Ecoinformatics: A Review of Approach and Applications in Ecological Research

Chau Chin Lin*

Society of Subtropical Ecology, Taipei, Taiwan

ABSTRACT

Ecological communities adapt the concept of informatics in the late 20 century and develop rapidly in the early 21 century to form Ecoinformatics as the new approach of ecological research. The new approach takes into account the data-intensive nature of ecology, the precious information content of ecological data, and the growing capacity of computational technology to leverage complex data as well as the critical need for informing sustainable management of complex ecosystems. It comprehends techniques for data management, data analysis, synthesis, and forecasting on ecological research. The present paper attempts to review the development history, studies and application cases of ecoinformatics in ecological research especially on Long Term Ecological Research (LTER). From the applications show that the ecoinformatics approach and management system have formed a new paradigm in ecological research

Keywords: Ecology, EML, Informatics, Information management system, LTER, Metadata

Introduction

Informatics is a distinct scientific discipline, characterized by its own concepts, methods, body of knowledge, and open issues. It covers the foundations of computational structures, processes, artifacts and systems; and their software designs, their applications, and their impact on society (CECE, 2017). Although Informatics was emerged in computer science, it has influenced many disciplines. Ecological communities adapt the concept of informatics in the late 20 century and develop rapidly in the early 21 century. The book "Ecological Informatics" (short as Ecoinformatics) was published in 2003 raised the concept of "understanding ecology by biologically-inspired computation"(Recknagel, 2003). The concept

takes into account the data-intensive nature of ecology, the precious information content of ecological data, and the growing capacity of computational technology to leverage complex data as well as the critical need for informing sustainable management of complex ecosystems. It comprehends techniques for data management, data analysis, synthesis, and forecasting on ecological research (Recknagel, 2008). The development of ecoinformatics has therefore built a new paradigm on ecological research (Porter & Lin, 2017). Bird banding studies are an early example of ecoinformatics: the combination of ecological research with informatics. Consider the history of the theory of bird migration. It was only in the early 1900s, after the advent of systematic bird-banding that a more accurate picture of migration emerged (Berthold *et al.*, 2001). In bird banding studies, coded tags or bands are placed on the legs of captured birds, allowing individual birds to be identified wherever else they might be found in the world. By maintaining a database that links a specific band code with the initial capture location and subsequent observations, a picture of the migration pattern

of individual birds can be assembled.

The purpose of the present paper attempts to review the development history, studies and application cases of ecoinformatics in ecological research especially on Long Term Ecological Research (LTER).

The History of Ecoinformatics

In the 2011 edition of Journal of Vegetation Science used "Ecoinformatics" as its special feature. The editorial cited Kareiva (2001) and Brunt *et al*. (2002) to mention "Ecoinformatics as a term and subfield of ecology first emerged from the biodiversity informatics initiatives of the U.S. Long-term Ecological Research Network (LTER) and the U.S. National Center for Ecological Analysis and Synthesis (NCEAS) in the late 1990s and early 2000s" (Dengler *et al*., 2011) . However, biodiversity informatics and ecoinformatics are largely overlapping fields, the first giving somewhat more emphasis to the taxonomic position of the analyzed species, the second more to interactions among taxa and between taxa and with their abiotic environment (Lin *et al*., 2008a).

Each discipline now has its own journal: Biodiversity Informatics started in 2004 and Ecological Informatics in 2006 (Dengler, 2011). Ecoinformatics not only overlaps with biodiversity informatics. It also related to bioinformatics. The need for a new bioinformatics has been suggested by Jone *et al*. (2006). The reason a new bioinformatics needed Jone *et al*. (2006) thought is "with the rapid growth of human populations and their impacts, it becomes critically important to better describe and understand natural processes. The increasing demands within ecology for greater access to more types of data emphasize the need for integrated data-management solutions that span biological sub disciplines from the gene to the biosphere. "

In the beginning of ecoinformtics introduced, it has emphasized primarily leveraging complex ecological data by computing. From the book of "Ecological Informatics" (Recknagel, 2003), the author mentioned "Ecosystems analysis, synthesis and forecasting in the past ten years were very much influenced by inventions in computational technology such as high performance computing and biologically-inspired computation. This computational approach allows discovering knowledge in complex". However, Jones *et al*. (2006) and Michener and Jones (2012) suggest archiving, sharing, and integrating ecological data as the key focus of ecoinformatics. Moreover, Hampton *et al*. (2017) raised the issue of the gap knowledge and skills of ecoinformatics among ecological researches and the next generation scientists is urging to initiate training programs. They suggest five key skills: (1) data management and processing, (2) analysis, (3) software skills for science, (4) visualization, and (5) communication methods

for collaboration and dissemination are needed.

When the third version of "Ecological Informatics" published in 2018, the scope of ecoinformatics focused on acquisition, archival, analysis, synthesis, and forecasting of ecological data by novel computational techniques (Recknagel and Michener, 2018). From the first chapter of the book, the authors mentioned the core as: at its core, ecological informatics combines developments in information technology and ecological theory with applications that facilitate ecological research and the dissemination of results to scientists and the public. Its conceptual framework links ecological entities (genomes, organisms, populations, communities, ecosystems, landscapes) with data management, analysis and synthesis, and communicating and informing decisions by following the course of a loop.

Today, ecology has joined a world of big data (Farley *et al*., 2018; Hampton *et al*., 2013). Ecological data can be organized into data systems. A data system usually comprises many data types. Different research communities work with different data systems, and each community has different levels of expertise and historical investments in ecoinformatics.

The Critical Role of Ecological Data

Data plays two critical roles in the process of scientific research (Porter & Lin, 2017). First, most scientific questions or theories have their basis in observations of some kind. The story of Charles Darwin during his travels on H.M.S. Beagle (1831-1836) observed the diverse flora and fauna of South America, the Galapagos Islands, New Zealand and Australia reveals that he observed in the plants and animals of these distant areas helped him reach the conclusions that were ultimately embodied in his theory of evolution through natural selection. Without those observations, Darwin would have had little reason to question the ruling paradigms of his day. This same relationship, with data and observations driving researchers to formulate new questions, occurs every day in science.

The second critical role played by data is in the testing of hypotheses. It is only repeatedly comparing what our data tells us with what our hypotheses predict will allow us to distinguish between hypotheses that are true from those that are false. A key feature of the data collected in most cases was that the data could have shown patterns that refuted the hypothesis proposed by the researcher conducting the study. Although data may be used to reject a hypothesis, it does not mean that all aspects of a hypothesis that are not rejected are true (Popper, 1959). A classic example is the hypothesis that "all swans are white." Data collected in most parts of the world would fail to reject this hypothesis. However, if data from Australia were included this hypothesis would be quickly

rejected because populations of black swans (*Cygnus atratus*) are found there. One of the challenges of testing ecological hypotheses is that ecological processes take place over a variety of time scales, from less than a second for a fish eating a minnow to hundreds of years for the growth of a forest (Porter & Lin, 2017).

