Evolution of natural history information in the 21st century – developing an integrated framework for biological and geographical data

Deborah A. Reusser
U.S. Geological Survey, dreusser@usgs.gov

Henry Lee II
U.S. EPA

Follow this and additional works at: http://digitalcommons.unl.edu/usepapapers

Part of the Civil and Environmental Engineering Commons

http://digitalcommons.unl.edu/usepapapers/90
Evolution of natural history information in the 21st century – developing an integrated framework for biological and geographical data

Deborah A. Reusser\textsuperscript{1}\textsuperscript{*} and Henry Lee II\textsuperscript{2}

ABSTRACT

Threats to marine and estuarine species operate over many spatial scales, from nutrient enrichment at the watershed/estuarine scale to invasive species and climate change at regional and global scales. To help address research questions across these scales, we provide here a standardized framework for a biogeographical information system containing queriable biological data that allows extraction of information on multiple species, across a variety of spatial scales based on species distributions, natural history attributes and habitat requirements. As scientists shift from research on localized impacts on individual species to regional and global scale threats, macroecological approaches of studying multiple species over broad geographical areas are becoming increasingly important. The standardized framework described here for capturing and integrating biological and geographical data is a critical first step towards addressing these macroecological questions and we urge organizations capturing biogeoinformatics data to consider adopting this framework.

Keywords
Biogeography, biogeoinformatics, biological information systems, chitons, classification schemas, ecoinformatics, hierarchical schemas, macroecology, marine ecosystems, natural history.

WHY A COMMON SCHEMA IS NEEDED

The extensive biological and ecological knowledge collected on flora and fauna of the world over the last two to three hundred years primarily resides in a wealth of books, reports and journal articles. Two recent examples of attempts to synthesize this diffuse information are Fishes of Alaska (Mecklenburg \textit{et al.}, 2002) and The Light and Smith Manual – Intertidal Invertebrates from Central California to Oregon (Carlton, 2007). Each of these contains over a thousand pages of information including habitat, distributions and life history characteristics. Today, advanced search engines available on the World Wide Web (WWW) also make it possible to discover a vast amount of text-based literature on life histories and habitat requirements for species one at a time. In addition, some of these data have been captured in electronic database format and can be accessed via the WWW (e.g. FishBase, http://www.fishbase.org/search.php and GBIF, http://data.gbif.org/). However, with a few exceptions (e.g. FishBase; Gulf of California Invertebrate Database; http://www.desertmuseum.org/center/seacoastcortez/searchdb.php), the existing formats do not provide a simple way to relate biotic attributes of marine and estuarine species to their biogeographical distributions and/or environmental preferences. Nonetheless, there is a growing need to synthesize information on multiple species from multiple locations to address regional and global issues, such as climate change and invasive species (Lee \textit{et al.}, 2008). Unfortunately, because of the time required to extract information for multiple species and/or locations from the plethora of text-based sources, existing data are often underutilized (Kerr \textit{et al.}, 2007). Another challenge in synthesizing these data is the idiosyncratic nature of much of the natural history data. For example, depth ranges of marine species are given in feet, fathoms and metres, while the term ‘littoral’ can mean either intertidal or the intertidal and shallow subtidal (Open University, 2000).

To better utilize the existing knowledge, integrated biological information systems are needed that allow the organization and extraction of both quantitative and qualitative natural history and distributional data. One approach to help in organizing such diverse types of information is to develop classification systems. There have been a few attempts to develop classification schemas for the natural history of marine/estuarine organisms. For example, Bush \textit{et al.} (2007)
and Bambach et al. (2007) developed a schema for ‘ecospace utilization’ in marine fossil assemblages based on life position relative to the sediment-water interface (tiering), motility and feeding mechanism. Fauchald & Jumars (1979) developed a schema for feeding guilds of polychaete worms based on three criteria: what the species feeds on, motility and feeding morphology. Both of these systems are multi-dimensional in that they combine several attributes to define each class or guild. Although these individual classification systems are useful in capturing specific types of natural history attributes, integrating multiple classification schemas within a single framework can substantially increase the utility of the information and provide insights not possible when evaluating a single type of biotic attribute.

**PROPOSED SCHEMA**

To address this need, we designed a biological information framework for the synthesis of natural history, environmental and geographical information for marine and estuarine species. Key aspects of this framework include: (1) consistent terminology for natural history requirements; (2) translation of specific numerical values for habitat/physiological requirements into classes; (3) either multi-dimensional or hierarchical classification schemas for natural history and environmental attributes and geographical distributions, depending on which schema best fits the data; and (4) an integration of the various biotic attributes to allow database queries on single or multiple species across different spatial scales. This integrated biological information framework was developed for marine/estuarine organisms in the ‘Pacific Coast Ecosystem Information System’ (PCEIS). However, the concepts presented here are applicable to biological information systems in general, including freshwater and terrestrial biota.

