

A Metadata-Based Framework for Multi-Language Ecological Information Management in East-Asia Pacific International Long-Term Ecological Research

Chau-Chin Lin¹⁾, John Porter²⁾, Meei-Ru Jeng¹⁾, Sheng-Shan Lu¹⁾, Chi-Wen Hsiao¹⁾, Chien-Wen Chen¹⁾

Introduction

Ecological data collection uses a variety of protocols in the field. As a result, heterogeneous datasets are produced and stored in very different ways. The long-term ecological data should be preserved and made available for future studies, for enlarging spatial and temporal scales, and for scientific conclusions.

Data sharing, exchange, and efficient synthesis by individual scientists as well as member networks within the East-Asia Pacific (EAP) region requires establishment of an Information Management System (IMS) for the EAP-ILTER. We have developed a framework of prototype ecological information management tools, based on a metadata standard. The framework was developed by Ecological Informatics Working Group of the Taiwan Forestry Research Institute (TFRI) to aid with editing, storing, and using documents in the multiple languages of Asian cultures that comprise the EAP-ILTER.

Rationale

Metadata describes the who, what, when, where, and how to every aspect of data. Several groups in the US have collaborated to define a standard format for documenting ecological data. This standard is called Ecological Metadata Language (EML). There are 4 major features of EML, modularity, detailed structure, compatibility, and strong type, all of which show the characteristics of this standard.

EML documents vary in their levels of content. To evaluate the implementation of EML, 5 levels are used to determine completeness. Each level adds more elements from the EML schema to provide a more-comprehensive description of the data resources documented by the metadata, and thereby supports higher functionality.

Due to the comprehensiveness of EML, the EAP-ILTER decided to use it as a standard and basis for establishing a regional IMS.



Fig. 1 The data discovery and analytical tool interfaces



Fig. 2 The outputs from the tools of the system

System

After conducting the resources available in the field of ecoinformatics review, we decided to adopt the systems developed by the National Center for Ecological Analysis and Synthesis (NCEAS) at the University of California, Santa Barbara which include Morpho, Metacat (short for metadata catalog), Google Map, and the EML2R from the Processing Techniques for Automated Harmonization (PTAH) project at the University of Virginia. Jointly these provide the tools for creating, editing, storing, retrieving, and using EML documents. Morpho and Metacat were modified to resolve language coding issues in Asian countries.

Framework

The conceptual framework of the information management system can be divided into 3 tiers. The first tier deals with datasets and related information. Data produced by automated sensors communicating through wired or wireless networks, or collected manually by scientists, are managed by this tier. In addition, all information related to a dataset is also edited in this tier. The second tier relates to information processing and management. Once datasets and other related information have been described, they will be stored in a standards-based, schema-independent database, and used to drive data discovery and analytical tools, including geographical information systems (GIS). The third tier consists of the full web-based interfaces that allow easy access to the functionality provided by the second tier. This tier also manages definitions of multiple user categories with different user rights.

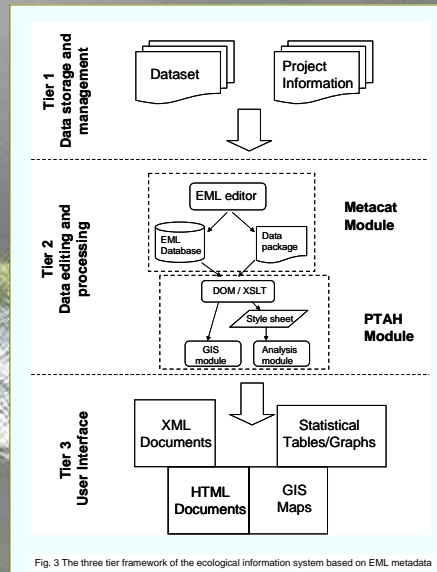


Fig. 3 The three tier framework of the ecological information system based on EML metadata

Software

OS: Red Hat Fedora Core 3

Data Catalog System:

Server: Apache Server, Tomcat Server, J2SE, OpenLDAP Server, PHP, PHPLDAPAdmin, PostgreSQL Server, Metacat Server.

