10-2022

Beyond Academia: A case for reviews of gray literature for science-policy processes and applied research

Yuki Yoshida
National Institute for Environmental Studies, Japan, yoshida.yuki@nies.go.jp

Nadia Sitas
Stellenbosch University, nadiasitas@gmail.com

Lelani Mannetti
Georgia State University, lmannetti@gsu.edu

Gabriela Arroyo-Robles
Universidad Nacional Autónoma de México

Marta Berbés-Blázquez
University of Waterloo, marta.berbes@uwaterloo.ca

See next page for additional authors

Follow this and additional works at: https://scholarworks.gsu.edu/urban_studies_institute

Part of the Urban Studies and Planning Commons

Recommended Citation

Authors

Yuki Yoshida, Nadia Sitas, Lelani Mannetti, Gabriela Arroyo-Robles, Marta Berbés-Blázquez, David Gonzalez Jimenez, Valerie Nelson, Aidin Niamir, and Zuzana V. Harmáčková

Beyond Academia:

A case for reviews of gray literature for science-policy processes and applied research

Yuki Yoshida[1], Nadia Sitas[2], Lelani Mannetti[3], Patrick O'Farrell[4], Gabriela Arroyo-Robles[5], Marta Berbés-Blázquez[6], David González-Jiménez[5], Valerie Nelson[7], Aidin Niamir[8], Zuzana V. Harmáčková[9,10]

Abstract

Gray literature is increasingly considered to complement evidence and knowledge from peer-reviewed literature for science-policy processes and applied research. On the one hand, science-policy assessments need to both consider a diversity of worldviews, knowledge types and values from a variety of sectors and actor groups, and synthesize policy-relevant findings that are salient, legitimate and credible. On the other hand, practitioners and scholars conducting applied research, especially in environmental and health-related fields, are affected by the time lag and documented biases of academic publication processes. While gray literature holds diverse perspectives that need to be integrated in science-policy processes as well as practical evidence unfiltered by commercial publication processes, its heterogeneity has made it challenging to access through conventional means for a literature review.

This paper details one endeavor within the Values Assessment of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) to review gray literature using Google's Programmable Search Engine. In the absence of a standardized approach, we build on a limited experiential knowledge base for reviewing gray literature and report on the potential applicability of our strategy for future reviews. Our results contrast the findings of our parallel review of academic literature, underlining the importance of mobilizing different knowledge bases in science-policy assessments, evidence-based practices, and applied research.

Keywords: Grey literature, Literature review, Transdisciplinary research, Methodology, Google Programmable Search Engine, Knowledge base

[1] National Institute for Environmental Studies, Japan

yoshida.yuki@nies.go.jp

[2] Centre for Sustainability Transitions, Stellenbosch University, South Africa

[3] Urban Studies Institute, Andrew Young School of Policy Studies, Georgia State University, USA

[4] Fitzpatrick Institute of African Ornithology, DST/NRF Centre of Excellence, Department of Biological Sciences, University of Cape Town, Rondebosch, South Africa

[5] National Autonomous University of Mexico, Mexico

[6] School of Planning and Faculty of Environment, University of Waterloo, Canada

[7] Livelihoods and Institutions Department, Natural Resources Institute, University of Greenwich, United Kingdom

[8] Senckenberg Biodiversity and Climate Research Institute, Frankfurt am Main, Germany

[9] Global Change Research Institute of the Czech Academy of Sciences, Brno, Czechia

[10] Stockholm Resilience Centre, Stockholm University, Sweden

Introduction

Current sustainability challenges require transformative approaches informed by multiple knowledge systems (Fazey et al. 2020). Ever growing interest from scholars, and increasingly policy makers, in assessing the environmental, economic, and social impacts of biodiversity loss and ecosystem degradation call for diverse perspectives and transdisciplinary approaches (Balvanera et al. 2020). Collective insight from science, policy, and practice as well as the arts, Indigenous and local perspectives can provide a better understanding of how societies relate to nature, and the state of nature and nature's contribution to people, offering options for action (Tengö et al. 2017; Pereira et al. 2019; Wyborn et al. 2019). Knowledge co-production is often viewed as a possible avenue in mobilizing findings from such collective insight since co-production bridges academic disciplines and social spheres (Zurba et al. 2021). But synthesizing existing knowledge on a specific theme relevant across such multiple stakeholder or actor groups remains challenging. Predominantly, peer-reviewed, or so-called academic publications, are considered when assessing frameworks and actionable approaches, with reviews and syntheses providing useful guidance for policy and decision-making. Yet a gap in mobilizing accessible , actionable, and equitable science - intercultural collaborations that bring all relevant knowledge to bear (Lazrus Id et al. 2022) - remains, possibly due to the inability of academic literature to pick up on richer and more nuanced material. For this reason, we complemented a systematic review of academic literature with a gray literature review in an assessment of values underlying projected futures (IPBES 2022a).

Limitations of academic publications have increased interest in gray literature as a valuable complement (Table 1). Notably, academic journals are more likely to publish "positive" results that are confirmative and have statistical significance (Dwan et al. 2008; Fanelli 2012; Scheel et al. 2021). This also means that certain study designs (e.g. smaller samples and effect sizes) are less likely to get published (Hopewell et al. 2007), and that researchers tend to publish only on supported hypotheses (selective reporting) (Cairo et al. 2020). This widespread practice raises concerns for Type I errors - the erroneous rejection of the null hypothesis - and inflated effect size estimates in various disciplines. A "replication crisis" has resulted, whereby established scientific results cannot be replicated (Cairo et al. 2020; Lange and Brick 2021; West and Bergstrom 2021). Further editorial biases by aspects of authors or their affiliations such as country development (Yousefi-Nooraie et al. 2006; Ekmekci 2017), as well as systemic power

imbalances such as Anglocentrism exist (Lynch et al. 2021). The peer-review process also creates a time lag in publication (Adams et al. 2017; Paez 2017), making academic publications largely unresponsive to current and developing situations.

		Peer-reviewed literature	Gray literature
Stakeholders	Primary authors/ contributors	Academics, experts in the field	Professionals, practitioners, planners, NGOs
	Primary target audience	English-reading academics	General public, policy makers, practitioners
Process	Timeliness	Slower (years)	Faster (months)
	Publication bias	Susceptible to biases towards positive results, and by author or affiliate characteristics e.g. country of origin.	Reports statistically nonsignificant results
	Quality	Blind peer-review: vetted by editors and reviewers, therefore regarded as having a consistent high quality	Varied
Format	Type of information/data	Analytical, synthesized	Practical, policy-relevant formats
	Structure	Standardized according to journal specifications and scholarly conventions	Varied, tailored to audience
	Language	Scientific, technical, prone to jargon	Varied, typically appropriate for general audience
Availability	Indexing	Scholarly databases	Limited or none
	Bibliography/ Referencing	Yes	Varied
	Access	Limited, protected by paywalls	Free

Table 1 Typical characteristics of peer-reviewed and gray literature (authors' interpretation)

Gray literature is most often defined as documents of various formats that are protected by intellectual property rights but "not controlled by commercial publishers i.e. where publishing is not the primary activity of the producing body" (p.12, Schöpfel, 2012), or those distributed through non-commercial channels (Saleh et al. 2014). In comparison to academic publications, gray literature tends to present a broader base of information and perspectives and may better

address concerns of non-academic stakeholders (Mahood et al. 2014; Adams et al. 2017; UBC Library 2022). Relevant information, such as practical, experiential details on the execution of studies, may be left out of journal articles due to word count limitations but be found in practitioner reports and consultancy documents (Adams et al. 2017; Collaboration for Environmental Evidence 2021). As such, researchers, too, have concluded that the added value of gray literature outweighs their possible lack of replicability and recommend their inclusion in reviews, in particular to validate, complement, and contextualize results of academic literature reviews (Benzies et al. 2006; Adams et al. 2017).

