

Application of the Inventory of Biodiversity Information and Social Networking Based Collaboration: An Implementation of Software Framework at Web Application Level

Hsin-Hui Wu, Richard Mayden

Department of Biology
Saint Louis University
St. Louis, MO 63103, USA
{hwu5, maydenr1}@slu.edu

Kevin Scannell

Department of Mathematics and Computer Science
Saint Louis University
St. Louis, MO 63103, USA
kscanne@gmail.com

Abstract— Biodiversity information is inherently complex and is highly distributed in collaborations from all over the world. The application of inventory of Biodiversity Information is a content management system and an end-user may have a technical threshold in maintaining this kind of application. We apply the idea of a software framework at the web application level and implement it as practical web application: Cypriniformes Commons (CypsCom), one inventory solution to store related biodiversity information for the fish Order Cypriniformes. In CypsCom, a biologist can provide multiple layers of information on a taxon through this web-based application and other biologists can review these contributions in an orderly scientific process built in CypsCom. Deploying this kind of software framework at the web application level makes it possible to implement this framework in different programming languages and be applied by different biodiversity research communities.

Keywords— *inventory of biodiversity information; social networking; software framework; collaborative review process; blog application*

I. INTRODUCTION

Biological information on any taxon composing biodiversity is essential information in our understanding of the catalogue of life. This kind of information provides basic for the identification of a taxon and associated support information such as occurrences, distribution, and other biological information. These kind of data are key components for further in-depth explorations in biology-related research [1]. However, there is no one standard and fixed storage structure system because of the variety and diversity of information desired for different taxa in different fields. Moreover, information is always updated dynamically due to continuous scientific discovery. For example, when a new characteristic is discovered or new evidence has been shown that the current description is no longer the most accurate depiction, related changes will be proposed, discussed, “recorded” and perhaps modified. In other words, how this information is updated and consistently maintained for information inventory structure is an important research issue in the developing field of biodiversity informatics.

Reviewing current developments in biodiversity informatics, in terms of relational database applications,

biologists representing different sub-fields have built database applications for decades. Large organizations have built centralized database applications in species information such as FishBase, eFlora and Actkey [2], [3]. On the other hand, some organizations have developed an integration standard or communication protocol to gather dispersed data sources on the web which are built by individuals or small organizations and communities. Global Biodiversity Facility (GBIF) has successful integrated millions of specimen and species checklist information [4], [5]. However, this integrated information inventory standard cannot contain all of the kinds of biodiversity information needed in all biological research fields. Moreover, these systems also lack an integrated process to rapidly collect and document new species information via cyberspace involving a community of professionals. How to keep track of new information with new discoveries under the evolving and rapid processing cycle is also a key issue in biodiversity informatics. It also needs academic review procedures for evaluating and documenting that new data are legal, especially in taxonomic information provenances. Emerging social networking technologies offer excellent opportunities to serve as a platform in the scientific review of information contributed by professionals via an open, public interface designed for sharing, exchanging and communicating, and driven by the members of the scientific community.

The main idea of this paper is applying the idea of a software framework at the web application level and implementing it as practical web application: Cypriniformes Commons (CypsCom), one inventory solution to store related biodiversity information under fish Order Cypriniformes, the most diverse order of freshwater fishes. In CypsCom, any biologist can provide multiple layers of information evidenced on a taxon through this web-based application and other scientists can review contributed information via a natural scientific process built in CypsCom. A flexible database structure serves as the meta pattern in the software framework. Biodiversity information is stored in two separate database tables: taxon account table and taxon document table. The application administrator can configure the inventory of Biodiversity Information structure in a metadata table. Another software framework design is a social networking-based collaborative tool to simulate the

scientific review process. This design is derived from the idea of a blog application with additional database design to control a review process. This hybrid review process can validate contributed data via professional reviewers and community contributors at the same time. Deploying this kind of software framework at the web application level makes it possible to implement this framework in different programming languages and be applied by different biodiversity research communities.

The structure of this paper is described as follows: First, we review related developments in the second section. Later, we propose the system architecture in the third section. System implementation and discussion are located in the fourth section. The last section is the conclusion and further development plan.

