FishTraits Version 2: Integrating Ecological, Biogeographic and Bibliographic Information

Zhiwu Xie
University Libraries
Virginia Tech
Blacksburg, VA 24061
zhiwuxie@vt.edu

Emmanual A. Frimpong
Department of Fish and Wildlife
Conservation
Virginia Tech
Blacksburg, VA 24061
frimp@vt.edu

Sunshin Lee
Department of Computer Science
Virginia Tech
Blacksburg, VA 24061
sslee777@vt.edu

ABSTRACT

In this paper we describe the new development of FishTraits. Originating from an ecological database that documents and consolidates more than 100 traits for 809 fish species, the new version focuses on the integration of these traits data with the bibliographic and biogeographic information. We explain the overall design as well as the implementation details.

Categories and Subject Descriptors

H.3.7 [**Information Storage and Retrieval**]: Digital Libraries

General Terms

Design.

Keywords

Digital library; architecture; specie traits.

1. INTRODUCTION

Species traits offer instructive lenses through which to view spatial and temporal variation in the distribution and abundance of species. Ecologists have a long history of using traits to study patterns of species distribution, abundance, and responses to anthropogenic impacts. Partially funded by the U.S. Geological Survey Aquatic Gap Analysis program, FishTraits version 1 compiled and consolidated more than 100 traits for 809 (731 native and 78 nonnative) fish species found in freshwaters of the conterminous United States, including 37 native families and 145 native genera [1, 2]. It contains information on four major categories of traits:

- trophic ecology

- body size, reproductive ecology, and life history

- habitat preferences

- salinity and temperature tolerances

Information on geographic distribution and conservation status was also compiled. The first version of the database [1] was released as a web application in 2009 and soon became the most cited web resource in the area.

The limitations of the original web development also started to

surface soon after its public release. Due to technical difficulties, the website was briefly taken offline in 2011 and was later resumed. Starting from 2012, a redesign and rebuild of both the frontend web interface and the backend database was underway. This paper briefly describes the architectural design as well as the implementation details of FishTraits Version 2.
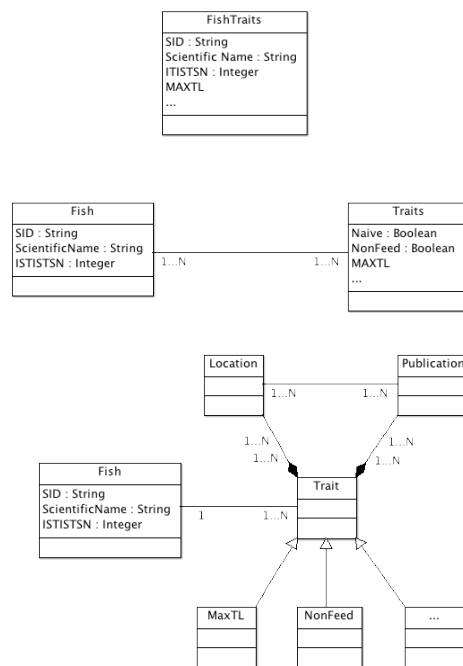

Figure 1. Evolving FishTraits Data Model.

2. INFORMATION INTEGRATION

The first version of FishTraits reports the consolidated data from various sources. These sources include the state and regional fish accounts, three general texts, and numerous specialized texts, peer-reviewed articles, technical reports such as species recovery plans, online catalogs, and state and federal agency websites. In version 1 design, a "NOTES" field was included to document the original sources and how these sources were consolidated to reach the resulting data values. This field, however, was not populated in version 1 and was left for future work.

During the version 2 development multiple limitations of the original design were exposed through the use-case analysis: First, a combined "NOTES" field cannot clearly differentiate the original sources from the consolidation process. Second, the per-

species granularity is too restrictive for the much more complicated referencing and interlinking relations between the traits and the resources.

We started by separating the bibliographic information from the NOTES field, and designate a N-to-N relation between the Species and the Publication objects. Further use case analysis exposed more complicated scenarios. In particular, since the state and regional fish accounts are organized by political boundaries whereas species traits transcend political and biogeographic boundaries, multiple resources may report different trait values on the same species but in different locations. This calls for the data model to further evolve. Figure 1 illustrates some of the data model changes during the development.