As a general rule, ecological data become more valuable over time as the length of time series grow longer (Michener and Jones, 2012). Many examples show that the processes of research that take decades to millennia to operate. For example, the Eurasian collared doves took 70 years to spread across Europe, and that the spread of maples and hemlocks across North America following glacial retreat took thousands of years. Short time scales are relatively easy to observe. However, longer time scales, especially those that exceed a single human lifespan are more difficult to study. Even processes that operate at intermediate scales can be difficult to observe - it's easy to observe the changes in a second-hand of a watch, but changes in the minute-hand are hard to see, and changes in the hour hand are almost impossible to see, even though all the hands are in constant motion.

From ecoinformatics point of view most data becomes gradually more valuable over time, because it becomes increasingly difficult to reproduce or assemble data from the past. Some data may decline in value over time because we may learn that the methods or instrumentation used were ineffective (Michener *et al*., 1997). Therefore, data management system is essential for ecological research.

Ecological Data Management

As mentioned in "the history of ecoinformatics" section previously, the scope of ecoinformatics focused on acquisition, archival, analysis, synthesis, and forecasting of ecological data by novel computational techniques. It means a process that starts at the conceptualization of the project and concludes after the data have been archived and the results have informed future research as well as resource management, conservation, and other types of decision-making. The process has been called ecological data management (Porter & Lin, 2017). It is conceptualized in terms of a data life cycle whereby: (1) projects are conceived and data collection and analyses are planned; (2) data are collected and organized, usually into data tables (e.g., spreadsheets) or databases; (3) data are quality assured using accepted quality assurance/quality control (QA/QC) techniques; (4) data are documented through the creation of metadata that describe all aspects of the data and research; (5) data are preserved in a data repository or archive so that they may be reused and shared; (6) data are discovered or made discoverable so that they may be used in synthesis efforts or to reproduce results of a study; (7) data are integrated with other data in order to answer specific questions such as examining the influence of climate extremes on pollination ecology; and (8) data are explored, analyzed and visualized, leading to new understanding that can then be communicated to other scientists and the public (Michener, 2018).

Ecoinformatics Approach of Ecological Data Management

Ecoinformatics approaches are required for ecological data management. Over time, according to the principle of entropy, systems tend to become increasingly disordered, unless external energy is applied. For ecological data it is called data entropy (Michener *et al*., 1997). This principle applies strongly to ecological data, which, in the absence of active efforts to preserve them, can be lost in a surprisingly short period of time. Often in the field, data are captured on a paper form, and then transcribed in to a computer-readable form (e.g., a spreadsheet) for analysis. However, neither the paper form nor a spreadsheet file stored on a computer has characteristics that support long-term archival storage or facilitate the sharing of data.

Data recorded on paper have the potential to last for long periods of time (Porter & Lin, 2017). For example, ancient scrolls from the Chinese Xia Dynasty, which were created in 2100 BCE, have survived for over 4,000 years. However, in practice, that potential is seldom realized. First, paper copies are easily misplaced or lost. Even if stored in file cabinets, those data may be lost when the files are discarded when a researcher retires. Second, paper is susceptible to both external and internal damage. Fires and water damage can both destroy paper and many types of modern paper have poor archival quality. Due to acid in the paper itself, the pages become increasingly brittle and yellowed over time. Additionally, because paper is unwieldy to duplicate and requires lots of storage space there are usually only a few copies of data forms made. This limits our ability to share data among a wide array of researchers and makes the data more susceptible to loss.

In many ways, electronic data are even worse than paper when it comes to having good archival qualities. The media of electronic data storage are also susceptible to damage from fire and water. Moreover, most storage media have a limited lifetime. Magnetic storage of data used hard drives and tapes tend to fade over time, so that data become unreliable after periods of time as short as a decade. Optical storage, such as compact disks (CDs) have a longer potential lifespan (up to 100 years), but rapid changes in technology make it unlikely that you will be able to find a reader for these disks in 20 years time. Rapid changes in the electronic formats associated with software also pose problems. Despite their poor ar-

chival qualities, digital data have strong advantages for data sharing, and if managed using the best practices of ecoinformatics, can overcome some of the archival limitations (Porter, 2000). Often this involves storing copies of the data in generic formats, such as text files, that can be read by many different kinds of software. However, some form of continued management is usually required to maintain access to digital data over the long term.

In addition to using ecoinformatics approaches for data management, some types of data require use of informatics tools from the beginning. Data collected by digital sensors are often produced at a rate that would defeat manual approaches (Porter *et al*., 2005). For example, carbon flux towers are used to measure productivity in the surrounding landscape. To do this, they measure both the wind speed and direction (so they know the source of air coming to the tower) and the amount of carbon dioxide in the air. The challenge, from the information management perspective, is that they take these measurements 10 times per second! That means 36,000 carbon dioxide measurements per hour, 864,000 measurements in a day and 314,712,000 measurements in a year. Automated computational tools are absolutely required to manage this flow of data (Lin & Hsia, 2010).

Ecoinformatics approaches also help facilitate the sharing of data (Vanderbilt *et al*., 2015). In the past ecological data is usually handled by individual researchers who do not use ecoinformatics approaches. Data is collected and analyzed, and publications prepared, but the raw data itself is often lost in the years following its use in a publication. When data is lost or discarded, this is often not intentionally. Researchers may store data after they have completed an analysis on a computer disk, or store a paper copy in a file cabinet or laboratory notebook. However, when the disk is replaced, or fails, the data may be lost. Similarly, when a researcher retires, their files and notebooks are typically packed away in boxes and these are eventually discarded. Even if the data itself is preserved, without metadata or documentation, over time even the researcher who originally collected the data may be unable to recall all the details needed to use the data (Michener *et al*., 1997) . In contrast if data is archived and shared, it lends itself to new analyses. This allows the data to be re-used for additional analyses, and perhaps more importantly, combined with other data to allow regional, national and global analyses and long-term studies.

The Key of Ecological Data Management

Ecological sciences are studies that attempt to understand complex questions of ecosystems, to engage in interdisciplinary collaboration, and to collect data in larger spatial and temporal scales. Ecological data collecting usually uses a variety of protocols on the field. The resulting heterogeneous datasets are produced and stored in very different ways, which may not be familiar to all scientists (Jones *et al*., 2006). These datasets are also dispersed only within small research communities. Practically, an ecological dataset might reside in files that are poorly documented. This leads to the consequence that the files often become unusable upon a scientist's departure or retirement. There is growing recognition that ecological data, especially long-term ecological data, should be networked and persevered for future studies in replicating and validating scientific conclusions, and in enlarging spatial and temporal scales (Porter, 2000).

