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
Available online 3 February 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),

* 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

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)	●.eps	●.eps	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	◒.eps	●.eps	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	◒.eps	●.eps	*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)	◒.eps	●.eps	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	◒.eps	◒.eps	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.)	◒.eps	◒.eps	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	○.eps	○.eps	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).

examining responses of fish assemblages to hydrologic alteration, found a statistically significant positive association (dependence) 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, semantically-enabled research findings can be used across the web (e.g. in environmental encyclopedias, models and decision support tools) without necessarily using databases.

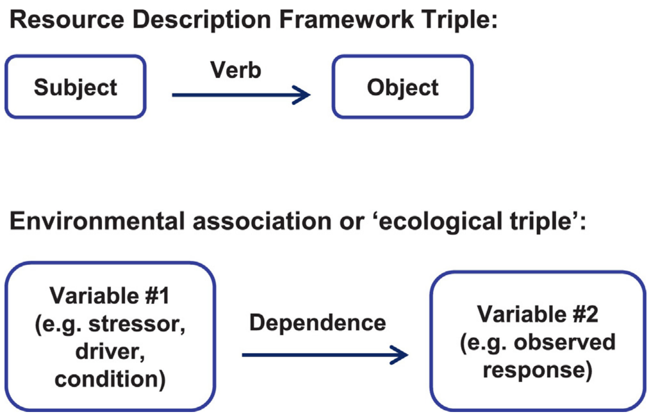
Resource Description Framework Triple:
Subject
Verb
Object
Environmental association or 'ecological triple':
Variable #1 (e.g. stressor, driver, condition)
Dependence
Variable #2 (e.g. observed response)

Fig. 1. Triples. The Resource Description Framework provides a three-part knowledge management model for storing web information; this is analogous to an "ecological triple" – for example, stressor–response relationships, often found in environmental science literature.

2. Analogs

The Plazi repository of taxonomic treatments (plazi.org; Agosti and Egloff, 2009) provides an analog to the concept of sharing associations. Plazi works to extract taxonomic information ('treatments', including species name, genus, descriptions, materials examined, distribution, etc.) from literature to populate a repository. This is implemented by 'marking up' or 'semantically enhancing' taxonomic manuscripts – that is, labeling pieces of text for identification and re-use by machines (Penev et al., 2010). The components of associations are generally more variable or heterogeneous than taxonomic treatment information, and therefore may be more technically challenging to catalog (see also Table 1). Plazi and related efforts are addressing copyright issues involved in sharing information extracted from publications (e.g. Agosti and Egloff, 2009; Patterson et al., 2014), and an open exchange of previously published associations will also require careful attention to copyright law. However, we question whether this will always be the case. The proliferation of open access publishing, including traditionally subscriber-only journals, is making more and more research freely available on the web. Similarly, many national funding bodies now require all research publications to be open access. Ultimately, the 'open sourcing' of ecological data (Parr, 2007), and collective buy-in of the greater scientific community – including publishers and researchers alike – will drive the sharing of associations.

The Resource Description Framework (RDF) provides a more abstract analog, namely, a data model for web-based information (Berners-Lee et al., 2001). An RDF expression has three parts – subject, verb and object – comprising a 'triple' (Fig. 1). Web information can be expressed as collections of triples. An association between environmental variables is very similar to an RDF triple. Furthermore, in environmental science, causal webs are sometimes visually represented as collections of associations (e.g. Fig. 2).

Ideally, efforts to share and use associations or 'ecological triples' will gain some benefit from information science disciplines, which have been exploring the concept of triples since the 1990s. At a minimum, RDF digital storage and visual display mechanisms may be transferable to ecological triples. Moreover, the RDF knowledge management analog demonstrates the power of depicting complex information about our world by breaking that information into smaller pieces. Consider, for example, the spectrum of potential interactions among causes of ecological responses, including additive and synergistic relationships, and myriad ways of describing those relationships. Our above example on responses of fish assemblages to hydrologic alteration was simple to catalog. In extracting ecological triples from diverse studies, it becomes increasingly important to dissect information to its simplest components. This does not limit the usefulness of cataloging associations; rather, complexity is achieved by connecting multiple triples (e.g., Fig. 2).

3. The rationale

There is a growing pool of under-used science (e.g. Bell et al., 2009; Howe et al., 2008) that could better contribute to environmental decision-making and research. For example, information syntheses – as part of systematic reviews (Khan et al., 2003; Pullin and Knight, 2001), their underlying meta-analyses (Osenberg et al., 1999; Vetter et al., 2013) and environmental assessment processes (Norris et al., 2012; Suter et al., 2010) – are often mired by two frustratingly inefficient steps: the search for relevant studies and extraction of findings from each study. The problem partly lies with the current process of scientific publication; that is, potentially relevant results are scattered widely across a prolific literature of variable media, language and quality, making discovery and screening for relevance time consuming and imprecise. Additionally, the heterogeneous characteristics of ecological data (Reichman et al., 2011), and associated conclusions that might underpin decision making, complicate attempts to organize and synthesize environmental evidence. A comprehensive repository of associations would help alleviate these problems, by largely eliminating search and extraction steps. Such a change in practice could cut costs and speed up synthesis, meta-analysis and assessment, ultimately enabling science to respond on timescales commonly associated with policy windows and environmental management (weeks rather than years).

Recent developments in online literature resources demonstrate how information technology can support and complement the traditional scientific publication paradigm, and thus highlight the timeliness of a repository of associations. Thompson Reuters recently posted a challenge, asking for new ideas for online interaction with scholarly content (2013). Similarly, an Elsevier contest, "Knowledge Enhancement in the Life Sciences" (http://www.elseviergrandchallenge.com/description.html, accessed 06.01.14) yielded entries such as Roderic Page's "Visualizing a scientific article" concept, which was aimed at technologically unlocking information from written manuscripts (Page, 2008).

