Fostering ecological data sharing: collaborations in the International Long Term Ecological Research Network

KRISTIN L. VANDERBILT,† CHAU-CHIN LIN, SHENG-SHAN LU, ABD RAHMAN KASSIM, HONGLIN HE, XUEBING GUO, INIGO SAN GIL, DAVID BLANKMAN, AND JOHN H. PORTER

1Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131 USA
2Forest Protection Division, Taiwan Forestry Research Institute, 53 Nanhai Road, Taipei 10666 Taiwan
3Forest Technology Service Division, Taiwan Forestry Research Institute, 53 Nanhai Road, Taipei 10666 Taiwan
4Forestry and Conservation Division, Forest Research Institute Malaysia, 52109 Kepong, Selangor Darul Ehsan, Malaysia
5Synthesis Center of CERN, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101 China
6Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131 USA
7Israel LTER, Jerusalem, Israel
8Department of Environmental Sciences, University of Virginia, Charlottesville, Virginia 22904 USA

Citation: Vanderbilt, K. L., C.-C. Lin, S.-S. Lu, A. R. Kassim, H. He, X. Guo, I. San Gil, D. Blankman, and J. H. Porter. 2015. Fostering ecological data sharing: collaborations in the International Long Term Ecological Research Network. Ecosphere 6(10):204. http://dx.doi.org/10.1890/ES14-00281.1

Abstract. The International Long Term Ecological Research (ILTER) Network was established in 1993 and is now composed of thirty-eight national networks representing a diversity of ecosystems around the globe. Data generated by the ILTER Network are valuable for scientists addressing broad spatial and temporal scale research questions, but only if these data can be easily discovered, accessed, and understood. Challenges to publishing ILTER data have included unequal distribution among networks of information management expertise, user-friendly tools, and resources. Language and translation have also been issues. Despite these 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. Throughout this article, we share 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.

Key words: collaboration; culture; data sharing; ecological networks; ILTER; information management; LTER; long term ecological research; metadata; multilingual; Special Feature: International LTER.

INTRODUCTION

The demand for access to diverse types of data is increasing within the ecological research community. Ecological patterns and processes, including those that drive global change, interact at many spatial and temporal scales (Levin 1992). Scientists seeking to address regional or global
scale research questions must often do so by integrating data collected at finer scales, from more than one discipline (Jones et al. 2006), and from several locations. Many research areas will benefit from increased availability of a diverse array of biophysical and socioeconomic data from around the globe. These areas include the spread of disease and invasive species (Crowl et al. 2008), loss of biodiversity (Willig et al. 2003), fire (Bowman et al. 2009) and dust (Field et al. 2010) impacts on ecosystems, global nitrogen cycling (Galloway et al. 2004), and ecosystem services (Carpenter et al. 2009), to name a few.

Scientists need to be able to discover, access, and understand data from many sites around the world in order to address critical broad-scale research questions. To meet this need, several national and international research networks have been created to provide data at a range of spatial and temporal scales (Peters et al. 2014). One such national network, the US Long Term Ecological Research (US LTER) Network, was established in 1980 to collect long-term data on ecological processes across a diversity of ecosystems and to publicly share that data to encourage cross-site analyses. The US LTER Network has 25 sites including one agricultural site, two urban sites and 22 sites located in relatively undisturbed ecosystems ranging from tundra to tropical forest (Hobbie et al. 2003). Recent data syntheses across multiple US LTER sites, as well as other North American sites (Peters et al. 2011, Jones et al. 2012, Hallett et al. 2014), have demonstrated the power of integrating long-term site-based data from many ecosystems. These cross-site projects could only be conducted because US LTER Network data are well-documented, discoverable, and publicly accessible.

To capture a greater extent of important environmental and socioeconomic gradients, US LTER scientists recognized the need for long-term data from ecosystems outside of the United States. During the 1990s, US scientists made visits to countries around the world, and the model of site-based, long-term research was enthusiastically embraced. In 1993, sixteen national LTER networks, including the US LTER Network, formed the loosely affiliated International Long Term Ecological Research (ILTER) Network (French and Shidong 1994). The ILTER Network is a “network of networks.” Currently, thirty-eight countries have established LTER or, for those emphasizing the human dimension of ecology, LTSER (Long Term Socioecological Research) (Mirtl et al. 2013) networks. ILTER national networks are organized into four regional networks to facilitate study of regional environmental concerns. The regions are (1) East Asia Pacific (EAP ILTER), (2) Europe (ILTER Europe), (3) Americas, and (4) Southern Africa. The ILTER Network (http://www.ilternet.edu) offers a wealth of ecological and socioeconomic settings for cross-site research at its 633 sites (Fig. 1). A central goal of the ILTER is to publicly share long-term ecological data about all the sites in the network to support research that will help solve international ecological and socioeconomic problems (Kaufman and Anderson 2006, Vihervaara et al. 2013).

In contrast to the US LTER Network, which has a centralized office to develop, support, and house its own data repository, the ILTER Network itself has few resources to support information management. Funding levels within ILTER member networks range widely, and the ability to manage data is highly variable. Nevertheless, despite the lack of an ILTER secretariat or a budget, ILTER information managers have made significant advances toward the ILTER data sharing goal through several grassroots projects. In this article we discuss the twenty-two-year evolution of the ILTER information management community in order to share what has made this international collaborative group successful. The lessons learned will inform: (1) information managers who address technical challenges to making data from heterogeneous sources discoverable and accessible, (2) scientists wishing to participate in international collaborative data syntheses, and (3) funding agencies seeking to optimize return on investment when supporting international collaborations.

Early information management outreach: training the trainers

Few of the ILTER member networks that formed in the 1990s had experience with information management. To educate ILTER personnel about information management, US LTER information managers offered workshops during
the 1990s. These workshops taught the basic tools and techniques of LTER data management (e.g., creating metadata, relational database management, quality control, and data archiving). Many workshops were one-off; that is, participants met for three to five days and there were no follow-up activities. The US LTER Network also offered “teach the teachers” workshops. In 1993, for instance, five Chinese Ecosystem Research Network (CERN) information managers attended a twelve-week workshop at the Sevilleta LTER site in New Mexico. The five participants returned to China and taught what they had learned to their colleagues. Other workshops from this period of outreach are listed in Table 1.

