comment

Software support for environmental evidence synthesis

Evidence-based environmental management is being hindered by difficulties in locating, interpreting and synthesizing relevant information among vast scientific outputs. But software developments that allow enhanced collation and sharing of data will help.

Martin J. Westgate, Neal R. Haddaway, Samantha H. Cheng, Emma J. McIntosh, Chris Marshall and David B. Lindenmayer

Ecological research is central to ensuring the provision of societal needs, from carbon abatement to averting biodiversity loss[1]. The number of publications in this area has increased enormously in recent years, yet this research is not always used to improve environmental management or policy[2]. This 'research-implementation gap' is sustained by many factors, including restricted access to scientific research outside academia[3], a lack of flexible decision-making structures[4] and mismatches between management and scientific priorities[5]. A key step towards bridging the research-implementation gap is to gather insight 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 takes scientific outputs (articles and reports) and uses them to understand the effectiveness of an intervention in a range of contexts[6].

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[7]. 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 in a sea of websites, reports and peer-reviewed articles[8]. If this problem is not resolved, there is significant risk of wasting effort and money by duplicating research and failing to capitalize on global investments in environmental science[8].

Evidence synthesis is 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 the 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[9,10]. Yet these methods are still rarely used. Topic models have only recently been advocated for investigating free text in ecology and evolution[11], for example, despite 15 years of testing in computer science[12]. This implies that many scientists are unaware of the software tools that 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 for future attempts to bridge the synthesis gap.

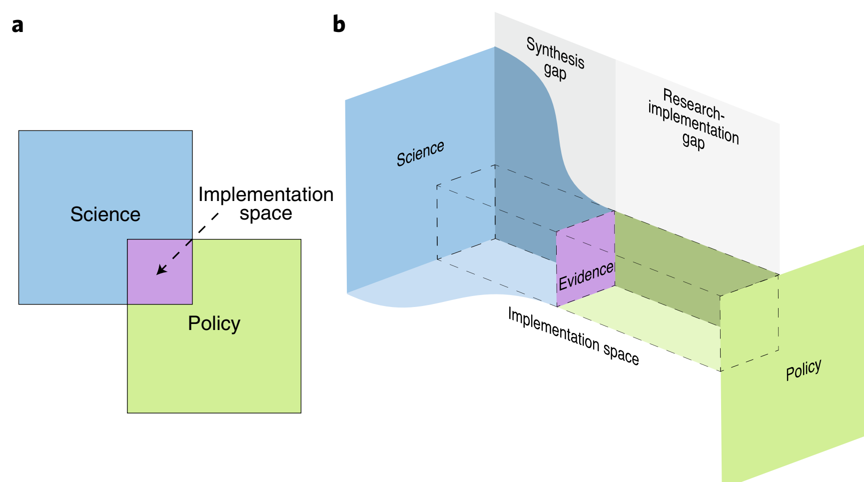

Fig. 1 | The synthesis gap. a, A simple model of science–policy interactions might present the 'implementation space' as the region where scientific information and policy concerns overlap. **b**, In practice, however, resolving poor communication between policymakers and scientists (the research-implementation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

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[13]. This could be addressed by research to validate and compare existing software tools[10]; such validation also minimizes the risk of new tools introducing new forms of bias. For example, 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[9]. Yet most papers tend to introduce new approaches rather than evaluate existing methods[10]. Scientists that use text mining during systematic reviews, for example, rarely report sufficient information to

Table 1 | Emerging methods for rigorous and efficient research synthesis

Stage	Problem	Description	Solution
Planning	Planning workflow	Large numbers of software tools are available, but their relative strengths and weaknesses are unclear	Online databases of relevant tools[15]
Searching	Data collection	Sources of 'grey literature' such as organizational websites often lack convenient download functions	Web scraping[27]
	Search record extraction	Downloading information from academic databases is slow and labour-intensive	No user-based solution: provider-dependent
	Incomplete search results	Downloading information from academic databases is slow and labour-intensive	Semantic analysis of key texts to locate additional search terms (synonyms)
Screening	Duplicates	Same content repeated many times in the dataset because of multiple databases searched	Duplicate detection algorithms[28]
	Classification	Need for overview of broad trends to ensure only relevant topics are included	Simple machine-learning approaches such as topic modelling[11]
	Inclusion of irrelevant material	Non-target subjects included in search results	Dynamic classification using machine learning[21]
	Locating full text articles	Download of full-text documents often requires manual searching and downloading	Built in to some software platforms[29]. 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[30,31]
	Meta-analysis	Appropriate statistical models, methods and workflows can be complex, particularly for new users	Many tools available[29,32]
	Data visualization	Presenting complex data for broad audiences is difficult	Open source/access to data. Interactive diagrams, such as evidence atlases, heat maps and visualizations[8]