Environmental assessment methods, in particular causal assessment, rely on associations from the literature to develop strength of evidence analyses to inform research and management (Greet et al., 2011; Hill, 1965; Norris et al., 2012; Suter et al., 2010). Syntheses of associations can help support or reject causal hypotheses (diagnostic or retrospective assessment). The hydrologic alteration and fish assemblages example mentioned above – combined with other environmental associations – might be used to determine potential candidate causes of river impairment (Ziegler, 2007). Moreover, in response to calls for ecology to become a more predictive science (Peters, 1991), such syntheses hold tremendous promise for supporting prospective assessment (e.g. risk analysis and scenario planning).

4. Existing methods and mechanisms for synthesizing environmental evidence

A global, concerted effort to share associations would build upon existing technological work in the publication arena and existing databases that have been developed to support causal assessment. At least three organizations from different parts of the globe have simultaneously and largely independently developed methods and tools for evaluating associations, along with databases that include atomized findings from environmental studies. The organizations

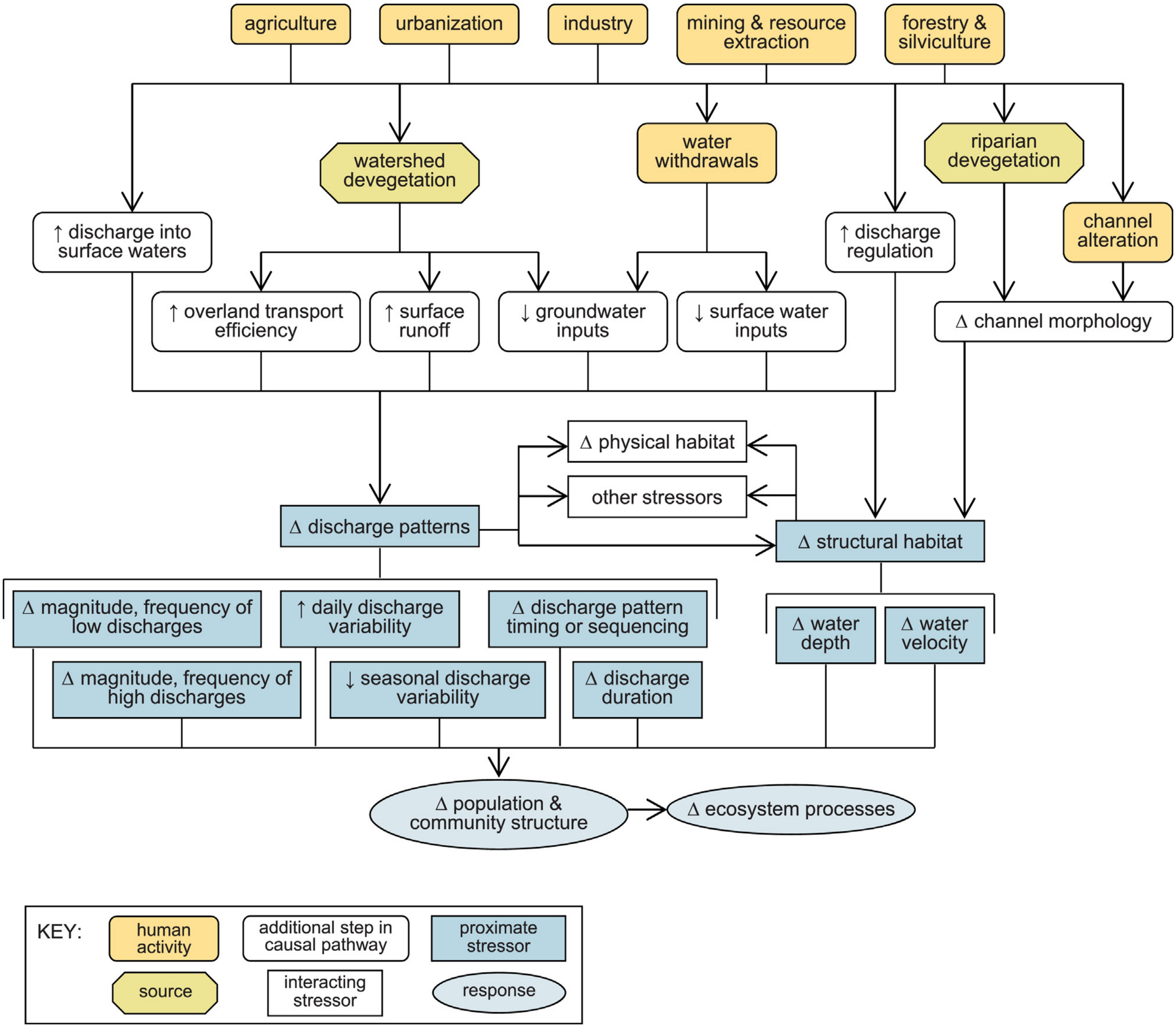

Fig. 2. Example conceptual diagram. Illustrates interconnected associations used to visualize the impacts of flow alteration on the biological condition of streams and rivers. Adapted from Schofield and Ziegler (2007).

are the United States Environmental Protection Agency's (US EPA) CADDIS project (Causal Analysis/Decision Diagnosis Information System), the European Union's (EU) WISER project (Water bodies in Europe: Integrative Systems to assess Ecological status and Recovery) and the eWater Cooperative Research Centre's Eco Evidence project in Australia (Table 2). Once the three projects became known to one another, the US EPA hosted an international workshop in October 2009, which brought together researchers, technology experts and peer-production/crowdsourcing specialists. The workshop resulted in a partnership of scientists and managers (of which we are members) committed to linking the existing databases, thereby creating a virtual exchange or digital repository. The partnership provides a foundation for continued tool development that leverages the association repository, as proposed, for example, under the EU-funded MARS project (Managing Aquatic ecosystems and water Resources under multiple Stress; see also http://www.mars-project.eu and Hering et al., 2014). The Collaboration for Environmental Evidence (CEE; also see Table 2) recently joined the partnership; we envision the repository supporting CEE systematic reviews. A core group of researchers (from three continents) continues to meet periodically to move the effort forward.

