BUILDING A BIODIVERSITY CONTENT MANAGEMENT SYSTEM FOR SCIENCE, EDUCATION, AND OUTREACH

CS Parr[1]*, R Espinosa[2], T Dewey[3], G Hammond[3], and P Myers [3,4]

[1]*Human-Computer Interaction Lab, UMIACS, Univ. of Maryland, College Park, MD 20742*
Email: csparr@umd.edu
[2]*Information Technology Central Services, University of Michigan, Ann Arbor, MI 48105*
Email: roger@umich.edu
[3]*Museum of Zoology, Univ. of Michigan, Ann Arbor, MI 48109*
Email: {lqb, gstarrh}@umich.edu
[4]*Department of Ecology and Evolutionary Biology, Univ. of Michigan, Ann Arbor, MI 48109*
Email: pmyers@umich.edu

ABSTRACT

We describe the system architecture and data template design for the Animal Diversity Web (http://www.animaldiversity.org), *an online natural history resource serving three audiences: 1) the scientific community, 2) educators and learners, and 3) the general public. Our architecture supports highly scalable, flexible resource building by combining relational and object-oriented databases. Content resources are managed separately from identifiers that relate and display them. Websites targeting different audiences from the same database handle large volumes of traffic. Content contribution and legacy data are robust to changes in data models. XML and OWL versions of our data template set the stage for making ADW data accessible to other systems.*

Keywords: Database design, Evolving schema, Education, Ontologies, Biodiversity, Interoperability

1   INTRODUCTION

Recent years have seen an explosion of digitally available information about biological diversity (Bisby, 2000). At this stage in the field of biodiversity informatics, there are multiple, often redundant databases, and work has begun in earnest to establish standards to allow them to be federated so that information retrieval across sources can be efficient. At the same time, access to natural history data about organisms is important to three distinct audiences with different needs: 1) the scientific community, especially those seeking coded data for large scale ecological or organismal analyses, 2) educators and learners in formal education settings, and 3) the general public. Our challenge has been to design a system that efficiently accommodates the data needs of these audiences. Below we describe our project, the Animal Diversity Web, and detail the implementation of a system architecture and data template design that supports highly scalable resource building and flexible delivery. The details of our architecture and data template design may serve as models to other biologists designing knowledge bases. In addition, though our system was not designed explicitly for interoperability, we believe that the technology is now available to make the contents of our database accessible to other computer systems.

1.1   Animal Diversity Web

The Animal Diversity Web (ADW) is an online resource providing information on extant taxa in the kingdom Animalia from all over the world. Content includes media, text, keywords, quantitative fields describing basic natural history and conservation status, a glossary, and a taxonomic database used for validating and organizing content. A large part of the content is provided by university undergraduates who submit reports on species as part of their course requirements. This content is edited by their instructors, and then edited again by a team of biologists at the University of Michigan. Experts at the University of Michigan and elsewhere provide content at higher taxonomic levels. The ADW project currently maintains

two parallel websites – the ADW, aimed at adults and intended primarily for undergraduate education and outreach, and the BioKIDS Critter Catalog, aimed at 10 to 12 year olds involved in an inquiry-learning biodiversity curriculum.

1.2   System requirements

A truly scalable, flexible biodiversity information system meets four main requirements: 1) it supports large numbers of authors and editors, 2) it allows managers to modify or add new data models (e.g. add, split or lump keywords, add new conservation lists or a physiology section, etc.) while preserving the integrity of legacy data, 3) it allows managers to deliver content to audiences with differing levels of subject expertise, or to otherwise change presentation at will, and 4) it supports sophisticated querying for inquiry learning or data harvesting for scientific studies.

1.3   Related work

Many web sites are designed to deliver natural history information about organisms. FishBase (Froese & Pauly, 2004) and AmphibiaWeb (2004), for example, provide in-depth information on particular subsets of related taxa. FWIE (Fish and Wildlife Information Exchange) (Conservation Management Institute, 2001) Master Species File, Ocean Biographic Information System (OBIS) (OBIS, 2004), and Global Biodiversity Information Systems (GBIF) (GBIF, 2005) have a broader taxonomic scope. These systems, maintained in large relational databases (distributed then federated in the cases of OBIS and GBIF), are created by and aimed primarily at experts. They often rely on an extensive controlled vocabulary of technical terms that is relatively static. The Tree of Life website includes full text descriptions more accessible to broad audiences. Its focus is on conveying information on evolutionary relationships among organisms and the characteristics supporting those hypotheses of relationships. The distributed nature of this system (taxonomically-related pages are maintained by experts on their local systems, then federated) offers high scalability.