Client: Morpho

GIS Maps System: Apache Server, PHP, Google Map API, DOM, Google Map, Google Earth.

Data Statistical Analytical System: Apache Server, PHP, R, R-phi, MySaj Server, ImageMagick Display, XSLT.

Outcome

A test of the suitability of the system for multiple languages was conducted through ten domestic workshops and two international workshops held from 2004 to 2006 (Tab 1). In total, 335 participants attended the workshops. Participants of the two international workshops were from Australia, China, Korea, Japan, Malaysia, the Philippines, Taiwan, Thailand, and the US. Results show that the system can be used in different Asian languages. In total, 58 EML documents were collected from the workshops. Most documents contained the minimum content required for adequate dataset discovery (Fig. 6). Twenty-eight documents (48%) reached the access level which includes details to support automated data retrieval. But none of these 58 EML documents yet supports the level of computer-assisted data integration and re-sampling.

Table 1. Workshops held to test the suitability of the system

Date	Location	Workshop type	Languages coded	No of participants
Aug. 2004	Taiwan	Domestic	2	20
Dec. 2004	Taiwan	Domestic	2	25
Feb. 2005	Taiwan	Domestic	2	20
Apr. 2005	Taiwan	Domestic	2	20
July 2005	China	International	5	80
Sept. 2005	Taiwan	Domestic	2	25
Oct. 2005	Taiwan	Domestic	2	25
Dec. 2005	Taiwan	Domestic	2	25
Dec. 2005	Taiwan	Domestic	2	10
Jan. 2006	Taiwan	Domestic	2	25
Feb. 2006	Taiwan	International	6	40
May 2006	Taiwan	Domestic	2	20

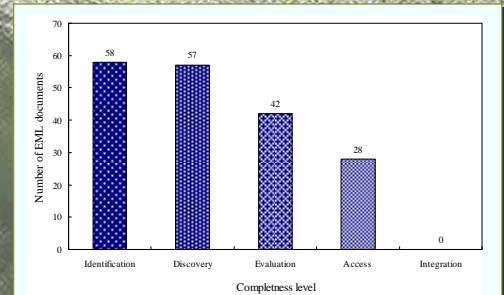


Fig. 6. Implementation of Ecological metadata Language (EML) documents

Conclusions

The compiled system should be useful to the ecological community of the Asian region for storing, managing, and retrieving research data through the metadata. It is obvious that our system provides the requirements of data legacy that needs sufficient documentation for tomorrow scientific use. Although the current system has not immediately solved the problem of integrating ecological data, it is a new method for data discovery, access, and analysis. In addition, the framework not only adopts the EML standard for documenting the structure and content of datasets, but has also developed data access, query, and analysis functions to facilitate the use of EML documents.

The data analysis function module consists of XSLT style sheets that translate EML documents into statistical programs (Fig. 4). The interface not only extends the capabilities of the transformation but also has become a prototype of a server-side system that allows researchers to access EML, upload data, and then run "R" code on the server. User can use it for data manipulation, calculation, and graphical display online without the need to install "R" locally.

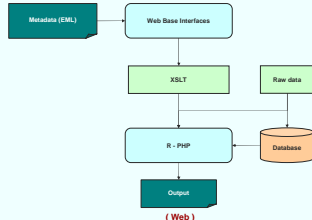


Fig. 4 Statistics module

The GIS function module (Fig. 5) uses DOM (Document Object Model), a platform- and language-neutral interface that can incorporate the result back into the presented page. The module takes information from an EML document and automatically displays locations on Google Map or Google Earth.

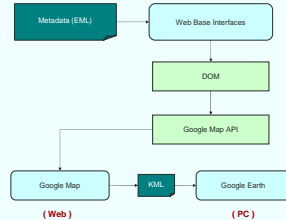


Fig. 5 GIS module