The difficulties of systematic reviews

Martin J. Westgate [a] * & David B. Lindenmayer [a, b]

[a] Fenner School of Environment and Society, The Australian National University, Canberra ACT 2601, Australia

[b] ARC Centre of Excellence for Environmental Decisions, The Australian National University, Canberra ACT 2601, Australia

*email martin.westgate@anu.edu.au

Keywords: synthesis, bias, synonymy, meta-analysis, text analysis

Abstract

The need for robust evidence to support conservation actions has driven the adoption of systematic approaches to research synthesis in ecology. However, applying systematic review to complex or open questions remains challenging, and this task is becoming more difficult as the quantity of scientific literature increases. We drew on the science of linguistics for guidance as to why the process of identifying and sorting information during systematic review remains so labor intensive, and to provide potential solutions. Several linguistic properties of peer-reviewed corpora – including nonrandom selection of review topics, small-world properties of semantic networks, and spatiotemporal variation in word meaning – greatly increase the effort needed to complete the systematic review process. Conversely, the resolution of these semantic complexities is a common motivation for narrative reviews, but this process is rarely enacted with the rigor applied during linguistic analysis. Therefore, linguistics provides a unifying framework for understanding some key challenges of systematic review and highlights two useful directions for future research. First, in cases where semantic complexity generates barriers to synthesis, ecologists should consider drawing on existing methods – such as natural language processing, or the construction of research thesauri and ontologies – that provide tools for mapping and resolving that complexity. These tools could help individual researchers to classify research material in a more robust manner, and provide valuable guidance for future researchers on that topic. Second, a linguistic perspective highlights that scientific writing is a rich resource worthy of detailed study, an observation that can sometimes be lost during the search for data during systematic review or meta-analysis. For example, mapping semantic networks can reveal redundancy and complementarity among scientific concepts, leading to new insights and research questions. Consequently, wider adoption of linguistic approaches may facilitate improved rigor and richness in research synthesis.

Introduction

The scientific literature is growing at an increasing rate (Ferreira et al. 2015), generating a corresponding need to collate and synthesize scientific knowledge (Westgate et al. 2015). This demand has been met in the biological and environmental sciences by the development of scientifically informed methods of data synthesis (i.e., systematic reviews and meta-analyses [Pullin & Knight 2009]). These methods differ in the nature of the collated data but share a series of steps used to identify and synthesize information from the peer-reviewed and grey literatures. Combined with a wider cultural push toward evidence-based policy (Sutherland et al. 2004), systematic reviews and their derivatives have become accepted as the gold standard of research synthesis in the environmental sciences (Dicks et al. 2014; Lortie 2014).

Despite widespread acceptance, however, the adoption of systematic review by ecologists has not been without controversy. A key problem has been the need to categorize the vast quantities of literature returned by search engines at the outset of the systematic review process, which can reach several thousand articles (e.g. Lindenmayer & Laurance 2016; Westgate et al. 2013). Further, systematic reviews are poorly suited to some qualitative tasks (such as synthesis of complex concepts), which are markedly different from the closed questions systematic review was designed to answer (Collaboration for Environmental Evidence 2013). Yet the potential benefits of reduced bias and increased robustness provided by systematic methods are clear (Pullin & Stewart 2006). Therefore, two key questions are: What factors make systematic review so difficult in practice, and how can researchers move past these issues to improve synthesis of environmental information?

We drew on the science of linguistics to outline why the problems discussed above have emerged, and to provide potential solutions. A linguistic perspective is needed because all reviews focus on a common medium (i.e. the scientific and grey literature) and unit of analysis (the written word). Consequently, methods for identifying and sorting scientific material are subject to the rules of linguistics, yet the goals and training of systematic review practitioners remain those of their particular specialization (i.e. the life sciences). As a result, there has been limited discussion of how the process of scientific review is affected by the technical structure of word use in the peer-reviewed literature. We considered which linguistic concepts could most strongly influence the review process, and what their implications might be for robust review in ecology and conservation.