A content management system similar to ADWs has been developed at University of Washington (Cherry, Washington, Fournier & Shuyler, 2003). Its goal is to provide a flexible learning platform supporting multiple authors. This system, zBento, is designed to accommodate multiple domains, but not multiple audiences. In addition, zBento is not explicitly designed to maintain long term data using evolving data models. SenseLab (Marenco, Tosches, Crasto, Shepherd, Miller & Nadkarni, 2003), uses an evolvable system designed to provide web access to an expert-oriented neuroscience database that is part of the Human Brain Project. Their semantic tagging approach is similar to ours.

2   IMPLEMENTATION

2.1   System implementation

The ADW approach can best be summarized as an application of the "loose coupling" philosophy (Weinberger, 2002) to content management. Content objects, or nodes, e.g. a photograph, sound, or other rich media file, or a paragraph of text or keyword pertaining to an organism, are managed together in a single object-oriented database. They are coupled, or related, by three kinds of identifiers. Semantic identifiers are concepts defined in an ontology/thesaurus and used to tag a node. From the user's perspective, this occurs via the process of filling out a data template. Taxon identifications tag the node with its biological taxonomic source -- a species or a higher level biological name. A route identifier, such as which audience education level or geographic region should see the node, specifies which website the node should appear in.

The "looseness" of the coupling refers to the fact that nodes are, in effect, managed separately from the identifiers used to relate and display them. Contributors and editors can manipulate the nodes and staff can modify data templates, taxonomic sources, and site display stylesheets. In practice, each tag on a node is merely an id number that points to the definition of the identifier.
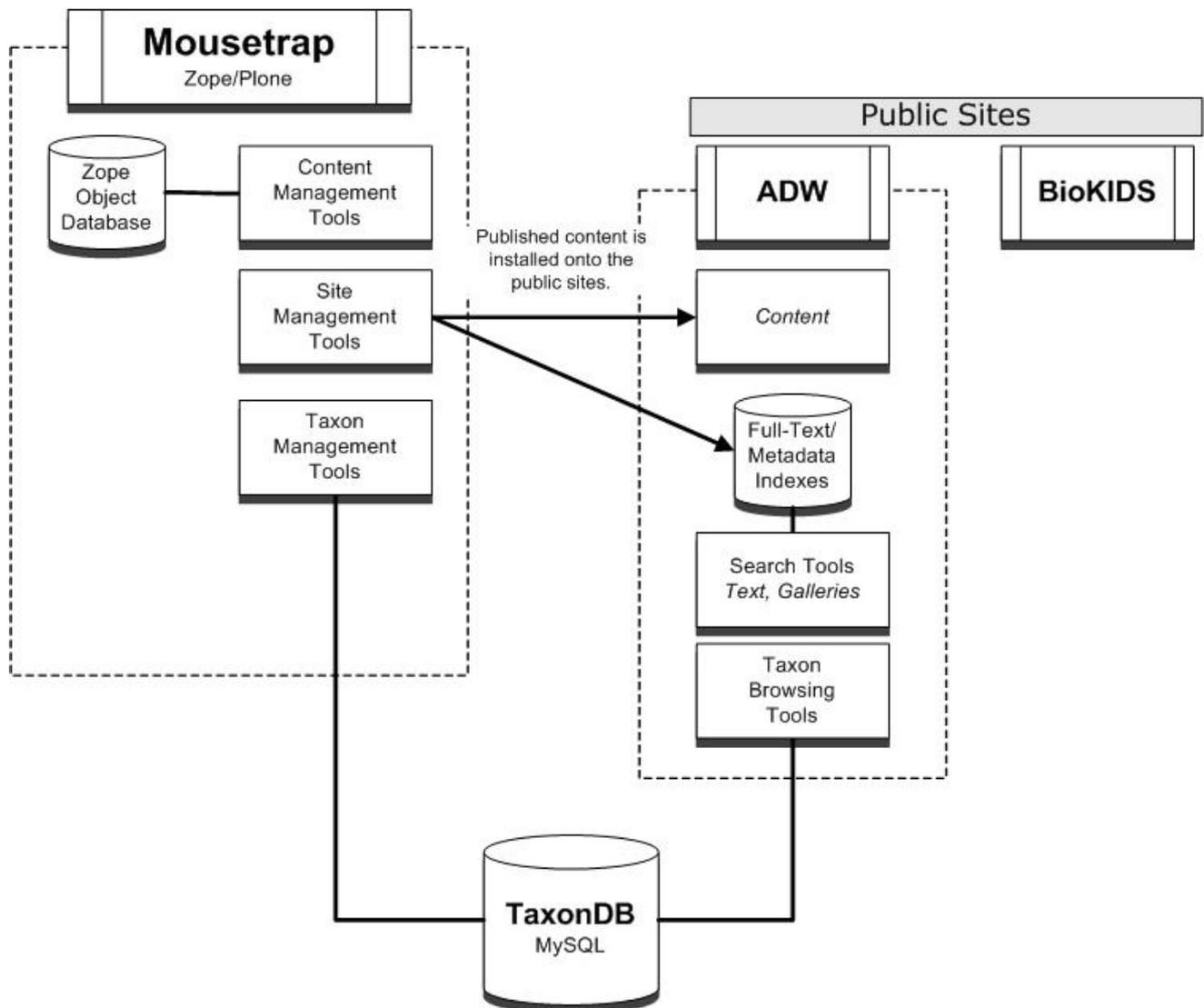
Figure 1. ADW architecture. Mousetrap is our online development environment, providing tools to allow contributors and editors to manipulate content. TaxonDB is a relational database providing both a taxonomic authority for content developers in Mousetrap, and a means of browsing the public sites taxonomically. The public sites are the content-rich pages and searching and browsing tools available to the general public, each customized to different audiences. As an example, the ADW site is expanded to show its subparts.