The degree to which gray literature has been taken up as evidence varies greatly from field to field. Positive result bias and extensive time lags in peer-review publication is of particular concern for evidence-based practices, such as the health sector. The Methodological Expectations of Cochrane Intervention Reviews, considered the medical field's "gold standard" for systematic reviews (Godin et al. 2015), explicitly recommend the review of relevant gray literatures "in order to reduce the risk of publication bias and to identify as much relevant evidence as possible (Higgins et al. 2021)". Accordingly, there are guides (Balshem et al. 2013; UBC Library 2022) and dedicated repositories (CADTH 2022) that are not as readily available in other fields. A similar guideline on searching for evidence from the Collaboration for Environmental Evidence (Collaboration for Environmental Evidence 2021) outlines inclusion of gray literature as its first standard: "searches for studies reporting non-significant results (most probably found in gray literature and studies in languages other than English) should be conducted in all systematic reviews and maps (Leimu and Koricheva 2005)." However, a recent review found a general lack of such practices in the field (Konno and Pullin 2020). Nonetheless, regional repositories exist for Australia (National Library of Australia 2022) and Europe (Inist-CNRS 2022), while the social sciences and public policy typically draw a lot of their evidence from gray literature (Curtin University Library).

Strategy	Examples	Strengths	Weaknesses	Sample past reviews	Use in present study
Databases, repositories	library catalogs, repository searches, conference proceedings, dissertation catalogs	Systematic (convenience, transparency, replicability)	Limited availability (geographic/topical scope). Limited search functions (e.g. title search only)	Godin et al. 2015; Mahood et al. 2014; Saleh et al. 2014	Not available
Generalized search engines	Google scholar, Google, DuckDuckGo	Availability, exhaustiveness, extensive search functions (e.g. Booleans, wildcards)	Dominated by academic findings. Can be biased by personalization e.g. browser version, location, past searches, institutional access. Ranked by popularity.	Haddaway et al. 2015; Mahood et al. 2014	Search results dominated by academic papers
Customized search engine	Google Programmable Search Engine	Systematic (convenience, transparency, replicability), availability, extensive search functions New: specification of target country or region, ranking based on specification	Dependent on design. Not all search results accessible or fully displayed.	Godin et al. 2015	**Applied**: both pre-existing and self made topical search engines
Expert knowledge	Consultations, requests of practitioners, policy experts	Availability	Variability; not systematic or replicable; requires careful planning	Godin et al. 2015; Mahood et al. 2014	**Applied**: Assessment authors and external reviewers
Manual search of organizational websites	Targeted websites	Availability	Variability; Not systematic or replicable; labor intensive	Godin et al. 2015; Haddaway et al. 2015	Reviewed those identified by expert knowledge
Snowballing from reference lists		Availability	Variability; Not systematic or replicable; labor intensive	Mahood et al. 2014	**Applied**: to documents identified by other means

Table 2. Review strategies for gray literature with reflections on current application

Researchers characterize reviews of gray literature as time and resource consuming in comparison to conventional reviews of academic literature (Mahood et al. 2014). Gray literature is transient and not widely indexed. Content is created and replaced at a much faster pace, sometimes in a matter of hours (Adams et al. 2017). Websites change, and titles can be catchy but misleading (Godin et al. 2015). Where databases do exist, many lack the search refinement functions of their academic counterparts (Mahood et al. 2014). In part, the documents themselves are less amenable to conventional archiving or analysis by means of data extraction and integration (Table 1; Adams et al. 2017). The absence of standards (e.g. abstracts, IMRaD format, word limits) imposed by commercial publishers also renders them more heterogeneous and often requires the entire document to be reviewed (Benzies et al. 2006; Mahood et al. 2014; Adams et al. 2017).

Past studies showcase various approaches to identifying relevant gray literature: technical or specialist online databases, generalized search engines, and reliance on expert knowledge through consultations or requests to practitioner and policy experts are prevalent (Table 2). However, the low number of documents yielded despite the use of multiple review strategies, and the ineffectiveness of conventional, systematic review strategies, demonstrate the difficulty of the review. Godin et al. (2015), despite combining a preexisting customized search engine, manual searches, and expert consultations, found only 15 (5%) of 302 identified gray literature publications met their criteria. None of these were found using the customized search engine. Similarly, Mahood et al. (2014) reviewed over 2,000 publications to identify 19 gray and 33 peer-reviewed publications. Additionally, while academic publications tend to be identifiable through multiple channels (e.g. databases), gray literature is often only found through a single source or channel, and through non-search means (Konno and Pullin 2020). Google Scholar reportedly listed gray literature around page 20-30 of search results in reviews of environmental science but missed "vital" information that could be found on organizational websites (Haddaway et al. 2015). In addition to being more time and effort-intensive, the search strategy for gray literature requires greater "care and creativity" on the part of the investigator (Mahood et al. 2014).

The gray literature review outlined in this paper contributed to the Values Assessment Report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), tasked with the assessment of diverse conceptualizations of multiple values of

nature and its contributions to people (IPBES 2022b). More specifically, it comprised a review of the state of knowledge on the types of values that underlie projected futures regarding nature, nature's contributions to people and good quality of life (Harmackova et al. 2020). While many global change assessments typically focus on assessing and synthesizing evidence from peer-reviewed, scientific journals, this does not account for the multiplicity of ways in which human-nature relationships are connected and shaped over time. The Assessment's thematic focus, in the context of increasing recognition that science-policy assessments need to consider a diversity of worldviews, knowledge types and values (Pereira et al. 2019; Lynch et al. 2021), meant it was essential to consider diverse modes of knowledge and associated plural visions of the future which are sometimes not articulated in a scientific context. Additionally, the IPBES Values Assessment applied a definition of scenarios that encompassed diverse futures, including scenarios, visions and outlooks. As such, we conducted a conventional review of peer-reviewed literature but supplemented it with a review of gray literature (the focus of this paper), arts-based materials, and materials based on Indigenous and local knowledge.

While much of the past gray literature reviews aimed at informing evidence-based practices, our review aimed to identify perspectives of non-academic stakeholder groups. In past reviews, conference proceedings and papers, theses, and working papers were the most frequently cited gray material (Adams et al. 2017). These materials help to address publication biases but originate in the academic community and thus failed to meet our needs. Further, despite calls for guidelines(Lynch et al. 2021), there is an absence of a standardized approach to reviewing gray literature. We hence report our approach and results to build upon the limited experiential knowledge base on reviewing gray literature, especially within the context of applied research and science-policy processes that are increasingly acknowledging the importance of mobilizing, assessing and synthesizing multiple evidence bases.
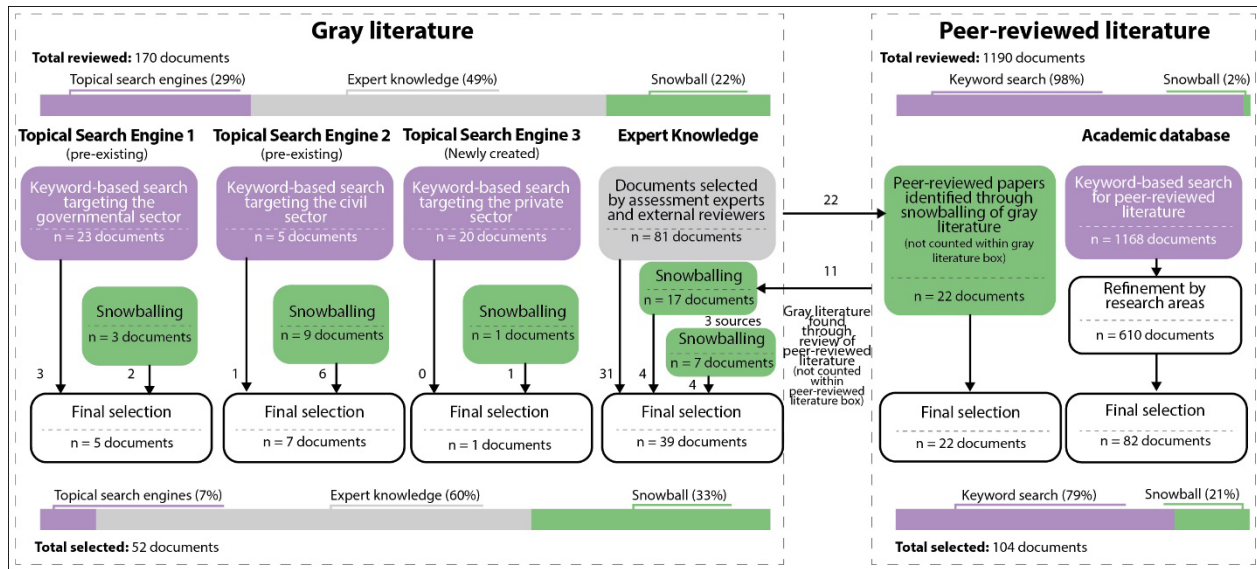
Figure 1 Overview of search methodology and outcomes