Linguistic properties of scientific corpora

At a fundamental level, research synthesis depends on scientists' ability to find and interpret information pertinent to their questions or field of study. However, any researcher's capacity to identify relevant material is reliant on a deep understanding of the relationships between concepts – known in linguistics as the study of semantics. For example, a fire ecologist is likely to know that the word *ash* is a result (meronym) of *fire*, which is a type (hyponym) of *disturbance*. Similarly, a botanist may be more likely to use the word *plant* as a noun (a plant grew), while a gardener may use it more frequently as a verb (to plant a tree). Conversely, identical concepts can be described using different words (synonyms), as when the verb *log* (as in logging) and *harvest* are used interchangeably in some ecological applications (see Fig. 1 for a graphical representation of these associations). All scientists have an intuitive understanding of the semantic relationships in their given field, which is why many of the previous examples will appear obvious to most readers. However, ignoring semantics can lead to nontrivial impacts on scientists' capacity to interpret and classify information.

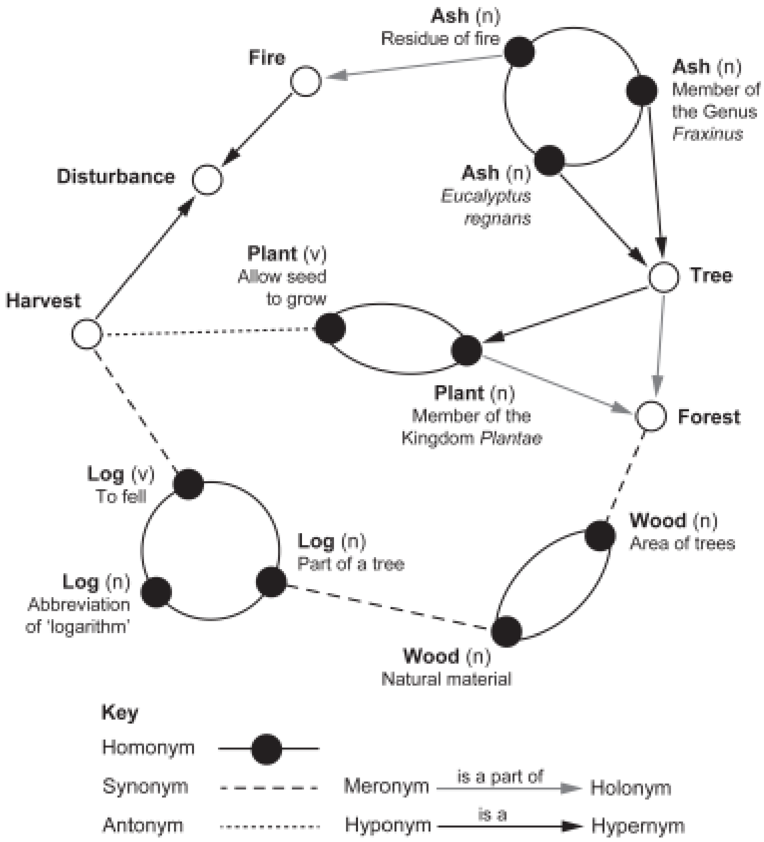

Fig. 1: Simplified semantic network showing selected associations between terms from forest ecology. For clarity, several possible homonyms have been omitted (e.g., *ash* can be a person's name), and not all connections are displayed (e.g. *Eucalyptus regnans* is a plant as well as a tree) (n, noun; v, verb).

Semantic networks have two properties that generate substantial problems for practitioners of systematic review. First, word meaning can be context dependent (the study of which is called *pragmatics*). For example, the words *forest* and *wood* are partially synonymous terms in the United Kingdom and United States (i.e., an area containing trees), but they are not synonyms in Australia. Similarly, the word used to designate a standing dead tree is *snag* in the Northern Hemisphere, but the equivalent Australian term is *stag* (Lindenmayer & Laurance 2016). This difference is important because reversing the context of these terms substantially alters their meaning: a *stag* is a male deer in North America, while a *snag* is an Australian sausage. Second, semantic networks display small-world architecture (sensu Watts

& Strogatz 1998), meaning that any two words can often be connected by only a small number of intervening concepts (Steyvers & Tenenbaum 2005). Returning to our earlier example, *ash* can be a residue of fire or (among other meanings) a tree that occurs in fire-prone environments (i.e., the mountain ash, *Eucalyptus regnans*). Therefore, the terms *ash* and *disturbance* are linked by two distinct – but equally short – semantic pathways (Fig. 1). These two properties of semantic networks (i.e. pragmatics and small-worldism) are not limited to a subset of words; rather, they are typically widespread through the corpus and can greatly increase the complexity of interpreting scientific information.