II. RELATED WORKS

A. Review of Current Solutions on Inventory of Biodiversity Information

1) *Application Solutions*: These include Client-based and Web-based Systems. These systems have fixed database design structure and are mostly designed by non-profit organizations, or a biologist. Most are used for a specific field or purpose. For example, the application Specify is a well-known open source biodiversity collection software that supports biological collections with data management software, data conversion services, helpdesk support and training [6]. Morpho is an ecological data management software supported by the Knowledge Network for Biocomplexity (KNB) [7]. To meet the needs of more complicated data structure of biodiversity and ecological information, this software must address the adoption of extensible database design. GBIF Integrated Publishing Toolkit is a newly published software using a java web framework to host biodiversity data and maintain data connectivity with the GBIF portal [8]. It also provides some integrated interfaces that can display collection occurrences, distributions, and statistical information. Bauble is a biodiversity collection application developed by a single programmer [9]. Content Management System (CMS)-Based Solutions are also adopted. They attempt to meet the requirements for different and dynamic data structures from various client users. A fundamental approach taken by these CMS's is to use metadata to structure their content. The advantage of metadata is that the content can be evolving and dynamic and even the content structure is always changing. Encyclopedia of Life (EOL) applies a CMS system, Drupal, to be their client data collection system, especially in its powerful taxonomy module [10]. The Natural History Museum, London applies Drupal to host an open source site "Scratchpads" that offer free space for scientists who are interested in online management, sharing and publishing of their taxonomic data [11].

2) *Storage Format Solution*: The structure of biodiversity information is dynamic in different fields.

Species 2000 is a loosely federated database standard that provides for a collaboration among different taxonomists and sponsoring agencies under the same standard [12]. GBIF employs Darwin Core, a metadata standard for historical and natural digital objects, and used to index the species distributions and specimen information [13]. These open standards give the individual biologist a method by which to publish their data in a public forum with minimal interaction between published and original data.

B. Review of Scientific Collaboration in Validating Biodiversity Information

This process includes the following behaviors: propose, review, approve, publish, expire, and archive. Each taxon is regarded as a unique account in a data bank. The account is created or updated with new discoveries or new characteristic information is explored. The review process will approve new observations. If it is approved, it will be published and added to the account. At the same time, the old one will be expired and archived to an historical storage pool. This is a dynamic and ongoing process.

Moreover, biodiversity information must be kept updated and made available for public discussion. Participants' interactions in their communities play a critical role in scientific communication. Either a scientific review process or a public discussion in community can promote the validation of new information on a taxon. To ensure the updated information is valid, a well-structured communication platform is essential for public discussion. In addition, this platform also needs to provide the function that can keep track of a changing history for each species [14].

C. Software Framework

A software framework is a reusable and semi-complete software application. It can be customized by a user-defined domain [15]. The software framework defines the specification at the architecture level and should also provide building procedures to create software applications [16]. At the code level, software will be composed of two parts: frozen spots and hot spots. Frozen spots are unchanged in the application instantiation stage; hot spots are made by a programmer during the instantiation stage [17]. Another feature of software framework is "Inversion of Control." This involves a software framework that will react with external input. When software receives input, the dispatcher of the software framework invokes its related, defined functions to reflect the input [18]. The software framework is also component-based software development in the field of software engineering [19]. In this paper, we apply the concept of software framework such that the program includes several software components [20].

III. APPLICATION DESIGN

A. Architecture of CypsCom

To design in component-based software, CypsCom is modularized into several parts such as data storage, scientific collaboration, etc. Taking scientific collaboration for

example, the user in the community can register as a member via simple registration procedure. A member can also volunteer to be the contributor or reviewer to contribute their knowledge of biodiversity information to CypsCom. The data storage provides a user-friendly interface allowing users to type/import data. Scientific collaboration provides the interface to post new information. Keeping newly posted information valid, a review procedure involves reviewers. The application architecture is shown in Figure 1.

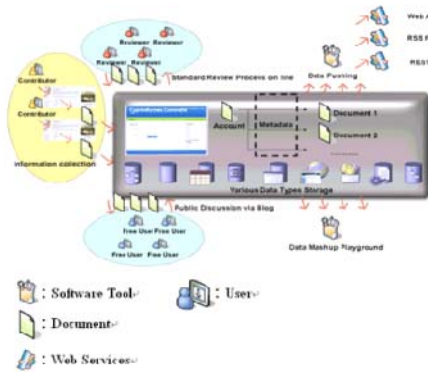


Figure 1. Architecture diagram of CypsCom.

B. Software Framework Design on Web application Level

Software framework is implemented in database design and interface design generated by the program. In database design, we separate biodiversity information into two parts: the first part is the core data, taxon account, and the second part is the supporting descriptive data, taxon document. The core data are fixed to store the taxon hierarchy information and the supporting descriptive data are kept dynamic to store the related information such as habitat, distribution, behavior, etc. To keep the data structure dynamic, we designed an additional “document metadata table” to control the structure of document table.