replicate their approach, or to evaluate software performance[14].

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 organization capable of routinely providing that service. At present organizations such as 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 producing systematic reviews and systematic maps[13]. These organizations 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, a practical solution may be to establish special interest groups who 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 methods. For example, directories such as the Systematic Review Toolbox[15] 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.

Broader adoption of open science

The continued development of software tools will require greater collaboration between developers and users. For example, the core task of sorting articles into relevant and irrelevant categories 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)[16]. A properly managed evidence-synthesis process generates an enormous amount of information on the sequence of practitioners' decisions, including which articles are included in the review, 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. New technological developments could be capitalized on much more effectively with more open sharing of outputs from evidence synthesis projects, and existing infrastructure such as the systematic review data repository (https://srdr.ahrq.gov) could be used to enable this.

Testing the need for article census

Systematic review guidelines typically advocate that all relevant studies must be included for the conclusions of that review to be valid[17], a condition that could hinder the wider adoption of new software tools. 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[18], and the extent of variability in the specified response variable between studies[19]. Further, there have been cases where a single new study has materially affected review outcomes (or the degree of confidence in that review)[20], suggesting that completeness of the evidence base can be important in some instances. Without more research it is impossible to know whether these cases are common or rare, and thus whether a complete census of scientific evidence is worth the effort in all instances, or 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 informed decisions about the effort needed to complete new reviews (or update old reviews) while accounting for trade-offs in cost and reliability.

Improved article-level metadata

The current system of scientific publication is highly inefficient for research synthesis, as it generates science that is inconsistently stored and indexed. Locating scientific articles by keyword-based searching is particularly laborious because it returns a large amount of irrelevant information[21], and this leads to enormous increases in the cost of bridging the synthesis gap[22]. 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 and complexity[23]. Organizations such as the Collaboration for Environmental Evidence could 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 — Semantic Scholar is one example[24].

Conclusions

The establishment of peer-reviewed protocols for systematic reviews in the healthcare and environment fields, and elsewhere, was motivated by the need to transparently, comprehensively and repeatably synthesize the evidence bases for 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 the wider adoption of evidence-based conservation[3,25,26]. Practical solutions to this problem depend on greater uptake of open science principles, and a new culture of working together to build a firm evidence base for best practice in evidence synthesis. ❐

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

[1]Fenner School of Environment and Society, The Australian National University, Acton, Australian Capital Territory, Australia. [2]Mistra EviEM, Stockholm Environment Institute, Stockholm, Sweden. [3]NCEAS, University of California, Santa Barbara, Santa Barbara, CA, USA. [4]Center for Biodiversity Outcomes, Julie Ann Wrigley School of Sustainability, Arizona State University, Tempe, AZ, USA. [5]School of Geography and the Environment, University of Oxford, Oxford, UK. [6]York Health Economics Consortium, University of York, York, UK.