The goals of ensuring ecological data availability and usability usually are achieved by the use of digital media to capture, store, and process increasingly larger volumes of data, but this has in turn created new challenges for indexing, navigating, and documenting this sudden wealth of information (Lin *et al*., 2008b). Metadata is the critical tool for dealing with this challenge (Michener, 2006). The term "meta" derives from the Greek word meaning "after, beside, between, or with". On the other hand, the term "data" derives from the Latin plural of datum meaning "something given, a fact, a proposition". So metadata can be define as "structured data which describes, explains, locates the characteristics of a source" (McCartney & Jones, 2002). There are three features a metadata generally include, which are content, context, and structure. Content relates to what the information contains. Context indicates all the background of information source concerning who, what, why, where, and how. Structure relates to the formal set of associations among information objects. Since metadata is a kind of standard of describing information sources, there are always a variety of metadata standards. Some have been developed to describe and provide access to a particular type of information resource, such as geospatial resources (the US Federal Geographic Data Committee, FGDC). There are others like the Dublin Core developed in 1995, which combined the Online Computer Library Center (OCLC) and the National Center for Supercomputing Application (NCSA). The Dublin Core has received widespread acceptance among many communities for resource discovery.

Ianella (1999) pointed out that "there is an obvious need for a unified information model that reflects the needs of many metadata communities". Since 1999, this need has been analyzed and conceptualized (Bearman *et al*., 1999). The conceptualized model consists of four components.

Resource—The resource that is being described.
Metadata—A number of metadata instances that describe the Resource, usually created by an Agent.
Agent—An entity (human, organization, etc) that is re-

sponsible for creating the metadata (and sometimes the Resource) and performs actions on the Resource.

Event—An action that occurred (or will occur) pertaining to the Resource.

As illustrated above, metadata consists of complex constructs to create and maintain. It can be expensive to proceed through the whole system. How then can one justify the costs and efforts involved? Fortunately, this challenge has enabled the creation of new projects related to developing tools and management of metadata from open source groups. Based on this information flow, Ecological Metadata Language (EML) and related tools are one of the significant examples of the integration and synthesis of ecological data at a global level. To provide the ecological community with an extensible, flexible metadata standard for use in data analysis and archiving that will allow automated machine processing, searching and retrieval, members of the ecological research community in the United States have worked on compiling metadata as part of the data archiving process for over a decade (Jones, 2003). In 1997, researchers at the National Center for Ecological Analysis and Synthesis (NCEAS) at the University of California, Santa Barbara began implementing the first version of EML, which was later revised several times and culminated in EML version 1.4 (McCartney and Jones, 2002). Currently, EML has been released into EML version 2.x which was an effort to revise by the Knowledge Network for Biocomplexity Project (KNB). To promote EML using, NCEAS started to develop the modular Metacat framework (short for "Metadata Catalog"). The system incorporated RDF-like (Resource Description Framework) methods for packaging data sets to allow researchers to customize and revise their metadata (Jones *et al*., 2001). Simultaneously, in summer 1999, the LTER information management committee of the United States evaluated the status of metadata within the LTER network in light of a series of long-term goals for the future of informatics in ecology (McCartney and Jones, 2002). The US LTER found that use of EML needed to standardize in both content and presentation format and a machine-readable form. Therefore, revision of EML was independently funded to begin the process.

Although EML revising work is still proceeding, the consensus on using EML as the standard within ecology community has formed. According to EML survey response summary report of US LTER (Servilla, 2005), 19 of the 26 LTER Network research sites responded to the use of EML. Several actions such as tools and standardization were proposed to take an action plan. Outside the United States, several LTER Network like Brazil, Costa Rica, Taiwan,Japan, Korea, Australia, and Malaysia reported that they had used EML to help managing ecological data (Kim *et al*., 2018; Vanderbilt *et al*., 2010; 2015).There are

several guiding principles that have been followed in the development of EML. First, EML needs to allow scientists to provide standardized descriptions of data from the various disciplines in an openly accessible text format. Second, EML should be encoded in a machine-readable format, such as eXtensible Markup Language (XML), which has strong industry support and is independent of particular platforms or softwares. Third, EML should be compatible with existing metadata standards. Finally, EML standard should serve to integrate, rather than dictate metadata only. Based on these principles, EML was designed to be (Jones, 2003):

1. Open to allow human readability and facilitate long-term data archiving

2. Modular to promote re-use of metadata sections and structures

3. Extensible to allow additional metadata that is not part of EML to be included

4. Structured to allow machine processing for analytical applications and other software applications

5. Easy to implement by minimizing required metadata

In order to be machine-parseable, EML chooses XML for its encoding format. XML is a text format language for marking up data and documents. It is similar to HyperText Markup Language (HTML). However, unlike HTML, which is just for displaying style, XML is designed for tagging the content of a document with a meaning for validating that content against a formal schema (McCartney & Jones, 2002). By adopting XML format, EML is implemented as a series of XML document types that can be used in a modular and extensible manner to document ecological data. The architecture of EML was designed based on previous work in other metadata standards. EML adopted the strengths of them, but also focused on improving the automated processing and integration of dataset resources. Currently, EML has updated to version 2.2.0. With 10 major features as following (Jones *et al*., 2019);

- Semantic Annotations
- Structured Funding Info
- Structured license info
- Fields for data papers
- Markdown support in text blocks
- Bibtex citation support
- New fields for literature cited, and reference publications, and usage citations
- Added identifiers for taxonClassifications
- EML namespace changed to use https
- New validation rules and reference implementation

EML Based Information Management Systems

EML documents vary in their levels of content. To

evaluate the implementation of EML, five levels are used to determine completeness (Chapman, 2004; Cook *et al.*, 2001). Each level adds more elements from the EML schema to provide a more-comprehensive description of the data resources documented by the metadata, and thereby supports higher functionality. Level 1 contains of only minimum content for identifying a dataset. Level 2 adds temporal, geographic, and taxonomic coverage for dataset discovery. Level 3 adds dataset details that enable end-user evaluation of the methodology and data entities. Level 4 describes data access to allow automated data retrieval. Level 5 includes complete attribute and quality control descriptions of the raw data to support computer-assisted data integration and re-sampling.