As ILTER member networks around the world built information management capacity, local information managers replaced visiting US LTER information managers as trainers. For instance, the Taiwan Ecological Research Network (TERN), which invested significantly in information management training in the US for its own personnel, has in turn led many workshops throughout the EAP ILTER region. Today, information managers trained by the Taiwanese are also trainers. At one workshop offered in the Philippines in June 2014, seventy participants were taught by instructors from ILTER networks in Taiwan, Malaysia, Japan and Israel.

The 1990s one-off style of US LTER-supported information management workshops was successful in terms of information transfer that benefited the individual participants, and sometimes colleagues whom they later instructed. No sustained interactions were expected, however, and there was little follow-up between participants and instructors. Without resources to start an information management program in most countries, many participants did not end up channeling their new knowledge directly into managing data. The lesson learned from this training strategy was that these one-off workshops were useful for raising awareness of the importance of data sharing in the ILTER Network and for providing insights into the levels of resources and commitment that are required for information management. They fell short, however, in establishing sustained interactions among ILTER information managers or triggering many information management activities in the national networks that participated.
Ninety day working visits
with US LTER information managers:
from training opportunity to collaboration

When TERN became interested in enhancing its information management capacity, a more grassroots, time-intensive, and ultimately more productive training mechanism was introduced. Multiple ninety-day visits by TERN IM personnel to Virginia Coast Reserve (VCR) and North Temperate Lakes (NTL) US LTER sites were made (Table 2), where the Taiwanese were mentored by US information managers. TERN information managers gained the knowledge necessary during the first few visits to adopt the same software stack used in the US LTER Network for managing data and metadata (Lin et al. 2006). This software includes Metacat (Berkley et al. 2001), a database for metadata and data, and Morpho, a tool for editing Ecological Metadata Language (EML; Fegraus et al. 2005). EML is the structured metadata standard used in the US LTER Network. TERN has improved on both, modifying Metacat to function with international characters, and translating the Morpho User Manual into Chinese.

This group then taught many workshops throughout the EAP ILTER region, which led to the adoption of this software by others. Metacat is now used in Taiwan, Malaysia, Japan, Thailand, Philippines and South Korea for storing ILTER metadata and data.

Table 2. Information management topics emphasized during 90-day internship-style visits by TERN personnel to US LTER sites.

| Year | Information management topic | Internship location (US LTER site) |
|------|-------------------------------|------------------------------------|
| 2004 | EML                           | North Temperate Lakes Virginia Coast Reserve |
| 2005 | EML and information management system architectures | Virginia Coast Reserve |
| 2005 | Wireless sensor networks      | Virginia Coast Reserve |
| 2006 | Metadata tool development (EML and data quality checker; EML to R web interface) | Virginia Coast Reserve |
| 2007 | Planning for EML implementation at TERN and manuscript development | Virginia Coast Reserve |
As TERN’s knowledge and expertise in information management matured, US LTER and TERN information managers became partners in collaboration. The adoption of the EML metadata standard by TERN and others in the EAP ILTER region created opportunities to generate tools that could be used throughout the ILTER Network by anyone also using EML. Quality assessment tools were developed based on EML that plot the locations of data points to see if they are reasonable, test for out-of-range values, and test for congruency between attribute descriptions in the EML document and attributes in the data table (Lin et al. 2008b). Another tool developed via this collaboration is a web-based interface for automatically creating R scripts (R Development Core Team 2013) based on the information provided in the associated EML document (Lin et al. 2008a; Fig. 2). Researchers can use the R code to upload data, check data quality, and conduct other analyses as desired. Other collaborative work by US LTER and TERN information managers resulted in four publications about sensor networks (Porter et al. 2005, 2010, 2012, Porter and Lin 2013).

The success of this US LTER-TERN information management collaboration is due to the long-term commitment of the core group of TERN and US LTER participants. This group has sustained a vision of how information management can enable ILTER science and has guided collaborators toward this goal. They have grown the partnership to include others from throughout the ILTER Network. Several US information managers have participated in workshops (discussed below) sponsored by TERN, the Malaysia LTER Network (MyLTER), CERN, and the Korea LTER (KLTER) Network. Participants in recent US LTER/EAP ILTER initiatives have included graduate students from the US and Asia (Vanderbilt and Porter 2010). Organizing video conferences for participants in many time zones and face-to-face meetings with scanty funding is challenging, but the strong personal commitment the core collaborators have to each other and to the ILTER information management vision has propelled the group forward for a decade.

The commitment made by all members of the collaboration is demonstrated by the way expenses for workshops have been shared. The local workshop host, whether it was TERN, CERN, Forest Research Institute Malaysia (FRIM, home of the MyLTER Network), or KLTER Network, has covered local costs. Participants obtained small travel grants from their own institute or funding agency. In some instances, one country was able to obtain funding to support scientists from another country who would not otherwise have been able to participate. This model for sharing expenses has allowed personnel from both more and less affluent ILTER member networks to take part. In an era where funding for science is not expanding, this grassroots approach to supporting face-to-face collaboration is a good strategy for fostering international interactions.

There are several lessons to take away from this training-turned-collaboration experience. The productive nature of the ninety-day TERN working visits to US LTER sites suggests that they were an efficient way to transfer technical expertise. Further, information management ‘trainees’ can become collaborators as their experience and knowledge matures. This is a sound reason for funding agencies to support technical training of international colleagues. In addition, international collaborators can support face-to-face meetings of participants from many countries, including those with fewer resources, by sharing expenses. Finally, collaborations can be long-lived. There may be a core group of people who have the bigger picture in mind who persist as collaborators and provide leadership toward the envisioned goal. Funding agencies should invest in collaborations with a long-term horizon in mind, in addition to investing for concrete short-term workshop products.