**Classification schemas for environmental and natural history attributes**

Quantifying patterns of similarity across multiple species is difficult because much of the existing natural history data are qualitative and often anecdotal. There is often a lack of accepted terminology across different taxa and/or habitat types. The use of standardized classes is a practical approach to capturing and organizing such natural history information and simplifies data queries to analyse patterns for multiple species and/or locations. Additionally, translating natural history attributes into standardized classes imposes a rigorous terminology for attributes that are challenging to quantify, such as wave energy. Another use of environmental or habitat classes is to fill in gaps when there are missing quantitative values. An example includes estimating temperature classes for a species from an analysis of the spatial locations where the species is found. The greatest limitation of classes is that they require simplification at the potential loss of some ecological nuances. However, the importance of the loss of species-specific or quantitative environmental information will depend upon the nature of the question as well as the inherent resolution of the classes.

The simplest type of classification structure is when an attribute can be divided into two or more unique classes without the need for subclasses. Examples of attributes that can be captured in a non-hierarchical class structure include temperature regimes, wave energy and population status of non-indigenous species (e.g. established, not established and stocked). The classification schemas adopted for temperature and wave/current energy are described below.

**Temperature classification schema**

Temperature classes used in PCEIS are based on annual and seasonal temperature ranges for geographical regions adapted from Hall (1964). These classes (Table 1) are based on mean monthly temperatures in the ecoregions occupied by the species and do not take into account local or habitat differences in temperature (e.g. upper intertidal versus subtidal). A more up-to-date analysis of coastal temperature regimes may change the criteria for the classes as well as how the classes are calculated (e.g. using site-specific temperature ranges instead of those based on regions). However, even a coarse classification such as this can be useful for certain questions. For example, such coarse temperature classes have been used in an environmental-matching risk assessment to identify marine/estuarine species in one location that could potentially invade another location based on similarity in temperature conditions (Gollasch, 2006). Assigning freshwater fishes to cold-water or warm-water thermal guilds was sufficient to identify habitat availability under climate induced changes in lakes and streams (Eaton & Scheller, 1996; Stefan et al., 2001).

**Wave energy classification schema**

Wave energy is an important habitat attribute for intertidal communities (e.g. Howes et al., 1999). However, actual wave
energy is difficult to quantify so classes have been substituted in many analyses. Lindegarth & Gamfeldt (2005) compared categorical wave exposure classes to continuous values and concluded that categorical values are meaningful substitutions. Accordingly, we adopted wave energy classes (Table 2) that are derived from maximum fetch and modified effective fetch as defined by Howes et al. (1999), who provide procedures for calculating wave heights from wind speed and direction. In addition, each classification contains descriptive text so that a wave energy class can be identified without having to calculate fetch.

Hierarchical schemas for environmental and natural history attributes

Hierarchical classification structures allow the capture of biotic attributes at increasing levels of detail, allowing information collected at different levels of resolution to be integrated into a single system. For example FishBase provides three broad classes for the salinity tolerances of fish (freshwater, brackish or marine). Other sources however, provide salinity information at a higher level of detail, such as the National Exotic Marine and Estuarine Species Information System (NEMESIS), which provides quantitative salinity tolerances for non-indigenous species in Chesapeake Bay (Fofonoff et al., 2003). The hierarchical classification schema for salinity discussed below provides the ability to capture both levels of information. When analysing data, an advantage of the hierarchical structure is that it provides the ability to extract information synthesized at an appropriate level of detail for a specific question. Compared to multi-class systems, hierarchical schemas also provide greater flexibility, particularly in allowing the addition of new higher resolution subclasses.

To evaluate the practicality of natural history hierarchies, we developed a multi-level classification schema for a range of life history characteristics and physiological tolerances. Some attributes were readily adapted to a hierarchical schema, such as taxonomy, salinity and depth. Other attributes are not as obviously hierarchical, such as reproductive and feeding modes. Nonetheless, we propose schemas for these attributes, recognizing that alternative topologies are possible. To the extent possible, this framework either utilizes existing classifications as one of the levels or a crosswalk is possible with the existing schema. For example, Faucald & Jumars (1979) had two classes of deposit feeders: surface-deposit feeding species and burrowers (= subsurface-deposit feeders). In our system, we added a higher level (deposit feeders) to capture the general feeding mode when the specifics are not known. With other attributes, we added finer resolution classes to capture more detailed information, as in the suspension feeding mode where the specifics are not known. With other attributes, we added a higher level (deposit feeders) to capture the general feeding mode when the specifics are not known. Without other attributes, we added lower level classes to define whether species are obligate or facultative suspension feeders. A portion of these hierarchical classification schemas are highlighted below as examples.