To help aid in managing ecological data for use by future researchers, a number of information systems have been established. These systems have as their goal of long term managing data and assuring their usability over a 20 to 100-year timeframe. Some of the archives are associated with specific research projects, and others focus on specific research topics and types of data. Researchers can contribute their data and metadata to these systems, and the archives curate the data, assuring that adequate backup copies are maintained, and that data formats are updated as needed, so that the data continues to be usable in new versions of software. They also provide search mechanisms or data catalogs, so that other researchers can discover data in the archive. Additionally some systems serve as clearinghouses. These do not curate data themselves, but instead provide search capabilities for locating data from many different archives, or even data made available by researchers themselves on their individual web sites.

Data Observation Network for Earth (DataONE) is the first EML based information management system derived from The Knowledge Network for Biocomplexity (KNB) project. It is a platform for environmental and ecological science, to provide access to Earth observational data (Allard, 2012; Michener *et al.*, 2011; 2012). DataONE is supported by funding from the US National Science Foundation as one of the initial DataNet programs in 2009. Funding was renewed in 2014 through April 2015. DataONE helps preserve, access, use, and reuse of multi-discipline scientific data through the construction of primary cyberinfrastructure and an education and outreach program. DataONE provides scientific data archiving for ecological and environmental data produced by scientists. DataONE's goal is to preserve and provide access to multi-scale, multi-discipline, and multi-national data. Users include scientists, ecosystem managers, policy makers, students, educators, librarians, and the public.

DataONE links together existing cyberinfrastructure to provide a distributed framework, management, and technologies that enable long-term preservation of multi-scale, multi-discipline, and multi-national observational data. The distributed framework is composed of Coordinating Nodes located at the Oak Ridge Campus at Tennessee, University of California Santa Barbara, and University of New Mexico, and member nodes. DataONE also provides resources including tools for accessing and using it.

LTER Europe, US LTER and Taiwan Forestry Research Institute (TFRI) , home of TERN, have become member nodes. Data can be contributed directly to one of the DataONE Member Nodes by ILTER data providers, and then the data will be discoverable through the DataONE web portal. DataONE will maintain a persistent archive of the data, so that ILTER networks need not have this responsibility (Vanderbilt *et al.*, 2015). This strategy obviates the need for countries to operate their own Metacat, which has been a barrier for some to publicly providing their data.

The Environmental Data Initiative (EDI) is another EML based information management system (Gries *et al.*, 2019). It began in the summer of 2016 as collaboration between two US National Science Foundation (NSF) grants, one awarded to the University of Wisconsin (UW) named NIMO and the other to the University of New Mexico (UNM) for PASTA+ (together, they are known as EDI). Both groups originate from the Long Term Ecological Research (LTER) Network and consist of highly motivated and experienced data practitioners, software developers, and research scientists. In addition to the LTER Network, EDI now supports a broad community of environmental and ecological scientists funded through the Long Term Research in Environmental Biology (LTREB), the Organization of Biological Field Stations (OBFS), and the Macrosystem Biology (MSB) programs at NSF. The goal of the LTER focused NIMO (National Information Management Office) project was to expand and enhance the support of informatics in the LTER program, while the goal of PASTA+ (Provenance Aware Synthesis Tracking Architecture – Plus) was to provide an open access data repository that was built using the PASTA software stack for communities other than LTER. To be more inclusive of all served communities, both goals are now part of EDI's vision. As such, EDI is a combination of informatics expertise and a production-level data repository for use by all four communities (and others). EDI also works closely with the LTER National Communications Office (NCO) and DataONE to promote data management best practices and stewardship, and supports two separate DataONE member nodes, one for LTER and the other for all non-LTER data (the EDI Member Node).

No matter DataONE or EDI are developed in the western ecological communities. In the eastern ecological communities, especially Eastern Asian ecological communities also worked together to develop a regional EML

based information management system (Kim *et al*., 2018; Lin *et al*., 2006; Lin *et al*., 2008a). The work started in 2004 under the support of Eastern Asian and Pacific International Long Term Ecological Network (EAP-ILTER) (Kim *et al*., 2018). The system is divided into three tiers. The first tier deals with datasets and related information. Data produced by automated sensors communicating through wired or wireless networks, or collected manually by scientists, are managed by this tier. In addition, all information related to a dataset is also edited in this tier. The second tier relates to information management. Once datasets and other related information have been described, they are stored in a schema-independent database. The third tier consists of the full web-based interfaces that allow easy access to the second tier. This tier also manages definitions of multiple user categories with different user rights.

The first step is to adopt the systems developed by the National Center for Ecological Analysis and Synthesis (NCEAS) at the University of California, Santa Barbara which include Morpho, Metacat (short for "metadata catalog"), and the EML2R from the Processing Techniques for Automated Harmonization (PTAH) project at the University of Virginia. Jointly these provide the tools for creating, editing, storing, retrieving, and using EML documents (Higgins *et al*., 2002; Lin *et al*., 2008b). Then Morpho and Metacat were modified to resolve language coding issues in Asian countries including China, Japan, Korea, Malaysia, Mongolia, the Philippines, Taiwan, and Thailand. Finally, the system has been tested through workshops held domestically and internationally since 2005.

The system includes an EML document database module, a data analysis function module, and a collection of EML documents. A Metacat server has been set up for the Taiwan contains the modules of the EML document database. The EML document database module is a Java servlet that acts as the interface to any SQL-compliant relational database (Jones *et al*., 2001). It handles storage, replication, query, validation, transformation, and authentication of EML documents and user management for researchers from Taiwan. Furthermore, by pointing directly to the referent locations in the database, the raw data can be stored with the EML documents.

The data analysis function module consists of "stylesheets" that translate EML documents into statistical programs. Building on the PTAH project, a web-based interface for creating "R" programs ("R" is a language and environment for statistical computing and graphics) was developed (Lin *et al*., 2008b). The interface altered the original UNIX-based PTAH engine so that it would work on a PC-based system. It not only extends the capabilities of the transformation but also has become a prototype of a server-side system that allows researchers to access

EML, upload data, and then run "R" code on the server. Since "R" provides a wide variety of statistical and graphical techniques and is highly extensible, researchers can use it for data manipulation, calculation, and graphical display online without the need to have their own copy of "R" locally.

Recently, the National Institute of Ecology (NIE) in Korea is developing a new EML based information management system called Ecological Information Bank (Eco-Bank) (Sung *et al*., 2018). EcoBank aims to develop into an online hub providing ecological information across the global that Enables such information to be accessible, usable, and publishable, by anyone in a free manner. Eco-Bank is also dedicated to involving into an international source that helps reward leading opportunities to global collaborative research within the ecological information communities. Currently, EcoBank is forming a new paradigm of the ecological information system which is different from the exiting systems in EAP. The system has a full version in Korean language and moving to an international version of the system. Testing has been among Korea, Taiwan, Thailand, Vietnam and Australia. It plans to fully set up in the near future and expanding to a global usage.

