Software support for environmental evidence synthesis

Martin J. Westgate[1], Neal R. Haddaway[2], Samantha H. Cheng[3,4], Emma J. McIntosh[5], Chris Marshall[6] & David B. Lindenmayer[1]

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

[2] Mistra EviEM, Stockholm Environment Institute, Box 24218, 104 51 Stockholm, Sweden

[3] National Center for Ecology Analysis and Synthesis (NCEAS), University of California, Santa Barbara, Santa Barbara, CA, 93101 USA

[4] Center for Biodiversity Outcomes, Julie Ann Wrigley School of Sustainability, Arizona State University, Tempe, AZ, 85281 USA

[5] School of Geography and the Environment, University of Oxford, South Parks Road, Oxford, OX1 3QY, UK

[6] York Health Economics Consortium, University of York, York, YO10 5NQ, UK

Keywords: systematic review, meta-analysis, text mining, natural language processing, decision support

Ecological research is central to efforts to ensure the provision of critical societal needs such as clean water, carbon abatement[1], and to avert the loss of biodiversity[2]. The amount of research published on these subjects has increased enormously in recent years[3], yet this research is not always used to improve environmental management or policy[4]. This 'research-implementation gap' is sustained by many factors including low access to scientific research outside of academia[5], a lack of flexible decision-making structures to incorporate new information[4], and mismatches between management and scientific priorities[6]. A key step towards bridging the research-implementation gap, however, is to gather insights from the entire body of available evidence to ensure that scientific advice is as consistent and accurate as possible[2]. This requires *evidence synthesis*; work by individuals or teams that take scientific outputs (articles and reports) and use them to understand the effectiveness of an intervention in a range of contexts[7]. Consequently, applied synthesis has become indispensable to the application of scientific information to socio-ecological problems[4].

Unfortunately, evidence synthesis is becoming increasingly difficult as the scientific literature continues to expand. In medicine, for example, the average systematic review takes five people 67 weeks to conduct[8], which extrapolates to nearly 12,000 person hours. We argue that the effort needed to locate, interpret and synthesize scientific information is so great that it requires a new term: the 'synthesis gap' (Fig. 1). This gap manifests as policy-relevant information being lost amongst a sea of websites, reports and peer-reviewed articles[9,10]. If this problem is not resolved, there is significant risk of wasting effort and money by duplicating research, and failing to capitalize on substantial global investments in environmental science[10].

Evidence synthesis is now undergoing methodological changes that – if more broadly adopted – will help to close the synthesis gap, even accounting for future increases in publication rates. Developments in software support (and particularly machine learning) that enable rapid sorting of large quantities of scientific information have the potential to revolutionize the synthesis process. For example, text-mining approaches have been used to distinguish between relevant and irrelevant papers during the literature sorting process, reducing effort by between 30 and more than 90% relative to manual sorting[11,12]. Yet these methods remain rarely used. Topic models have only recently been advocated for investigating free text in ecology and evolution[13,14], for example, despite 15 years of testing in computer science[15]. This implies that substantial gaps remain in natural scientists' knowledge of what software tools are available, and how best to apply them during synthesis projects. Consequently, scientists are wasting effort, time and money on research synthesis projects that could be made cheaper and more efficient by the adoption of recent technological advancements (Table 1). Here we discuss five actions that are important to future attempts to bridge the synthesis gap.

1. Better validation of software tools

The speed with which new computational tools are being developed makes it difficult for users to determine their reliability and utility for synthesis projects[16]. This could be addressed by research to validate and compare existing software tools[12]. Research on software validation can be unglamorous[17], but is needed because there is a risk that untested approaches may introduce new forms of bias[18]. For example, the before mentioned text-mining approaches are associated with a risk of missing up to 5% of relevant studies (only 95% recall) when compared with manual screening processes[11]. Yet, most new papers tend to introduce new approaches rather than evaluate existing methods[12]. Scientists that use text mining during systematic reviews, for example, rarely report sufficient information to replicate their approach, or to evaluate software performance[19].