Methodology

This report details the methods of our review of gray literature, drawing on the review of peer-reviewed literature for comparison. Our primary strategy for reviewing peer-reviewed literature was a keyword search of an academic search engine (Clarivate Analytics Web of Science), using an extensive search term with Boolean operators and wildcards, as well as filter functions based on research area (Harmackova et al. 2020). This approach identified 1168 studies from the academic literature addressing our topic from diverse fields, out of which 82 were selected and yielded 257 future scenarios (Harmackova et al. 2020). As detailed in Table 2, we explored various strategies for complementing this systematic review with gray literature before finding that Google's Programmable Search Engines (formerly: Custom Search Engines) provided adequate levels of relevance and searchability for our purpose. Programmable Search Engines are often used by web developers to search within a host website, but can also be developed as a standalone, topical search engine that searches a specified list of web domains (Google 2022a) We found pre-developed engines that addressed relevant governmental and civil sector (NGO) websites, and developed our own to target the private sector (Search engine ID: 007984709256071293844:pgympkm2sah).

Three topical engines were employed and search outcomes augmented with snowball sampling (Figure 1). The search engines targeted the 1) governmental, 2) civil, and 3) private

sectors, and searched, respectively, domains of: 1) 415 International Governmental Organizations (American Library Association 2022), 2) 1,584 Nongovernmental Organizations (American Library Association 2022), and 3) the top 100 Fortune Global 500 companies (Fortune Media IP Limited 2019). The extensive keyword string that was utilized for the search of an academic database was applied, but shortened for the 3) private sector search engine, as the full set returned only one search result (Harmackova et al. 2020). The review also drew on expert knowledge of the assessment authors and external reviewers.

Search results and expert recommendations were first filtered for topical relevance, and then more closely assessed according to the review criteria: did the document contain a future vision or scenario with explicit or implicit underlying values and their implications on nature, nature's contributions to people and good quality of life? Relevant citations in the documents were also followed and assessed in the same manner.

Given the supplementary role of this review, search results were considered with the aim of identifying a sample of relevant documents from the respective sectors. More or less the top 20 governmental and private sector search results were reviewed, while only the top 5 NGO search results were reviewed, since many relevant citations were already identified in the first few search results, resulting in a larger sample for this sector. All expert recommendations were considered, in part because, similar to an academic peer-review, all external reviewer comments needed to be addressed according to IPBES assessment protocols.

For the IPBES Values Assessment's review of types of values underlying different potential futures, scenarios within selected documents were cataloged for further analysis (Harmackova et al. 2020). The database includes both academic and gray literature and encompasses descriptive characteristics of the scenarios (e.g. time horizon), their development (e.g. stakeholders involved), actions (e.g. decision scale), outcomes (e.g. for nature), and implicit or explicit values. Funding sources were newly cataloged for this report. Chi-Square Tests of Independence were conducted to assess the relationship between the literature type and various thematic attributes. Effect sizes were interpreted according to Cohen (1988).

Results

Review methods and outcomes

Figure 1 provides an overview of the review queries. In total, $n_{TSE} = 48$ documents were identified by the topical search engines ($n_{TSE1}=23$, $n_{TSE2}=5$, $n_{TSE3}= 20$). Four ($n_{TSE1}=3$, $n=1$, $n_{TSE3}= 0$; 8%) of these met our criteria. An additional 9 documents ($n_{TSE1}=2$, $n_{TSE2}=6$, $n_{TSE3}= 1$) were identified while reviewing these search results and met our criteria. Of $n_{EK}= 81$ documents identified through expert knowledge, 31 (38%) met our review criteria. An additional 24 gray literature documents, including 11 that were identified during the review of peer-reviewed literature, were identified in the review process, resulting in a total of 37 snowballed documents ($n_{TSE1}=3$, $n_{TSE2}=9$, $n_{TSE3}= 1$, $n_{EK}= 24$). Of the 52 gray literature documents that ultimately met our criteria, 4 (8%) were directly identified by topical search engines, 39 (75%) did not rely on any search engine, and 17 (33%) were identified through snowballing from search results or documents selected by experts.

Our review of peer-reviewed literature utilized an academic database and augmented results with peer-reviewed documents identified through the search for gray literature ($n=22$) (Figure 1). The overview illustrates increased efficiency of the review due to database search refinement functions. The selection rate of documents identified by keyword-based search improved from 7% before limiting the search to relevant research areas, to 13%.

Review content

The following Figures depict some of the review outcomes of gray and peer-reviewed literature. Evident contrasts are found across the board: 1) sectoral associations of the document, 2) the way the term "value" was understood, and 3) procedural (stakeholder involvement) and 4) descriptive (e.g. scale, methods) aspects of the scenarios within the documents.

1) Sectoral associations

Unsurprisingly, the vast majority (99%) of peer-reviewed literature originated in academia. In contrast, 85% of gray literature originated elsewhere: science policy interface (17%), policy-making and support, such as governments and NGOs (33%), and private sector (35%) (Figure 2a; $X^2(3, N = 156) = 118.34$, $p \leq .001$, Cramer's $V = .871$: large effect size).

Figure 2. Sectoral Associations.
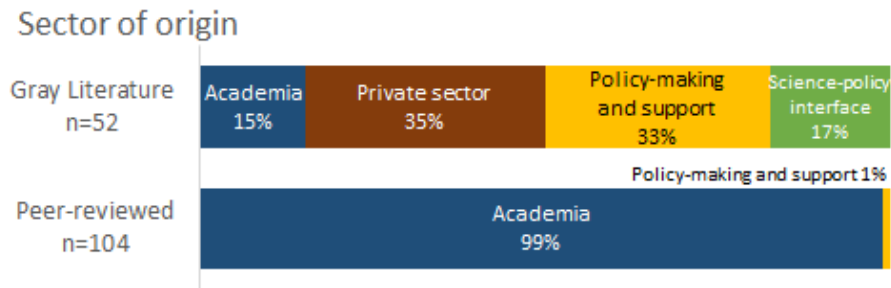
**Figure 2a. Sector of origin.** The majority of gray literature originated outside of academia.
*Policy-making and support: governments, intergovernmental organizations, NGOs
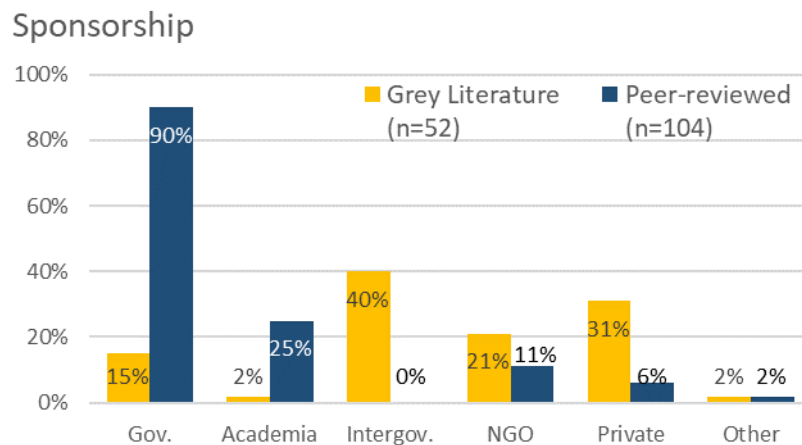
Figure 2b. Sponsorship. Sponsorship of gray literature publications are relatively dispersed, while peer-reviewed publications were overwhelmingly funded by government.