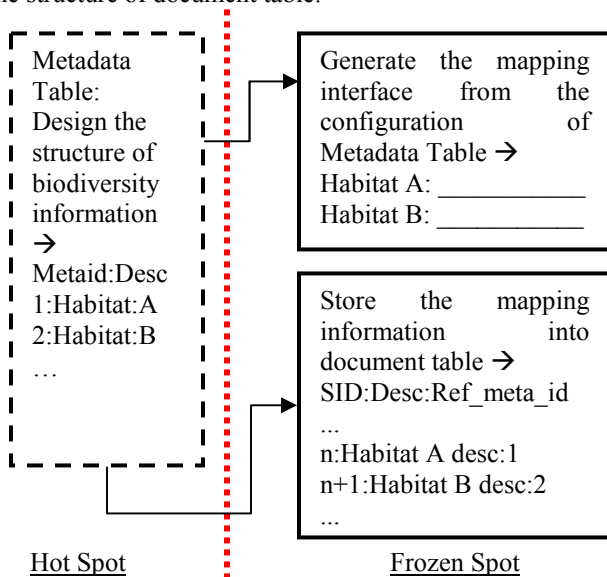


Figure 2. Software Framework Design on Data Storage Module.

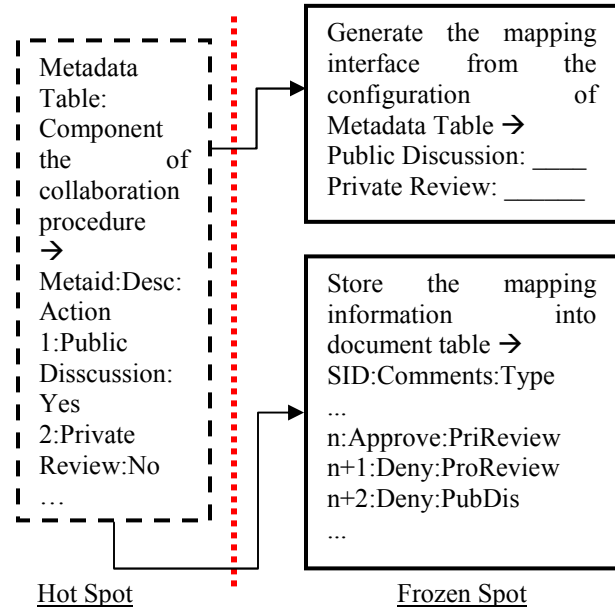


Figure 3. Software Framework Design on Scientific Collaboration Module.

IV. APPLICATION IMPLEMENTATION

The web front end to CypsCom is implemented in PHP

A. Administrator Functions

The administrator of CypsCom must establish a metadata table initially to define the supporting descriptive data structure. Within CypsCom, two kinds of data are stored: core and supporting descriptive data. Core data are in a database table storing basic taxon information such as the hierarchy, name, author and associated information. The structure of the supporting descriptive data is dynamically controlled by the definitions in the metadata table. The administrator also functions in approving all new biodiversity information for validation.

B. Registered User Functions

A user in the community can register for an account and log into the system to contribute data to te CypsCom. Users can volunteer to be the author, reviewer, or editor for specific taxon account(s). CypsCom automatically gives the user a period of time to complete their contribution before it is returned to the pool of available taxa for contributions. Users can choose to review and validate new information on a taxon, a fundamental need in assisting the community.

C. Registered Reviewer and Review Procedure Function

When a user contributes taxon information to CypsCom and proceeds to the submission stage, the system will send a message to the administrator to initiate the review process. An automated process will select the appropriate reviewer(s) and editor(s). An automated event will schedule an appropriate period of time for the review cycle. An automated event will facilitate blog communications. Registered users may post opinions and offer feedback in the

collaborative process. At the end of the review period, a systematic event will collect the related information and provide a “current view” for all associated collaborators to review. The review process flow is shown in Figure 4.

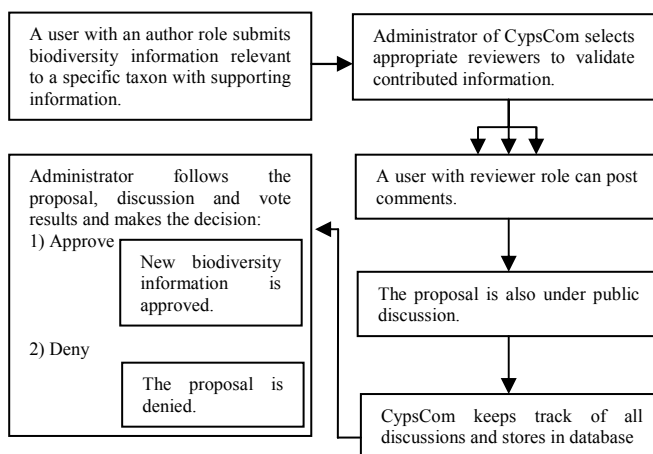


Figure 4. Scientific collaboration example: review procedure.