Our system architecture is shown in Figure 1. Mousetrap, available online to registered contributors and editors, is a customization of the Plone content management system. Its content management tools provides services to manage contributor information and access, file uploading and image processing, content metadata, and routing of nodes to particular websites. Nodes are managed as Zope objects. Mousetrap provides tools for customizing our public sites, such as style sheets. It also includes tools for managing the taxonomic database. TaxonDB is a MySQL relational database of biological names and their hierarchical or parent-child relationships. TaxonDB was built by integrating a number of publicly available datasets (Parr, Lee, Campbell & Bederson., 2004). It serves both as an authority for taxonomic identification and as a source of page organization in the published sites. Support in TaxonDB for multiple hierarchies provides flexibility in how we present the tree of life. The public sites, each built to serve a particular audience, are the third major part of the system. Figure 1 shows our current ADW and BioKIDS sites, but any number of targeted sites are possible. The sites share some

content, but also house content and tools specific to their intended audiences. Full-text and metadata searches for public sites are serviced by Swish-E indexes（http://swish-e.org）.
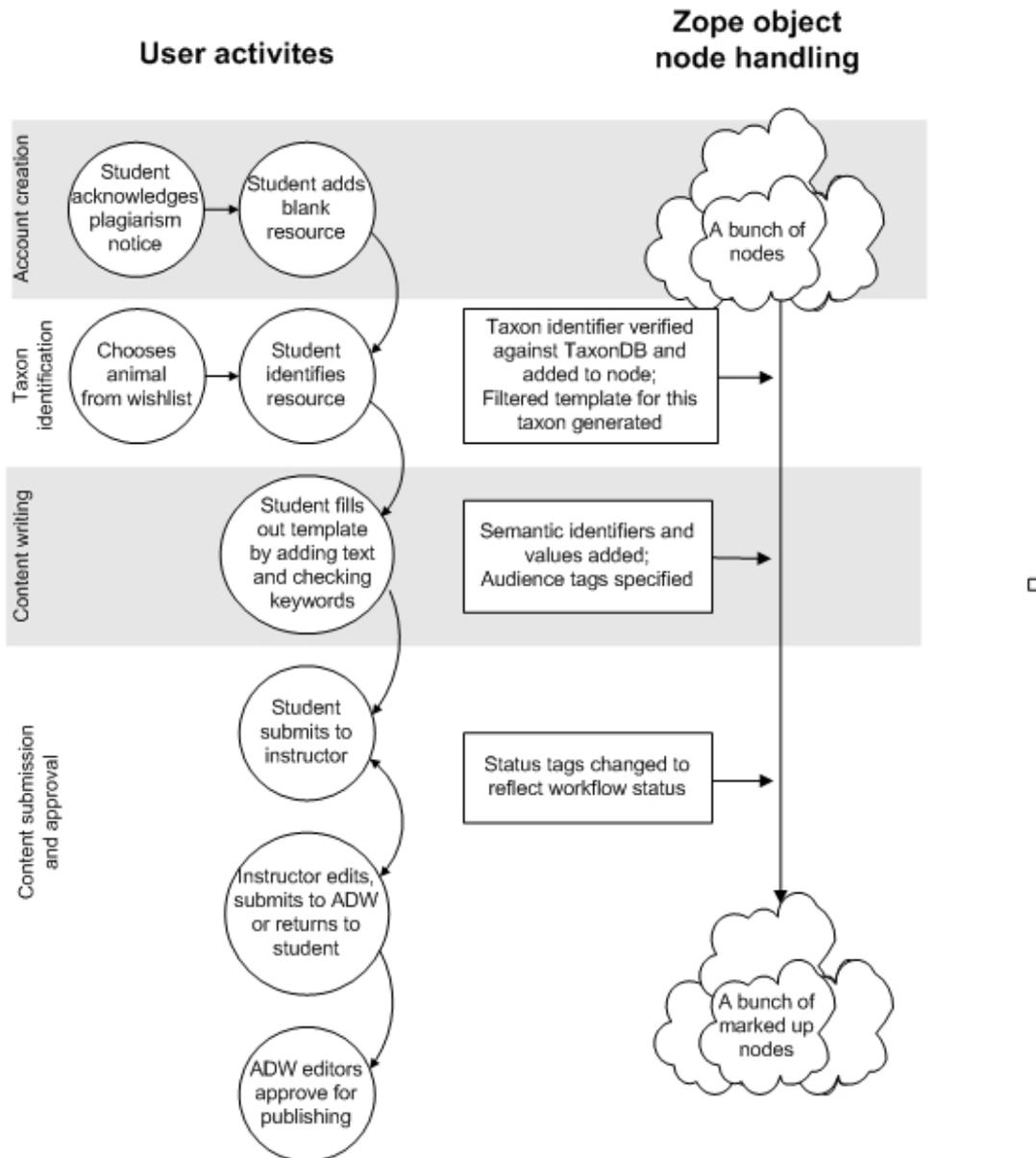

Figure 2. Content creation workflow in ADW. Workflow from the user perspective is shown on the left. Workflow from the content management system perspective is on the right. A "resource" is any content type managed by the system; this could be a media file or a special topic page or, as shown in this example, a taxon account.

Figure 2 shows the workflow that supports the creation and manipulation of nodes by contributors and editors. First, student contributors must check a box that they have read the policy against plagiarism. Each resource node or collection of nodes (a media file, taxon account, etc.) must be identified with a valid taxonomic identifier from TaxonDB. Students may choose an animal from our "wishlist" of names for which we have media but no text account, or they may enter any name of their choosing. Mousetrap prevents redundant taxon accounts and enforces that Latin name spellings be consistent with our authority. If the name is not in our taxonomic database (which is large but not comprehensive) an editor must approve it before the student can proceed.

If the node is to be a taxon account, an appropriate data template is generated which is customized for the taxonomic group (see *taxon filters*, below). By filling in a blank data template including checking keyword boxes, contributors are actually creating a collection of text nodes and adding semantic identifiers to them. Editors may indicate the audience of a particular paragraph is college-level only and add a parallel paragraph appropriate for other audiences. Appropriate status tags are added as the nodes pass through the workflow from contributor to instructor to ADW editors, ensuring that content is available to appropriate people for modification. Other kinds of nodes (sound clips, photographs) may have entirely different templates but pass through similar workflow.