Sponsorship for the vast majority (90%) of academic publications was linked to a federal government (Figure 2b), typically in the form of research grants (acknowledged in 82% of peer-reviewed publications). In contrast, the sponsor or commissioner of gray literature publications varied amongst intergovernmental bodies (40%; X^2(1, N = 156) = 48.533, $p \leq$ .001, Cramer's V = .558), private sector entities (31%; X^2(1, N = 156) = 20.057, $p \leq$ .001, Cramer's V = .359), non-governmental organizations such as charities (21%; X^2(1, N = 156) = 3.962, p = .047, Cramer's V = .159) as well local levels of government (15%; X^2(1, N = 156) = 83.251, $p \leq$ .001, Cramer's V = .731). While most (62%) peer-reviewed publications acknowledged one or more funding bodies, sponsorship of the majority (69%) of gray literature publications was attributed to a partnership, such as between an NGO and a governmental body, or between countries, such as by

an intergovernmental body (Figure 2c; X^2(2, N = 156) = 96.81, $p \leq .001$, Cramer's V = .788: large effect size).
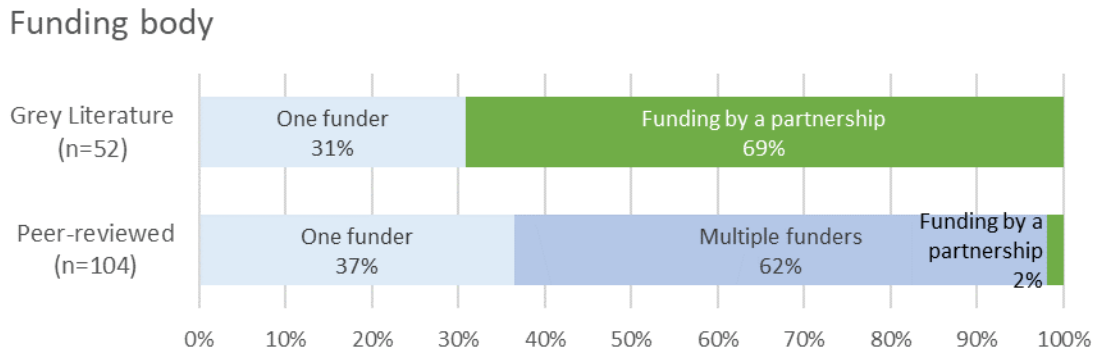
Funding body

Figure 2c. Funding body. The majority of gray literature was funded through a body of multiple organizations.

2) Terminology

Values, the central theme of the IPBES Values Assessment, was understood differently by the two literature groups (Figure 3). We surveyed the literature for any of four definitions of the term: as a 1) principle or core belief, 2) preference for something or for a particular state of the world, 3) the importance of something for itself of for other things, and 4) a measure, such as the monetary value of something or the number of species (IPBES 2015). The majority (60%) of gray literature referred to values as 1) principles or core beliefs, while this was by far the least used definition in the peer-reviewed literature. Gray literature was 9.49 times more likely to refer to values as a principle than peer-reviewed literature was (X^2(1, N = 156) = 38.54, $p \leq .001$, Cramer's V = 0.497: large effect size). There was little to no statistical significance in the difference in the usage of values as 2) preference (X^2(1, N = 156) = 4.16, p = 0.041, Cramer's V = 0.163), 3) importance (X^2(1, N = 156) = 0.214, p = 0.644), or 4) measure (X^2(1, N = 156) = 2.17, p = 0.141).
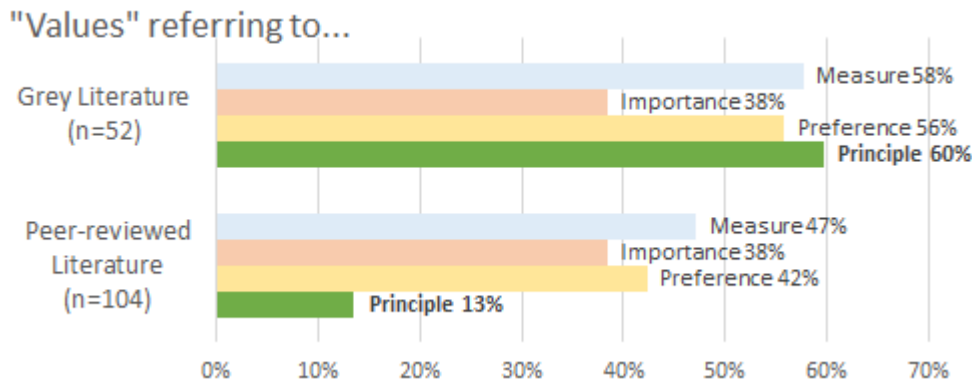
Figure 3. Terminology. "Value" was most often understood as "principle" in gray literature, and least understood as "principle" in peer-reviewed literature.

3) Stakeholder involvement

The scenario development process also differed between the literatures. All of the gray literature involved at least one stakeholder other than the author (team) in their scenario development, while a third (36%) of the peer-reviewed literature did not (Figure 4a; X^2(1, N = 156) = 24.252, $p \leq .001$, Cramer's V = .394: medium to large effect size). Peer-reviewed literature was more likely to involve individuals (Figure 4b; X^2(1, N = 156) = 8.05, p = 0.005, Cramer's V = .227), while gray literature was more likely to involve businesses or firms (X^2(1, N = 156) = 6.99, p = 0.008, Cramer's V = .21) and other stakeholders (X^2(1, N = 156) = 9.88, p = 0.002, Cramer's V = .252). Additionally, peer-reviewed literature was 3.46 times more likely to include indigenous and local knowledge than gray literature was (Figure 4c; X^2(1, N = 156) = 6.23, p = .013, Cramer's V =.200). There were no statistical differences in the proportion of documents that involved households (X^2(1, N = 156) = 0.000, p = 1.000), communities and organized groups (X^2(1, N = 156) = 1.84, p = 0.175), governments and authorities (X^2(1, N = 156) = 0.35, p = 0.554), or unspecified stakeholders (X^2(1, N = 156) = .33, p = 0.565).

Figure 4. Stakeholder involvement.

Stakeholder Involvement

Grey Literature

Peer-reviewed Literature

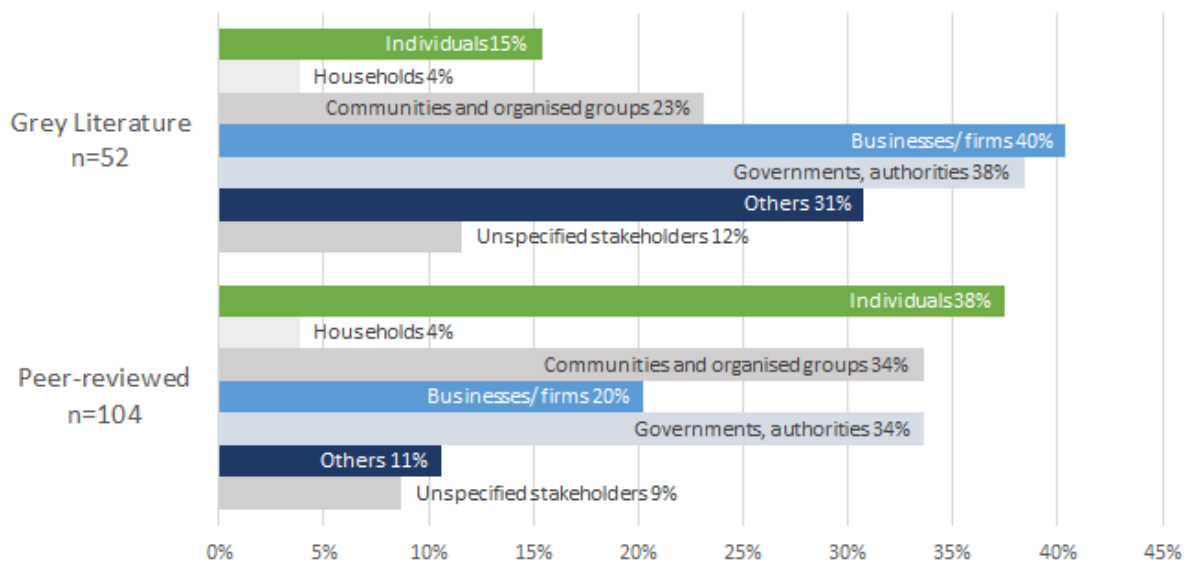■ No stakeholder involvement   ■ Stakeholders involved
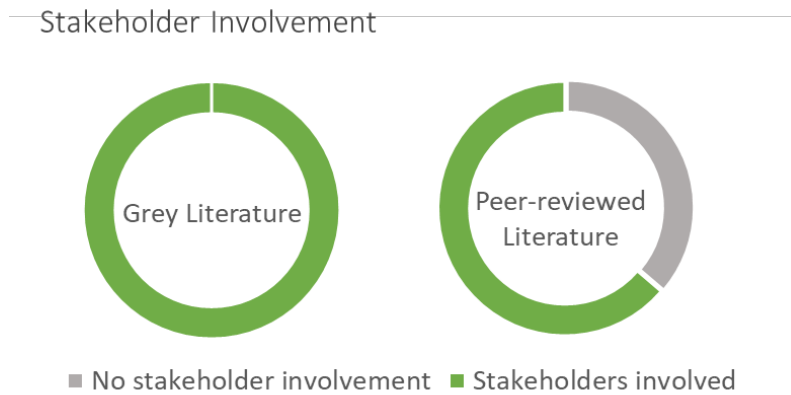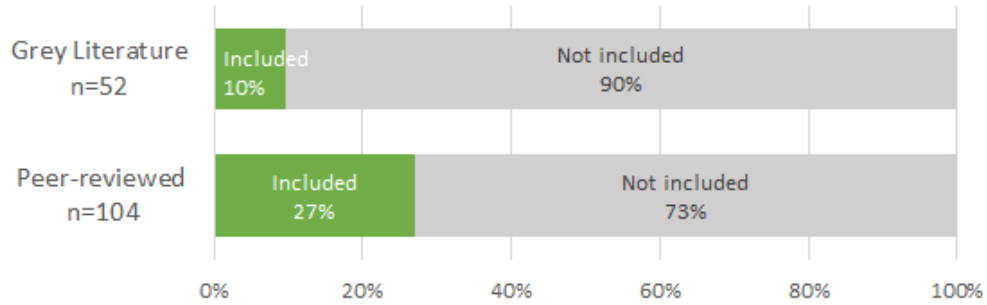
Figure 4a. Stakeholder involvement. Stakeholders other than the author(s) were involved in all of the gray literature.

