Running head: Ecoinformatics in Environmental Monitoring

Beyond Data Management: How Ecoinformatics Can Benefit Environmental Monitoring Programs

Stephen S. Hale and Jeffrey W. Hollister

Atlantic Ecology Division, National Health and Environmental Effects Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency

27 Tarzwell Drive, Narragansett, RI 02882 USA

Tel. 401-782-3048
Fax 401-782-3030
Email: hale.stephen@epa.gov

Published version of this paper is available from:
http://dx.doi.org/10.1007/s10661-008-0675-x

Abstract We review ways in which the new discipline of ecoinformatics is changing how environmental monitoring data are managed, synthesized, and analyzed. Rapid improvements in information technology and strong interest in biodiversity and sustainable ecosystems are driving a vigorous phase of development in ecological databases. Emerging data standards and protocols enable these data to be shared in ways that have previously been difficult. We use the U.S. Environmental Protection Agency's National Coastal Assessment (NCA) as an example. The NCA has collected biological, chemical, and physical data from thousands of stations around the U.S. coasts since 1990. NCA data that were collected primarily to assess the ecological condition of the U.S. coasts can be used in innovative ways, such as biogeographical studies to analyze species invasions. NCA application of ecoinformatics tools leads to new possibilities for integrating the hundreds of thousands of NCA species records with other databases to address broad-scale and long-term questions such as environmental impacts, global climate change, and species invasions.

Keywords Ecoinformatics; Ecological databases; Environmental monitoring; Taxonomic databases; U.S. National Coastal Assessment

Published version of this paper is available from:
http://dx.doi.org/10.1007/s10661-008-0675-x

1 Introduction

Compelling needs for environmental monitoring arise from the numerous and complex problems associated with global environmental change, such as climate change, loss of biotic diversity, nutrient enrichment, and land-use change (Bruns and Wiersma, 2004). Because monitoring is an essential component of environmental science and policy (Lovett et al., 2007), the U. S. Environmental Protection Agency began its Environmental Monitoring and Assessment Program (EMAP) in 1990. Large monitoring programs such as EMAP create large volumes of complex data. Managing and sharing these data with other monitoring programs to address regional, national, and global environmental issues provides a serious challenge to traditional data and information management systems.

1.1 U.S. National Coastal Assessment

We will use as a case study the U. S. National Coastal Assessment (NCA), the marine component of EMAP, which aims to assess the ecological condition of U.S. estuaries and coastal waters (NCA, 2007). Sampling began in 1990 in the Virginian Biogeographic Province and has since sampled thousands of probability-based stations in seven biogeographic provinces along the three major U.S. coasts, as well as the Gulf of Alaska, Hawaii, Guam, and Puerto Rico. NCA uses a random survey design to estimate condition of all estuaries in a region (such as a biogeographic province). NCA holds hundreds of thousands of records of distribution and abundance data for thousands of macroinvertebrate and fish species. There are concurrent data on physical, chemical, and habitat measurements, sediment and tissue contaminants, and sediment toxicity tests.

These data provide an unbiased baseline estimate of ecological conditions (USEPA, 2007) and can also be used in innovative ways to address long-term and broad-scale questions such as environmental impacts, global climate change, biodiversity, species invasions, and conservation efforts. They can be used to develop biogeography and biodiversity models (Halpin et al., 2006) and to test hypotheses regarding underlying principles that govern complex biological phenomena (Colwell, 1998).

To manage, analyze, report, and share these multi-dimensional data, NCA has used a variety of information management techniques including a web-based Oracle database management system to manage and distribute most of the data, ArcGIS for GIS mapping, and SAS for quality-assurance procedures and statistical analyses (NCA, 2007). We have now turned to some of the new tools of ecoinformatics.

1.2 Ecoinformatics

Ecoinformatics is broadly defined as the science of information in ecology, or “the study of the inherent structure of ecological information in order to create and apply computer technology for its management and analysis” (SEEK, 2008). The purpose of ecoinformatics is to increase the ease by which complex data can be managed and understood and to improve the scale and type of ecological science that can be addressed.

In the same manner that genomics researchers developed bioinformatics as a way to cope with ever-increasing volumes of genomics data, ecologists and computer scientists are developing ecoinformatics to deal with the even more complex ecological data from a discipline that includes far more numerous data types and covers vast temporal and spatial scales.