The causal assessment methods mentioned above employ systematic review of literature (Khan et al., 2003) to address questions of causes and effects in environmental systems. Systematic review combines results of multiple studies on the same question to provide an objective and transparent assessment of available evidence. Although most prevalent in the health sector, systematic review has now spread to other disciplines where research is conducted to inform decision making in policy or practice (e.g. social sciences and environmental management).

The CEE has developed guidelines for systematic review in environmental science (Collaboration for Environmental Evidence, 2013), hosts an online library of completed systematic reviews, and has established an open-access journal *Environmental Evidence*. Separately and independently of CEE, the three other organizations from around the globe established methods and tools for evaluating evidence of causes and effects in aquatic ecosystems. These range from questions about specific causation (CADDIS) to general causation (WISER and Eco Evidence) (Table 2). All four applications use

Table 2
Existing applications and their purpose.

Application	Purpose
CEE (see: http://environmentalevidence.org) (Pullin and Knight, 2009)	Open collaboration that promotes and facilitates the conduct and dissemination of systematic reviews in environmental management
US: CADDIS – Causal Analysis/Decision Diagnosis Information System (see: http://epa.gov/caddis) (Norton et al., 2009)	Method and tools that help investigators systematically identify probable causes of specific undesirable biological effects in aquatic systems (e.g. a fish kill on a particular river)
EU: WISER (see: http://wiser.eu/download/D5.1-2.pdf and http://wiser.eu/results/conceptual-models/) (Feld et al., 2011)	Specific systematic review of evidence linking common restoration efforts and their environmental implications to four biological quality elements (fish, benthic invertebrates, macrophytes and phytobenthos)
Australia: Eco Evidence (see: http://toolkit.net.au/tools/eco-evidence) (Norris et al., 2012)	Method and software that help investigators systematically synthesize evidence for more general cause-effect hypotheses (e.g. whether habitat enhancement can increase fish abundance)

literature searches to identify relevant research. Associations are then manually extracted and synthesized using various approaches from simple tallying of papers to quantitative meta-analysis. Conceptual diagrams (e.g. Fig. 2) are used to guide literature extractions (Eco Evidence), display results online (CADDIS and WISER), and synthesize evidence of causes and effects (WISER).

The CADDIS website provides a general idea about the type of information included in its database of associations (http://epa.gov/caddis, and specifically navigate to the “ICD Application”). The Eco Evidence software (Webb et al., 2011) uses web services to directly retrieve information from the online Eco Evidence database (Webb et al., 2012a). The Eco Evidence database catalogs more specific information than CADDIS. Both databases store and utilize online, machine-readable environmental associations, similar to the “flow regulation” association example described above.

5. Benefits of a digital repository of associations

> “. . .[A] traditional biological journal will become just one part of various biological data resources as the scientific knowledge in published papers is stored and used more like a database.” (Bourne, 2005)

> “The once-sharp distinction between journals and databases is beginning to blur.” (Seringhaus and Gerstein, 2007)

The exchange of research findings that we envision will provide a more comprehensive foundation for environmental resources management, help inform development of ecological theory, illuminate knowledge gaps to help guide research planning and help combat the scientific information landslide (Attwood et al., 2009). Table 3 illustrates some of the potential uses of the repository. Generally speaking, the academic community benefits from tools that facilitate rapid summarization of high volumes of relevant findings. Researchers are typically not looking for papers to read, “but rather to find, assess, and exploit a range of information by scanning portions of many articles” (Renear and Palmer, 2009). In addition to assisting the academic community, online interfaces have been developed to retrieve machine-readable associations (see above), and subsequently provide summarized research findings in a format more readily understood by managers and decision makers (Webb et al., 2011). This facilitates the ongoing move toward evidence-based environmental management. It also improves transparency, as decisions can be traced back to scientific evidence – in this case, open-access associations.

Looking beyond applications that are already possible, there are many potential future benefits of an association repository. Web-based decision support tools, scenario planning tools and simulation games might one day be connected to machine-readable associations to improve their scientific underpinning. Users and contributors of the proposed repository will be able to see who is entering and using associations, allowing for example, researchers to more readily find others working in specific areas, with greater resolution than by using bibliographic information alone. To facilitate this, the proposed digital repository of associations could be connected to researchers' online profiles (e.g. http://www.researchgate.net/). Notably, a repository of associations would likely mirror the existing direction of research, including potential publication biases. Reuse or analysis of the repository's information would need to account for possible biases in a similar way that any other literature synthesis must. Ideally, however, a repository of associations might help illuminate where such biases exist, along with research gaps and understudied – but important – associations.

At another level, we believe the repository provides a building block toward more evolved metrics of research impact beyond currently dominant bibliometric indicators and more recent ‘altmetrics’ (Piwowar, 2013; Priem et al., 2010). Unlike other emerging forms of information sharing in environmental sciences (Table 1), associations often provide a more direct link to natural resources management and decision-making. Consider, for example, how systematic reviews and assessment studies rely on associations and then provide decision makers with evidence-based guidance (see above). Such information flows can be tracked in our Web era (Priem, 2013), potentially with support from existing online mechanisms being advanced by the altmetrics development community (e.g. http://impactstory.org/). It is conceivable that explicit connections will one day trace the path of how a researcher's work informed a decision, and subsequently to what extent that research and decision led to actual outcomes or improvements in environmental quality (Fig. 3, Egghe, 2006; Hirsch, 2005). This may ultimately assist in the struggle to better assess the impact of specific environmental studies and – on the other hand – to attribute societal benefits to specific research (Sutherland et al., 2011).