Ecological Data Sharing

As ecological research becomes increasingly multidisciplinary way, it means that the research is a data-intensive, and multifaceted approach. Therefore, the need of sharing data is manifest since no individual scientist, or even small group of scientists, can collect all the data that are needed to address the major ecological research questions (Porter, 2010). Sharing data that support publications facilities the scientific ideals of replication, building on previous work and syntheses is an obvious benefit (Parr and Cummings, 2005). Why is data sharing not yet common practice? Many reasons found. One of them has been recognized that the logistical barriers to data sharing exit (Parr & Cummings, 2005). Fortunately, not only national investment but also technology developed is benefited to the movement of data sharing. For example, Linked Data approach (Berners-Lee, 2006) which is a style of publishing and interlinking structured data on the web provides a new way to dissemination of scientific data for sharing and reuse (Bizer *et al*., 2009).

An example from Taiwan to report and discuss the use of Linked Data approach on the exiting relational databases on forest fire, plant specimen, insect collection, forest dynamics plot census, and Taiwanese species checklist (Mai *et al*., 2011). The application adopts Linked Data approach to connect together data intrinsically related from distributed divisions. The approach develops a workflow through 4 steps to integrate and publish human and ma-

chine readable ecological data as the Linked Open Data on the web. The example concludes that Linked Data approach is a new way to improve and advance ecological data sharing.

For science in general, data sharing is a new concept in today (Reichhardt, 1999). It is a good thing. Data sharing has been promoted in several fields, particularly genomics, that have had a great success under the development of bioinforamtics. The genomics field has founded that growth on access to shared data. Genbank and other genome databases has lead the genomic scientist to to conduct cutting-edge research, and most of their time would be spent duplicating data already collected by others (Benson et al., 2005). Some advantages of sharing data are:

• Elimination of duplication of effort. Researchers can reuse existing data, rather than collecting new, but duplicative, data.
• Improvement in data quality. When data is analyzed by multiple researchers, problems or inconsistencies within the data are more likely to be detected and fixed.
• New science. Combining data in new and interesting ways allows researchers to answer scientific questions at larger spatial and temporal scales.

Why data sharing is so important? It is because the new types of scientific inquiry made possible have to Integrating data from a large number of sites allows researchers to reach more general conclusions and observe larger-scale patterns (Borregaard & Hart, 2016). Meanwhile, combining data from past studies with ongoing studies contributes a historical perspective that may explain aspects of the new data that would otherwise be hidden.

However, the issue of "data ethics" is concerned on data shaing (Porter & Lin, 2017). Without ethical principles for sharing data, few researchers would be willing to share data. The unethical approach of "if you give me your data, I will publish it" has all the ethical appeal of "if you give me your money, I will spend it." Instead, when using data generated by others, we need to adhere to ethical principles that are similar to those used when dealing with published material. All students know that it is plagiarism to copy, without proper acknowledgement, text and ideas from published materials. Similarly, it is unethical to take someone else's data and present it as your own. Most journals now have mechanisms for directly citing data stored in repositories, and if not, data sources can still be acknowledged in the text of an article or in an acknowledgements section.

To help data collectors, users and repository operators adhere to ethical principles, most data repositories or archives have data access policies. These policies typically address three groups: data users, data providers and the data repository itself (Michener, 2015). For data users, policies almost universally require the proper acknowledgement of the data collector in publications that use the data. Additionally, policies may dictate additional conditions for the use of the data, perhaps limiting sale of the data, requiring the data user to identify themselves, or requiring that the author of the data be contacted regarding possible co-authorship. For data providers, policies often dictate the form and content of the metadata required.

Ethical use of data collected by others is a requirement for an ecological researcher to maintain their standing in the scientific community. Researchers who fail to abide by data policies and licenses are subject to legal actions, sanctions by funding sources and, more importantly, loss of their reputation in the scientific community. Fortunately, most researchers are more than happy to properly acknowledge data collectors and to abide by other elements in data policies and licenses. A survey of LTER sites found that problems regarding violations of data policies occurred in less than 0.1% data downloads (Porter, 2010). Moreover, a researcher who publishes data in a repository or archive can clearly document that they were the original data source, so that conflicting claims of priority can be easily resolved.

Despite major significant obstacles, ILTER information managers have formed grassroots partnerships and collaborated to provide information management training, adopt a common metadata standard, develop information management tools useful throughout the network, and organize scientist/information manager workshops that encourage scientists to share and integrate data (Vanderbilt et al., 2015). Throughout this effort, ILTER has shared lessons learned from the successes of these grassroots international partnerships to inform others who wish to collaborate internationally on projects that depend on data sharing entailing similar management challenges.

Application of Ecoinformatics

Carbon flux data management

There is no universally accepted method of carbon flux data management system which uses a metadata approach has yet been established for data archiving, curation, discovery, retrieval, and calculation. Instead, each flux research group has formed their own regional network such as CarboEurope, AmeriFlux, and AsiaFlux and each has developed software to address data management issues. Since 2004, Taiwan Ecological Research Network (TERN) has attempted to collect exiting EML-based tools to assemble them as a data management system that could be used universally in carbon flux research (Lin et al., 2009a; Lin and Hsia, 2010).
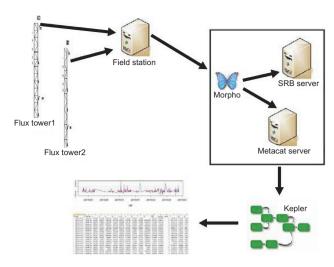
Using this EML-based data management system, a

Fig. 1. Using EML-based tools for carbon flux data management (Lin & Hsia, 2010).

conceptual framework has been developed for flux data management that can be divided into three tiers (Fig. 1). The first tier deals with datasets and related information. Data produced by eddy covariance sensors communicating automatically through wired or wireless networks are managed by this tier. In this first tier, all information related to a flux dataset is documented in EML using the Morpho EML editor. The second tier relates to information management. Once metadata and data quality have been described and checked, they are stored in a schema-independent database called Metacat and Storage Resource Broker (SRB) released by San Diego Supercomputer Center (SDSC). The third tier consists of web service based scientific workflows that allow easy access to the second tier. The Kepler workflow system was adapted for use in this layer.