This includes the development and application of computer technologies for storing, handling, integrating, analyzing, modeling, and communicating the massive and ever-increasing amounts of data emerging from multi-dimensional ecological data (Ecoinformatics, 2007). Ecoinformatics merges data management with data analysis and modeling and includes ecological synthesis, simulation, forecasting, visualization, and ecological pattern analysis. It covers machine learning and artificial intelligence and may provide for automated data handling.

Ecological data can include information from genomes, individuals, populations, communities, and ecosystems, along with physical, chemical, and habitat data. A tremendous amount of information is continually being processed in interactions within and between these biological levels and their physical-chemical environment. Models explaining these interactions can contribute to ecological theory, such as a better understanding of food webs (Vos et al., 2006) and the transfer of chemical information in freshwater ecosystems (Van Donk, 2007).

Questions at regional and global scales that were intractable just a few years ago are now being studied. For example, the Encyclopedia of Life aims to serve as an online reference source and database for all the 1.8 million species named and known on the planet (EOL, 2007). In marine areas, the Ocean Biogeographic Information System (OBIS, 2007) integrates species distribution data from numerous studies, enabling new biogeographic analyses and hypothesis-testing. Synthetic databases compiled from many disparate data sources, for example, the synthetic studies of Jackson et al. (2001) on historical overfishing and Worm et al. (2006) on impacts of biodiversity loss, required a great deal of effort to find and integrate diverse data; now, tools of ecoinformatics that can automate some of these steps are becoming available (Jones, 2006).

Two examples illustrate some of the potential of ecoinformatics and new technology. Clothier and Bailey (2006) combined portable computers, GPS, wireless Internet, and Augmented Reality (akin to watching a football field with the first down line superimposed on the screen) to superimpose data on top of what a scientist sees in the field. Zones of air temperature could be displayed on top of a view of a natural mountain-side landscape to see how natural ecozones lined up with temperature data. Another example is the Biophony Grid Portal that manages acoustic data from numerous field-deployed microphones (Butler et al. 2006). Combining recorded bird songs from these microphones, a database of known bird acoustic signatures, and distributed GIS coverages allows researchers to identify a probability of presence for bird species.

The purpose of this paper is to review recent developments in ecoinformatics and describe how these have been and could be applied to the National Coastal Assessment and other environmental monitoring programs to advance ecological syntheses and analyses. We focus on four areas where ecoinformatics methods show particular promise: data discovery and integration, taxonomy and species identification, data analysis and modeling, and biogeography and ecological niche modeling.

2 Data Discovery and Integration

As ecological data are being collected on increasingly finer spatial scales, across broader spatial extents and at higher temporal frequencies, generating huge volumes of multi-dimensional data, one of the main difficulties is in locating data of an appropriate type and scale for a given research question and then merging those data into a synthetic dataset for analysis and modeling. The essential component for overcoming this difficulty and allowing for efficient discovery and integration is metadata.

2.1 Metadata

Metadata include information on measurements, methods, location, and time, information necessary for compiling a synthetic dataset, and are essential for merging data from multiple sources and are needed to counteract the tendency for data to become less useful with time (Michener, 2006). Currently, metadata standards are in a state of flux. When NCA began collecting data in 1990, we used the NASA Directory Interchange Format standard. Since then, we added the Federal Geographic Data Committee standard and are now beginning to use Ecological Metadata Language (EML), a key component of ecoinformatics efforts. EML is defined as a “metadata specification developed by the ecology discipline and for the ecology discipline” (KNB, 2007; Michener et al., 1997). The Knowledge Network for Biocomplexity has developed Morpho, a data management tool that eases EML creation and data discovery, and Metacat, a metadata database (KNB, 2007).

In addition to the Data Directory of the Environmental Monitoring and Assessment Program (EMAP, 2007), NCA metadata appear in EPA’s Environmental Information

Management System (EIMS, 2007), EPA's STORET/WQX water quality system (STORET, 2007), the Global Change Master Directory (GCMD, 2007), and the Knowledge Network for Biocomplexity (KNB, 2007). Additionally, NCA has stored Virginian Biogeographic Province data in *Ecological Archives* (Hale et al., 2002) under a common metadata format and plans to upload data from the remaining biogeographic provinces. To compare these and several other metadata standards, the Metadata Expert Team of the U. S. Integrated Ocean Observing System put together a Mega Metadata Matrix that crosswalks several of these standards (available on the Marine Metadata Interoperability website, MMI, 2007).