6. Sharing associations faces different challenges than sharing data

The past decade has seen increasing calls for on-line publishing and sharing of scientific data (Arzberger et al., 2004; Costello, 2009; De Mesquita et al., 2003; Piwowar et al., 2008), including ecological data (Chavan and Ingwersen, 2009; Parr, 2007; Reichman et al., 2011). However, ecologists have been slow to adopt a more open attitude to data sharing (Costello, 2009). Challenges include lack of well-accepted incentive systems for researchers, equivalent to literature citations (Thorisson, 2009), political interference (Goldston, 2008), difficulty of providing easy access to data (De Mesquita et al., 2003), confidentiality (Piwowar et al., 2008), existing research culture (Costello, 2009) and the fear of having novel conclusions “scooped” by others (Parr, 2007).

Institutionally, researchers may be more willing to share their findings rather than primary data. Because associations are atomized information from previously published sources, citation of the original publication is necessary when that information is used. Sharing of primary data is associated with increased citation rates

Table 3
Example use cases of digital repository of associations.

Use case description	Challenge or question addressed	Client or beneficiary	Output	Outcome
EU assessment to guide river rehabilitation (also see Panel 2)	What are the efficacy and ecological consequences of existing restoration projects?	EU member countries attempting to improve riparian conditions	Meta-analysis/review of 100 peer reviewed international articles, 60% of which were US flow studies	Better knowledge among decision makers when attempting to rehabilitate rivers in EU; better understanding of geographic-based knowledge gaps in the literature
Site specific causal assessments of degraded fish or macroinvertebrate assemblages in streams	What are candidate causes of biological impairments? What stressor levels have been associated with similar effects in other studies?	Public or private sector scientists investigating causes of undesirable effects	Summary of relevant findings from other locations	More confident attribution of cause; management actions more likely to be directed at true cause(s), and result in improved biological condition
Systematic review	What is the impact of wooded riparian buffer strips on stream temperature?	Government policy and private sector fishing interest groups	Meta-analysis of effects and identification of knowledge gaps	Informed policy development on climate mitigation measures for rivers and streams
Systematic review	What is the impact of wind farm installation on bird populations?	Government policy and local planning authorities; public stakeholder groups	Meta-analysis of effects and identification of knowledge gaps	Informed decisions on location of wind farm installations
Facilitate 'horizon scanning' or trend analysis for emerging environmental issues (Sutherland and Woodroof, 2009)	What are emerging stressors and threats to the environment, and how should they be prioritized for research and policy action?	Governments and environmental organizations such as US EPA (National Advisory Council for Environmental Policy and Technology, 2002)	Emerging environmental threats are preemptively discussed and incorporated into governmental strategies and research plans	Threats to the environment and degradation are proactively avoided
PhD thesis aid, to foster efficient and targeted research	Why is pollinator insect abundance dropping?	Primarily academic initially	More targeted literature review in the thesis, publishable as a peer-reviewed paper	Improved research planning for the PhD project. The review identifies critical research gaps that can be filled by targeted research in the project

(Piwowar et al., 2007), and sharing of environmental associations also provides a more direct pathway to increased citation rate (e.g. via their use, and consequent citation, in synthesis studies). Opportunities for increased citation that result from shared associations may prove a powerful incentive for authors to make such findings available. Also, because associations are drawn from previously published sources already deemed appropriate for the public domain, concerns over confidentiality or political interference are reduced. Moreover, other researchers cannot scoop conclusions from previously published research. Instead, new syntheses of associations drawn from multiple studies will yield novel conclusions.

Sharing associations may even provide a stepping-stone toward greater acceptance and practice of data sharing. As explained above, we believe that researchers will be willing to share published associations. This could combat cultural and institutional resistance toward scholarly sharing in general. Sharing causal associations also requires technical and logistical innovation (e.g. new and improved ontologies) that may be transferable to data sharing and probably other forms of information exchange as well.

7. Next steps

7.1. Mechanisms for sharing associations

To date, associations have been extracted from scientific articles by research and contractor staff, who read chosen studies and manually enter information into existing association databases (see above). This is a labor-intensive process (Webb et al., 2012b), and different methods are required to achieve the goal of a large-scale, sustainable exchange. Concurrently, populating databases via any mechanism raises issues of credibility, reliability and bias, particularly if contributed information leads to an environmental decision involving risk, liability or attribution of blame.

Online crowdsourcing may offer an efficient model for accurately populating and moderating the repository (Fraternali et al., 2012). In this case, the crowd might be limited to a combination of study authors, colleagues and students, or enthusiasts among the general public interested in specific research topics. Contributor authentication and quality control mechanisms employed by existing online citizen science and crowdsourcing tools (e.g. Wikipedia and Old Weather quality control: http://en.wikipedia.org/wiki/Wikipedia:Quality_control and http://www.oldweather.org/faq, respectively) are transferable to crowdsourcing of associations. Moreover, crowdsourcing mechanisms have produced comparable resources of sometimes better quality and more up-to-date than those produced by more traditional publication mechanisms (e.g. Giles, 2005; Reavley et al., 2012).

Manuscript mark-up and text-mining may be used to assist in extracting associations from legacy research and new studies (Attwood et al., 2009). Semantic or descriptive mark-up, whereby terms, phrases, interconnections and other pieces of documents are annotated and made machine-readable, can be facilitated by tools such as the Utopia Documents PDF reader (getutopia.com; Attwood et al., 2010). However, mark-up processes are often difficult and labor-intensive (Attwood et al., 2009). Natural language processing (NLP) computer software (Joshi, 1991) has been used for text-mining medical research (Demner-Fushman et al., 2009), and may facilitate semi-automatic extraction of associations. We have had some success using NLP to extract study-level metadata (Willett et al., 2012), though associations are notably more heterogeneous, and therefore more technically challenging to extract.