This data management model has been applied in Chi-lan, a TERN site where two flux towers have been set up since 2000. The two towers equipped with vertical and horizontal wind vectors and CO_2 mixing ratio at 10-20 Hz are measured with a sonic anemometer mounted 5 m above the forest canopy, beside an intake port from which air is pumped to a closed-path infrared gas analyzer. A desktop computer collects these data. Every 30 minutes, the computer stores the raw data which is downloaded weekly and loaded to a SRB server to be retrieved to calculate from the lab. Metadata of these raw data are created and stored in the Metacat. Then, using the Kepler system, five workflows are run that search data from the Metacat, download data from the SRB, rotate data coordinates, QA/QC the data, and Web-Pearman-Leuning (WPL) are created to standardize the flux data calculation process based on each 30 minutes data collected.

Output of the final calculation of all flux data are dis-

played in a text file which reports all the variables and a graphical file which shows the flux trend of a specific period. These secondary data can be saved locally or remotely.

The adaptation of the existing tools based on EML from flux data management experiment has achieved the goal that analyses of sequential ecological data be accompanied by formal process metadata.

Automated analysis of senor data

The development of senor networks allows sensors to gather data in the field and deliver them to the laboratory automatically. Already a variety of ecological important data such as generic meteorological measurements, soil and water temperatures and acoustical records are being collected by senor networks (Porter and Lin, 2013). Examples like high-frequency observations of aquatic systems, intensive and extensive sampling of watershed ecosystems, and unobtrusive observation of animal behavior have showed advantages of using senor networks, especially using wireless senor networks (Porter et al., 2005; 2012). However, the shift toward this data collection paradigm in turn creates new challenges for data management including documentation, quality assurance, discovery, and analysis. In order to be useful, data must provide relevant and reliable information for scientific queries, which usually means that it is documented and has undergone quality assurance checks. Data of unknown quality are essentially useless, and because inaccurate, low quality data can potentially bias results and lead to erroneous conclusions.

Based on the Processing Techniques for Automated Harmonization (PTAH) project of University of Virginia, a web-based interface for checking EML defined datasets has been developed and used since 2006 (Lin et al., 2008b). It not only has the capabilities for transforming EML documents into working statistical programs, but also is a prototype of a server-side system that allows researchers to access EML, upload data, and then check data types and ranges based on the specifications defined in EML metadata (Lin et al., 2007). The interface provides functions to let users correct the errors online and save the corrected data. Researchers also can continue to run the "R" statistical program code on the server. Since "R" provides a wide variety of statistical and graphical techniques and is highly extensible, researchers can use it for data quality control, data manipulation, calculation, and graphical display online without the need to have their own copy of "R" locally. The framework of automating analysis system provides a researcher access to tools that aid in the documentation and analysis of ecological data (Fig. 2).
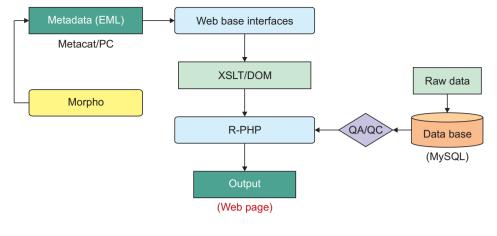
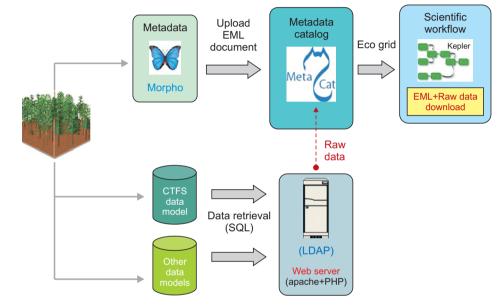**Fig. 2.** The framework of automating analysis system (Lin *et al*., 2008b).

Fig. 3. EML based conceptual model for forest dynamics plot data management (Lin *et al*., 2011).

Forest dynamic plot data management

Several forest dynamics plot research projects in the East-Asia Pacific (EAP) region of the International Long-Term Ecological Research (ILTER) network actively collect long-term data, and some of these large plots are members of the Center for Tropical Forest Science (CTFS) network. The CTFS is a network of forest plots that monitors trees 25-50-ha plots (Ashton *et al*., 1999). CTFS plots involve hundreds of scientists from more than 40 institutions worldwide and share a common methodology as to measurements taken, periodicity of surveys, and identification of tree species (Condit, 1995). There are also many other large forest research plots in the EAP-ILTER region that have comparable data.

In order to facilitate the management of these data, a Forest Dynamics Plot Database and Application Workshop which adapted ecoinformatics approach was held in Taiwan 2009. The results of the workshop produced and tested an integrated information management framework

(Lin *et al*., 2011; Vanderbilt *et al*., 2015). The goal for the framework was to demonstrate how fully documented data archives can be effectively used for data discovery, access, retrieval, analysis, and integration. Results from the work included setting up a database based on the Center for Tropical Forest Science structure on a local relational database (MySQL) server, an authentication interface, a metadata query web page, and 3 workflows to test the framework (Fig. 3).

The case concluded that the framework prototyped based on ecoinformatics approach should be useful to the forest dynamics research community through the establishment of mutualitic relationships between scientists and information managers (Lin *et al*., 2015). Although the functions of this framework have not immediately resolved all metadata and data sharing problems, it provides a collaborative way to link CTFS databases.

Forest healthy data management

Forest healthy management at national and regional scales is important component of providing what society wants and needs from the forest ecosystem. From ecological point of view, a healthy forest nourishes its unique species and processes, while maintaining its basic structures, compositions and functions. From sociological point of view, on the contrary, a healthy forest has an ability to accommodate current and future society needs for values, products and services. Regardless of how different concepts people view these complex forests systems, maintaining the balance between forest sustainability and production of goods and services is the challenge for the forest administrative agency. However, fundamental questions do arise, such as: what is a truly healthy forest and compares to what, by what criteria can we specify a healthy forest, and how do we manage a healthy forest? To answer these questions we need to have multidiscipline scientific data. Without these data, it is difficult for decision makers and forest managers to formulate technically sound policies and address forest health management issues. In Taiwan, relevant data related to forest health are often collected by different scientific groups and consists of a variety of formats and in many geographic locations. To obtain integrated and high quality information to help decision making is often hindered by the lack of standard methodologies for data collection, data management practices, and detailed metadata documentation across groups.

Therefore, building a health forest information management research was initiated in 2004 (Lin *et al*., 2008b; 2009b; Mai *et al*., 2011). The initial goals were: (1) to aggregate the existing dispersed databases including biodiversity, insect and disease, fire, and invasive species, (2) to develop a web-based portal that would streamline the discovery and exploration of forest health information, and (3) to provide data analysis application. Through choosing ecoinformatics approach, a web portal using Java servlet, user authentication, and backend schema-independent metadata repository was designed to be used for a data catalog. A scientific work flow system was recommended for integrating and analyzing data to generate information of forest health management needs. The system includes datasets on biodiversity, insect and disease, fire, and invasive species. They are transformed and reorganized using species name and/or spatial attributes. Those data is archived using EML as metadata standard to combine the raw data stored in repository server called Metacat which is a schema-independent database. Finally, a full web-based interfaces that allow easy access to the second tier. This tier also manages definitions of multiple user categories with different user rights (Fig. 4).