Taxonomic hierarchical schema

The grandfather of all biological hierarchical classification schemas is the Linnaean Taxonomic Hierarchy (Linnaeus, 1758). The current version of Linnaeus’s scientific classification schema has undergone modifications through the centuries, but is still the recognized standard. Web-based versions of this standard taxonomic structure are available from several sources, such as the Integrated Taxonomic Information System (ITIS; http://www.itis.gov/) and the World Register of Marine Species (WoRMS; http://www.marinespecies.org/).

Biological information systems should capture the full taxonomic hierarchy for species, as do ITIS and WoRMS, to allow analyses of community assemblages using metrics based on taxonomic structure such as the taxonomic distinctness measure of biodiversity (Clarke & Warwick, 1999) and analyses within and among taxonomic groups (Koleff et al., 2003). However, the constant revision of species names and their taxonomies is a major challenge. The literature on chitons in the Northeast Pacific spans more than a century during which time species names changed, merged or split into a compendium of new names. Almost half of the names for the California chitons listed in Hemphill (1890) have changed and Eernisse (1986) gives a detailed example of the tortuous changes in the names of three species of Lepidochitona from 1892 to 1983 (Eernisse, 1986). One solution is for taxonomic databases to capture and provide search utilities for synonyms and common misspellings as well as the current name, as WoRMS does. This provides a mechanism for conversion of historical names to current naming conventions for the standardization of species occurrences through time and

Table 2 Wave energy classes for marine/estuarine species based on methods for estimating wave heights using wind speed and direction following Howes et al. (1999). Note that in most cases it is possible to assign a wave energy class from the qualitative descriptions so that the calculation of the actual fetch in not necessary.

<table>
<thead>
<tr>
<th>Wave energy classes</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>‘High ambient wave conditions usually prevail within this exposure category, which is typical of open-Pacific type conditions’. Max. fetch distances &gt; 500 km</td>
</tr>
<tr>
<td>Semi-exposed</td>
<td>‘Swells, generated in areas distant from the shore unit create relatively high wave conditions. During storms, extremely large waves create high wave exposures’. Max. fetch distance between 50 and 500 km</td>
</tr>
<tr>
<td>Semi-protected</td>
<td>‘Waves are low most of the time except during high winds’. Fetch in range of 10–50 km</td>
</tr>
<tr>
<td>Protected</td>
<td>‘Usually areas of provisional anchorages and low wave exposure except in extreme winds’. Fetch &lt; 10 km</td>
</tr>
<tr>
<td>Very protected</td>
<td>‘Usually the location of all-weather anchorages, marinas and harbours’. Max. fetch &lt; 1 km</td>
</tr>
</tbody>
</table>
Salinity hierarchical schema

Salinity is considered a ‘master variable’ in controlling the distribution of marine/estuarine species (Hodgkin, 1987) and provides another good example of a hierarchical classification schema. The Venice system was defined at the ‘Symposium on the Classification of Brackish Waters’ (Anonymous, 1958) and consists of freshwater, oligohaline, mesohaline, polyhaline, euhaline and hypersaline. This symposium also defined a higher resolution schema that divides the oligohaline, mesohaline and polyhaline classes into two subclasses (Fig. 1). Because the subclasses were based on the lower salinity Baltic, the euhaline class was not subdivided. Accordingly, we divided the euhaline class into two subclasses, using 36 practical salinity units (psu) as the breakpoint. To allow more flexibility, we then integrated the different salinity classifications into a three-level hierarchical classification schema, from the coarse scale (freshwater, brackish and marine) to the six-level Venice system, and finally to the ten-level ‘modified Venice system’ (Fig. 1).

To assess the extent of information lost by going to a classification system versus using numerical salinity values, we evaluated the relationship between salinity and species richness in soft-bottom benthos. For this analysis, we used data from the U.S. Environmental Protection Agency’s (EPA) Benthic indicator database which integrates published data from regional surveys of the near-coastal waters of California, Oregon and Washington from 1997 to 2006 (see Nelson et al., 2005 for description of sampling methodology). Only samples classified as undisturbed were used in this analysis to reduce the potential for anthropogenic influences on species richness ($n = 531$). The number of species per sample (0.1 m$^2$ grab) was regressed on continuous salinity (psu) measurements taken at the time of sampling (Fig. 2). The numerical salinity values for each of these samples were then converted to salinity classes based on the modified Venice system defined in Fig. 1. The analysis was repeated using an ANOVA because the