Stakeholder type

Grey Literature n=52
- Individuals 15%
- Households 4%
- Communities and organised groups 23%
- Businesses/firms 40%
- Governments, authorities 38%
- Others 31%
- Unspecified stakeholders 12%

Peer-reviewed n=104
- Individuals 38%
- Households 4%
- Communities and organised groups 34%
- Businesses/firms 20%
- Governments, authorities 34%
- Others 11%
- Unspecified stakeholders 9%

Figure 4b. Stakeholder type. Types of stakeholders involved in the scenario development process differed between the literatures.
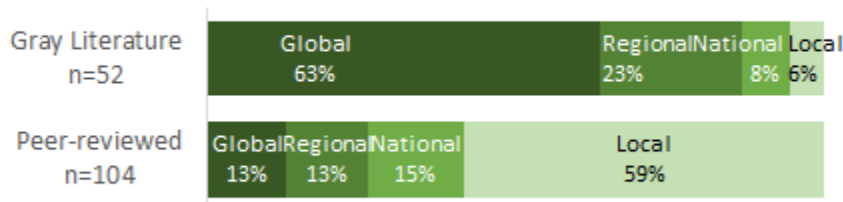
Inclusion of Indigenous and Local Knowledge

Grey Literature n=52	Included 10%	Not included 90%
Peer-reviewed n=104	Included 27%	Not included 73%

0%   20%   40%   60%   80%   100%

Figure 4c. Inclusion of Indigenous and Local Knowledge. Peer-reviewed literature was more likely to address Indigenous and Local Knowledge.

4) Descriptive characteristics of the future works

The focal scale varied significantly between the two literatures, with gray literature focused primarily (63%) on the global scale, and the majority (59%) of peer-reviewed literature focused on the local scale (Figure 5a; X^2(3, N = 156) = 57.69, $p \leq$ .001, Cramer's V = 0.608: large effect size). Gray literature was 11.89 times more likely to focus on the global scale than peer-reviewed literature was; peer-reviewed literature was 23.17 times more likely to focus on the local scale than gray literature was.

**Figure 5. Descriptive characteristics of the future works.**

Geographic scale

Gray Literature n=52	Global 63%	Regional 23%	National 8%	Local 6%
Peer-reviewed n=104	Global 13%	Regional 13%	National 15%	Local 59%

Figure 5a. Scale. Gray literature focused on larger geographic scales

The location of the scenarios also differ starkly (Figure B5; X^2(4, N = 156) = 52.66, $p \leq$ .001, Cramer's V = 0.581: large effect size; omitting global-scale scenarios, X^2(3, N = 108) = 14.55, p = .002, Cramer's V: 0.364: large effect size), with the African continent (32% of gray

literature documents and 7% of peer-reviewed documents with sub-global focus) and Europe and Central Asia (42% and 27%, respectively) more strongly represented in gray literature, and the Americas (16% and 29%), Asia-Pacific (11% and 37%), more strongly represented in the peer-reviewed literature.
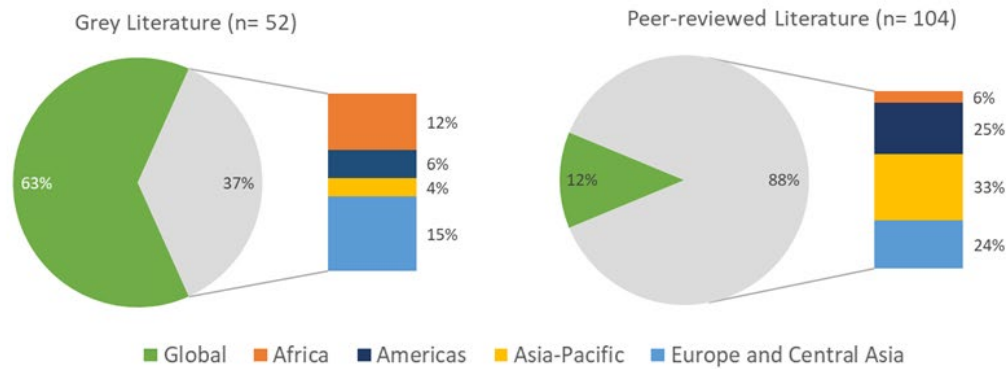

Figure 5b. Region. The two literatures focused on different continental regions.

The methodological approach taken in gray literature was relatively evenly spread between qualitative (38%), quantitative (18%), and mixed (44%) methods. In contrast, peer-reviewed studies strongly tended towards quantitative (69%) methods, with mixed (7%) methods being the minority (Figure 5c; $X^2(2, N = 125) = 34.42$, $p \leq .001$, Cramer's $V = 0.525$: large effect size.
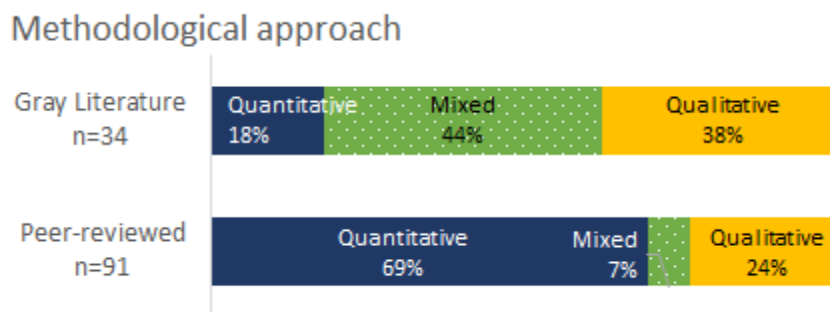

Figure 5c. Methodological approach. Most peer-reviewed literature used quantitative methods, while most gray literature used qualitative or mixed methods.

Gray literature documents were significantly more likely to portray multiple scenarios than peer-reviewed literature (Figure 5d; $X^2(1, N = 156) = 38.54$, $p \leq .001$, Cramer's $V = .497$: large effect size). The odds of gray literature using a single, rather than multiple, scenarios, was 9.09 times higher than for peer-reviewed literature.

