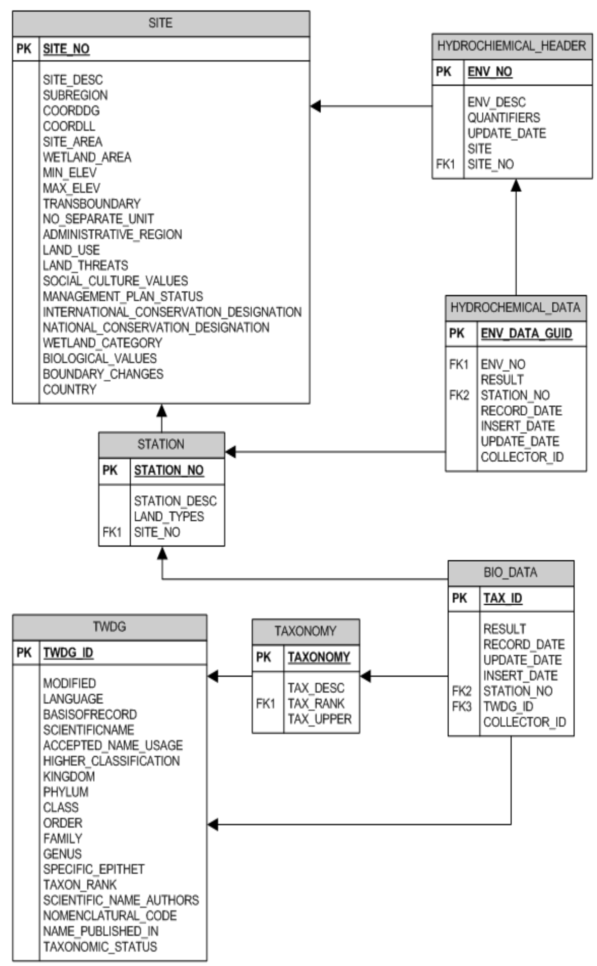

Figure-2. Entity relationship diagram of the aquatic data warehouse system.

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

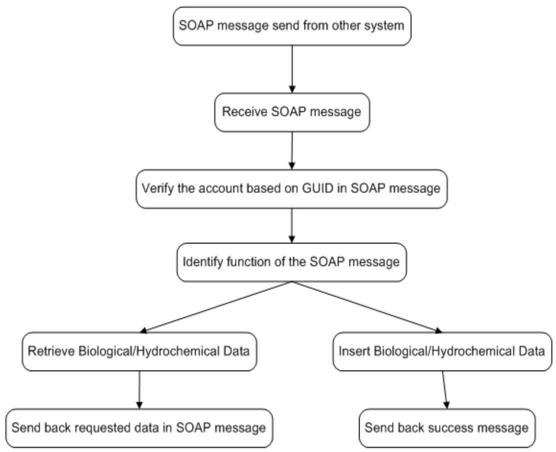

Figure-3. Illustration of SOAP message transfer in aquatic data warehouse system.

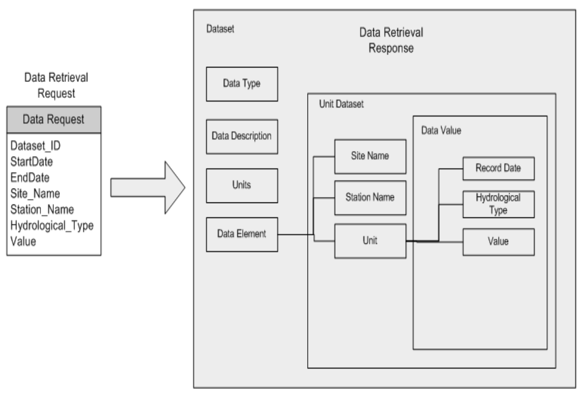

Figure-4. XML document structure for data retrieval request and response.

Data needs to be transformed into suitable format for data mining. The aquatic data warehouse enables data transformations into suitable formats that can be used by data mining tools. These data mining tools includes Hybrid Evolutionary Algorithm (HEA), Artificial neural Networks (ANNs) and Kohonen Self Organizing Feature Map (SOM). As visualization of aquatic data is important to detect pattern and trend in data for effective environmental planning the aquatic data warehouse also deploys GIS for visualization and modeling of aquatic data distribution over region, location and time frame.

RESULT AND DISCUSSIONS

The development of visualizations from a dataset is essential to discover relationships, trends, and temporal patterns. Researcher should be able to use the dataset to observe relationships between variables, answer research question, generate reports and charts, time series data visualizations, real time data entry, data query, and modification to produce informative statistical report via internet browser or web services. The temporal and spatial dataset of aquatic ecosystem should also be formatted into standard formats that can be used for analysis such as Ramsar classification. This is illustrated in Figure-5.