![Figure 1](http://www.fishbase.org/search.php) Three level hierarchical schema for salinity. The first 3 classes (Freshwater, Brackish and Marine) of Level 1 are used in FishBase (http://www.fishbase.org/search.php) while Level 2 is the Venice system developed by the International Union for Biological Sciences (Anonymous, 1958). Level 3 is the modified Venice system based on the Baltic system (Anonymous, 1958) enhanced in this schema to include beta- and alpha-euhaline levels for marine waters. Salinity is given as practical salinity units (psu). Numbers in the circles in the lower right hand side of the class boxes refer to the salinity classes used in the analysis shown in Fig. 3.
Figure 2  Number of species per benthic sample (0.1 m$^2$ grab) relative to the salinity value at the time of sampling. Data are from the U.S. EPA’s benthic indicator database of regional surveys of soft-bottom communities in California, Oregon and Washington over a 10-year period from 1997 to 2006. Predictive equation based on least squares regression model ($R^2 = 0.21$; $n = 526$; d.f. = 1, 524; $t$-value = 11.78; $P < 0.001$). Salinity is given as practical salinity units (psu).

Figure 3  Salinity values from benthic samples (0.1 m$^2$ grabs) in Fig. 2 converted to modified Venice salinity classes. Modified Venice class numbers relate to the salinity classes shown in Fig. 1, where $I$ = 0 to $< 0.5$ practical salinity units (psu); $2$ = 0.5 to $< 3$ psu; $3$ = 3 to $< 5$ psu; $4$ = 5 to $< 10$ psu; $5$ = 10 to $< 18$ psu; $6$ = 18 to $< 25$ psu; $7$ = 25 to $< 30$ psu; $8$ = 30 to $< 36$ psu. Values and horizontal bars indicate mean number of species within each salinity class (ANOVA; $R^2 = 0.22$; $n = 531$; d.f. = 7, 523; $F = 23.71$; $P < 0.001$). The $R^2$ value of 0.21 from Fig. 2 was compared with the $R^2$ value of 0.22 here to evaluate the loss of predictive power for species richness if salinity classes were used rather than actual salinity values.

An integrated framework for biological and geographical data

**Number of species by salinity psu**

![Number of species by salinity psu](image)

**Number of species by modified Venice**

![Number of species by modified Venice](image)

**Depth hierarchical schema**

Depth is another key factor affecting the distribution of marine and estuarine species (Madden et al., 2005). Historically, different bathymetric classification systems have been used for benthic versus pelagic species, and we maintain this dichotomy as the first level of the depth hierarchy (Fig. 4). The pelagic zone is divided into five generally recognized classes (epipelagic to hadopelagic). To better capture the distribution of near-surface pelagic flora and fauna, we adopted a three level split of the epipelagic class into surface, shallow and deep zones. For benthic organisms, including both epibenthic and infaunal species, the oceanic environment is commonly divided into five main depth classes, from the intertidal to the hadal. To expand the range to include semi-terrestrial environments, we incorporated the supralittoral zone and coastal fringe (e.g. dunes). Other than splitting the intertidal into upper, middle and lower zones, there is no generally agreed upon subdivision of the neritic, bathyal and abyssal zones. The 30-m threshold for shallow versus deep subtidal habitats in the neritic zone is based on Madden et al. (2005), although some other classification systems use 20 m as the threshold (e.g. Biological Traits Information Catalogue, http://www.marlin.ac.uk/biotic/).

With the potential effects of ocean acidification on deeper water calcareous species (Kleypas et al., 2006), higher resolution depth subclasses for deep oceanic waters have become important to help identify species at greatest risk. To capture this higher resolution, the subdivisions of the bathyal and abyssal zones proposed by Berggren & Miller (1989) for benthic foraminifera were incorporated with one modification. Berggren & Miller (1989) divided the bathyal zone into three subclasses but the abyssal into only two. The limitation with this division of the abyssal zone is that the split does not adequately resolve the depth of the calcite compensation depth...
(CCD, the depth below which the rate of dissolution of calcium carbonate is equal to the rate of its formation). The depth of the CCD varies by location but ranges from about 3600 to 4800 m in the North Pacific (Berger et al., 1976); thus, we created an additional subdivision at 4000 m (Fig. 4). The 4000-m break is considered preliminary and will be modified as additional information becomes available on the distribution of calcareous and non-calcareous deep-sea organisms as well as the predicted effects of climate change on the depth of the CCD. Because of the possibility of changes in thresholds, we suggest that biological information systems should capture both quantitative bathymetric depth ranges as well as depth classes simplifying reclassification if class thresholds change.

Reproductive mode hierarchical schema

Some characteristics of a species’ reproductive mode are inherently categorical, such as whether a species has an asexual or sexual reproductive strategy. While this initial split is straightforward, reproductive strategies can be further subdivided to capture more details such as whether a sexual species is monocious or dioecious or whether an asexual species is one of five different classes (such as parthenogenetic). These kinds of natural history attributes are ‘messy’ and different hierarchical structures are possible. Nonetheless, we suggest that the schema presented in Fig. 5 imposes a useful structure in organizing reproductive modes. One example of how this information can be used is in the identification of species that have characteristics of successful invaders, such as parthenogenetic reproduction as in the New Zealand mud snail, *Potamopyrgus antipodarum* (Kerans et al., 2005).