Number of scenarios in document
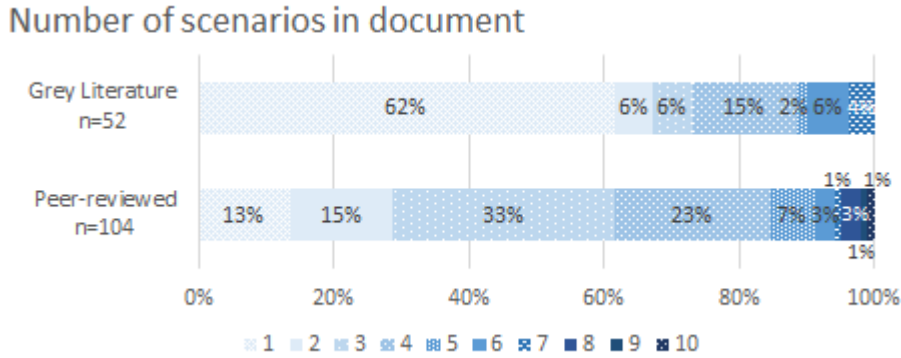
Figure 5d. Number of scenarios. Most gray literature portrayed one scenario, while most peer-reviewed literature portrayed multiple scenarios.

Gray literature most often used normative scenarios, with the majority comprised of normative or target-seeking scenarios (48%) and pathways (19%) toward future goals (Figure 5e; $X^2(2, N = 151) = 55.86, p \leq .001$, Cramer's $V = 0.608$: large effect size, omitting "other" ($n = 5$)). The vast majority (87%) of peer-reviewed literature used exploratory scenarios.

Type of futures work
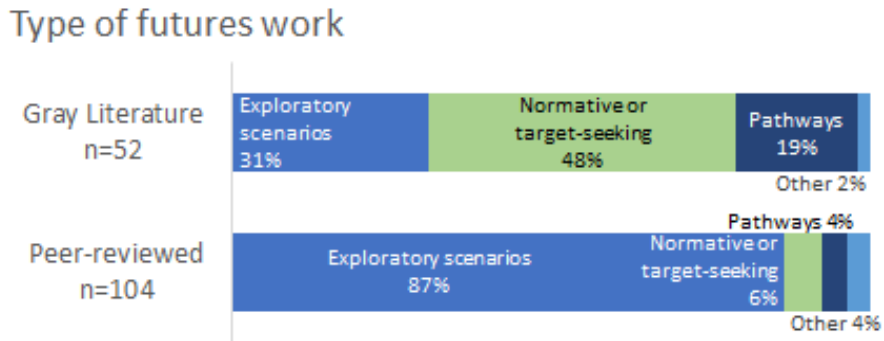
Figure 5e. Type of futures work. Gray literature used more target-oriented scenarios, while peer-literature used more exploratory scenarios

A larger majority of scenarios in the reviewed gray literature could be characterized as desirable futures (Global or Regional Sustainable Development). Undesirable scenarios (Business as Usual, Breakdown, Inequality) were more often found in peer-reviewed literature (Figure 5f; $X^2(6, N = 440) = 22.04, p = .001$, Cramer's $V = .224$: small to medium effect size.

12
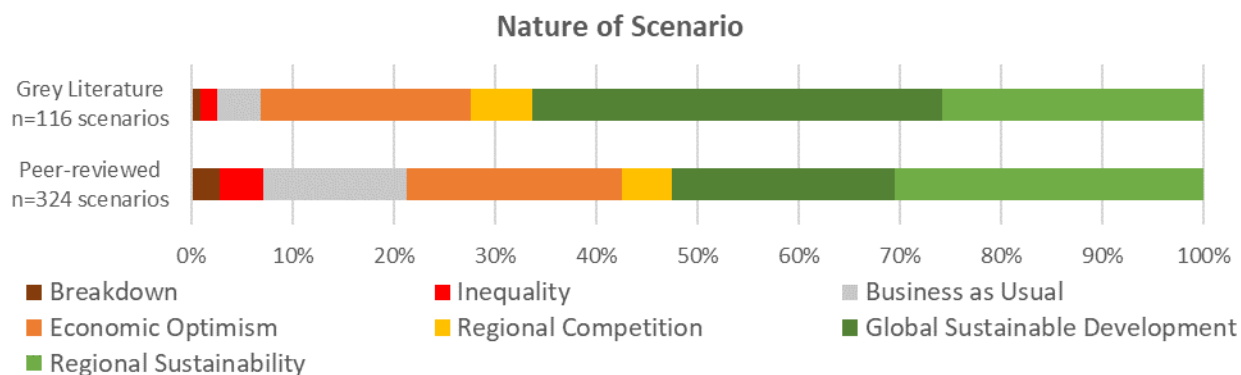
Figure 5f. Nature of scenario. Gray literature more often portrayed desirable futures.

Discussion

Review methods

In the absence of relevant, pre-existing databases, topically customized search engines greatly aided our review. Peer-reviewed literature dominated general search engine results, making it difficult to review non-academic sectors. Unlike existing archives (e.g. (Inist-CNRS 2022; National Library of Australia 2022), topical search engines offered detailed search functions (e.g. Booleans, wildcards) and searched within document contents (Table 1). As they search only specified domains, the topical search engines enabled targeted searches of specific sectors, improving search outcomes to levels comparable to our review of peer-reviewed literature: 13% of peer-reviewed search results, filtered by research area, met our review criteria; topical search engines varied in selection rate, but also aggregated to 13%. This comparison does not negate the aforementioned difficulties of reviewing gray literature, indicated also by the relative numbers of documents that were directly identified by these searches (n_{TSE}=6 gray vs n_{PR}=78 peer-reviewed publications). Nonetheless, topical search engines dramatically facilitated the review process for gray literature and can be recommended for future reviews due to their customizability. Reviews centered on a subset of sectors or domains as well as sustainability-related assessments that increasingly explore evidence from diverse contexts (e.g. upcoming IPBES assessments linked to nexus issues, Intergovernmental Panel on Climate Change (IPCC), Human Development Report) may find them particularly helpful.

It should be noted, however, that search results comprise a small fraction (8%) of the final selection of gray literature documents (Figure 1). The majority (60%) of selected documents were identified by expert knowledge, which also had a much higher selection rate (38%). While the context of a formal science-policy assessment process gave us access to a broad spectrum of relevant experts from within and beyond academia, reviews based on expert knowledge are irreplicable and normally require careful planning (Saleh et al. 2014).

As quality control has been raised as another challenge in reviewing gray literature, one approach in categorizing gray literature is by tiers of retrievability and credibility by outlet control and source expertise (Adams et al. 2017). Typical examples of the highest tier include books, government reports and think tank publications; middle tier examples include annual reports, news articles, company publications and NGO studies, and the lowest tier includes blogs, emails, tweets and catalogs. Given our approach, we did not encounter or review any of the lowest tier documents and did not experience difficulties due to lack of quality. Most of our sources fall under the highest tier. A handful of selected NGO and company publications fall under the middle tier in this categorization but were not of noticeably lower quality. Thus, while quality control has been considered a challenge in reviewing gray literature, this was not part of our experience. This may have be attributed to our highly targeted review methodology, as our sources were high in retrievability and credibility.

Comparing the two literatures

Gray and peer-reviewed literature differed profoundly across all aspects we reviewed. In terms of sectoral and participant characteristics, gray literature proved much more heterogenous. While peer-reviewed works virtually exclusively originated in academia and were government funded, gray literature was associated with various sectors and tended to be rooted in multi sectoral partnerships (Figure 2). All gray literature, and a minority of peer-reviewed literature, involved stakeholders. However, peer-reviewed literature more often included Indigenous and local knowledge, perhaps owing to its more local focus (Figure 4).