After being entered and edited on Mousetrap, content is published via the following process. The simple plain-text markup language our contributors use (based on reStructuredText) is rendered into HTML. Search indices are updated to allow searching, and external and internal links are created and tested. The account is transformed to semantically mark up the content, resolving pointers to identifiers, and routed to the presentation stylesheet appropriate for each site. The system also uses TaxonDB to organize the resources for display in a Linnean hierarchy. Dynamic pages (e.g. image galleries and "feature" pages) are first generated when a user asks for a page, through a servlet reading the indices. The page is then written to the filesystem so the next request is static. Thus the public sites write themselves as they are used.

Because legacy data templates remain archived, legacy data remains semantically related. For example, in the current version of our data template, we might attach the semantic identifier "hermaphrodite" to nodes identified to the taxon "woolly slug." Later, we might decide to require contributors to specify "simultaneous" or "sequential" hermaphrodism, so we alter our data template. The legacy data entered under the previous template remains semantically tagged, so queries can still find these nodes and we can continue to display them, if we desire, in appropriate places on a web page. In addition, the new keywords are available to editors of legacy content. We may decide that this information is not appropriate for display to younger audiences, and so remove routing identifiers so as not to show it to them, or to show them a simpler synonym. We may re-identify all the nodes from the "woolly slug" to a more recent name simply by managing the taxonomic database.

The integration of semantic markup and nodes occurs at the lowest level; most of the system works with this combined XML so this approach could therefore be achieved in any environment with good XML support.

2.2   Data template implementation

Taxon accounts form the core data objects in the ADW natural history database.   Information in the current taxon account template is organized into as many as 18 sections describing important aspects of animal biology. Section topics include distribution, physical description, reproductive biology, lifespan, behavior, food habits, predators, ecosystem roles, economic importance to humans, and conservation status. Template section choice was driven primarily by the goal of organizing the incredible breadth of natural history patterns in the animal kingdom into manageable, related pieces that could be consistently recorded across a wide range of animal taxa. The organization of the template in this way facilitates two activities. First, it allows the use of the ADW by both scientific researchers and educators as a source of data on animal behavior, ecology, and evolution. Second, it supports reliable addition of new content to the ADW by student contributors who are not technical experts and often lack access to some kinds of sources. For example, although we could add sections to the ADW data template covering population genetics, physiology, etc., those kinds of information are often only available for a limited set of organisms, may be available only in primary literature, and often require advanced training to understand and summarize. The dynamic features of the template and its legacy consistency make it possible for the template to be continually modified for new purposes.

The most important part of each section of the template is a block of searchable text, written by the account author. This text contains all the information presented in the section. Each section also has a list of controlled vocabulary keywords, unique to the section and may include data fields, where authors enter numerical data (e.g. mass, basal metabolic rate) or small items of text that address particular points (e.g. names of known predators, breeding season). The use of controlled vocabulary keywords avoids problems of synonymy and varying parts of speech (e.g. "hibernates" vs. "hibernation"), thus improving the

accuracy of data searches. Hierarchical keywords are employed as appropriate, for instance, a taxon coded as eating mollusks (molluscivore) is automatically tagged as a carnivore as well.

This template structure facilitates accurate data searches by allowing users to search in specific natural history fields, for particular natural history descriptors (keywords), for data ranges (e.g. birds with wingspan 25 to 50 cm), and for combinations of these. In addition, the template structure acts as a guide to contributors, ensuring that a broad suite of natural history data is considered in writing about an animal taxon.

Contributors provide standard-format reference entries to document all information used in creating accounts. These references are managed separately and directly linked to the relevant taxon account section. Contributors select from a list of reference types (journal article, book, web resource, etc.) and are then supplied with a reference template with the fields and format appropriate for that reference type. Once all references are entered contributors then select relevant references from a list appearing within each template section. This process facilitates uniformity and consistency of both reference format and citation style within text sections. Online references are available as hotlinks.

Table 1. Taxon account template sections and examples of controlled-vocabulary keywords and data fields. Keywords are defined in a glossary with synonyms provided. Keywords or data field labels may be defined differently depending on audience or taxon. Some keywords are hierarchical. A full list of current keywords and data fields can be found at http://animaldiversity.ummz.umich.edu/site/about/technology/index.html.

