Ecological Indicators 53 (2015) 61–69

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Short communication

Digital repository of associations between environmental variables: A new resource to facilitate knowledge synthesis

C. Richard Ziegler[a,*], J. Angus Webb[b,c], Susan B. Norton[a], Andrew S. Pullin[d], Andreas H. Melcher[e]

[a] Office of Research and Development, United States Environmental Protection Agency, 1200 Pennsylvania Avenue, NW (8623-P), Washington, DC 20460, United States
[b] Department of Resource Management and Geography, The University of Melbourne, Parkville 3010, Victoria, Australia
[c] Department of Infrastructure Engineering, The University of Melbourne, Parkville 3010, Victoria, Australia
[d] Centre for Evidence-Based Conservation, School of Environment, Natural Resources and Geography, Bangor University, Bangor LL57 2UW, United Kingdom
[e] Department of Water, Atmosphere and Environment, University of Natural Resources and Life Sciences (BOKU), Max Emanuel-Strasse 17, 1180 Vienna, Austria

ARTICLE INFO

Article history:
Received 6 February 2014
Received in revised form
24 December 2014
Accepted 2 January 2015

Keywords:
Environmental assessment
Systematic review
Meta-analysis
Bibliometrics
Informatics
Crowdsourcing

ABSTRACT

Responsible care and management of Earth's resources requires scientific support, but the pool of underused research is growing rapidly. Environmental science research studies describe associations between variables (e.g. statistical relationships between stressors and responses). We propose open-access and online sharing of such associations. This concept differs from various efforts around the world to promote sharing of primary research data, but holds similar goals of improved use of existing knowledge. The initiative is made possible by recent developments in information technology and evolving online culture (e.g. crowdsourcing and citizen science). We have begun to connect existing projects that catalog and store associations, thereby moving toward a single virtual repository. Researchers and decision makers may share and re-use associations for myriad purposes, including: increasing efficiency and timeliness of systematic reviews, environmental assessments and meta-analyses, identifying knowledge gaps and research opportunities, providing evolved metrics of research impact, and demonstrating connections between research and environmental improvement.

1. Introduction

Environmental managers and policy makers require timely and quality scientific support for effective assessments, decision making and actions (e.g. Abbot, 2009; Cane, 2010). There is a critical need for mechanisms to help organize and distil the vast scientific literature to support these activities (e.g. Parr et al., 2012). However, while the published paper has long been the accepted means of disseminating research findings, "It isn't the documents which are actually interesting, it is the things they are about!" (Berners-Lee, 2007).

Imagine therefore being able to efficiently access summarized findings of all research studies on a chosen environmental topic. Findings from studies can be extracted, atomized, and stored, thereby facilitating retrieval, synthesis and sharing with wide audiences beyond what is easily achievable with a collection of written manuscripts. The challenge is to manage and/or summarize research findings so that they can be discovered and re-used by investigators asking new or different questions. Multiple types of information from the fields of ecology and environmental science have been, or could be, cataloged and shared (Table 1). Our focus is on a specific sub-set of research findings – associations between two variables.

Associations are of particular interest because they often provide evidence of underlying causal processes that produced them. For example, one variable may directly cause another, or the exact causal web may be complex (Pearl, 2009). Importantly, associations are raw findings from research studies rather than the study author's interpretation of those findings. In environmental studies, an association typically has three parts: the statistical dependence (1) between a stressor, driver or condition (2) and an observed response (3). For example, Mims and Olden (2012), examining responses of fish assemblages to hydrologic alteration, found a statistically significant positive association (dependence)

* Corresponding author. Tel.: +1 703 347 8554; mobile: +1 202 577 9031.
E-mail addresses: ziegler.rick@epa.gov, zieglermail@yahoo.com (C.R. Ziegler), angus.webb@unimelb.edu.au (J.A. Webb), norton.susan@epa.gov (S.B. Norton), a.s.pullin@bangor.ac.uk (A.S. Pullin), andreas.melcher@boku.ac.at (A.H. Melcher).

http://dx.doi.org/10.1016/j.ecolind.2015.01.003
1470-160X/

Table 1
Examples of environmental science contributions, which can be cataloged and shared.