2.2 Data

Ecoinformatics methods are also useful for extracting information from the numerous intricacies of ecological data. McGuire et al. (2006) designed a data warehouse and online analytical processing tools to observe spatiotemporal patterns in benthic invertebrates and fish distributions. Flemons et al. (2007) describe a web-based GIS tool for exploring the over 110 million records of the Global Biodiversity Information Facility. Parr et al. (2007) developed an interactive visualization tool for data exploration of complex, multi-dimension ecological datasets so that patterns can be discerned and subsets chosen for further analysis.

For data integration, NCA data are translated to the formats of EPA's STORET/WQX system, which enable NCA data to be overlaid with that from numerous other EPA and state programs. STORET/WQX uses EPA's Data Exchange Network, a system of

Extended Markup Language (XML) templates with standard ontologies for machine-to-machine data exchange. To further enhance data integration, XML templates could be extended to field computers to provide metadata-driven field data entry (Jones, C., 2007).

Integration of NCA data with data from other monitoring programs is enhanced by use of common standards for parameter names and codes. For example, NCA and STORET, and several other monitoring programs, use the common taxonomic codes of the Integrated Taxonomic Information System (ITIS, 2007), which helps integrate biological data. Additionally, NCA shares species distribution data with the Ocean Biogeographic Information System (OBIS, 2007). Better registration and integration of NCA data with data from other monitoring programs, including linking with a future biodiversity Grid (Jones, A. C., 2007), would help synthetic analyses.

3 Taxonomy and Species Identification

Taxonomic databases, such as that of the NCA, illustrate the challenges and promise of ecoinformatics. Taxonomic names are ambiguous and change over time; experts disagree on classifications; new species are found—often, these issues make it difficult to synthesize data on species groups from different sources (Kennedy et al., 2005; White, 2007). The field of taxonomy is rich in data, with a literature spanning close to 250 years (Godfray, 2002; Scoble and Berendsohn, 2007). While much of the raw data at molecular and ecosystem level are digital, at the species level (such as taxonomic descriptions of morphological characteristics), data are often analogue and descriptive (Lane and Edwards, 2007).

Published version of this paper is available from:
http://dx.doi.org/10.1007/s10661-008-0675-x

Currently, NCA handles taxonomy in a normalized Oracle database. A user interface to the species data allows users to select an area, such as a biogeographic province, state or estuary, and retrieve a list of species documented in that area by NCA. The user can also select a species, by scientific name or by common name, or a higher taxon, and get a list of areas where this taxa has been found by NCA. Users with a list of species can then link to other databases, such as FishBase (FishBase, 2007) or the International Union for the Conservation of Nature's Red List (IUCN, 2007), for species-specific information on such things as distribution, habitat preferences, food items, and conservation status. Future NCA uses of ecoinformatics could include automating the integration of NCA species list with other databases containing additional information about those species (such as habitat preferences) through use of onward links (White, 2007).

3.1 Advances in taxonomic computing

Recent advances in ecoinformatics are addressing some of the challenges of taxonomic data by computerizing taxonomic data, developing tools for rapid identification of species, enabling data integration by managing taxonomic names and concepts, tracking changes in nomenclature, and monitoring changes in spatial distributions. Emerging from the Taxonomic Database Working Group (TDWG, 2007) is the Globally Unique Identifier system for taxonomic concepts that will simplify sharing and integration of taxonomic data. Taxonomic concepts (name, elements, relationships, publication, authority) are being developed as a solution to issues with taxonomic names; visualization software shows the relationships among different taxonomic treatments

(Kennedy et al., 2005; Berendsohn and Geoffroy, 2007). Quality-assuring NCA taxonomic names by using software for taxonomic concepts (Scoble and Berendsohn, 2007) could speed up a time-consuming process and reduce errors.