Engaging scientists with information managers for mutual benefit

Information managers throughout the ILTER strive to build tools that enable science, but it is often difficult to get ecological researchers to articulate what their visualization or computational needs are. Research into the dynamics of collaboration between information technology practitioners and scientists has demonstrated that domain scientists and information managers may have difficulty understanding one another. Scientists may be unable to envision how new technologies can help them while information managers may communicate in ways that are too
Fig. 2. Information managers from TERN and the US LTER Network collaborated on tools that parse EML documents and the data tables they describe to produce R code with which to analyze and manipulate the data. An EML document can be ingested (blue ovals) via a web service, from a URL, or pasted into a web form for upload. The EML document describes the structure of the data table (lower white box) and can be machine-parsed to produce R code (upper white box). Attributes in the EML document are parsed (red, bottom white box) to represent columns in the dataset within the R code (red, top white box). The R code can be used for quality assurance purposes to provide statistical summaries and scatterplots of the data, and to generate maps illustrating locations where the data were collected. The R code can also be downloaded for use in further data analysis. R and EML in the figure are abstractions.
technical (Pennington 2011). Recognizing that this communication problem is a barrier to developing information management systems in the ILTER Network, TERN and US LTER Network information managers undertook to learn more about this issue. They brought together scientists and information managers for data synthesis workshops to learn what the scientists needed from the information managers in order to facilitate the data integration process.

The ILTER scientists targeted in these scientist/information manager workshops engage in long-term monitoring of large forest plots (25–50 ha) in the EAP ILTER region. Several of the plots are part of the Center for Tropical Forest Science (CTFS), an international network of forest plots around the world (Ashton et al. 1999). CTFS sites use common tree census protocols and the data are thus suitable for integration across sites in order to detect patterns that might otherwise not be recognizable with fewer datasets. The CTFS forest plot datasets are large and complex, making them a challenge to work with. In the Pasoh, Malaysia plot, for instance, over 400,000 individual trees from 817 species have been monitored since 1987. Datasets from different sites vary significantly in many ways, including whether the trees are identified to genus and species, how codes are used to describe the trunk structure, and even the physical structure of the files. Synthesis of multiple forest plot datasets is a challenge to scientists and was an opportunity for information managers to demonstrate how informatics tools could ease the data integration burden.

In 2009, information managers and forest plot scientists from US LTER, TERN, MyLTER, and Japan LTER Network (JaLTER) met during the “First Forest Dynamic Plot Information Application Workshop” held in Taiwan to identify information management solutions that would help the scientists integrate CTFS datasets. The group first brainstormed about what was possible and what the scientists wanted in the way of tools to expedite their analyses. Scientists desired user-friendly analytical and visualization tools, a system that would hide the complexity of programming from the user. Information managers responded by developing a framework based on the Kepler workflow program (Altintas et al. 2004), which allows users to chain analytical steps together to produce a reusable analytical workflow. Data documented in EML can be automatically imported into Kepler, and then subsequently manipulated, analyzed and visualized with programs such as R. A workflow can be reused multiple times for different forest plots, simplifying the analytical process. Scientists deemed a proof of concept workflow (Fig. 3), produced during the workshop, a promising solution for ensuring that analytical steps were documented and easy to replicate. The success of this workshop has been attributed to the brainstorming period, when information managers and scientists talked creatively about their needs and started to form a mutualistic relationship (Lin et al. 2011).

The follow-up “Second Forest Dynamic Plot Data Application Workshop” was held in Malaysia in 2011, and included scientists and information managers from MyLTER, KLTER, Singapore, Vietnam Biodiversity Center, and the US LTER Network. The objective of this meeting was to attempt comparative ecological analyses of data from CTFS plots in Malaysia, Taiwan, and Puerto Rico while using workflows developed in the previous workshop to make the process more efficient. Scientists felt that, although the R statistical language and Kepler learning curves were steep, workflow reuse ultimately accelerated the data processing steps. Workflows were also perceived to be an effective mechanism for collaboration between workshop members who would ultimately conduct most of their analyses from their home institution, rather than at the workshop. Workflows created in Malaysia, for instance, can be run in the US or Taiwan, allowing scientists to easily replicate the analytical steps of their collaborators in Malaysia. This series of two CTFS-oriented workshops illustrated that information managers and scientists can benefit from working together during the analytical part of the research process, although the international nature of the group sometimes posed challenges.

The “Forest Dynamic Plot” workshops consisted not only of people from two different disciplines, but also from different cultural contexts. On top of the challenges of learning to communicate across disciplines and languages, there were subtler, culture-dependent ways of seeing and understanding the world to negotiate.
As with any newly formed group, it takes some time for the members to get to know each other, to feel comfortable working together, and to understand the capabilities and needs of everyone involved. International gatherings require even longer ramp up times, as cultural differences, such as approaches to information sharing and how relationships of trust are developed (Sloan and Arrison 2011), come into play. The organizers of these two scientist/information manager workshops planned for the group to interact over an extended period, face-to-face and via the internet, so the group could gel and be productive.

Participants in the “Forest Dynamic Plot” workshops encountered an unexpected cross-cultural difference that had to be resolved before significant progress could be made. The language used in the ILTER scientific community is English, but the meaning of the English words ‘data sharing’ meant different things to different people based on the culture of science they belonged to. For some, the concept of ‘data sharing’ was like being asked to share confiden-

Fig. 3. This abstraction of a Kepler workflow illustrates how several analytical steps can be chained together. The process first uploads a CTFS dataset, uses the accompanying EML file to identify columns, and then displays the data. R code is generated to manipulate the dataset, and the data are then piped through other R scripts that summarize tree measurements by family, genus, and species. The workflow can be shared with other researchers, and they can easily run it on their own CTFS data by merely changing the inputs.
tial information. Others felt they weren’t free to ‘share data’, because in their culture someone else has the authority to do that. Still others felt that sharing the data would not benefit them, but would in fact damage their ability to do research in their own institute. A mutual understanding of the meaning of ‘data sharing’ had to be arrived at during the workshops. Through many conversations, emphasizing commitment to the project and to the other collaborators, a trust relationship was established. The definition of ‘data sharing’ came to mean the same thing for all participants: that the data were to be shared for the mutual benefit of the workshop scientists. Once this very important psychological milestone was reached, the participants became more open to giving and receiving technological training, so that the same tools could be used throughout this ILTER research group, and also to accepting critiques of their plot data to improve its quality. This new ‘ILTER collaboration culture’ that developed through face-to-face meetings meant that group members could continue to work together remotely after the workshop and understand each other without having to be in the same room. This intangible culture of trust has sparked international friendships, co-authorship on publications, technology transfer, and the sharing and reuse of data.