2. Rapid communication of novel methods

Research to validate new software tools will not reduce the synthesis gap unless it is combined with a mechanism for rapid, independent confirmation and publication of validation results. This is challenging as there are no widely agreed-upon standards for testing synthesis tools, and no organisation capable of routinely providing that service. Currently, central organisations – namely the Collaboration for Environmental Evidence, and the Campbell and Cochrane Collaborations (for social welfare and healthcare, respectively) – act as arbiters of which tools and workflows are deemed 'rigorous' for the production of systematic reviews and systematic maps[16]. These organisations are not equipped for independent validation, nor should they be expected to regulate new methods given that they are composed largely of volunteer researchers. In the short term, therefore, a practical solution may be to establish special interest groups who then become responsible for evaluating the evidence supporting (or refuting) the use of new software tools. An alternative is to rely on more flexible methods of community involvement to screen new methods. For example, directories such as the Systematic Review Toolbox[20] can be valuable for locating relevant software. Community-managed projects such as Wikipedia provide another model that could be adapted for listing software options and their relative strengths and weaknesses.

3. Broader adoption of open science principles

New software tools can maintain the rigor of evidence synthesis while reducing effort; but the continued development of these tools will require greater collaboration between developers and users. For example, the core task of sorting information into relevant and irrelevant information is highly amenable to machine learning solutions (by developers), yet the best way to validate these tools is to compare their performance against human decisions (provided by users)[21]. A properly managed evidence-synthesis process generates an enormous amount of information on the sequence of decisions that practitioners make, including not just which articles are included in the review, but also what data are contained within selected articles, and at which screening stage material is deemed irrelevant and excluded from consideration. However, there is currently no standard format for storing or sharing data of this kind. Nor is there a general appreciation of the enormous value of such data for improving research synthesis methods (such as training machine learning algorithms), despite similar information (such as search protocols or the list of final included articles) being routinely supplied during the systematic review process. Therefore, capitalizing on new technological developments could benefit enormously from more open sharing of outputs from evidence synthesis project, a process that could capitalize on existing infrastructure (such as the systematic review data repository, https://srdr.ahrq.gov).

4. Investigating 'completeness' in evidence synthesis

Systematic review guidelines typically advocate that all relevant studies must be included for the conclusions of that review to be valid[22], a condition that could hinder wider adoption of new software tools[23]. Research from healthcare has shown that the effect of a single extra study on the conclusions of a review can depend on both the statistical power of the added study[24], and the extent of inter-study variability in the specified response variable[25]. Further, there have been cases where a single new study has materially affected review outcomes (or the degree of confidence in that review)[26], suggesting that completeness of the evidence base can be important in some instances. Without new research, however, it is impossible to know whether these cases are common or rare. Therefore, we remain some way from being able to assess whether a complete census of scientific evidence is worth the effort in all instances, or conversely, whether it is ever acceptable to use simpler search protocols and risk missing some articles during evidence synthesis projects. Certainty on this question would help synthesists to make rapid, informed decisions about the effort needed to complete new reviews (or update old reviews) while accounting for tradeoffs in cost and reliability.

5. Improved article-level meta-data

In the long-term, the current system of scientific publication is highly inefficient for research synthesis, as it generates science which is inconsistently stored and indexed, meaning that later synthesis projects must expend considerable effort to locate and interpret that information. Locating scientific articles by keyword-based searching is particularly inefficient because it returns a large amount of irrelevant information[27], and this leads to enormous increases in the cost of bridging the synthesis gap[28]. Furthermore, there are limits to how much more efficient academic databases can become at locating relevant material without investment in more effective tools (such as thesauri) to navigate documents that incorporate considerable linguistic variability[29] and complexity[30]. Organisations like the Collaboration for Environmental Evidence could therefore consider advocating for change in the way articles are presented, for example by providing systems for enhanced data and metadata storage. Alternatively, there is the potential to establish open databases that collate published information in a rich yet systematic way, a goal that is already being attempted in some groups and subsets of the literature (e.g. Semantic Scholar[31]).