Conclusions

Ecoinformatics development becomes a discipline to foster and change the research on ecology into a data intensive field. Ecological data from data collection to permanent archived through a data life cycle takes into account the ecoinformatics nature of information technology. The valuable content of a ecological repository based on the approach of ecoinformatics techniques has benefited ecological communities. It has formed as a new paradigm of ecology. At the core of ecoinformatics, the information system that combine developments in information technology and ecological theory with applications has shown not only facilitate ecological research but also links ecological entities from organisms to ecosystems with data integration, analysis and synthesis.
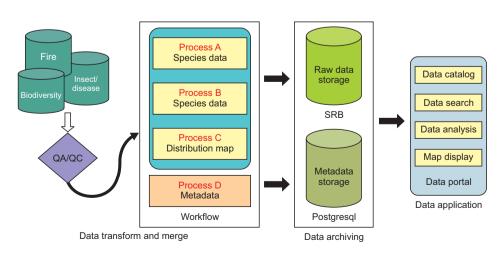
Acknowledgments

Fig. 4. Using EML to integrate traditional databases and applying data archiving and data visualization into a system (Lin *et al*., 2009b).

efforts between US LTER and TERN. The collaboration then expanded to EAP-ILTER. I also thank colleagues of TERN and EAP-ILTER to participate and support the information management activities which have held in China, Taiwan, Korea, Japan, Thailand, Malaysia, Philippines and Vietnam since 2005.

Conflict of Interest

The authors declare that they have no competing interests.

References

Allard, S. (2012). DataONE: facilitating eScience through collaboration. *Journal of eScience Librarianship,* 1, e1004.

Ashton, P.S., Boscolo, M., Liu, J., and LaFrankie, J.V. (1999). A global programme in interdisciplinary forest research: the CTFS perspective *Journal of Tropical Forest Science*, 11, 180-204.

Bearman, D., Rust, G., Weibel, S., Miller, E., Trant, J. (1999). A common model to support interoperable metadata: progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI communities. *D-Lib Magazine*, 5.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2005). GenBank. *Nucleic Acids Research,* 33, D34-D38.

Berners-Lee, T. (2006). Linked Data- design issues. Retrieved October 17, 2020 from https://www.w3.org/DesignIssues/LinkedData.html.

Berthold, P.E., Bauer, H.G., and Westhead, V. (2001). *Bird Migration*: *A General Survey*, 2nd ed. Oxford: Oxford University Press.

Bizer, C., Heath, T., and Berners-Lee, Tim. (2009). Linked Data- the story so far. *International Journal on Semantic Web and Information Systems,* 5, 1-22.

Borregaard, M.K., and Hart, E.M. (2016). Towards a more reproducible ecology. *Ecography,* 39, 349-353.

Brunt, J.W., McCartney, P., Baker, K., and Stafford, S.G. (2002, July 14-18). The future of ecoinformatics in long term ecological research. In N. Callaos, J. Porter, & N. Rishe (Eds.), *The 6th World Multiconference on Systemics, Cybernetics and Informatics* (pp. 367-372). International Institute of Informatics and Systemics.

Chapman, R.F. (2004). EML Best Practices Document 2004, 3rd ed. New York: Elsevier.

Condit, R. (1995). Research in large, long-term tropical forest plots. *Trends in Ecology and Evolution,* 10, 18-22.

Cook, R.B., Olson, R.J., Kanciruk, P., and Hook, L.A. (2001). Best practices for preparing ecological data sets to share and archive. *Bulletin of the Ecological Society of America,* 82, 138-141.

Dengler, J., Ewald, J., Kühn, I., and Peet, R.K. (2011). Ecoinformatics and global change- an overdue liaison. *Journal of Vegetation Science,* 22, 577-581.

Farley, S.S., Dawson, A., Goring, S.J., and Williams, J.W. (2018). Situating ecology as a big-data science: current advances, challenges, and solutions. *BioScience,* 68, 563-576.

Gries, C., Servilla, M., O'Brien, M., Vanderbilt, K., Smith, C.,

Costa, D., *et al*. (2019). Achieving FAIR data principles at the environmental data initiative, the US-LTER data repository. *Biodiversity Information Science and Standards*, 3, e37047.

Hampton, S.E., Jones, M.B., Wasser, L.A., Schildhauer, M.P., Supp, S.R., Brun, J., *et al*. (2017). Skills and knowledge for data-intensive environmental research. *BioScience,* 67, 546-557.

Hampton, S.E., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., *et al*. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment,* 11, 156-162.

Higgins, D., Berkley, C., Jones, M. (2002). *Managing heterogeneous ecological data using Morpho*. Paper presented at Proceedings 14th International Conference on Scientific and Statistical Database Management 24-26 July 2002, Edinburgh, UK.

Iannella, R. (1999). *The future of Internet metadata*. Paper presented at 5th Hong Kong Web Symposium, 4-6 October 1999, Hong Kong, China.

Jones, M.B. (2003). A brief overview of ecological metadata language. *DataBits,* Spring.

Jones, M.B., Berkley, C., Bojilova, J., and Schildhauer, M. (2001). Managing scientific metadata. *IEEE Internet Computing,* 5, 59-68.

Jones, M.B., O'Brien, M., Mecum, B., Boettiger, C., Schildhauer, M., Maier, M., *et al*. (2019). Ecological metadata language version 2.2.0. *KNB Data Repository*, doi: 10.5063/f11834t2.

Jones, M.B., Schildhauer, M.P., Reichman, O.J., and Bowers, S. (2006). The new bioinformatics: integrating ecological data from the gene to the biosphere. *Annual Review of Ecology, Evolution*, *and Systematics*, 37, 519-544.

Kareiva, P. (2001). Ecoinformatics: facilitating access to existing data sets. *Trends in Ecology and Evolution,* 16, 226.

Kim, E., Trisurat, Y., Muraoka, H., Shibata, H., Amoroso, V., Boldgiv, B., *et al*. (2018). The International Long-Term Ecological Research-East Asia-Pacific Regional Network (ILTER-EAP): history, development, and perspectives. *Ecological Research,* 33, 19-34.