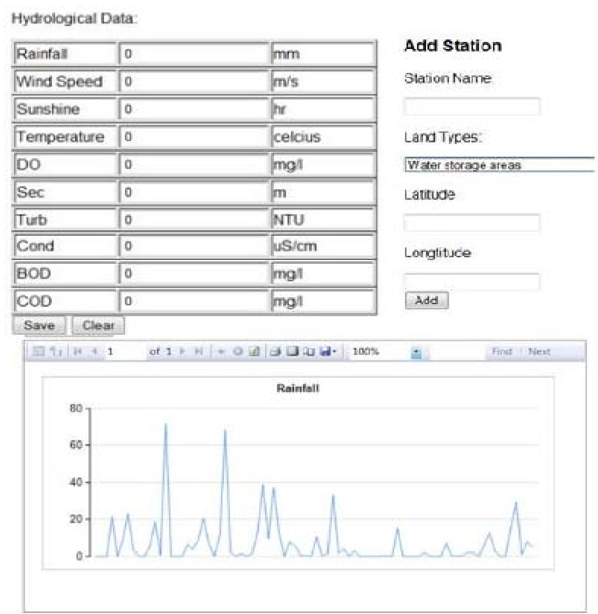

Figure-5. The aquatic data warehouse system GUI for report generation.

The aquatic data warehouse system developed in this study is able to generate suitable data formats for example .xls, .txt, which can be used by data mining tools such as ANNs, HEA, GIS and SOM. Data mining tools can be used for effective aquatic ecosystem management. HEA for example are able to generate rule sets or arithmetic functions to forecast hydrological data patterns. HEA offers extremely comprehensible rule set models and important information based on threshold values of hydrological datasets [18]. SOM meanwhile are used for modeling hydrological data to identify relationships between dataset that can be used for monitoring and managing aquatic ecosystem quality [19]. Combining both SOM and HEA might increase understanding of complex relationships for targeted aquatic ecosystem [20]. GIS are widely used for clustering, visualization, and discovery of relationship in ecological data for analysis and decision making by using the suites of predictions [21][22] [23]. GIS is deployed in the aquatic data warehouse system to visualize distribution of hydrological data as illustrated in Figure-5.

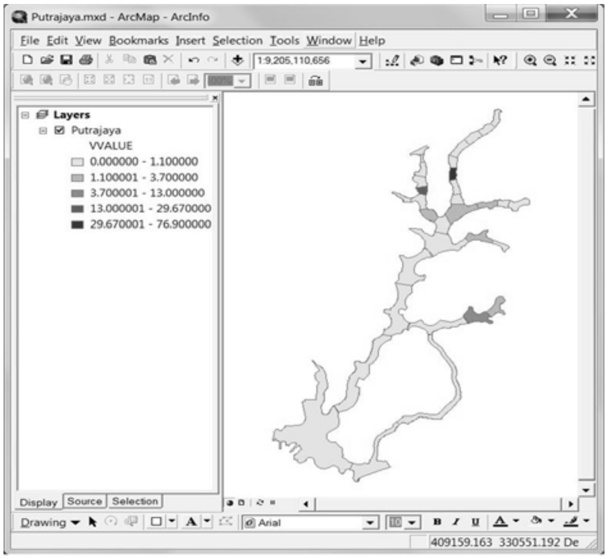

Figure-5. Hydrological data distribution of a Putarajaya lake using GIS deployed by the aquatic data warehouse system.

The aquatic data warehouse system allows skilled users to integrate web services into the user's database system to enable data sharing and transfer. A web services is a platform for users for data communication within systems. Figure-6 illustrates XML schema for data retrieval request and the resulting dataset in XML format sent by web server in this study.


```
<diffgr:diffgram>
  <NewDataSet>
    <Table1 diffgr:id "Table11" msdata:rowOrder "0">
      <modified>2006-05-04T18:13:51.07</modified>
      <language>en</language>
      <basisOfRecord>Taxon</basisOfRecord>
      <scientificName>Centropyge flavicauda Fraser-Brunner 1933</s
      <acceptedNameUsage>Centropyge fisheri (Snyder 1904)</accep
      <parentNameUsage>Centropyge Kaup, 1860</parentNameUsa
      <higherClassification>
        Animalia;Chordata;Vertebrata;Osteichthyes;Actinopterygii;Neo
      </higherClassification>
      <kingdom>Animalia</kingdom>
      <phylum>Chordata</phylum>
```

```
POST /EDWWS.asmx HTTP/1.1
Host: localhost
Content-Type: text/xml; charset=utf-8
Content-Length: length
SOAPAction: "http://tempuri.org/DarwinCoreSimpleGet"

<?xml version="1.0" encoding="utf-8"?>
<soap:Envelope xmlns:xsi="http://www.w3.org/2001/XMLSchema-insta
  <soap:Body>
    <DarwinCoreSimpleGet xmlns="http://tempuri.org/">
      <Species_No>string</Species_No>
    </DarwinCoreSimpleGet>
  </soap:Body>
</soap:Envelope>
```


Figure-6. XML SOAP message protocol results.

Aquatic data are naturally spatial and if databases are constructed using common and interchangeable representations they can be easily managed, provide metadata and offer common variables that allow dissimilar data sets to be joined. Furthermore functional data can be uploaded in an ad hoc manner, without interfering with performance of structure-based templates.