**Going global: products from which many ILTER networks can benefit**

Ecological researchers studying large-scale phenomena need to be able to access data from all over the world. This need would be best met by the ILTER Network by providing easy discovery and access to data, preferably from a single portal. If all ILTER member networks were to use the same machine-readable metadata standard, then tools could be built that could search across all ILTER metadata records. By 2005, ILTER data managers were considering how to make this vision a reality in a network with heterogeneous data documentation and archival practices. The US LTER Network and other members of the ILTER Network (South Africa, Brazil, Spain, Taiwan, Japan) were utilizing EML and the Metacat database for managing data. Other ILTER networks had, however, chosen a different path. LTER Europe had invested in an ontology-based system (Mirtl 2010), and CERN was using a metadata standard developed strictly for CERN. This intra-network heterogeneity made it difficult to build a harmonized information management system for the ILTER Network that could provide researchers with a comprehensive catalog of ILTER data. To resolve this issue, ILTER information managers from US LTER, LTER Europe, and the EAP ILTER region met in China in 2008. They discussed whether to choose the ontology or EML path for the whole ILTER. They agreed that all members should supply discovery-level metadata (e.g., title, abstract, keywords, author) in EML in English (Vanderbilt et al. 2010). EML was a mature standard and tools already existed to create EML documents, factors which played heavily into the group’s decision.

The agreement to use EML as the ILTER metadata standard was straightforward, but the implementation was much more complex in light of the multilingual aspects of creating a global information management system. Most metadata are generated in the local language of each country’s LTER, and a way to search metadata in several languages was needed. Representatives from throughout the ILTER convened in Shanghai in 2012 at the CERN-sponsored “ILTER Information Management Workshop on Semantic Approaches to Discovery of Multilingual ILTER Data” to address this issue. They agreed to use a multilingual thesaurus to help resolve semantic ambiguities and to create a prototype tool to query a Metacat in the local language in which it was implemented. The multilingual thesaurus selected is EnvThes (Schantz and Peterseil 2013), which was developed in part by LTER Europe. It contains terms for categorizing LTER monitoring and experiments, including the controlled vocabulary developed by the US LTER Network (Porter 2010b). The prototype tool for searching Metacats in different languages that was developed at the workshop (Fig. 4) holds promise for the ILTER community as well as other global data collections, such as the Global Biodiversity Information Facility (GBIF) (Edwards 2004). Further development of this prototype will lessen the language barrier to data discovery, making ILTER data easier to find.

High performing collaborative research teams, such as the group that met in the 2012 Shanghai workshop, have been shown to have a high
degree of diversity, a multidimensional factor including culture, ethnicity, gender and religion of the participants (Cheruvelil et al. 2014). This workshop capitalized on the unique expertise of participants to produce outcomes that would not have been possible without multinational participation. Translations of keywords for inclusion in the multilingual thesaurus were swiftly made for Japanese, Korean, Traditional Chinese, and Simplified Chinese. A European representative with experience using Resource Description Framework (RDF) tools designed for managing a multilingual vocabulary was readily able to apply these skills to the ILTER thesaurus project. The Asian participants knew how to internationalize software to make it accommodate Asian.
character sets as well as Latin characters. The primary lesson learned in this workshop was that participants have different skill sets based on their international experience, and that these complementary strengths yielded a result that no national team of information managers could have produced alone.

**Leveraging open source software when resources are scarce**

Despite all the above-mentioned collaborations between members of the ILTER information management community, the fact remains that some countries simply do not have the resources to participate in an ILTER information management system. Such networks would benefit from a free, easy-to-implement, web-based metadata editor and data catalog. ILTER information managers have begun to capitalize on open source software, collaborating with a global community of developers, as a means of creating such a tool.

The Drupal Ecological Information Management System (DEIMS) is a web-based framework for storage, documentation, and distribution of data products from an ecological research site (San Gil et al. 2010b). It was developed using the open source Drupal Content Management System (http://www.drupal.org) which is sustained by an international community of thousands of users and developers. DEIMS was inspired by the need of several US LTER sites to have a web-based tool to produce EML to feed in to the US LTER Network’s centralized data catalog. Information managers from several US LTER sites collaborated on this grassroots development project, employing modules created by others in the Drupal community (to manage publications, for instance) and writing custom code for LTER-specific tasks. The first version of DEIMS was deployed in five US LTER sites in 2010 (Gries et al. 2010). The core of DEIMS is an easy-to-use web-based metadata editor and metadata generator that will produce EML, Biological Data Profile (BDP) and ISO metadata (San Gil et al. 2010a; Fig. 5). Other features include a queryable data catalog, and the capacity to create a harvest list that would facilitate the transfer of local data to a centralized data repository.