Geographical hierarchical schemas

Location matters; spatial relationships are fundamental to the evolving organization of life on Earth (Gaile & Willmott,
2003), with ecological processes operating across many different spatial scales. To address this phenomenon, we developed hierarchical spatial groupings to allow us to link macroecological patterns and processes at a range of scales to address a variety of ecological questions and management issues.

**Biogeographical hierarchical schema**

Biogeographical frameworks attempt to define regions with similar faunas and floras to those in other geographical regions (Briggs, 1995) and a number of biogeographical schemas have been developed. The boundaries defined by some of these schemas are very large ocean areas designed for fisheries management and conservation purposes. For example, the Large Marine Ecosystems (LMEs) schema developed by the National Oceanic and Atmospheric Administration (NOAA; http://www.lme.noaa.gov/) consists of 64 marine areas with delineations based on a combination of political and biogeographical boundaries. In comparison, the International Union for Conservation of Nature (IUCN) schema consists of more than four times as many delineations, with 264 marine ecoregions (Kelleher et al., 1995).

These different biogeographical systems are usually presented as static alternatives. However, the recently developed Marine Ecoregions of the World (MEOW) (Spalding et al., 2007) uses a three-tiered schema (realm, province and ecoregion). Thus, using the MEOW schema, species distributions can be entered at the ecoregion scale if available and aggregated to higher levels or entered at the higher level if detailed location information is not known. The three MEOW levels are closely related to the boundaries identified in other schemas, with the province level approximating the LMEs and the ecoregions approximating the IUCN ecoregions, so it is not necessary to treat the different systems as distinct alternatives.

A limitation of the MEOW schema is that it does not split the Pacific and Atlantic into east/west components. These wide

---

**Figure 5** Hierarchical schema for capturing reproduction information on marine and estuarine species. Spermcast refers to the discharge of male gametes into the water column that fertilize eggs retained by the female (Pemberton et al., 2003) while a broadcast spawner refers to species that discharge both sperm and eggs into the water column.
stretches of ocean act as effective barriers for near-coastal species (Cox & Moore, 2005), separating species on opposite sides of oceans. Additionally, most invasions in near-coastal waters in the Northern Hemisphere have occurred across oceans rather than from lower or higher latitudes (Ruiz et al., 2000) and the current version of MEOW ecoregions does not facilitate capturing such across-ocean invasions. Therefore, as shown in Fig. 6, a new ‘region’ level was added to the MEOW schema between the realm and province level to provide east–west breaks in the Pacific and Atlantic Ocean realms. Adding this region level simplifies addressing questions such as the relative importance of eastern versus western sides of oceans as donor regions for non-indigenous species (Chapman, 2000).

Land/sea connection hierarchical schema

Biogeographical schemas are useful for addressing questions related to evolutionary history, patterns of invasion, and potential effects of climate change. However, a finer resolution schema is needed to address questions related to the interactions between watersheds and near-coastal water bodies, such as nutrient runoff (Bricker et al., 2007), or local distributions of species. To address these issues, we developed a five-tier land/sea hierarchical classification schema, which is linked to the biogeographical hierarchical schema (Fig. 6). Fundamental units for the land/sea hierarchy are estuaries, coastal segments, and their associated watersheds. In the United States, estuaries can be defined by the presence of National Wetland Inventory (NWI; http://www.fws.gov/wetlands/) estuarine and tidal riverine polygons (see Lee & Brown, 2009). Coastal segments are stretches of the coast that drain directly into the ocean or water body system rather than an estuary, and are equivalent to NOAA’s coastal drainage areas (CDA; http://coastalgeospatial.noaa.gov/). The defining characteristic of both estuaries and coastal segments is that a single watershed can be defined for them. For coastal creeks, the watershed may be very small (<1 km²) while the watersheds draining into the Columbia River and San Francisco Estuary are greater than 120,000 km². Regardless of size, the key aspect is that there is a one-to-one relationship between the estuary or coastal segment and the watershed draining into it. At a finer spatial scale, the hierarchy includes the embayments and tributaries that make up a single estuary.