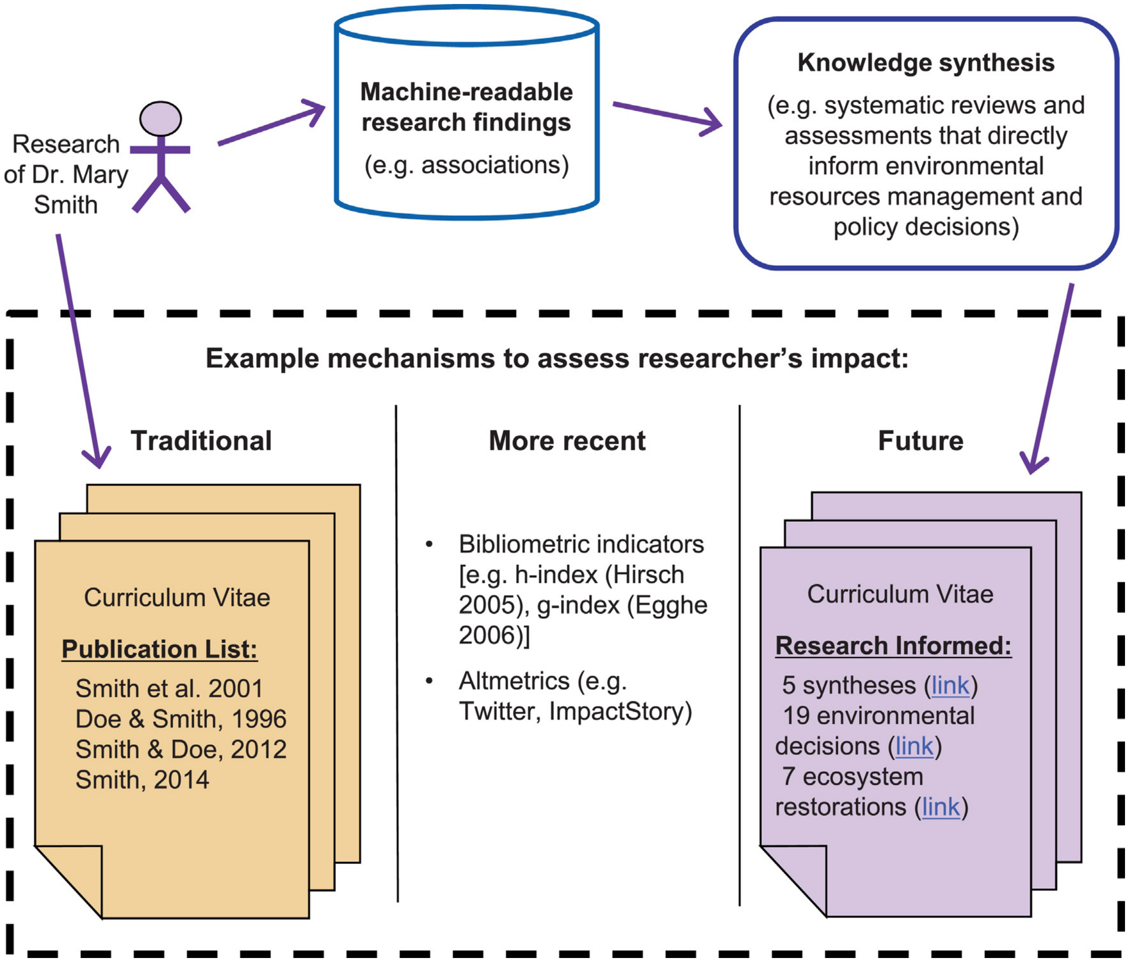
Research of Dr. Mary Smith
Machine-readable research findings
(e.g. associations)
Knowledge synthesis
(e.g. systematic reviews and assessments that directly inform environmental resources management and policy decisions)
Example mechanisms to assess researcher's impact:
Traditional
More recent
Future
Curriculum Vitae
Publication List:
Smith et al. 2001
Doe & Smith, 1996
Smith & Doe, 2012
Smith, 2014
Bibliometric indicators [e.g. h-index (Hirsch 2005), g-index (Egghe 2006)]
Altmetrics (e.g. Twitter, ImpactStory)
Curriculum Vitae
Research Informed:
5 syntheses (link)
19 environmental decisions (link)
7 ecosystem restorations (link)

Fig. 3. From measuring outputs to actual outcomes or improvements in environmental quality.

Some combination of mark-up, text-mining and crowdsourcing may offer the best hope for widespread cataloging of associations.

7.2. Interoperability

Standard data structures and web services allow for sharing of machine-readable associations among different sources (e.g. the three existing databases described above) and facilitate the development of different tools that search for, download and synthesize research findings. US EPA and eWater Cooperative Research Center (CRC) representatives met in March 2011, to begin developing a standard 'Ecological Exchange Language' (EEL) for the repository of associations. Database fields that comprise an association are sometimes qualitative (e.g. classification of cause and effect), requiring a standardized and controlled vocabulary for options in such fields (e.g. a list of recognized causes with definitions of each). Terms and definitions have been largely drawn from an existing environmental ontology, EnvO (http://environmentontology.org). Using terms from EnvO avoids duplication of effort and aligns the vocabulary in the repository with a more widely used ontology.