DEIMS, now in its second version, is free to all and available at https://www.drupal.org/project/deims. It is an attractive option for the ILTER Network because Drupal supports content and user interface translation automatically. Members of the ILTER outside the US LTER have started to use the tool and also contribute new features. LTER Europe has adapted it for managing site-related information for its twenty member countries. TERN has improved on DEIMS by implementing an Apache Solr search engine (http://lucene.apache.org/solr/) that enhances data discoverability (Lin 2013). DEIMS has the potential to become the tool that allows any LTER site to document and distribute their data using a common model. The combination of grassroots innovation and leveraging of existing open source software has yielded this promising “out-of-the-box” information management solution. Through this project, the ILTER information management community learned that they need not develop custom solutions for all ILTER information management needs. There are other groups addressing the same needs, and through grassroots efforts co-development of software useful to a broader community can be achieved. By partnering with others developing open source software, some of the costs of new software development and maintenance can be shared.

**The future**

The ILTER information management community has made steady progress toward goals of improving data accessibility and use, but still have many avenues for collaboration into the future. Possibilities include capitalizing on the Semantic Web, partnering with cyberinfrastructure (CI) initiatives that facilitate data sharing and discovery, managing “big data” (e.g., high frequency and volume streaming sensor data), and training graduate students and scientists to integrate and analyze the wide variety of data now available online. Synergies with other international networks, such as the grassroots Global Lake Ecological Observatory Network (GLEON) (Weathers et al. 2013), will also be explored. Finally, ILTER information managers will continue to promote a culture of data sharing in the ILTER and beyond.

**Semantic web.**—One way to improve ILTER data discovery and sharing may be to define relationships between data on the WWW using
the Linked Data method. With this approach, machine-readable RDF documents specify how one piece of data is related to another piece of data, making the data amenable to automated interpretation by computers (Bizer et al. 2009). Data described in this way become part of the Web of Data, also known as the Semantic Web. This approach has the potential to allow researchers to make connections among datasets that were not previously possible. TERN has already demonstrated how four disconnected fire ecology, herbarium, insect collection, and biodiversity databases can be integrated once all four have been described using the Linked Data method (Mai et al. 2011). The potential of Linked Data to improve discoverability of ILTER data...
Partnering with other CI initiatives.—To create an online ILTER data portal that would combine all datasets from throughout the network, ILTER had at one time envisioned a system of linked, replicating Metacats that would have been maintained by national ILTER member networks or by one member of each regional ILTER network. Recent developments in CI may make the network of Metacats unnecessary. DataONE is a federated network that connects Member Nodes (data repositories) with Coordinating Nodes (data service providers). It offers data discovery, access and preservation services for Earth observational data from around the world (Michener et al. 2011). LTER Europe, US LTER and Taiwan Forestry Research Institute (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. This strategy obviates the need for countries to operate their own Metacat, which has been a barrier for some to publicly providing their data.

Big data.—Data that come in large volumes or with large amounts of variety are referred to as “big data” (Lynch 2008). Most data produced by ILTER qualify as “big data”, especially with respect to the variety of data and ecosystems in which the data are collected (Hampton et al. 2013b). High volume “big data” collected by ILTER ecologists include advanced imaging technologies, such as high-resolution satellite imagery, LiDAR, and unmanned aerial vehicles (UAVs), which can collect imagery at both high spatial and temporal resolution (Anderson and Gaston 2013). Flux towers and other ground-based environmental sensors may operate at high frequencies (>10 Hz), generating data at high velocities and volumes (Porter et al. 2012). The diversity of data collected by ecologists has also increased over the last decade to include genetic sequences, acoustic and video data (Jones et al. 2006). The ever increasing variety and volume of ecological data will continue to supply data management and integration challenges for ILTER informaticians and scientists to tackle together.

ILTER information managers will support expanding the use of big data by ILTER scientists by offering training for graduate students in advanced information integration and analysis techniques. Few courses are available that offer training on topics such as data harmonization across temporal and spatial scales, gap-filling, and how to make analyses reproducible using tools such as R and Kepler. The ILTER information management community is seeking funding to support the first such workshop for graduate students which will be held in the EAP ILTER region. Students will emerge from the three-week course with a toolkit for analyzing complex datasets, as well as valuable international connections.

Cooperating with other networks.—There are other networks that are developing CI and engaging in international research and training with which the ILTER information management community can synergize in the future. The Global Lake Ecological Observatory Network (GLEON), for instance, is a grassroots international network that has a successful program to provide graduate students with education and training opportunities (http://gleon.org/). The ILTER Network may partner with GLEON, or learn by example to expand graduate student information management training opportunities. In addition, some ILTER sites are co-located with sites that are part of new networks of intensely instrumented sites (e.g., National Ecological Observatory Network (NEON) in the US (http://www.neoninc.org/); Korea Ecological Observatory Network (KREON) in South Korea) that will provide standardized environmental data for selected ecosystems. Integration of these new, often large scale, sensor-collected data with long-term, fine-scale ILTER data will offer new opportunities for collaboration between scientists and information managers.

Data sharing.—One challenge for the ILTER community, as for other scientific communities (Nelson 2009), is finding ways to inspire more scientists to share their data. The ILTER Network has a data sharing policy that states that data products collected under the auspices of an ILTER member network will be archived and publicly shared. Yet, while the importance of data sharing is widely recognized, many ecologists (in the ILTER or elsewhere) still do not
share their data. Hampton et al. (2013a) found that 57% of ecological research projects did not share data, and 81% of what were shared were genetic sequence data. In the case of the ILTER Network, some national networks are not sharing data because they do not have the resources to support information management. In other cases, however, data are not shared because the culture of data sharing is still developing.

The US LTER, TERN, MyLTER and CERN networks all have data sharing policies that, with varying degrees of success, encourage scientists to share their data. The US LTER Network has had a network-wide data sharing policy since 1997 (Porter 2010a) stating that data must be shared within two years of collection. Still, some scientists withhold their data, citing concerns about being “scooped”, having their data misinterpreted, and not receiving credit (Costello 2009). US Funding agencies, such as the National Science Foundation which funds the US LTER Network, are now insisting that data collected on projects supported with government funds be published and available for public use (NSF 2010). Some journals, such as Science and Nature, make data deposition in a public repository a condition of publication. With this encouragement, the US LTER data archive is likely to capture more and more of the projects funded by this program.