Template section	Sample keywords	Sample data fields
Diversity (higher taxa only)		
Geographic range	Nearctic, Neotropical, Antarctica, Indian Ocean, Mediterranean Sea, island endemic, cosmopolitan	
Habitat	temperate, tropical, polar, terrestrial, saltwater/marine, freshwater, desert, rainforest, pelagic, rivers and streams, urban, intertidal	elevation, depth
Systematic and Taxonomic History (higher taxa only)		synonyms, synapomorphies
Physical description	ectothermy/endothermy, type of symmetry, sexual dimorphism, polymorphism, poisonous/venomous	mass, length, basal metabolic rate
Development	neotenic/paedomorphic, metamorphosis, colonial growth, indeterminate growth	
Reproduction: mating systems	monogamous, polygamous, eusocial, cooperative breeding	
Reproduction: general behavior	semelparity/iteroparity, seasonal/year round breeding, gonochoric, hermaphroditic, parthenogenic, sexual/asexual, internal/external fertilization, oviparous/viviparous	breeding season, number offspring, time to hatching, age at maturity
Reproduction: parental investment	presence of parental care, types of parental investment by males and females, altricial/precocial, extended period of juvenile learning	
Lifespan/longevity		expected and maximum lifespan in captivity and in the wild
Behavior	degree of sociality, diurnal/nocturnal, migration, mode of locomotion or dominant way of living (scansorial, fossorial, natatorial), sessile/motile, hibernation/aestivation	territory and home range size
Communication/Perception	visual, chemical, tactile, acoustic, electrical, magnetic, heat, ultrasound, bioluminescence, mimicry, scent marking, pheromones	
Food habits	dominant food type (carnivore, herbivore, other) along with a more specific designation (molluscivore, scavenger, nectarivore, coprophage), list of all foods eaten, special food behaviors including caching and filter feeding	
Predation	mimicry, crypsis, aposematism	list of predators
Ecosystem roles	seed dispersal, pollination, biodegradation, soil aeration, creates habitat, keystone species	lists of mutualists, commensal species, and hosts
Economic Importance for humans: positive	pet trade, food, research, ecotourism, medicine, pollinates crops, controls pests	
Economic Importance for humans: negative	injures humans, crop pest, household pest, causes or carries domestic animal disease	
Conservation status	status on IUCN Redlist and U.S. E.S.A., CITES category	
Other comments	an unstructured section, including cultural significance, synonyms, fossil	

	history, scientific name etymology, genetics, etc.	

Diversity across the animal kingdom is accommodated by allowing the definition of multiple life stages, each of which then has its own set of descriptors, by allowing users to select units appropriate to their organisms, and by using *taxon filters* to control the visibility, content, and labeling language of natural history sections. For example, higher taxa where fertilization and gestation are internal across the entire taxon will have the appropriate keywords permanently filtered "on", whereas other sections or keywords may be filtered "off" for higher taxa to which they don't apply. ADW content specialists determine the definitions of these taxon filters.

While species-level taxon accounts are the core units addressed in the ADW, higher-level taxonomic coverage provides an important framework within which differences among species and groups are understood in an evolutionary context. Higher taxon accounts often function as sets of educational support materials for instructors in animal diversity or taxon-specific courses. For some taxa, where species are either poorly defined or species information is unavailable, higher level taxon accounts (genus or family levels, for example) may be the lowest level of resolution available. We designed a modified template for taxonomic information above the species level in order to maximize the breadth of natural history information recorded while taking into account the non-specific nature of data on animal taxa that, though related, may be quite diverse.

ADW template design was and is an iterative process and we have seen a progression from simpler to more complex template designs. However, design decisions are ultimately based on the need to balance the challenges of describing the biological complexity of all animals with the essential goal of building a useful resource for data mining and inquiry-learning and the limitations of animal natural history data availability and accessibility to most users.

An XML representation of our data template is available at http://animaldiversity.ummz.umich.edu/site/about/technology/index.html. We do not include reference or authorship management details, as they are essentially consistent with Dublin and Darwin Core metadata standards. We have also drafted an ontology using Protégé that captures most of the natural history concepts and their relationships. It lacks the taxon filters and help text found in the XML document. The ADW Ontology is archived for public use at the above URL and at Open Biological Ontologies (http://obo.sourceforge.net/).