In addition to estuaries and coastal segments with a one-to-one relationship with their associated watershed, there are compound systems that are defined as water body systems that receive runoff from multiple watersheds (e.g. Puget Sound and Monterey Bay) (Fig. 6). In compound systems the water body may consist of basins and sub-basins, both of which may be composed of one or more estuaries and/or coastal segments. For example, the entire Puget Sound is considered a single water body system which is composed of basins such as Whidbey Basin, which in turn are composed of sub-basins such as Possession Sound and Skagit Bay, which in turn have

![Diagram](image_url)

**Figure 6** Integration of the hierarchical schemas for marine biogeographical areas and land/sea connections. The biogeographical schema is based on the Marine Ecosystems of the World (MEOW; Spalding et al., 2007) with the modification that a Region level is added between the Realm and Province levels to differentiate eastern and western sides of the Atlantic and Pacific. Estuaries and coastal segments are characterized by a one-to-one relationship with a watershed. Compound systems consist of several waterbodies and watersheds, which in turn may consist of one or more basins and sub-basins.
individual estuaries and coastal segments embedded within them.

This connectivity framework provides the ability to analyze differences in estuarine and near-shore communities based on watershed characteristics and land-use practices (e.g. Lee & Brown, 2009). In addition, the aggregation levels for compound systems provide a framework to conduct analyses within segments or over an entire system, allowing researchers to match the scale of biological data to the scale of available environmental data.

**POPOPULATING THE INFORMATION SYSTEM**

Developing an integrated framework is only part of the task; the other is populating the information system. Graphical interfaces, compared to a spreadsheet format, can increase the efficiency of capturing information from text-based sources. For example, data entry of species habitat regimes based on the classification structure seen in Table 3 can be implemented in a graphical user interface (Fig. 7). Regardless of how efficiently information can be entered for a single species, a major

---

**Table 3** Habitat regimes. Regimes are the broad physical and environmental divisions based on a combination of salinity, geomorphology and depth (Madden et al., 2005).

<table>
<thead>
<tr>
<th>Regime class</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terrestrial</td>
<td>Land areas not directly impinging upon aquatic ecosystems</td>
</tr>
<tr>
<td>Lakes and ponds (lentic)</td>
<td>Bodies of standing fresh water, including freshwater wetlands</td>
</tr>
<tr>
<td>Rivers, streams and creeks (lotic)</td>
<td>Flowing bodies of freshwater, including riparian zones</td>
</tr>
<tr>
<td>Estuaries and lagoons</td>
<td>Estuaries are semi-enclosed coastal water bodies with one or more rivers or streams flowing into them and with a connection to the ocean. Salinities in estuaries are normally below that of the bordering ocean water. Lagoons are shallow coastal water bodies separated from the ocean by a barrier island or by shallow or exposed sandbanks or coral reefs. Depending upon freshwater inputs and connection to the ocean, salinity in lagoons can range from essentially fresh to hypersaline</td>
</tr>
<tr>
<td>Coastal fringe</td>
<td>Area between terrestrial and nearshore or estuarine ecosystems with primarily terrestrial characteristics but strongly affected by bordering aquatic ecosystem (e.g. sand dunes, estuarine shrub/scrub wetland, estuarine forest wetland)</td>
</tr>
<tr>
<td>Coastal bay</td>
<td>An area of water mostly surrounded by land on the open coast, creating calmer waters than the open sea</td>
</tr>
<tr>
<td>Nearshore</td>
<td>0–30 m. The outer coast from the intertidal to 30 m bathymetric isopleths, including both the benthos and overlying water</td>
</tr>
<tr>
<td>Shelf</td>
<td>30–200 m bathymetric isopleths. The benthos and overlying water that borders the continent and extends out to where there is an increased slope of the sea floor, approximately 200 m depth</td>
</tr>
<tr>
<td>Oceanic</td>
<td>&gt;200 m bathymetric isopleths. Includes the benthos and overlying water of the continental slope and ocean floor</td>
</tr>
</tbody>
</table>

---

**Figure 7** Graphical interface for data entry and display of regime data for a single species (*Leptochiton alveolus*) in the Pacific Coast Ecosystem Information System (PCEIS). A check in a box with a red border indicates a regime the species has been observed in, while a check in a box with a green border indicates the preferred habitat for the species. Use of observed and preferred options for several attributes allows a separation of the ‘normal’ range for a species versus the total range, which may include individuals displaced out of their normal habitat.
challenge remains – the Census of Marine Life estimates that there are about 230,000 described marine species (http://www.coml.org/about). One approach that can often be used is to capture information for multiple species at higher taxonomic levels. For example, feeding type can often be captured at the genus or family level. As mentioned, Fauchald & Jumars (1979) assigned polychaete feeding types by family and genus, and the biogeographical pattern of carnivorous marine snails was determined by assigning feeding type by family (Valentine et al., 2002). Other natural history attributes that can often be captured at higher taxonomic levels include general habitat type (e.g. pelagic versus benthic), reproductive mode (e.g. sexual versus asexual), and certain aspects of development type (e.g. pelagic larvae versus direct development).