Accelerating evidence-based synthesis

The motivating factor behind the establishment of peer-reviewed protocols for systematic reviews in healthcare, environment and elsewhere was the need to transparently, comprehensively and repeatably synthesise evidence bases on particular policies or management actions[2]. These principles now need to be applied to the process of synthesis itself, to further entrench evidence-based practice in research synthesis. Although testing and adopting new methods will take time, it does not constitute a fundamental change in research practice, because this field has always been progressive. Further, these software tools will only become more important as the rate of scientific publication continues to increase. Indeed, low uptake of tools for locating, interpreting and classifying scientific information has been described as a major barrier to wider adoption of evidence-based conservation[5,32,33].

Practical solutions to this problem depend on wider adoption of open science principles, and a new culture of working together to build a firm evidence base for best practice in evidence synthesis.

References

1 Cardinale, B. Impacts of Biodiversity Loss. *Science* **336**, 552-553 (2012).
2 Sutherland, W. J., Pullin, A. S., Dolman, P. M. & Knight, T. M. The need for evidence-based conservation. *Trends Ecol. Evol.* **19**, 305-308, doi:10.1016/j.tree.2004.03.018 (2004).
3 Fox, C. W., Paine, C. E. T. & Sauterey, B. Citations increase with manuscript length, author number, and references cited in ecology journals. *Ecol. Evol.* **6**, 7717-7726, doi:10.1002/ece3.2505 (2016).
4 Sutherland, W. J. & Wordley, C. F. Evidence complacency hampers conservation. *Nature Ecology & Evolution* **1**, 1215 (2017).
5 Sunderland, T., Sunderland-Groves, J., Shanley, P. & Campbell, B. Bridging the Gap: how can information access and exchange between conservation biologists and field practitioners be improved for better conservation outcomes? *Biotropica* **41**, doi:10.1111/j.1744-7429.2009.00557.x (2009).
6 Toomey, A. H., Knight, A. T. & Barlow, J. Navigating the Space between Research and Implementation in Conservation. *Conserv. Lett.*, in press, doi:10.1111/conl.12315 (2017).
7 Gough, D., Oliver, S. & Thomas, J. *An introduction to systematic reviews*. (Sage, 2017).
8 Borah, R., Brown, A. W., Capers, P. L. & Kaiser, K. A. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open* **7** (2017).
9 Grubert, E. & Siders, A. Benefits and applications of interdisciplinary digital tools for environmental meta-reviews and analyses. *Environmental Research Letters* **11**, 093001 (2016).
10 McKinnon, M. C., Cheng, S. H., Garside, R., Masuda, Y. J. & Miller, D. C. Sustainability: Map the evidence. *Nature* **528**, 185-187 (2015).
11 Olorisade, B. K., de Quincey, E., Brereton, P. & Andras, P. in *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering.* 14 (ACM).
12 O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M. & Ananiadou, S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic Reviews* **4**, 5, doi:10.1186/2046-4053-4-5 (2015).
13 Nunez-Mir, G. C., Iannone, B. V., Pijanowski, B. C., Kong, N. & Fei, S. Automated content analysis: Addressing the big literature challenge in ecology and evolution. *Methods Ecol. Evol.* **7**, 1262-1272, doi:10.1111/2041-210X.12602 (2016).
14 Westgate, M. J., Barton, P. S., Pierson, J. C. & Lindenmayer, D. B. Text analysis tools for identification of emerging topics and research gaps in conservation science. *Conserv. Biol.* **29**, 1606-1614, doi:10.1111/cobi.12605 (2015).
15 Blei, D. M., Ng, A. Y. & Jordan, M. I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **3**, 993-1022 (2003).
16 Thomas, J., McNaught, J. & Ananiadou, S. Applications of text mining within systematic reviews. *Research Synthesis Methods* **2**, 1-14 (2011).
17 Mislan, K. A. S., Heer, J. M. & White, E. P. Elevating The Status of Code in Ecology. *Trends Ecol. Evol.* **31**, 4-7, doi:https://doi.org/10.1016/j.tree.2015.11.006 (2016).