3   RESULTS

Versions of the data template and Mousetrap have been used during three cycles of contribution beginning in January 2002. Eight instructors at eight institutions have worked with 292 contributors from both introductory and advanced undergraduate biology courses. Advanced students had no difficulty with the new data template, which had been far more complex than our previous templates. Introductory students, however, found its complexity daunting and need more help learning to use it.

An example illustrates the value of our loose coupling structure. In January 2004 we decided to alter our data template shortly before a round of student contributions. We decided to expand our keyword coverage of parental investment in offspring. We renamed the subsection (from Parental Care to Parental Investment), and replaced a pair of keywords with a longer list of hierarchical keywords, some specifying protection or provisioning of offspring by males or females at various stages in offspring development (e.g. pre/post hatching, fledging) and a few for particular complex behaviors (e.g. "inherits maternal/paternal territory"). We also revised section instructions to address the new terms. These changes were made successfully about a week before students began working on taxon accounts. The system retains the old keywords for legacy accounts, makes the new keywords available if these old accounts are revised, and continues to effectively search and display these accounts.

Currently six staff editors are using Mousetrap to edit taxon accounts, install multimedia, and create and maintain general site pages such as FAQs and special topics pages. Not including the database of over 196,000 animal names, we are managing information on over 4700 different animals. This includes 8722 media files (photographs, illustrations, and sounds) and 2363 detailed taxon accounts. Of these taxon accounts, about 1800 were created under prior back-end conditions (a traditional relational database) while the rest were created under the new system. Almost 600 accounts are currently in progress. Editors report no problems using the same system to modify both legacy and new accounts. In addition, legacy content is easily identified by the system and can be presented to editors as candidates for revision.

Two public sites generated by this technology have been successfully deployed. BioKIDS was first launched under earlier prototypes of this system and has been used effectively by about 2000 5th and 6th grade students involved in the BioKIDS curriculum program (Songer, 2004). It includes a 165-animal subset of the total number of accounts, uses alternate labeling for sections and keywords, and employs a simplified navigation structure that skips some taxonomic levels. Animal Diversity Web, aimed primarily at a worldwide adult audience, began using this new infrastructure in January 2004 to present our full complement of multimedia and text accounts. In addition, we now display on a classification tab a way to explore all

196,078 biological names in our taxonomic database. Animal Diversity Web receives a large volume of traffic; in the month of April 2004 the new ADW served over 220,000 pages daily to 10,000 unique IP addresses.

Several issues remain to be tackled in future work. First, we have not formally studied the usability of the templates or of the public sites. In particular, it is a challenge to present such a complex template to contributors who are unlikely to provide content for more than one or a few taxa. Second, certain management functions, such as changing many node attributes at once, are not supported well in Zope. In addition, we ran into a scaling wall when using Plone for the public sites. Third, we are actively researching better ways to store and manage our nodes. Zope is inherently hierarchical, and our data is increasingly "placeless" --- organized more by its metadata and identification than its filesystem location.

4   DISCUSSION

This approach is highly scalable in a number of dimensions. First, storage is efficient because only relevant concepts are stored with each node. This is similar to the Entity-Attribute-Value approach taken by SenseLab (Marenco et al., 2003). Second, legacy data need never be discarded because of a data template redesign. Third, managing data for several audiences does not typically require duplication of data, merely different stylesheets. Fourth, with the right tools, management of data templates and taxon filters can be achieved by staff biologists rather than by programmers. Fifth, different components can be substituted as new, more scalable technology matures. For example, we could replace the current TaxonDB component with a taxonomic name service such as uBio (MBL/WHOI, 2005).

Our approach has specific advantages for our formal and informal education audiences. We provide highly structured data suitable for inquiry learning, supporting searches for patterns and testing of hypotheses. Future work at the undergraduate level will examine the success of this notion. We can restructure the displays to support younger students who require certain amounts of structure or scientific terminology and who are being scaffolded as they develop scientific skills. At the same time, we can display our content to appear more readable to audiences that do not require high structure or controlled vocabulary keywords. This system ensures that improvements to our content simultaneously reach all of our audiences.