Drawing on the precept that 'geography is biology', another practical approach for capturing information on multiple species is to utilize automated routines to extract habitat ranges from databases of field surveys. From the benthic database used for the salinity analysis, we automated the extraction of environmental ranges for salinity, depth and percentage silt/clays for over 3000 benthic taxa from more than 4000 samples. Development of flexible tools to allow such bulk data collection and entry from different types of monitoring studies and downloads from biotic web sites will allow the generation of environmental ranges for thousands of species.

**PROOF OF CONCEPT**

To illustrate the utility of the hierarchical framework to address regional environmental concerns, life history and geographical distribution information on chitons was collected from the monographs by Kaas & Van Belle (1985) and approximately 30 additional publications (e.g. Stebbins & Eernisse, 2009). Once the chiton information was collated in PCEIS, the data could be extracted using queries that tallied the number of chitons by a variety of attributes. Figure 8 illustrates the number of chiton species in the Northeast Pacific and Arctic by depth class. Chitons that occur only in the intertidal or shallow subtidal are presumably at the greatest risk from both global warming and sea level rise, while species extending into the deep subtidal to abyssal depths are at a lower risk. Conversely, because chitons have calcareous shells, those species that only occur in the deeper depth zones are at greater potential risk to the shallowing of the CCD in response to ocean acidification.

A biogeographical analysis is illustrated using the distribution of the total number of chiton species and number of endemic chitons in the Northeast Pacific and Arctic by ecoregion (Figs 9 & 10). While this distributional pattern raises a number of interesting biogeographical questions, the analysis was conducted to help identify species at risk to climate change (H. Lee II, D.A. Reusser & E.V. Saarinen, unpublished data). Because of their limited range, the 42 endemic species are assumed to be more vulnerable to climate changes than species with broader ranges (Rabinowitz, 1981), with the Southern California and Cortezian (=Gulf of California) ecoregions having the greatest number of endemics at risk. The species in the Chukchi and Bering seas are also assumed to be at greater risk because of the greater projected temperature increases in the Arctic (Corell, 2006), the limited potential for Arctic species to migrate northward as climate changes, and because the CCD occurs at shallower depths in colder water (Andersson et al., 2008).

One of the significant contributions of the proposed framework is that it facilitates matrix overlays between and among the natural history and environmental ranges for individual or multiple species across different geographical scales. An example of a multi-factor analysis is the crosswalk between the biogeographical ranges of chiton species in the Northeast Pacific and Arctic by their bathymetric distributions (Table 4). Combining these two attributes identifies the shallow-water chitons that are presumably at the greatest risk to climate change due to the combination of restricted distributions and exposure to sea level rise and global warming. In addition, the multi-factor analysis identifies the

![Figure 8](http://www.coml.org/about) The number of chiton species \(n = 167\) in the Northeast Pacific and Arctic (Gulf of California to Chukchi Sea) by depth range using the depth classification schema presented in Fig. 1. For this analysis the level 2 classifications were used except for the neritic class where the level 3 classifications of shallow and deep subtidal were used. The bars span the depth classes for the given number of species. Panel (a) shows the distributions for species that occur in only one depth class whereas panel (b) shows the distributions for species that occur in multiple depth classes.
CONCLUSIONS AND RECOMMENDATIONS

The field of bioinformatics has made great strides since the 1980s, particularly in the field of genetics, as exemplified by GenBank (http://www.ncbi.nlm.nih.gov/genbank/). In comparison, marine biogeography, ecology and conservation biology have lagged behind in this information revolution, in part because of the lack of standardization. However, the ability to conduct multi-factor syntheses is becoming increasingly important with the need to assess and manage biodiversity at regional and global scales. To achieve this end, both geographical and biological information must be captured in a standardized format to allow the union and intersection of both types of information.

Development of competing biogeographical schemas is not surprising as different authors have evaluated different taxa and applied different criteria to defining boundaries driven by different research needs. However, as a nested hierarchical schema that merges several of the existing biogeographical schemas, the MEOW system is the most flexible of the current systems for near-coastal environments. Although relatively recent, the MEOW schema has been used to assess invasive species globally (Molnar et al., 2008) and in the North Pacific...
It has also been used to assess the extent of marine protected areas globally (Spalding et al., 2008; Toropova et al., 2010), was adopted by the international Ramsar Convention on wetland conservation (Ramsar, 2008) and is currently being used as a framework to assess species’ vulnerability to climate change (H. Lee II et al., unpublished data). Because of this flexibility, we recommend the adoption of the MEOW schema as modified here for near-coastal biodiversity studies and conservation efforts, including invasive species and climate change. We recognize that other schemas have utility in addressing other types of questions; nonetheless, we strongly encourage researchers to provide sufficient geographical detail to allow mapping species distributions to the level of the MEOW ecoregions.