18 Caliskan, A., Bryson, J. J. & Narayanan, A. Semantics derived automatically from language corpora contain human-like biases. *Science* **356**, 183-186, doi:10.1126/science.aal4230 (2017).
19 Olorisade, B. K., Brereton, P. & Andras, P. Reproducibility of studies on text mining for citation screening in systematic reviews: Evaluation and checklist. *Journal of Biomedical Informatics* **73**, 1-13, doi:http://dx.doi.org/10.1016/j.jbi.2017.07.010 (2017).
20 Marshall, C. & Brereton, P. in *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering.* 23 (ACM).
21 Chang, J., Boyd-Graber, J. L., Gerrish, S., Wang, C. & Blei, D. M. in *Nips.* 1-9.
22 Keele, S. in *Technical report, Ver. 2.3 EBSE Technical Report. EBSE* (sn, 2007).
23 Hemens, B. J. & Iorio, A. Computer-Aided Systematic Review Screening Comes of Age. *Annals of Internal Medicine* **167**, 210-211 (2017).
24 Turner, R. M., Bird, S. M. & Higgins, J. P. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. *PloS one* **8**, e59202 (2013).
25 Valentine, J. C., Pigott, T. D. & Rothstein, H. R. How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics* **35**, 215-247 (2010).
26 Garner, P. *et al.* When and how to update systematic reviews: consensus and checklist. *BMJ* **354**, doi:10.1136/bmj.i3507 (2016).
27 Gerstner, K. *et al.* Will your paper be used in a meta-analysis? Make the reach of your research broader and longer lasting. *Methods Ecol. Evol.* **8**, 777-784, doi:10.1111/2041-210X.12758 (2017).
28 Ioannidis, J. P. A. Why Most Clinical Research Is Not Useful. *PLOS Medicine* **13**, e1002049, doi:10.1371/journal.pmed.1002049 (2016).
29 Westgate, M. J. & Lindenmayer, D. B. The difficulties of systematic reviews. *Conserv. Biol.* **31**, 1002-1007 (2017).
30 Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C. & Thompson, W. H. The readability of scientific texts is decreasing over time. *eLife* **6**, e27725, doi:10.7554/eLife.27725 (2017).
31 Ziegler, C. R., Webb, J. A., Norton, S. B., Pullin, A. S. & Melcher, A. H. Digital repository of associations between environmental variables: a new resource to facilitate knowledge synthesis. *Ecol. Indic.* **53**, 61-69 (2015).
32 Ma, X., Fox, P., Narock, T. & Wilson, B. Semantic e-Science. *Earth Science Informatics* **8**, 1-3, doi:10.1007/s12145-015-0212-8 (2015).
33 Wolkovich, E. M., Regetz, J. & O'Connor, M. I. Advances in global change research require open science by individual researchers. *Glob. Change Biol.* **18**, 2102-2110, doi:10.1111/j.1365-2486.2012.02693.x (2012).
34 Haddaway, N. R., Collins, A. M., Coughlin, D. & Kirk, S. A rapid method to increase transparency and efficiency in web-based searches. *Environmental Evidence* **6**, 1, doi:10.1186/s13750-016-0079-2 (2017).
35 Rathbone, J., Carter, M., Hoffmann, T. & Glasziou, P. Better duplicate detection for systematic reviewers: evaluation of Systematic Review Assistant-Deduplication Module. *Systematic Reviews* **4**, 6, doi:10.1186/2046-4053-4-6 (2015).
36 Shemilt, I. *et al.* Pinpointing needles in giant haystacks: use of text mining to reduce impractical screening workload in extremely large scoping reviews. *Research Synthesis Methods* **5**, doi:10.1002/jrsm.1093 (2013).

37 Marshall, I. J., Kuiper, J. & Wallace, B. C. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. *Journal of the American Medical Informatics Association* **23**, 193-201, doi:10.1093/jamia/ocv044 (2016).
38 Pick, J. L., Nakagawa, S. W. & Noble, D. W. Reproducible, flexible and high throughput data extraction from primary literature: The metaDigitise R package. *bioRxiv*, doi:10.1101/247775 (2018).
39 Lajeunesse, M. J. Facilitating systematic reviews, data extraction and meta-analysis with the metagear package for r. *Methods Ecol. Evol.* **7**, 323-330, doi:10.1111/2041-210X.12472 (2016).
40 Wallace, B. C. *et al.* OpenMEE: Intuitive, open-source software for meta-analysis in ecology and evolutionary biology. *Methods Ecol. Evol.* **8**, 941-947, doi:10.1111/2041-210X.12708 (2017).