Like the US LTER, TERN has also promoted data sharing since the early 1990s. By 2004, 60% of TERN scientists acknowledged that data sharing was important, yet few voluntarily submitted their data. To push the issue, the head of the Taiwan Forestry Research Institute (TFRI), where TERN is housed, imposed a data sharing policy on the whole institution. By TFRI policy, all data must be archived at the end of a project. Scientists failing to archive data may experience restricted funding when applying for future grants, a threat which has increased the rate of data submission slightly. For TERN, the barrier to data sharing has been at least partially technical. With no dedicated site information managers (like US LTER sites provide), scientists must archive the data themselves. The availability of easy-to-use tools has simplified the process and made data archival for TERN a less onerous task. Taiwan endorsed the Open Government Data Initiative in 2012 (Yang et al. 2013), as did the US in 2013 (Burwell et al. 2013), acknowledging that data collected by government agencies should be publicly accessible. This move is expected to put additional pressure on TERN scientists to share their data.

Although the Chinese and Malaysian governments have not yet endorsed the Open Government Data Initiative, both the MyLTER Network and CERN are promoting a data sharing culture. CERN, which was established in 1993, has long had a data sharing policy that asks CERN scientists to archive data within two years of collection. Digital data products maintained by the centralized CERN information management system are currently shared among CERN scientists, and hardcopy data from environmental sensors and taxonomic lists are also available. Nevertheless, many CERN scientists still prefer to keep their data private, as they have always done. CERN recognizes that additional pressure will be required to motivate scientists to share their data, and anticipates an increase in data sharing as new funding measures are implemented that require data be shared at the end of a project. The new “National Science and Technology Infrastructure (NSTI)” initiative, which aims to foreground technical and scientific resources, including data, is expected to make data sharing more common. Several CERN sites receive funding from this source.

The Forest Research Institute of Malaysia (FRIM) established a data sharing policy in 2014 for all scientists, many of whom are associated with the MyLTER Network, a new ILTER member as of 2013. Scientists are asked to submit data to the FRIM archive at the end of a project, which typically runs two to three years. Since only a year has passed since the policy went into effect, it is too early to know how well scientists will comply. The policy does not state that noncompliance will be penalized, although this is the spirit in which the policy was implemented.

From these four examples, it is clear that ILTER member networks are at different places with respect to developing a data sharing culture. They all recognize the importance of sharing data, however, and are taking steps to encourage the practice. Three of the four countries are now tying data sharing from today’s research project to future access to more research funds. As
governments, funding agencies, and institutes apply more pressure, data sharing will almost certainly become a more common practice throughout the ILTER Network. To facilitate ILTER data sharing, the ILTER information management community will continue to develop tools that make data management and sharing simple, and that help transform ILTER data into information and knowledge. The group will also promote data sharing by hosting workshops that bring technology specialists together with scientists to tackle data integration challenges and produce synthetic research products.

**CONCLUSION**

The ILTER Network was created in 1993 to foster international research on processes operating at broad spatial and temporal scales, and data sharing in the ILTER Network is thus of paramount importance. Since that time, the ILTER information management community has addressed many challenges related to making ILTER data discoverable, accessible, and well-documented. Among their accomplishments are adoption of an ILTER metadata standard, use of a common set of data management tools in many national networks, and development of prototype software to support multilingual data discovery. During many years of cooperation, the group has learned lessons about what has made this international collaboration successful (Table 3). The collaborations within the ILTER information management community have been at the grassroots level, inspired by information managers who needed tools in their national networks and who adopted a collective willingness to generalize for multinational use. These partnerships grew out of face-to-face meetings where participants recognized that they could accomplish more as an international team than by working independently. By sharing resources and expertise, and because of a strong personal commitment to other team members, these collaborators have advanced toward a shared vision of ILTER information management that supports science. In an environment where resources are limited and research is global in scale, international grassroots collaborations such as these may become an emerging norm in ecology.

**Table 3. Lessons learned about what has made the long-term collaboration of ILTER information managers successful. Relevance to stakeholders other than information managers is noted.**

| Lessons learned | Relevance to specific stakeholders |
|-----------------|-----------------------------------|
| International collaborators bring together different skill sets and these complementary strengths allow multinational teams to produce outcomes that no single network could have produced alone. | Scientists and funding agencies should support international working groups to increase diversity of collaborators. |
| The grassroots workshops of the ILTER information management community have been possible because collaborators have shared expenses by: • the local host picking up participant local costs, and • participants acquiring their own travel funds from their own institute or government. | Funding agencies should recognize that, although not all participants may pay the same amount, this cost sharing mechanism: • demonstrates the commitment of participants to the collaboration, and • keeps costs reasonable for all participants. |
| Collaborations may yield short-term workshop products, or there may be longer term outputs. A core group of collaborators may provide continuity to the collaboration, sustaining it with their commitment, passion and vision and leading the group toward a long-term goal. | Funding agencies should: • plan for short-term product-oriented outcomes, and • plan for longer-term support that allows the core collaborators with the big picture in mind to pursue long-term goals. |
| New international, interdisciplinary collaborations bring together individuals from different domains and with different cultural ways of seeing and interpreting the world. International collaborators will have additional challenges when learning to communicate effectively in order to understand what each participant can contribute. | International collaborators should recognize the longer ramp up time for their groups and build sufficient face-to-face meeting time into proposals. |
| Information managers need not develop every new tool from scratch to meet the needs of their network. They should partner with other cyberinfrastructure initiatives whenever possible, to cross-pollinate ideas and take advantage of the pool of developers in other organizations. | Funding agencies may capitalize on the potential cost-savings for collaborators proposing to utilize open source solutions or partner with other CI developers with similar needs. |
ACKNOWLEDGMENTS

We gratefully acknowledge ILTER hosts in Taiwan, Malaysia, South Korea and China who generously covered local costs so that ILTER information managers could hold face-to-face workshops. We also thank our own institutions or funding agencies for providing travel funds. Supplemental grant funds from the US National Science Foundation to the Virginia Coast Reserve LTER site and the Sevilleta LTER site supported travel by US participants. We also thank an anonymous reviewer whose comments greatly improved this manuscript.