Lin, C.C., and Hsia, Y.J. (2010). Carbon flux data computing and sharing using ecoinformatics techniques. *AsiaFlux Newsletter,* 31, 9-11.

Lin, C.C., Hsu, C.H., Hsia, Y.J., Chu, H.S., and Lu, S.S. (2009a). Scientific workflow system (Kepler) for carbon dioxide flux computing. *Quarterly Journal of Chinese Forestry,* 42, 75-88.

Lin, C.C., Jeng, M.R., and Lu, S.S. (2008a). A review of the development and applications of ecoinformatics in Taiwan. *Taiwan Journal of Forest Science,* 23(Supplement), S1-S10.

Lin, C.C., Kassim, A.R., Vanderbilt, K., Henshaw, D., Melendez-Colom, E.C., Porter, J.H., *et al*. (2011). An ecoinformatics application for forest dynamics plot data management and sharing. *Taiwan Journal of Forest Science,* 26, 357-369.

Lin, C.C., Lu, S.S., and Jeng, M.R. (2009b). Building a network-based information system for forest health management: a Taiwanese experience. *Ecological circuits*, 19-20.

Lin, C.C., Porter, J.H., and Hsiao, C.W. (2007). Automating analysis of sensor data using Ecological Metadata Language (EML). Paper presented at Conference on Coastal Environmental Sensor Networks 2007 Apr 12-13 University of Massachusetts, Boston, USA.

Lin, C.C., Porter, J.H., and Lu, S.S. (2006). A metadata-based framework for multilingual ecological information management. *Taiwan Journal of Forest Science, 21*, 377-382.

Lin, C.C., Porter, J.H., Hsiao, C.W., Lu, S.S., and Jeng, M.R. (2008b). Establishing an EML-based data management system for automating analysis of field sensor data. *Taiwan Journal of Forest Science, 23*, 279-285.

Lin, C.C., Wang, Y.H., and Lu, S.S. (2015). Forest dynamics plot database of Taiwan: a content management approach. *Taiwan Journal of Forest Science, 30*, 281-288.

Mai, G.S., Wang, Y.H., Hsia, Y.J., Lu, S.S., and Lin, C.C. (2011). Linked Open Data of Ecology (LODE): a new approach for ecological data sharing. *Taiwan Journal of Forest Science, 26*, 417-424.

McCartney, P., and Jones, M.B. (2002, July 14-18). Using XML-encoded metadata as a basis for advanced information systems for ecological research. In N. Callaos, J. Porter, & N. Rishe (Eds.), *The 6th World Multiconference on Systemics, Cybernetics and Informatics* (pp. 379-384). International Institute of Informatics and Systemics.

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

Michener, W.K. (2015). Ecological data sharing. *Ecological Informatics, 29*, 33-44.

Michener, W.K., Allard, S., Budden, A., Cook, R.B., Douglass, K., and Frame, M. (2012). Participatory design of DataONE- enabling cyberinfrastructure for the biological and environmental sciences. *Ecological Informatics, 11*, 5-15.

Michener, W.K., and Jones, M.B. (2012). Ecoinformatics: supporting ecology as a data-intensive science. *Trends in Ecology and Evolution, 27*, 85-93.

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

Michener, W.K., Porter, J., Servilla, M., and Vanderbilt, K. (2011). Long term ecological research and information management. *Ecological Informatics, 6*, 13-24.

Parr, C.S., and Cummings, M.P. (2005). Data sharing in ecology and evolution. *Trends in Ecology and Evolution, 20*, 362-363.

Popper, K.R. (1959). *The Logic of Scientific Discovery,* 2nd ed. London: Hutchinson.

Porter, J., and Lin, C.C. (2017). Ecoinformatics: a new paradigm of ecology. In Chao JT, King HB (eds) Ecology: concept and application version 7 Chinese translation pp. 555-568. McGraw Hill Edcation.

Porter, J.H. (2000). Maternal control of anterior development in the Drosphila embryo. In W.K. Michener, and J.W., Brunt (Eds.), *Ecological Data*: *Design*, *Management and Processing* (pp. 48-69). London: Blackwell Science.

Porter, J.H. (2010). A brief history of data sharing in the U.S. long term ecological research network. *Bulletin of the Ecological Society of America, 91*, 14-20.

Porter, J.H., and Lin, C.C. (2013). Hybrid networks and ecological sensing. In S.C. Mukhopadhyay, and J.A., Jiang (Eds.), *Wireless Sensor Networks and Ecological Monitoring* (pp. 99-124). Heidelberg: Springer.

Porter, J.H., Arzberger, P., Braun, H.W., Bryant, P., Gage, S., and Hansen, T., *et al*. (2005). Wireless sensor networks for ecology. *BioScience, 55*, 561-572.

Porter, J.H., Hanson, P.C., and Lin, C.C. (2012). Staying afloat in the sensor data deluge. Trends in Ecology and Evolution, 27, 121-129.

Recknagel, F. (2003). *Ecological Informatics Understanding Ecology by Biologically-Inspired Computation*, Heidelberg: Springer.

Recknagel, F. (2008). Ecological informatics: overview. In S.E. Jorgensen, and B.D., Fath (Eds.), *Encyclopedia of Ecology* (pp. 1041-1058). Oxford: Elsevier.

Recknagel, F., and Michener, W.K. (2018). Ecological informatics: an introduction. In F. Recknagel, and W.K., Michener (Eds.), Ecological Informatics (pp. 3-12). New York: Springer International Publishing.

Recknagel, F., and Michener, W.K. (2018). Ecological Informatics: *Data Management and Knowledge Discovery*, 3rd ed. Cham: Springer International Publishing.

Reichhardt, T. (1999). It's sink or swim as a tidal wave of data approaches. *Nature, 399*, 517-520.

Servilla, M. (2005). EML survey response summary report. *DataBits*, Spring.

Sung, S., Kwon, Y.S., and Kim, K.D. (2018). Development and applications of ecological data portal service (EcoBank) for sharing ecological information of Korea. *Korean Journal of Ecology and Environment, 51*, 212-220.

The Committee on European Computing Education (CECE). (2017). *Informatics Education in Europe*: *Are We all in the Same Boat*? New York: Association for Computing Machinery.

Vanderbilt, K.L., Blankman, D., Guo, X., He, H., Lin, C.C., Lu, S.S., *et al*. (2010). A multilingual metadata catalog for the ILTER: issues and approaches. *Ecological Informatics, 5*, 187-193.

Vanderbilt, K.L., Lin, C.C., Lu, S.S., Kassim, A.R., He, H., Guo, X., *et al*. (2015). Fostering ecological data sharing: collaborations in the International Long Term Ecological Research Network. *Ecosphere*, 6, 204.