7.3. Sustainable and interdisciplinary business model

The project partners would like to develop this concept to the point where it can be transferred to an existing or yet-to-be founded international non-profit organization or consortium. An international organization, such as CEE or the United Nations Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES; http://www.ipbes.net/), may provide a framework and infrastructure that could support and nurture the concept. Non-profit organizations with similar overall goals, such as medicine's Cochrane Collaboration (http://cochrane.org) may provide appropriate business model examples. Disciplinary individualism (Abbot, 2009) complicates matters. In addition to expertise in environmental science, technical, managerial and executive understanding of new and emerging technology will be required for the proposed concept to succeed (Kietzmann et al., 2011). Genomics has managed to take advantage of new information technology tools, although this may partly be due to genomics' more manageable data (Parr et al., 2012) compared to challenges associated with ecological information. In our opinion, concerted sharing of associations will help ecology mature as a science, with help from the information science community and more digitally advanced disciplines such as genomics.

8. Conclusion

We see tremendous potential for digitally sharing associations – namely, to increase the value of published research, catalyze synthesis studies, derive new knowledge from existing literature and better connect science to decision making. We have described existing and potential benefits of a repository of associations, and we believe other benefits will emerge as the concept evolves. Although we concentrate on ecology and environmental science, sharing associations would apply equally well to other disciplines. An initiative such as we describe would not have been possible 20, or even 10 years ago, because today's information technology tools and online culture did not exist. We believe the community of environmental scientists and managers would benefit by embracing these new counterparts to the traditional research paper. This would allow a repository of associations to reach the critical mass necessary to become self-sustaining, and ultimately provide the foundation of a more evolved research process.

Acknowledgements

Progress to date in developing a repository of associations has been driven by a number of people other than the authors. We thank Mike Barbour, Christian Feld, Dave Feldman, Michael Griffith, Sue Nichols, Richard Norris, Ralph Ogden, Astrid Schmidt Kloiber, Kate Schofield, Mike Stewardson, Steve Wealands, Jeff White, Lester Yuan and European Union's WISER team. We also thank those who attended the symposium in Washington DC, contributing their ideas to the nascent collaboration. We thank Donat Agosti and Willi Egloff for feedback about Plazi efforts in the context of the proposed concept. We thank Dr. Isabella Bertani for general consultation on technological advancements surrounding data sharing and meta-analysis. This work benefited from EU 7th Framework Directive support, specifically the WISER project (contract number 226273) and the MARS project (contract number 603378).

References

Abbot, M.R., 2009. A new path for science? In: Hey, T., Tansley, S., Tolle, K. (Eds.), The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research. Redmond, Washington, pp. 111–116.

Agosti, D., Egloff, W., 2009. Taxonomic information exchange and copyright: the Plazi approach. BMC Res. Notes 2, 53.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P., Wouters, P., 2004. An international framework to promote access to data. Science 303, 1777–1778.

Attwood, T.K., Kell, D.B., McDermott, P., Marsh, J., Pettifer, S.R., Thorne, D., 2009. Calling international rescue: knowledge lost in literature and data landslide! Biochem. J. 424, 317–333, http://dx.doi.org/10.1042/bj20091474.

Attwood, T.K., Kell, D.B., McDermott, P., Marsh, J., Pettifer, S.R., Thorne, D., 2010. Utopia documents: linking scholarly literature with research data. Bioinformatics 26, i568–i574, http://dx.doi.org/10.1093/bioinformatics/btq383.

Bell, G., Hey, T., Szalay, A., 2009. Beyond the data deluge. Science 323, 1297–1298, http://dx.doi.org/10.1126/science.170411.

Berners-Lee, T., 2007. Levels of Abstraction: Net, Web, Graph [WWW Document], http://www.w3.org/DesignIssues/Abstractions.html (accessed 30.08.13).

Berners-Lee, T., Hendler, J., Lassila, O., 2001. The semantic web: scientific American. Sci. Am. 284, 28–37.

Bourne, P., 2005. Will a biological database be different from a biological journal? PLoS Comput. Biol. 1, e34.

Cane, M.A., 2010. Climate science: decadal predictions in demand. Nat. Geosci. 3, 231–232, http://dx.doi.org/10.1038/ngeo823.

Chavan, V.S., Ingwersen, P., 2009. Towards a data publishing framework for primary biodiversity data: challenges and potentials for the biodiversity informatics community. BMC Bioinform. 10, S2, http://dx.doi.org/10.1186/1471-2105-10-s14-s2.

Collaboration for Environmental Evidence, 2013. Guidelines for systematic review and evidence synthesis in environmental management, version 4.2. Environ. Evid.

Costanza, R., Chichakly, K., Dale, V., Farber, S., Finnigan, D., Grigg, K., Heckbert, S., Kubiszewski, I., Lee, H., Liu, S., Magnuszewski, P., Maynard, S., McDonald, N., Mills, R., Ogilvy, S., Pert, P.L., Renz, J., Wainger, L., Young, M., Richard Ziegler, C., 2014. Simulation games that integrate research, entertainment, and learning around ecosystem services. Ecosyst. Serv. 10, 195–201, http://dx.doi.org/10.1016/j.ecoser.2014.10.001.

Costello, M.J., 2009. Motivating online publication of data. Bioscience 59, 418–427, http://dx.doi.org/10.1525/bio.2009.59.5.9.

De Mesquita, B.B., Gleditsch, N.P., James, P., King, G., Metelits, C., Ray, J.L., Russett, B., Strand, H., Valeriano, B., 2003. Symposium on replication in international studies research. Int. Stud. Perspect. 4, 72–107.

Demner-Fushman, D., Chapman, W.W., McDonald, C.J., 2009. What can natural language processing do for clinical decision support? J. Biomed. Inform. 42, 760–772, http://dx.doi.org/10.1016/j.jbi.2009.08.007.

Egghe, L., 2006. Theory and practise of the g-index. Scientometrics 69, 131–152, http://dx.doi.org/10.1007/s11192-006-0144-7.