Gray literature tended to use scenarios or visions for illustrative purposes, while peer-reviewed works tended to use them for analytical purposes (Figure 5). The majority (87%) of peer-reviewed works used multiple scenarios and discussed their varying implications; over half (62%) of gray literature works portrayed one vision (Figure 5d). The majority (87%) of peer-

reviewed works used exploratory scenarios; normative or target-seeking scenarios and visions (48%) comprised the biggest portion of reviewed gray literature (Figure 5e). These tendencies are also consistent with the general characteristics of the scenarios, whereby a larger majority of scenarios in the reviewed gray literature were sustainability oriented, and undesirable scenarios were more often found in peer-reviewed literature (Figure 5f). Perhaps most tellingly, "values," was most often understood as principle or core beliefs in the gray literature, while this was the least common use of the term in peer-reviewed literature (Figure 3). The fact that certain terms may be interpreted differently depending on the user or author group is critical to consider in works targeting a diverse audience.

The reviewed gray literature had a more global focus, tended to use mixed or qualitative approaches, and involved more diverse groups of stakeholders than the peer-reviewed literature. That the majority of gray literature documents had a global scope contrasts the local focus of peer-reviewed studies (Figure 5a). While our topical search engine targeting the governmental sector searched web domains of International Governmental Organizations and did not include national or subnational governments (a possible consideration for future studies), this is a poor explanation for the contrast. Many assessment reviewers represented countries, and expert knowledge identified the majority of our final selection of documents (Figure 1). Additionally, the regional focus of sub-global works also contrast between the two literatures (Figure 5b).

The distinctiveness of the methodological and procedural approaches is consistent with the target-oriented approach in gray literature. While gray literature accounted for 33% of source documents in the assessment, 81% (n=25) of normative visions and 71% (n=10) of pathways in the assessment originated in gray literature. Similarly, gray literature accounted for 72% (n=33) of global-level studies, and 71% (n=15) of mixed methods studies, and all but one of studies originating outside of academia (28% (n=44) of source documents). These numbers speak to the considerable extent to which the addition of gray literature complemented the peer-reviewed literature and enhanced the scope of the assessment.

Going forward

Lynch et al. (2021) highlight the need for diverse linguistic inclusion for processes undertaking knowledge syntheses in order to counter existing power imbalances in science. The authors offer a reflection on best practices to enhance current peer-reviewed searches to access

literature in other languages through translation technology and searching in non-English databases. They go further to stress the need for the development of step-by-step guidelines for how to include other evidence, which is what we hope this contribution can start to address. Including a plurality of evidence is especially important in the context of enhancing the inclusion of Indigenous and local knowledge which is often captured in other languages, and in gray literature reports (Tengö et al. 2017; Sutherland et al. 2019; Lynch et al. 2021). The inclusion of diverse evidence also requires systemic shifts in ideologies of what is considered "valuable knowledge" and that indigenous knowledge does not need to be validated by science, pointing to the need for the decolonization of science (Johnson et al. 2015; Lynch et al. 2021; Adendorff and Blackie 2022). Our experience with working with gray literature is consistent with IPBES ILK guideline that highlights an approach through online calls across networks to mobilize gray literature not available through academic systematic searches (IPBES 2021). Here, shifts towards more transdisciplinary approaches are needed that see all knowledge holders as active and equal participants in co-designing the research, co-producing the knowledge and collaborating on a variety of outputs based on the work e.g. plain language summaries (Lynch et al. 2021), non-academic outputs such as arts-based materials (Pereira et al. 2019; Heras et al. 2021) and social media (Toivonen et al. 2019; Calcagni et al. 2019). Academics holding powerful roles on journal editorial boards, and handling special issues could also benefit from reflecting on potential biases during publication processes e.g. only publishing successful/ positive results (Ekmekci 2017). There is also a need to support and promote the use and inclusion of multiple language journals (Alves and Pozzebon 2013).

Conclusion

There is increasing recognition that science-policy assessments need to both consider a diversity of worldviews, knowledge types and values, and be implemented in ways that synthesize findings that are salient, legitimate, and credible. This requires working across and beyond disciplinary knowledge bases and mobilizing knowledge from a variety of sectors and actor groups working on understanding how to catalyze change for more just and sustainable futures. However, many global change assessments typically focus on assessing and synthesizing evidence from peer-reviewed, scientific journals which do not account for the multiplicity of

ways in which human-nature relationships are connected and shaped over time. In order to increase current understanding about the state of knowledge of how nature contributes to people in a diversity of ways and what interventions, knowledge and response options are available to a variety of decision makers, we need to widen the scope of evidence that is currently used in science-policy assessments. However, details on how to do this are limited, especially linked to maximizing computing capabilities and opportunities provided by search engines.

Our work provides a methodological precedent, as well as justification, for gray literature reviews not only for applied research but also science-policy interface assessments and similar processes. Methodologically, our exploratory study demonstrates the potential of Programmable Search Engines as an enabler of systematic searches for gray literature, independent of the availability or quality of existing databases. While we did not conduct an exhaustive review, our experience suggests that a more systematic and extensive scan of the search results would easily improve the effectiveness of searches. The customizability that allowed us to target specific sectors also shows promise for adaptation in different contexts. For specialized topics such as ours, where generalized browser search results are dominated by peer-reviewed studies, the customized search engines could also be useful in identifying possible stakeholders. Importantly, the engine has already been updated to enable specification of the target schema, such as medium type (Table 2; Google 2022; Schema.org 2022). This further facilitates the use of topical engines, not only for exploration of reports and publications, but also for audiovisual contents through their metadata. One also wonders about further enhancements. For example, search of caches, restricted by time period, or of automatic translations, would drastically broaden the horizons of accessible information. A posibile development in the nearer future might be the use of wildcards in the domain list, such that a search could target all governmental domains within a country, or .edu or .org addresses.

Our study also strongly justifies gray literature reviews themselves. The addition of gray literature diversified the range of perspectives covered by our assessment. Gray literature contrasted peer-reviewed literature in all aspects considered: language, geographic focus and scale, type, number, use and nature of scenarios, as well as sector and stakeholders involved. Differing use of terminology warrants particular emphasis, as this has critical implications on every aspect of a given study e.g. review, analysis, interpretation by readers as well as dissemination and uptake.

In the context of assessing and synthesizing knowledge on human-nature relationships as determinants of sustainable and just futures for nature and people, this study highlights how including gray literature into structured reviews helps reach broader and more diversified knowledge bases. This supports current calls for knowledge plurality in transdisciplinary processes to address societies' grand challenges. For the purposes of our study, complementing evidence and knowledge from gray literature sources in a systematic manner, not only improves current understanding of the types of values underlying projected futures regarding nature, nature's contributions to people and good quality of life but has also cataloged more nuanced perspectives of a more diverse set of sectors and actor groups. Providing insight on underlying human-nature connections, such a systematic consideration of gray literature contributes to science-policy processes and applied research by meeting state of the art principles and offering verified options for action. This in turn holds the potential to mobilize collective insight needed for transformative change toward just and sustainable futures.

Acknowledgements

References

Adams RJ, Smart P, Huff AS (2017) Shades of Grey: Guidelines for Working with the Grey Literature in Systematic Reviews for Management and Organizational Studies. Int J Manag Rev 19:432–454. https://doi.org/10.1111/ijmr.12102

Adendorff H, Blackie MAL (2022) Decolonization and science education : What is at stake? Decolonising Knowl Knowers 83–102. https://doi.org/10.4324/9781003106968-6

Alves MA, Pozzebon M (2013) How to resist linguistic domination and promote knowledge diversity? Rev Adm Empres 53:629–633. https://doi.org/10.1590/S0034-759020130610

American Library Association (2022) IGO Search. https://www.ala.org/rt/godort/taskforces/internationaldocuments/idtf/igo-search. Accessed 19 Sep 2022

Balshem H, Stevens A, Ansari M, et al (2013) Finding Grey Literature Evidence and Assessing for Outcome and Analysis Reporting Biases When Comparing Medical Interventions: AHRQ and the Effective Health Care Program. Rockville, MD

Balvanera P, Jacobs S, Nagendra H, et al (2020) The science-policy interface on ecosystems and people: challenges and opportunities. https://doi.org/101080/2639591620201819426 16:345–353. https://doi.org/10.1080/26395916.2020.1819426

Benzies KM, Premji S, Hayden KA, Serrett K (2006) State-of-the-Evidence Reviews: Advantages and Challenges of Including Grey Literature. Worldviews Evidence-Based Nurs 3:55–61. https://doi.org/10.1111/j.1741-6787.2006.00051.x