Temporal shifts in the relationship between words and their meanings are similarly common. In some cases, it may be possible simply to ignore redundant word meanings. For example, *ash* (Fig. 1) was once synonymous with *spear* (because the latter was often constructed from the former), but this synonymy is now obsolete (OED Online 2016). Unfortunately, shifts in semantics can be substantially more subtle and widespread than this. Analysis of language growth highlights that variance in the frequency of word use is reduced as publication rates increase (Petersen et al. 2012) but also that new words tend to emerge within clusters of semantically-related terms. That is, new words tend clarify the meaning of existing words that are already strongly semantically connected (Steyvers & Tenenbaum 2005). Consequently, one might expect new fields to become more semantically complex as they become more common. This is certainly true for several ecological terms that are subject to debates over terminology, including density dependence (Herrando-Pérez et al. 2012) and adaptive management (Westgate et al. 2013). Alternatively, changes in word frequency and meaning can be the deliberate outcomes of scientific research, such as when authors attempt to clarify the language on a particular topic (e.g. Nimmo et al. 2015). Regardless of the cause, changes in word meaning can confuse the reader and violate the assumption that search-term-based article identification methods will return relevant material (see "Implications for systematic review").

Although not all words are equally prone to the complex semantic associations outlined above, in practice most reviewers will encounter these problems very frequently. This is because the topics that scientists deem to be worthy of review are often linguistically non-random. Indeed, a frequently cited reason to begin a review is that a topic has developed redundant or incompatible meanings (e.g. Pulsford et al. 2016); that is, that the topic in question is semantically complex. Similarly, words that are rare in the corpus are less likely to be subject to review than common words, either because users require sufficient data for a review to be feasible, or because those topics are unlikely to be considered sufficiently well developed to warrant the effort of a full review (but see Doerr et al. 2015). Consequently, practitioners of research synthesis tend to select nodes in the semantic network (i.e. words or concepts) that are both frequently occurring and strongly connected and are therefore more difficult to review. This trend is analogous to the situation in ecology where common species have stronger ecological interactions with co-occurring taxa than rare species (Aizen et al. 2012; Poisot et al. 2015). A further challenge emerges from the fact that both the frequency with which words are used and the number of semantic connections those words possess follow power-law distributions (Zipf 1965). That is, common words are not marginally more common than rare words; rather, they are exponentially more common and therefore exponentially more difficult to summarize.

Implications for systematic review

The semantic properties of scientific corpora are particularly important during systematic review because the academic search engines used to identify relevant articles ignore semantic connections between concepts. For example, homonyms have the effect of adding irrelevant hits to search results because redundant meanings are provided by the search engine (leading to low specificity). Conversely, synonyms reduce search-term sensitivity by hindering the inclusion of relevant concepts. Homonyms cannot be readily excluded from keyword-based searches, but one can avoid problems with synonyms by identifying them early in the review process and including them as additional search terms. However, this can cause further problems if the synonym has its own homonyms. For example, a search related to forest ecology that incorporates the synonyms *forest* and *wood* might return information on carpentry as well as ecology (because wood can be a material as well as a land-use classification). Further, the small-world property of semantic networks acts in a similar way to homonymy: it increases the number of irrelevant hits returned by search engines. This occurs because pairs of words that are closely semantically linked have increased probabilities of co-occurring within the same articles. Combined with biased selection of research on semantically complex topics, these properties exponentially increase the number of tangentially related concepts returned by keyword-based searches and thus contribute to the difficulty of systematic reviews.

In extreme cases, semantic complexity can lead to situations where there is no optimal combination of sensitivity and specificity of search terms for a given topic. This conflicts with the longstanding view that search-term selection is an optimization problem, namely, of how to find the combination of keywords that balances sensitivity (identification of all relevant literature) and specificity (identification only of relevant literature [Pullin & Stewart 2006]). Instead, the semantic properties of scientific terms, combined with nonrandom selection of review topics, combine to ensure most topics will generate search terms that return large numbers of articles. In some circumstances, the advice that researchers should seek a balance between sensitivity and specificity can be achieved only by using *NOT* statements to exclude superfluous semantic connections. However, this requires huge numbers of search terms, exponentially increasing both effort and the probability of missing relevant articles by mistake. Therefore, researchers should consider that there will be some topics that are simply not amenable to efficient systematic review (i.e., that there may be no sensible set of simple search terms that will provide high sensitivity and specificity on a given topic). Instead, manual sorting of large corpora may be the only way to ensure systematic inclusion of relevant material, although this approach comes with the risk of fatigue-induced bias (Danziger et al. 2011).

