SHORT COMMUNICATION   **Open Access**

CrossMark

A case study of ECN data conversion for Korean and foreign ecological data integration

Hyeonjeong Lee[1], Miyoung Shin[1*] and Ohseok Kwon[2]

Abstract

In recent decades, as it becomes increasingly important to monitor and research long-term ecological changes, worldwide attempts are being conducted to integrate and manage ecological data in a unified framework. Especially domestic ecological data in South Korea should be first standardized based on predefined common protocols for data integration, since they are often scattered over many different systems in various forms. Additionally, foreign ecological data should be converted into a proper unified format to be used along with domestic data for association studies. In this study, our interest is to integrate ECN data with Korean domestic ecological data under our unified framework. For this purpose, we employed our semi-automatic data conversion tool to standardize foreign data and utilized ground beetle (*Carabidae*) datasets collected from 12 different observatory sites of ECN. We believe that our attempt to convert domestic and foreign ecological data into a standardized format in a systematic way will be quite useful for data integration and association analysis in many ecological and environmental studies.

Keywords: Ecological data, Ecological data conversion tool, Data standardization, Data integration

Introduction

In recent decades, it becomes increasingly significant to monitor and research long-term ecological changes (Michener and Jones 2012). Massive volume of ecological data are being collected according to predefined protocols throughout the world and demand new techniques for exploring hidden patterns at large spatial and temporal scales (Kelling et al. 2009). Some observatory networks, such as Long Term Ecological Research network (LTER) (San Gil et al. 2009) and the National Ecological Observatory Network (NEON) (Keller et al. 2008), have been providing platforms to access and share plenty of ecological data in the form of Ecological Metadata Language (EML) (Fegraus et al. 2005).

Despite these worldwide trends and attempts to integrate and manage long-term ecological data in a unified framework, domestic data in South Korea are still mostly scattered over many different systems in various forms, making it difficult for long-term ecological research. For this reason, they should be first standardized in a common form for data integration and further analyses (Bonet et al. 2014). This need has led to the development of the Korean ecological data conversion tool (Lee et al. 2017) which converts raw data into common form based on some protocols. Moreover, foreign ecological data will be used along with domestic data for association studies if they could be converted into a proper unified format.

To find the feasibility of integration of foreign ecological data with Korean domestic data under our unified framework, we conducted foreign ecological data conversion with our tool and datasets of Environmental Change Network (ECN) (Morecroft et al. 2009). Brief process and result of ECN data conversion are described in the next sections.

Materials and methods

The protocols about ground beetles (*Carabidae*) selected by the long-term ecological research of Kyungpook National University in Korea were used as a predefined protocol (refer to Table 1). We utilized ground beetle (*Carabidae*) datasets collected from 12 different

Lee *et al. Journal of Ecology and Environment* (2017) 41:20

Page 2 of 3

Table 1 Protocols of ground beetles (*Carabidae*) used in this study

Species (measurement)	Protocol	Attribute
Ground beetles (*Carabidae*)	SC	Start date, start time, search week, recorder, end date, end time, hour, maximum temperature, minimum temperature, average temperature, humidity, wind speed, wind direction, weather, reference, description
	SO	Start date, trap ID, species, count, description

SC survey condition, *SO* species observed

observatory sites of ECN (Rennie et al. 2015). They contain two types of files: recording tables of species from each 12 sites and record date information table. Considering that domestic predefined protocol of *Carabidae* consists of two "species observed (SO)" and "survey condition (SC)," the form of ECN dataset is very similar to that of Korea. However, since foreign ecological data were collected according to their own protocols and therefore attributes (variables) or data types are generally different from ours, data conversion for standardization should be done before integrating foreign data with domestic data.

In ECN data conversion process, the first step is to upload ECN data file and select our predefined protocol. Input data is converted into common form based on chosen protocol which clarifies attributes and data types for target species, in our case, all species belonging to the ground beetles. Next, only some part of data which contain the target species are selectively extracted from input data. As the ECN data used in this conversion consists only of a number of ground beetle species corresponding to the protocol, no specific species selection is necessary at this step because all the input data should be used in the next conversion process. In the third step

Table 2 Data conversion results from 12 datasets of ECN. Numbers outside the parentheses indicate the total number of attributes defined in each data table, and the numbers inside the parentheses indicate how many attributes real records contain