Scoble and Berendsohn (2007) propose a common format using Extended Markup Language (XML) to network biological collection databases. Curry and Connor (2007) describe a system for automatic extraction of data from taxonomic descriptions in the literature, which are often written in a standardized format, and include information on taxonomic name, morphology, authorship, location of specimens, illustrations, and biogeographic ranges. Although taxonomic names may have changed, the morphological descriptions remain valid and Chandramouli and Gauch (2006) developed a text mining system with a web spider that searches, for a certain taxa, literature that has been published after the original taxonomic description. This helps taxonomists keep up with new taxonomic treatments for the taxa, which may appear in a variety of different journals.

3.2 Species identification

Another advancement in taxonomy is the development of automated species identification systems (Morris and Boddy, 2006) and electronic field guides (Stevenson and Shrewsbury, 2006). Computer modeling tries to replicate what human eyes and brains do in recognizing and identifying species (Culverhouse, 2007). MacLeod et al. (2007) used both discriminant analysis on morphometric data and an artificial neural network trained with images of known species. As taxonomy becomes more automated, it will broaden the ability to take on large-scale biodiversity projects; re-invigorate the discipline of morphological systematics (which, in the face of DNA bar-coding and Gene Bank, must become more automated to survive); and capture knowledge of expert taxonomists who are retiring and not being replaced (MacLeod et al., 2007).

4 Data Analysis and Modeling

Ecoinformatics has fostered methods to help analyze large, complex synthetic databases. For instance, Pascoe et al. (2006) developed BugML, an XML standard for sharing aquatic biomonitoring data from distributed data sources in order to construct a national indicator of ecological condition of freshwater ecosystems. Williams and Poff (2006) evaluated the application of artificial neural networks, evolutionary algorithms, and classification/regression trees to the USEPA’s Environmental Monitoring and Assessment Program stream monitoring data in the U.S. mid-Atlantic area to calculate ecological indices. The ecological condition of stream benthic macroinvertebrates has also been evaluated with self-organizing maps (Horrigan and Baird, 2006; Song et al. 2006), a data visualization technique that reduces the dimensions of data through the use of self-organizing neural networks. Such tools help ecologists without extensive knowledge of computational science to extract more information from multi-dimensional, interrelated datasets (Williams and Poff, 2006).

When observing a disturbed community, it is often difficult to identify what stressor(s) caused the condition; Van den Brink et al. (2006) addressed this with a tool that links taxa sensitivity to stressors database with a database on species traits such as

Published version of this paper is available from:
http://dx.doi.org/10.1007/s10661-008-0675-x

being a water breather, body size, and maximum age. Lastly, synthetic datasets created from NCA data and other data sources could enhance the usefulness of NCA data in discerning larger-scale patterns. Future landscape ecology studies could take advantage of ecoinformatics tools such as that of Olden et al. (2006) who used Artificial Neural Networks with EMAP data and landscape data across multiple spatial scales.

5 Biogeography and Ecological Niche Modeling

Biogeography and ecological niche modeling are areas that have had significant ecoinformatics developments. Best et al. (2007) described a system that takes advantage of recent advances in GIS, open-source software development, Open Geospatial Consortium data standards, and content management systems to stimulate new approaches in biogeographic and conservation research. For biodiversity data, Jones, A. C. (2007) proposed the use of Grid-based computing with interconnected databases, computer processing, and analytical tools located at different physical locations. This would provide a large amount of computational power to study such things as migrational patterns, range shifts, and species invasions.

Ecoinformatics makes possible, through increasingly sophisticated ecological niche modeling tools such as artificial neural networks (e.g., Olden et al., 2006; Horrigan and Baird, 2006) and self-organizing maps (e.g., Song et al., 2006; Worner and Watts, 2006), techniques that can be used to predict potential species distributions under climate change, different land management schemes, and new environments encountered by invasive species (Pennington et al., 2006).

In each of the U.S. biogeographic provinces, NCA has developed or is developing benthic indices that reflect the impact of stressors such as chemical contaminants, eutrophication, and hypoxia on species distribution patterns. To accomplish this, species distributions, which are patterned by latitude and longitude, depth, temperature, salinity, and sediment grain size, are needed for non-polluted reference sites. Additionally, the NCA indices from separate biogeographical provinces need to be inter-calibrated to be sure they are measuring the same thing at the same scale. These efforts require considerable effort in discovery and integration of data from multiple sources. New ecoinformatics tools for integrating taxa, such as software for handling taxonomic concepts like the Globally Unique Identifier (TDWG, 2007), could save time and reduce errors.