Fortunately, some forms of existing data (besides keywords) can act as proxies for semantic information in the article-identification stage of the systematic review process. First, expert knowledge provides an excellent source of information on the relatedness of distinct concepts. Combined with careful reading of the literature, therefore, we suggest there is often considerable merit in discussing proposals for a new systematic review with the authors of past reviews. This can provide useful ecological context and help identify associated concepts and key literature that might otherwise be overlooked. For example, there is a large but nonoverlapping literature on both ecological traps (Schlaepfer et al. 2002) and artificial reefs (Baine 2001), although there are times when artificial reefs will act as ecological traps.

Second, checking the citation lists of relevant articles is a useful method for identifying articles lacking search keywords. This approach was recently undertaken in a major review of the factors influencing the success of forest restoration (McAlpine et al. 2016) and in a global meta-analysis of cross-taxon congruence (Westgate et al. 2016). Alternatively, articles that share an intellectual tradition are more likely to display shared semantics, so identifying the group of articles that derive from a single foundational work can be a further way to identify articles (Westgate et al. 2015), particularly where single references are identified during expert elicitation.

Finally, modern search engine algorithms (e.g. Google) already seek to incorporate users' intended meanings in their page searches, and the use of semantic information by search providers will only increase in future. One key problem is that these methods currently only rank articles, meaning they lack the advantage of current systematic review methods in which the entire list of articles is searched systematically. Instead, guidance will be needed as to the maximum number of pages users should search to trade-off comprehensiveness against realism (Haddaway et al. 2015b).

Future directions

We argue that semantic complexity can hinder systematic review, but an alternative view is that this complexity is worthy of study in its own right. From this perspective, linguistics provides not only a framework for understanding existing problems in classifying scientific literature but also gives a suite of tools that actively facilitate scientific progress. This is because a key component of scientific development is to think deeply about how terms can be used to describe nature in the clearest and most accurate way possible. Traditionally, ecologists have tackled this problem (i.e., semantic development of the scientific vocabulary) via narrative reviews, but these methods lack the rigor of systematic methods (Lortie 2014). One solution to this problem is to simply apply more robust search methods to ensure narrative reviews reliably sample the available literature (Haddaway et al. 2015a), but a broader approach is to systematically map the semantic networks within a particular area of study. This method has a long history in applied linguistics (e.g. Fellbaum 1998), and the concept of robust vocabularies is fundamental to modern information management (National Information Standards Organization 2005). Despite strong uptake of these ideas for some environmental applications (see Zhang et al. 2015), however, the potential of semantic mapping for clarifying and resolving ecological debates remains underdeveloped.

Greater adoption of semantic network mapping in ecology would build on a long tradition within the life sciences. Perhaps the best example is taxonomic research, where the nature of the connections (i.e. shared phylogeny) between concepts (species) is the primary research goal of the discipline, and rules for updating and storage of those connections are similarly robust (e.g. see ICZN 1999). More broadly, ecologists often use conceptual models to define the key components of a given system and to determine how those components relate to one another (Gentile et al. 2001). Yet the classification of ecological meanings is rarely a systematic process, despite calls for more widespread use of conceptual diagrams in ecology (Lindenmayer & Likens 2009). One example of how increased systematization might be achieved is through the creation of standard thesauri to index scientific information, which has recently been promoted as a means of facilitating data sharing and re-use in ecology (Garnier et al. 2016; Wright et al. 2015). A second useful approach is systematic mapping, which seeks to define the major research areas and points of debate relevant to a particular research question (Haddaway et al. 2016) and could readily be expanded to assess semantics. Therefore, a range of existing methods – both from within the life sciences and sourced from other disciplines – are available to increase the clarity and rigor of environmental synthesis.