V. CONCLUSION

In this paper, we design a software framework for the inventory of biodiversity information and implement it in a practical web application using PHP scripting language. At the web application level, we designed a flexible database structure by separating the taxon information into two parts: account and document. With the design of a metadata table, it can generate dynamic structure in the document. With the support of the metadata table, we can store the different structures of data by adjusting the metadata table. In scientific collaboration, we also use a metadata table to configure components of standard scientific procedures such as public discussion, private reviews and committee review.

Another feature of CypsCom is the integration of the academic scientific collaboration process and social networking. Our blog interface provides for a public discussion by all registered users and a chronicle-based discussion is applied in a review mechanism for validating newly proposed biodiversity information. Based on the implementation software framework, CypsCom is easy to implement in different programming languages and has the potential to be adopted in different biodiversity research communities.

ACKNOWLEDGMENT

The research of the paper was supported by the National Science Foundation (NSF) research awards to Richard L. Mayden, Saint Louis University: EF 0431326. AToL:

Collaborative Research: Systematics of Cypriniformes, Earth's Most Diverse Clade of Freshwater Fishes; DEB 0817027. Inventory and Evolution of Fishes from Streams of the Sierra Madre Occidental, Mexico; and EF-0431326. Supplement to CTOL project by NSF.

REFERENCES

- [1] S. K. Wiser, P. J. Bellingham, and L. E. Burrows, "Managing biodiversity information: development of New Zealand's National Vegetation Survey databank," *New Zealand Journal of Ecology*, vol. 25, pp. 1-17, 2001.
- [2] Froese, R. And D. Pauly, eds., *Fishbase*. International Center for Living Aquatic Resources (ICLARM), 2004.
- [3] A. Brach and H. Song, "ActKey: a Web-based interactive identification key program," *Taxon*, vol. 54, pp. 1041-1046, 2005.
- [4] *Global Biodiversity Information Facility*. Available: <http://www.gbif.org/>
- [5] A. Kirchhoff, A. Hahn, J. Holetschek, P. Kelbert, R. Jahn, and W. Berendsohn, "Open Access to Biodiversity Collection Data—GBIF Germany and the Botanical Node," 2007.
- [6] *Specify Software Project*. Available: <http://www.specifysoftware.org/Specify>
- [7] D. Higgins, C. Berkley, and M. B. Jones, "Managing heterogeneous ecological data using Morpho," in *Scientific and Statistical Database Management, Proceedings. 14th International Conference on*, pp. 69-76, 2002.
- [8] *GBIF Integrated Publishing Toolkit (IPT)*. Available: <http://ipt.gbif.org/index.html>
- [9] *Bauble Biodiversity Collection Manager*. Available: <http://bauble.belizebotanic.org/>
- [10] S. Mayo, R. Allkin, W. Baker, V. Blagoderov, I. Brake, B. Clark, R. Govaerts, C. Godfray, A. Haigh, and R. Hand, "Alpha e-taxonomy: responses from the systematics community to the biodiversity crisis," *Kew Bulletin*, vol. 63, pp. 1-16, 2008.
- [11] *Scratchpads: Biodiversity Online*. Available: <http://scratchpads.eu/>
- [12] *Species 2000 Standard*. Available: <http://www.sp2000.org/>
- [13] D. Vieglais, *The Darwin Core. Revision 1.5*. Available: <http://speciesanalyst.net/docs/dwc/>
- [14] D. Agosti, C. Klingenberg, G. Sautter, N. Johnson, C. Stephenson, and T. Catapano, "Why not let the computer save you time by reading the taxonomic paper for you?," *Biológico, São Paulo*, vol. 69(2), pp. 545-548, 2007.
- [15] R. Johnson, B. Foote, "Designing reusable classes," *Journal of Object-Oriented Programming*, vol. 1(2), pp. 22--35, 1998.
- [16] F. Buschmann, R. Meunier, H. Rohnert, P. Sommerlad, M. Stal, "Pattern-oriented software architecture: a system of patterns," John Wiley & Sons, Inc., New York, NY, 1996
- [17] W. Pree, "Meta Patterns - A means for capturing the essential of reusable object-oriented design," *Proceedings of the 8th European Conference on Object-Oriented Programming, Bologna, Italy, 1994*.
- [18] C. Larman, "Applying UML and Patterns: an Introduction to Object-Oriented Analysis and Design and Iterative Development," Prentice Hall PTR, 1994
- [19] R. Pressman and D. Ince, "Software engineering: a practitioner's approach," McGraw-Hill, New York, 1997.
- [20] L. Li and I. Horrocks, "A software framework for matchmaking based on semantic web technology," *International Journal of Electronic Commerce*, vol. 8, pp. 39-60, 2004.