Biogeographic boundaries in the past have sometimes been established based on a limited number of taxa, such as mollusks. NCA enables all taxa of benthic macroinvertebrates (biogeography of communities) to be used. NCA data in the Ocean Biogeographic Information System allow integration with other benthic studies. Future work on the biogeography of coastal ecosystems will use ecoinformatics tools to synthesize data from studies on phytoplankton, zooplankton, benthic invertebrates, and fishes to achieve a biogeography based on ecosystems. The field is ripe for use of synthetic databases, automated tools, and Grid technology.

6 Conclusions

The need for ecological data to address broad-scale and long-term questions such as environmental impacts, global climate change, and biodiversity is urgent and increasing. Ecoinformatics provides the means to manage, integrate, analyze, and share these data. Present-day ecological databases are often smaller than those used for business, which frequently involve huge numbers of transactions. For example, systems for handling money transfers or telephone calls may log millions of transactions per day. This creates databases that are constantly changing, unlike NCA and many marine ecology programs that take only snapshots of environmental conditions. But ecological systems themselves experience an enormous number of transactions per day—for example, sunlight shines on plants, plants take up nutrients and photosynthesize, grazers graze, and predators seize prey—all patchily distributed across a marine landscape. Heretofore, we have not simultaneously measured these immense ecosystem transactions because our field studies are limited in time, space, and scope of variables. This may change in the future as our knowledge of ecosystem processes and our ability to measure them increases. Wireless sensor networks for habitat monitoring will greatly increase the stream of incoming ecological data (Suri et al., 2006) and projects such as the National Ecological Observatory Network (NEON, 2007) will provide a national system of ecological sensor networks connected by advanced cyberinfrastructure. Linking these volumes of information flows from the level of cells to ecosystems over broad and variable geospatial domains requires the advanced data management and computational techniques that are emerging from the field of ecoinformatics (Vos et al., 2006).

It is clear there will be abundant challenges for managing ecological data in the future (Michener, 2000). Significant expansion of our understanding of ecological systems is dependent on further improvements in cyberinfrastructure. Ecoinformatics has already improved the scale and type of ecological science that can be addressed. Ecological databases holding millions of records and gigabytes to terabytes of data are now common and will continue to grow in volume, breadth, and complexity (Michener, 2006). Astronomers are considering petabyte (a quadrillion, or 10^{15} bytes) databases; the Large Synoptic Survey Telescope will produce over seven petabytes of data per year and fundamentally change the way astronomy is done (USNVO, 2007). Ecological insights that we can not yet imagine may emerge when ecological studies and databases reach this level.

Acknowledgments

This paper was improved by reviews from E. Dettmann, N. Detenbeck, H. Buffum, L. Kracker, and J. Paul. Although the research described in this article has been funded wholly by the U.S. Environmental Protection Agency, it has not been subjected to Agency review. Therefore, it does not necessarily reflect the views of the Agency. This is contribution number AED-07-077 of the U.S. Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Atlantic Ecology Division, Narragansett, RI. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

Berendsohn, W. G,. & Geoffroy, M. (2007). Networking taxonomic concepts—uniting without 'unitary-ism'. (In G.B. Curry & C.J. Humphries (Eds.), *Biodiversity databases: Techniques, politics, and applications* (pp. 13–22). Boca Raton: CRC Press.)

Best, B. D., Halpin, P. N., Fujioka, Ei, Read, A. J., Song, S. Q., Hazen, L. J., & Schick, R. S. (2007). Geospatial web services within a scientific workflow: Predicting marine mammal habitats in a dynamic environment. *Ecological Informatics*, 2, 210–223.

Bruns, D. A. & Wiersma, G. B. (2004). Conceptual basis of environmental monitoring systems: a geospatial perspective. (In G. B. Wiersma (Ed.), *Environmental Monitoring* (pp. 1–35). Boca Raton: CRC Press.)