The limitations associated with synthesizing natural history attributes of marine/estuarine organisms are the paucity of Table 4 Crosswalk of the biogeographical and bathymetric ranges of chitons in the Northeast Pacific and Arctic (Gulf of California to Chukchi Sea; \(n = 167\) species). Biogeographical ranges are defined by the number of MEOW (Marine Ecosystems of the World; Spalding et al., 2007) ecoregions occupied in the world. The bathymetric ranges are based on the depth classes in Fig. 4. The boxes with solid black boundary lines highlight the number of species assumed to be at greatest risk while the boxes with dashed black boundary lines indicate high risk combinations. The outlined boxes with a white background are at risk due to temperature increases and sea level rise. The outlined boxes with grey background are at risk due to ocean acidification.

<table>
<thead>
<tr>
<th>Number of MEOW ecoregions</th>
</tr>
</thead>
<tbody>
<tr>
<td>One ecoregion</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Intertidal only</td>
</tr>
<tr>
<td>Intertidal to shallow subtidal</td>
</tr>
<tr>
<td>Intertidal to deep subtidal</td>
</tr>
<tr>
<td>Intertidal to bathyal</td>
</tr>
<tr>
<td>Shallow subtidal only</td>
</tr>
<tr>
<td>Shallow to deep subtidal</td>
</tr>
<tr>
<td>Shallow to bathyal</td>
</tr>
<tr>
<td>Deep subtidal only</td>
</tr>
<tr>
<td>Deep subtidal to bathyal</td>
</tr>
<tr>
<td>Deep subtidal to abyssal</td>
</tr>
<tr>
<td>Bathyal only</td>
</tr>
<tr>
<td>Abyssal only</td>
</tr>
</tbody>
</table>

Figure 10 Number of chiton species \((n = 167)\) by MEOW (Marine Ecosystems of the World; Spalding et al., 2007) ecoregion in the Northeast Pacific and Arctic. The grey area and numbers indicate the number of chiton species in the ecoregion that occur in two or more ecoregions. The black area and numbers indicate the number of chiton species that are endemic to that ecoregion. The total number of chiton species in an ecoregion is the sum of both numbers.
of efforts towards standardization and developing a qui- rhythmic natural history topology. The hierarchical schema presented addresses that limitation and forms the basis for capturing life history and habitat information on marine/estuarine species. Combined with the MEOW biogeographical schema, this biogeoinformatics system integrates and standardizes both distributional and natural history attributes in a systematic fashion. The topology has been designed to be consistent with most of the existing efforts at synthesizing natural history information so that information from efforts such as FishBase can be directly incorporated into the proposed system. The natural history schema presented here has been adopted by the multi-national North Pacific Marine Science Organization (PICES) for generating an atlas of the non-indigenous species in the North Pacific along with a supporting database (Otani et al., 2009; H. Lee II & D.A. Reusser, unpublished data). We recommend adoption of this biogeoinformatics framework by other international organizations and research groups to enhance the sharing, synthesis and systematic use of natural history data for addressing key conservation issues facing near-coastal ecosystems now and in the future. To promote such adoption, the architectural entity relationship diagram along with the data dictionary showing table structures and data types for this framework as implemented in the PICES non-indigenous database is available from the authors upon request.

ACKNOWLEDGEMENTS

D.A.R. was partially funded through AMI/GE OSS IAG no. DW-14-92231501-0 from the U.S. EPA and by the North Pacific Marine Science Organization (PICES) through funding from the Japanese Government. Special thanks to Melanie Frazier, U.S. EPA, for assistance with statistical analyses; Jody Stecher, Katie Marko and Karen Ebert for their continued efforts populating the PCEIS database; and Rachel Nehmer and Tad Larsen for developing PCEIS. Two anonymous referees provided insightful suggestions that improved the manuscript. This publication was subjected to review by the National Health and Environmental Effects Research Laboratory’s Western Ecology Division of the EPA and the Western Fisheries Research Centre of U.S. Geological Survey, and is approved for publication. However, approval does not signify that the contents reflect the views of the U.S. EPA. The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader; such use does not constitute official endorsement or approval by the U.S. Department of Interior, the U.S. Geological Survey, or the U.S. EPA of any product or service to the exclusion of others that may be suitable.

REFERENCES


**BIOSKETCHES**

**Deborah Reusser** is a geographer interested in ecoinformatics and biogeography. She is actively engaged in research to quantify the biogeographical patterns of native and non-indigenous species and how distribution patterns may be altered by a changing climate.

**Henry Lee II** is a marine biologist specializing in soft-bottom benthic organisms. He is interested in the biogeography of native and non-indigenous marine and estuarine species and is actively engaged in research to identify species at risk due to climate change and sea level rise.

Editor: Craig McClain