**e-mail: martin.westgate@anu.edu.au*

Published online: 23 March 2018
https://doi.org/10.1038/s41559-018-0502-x

References

1. Cardinale, B. J. et al. *Nature* **486**, 59–67 (2012).
2. Sutherland, W. J., Pullin, A. S., Dolman, P. M. & Knight, T. M. *Trends Ecol. Evol.* **19**, 305–308 (2004).
3. Sunderland, T., Sunderland-Groves, J., Shanley, P. & Campbell, B. *Biotropica* **41**, 549–554 (2009).
4. Sutherland, W. J. & Wordley, C. F. *Nat. Ecol. Evol.* **1**, 1215–1216 (2017).
5. Toomey, A. H., Knight, A. T. & Barlow, J. *Conserv. Lett.* **10**, 619–625 (2017).
6. Gough, D., Oliver, S. & Thomas, J. *An Introduction to Systematic Reviews* (Sage, London, 2017).
7. Borah, R., Brown, A. W., Capers, P. L. & Kaiser, K. A. *BMJ Open* **7**, e01254 (2017).
8. McKinnon, M. C., Cheng, S. H., Garside, R., Masuda, Y. J. & Miller, D. C. *Nature* **528**, 185–187 (2015).
9. Olorisade, B. K., de Quincey, E., Brereton, P. & Andras, P. A critical analysis of studies that address the use of text mining for citation screening in systematic reviews. In *Proc. 20th International Conference on Evaluation and Assessment in Software Engineering* 14 (ACM, New York, 2016).
10. O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M. & Ananiadou, S. *Syst. Rev.* **4**, 5 (2015).
11. Westgate, M. J., Barton, P. S., Pierson, J. C. & Lindenmayer, D. B. *Conserv. Biol.* **29**, 1606–1614 (2015).
12. Blei, D. M., Ng, A. Y. & Jordan, M. I. *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
13. Thomas, J., McNaught, J. & Ananiadou, S. *Res. Synth. Methods* **2**, 1–14 (2011).
14. Olorisade, B. K., Brereton, P. & Andras, P. *J. Biomed. Inform.* **73**, 1–13 (2017).
15. Marshall, C. & Brereton, P. Systematic review toolbox: a catalogue of tools to support systematic reviews. In *Proc. 19th International Conference on Evaluation and Assessment in Software Engineering* 23 (ACM, New York, 2015).
16. Chang, J., Boyd-Graber, J. L., Gerrish, S., Wang, C. & Blei, D. M. in *Advances in Neural Information Processing Systems* Vol. 22. (eds Bengio, Y., Schuurmans, D., Lafferty, J. D., Williams, C. K. I. & Culotta, A.) 288–296 (Curran Associates, Red Hook, 2009).
17. Kitchenham, B. & Charters, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering* EBSE Technical Report Version 2.3 (EBSE, 2007).
18. Turner, R. M., Bird, S. M. & Higgins, J. P. *PLoS ONE* **8**, e59202 (2013).
19. Valentine, J. C., Pigott, T. D. & Rothstein, H. R. *J. Educ. Behav. Stat.* **35**, 215–247 (2010).
20. Garner, P. et al. *BMJ* **354**, i3507 (2016).
21. Shemilt, I. et al. *Res. Synth. Methods* **5**, 31–49 (2013).
22. Ioannidis, J. P. A. *PLoS Med.* **13**, e1002049 (2016).
23. Westgate, M. J. & Lindenmayer, D. B. *Conserv. Biol.* **31**, 1002–1007 (2017).
24. Ziegler, C. R., Webb, J. A., Norton, S. B., Pullin, A. S. & Melcher, A. H. *Ecol. Indic.* **53**, 61–69 (2015).
25. Ma, X., Fox, P., Narock, T. & Wilson, B. *Earth Sci. Inform.* **8**, 1–3 (2015).
26. Wolkovich, E. M., Regetz, J. & O'Connor, M. I. *Glob. Change Biol.* **18**, 2102–2110 (2012).
27. Haddaway, N. R., Collins, A. M., Coughlin, D. & Kirk, S. *Environ. Evid.* **6**, 1 (2017).
28. Rathbone, J., Carter, M., Hoffmann, T. & Glasziou, P. *System. Rev.* **4**, 6 (2015).
29. Lajeunesse, M. J. *Methods Ecol. Evol.* **7**, 323–330 (2016).
30. Marshall, I. J., Kuiper, J. & Wallace, B. C. *J. Am. Med. Inform. Assoc.* **23**, 193–201 (2016).
31. Pick, J. L., Nakagawa, S. W. & Noble, D. W. Preprint at https://doi.org/10.1101/247775 (2018).
32. Wallace, B. C. et al. *Methods Ecol. Evol.* **8**, 941–947 (2017).

Competing interests

The authors declare no competing interests.