Table 1 | Emerging methods for rigorous and efficient research synthesis.

Stage	Problem	Solution
Planning	**Planning workflow:** Large numbers of software tools available, relative strengths and weakness unclear	Online databases of relevant tools[20]
Searching	**Data collection:** Organisational websites often lack convenient download functions	Web scraping[34]
	Search record extraction: Downloading information from academic databases is intentionally slow and labour-intensive	No user-based solution: provider-dependent
	Incomplete search results: Non-detection of known relevant content	Semantic analysis of key texts to locate potential search terms (synonyms)
Screening	**Duplicates:** Same content repeated many times in the dataset because of multiple databases searched	Duplicate detection algorithms[35]
	Classification: Need for overview of broad trends to ensure only relevant topics are included	Simple machine-learning approaches such as topic modelling[13,14]
	Inclusion of irrelevant material: Non-target fields or journals included in search results	Dynamic classification using machine learning[36]
	Locating full text articles: Download of full-text documents often requires manual searching and downloading	Built in to some software platforms. Limited by copyright and access issues
Synthesis	**Data Extraction:** Information located in a combination of text, tables and figures, requiring manual checking	Automated image and natural language processing[37,38]
	Meta-analysis: Appropriate statistical models, methods and workflows can be complex, particularly for new users.	Many tools available[39,40]
	Data visualization: Presenting complex data for broad audiences is difficult	Open source/access to data. Interactive diagrams, such as evidence atlases, heat maps[10], and visualizations (e.g. R Shiny)

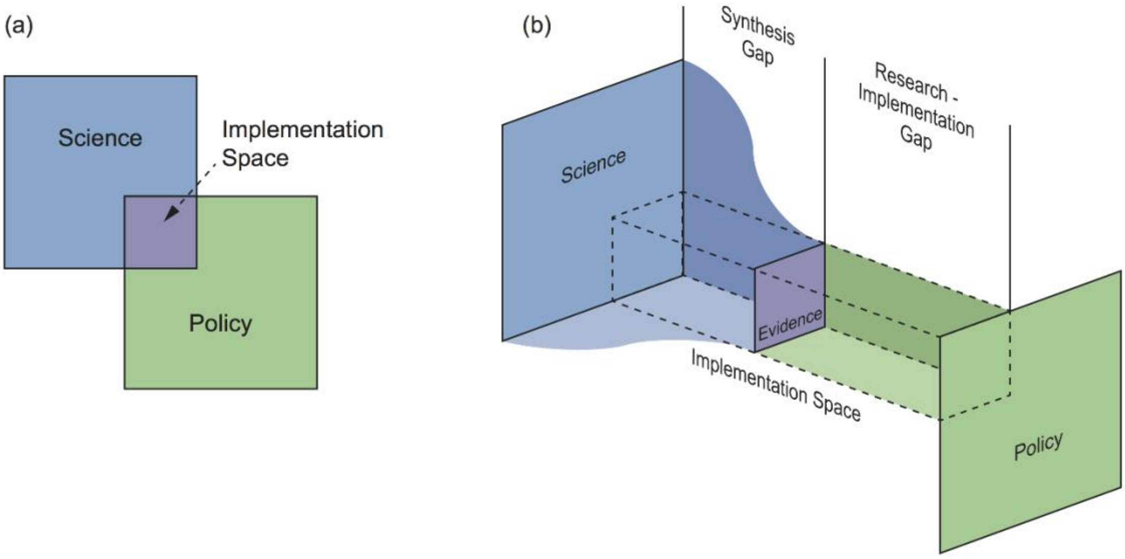

Fig. 1 | The conceptual space of the synthesis gap. A simple model of science-policy interactions might conceive of the 'implementation space' as the region where scientific information and policy concerns overlap (panel a). In practice, however, resolving poor communication between policy-makers and scientists (the research-implementation gap) depends on a process for collapsing primary scientific information into relevant evidence (the synthesis gap; panel b). This synthesis gap becomes increasingly difficult to bridge as the volume of scientific literature increases.