Site	Dataset	Year(s)	Before		Species	Protocol	After	
			Attributes	Records			Attributes	Records
ALI	IG_date	1994–2012	5 (5)	804	Ground beetles (*Carabidae*)	SC	16 (3)	804
	IG_record	1994–2012	9 (9)	18,252		SO	5 (5)	18,252
CAI	IG_date	1999–2012	5 (5)	601	Ground beetles (*Carabidae*)	SC	16 (3)	601
	IG_record	1999–2012	9 (9)	6816		SO	5 (5)	6816
DRA	IG_date	1993–2009	5 (5)	631	Ground beetles (*Carabidae*)	SC	16 (3)	631
	IG_record	1993–2009	9 (9)	9709		SO	5 (5)	9709
GLE	IG_date	1994–2012	5 (5)	879	Ground beetles (*Carabidae*)	SC	16 (3)	879
	IG_record	1994–2012	9 (9)	10,320		SO	5 (5)	10,320
HIL	IG_date	1993–2012	5 (5)	402	Ground beetles (*Carabidae*)	SC	16 (3)	402
	IG_record	1993–2012	9 (9)	12,836		SO	5 (5)	12,836
MOO	IG_date	1993–2012	5 (5)	1005	Ground beetles (*Carabidae*)	SC	16 (3)	1005
	IG_record	1993–2012	9 (9)	7935		SO	5 (5)	7935
NOR	IG_date	1993–2010	5 (5)	744	Ground beetles (*Carabidae*)	SC	16 (3)	744
	IG_record	1993–2010	9 (9)	12,982		SO	5 (5)	12,982
POR	IG_date	1994–2010	5 (5)	648	Ground beetles (*Carabidae*)	SC	16 (3)	648
	IG_record	1994–2012	9 (9)	9322		SO	5 (5)	9322
ROT	IG_date	1992–2009	5 (5)	523	Ground beetles (*Carabidae*)	SC	16 (3)	523
	IG_record	1992–2009	9 (9)	18,188		SO	5 (5)	18,188
SNO	IG_date	1999–2012	5 (5)	660	Ground beetles (*Carabidae*)	SC	16 (3)	660
	IG_record	1999–2012	9 (9)	10,748		SO	5 (5)	10,748
SOU	IG_date	1994–2012	5 (5)	806	Ground beetles (*Carabidae*)	SC	16 (3)	806
	IG_record	1994–2012	9 (9)	11,377		SO	5 (5)	11,377
WYT	IG_date	1993–2012	5 (5)	840	Ground beetles (*Carabidae*)	SC	16 (3)	840
	IG_record	1993–2012	9 (9)	11,754		SO	5 (5)	11,754

Lee *et al. Journal of Ecology and Environment* (2017) 41:20

Page 3 of 3

of attribute mapping, we specified the relations between ECN data attributes and standard attributes of the protocol by manually matching them with each other. Unlike the domestic protocol, the ECN data table does not contain some attributes related to the survey environment (temperature, humidity, wind, weather, etc.). Therefore, there is a little data to be converted under the survey environment attributes while the attributes for the species observation (search dates, species, and count) are all matched up and included in the following process. Finally, data type and unit of each attribute are properly transformed to suit the protocol. Attributes of "search start date" and "search end date" are standardized in the format of "YYYY-MM-DD" from "DD-MM-YYYY" by using date conversion function, which adjusts the order of year, month, and day to fit the desired order of protocol. The substitution function converts all months written in English into numbers. For example, search date such as 01-MAR-2000 was changed to 2000-03-01 with date conversion and substitution functions. The standardized data is finally stored in a tabular csv file. The conversion results are summarized in Table 2.

Results and discussion

Using our semi-automatic data conversion tool, we could easily and efficiently standardized ECN datasets of ground beetles (*Carabidae*) which are different from our predefined protocol in various aspects such as data attributes and types. We expect that our attempt to convert domestic and foreign ecological data into a standardized format in a systematic way could provide broad usability for data integration and association analysis in many ecological and environmental studies.

Abbreviations
ECN: Environmental Change Network; EML: Ecological Metadata Language; LTER: Long Term Ecological Research network; NEON: National Ecological Observatory Network

Availability of data and materials
Data are not publicly available to this article because they are used under license for the current study.

Authors' contributions
HL carried out the studies, performed the analysis, and wrote/reviewed the manuscript; MS participated in the design of the study and wrote/reviewed the manuscript; OK participated in the design of the study. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Bio-Intelligence & Data Mining Laboratory, Graduate School of Electronics Engineering, Kyungpook National University, Daegu 41566, Korea. [2]School of Applied Bioscience, College of Agriculture and Life Sciences, Kyungpook National University, Daegu 41566, Korea.

References
Bonet, F. J., Pérez-Pérez, R., Benito, B. M., De Albuquerque, F. S., & Zamora, R. (2014). Documenting, storing, and executing models in Ecology: a conceptual framework and real implementation in a global change monitoring program. *Environmental Modelling & Software, 52*, 192–199.

Fegraus, E. H., Andelman, S., Jones, M. B., Schildhauer, M. (2005). Maximizing the value of ecological data with structured metadata: an introduction to Ecological Metadata Language (EML) and principles for metadata creation. *Bulletin of the Ecological Society of America, 86*, 158–168.

Keller, M., Schimel, D. S., Hargrove, W. W., Hoffman, F. M. (2008). A continental strategy for the National Ecological Observatory Network. *The Ecological Society of America, 6*, 282–284.

Kelling, S., Hochachka, W. M., Fink, D., Riedewald, M., Caruana, R., Ballard, G., & Hooker, G. (2009). Data-intensive science: a new paradig m for biodiversity studies. *BioScience, 59*(7), 613–620.

Lee, H., Jung, H., Shin, M., & Kwon, O. (2017). Developing a semi-automatic data conversion tool for Korean ecological data standardization. *Journal of Ecology and Environment, 41*, 11.

Michener, W. K., & Jones, M. B. (2012). Ecoinformatics: supporting ecology as a data-intensive science. *Trends in Ecology and Evolution, 27*(2), 85–93.

Morecroft, M. D., et al. (2009). The UK Environmental Change Network: emerging trends in the composition of plant and animal communities and the physical environment. *Biological Conservation, 142*, 2814–2832.

Rennie, S., et al. (2015). *UK Environmental Change Network (ECN) carabid beetle data: 1992-2012.* https://doi.org/10.5285/4c9613ce-de52-41b1-9fde-7c41f9199686.

San Gil, I., et al. (2009). The Long-Term Ecological Research community metadata standardisation project: a progress report. *International Journal of Metadata, Semantics and Ontologies, 4*, 141–153.