Feld, C.K., Birk, S., Bradley, D.C., Hering, D., Kail, J., Marzin, A., Melcher, A., Nemitz, D., Pedersen, M.L., Pletterbauer, F., Pont, D., Verdonschot, P.F.M., Friberg, N., 2011. Chapter three – from natural to degraded rivers and back again: a test restoration ecology theory and practice. Adv. Ecol. Res. 44, 119–209.

Fraternali, P., Castelletti, A., Soncini-Sessa, R., Vaca Ruiz, C., Rizzoli, A.E., 2012. Putting humans in the loop: social computing for water resources management. Environ. Model. Softw. 37, 68–77.

Giles, J., 2005. Internet encyclopaedias go head to head. Nature 438, 900–901, http://dx.doi.org/10.1038/438900a.

Goldston, D., 2008. Big data: data wrangling. Nature 455, 15, http://dx.doi.org/10.1038/455015a.

Greet, J.M., Webb, J.A., Cousens, R.D., 2011. The importance of seasonal flow timing for riparian vegetation dynamics: a systematic review using causal criteria analysis. Freshw. Biol. 56, 1231–1247.

Hering, D., Carvalho, L., Argillier, C., Beklioglu, M., Borja, A., Cardoso, A.C., Duel, H., Ferreira, T., Globevnik, L., Hanganu, J., Hellsten, S., Jeppesen, E., Kodeš, V., Solheim, A.L., Nõges, T., Ormerod, S., Panagopoulos, Y., Schmutz, S., Venohr, M., Birk, S., 2014. Managing aquatic ecosystems and water resources under multiple stress – an introduction to the MARS project. Sci. Total Environ., 106, http://dx.doi.org/10.1016/j.scitotenv.2014.06.

Hill, A.B., 1965. The environment and disease – association or causation? Proc. R. Soc. Med. 58, 295–300.

Hirsch, J.E., 2005. An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. U. S. A. 102, 16569–16572, http://dx.doi.org/10.1073/pnas.0507655102.

Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., Hill, D.P., Kania, R., Schaeffer, M., St Pierre, S., Twigger, S., White, O., Rhee, S.Y., 2008. Big data: the future of biocuration. Nature 455, 47–50, http://dx.doi.org/10.1038/455047a.

Joshi, A.K., 1991. Natural-language processing. Science 253, 1242–1249.

Khan, K.S., Kunz, R., Kleijnen, J., Antes, G., 2003. Five steps to conducting a systematic review. JRSM 96, 118–121, http://dx.doi.org/10.1258/jrsm.96.3.118.

Kietzmann, J.H., Hermkens, K., McCarthy, I.P., Silvestre, B.S., 2011. Social media? Get serious! Understanding the functional building blocks of social media. Bus. Horiz. 54, 241–251, http://dx.doi.org/10.1016/j.bushor.2011.01.005.

Mims, M.C., Olden, J.D., 2012. Life history theory predicts fish assemblage response to hydrologic regimes. Ecology 93, 35–45, http://dx.doi.org/10.1890/11-0370.1.

National Advisory Council for Environmental Policy and Technology, 2002. The Environmental Future: Emerging Challenges and Opportunities for EPA. Washington DC.

Nešić, S., Rizzoli, A., Athanasiadis, I., 2011. Towards a semantically unified environmental information space. Environ. Softw. Syst.

Norris, R.H., Webb, J.A., Nichols, S.J., Stewardson, M.J., Harrison, E.T., 2012. Analyzing cause and effect in environmental assessments: using weighted evidence from the literature. Freshw. Sci. 31, 5–21, http://dx.doi.org/10.1899/11-027.1.

Norton, S.B., Cormier, S.M., Suter II, G.W., Schofield, K., Yuan, L., Shaw-Allen, P., Ziegler, C.R., 2009. CADDIS: the causal analysis/diagnosis decision information system. In: Marcomini, A., Suter II, G.W., Critto, A. (Eds.), Decision Support Systems for Risk-based Management of Contaminated Sites. Springer, New York, pp. 351–374.

Osenberg, C.W., Sarnelle, O., Cooper, S.D., Holt, R.D., 1999. Resolving ecological questions through meta-analysis: goals, metrics, and models. Ecology 80, 1105–1117.

Page, R., 2008. Visualising a scientific article. Nat. Preced., http://dx.doi.org/10.1038/npre.2008.2579.1.

Parr, C.S., 2007. Open sourcing ecological data. Bioscience 57, 309–310, http://dx.doi.org/10.1641/b570402.

Parr, C.S., Guralnick, R., Cellinese, N., Page, R.D.M., 2012. Evolutionary informatics: unifying knowledge about the diversity of life. Trends Ecol. Evol. 27, 94–103.

Patterson, D.J., Egloff, W., Agosti, D., Eades, D., Franz, N., Hagedorn, G., Rees, J.A., Remsen, D.P., 2014. Scientific names of organisms: attribution, rights, and licensing. BMC Res. Notes 7, 79, http://dx.doi.org/10.1186/1756-0500-7-79.

Pearl, J., 2009. Causality: Models, Reasoning and Inference, 2nd ed. Cambridge University Press, New York, NY, USA.

Penev, L., Agosti, D., Georgiev, T., Catapano, T., Miller, J., Blagoderov, V., Roberts, D., Smith, V.S., Brake, I., Ryrcroft, S., Scott, B., Johnson, N.F., Morris, R.A., Sautter, G., Chavan, V., Robertson, T., Remsen, D., Stoev, P., Parr, C., Knapp, S., Kress, W.J., Thompson, C.F., Erwin, T., 2010. Semantic tagging of and semantic enhancements to systematics papers: Zookeys working examples. Zookeys 30 (June), 1–16, http://dx.doi.org/10.3897/zookeys.50.538.