LITERATURE CITED

Altintas, I., C. Berkley, E. Jaeger, M. Jones, B. Luda¨scher, and S. Mock. 2004. Kepler: an extensible system for design and execution of scientific workflows. Pages 423–424 in Proceedings of the 16th Annual Conference on Scientific and Statistical Database Management, Santorini Island, Greece, June 21–23, 2004. IEEE Computer Society, Washington, D.C., USA.

Anderson, K., and K. J. Gaston. 2013. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. Frontiers in Ecology and the Environment 11:138–146.

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

Bailey, R. G., and H. C. Hogg. 1986. A world ecoregions map for resource reporting. Environmental Conservation 13:195–202.

Berkley, C., M. Jones, J. Bojilova, and D. Higgins. 2001. Metacat: a schema-independent XML database system. Pages 171–179 in Proceedings of the 13th International Conference on Scientific and Statistical Database Management, George Mason University, Virginia, July 18–20, 2001. IEEE Computer Society, Washington, D.C., USA.

Bizer, C., T. Heath, and T. Berners-Lee. 2009. Linked data: the story so far. International Journal of Semantic Web and Information Systems 5:1–22.

Bowman, D. M. J. S., et al. 2009. Fire in the Earth system. Science 324:481–484.

Burwell, S., S. VanRoeckel, T. Park, and D. J. Mancini. 2013. OMB memorandum M-13-13, open data policy-managing information as an asset. Executive Office of the President, Office of Management and Budget, Washington, D.C., USA. http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13

Carpenter, S. R., et al. 2009. Science for managing ecosystem services: beyond the Millennium Ecosystem Assessment. Proceedings of the National Academy of Sciences USA 106:1305–1312.

Cheruvil, K. S., P. A. Soranno, K. C. Weathers, P. C. Hanson, S. J. Goring, C. T. Filstrup, and E. K. Read. 2014. Creating and maintaining high-performing collaborative research teams: the importance of diversity and interpersonal skills. Frontiers in Ecology and the Environment 12:31–38.

Costello, M. J. 2009. Motivating online publication of data. BioScience 59:418–427.

Crowl, T. A., T. O. Crist, R. R. Parmenter, G. Belovsky, and A. E. Lugo. 2008. The spread of invasive species and infectious disease as drivers of ecosystem change. Frontiers in Ecology and the Environment 6:238–246.

Edwards, J. L. 2004. Research and societal benefits of the Global Biodiversity Information Facility. BioScience 54:485–486.

Fegraus, E., S. Andelman, M. B. Jones, and M. P. Schildhauer. 2005. Maximizing the value of ecological data with structured metadata: an introduction to the Ecological Metadata Language (EML) and principles for metadata creation. Ecological Society of America Bulletin 86:158–68.

Field, J. P., J. Belnap, D. D. Breshears, J. C. Neff, G. S. Okin, J. J. Whicker, T. H. Painter, S. Ravi, M. C. Reheis, and R. L. Reynolds. 2010. The ecology of dust. Frontiers in Ecology and the Environment 8:423–430.

French, C., and Z. Shidong. 1994. Developing LTER programs worldwide. Pages 12–13 in R. W. Nottrott, J. F. Franklin, and J. R. Vande Castle, editors. Proceedings of an International Summit: International Networking in Long-Term Ecological Research, Estes Park, Colorado, USA, September 23–24, 1993. LTER Network Office, University of Washington, Seattle, Washington, USA.

Galloway, J. N., et al. 2004. Nitrogen cycles: past, present, and future. Biogeochemistry 70:153–226.