Butler, R., Servilla, M., Gage, S., Basney, J. Welch V., Baker, B., Fleury, T., Duda, P., Gehrig, D., Bletzinger, M., Tao, J. and Freemon, D.M. (2006). Cyberinfrastructure for the analysis of ecological acoustic sensor data: a use case study in grid deployment. *Proceedings of Challenges of Large Applications in Distributed Environments*, 2006, Institute of Electrical and Electronics Engineers, pp. 25–33.

Chandramouli, A. & Gauch, S. (2006). Semi-automatic update of existing taxonomy using text mining [abstract]. In: *5th International Conference on Ecological Informatics*; Dec 4–6, 2006; Santa Barbara, CA. [014].

Clothier, M. & Bailey, M. (2006). Visualization in outdoor ecological environments [abstract]. In: *5th International Conference on Ecological Informatics*; Dec 4–6, 2006; Santa Barbara, CA. [P1].

Colwell, R. (1998). Balancing the biocomplexity of the planet's living systems: A twenty-first century task for science. *BioScience*, 48, 786–787.

Culverhouse, P. (2007). Human and machine factors in algae monitoring performance. *Ecological Informatics*, 2, 361–366.

Curry, G. B. & Connor, R. J. (2007). Automated extraction of biodiversity data from taxonomic descriptions. (In G.B. Curry & C.J. Humphries (Eds.), *Biodiversity databases: Techniques, politics, and applications* (pp. 63–81). Boca Raton: CRC Press.)

Ecoinformatics. (2007). Retrieved June 01, 2007 from www.ecoinformatics.org.

EIMS [Environmental Information Management System]. (2007). Retrieved June 01, 2007 from www.epa.gov/eims.

EMAP [Environmental Monitoring and Assessment Program]. (2007). Retrieved June 01, 2007 from www.epa.gov/emap.

EOL [Encyclopedia of Life]. (2007). Retrieved June 01, 2007 from www.eol.org.

FishBase. (2007). Retrieved June 01, 2007 from www.fishbase.org.

Flemons, P., Guralnick, R., Krieger, J., Ranipeta, A., & Neufeld, D. (2007). A web-based GIS tool for exploring the world’s biodiversity: The Global Biodiversity Information Facility Mapping and Analysis Portal Application (GBIF-MAPA). *Ecological Informatics*, 2, 49–60.

GCMD [Global Change Master Directory]. (2007). Retrieved June 01, 2007 from gcmd.nasa.gov.

Godfray, H. C. J. (2002). Challenges for taxonomy. *Nature*, 417, 17–19.

Hale, S.S., Hughes, M.M., Strobel, C.J., Buffum, H.W., Copeland, J.L., & Paul, J.F. (2002). Coastal ecological data from the Virginian Biogeographic Province, 1990–1993. *Ecology*, 83, 2942 and *Ecological Archives*, E083-057.

Halpin, P.N., Read, A.J., Best, B.D., Hyrenback, K.D., Fujioka, E., Coyne, M.S., Crowder, L.B., Freeman, S.A., Spoerri, C. (2006). OBIS-SEAMAP: developing a biogeographic research data commons for the ecological studies of marine mammals, seabirds, and sea turtles. *Marine Ecology Progress Series*, 316, 239–246.

Horrigan, N. & Baird, D. J. (2006). Using artificial neural network models to diagnose potential causes of impairment in stressed river invertebrate communities [abstract]. In: *5th International Conference on Ecological Informatics;* Dec 4–6, 2006; Santa Barbara, CA. [O55].

ITIS [Integrated Taxonomic Information System]. (2007). Retrieved June 01, 2007 from www.itis.gov.

IUCN [International Union for the Conservation of Nature]. (2007). Retrieved June 01, 2007 from www.iucn.org.

Jackson, J. B.C., Kirby, M. X., Berger W. H, ., Bjorndal, K.A., Botsford, L.W., Bourque, B.J, Bradbury, R. H., Cooke, R. G., Erlandson, J., Estes, J. A., Hughes, T. P., Kidwell, S, Lange, C. B, Lenihan, H. S, Pandolfi, J. M., Peterson, C. H., Steneck, R. S., Tegner, M. J., & Warner, R. R. (2001). Historical overfishing and the recent collapse of coastal ecosystems. Science, 293(5530), 629–637.