Peters, R.H., 1991. A Critique for Ecology. Cambridge University Press.

Piwowar, H., 2013. Altmetrics: value all research products. Nature 493, 159.

Piwowar, H.A., Becich, M.J., Bilofsky, H., Crowley, R.S., 2008. Towards a data sharing culture: recommendations for leadership from academic health centers. PLoS Med. 5, 1315–1319, http://dx.doi.org/10.1371/journal.pmed.0050183.

Piwowar, H.A., Day, R.S., Fridsma, D.B., 2007. Sharing detailed research data is associated with increased citation rate. PLOS ONE 2, e308, http://dx.doi.org/10.1371/journal.pone.0000308.

Priem, J., 2013. Scholarship: beyond the paper. Nature 495, 437–440.

Priem, J., Taraborelli, D., Groth, P., Neylon, C., 2010. Altmetrics: A Manifesto.

Pullin, A.S., Knight, T.M., 2001. Effectiveness in conservation practice: pointers from medicine and public health. Conserv. Biol. 15, 50–54.

Pullin, A.S., Knight, T.M., 2009. Data credibility: a perspective from systematic reviews in environmental management. New Dir. Eval. 2009, 65–74.

Reavley, N.J., Mackinnon, A.J., Morgan, A.J., Alvarez-Jimenez, M., Hetrick, S.E., Killackey, E., Nelson, B., Purcell, R., Yap, M.B., Jorm, A.F., 2012. Quality of information sources about mental disorders: a comparison of Wikipedia with centrally controlled web and printed sources. Psychol. Med. 42, 1753–1762.

Reichman, O.J., Jones, M.B., Schildhauer, M.P., 2011. Challenges and opportunities of open data in ecology. Science 331, 703–705.

Renear, A.H., Palmer, C.L., 2009. Strategic reading, ontologies, and the future of scientific publishing. Science 325, 828–832, http://dx.doi.org/10.1126/science.1157784.

Schofield, K.A., Ziegler, C.R., 2007. Common Candidate Cause: Simple Conceptual Diagram for Flow Alteration [WWW Document]. US EPA Causal Anal. Diagnosis Decis. Inf. Syst, http://www.epa.gov/caddis/ssr_flow4s.html (accessed 01.06.14).

Seringhaus, M., Gerstein, M., 2007. Publishing perishing? Towards tomorrow's information architecture. BMC Bioinform. 8, 17.

Suter, G.W., Norton, S.B., Cormier, S.M., 2010. The science and philosophy of a method for assessing environmental causes. Hum. Ecol. Risk Assess. 16, 19–34.

Sutherland, W.J., Goulson, D., Potts, S.G., Dicks, L.V., 2011. Quantifying the impact and relevance of scientific research. PLOS ONE 6, e27537, http://dx.doi.org/10.1371/journal.pone.0027537.

Sutherland, W.J., Woodroof, H.J., 2009. The need for environmental horizon scanning. Trends Ecol. Evol. 24, 523–527, http://dx.doi.org/10.1016/j.tree.2009.04.008.

Thompson Reuters, 2013. Thomson Reuters Hosts Challenge to Generate Ideas for Expanding the Scientific Discovery Experience on the Web of Knowledge [WWW Document], http://thomsonreuters.com/press-releases/012013/thomson_reuters_hosts_challenge_to_generate_ideas_for_expanding_the_scientific_discovery_experience_on_the_web_of_knowledge (accessed 29.08.13).

Thorisson, G.A., 2009. Accreditation and attribution in data sharing. Nat. Biotechnol. 27, 984–985, http://dx.doi.org/10.1038/nbt1109-984b.

Vetter, D., Rücker, G., Storch, I., 2013. Meta-analysis: a need for well-defined usage in ecology and conservation biology. Ecosphere 4, http://dx.doi.org/10.1890/ES13-00062.1, art74.

Webb, J.A., de Little, S.C., Stewardson, M.J., 2012a. Eco evidence database: a distributed modelling resource for systematic literature analysis in environmental science and management. In: Seppelt, R., Voinov, A.A., Lange, S., Bankamp, D. (Eds.), 2012 International Congress on Environmental Modelling and Software. Managing Resources of a Limited Planet: Pathways and Visions under Uncertainty, Sixth Biennial Meeting. International Environmental Modelling and Software Society (iEMSs), Leipzig, Germany.

Webb, J.A., Nichols, S., Norris, R., Stewardson, M., Wealands, S., Lea, P., 2012b. Ecological responses to flow alteration: assessing causal relationships with Eco Evidence. Wetlands 32, 203–213, http://dx.doi.org/10.1007/s13157-011-0249-5.

Webb, J.A., Miller, K.A., Stewardson, M.J., de Little, S.C., Nichols, S.J., Wealands, S.R., 2015. An online database and desktop assessment software to simplify systematic reviews in environmental science. Environ. Model. Softw. 64, 72–79, http://dx.doi.org/10.1016/j.envsoft.2014.11.011.

Webb, J.A., Wealands, S.R., Lea, P., Nichols, S.J., de Little, S.C., Stewardson, M.J., Norris, R.H., 2011. Eco Evidence: using the scientific literature to inform evidence-based decision making in environmental management. In: 19th International Congress on Modelling and Simulation, pp. 2472–2478.

Willett, J., Baldwin, T., Martinez, D., Webb, A., 2012. Classification of study region in environmental science abstracts. In: Proceedings of the Australasian Language Technology Association Workshop 2012, Dunedin, New Zealand, pp. 118–122.

Ziegler, C.R., 2007. Common candidate cause: flow alteration [WWW Document]. USEPA Causal Anal. Diagnosis Decis. Inf. Syst., http://www.epa.gov/caddis/ssr_flow_int.html (accessed 01.06.14).