Gries, C., I., San Gil, K. Vanderbilt, and H. Garritt.
2010. Drupal developments in the LTER network. http://databits.lternet.edu/spring-2010/drupal-semanticdevelopments-lter-network
Hallett, L. M., J. S. Hsu, E. E. Cleland, S. L. Collins, T. L. Dickson, E. C. Farrer, L. A. Gherardi, K. L. Gross, R. J. Hobbs, L. Turnbull, and K. N. Suding. 2014. Biotic mechanisms of community stability shift along a precipitation gradient. Ecology 95:1693–1700.
Hampton, S. E., C. A. Strasser, and J. J. Tewsksbury. 2013a. Growing pains for ecology in the twenty-first century. BioScience 63:69–71.
Hampton, S. E., C. A. Strasser, J. J. Tewsksbury, W. K. Gram, A. E. Budden, A. L. Batcheller, C. S. Duke, and J. H. Porter. 2013b. Big data and the future for ecology. Frontiers in Ecology and the Environment 11:156–162.
Hobbie, J. E., S. R. Carpenter, N. B. Grimm, J. R. Gosz, and T. R. Seastedt. 2003. The US Long Term Ecological Research Program. BioScience 53:21–32.
Jones, J. A., et al. 2012. Ecosystem processes and human influences regulate streamflow response to climate change at Long-Term Ecological Research sites. BioScience 62:390–404.
Jones, M. B., M. P. Schildhauer, O. J. Reichman, and S. Bowers. 2006. The new bioinformatics: integrating ecological data from the gene to the biosphere. Annual Review of Ecology, Evolution, and Systematics 37:519–44.
Kaufman, H., and M. Anderson. 2006. International Long-Term Ecological Research Network (ILTER) final draft strategic plan. http://www.ilternet.edu/about/key-documents/FINAL-DRAFT-STRATEGIC-PLAN-Aug2006.pdf/view
Levin, S. A. 1992. The problem of pattern and scale in ecology. Ecology 73:1943–1967.
Lin, C.-C. 2013. A tool for many tasks: the drupal ecological information management system. In Proceedings of the International Conference on Open Data in Biodiversity and Ecological Research, Taipei, Taiwan, November 20–22, 2013. Academia Sinica, Taipei, Taiwan.
Lin, C.-C., 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., J. H. Porter, C.-W. Hsiao, S.-S. Lu, and M.-R. Jeng. 2008a. 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., J. H. Porter, and S.-S. Lu. 2006. A metadata-based framework for multilingual ecological information management. Taiwan Journal of Forestry 21:377–382.
Lin, C.-C., J. H. Porter, S.-S. Lu, M.-R. Jeng, and C.-W. Hsiao. 2008b. Using structured metadata to manage forest research information: a new approach. Taiwan Journal of Forest Science 23:133–143.
Lynch, C. 2008. Big data: How do your data grow? Nature 455:28–29.
Mai, G.-S., Y.-H. Wang, Y.-J. Hsia, S.-S. Lu, and C.-C. Lin. 2011. Linked Open Data of Ecology (LODE): a new approach for ecological data sharing. Taiwan Journal of Forest Science: 26:417–424.
Michener, W. K., T. J. Vision, P. Cruse, D. Vieglais, J. Kunze, and G. Janée. 2011. Data-ONE: Data Observation Network for Earth—preserving data and enabling innovation in the biological and environmental sciences. D-Lib Magazine 17. http://dx.doi.org/10.1045/january2011-michener
Mirtl, M. 2010. Introducing the next generation of ecosystem research in Europe: LTER-Europe’s multi-functional and multi-scale approach. Pages 75–94 in F. Müller, C. Baessler, H. Schubert, and S. Klotz, editors. Long-Term Ecological Research: between theory and application. Springer, New York, New York, USA.
Mirtl, M., D. E. Orenstein, M. Wildenberg, J. Peterseil, and M. Frenzel. 2013. Development of LTSER platforms in LTER-Europe: challenges and experiences in implementing place-based long-term socio-ecological research in selected regions. Pages 409–442 in S. J. Singh, H. Haberl, M. Chertow, M. Mirtl, and M. Schmid, editors. Long term socio-ecological research studies in society–nature interactions across spatial and temporal scales. Springer, Dordrecht, The Netherlands.
National Science Foundation. 2010. Press release 10-077. Scientists seeking NSF funding will soon be required to submit data management plans. http://www.nsf.gov/news/news_summ.jsp?cnn_id=116928
Nelson, B. 2009. Data sharing: empty archives. Nature 461:160–163.
Pennington, D. D. 2011. Collaborative, cross-disciplinary learning and co-emergent innovation in eScience teams. Earth Science Informatics 4:55–68.

Peters, D. P. C., H. W. Loescher, M. D. San Clements, and K. Havstad. 2014. Taking the pulse of a continent: expanding site-based research infrastructure for regional- to continental-scale ecology. Ecosphere 5:art29.

Peters, D. P. C., A. E. Lugo, F. S. Chapin, S. T. A. Pickett, M. Duniway, A. V. Rocha, F. J. Swanson, C. Laney, and J. Jones. 2011. Cross-system comparisons elucidate disturbance complexities and generalities. Ecosphere 2:art81.

Porter, J. H. 2010a. 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. 2010b. A controlled vocabulary for LTER datasets. 2010. http://databits.lternet.edu/spring-2010/controlled-vocabulary-lter-datasets

Porter, J. H., et al. 2005. Wireless sensor networks for ecology. BioScience 55:561–572.

Porter, J. H., P. C. Hanson, and C.-C. Lin. 2012. Staying afloat in the sensor data deluge. Trends in Ecology and Evolution 27:122–129.

Porter, J. H., and C.-C. Lin. 2013. Hybrid networks and ecological sensing. Pages 99–124 in S. C. Mukhopadhyay and J. A. Jiang, editors. Wireless sensor networks and ecological monitoring. Springer-Verlag, Berlin, Germany.

Porter, J. H., C.-C. Lin, D. E. Smith, and S.-S. Lu. 2010. Ecological image databases: from the webcam to the researcher. Ecological Informatics 5:51–58.

R Development Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

San Gil, I., V. Hutchinson, M. Frame, and G. Palanisamy. 2010a. Metadata activities in biology. Journal of Library Metadata 10:99–118.

San Gil, I., M. White, E. C. Melendez-Colom, and K. L. Vanderbilt. 2010b. Case studies of ecological integrative information systems: the Luquillo and Sevilleta information management systems. Metadata and Semantic Research 108:18–35.

Schentz, H., and J. Peterseil. 2013. ENVTHES: Environmental Thesaurus Short Description. http://www.enveurope.eu/Project/Detailed_actions/EnvThes_ShortDescription_Rev0_1_20130122.pdf

Sloan, S. S., and T. S. Arrison. 2011. Examining core elements of international research collaboration: summary of a workshop. National Academies Press, Washington, D.C., USA.

Vanderbilt, K. L., D. Blankman, X. Guo, H. He, C.-C. Lin, S.-S. Lu, A. Ogawa, E. O’Tuama, H. Schentz, and W. Su. 2010. A multilingual metadata catalog for the ILTER: issues and approaches. Ecological Informatics 5:187–193.

Vanderbilt, K., and J. Porter. 2010. Synthesizing large datasets. LTER Network News, Fall 2010. http://news.lternet.edu/Article2270.html

Vihervaara, P., et al. 2013. Using long-term ecosystem service and biodiversity data to study the impacts and adaptation options in response to climate change: insights from the global ILTER sites network. Current Opinion in Environmental Sustainability 5:53–66.

Weathers, K. C., et al. 2013. The Global Lake Ecological Observatory Network (GLEON): the evolution of grassroots network science. Limnology and Oceanography Bulletin 22:71–73.

Willig, M. R., D. M. Kaufman, and R. D. Stevens. 2003. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. Annual Review of Ecology, Evolution, and Systematics 34:273–309.

Yang, T.-M., J. Lo, H.-J. Wang, and J. Shiang. 2013. Open data development and value-added government information: case studies of Taiwan e-government. Pages 238–241 in ICEGOV ’13 Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance, Seoul, Republic of Korea, October 22–23, 2013. ACM, New York, New York, USA.