CONCLUSIONS

This study proposes a framework for aquatic ecosystem data warehouse integrating web services. The data warehouse was developed based on the need for fragmented aquatic data to be presented into a standardized format for data transfer and manipulation to solve the issues of data standard, data sharing and data incompleteness. Users and researchers can access the data warehouse for data exchange and report generation. The data warehouse system can be integrated into other aquatic data system because it has metadata for data migration using web services.

REFERENCES

[1] National Research Council. 1992. Restoration of Aquatic Ecosystems National Academy Press: Washington, D.C1

[2] United Nations Environment Programme. 1994. The Pollution of Lakes and Reservoirs: UNEP Environment Library No. 12. United Nations Environment Programme: Nairobi.

[3] Dinar A., Seidl S., Olem H., Jordan V., Duda A. and Johnson R. 1995. Restoring and Protecting the World's Lakes and Reservoirs. World Bank Technical Paper No. 289. World Bank: Washington, D.C.

[4] Ayres W., Busia A., Dinar A., Hirji R., Lintner S., McCalla A. and Robelus R. 1996. Integrated Lake and Reservoir Management: World Bank Approach and Experience World Bank Technical Paper No. 358. World Bank: Washington, D.C

[5] Nakamura M. 1997. Preserving the health of the world's lakes. Environment, Vol. 39, No. 5, pp. 16-39.

[6] Duker L. 2001. A literature review of the state of the world's lakes and a proposal for a new framework for prioritizing lake conservation work. LakeNet Working Paper Series, No.1. LakeNet: Annapolis, Maryland.

[7] Ayres W., Busia A., Dinar A., Hirji R., Lintner S., McCalla A. and Robelus R. 1996. Integrated Lake and Reservoir Management: World Bank Approach and Experience World Bank Technical Paper No. 358. World Bank: Washington, D.C

[8] International Lake Environment Committee Foundation (ILEC), 2013 World Lake Database – Accessed on November 21 http://wldb.ilec.or.jp

[9] Ramsar Convention on Welands. 2013. Accessed on December 4 - http://www.ramsar.org

[10] Kirill Ya. Kondratyev and N. N. Filatov. Limnology and Remote Sensing:a contemporary approach , Spinrger, 1999.

[11] LakeNet. 2013 World Lakes Website. Accessed on October 5 - http://www.worldlakes.org

[12] Seligman L and Roenthal: XML's impact an databases and data sharing. Computer and Processing 2001, IEEE Computer Society. Vol. 34, pp. 59-67.

[13] Mineter M.J., Jarvis C.H. and Dowers S. 2003. From stand-alone programs towards grid-aware servicess and components: a case study in agricultural modelling with interpolated climate data. Environmental Modelling & Software, Vol. 18, No. 4, pp. 379e391.

[14] Liang S.H.L., Croitoru A. and Tao C.V. 2005. A distributed geospatial infrastructure for Sensor Web. Computers & Geosciences, Vol. 31, No. 2, pp. 221e231.

[15] Frehner M. and Brandli M., 2006. Virtual database: spatial analysis in a webbased data management system for distributed ecological data. Environmental Modelling & Software, Vol. 21, No. 11, pp. 1544e1554.

[16] Alberto Labarga *et al.* 2007. Web Servicess at European Bioinformatics Institute, Nucleic Acids Research, Vol. 35, Cambridge.

[17] Global Biodiversity Information Facility, 2013 Accessed October 10 - http://www.gbif.org/

[18] Hongqing Cao, Friedrich Recknagel and Philip T. Orr. 2012. Enhanced functionality of the redesigned hybrid evolutionary algorithm HEA demonstrated by predictive modelling of algal growth in the Wivenhoe Reservoir, Queensland (Australia), Ecological Modelling, Elsevier.

[19] Sorayya M., Aishah S. and Sharifah MSA. 2009. Analysis of Algal Growth Using Kohonen Self Organizing Feature Map (SOM) and its Prediction Using Rule Based Expert System. Proceeding ICIME '09 Proceedings of the 2009 International Conference on Information Management and Engineering. IEEE Computer Society, Washington DC.

[20] Tae-Soon Chon, Self-Organizing Maps applied to ecological science, Ecological Informatics, ScienceDIrect, 2011.

[21] Xavier S., Jose CB, Neftali S. Juan MP, Gustavo AL, Soumia F. and Xavier P. 2006. Inferring habitat-suitability areas with ecological modeling techniques and GIS: A contribution to assess the conservation status of Vipera Latastei. Biol Cons, pp. 416-425.

[22] Richard A. and Diane P. 2000. Integrated geographical assessment of environmental condition in water catchments: Linking landscape ecology, environmental modeling and GIS. J Environ Manage Vol. 59, pp. 299-319.

[23] E. Ashley *et al.* 2008. A Spatiallly Explicit Decision Support System for Watershed-scale Management of Salmon, Ecology and Society, Vol. 13, No. 2.