Jones, A. C. (2007). The Grid and biodiversity databases. (In G.B. Curry & C.J. Humphries (Eds.), *Biodiversity databases: Techniques, politics, and applications* (pp. 83–97). Boca Raton: CRC Press.)

Jones, C., Blanchette, C., Brooke, M., Harris, J., Jones, M., & Schildhauer, M. (2007). A metadata-driven framework for generating field data entry interfaces in ecology. *Ecological Informatics*, 2, 270–278.

Published version of this paper is available from:
http://dx.doi.org/10.1007/s10661-008-0675-x

Jones, M. (2006). Using metadata and ontologies to facilitate ecological analysis and modeling [abstract]. In: *5th International Conference on Ecological Informatics*; Dec 4–6, 2006; Santa Barbara, CA. [K1].

Kennedy, J. B., Kuhla, R., & Paterson, T. (2005). Scientific names are ambiguous identifiers for biological taxa: Their context and definition are required for accurate data integration. (In B. Lud≅scher & L. Raschid (Eds.), *Data Integration in the Life Sciences* (pp. 80–95). Berlin: Springer.)

KNB [Knowledge Network for Biocomplexity]. (2007). Retrieved June 01, 2007 from knb.ecoinformatics.org.

Lane, M .A. & Edwards, J .L. (2007). The Global Biodiversity Information Facility (GBIF). (In G.B. Curry & C.J. Humphries (Eds.), *Biodiversity databases: Techniques, politics, and applications* (pp. 1–4). Boca Raton: CRC Press.)

Lovett, G. M., Burns, D. A., Driscoll, C. T., Jenkins, J. C., Mitchell, M. J, Rustad, L., Shanley, J. B., Likens, G. E., & Haeuber, R. (2007). Who needs environmental monitoring? *Frontiers in Ecology and Environment*, 5, 253–260.

LTER [Long-Term Ecological Research]. (2007). Informatics. Retrieved June 01, 2007 from www.lternet.edu/informatics.

MacLeod, N., O’Neill, M., & Walsh, S. A. (2007). A comparison between morphometric and artificial neural network approaches to the automated species recognition problem in systematics. (In G.B. Curry & C.J. Humphries (Eds.), *Biodiversity databases: Techniques, politics, and applications* (pp. 37–62). Boca Raton: CRC Press.)

Published version of this paper is available from:
http://dx.doi.org/10.1007/s10661-008-0675-x

McGuire, M. P., Komlodi, A., & Gangopadhyay, A. (2006). A user-centered design of a spatiotemporal data warehouse for data exploration in ecological research [abstract]. In: *5th International Conference on Ecological Informatics*; Dec 4–6, 2006; Santa Barbara, CA. [18].

Michener, W. K. (2006). Meta-information concepts for ecological data management. *Ecological Informatics*, 1, 3–7.

Michener, W. K., Brunt, J. W., Helly, J., Kirchner, T. B., & Stafford, S. G. (1997). Non-geospatial metadata for the ecological sciences. *Ecological Applications*, 7, 330–342.

Michener, W.K. (2000). Ecological knowledge and future data challenges. (In W.K. Michener & J. W. Brunt (Eds.), *Ecological Data: Design, Management and Processing* (pp. 162–174)., Oxford: Blackwell Science.)

MMI [Marine Metadata Interoperability]. (2007). Retrieved June 01, 2007 from marinemetadata.org.

Morris, C. & Boddy, L. (2006). Some inherent problems in biological identification from images [abstract]. In: *5th International Conference on Ecological Informatics*; Dec 4–6, 2006; Santa Barbara, CA. [063].

NEON [National Ecological Observatory Network]. (2007). Retrieved June 01, 2007 from neoninc.org.

NCA [National Coastal Assessment]. (2007). Retrieved June 01, 2007 from www.epa.gov/emap/nca.

OBIS [Ocean Biogeographic Information System]. (2007). Retrieved June 01, 2007 from www.iobis.org.

Olden, J. D., Poff, N. L., & Bledsoe, B. P. (2006). Incorporating ecological knowledge into ecoinformatics: An example of modeling hierarchically structured aquatic communities with neural networks. *Ecological Informatics*, 1,33–42.