Bibliographic information is an important part of the Version 2 development. Although we need many functionalities from a bibliographic information management system, such as cross-referencing, it would be overkill to build a full-fledged system just to manage the publications. We decided to leverage the existing developments in the field and build an interface with Zotero to import bibliographic information through the standard BibTeX format.

As illustrated above, biogeographic information can easily complicate the architectural and the model design. We therefore need not only a flexible data model, but also sufficient geographic information support at the database level. This puts restrictions on our technology options. Furthermore, we expect the data model to continue evolving when more complicated relations are discovered between traits, their referencing publications and geographic locations. Such evolution usually put pressures on technology development because most frontend code and backend queries may need to be fully rewritten if data model changes.

FishTraits Version 1 was a simple ASP application built on Access database and IIS. It lacks many crucial web components and cannot handle heavy query load. When the original hosting server was upgraded, many old technologies became obsolete and the support was eventually dropped. Learning from the past experience, we decided to rebuild the system with mature, stable, and well-adopted open source technologies. We picked Python/Django as our web framework due to its excellent capabilities to evolve code with the data model. Its Object-Relation Model allows the development to be much more agile and flexible. As for the geospatial database, we chose the open source PostGIS, which adds spatially-explicit biogeographic information support to PostgreSQL database and allows many routine geospatial operations. Figure 2 shows the query interfaces of the new version. Compared with its predecessor, the new version also supports autocomplete that allows the users to easily retrieve data values including the species names.

FishTraits Version 2 will eventually export all its data to be merged into USGS's computing infrastructure, but the frontend functionalities will not be replicated. It will continue to run as a standalone web application. FishTraits Version 2 also adds support to easily backup and restore from VTechWorks, Virginia Tech's institutional repository via the SWORD protocol [3]. Through VTechWorks planned interfacing with DataOne, FishTraits data may be further disseminated.
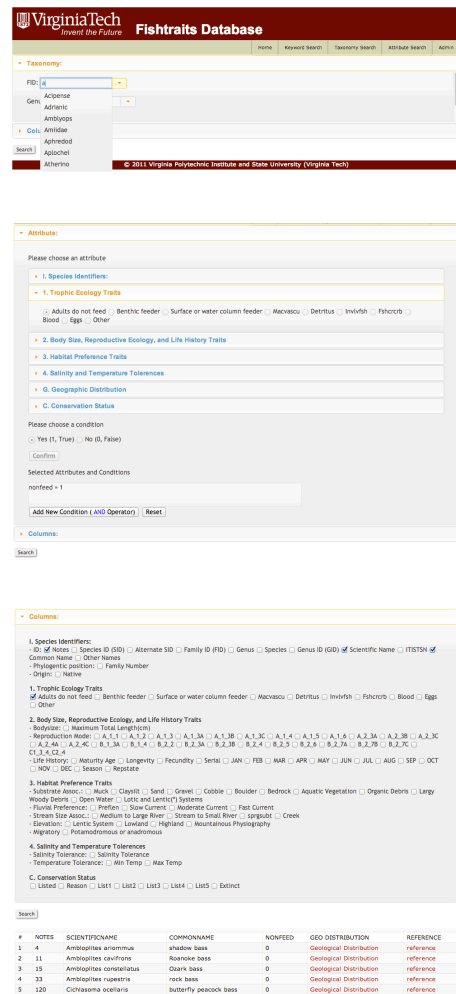

Figure 2. FishTraits Version 2 Screenshots.

3. CONCLUSION

In this paper we describe the new development of FishTraits. Our emphasis is on its architectural design and implementation details that integrate ecological information with biogeological and bibliographic information.

4. ACKNOWLEDGEMENTS

5. REFERENCES

[1] FishTraits. http://ww.fishtraits.info

[2] Frimpong, E. A. and Angermeier, P. L. 2009. Fish Traits: A Database of Ecological and Life-history Traits of Freshwater Fishes of the United States. *Fisheries*, vol. 34, no. 10, 2009, 487–495.

[3] Allinson, J., François, S. and Lewis, S. 2008. SWORD: Simple Web-service Offering Repository Deposit. *Ariadne*, Jan, 2008.