Wider adoption of linguistic approaches would improve ecologists' capacity to answer existing questions in a more robust manner. For example, we have already discussed situations where the same concept is given different names in different contexts (see the example above regarding terms for standing dead trees), but the same process can occur with any ecological term or concept. Therefore, mapping semantic relations may help to reduce the number of terms applied to the same ecological phenomenon (synonymy) and identify commonalities among distinct concepts, locations, or ecosystems that might otherwise remain hidden (Driscoll & Lindenmayer 2012; Pulsford et al. 2016). Conversely, mapping pragmatic relations (i.e. differences in word usage between contexts) is a useful way to quantify the distribution and prevalence of distinct schools of thought, which can assist in understanding the process of scientific development (McFadden et al. 2011). Finally, text mining of the scientific literature can generate usable insights into the dominant trends within a corpus and reveal how these trends vary over space or time (Nunez-Mir et al. 2016; Westgate et al. 2015). Each of these approaches recognizes that science is – in part – a literary exercise and that treating it as such can increase the richness of our understanding of the scientific process.

Creating robust yet flexible ways to collate and synthesize scientific information is an ongoing challenge. We suggest linguistics provides a framework for understanding why systematic approaches to research synthesis remain challenging and a set of tools for investigating complex problems in a robust manner. Therefore, increased engagement with methods for resolving semantic disputes may facilitate improved conceptual development in ecology, continuing the development of systematic approaches for environmental synthesis.

Acknowledgments

We thank M. Burgman for encouraging us to write this essay and N. Haddaway for discussion of the properties of systematic review.

Literature Cited

Aizen, M. A., M. Sabatino, and J. M. Tylianakis. 2012. Specialization and rarity predict nonrandom loss of interactions from mutualist networks. Science **335**:1486-1489.

Baine, M. 2001. Artificial reefs: a review of their design, application, management and performance. Ocean & Coastal Management **44**:241-259.

Collaboration for Environmental Evidence. 2013. Guidelines for systematic review and evidence synthesis in environmental management. Version 4.2. Environmental Evidence, Bangor, United Kingdom. Available from http://www.environmentalevidence.org/wp-content/uploads/2014/06/Review-guidelines-version-4.2-final.pdf (accessed November 2016).

Danziger, S., J. Levav, and L. Avnaim-Pesso. 2011. Extraneous factors in judicial decisions. Proceedings of the National Academy of Sciences **108**:6889-6892.

Dicks, L. V., J. C. Walsh, and W. J. Sutherland. 2014. Organising evidence for environmental management decisions: a '4S' hierarchy. Trends in Ecology & Evolution **29**:607-613.

Doerr, E. D., J. Dorrough, M. J. Davies, V. A. J. Doerr, and S. McIntyre. 2015. Maximizing the value of systematic reviews in ecology when data or resources are limited. Austral Ecology **40**:1-11.

Driscoll, D. A., and D. B. Lindenmayer. 2012. Framework to improve the application of theory in ecology and conservation. Ecological Monographs **82**:129-147

Fellbaum, C., editor. 1998. Wordnet: an electronic lexical database. MIT Press, Cambridge, Massachusetts.

Ferreira, C., et al. 2015. The evolution of peer review as a basis for scientific publication: directional selection towards a robust discipline? Biological Reviews **91**:597-610.

Garnier, E., et al. 2016. Towards a thesaurus of plant characteristics: an ecological contribution. Journal of Ecology:in press.

Gentile, J. H., M. A. Harwell, W. Cropper Jr, C. C. Harwell, D. DeAngelis, S. Davis, J. C. Ogden, and D. Lirman. 2001. Ecological conceptual models: a framework and case study on ecosystem management for South Florida sustainability. Science of The Total Environment **274**:231-253.

Haddaway, N., P. Woodcock, B. Macura, and A. Collins. 2015a. Making literature reviews more reliable through application of lessons from systematic reviews. Conservation Biology **29**:1596-1605.

Haddaway, N. R., C. Bernes, B.-G. Jonsson, and K. Hedlund. 2016. The benefits of systematic mapping to evidence-based environmental management. Ambio:1-8.

Haddaway, N. R., A. M. Collins, D. Coughlin, and S. Kirk. 2015b. The role of Google Scholar in evidence reviews and its applicability to grey literature searching. PloS one **10** (e0138237) DOI: https://doi.org/10.1371/journal.pone.0138237

Herrando-Pérez, S., S. Delean, B. Brook, and C. A. Bradshaw. 2012. Density dependence: an ecological Tower of Babel. Oecologia **170**:585-603.

Ride, W. D. L., Cogger, H. G., Dupois, C., Kraus, O., Minelli, A., Thompson, F. C., and P. K. Tubbs. 1999. International code of zoological nomenclature. 4th edition. The International Trust for Zoological Nomenclature, London.