Parr, C. S., Lee, B., & Bederson, B. B. (2007). EcoLens: Integration and interactive visualization of ecological datasets. *Ecological Informatics*, 2, 61–69.

Pascoe, T. J., Kralidis, T., Cree, J., & Baird, D. (2006). BugML: Implementing XML standards for sharing and interoperability of aquatic biomonitoring data. [abstract]. In: *5th International Conference on Ecological Informatics*; Dec 4–6, 2006; Santa Barbara, CA. [P6].

Pennington, D., Michener, W. K., Higgins, D., & Peterson, A. T. (2006). Ecological niche modeling with the Kepler workflow system [abstract]. In: *5th International Conference on Ecological Informatics*; Dec 4–6, 2006; Santa Barbara, CA. [010].

Scoble, M. J. & Berendsohn, W. G. (2007). Networking biological collections databases: Building a European infrastructure. (In G.B. Curry & C.J. Humphries (Eds.), *Biodiversity databases: Techniques, politics, and applications* (pp. 23–35). Boca Raton: CRC Press.)

SEEK [Science Environment for Ecological Knowledge]. (2008). Retrieved January 2, 2008 from http://seek.ecoinformatics.org.

Song, M., Park, Y., Kwak, I., Woo, H. & Chon, T. (2006). Characterization of benthic macroinvertebrate communities in a restored stream by using self-organizing map. *Ecological Informatics*, 1, 295–305.

Published version of this paper is available from:
http://dx.doi.org/10.1007/s10661-008-0675-x

Stevenson, R.D. & Shrewsbury, S. (2006). Identifying and quantifying the usage of information elements found in field guides [abstract]. In: *5th International Conference on Ecological Informatics*; Dec 4–6, 2006; Santa Barbara, CA. [O70].

STORET. (2007). Retrieved June 01, 2007 from www.epa.gov/storet.

Suri, A., Iyengar, S. S., & Cho, E. (2006). Ecoinformatics using wireless sensor networks: An overview. *Ecological Informatics*, 1, 287–293.

TDWG [Taxonomic Database Working Group]. (2007). Retrieved June 01, 2007 from www.tdwg.org.

USEPA. (2007). National Coastal Condition Report III. EPA-620/R-XX/XXX. Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC.

USNVO [The U.S. National Virtual Observatory]. (2007). Retrieved June 01, 2007 from www.us-vo.org.

Van den Brink, P., Rubach, M. Baird, D., & Maund, S. (2006). Assessing ecosystem health and impairment by species traits and their relation to stressors [abstract]. In: *5th International Conference on Ecological Informatics*; Dec 4-6, 2006; Santa Barbara, CA. [P7].

Van Donk, E. (2007). Chemical information transfer in freshwater plankton. *Ecological Informatics*, 2, 112–120.

Vos, M., Vet, L.E.M., W≅ckers, F.L., Middelburg, J.J., van der Putten, W.H., Mooij, W.M., Heip, C.H.R., & van Donk, E. (2006). Infochemicals structure marine, terrestrial and freshwater food webs: Implications for ecological informatics. *Ecological Informatics*, 1, 23–32.

White, R. J. (2007). Linking biodiversity databases: Preparing species diversity information sources by assembling, merging and linking databases. (In G.B. Curry & C.J. Humphries (Eds.), *Biodiversity databases: Techniques, politics, and applications* (pp. 111–128). Boca Raton: CRC Press.)

Williams, J. B. & Poff, N. L. (2006). Informatics software for the ecologist's toolbox: A basic example. *Ecological Informatics*, 1, 325–329.

Worm, B., Barbier, E. B., Beaumont, N., Duffy, J. E., Folke, C., Halpern, B. S., Jackson, J. B. C., Lotze, H. K., Micheli, F., Palumbi, S. R., Sala, E., Selkoe, K. A., Stachowicz, J. J., & Watson, R. (2006)**.** Impacts of biodiversity loss on ocean ecosystem services. *Science*, 314, 787–790.

Worner, S. & Watts, M. (2006). Null model analysis of a self organising map of invasive species distributions [abstract]. In: *5th International Conference on Ecological Informatics;* Dec 4–6, 2006; Santa Barbara, CA. [O26]