A significant audience we have not yet formally served is the scientific community (but see (Norris, Zhou, Zhou, Yang, Kirkpatrick & Honeycutt., 2004) and seventeen Enhanced Perspectives articles in the online version of *Science*). We know through site feedback that scientists in developing countries often use our site due to lack of access to good libraries. However, we are approaching the point where comparative biologists all over the world could harvest our raw, structured data and use it in comparative studies. For example, a molecular biologist studying genetic basis of trait that varies across Animalia could, with the right data mining application, search for explanatory patterns in our database of reproductive and life history characteristics. However, it should be noted that although we provide two levels of editorial oversight, we do not formally check all facts. We do provide a "Report error" link so that outside users can bring errors to our attention, and the system allows us to quickly publish corrections.

In addition to extending Dublin and Darwin Core metadata standards, ADW's hierarchical data template and associated glossary of terms can be represented as a simple ontology. Concepts are organized into sections (classes) and subsections (subclasses) within which are specific keywords and data fields (slots). Relationships are typically IS-A or HAS-A in nature. For example, "Simultaneous hermaphrodism" is a kind of "hermaphrodism" is a "general reproductive characteristic." Our taxon filters are analogous to facets limiting the allowable values for particular instances. It should be straightforward to implement a new web stylesheet that takes advantage of the OWL version of our data template to generate web pages that include our semantic markup. These new pages will then be available on the semantic web (Hendler, 2003) where intelligent agents can assist users with varying levels of content expertise to more effectively retrieve information.

With or without the semantic web, our approach begins to make it possible to always provide the most current contents of our database to the public and to web agents and to federation efforts. Our data template definitions extend Dublin and Darwin Core and should be able to interact with distributed querying protocols for biological collections such as DiGIR and BioCASE. Much work remains to be done to realize this potential, but we believe that our system provides an infrastructure that encourages rather than limits long-term growth and access by many audiences.

5   ACKNOWLEDGEMENTS

6   REFERENCES

AmphibiaWeb (2004) Homepage of AmphibiaWeb. Available from: http://amphibiaweb.org/.

Bisby, F. A. (2000) The Quiet Revolution: Biodiversity Informatics and the Internet. *Science 289*, 2309-2312.

Cherry, G., Washington, W., Fournier, J., & Shuyler, K. (2003) Learners on the back end: students contributing to web-based information systems. *Proceedings of CHI '03, Conference on Human Factors in Computing Systems* . Ft. Lauderdale, FL.

Conservation Management Institute (2001)Homepage of Fish and Wildlife Information Exchange. Available from: http://fwie.fw.vt.edu/WWW/spp.htm.

Froese, R. & Pauly, D. (2004) Homepage of FishBase. Available from: http://www.fishbase.org.

GBIF (2005) Global Biodiversity Information Facility Data Portal. Retrieved April 12, 2005 from the World Wide Web: http://www.gbif.net/.

Hendler, J. (2003) Science and the Semantic Web. *Science* 199, 520-521.

Marenco, L., Tosches, N., Crasto, C. J., Shepherd, G. M., Miller, P. L., & Nadkarni, P. M. (2003) Achieving evolvable web-database bioscience applications using the EAV/CR framework. *JAMIA 10*, 444-453.

MBL/WHOI Library (2005) Homepage of uBio: Universal Biological Indexer and Organizer. Available from the World Wide Web: http://www.ubio.org/.

Norris, R. W., Zhou, K., Zhou, C., Yang, G., Kirkpatrick, C. W., & Honeycutt, R. L. (2004) The phylogenetic position of the zokors (Myospalacinae) and comments on the families of muroids (Rodentia). *Molecular Phylogenetics and Evolution* 31, 972-978.

OBIS (2004) Homepage of Ocean Biogeographic Information System. Available from: http://www.iobis.org.

Parr, C. S., Lee, B., Campbell, D., & Bederson, B. (2004) Tree visualizations for taxonomies and phylogenies. *Bioinformatics 20*, 2997-3004.

Songer, N.B. (2004) Persistence of inquiry: evidence of complex reasoning among inner city middle school students. *American Educational Research Association (AERA) annual meeting*. San Diego, CA.

Weinberger, D. (2002) *Small Pieces Loosely Joined: A Unified Theory of the Web.* New York: Perseus Books Group.