Lindenmayer, D. B., and W. F. Laurance. 2016. The ecology, distribution, conservation and management of large old trees. Biological Reviews: in press.

Lindenmayer, D. B., and G. E. Likens. 2009. Adaptive monitoring: a new paradigm for long-term research and monitoring. Trends in Ecology & Evolution **24**:482-486.

Lortie, C. J. 2014. Formalized synthesis opportunities for ecology: systematic reviews and meta-analyses. Oikos **123**:897-902.

McAlpine, C., et al. 2016. Integrating plant- and animal-based perspectives for more effective restoration of biodiversity. Frontiers in Ecology and the Environment **14**:37-45.

McFadden, J. E., T. L. Hiller, and A. J. Tyre. 2011. Evaluating the efficacy of adaptive management approaches: Is there a formula for success? Journal of Environmental Management **92**:1354-1359.

National Information Standards Organization 2005. Guidelines for the construction, format, and management of monolingual controlled vocabularies. NISO Press, Baltimore, Maryland.

Nimmo, D. G., R. Mac Nally, S. C. Cunningham, A. Haslem, and A. F. Bennett. 2015. Vive la résistance: reviving resistance for 21st century conservation. Trends in Ecology & Evolution **30**:516-523.

Nunez-Mir, G. C., B. V. Iannone, B. C. Pijanowski, N. Kong, and S. Fei. 2016. Automated content analysis: Addressing the big literature challenge in ecology and evolution. Methods in Ecology and Evolution **7**:1262-1272

OED Online. 2016. ash, n.1. Oxford University Press, Oxford, United Kingdom. Available from http://www.oed.com/view/Entry/11435?rskey=T0XyBH&result=2 (accessed November 2016).

Petersen, A. M., J. N. Tenenbaum, S. Havlin, H. E. Stanley, and M. Perc. 2012. Languages cool as they expand: Allometric scaling and the decreasing need for new words. Scientific Reports **2**:943.

Poisot, T., D. B. Stouffer, and D. Gravel. 2015. Beyond species: why ecological interaction networks vary through space and time. Oikos **124**:243-251.

Pullin, A. S., and T. M. Knight. 2009. Doing more good than harm—Building an evidence-base for conservation and environmental management. Biological Conservation **142**.

Pullin, A. S., and G. B. Stewart. 2006. Guidelines for systematic review in conservation and environmental management. Conservation Biology **20**:1647-1656.

Pulsford, S. A., D. B. Lindenmayer, and D. A. Driscoll. 2016. A succession of theories: purging redundancy from disturbance theory. Biological Reviews **91**:148-167.

Schlaepfer, M. A., M. C. Runge, and P. W. Sherman. 2002. Ecological and evolutionary traps. Trends in Ecology & Evolution **17**:474-480.

Steyvers, M., and J. B. Tenenbaum. 2005. The Large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. Cognitive science **29**:41-78.

Sutherland, W. J., A. S. Pullin, P. M. Dolman, and T. M. Knight. 2004. The need for evidence-based conservation. Trends in Ecology & Evolution **19**:305-308.

Watts, D. J., and S. H. Strogatz. 1998. Collective dynamics of `small-world' networks. Nature **393**:440-442.

Westgate, M. J., P. S. Barton, J. C. Pierson, and D. B. Lindenmayer. 2015. Text analysis tools for identification of emerging topics and research gaps in conservation science. Conservation Biology **29**:1606-1614.

Westgate, M. J., G. E. Likens, and D. B. Lindenmayer. 2013. Adaptive management of biological systems: a review. Biological Conservation **158**:128-139.

Westgate, M. J., A. I. T. Tulloch, P. S. Barton, J. C. Pierson, and D. B. Lindenmayer. 2016. Optimal taxonomic groups for biodiversity assessment: a meta-analytic approach. Ecography **40**: 539-548.

Wright, D. G., K. A. Harrison, and J. Watkins. 2015. Automated tagging of environmental data using a novel SKOS formatted environmental thesaurus. Earth Science Informatics **8**:103-110.

Zhang, Y., A. Ogletree, J. Greenberg, and C. Rowell. 2015. Controlled vocabularies for scientific data: Users and desired functionalities. Proceedings of the Association for Information Science and Technology **52**:1-8.

Zipf, G. K. 1965. Human behavior and the principle of least effort. Hafner, New York.