Contribution	Example components	Challenges[a]		Example existing mechanisms for sharing and re-use [outside of publications]	Example benefits of contributions being shared openly through online systems
		Institutional (e.g. ownership issues)	Technical (complexity and heterogeneity of information)		
Underlying data	Data itself (e.g. geospatial species distributions, water quality time series)	●	●	Collectors/owners upload data to repositories (e.g. Dryad), and/or describe data in registries (e.g. Ecological Society of America Data Registry)	Conduct different analyses than originally intended; combine multiple datasets to conduct meta-analyses
Quantitative descriptions	Models; equations	◒	●	Modelers develop code (or extract from literature) and share using source code repositories (e.g. Github; iemhub.org)	Re-use and/or edit models for purposes not originally intended; combine multiple models into larger simulations [e.g. integrated assessment models, virtual or augmented-reality games (Costanza et al., 2014)]
Ideas for future research	"Next steps" section from manuscripts	◒	●	*No known dedicated mechanisms*	Organize "next steps" from multiple studies to identify knowledge gaps, and guide future research directions and funding
Associations (*focus of this manuscript*)	Stressors, responses, dependence characteristics, supplemental information (e.g. effect size, level of replication)	◒	●	Scientists manually extract associations from literature and populate the proposed database herein and/or the semantic web	Facilitate information syntheses (e.g. systematic reviews, meta-analyses, assessments) identify knowledge gaps; identify direct connections between research[ers] and environmental improvements
Qualitative descriptions	Definitions; explanations of meaning	◒	◒	Descriptions are chosen, sometimes from what might be considered seminal, authoritative or original sources of information	Reconcile otherwise controversial meanings (e.g., "biodiversity", "sustainability")
Species specific information	Traits; taxonomic treatments (species, genus, etc.)	◒	◒	Information is extracted from literature and/or literature is semantically enhanced to populate curated databases and/or repositories (e.g. AnAge Database of Animal Aging and Longevity; Plazi taxonomic treatments)	Feed bio-encyclopedias (e.g. Encyclopedia of Life) with contributions; facilitate meta-analyses; increase re-use of published information
Basic metadata, study attributes	Authors; dates; keywords; locations; supporting citations	○	○	Citation indexing services develop and manage bibliographic databases (e.g. Elsevier's SciVerse Scopus; Thomson Reuters' Web of Science; Google Scholar)	Search for manuscripts using authors, keywords, dates, etc.; calculate impact factor and h-index

[a] ● = high, ◒ = medium, ○ = low (qualitative judgment of authors).

between the seasonality of flow regimes (stressor) and the prevalence of 'periodic' life-history strategists (response) using data from across the continental US. A single research study may report multiple associations, with several potential causal agents associated with the response, potentially indicative of additive or interactive causation. Supplemental information and study attributes (e.g. location, study design, level of replication, effect size, quality and strength of the dependence) further aid in interpreting and weighting individual associations in the context of new hypotheses and analyses. The extraction of such information from Mims and Olden (2012) is detailed in Webb et al. (2015).

We are developing an open-access, online and machine-readable repository of associations, external, but complementary, to the traditional written manuscript and scientific publication paradigm. The underlying framework of this exchange (including for example, databases and database fields) aims to facilitate syntheses of multiple studies, allowing derivation of general and specific ecological responses to a multitude of stressors. Database fields for an association include two variables (e.g. stressor, driver, or condition, and the response), their statistical dependence and supplemental information described above. Existing databases of associations provide a tangible starting point for determining how to share this type of information and for demonstrating the usefulness of sharing associations (see further below). However, we conceive of an online repository of associations as part of the semantic web (Berners-Lee et al., 2001; Nešić et al., 2011), either with or without centralized databases. Any individual association (and its sub-component parts) may ultimately be uniquely identified at its source and made machine-readable, to be used and re-used for various knowledge synthesis purposes. Thus,