Cairo AH, Green JD, Forsyth DR, et al (2020) Gray (Literature) Matters: Evidence of Selective Hypothesis Reporting in Social Psychological Research. Personal Soc Psychol Bull 46:1344–1362. https://doi.org/10.1177/0146167220903896

Calcagni F, Amorim Maia AT, Connolly JJT, Langemeyer J (2019) Digital co-construction of relational values: understanding the role of social media for sustainability. Sustain Sci 14:1309–1321. https://doi.org/10.1007/S11625-019-00672-1/FIGURES/5

Canada's Drug and Health Technology Agency (2022) Grey Matters: a practical tool for searching health-related grey literature. https://www.cadth.ca/grey-matters-practical-tool-searching-health-related-grey-literature. Accessed 21 Aug 2022

Cohen J (1988) Statistical power analysis for the behavioral sciences, Second edi. Lawrence Erlbaum Associates, New York, NY

Collaboration for Environmental Evidence (2021) Planning a CEE Evidence Synthesis. In: Resour. Authors. https://environmentalevidence.org/information-for-authors/3-planning-a-cee-evidence-synthesis/. Accessed 21 Aug 2022

Curtin University Library Systematic & scoping reviews: What is grey literature? In: Res. Toolkit. https://libguides.library.curtin.edu.au/c.php?g=863554&p=6191902. Accessed 21 Aug 2022

Dwan K, Altman DG, Arnaiz JA, et al (2008) Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One 3:. https://doi.org/10.1371/JOURNAL.PONE.0003081

Ekmekci PE (2017) An increasing problem in publication ethics: Publication bias and editors' role in avoiding it. Med Health Care Philos 20:171–178. https://doi.org/10.1007/S11019-017-9767-0

Fanelli D (2012) Negative results are disappearing from most disciplines and countries. Scientometrics 90:891–904. https://doi.org/10.1007/S11192-011-0494-7/FIGURES/5

Fazey I, Schäpke N, Caniglia G, et al (2020) Transforming knowledge systems for life on Earth: Visions of future systems and how to get there. Energy Res Soc Sci 70:101724. https://doi.org/10.1016/J.ERSS.2020.101724

Fortune Media IP Limited (2019) Global 500. In: Fortune. https://fortune.com/global500/. Accessed 22 Mar 2022

Godin K, Stapleton J, Kirkpatrick SI, et al (2015) Applying systematic review search methods to the grey literature: a case study examining guidelines for school-based breakfast programs in Canada. Syst Rev 4:138. https://doi.org/10.1186/s13643-015-0125-0

Google (2022a) Topical engine. In: Program. Search Engine. https://developers.google.com/custom-search/docs/topical. Accessed 21 Aug 2022

Google (2022b) Programmable Search Engine Help. https://support.google.com/programmable-search/answer/4544182?hl=en. Accessed 1 Oct 2022

Haddaway NR, Collins AM, Coughlin D, Kirk S (2015) The role of google scholar in evidence reviews and its applicability to grey literature searching. PLoS One 10:. https://doi.org/10.1371/journal.pone.0138237

Harmackova Z, Yoshida Y, Mannetti L (2020) IPBES VA Chapter 5. Systematic review of association between values of nature, nature's contributions to people and good quality of life and futures in scenarios, visions and pathways. https://doi.org/10.5281/ZENODO.6543964

Heras M, Galafassi D, Oteros-Rozas E, et al (2021) Realising potentials for arts-based sustainability science. Sustain Sci 16:1875–1889. https://doi.org/10.1007/S11625-021-01002-0/FIGURES/4

Higgins J, Lasserson T, Chandler J, et al (2021) Methodological Expectations of Cochrane Intervention Reviews (MECIR). In: MECIR Man. https://community.cochrane.org/mecir-manual. Accessed 7 Feb 2022

Hopewell S, McDonald S, Clarke MJ, Egger M (2007) Grey literature in meta-analyses of randomized trials of health care interventions. Cochrane Database Syst Rev. https://doi.org/10.1002/14651858.MR000010.pub3

Inist-CNRS (2022) OpenGrey. In: DANS. https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:200362. Accessed 21 Aug 2022

IPBES (2022a) Summary for policymakers of the methodological assessment of the diverse values and valuation of nature of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. In: Pascual U, Balvanera P, Christie M, et al. (eds). IPBES secretariat, Bonn, Germany

IPBES (2022b) Methodological assessment of the diverse values and valuation of nature of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. In: Balvanera P, Pascual U, Christie M, et al. (eds). IPBES secretariat, Bonn, Germany

IPBES (2015) Preliminary guide regarding diverse conceptualization of multiple values of nature and its benefits, including biodiversity and ecosystem functions and services (deliverable 3 (d)). Bonn, Germany

IPBES (2021) Methodological guidance for recognizing and working with indigenous and local knowledge in IPBES (draft). Bonn, Germany

Johnson JT, Howitt R, Cajete G, et al (2015) Weaving Indigenous and sustainability sciences to diversify our methods. Sustain Sci 2015 111 11:1–11. https://doi.org/10.1007/S11625-015-0349-X

Konno K, Pullin AS (2020) Assessing the risk of bias in choice of search sources for environmental meta-analyses. Res Synth Methods 11:698–713. https://doi.org/10.1002/jrsm.1433

Lange F, Brick C (2021) Changing Pro-Environmental Behavior: Evidence from (Un)Successful Intervention Studies. 13:7748. https://doi.org/10.3390/su13147748

Lazrus Id H, Maldonado Id J, Id PB, et al (2022) Culture change to address climate change: Collaborations with Indigenous and Earth sciences for more just, equitable, and sustainable responses to our climate crisis. PLOS Clim 1:e0000005. https://doi.org/10.1371/JOURNAL.PCLM.0000005

Leimu R, Koricheva J (2005) What determines the citation frequency of ecological papers? Trends Ecol Evol 20:28–32. https://doi.org/10.1016/j.tree.2004.10.010

Lynch AJ, Fernández-Llamazares Á, Palomo I, et al (2021) Culturally diverse expert teams have yet to bring comprehensive linguistic diversity to intergovernmental ecosystem assessments. One Earth 4:269–278. https://doi.org/10.1016/J.ONEEAR.2021.01.002

Mahood Q, Van Eerd D, Irvin E (2014) Searching for grey literature for systematic reviews: challenges and benefits. Res Synth Methods 5:221–234. https://doi.org/10.1002/jrsm.1106

National Library of Australia (2022) Trove. https://trove.nla.gov.au/. Accessed 21 Aug 2022

Paez A (2017) Gray literature: An important resource in systematic reviews. J Evid Based Med 10:233–240. https://doi.org/10.1111/jebm.12266

Pereira L, Sitas N, Ravera F, et al (2019) Building capacities for transformative change towards sustainability: Imagination in Intergovernmental Science-Policy Scenario Processes. Elementa 7:35. https://doi.org/10.1525/elementa.374

Saleh AA, Ratajeski MA, Librarian R, Bertolet M (2014) Grey Literature Searching for Health Sciences Systematic Reviews: A Prospective Study of Time Spent and Resources Utilized HHS Public Access. Evid Based Libr Inf Pr 9:28–50

Scheel AM, Schijen MRMJ, Lakens D (2021) An Excess of Positive Results: Comparing the Standard Psychology Literature With Registered Reports: https://doi.org/101177/25152459211007467 4:. https://doi.org/10.1177/25152459211007467

Schema.org (2022) Schema.org. In: Schema.org. https://schema.org/. Accessed 1 Oct 2022

Schöpfel J (2012) Towards a Prague Definition of Grey Literature. In: Twelfth International Conference on Grey Literature: Transparency in Grey Literature. Grey Tech Approaches to High Tech Issues. Prague, Czech Republic, pp 11–26

Sutherland WJ, Taylor NG, MacFarlane D, et al (2019) Building a tool to overcome barriers in research-implementation spaces: The conservation evidence database. Biol Conserv 238:108199. https://doi.org/10.1016/J.BIOCON.2019.108199

Tengö M, Hill R, Malmer P, et al (2017) Weaving knowledge systems in IPBES, CBD and beyond— lessons learned for sustainability. Curr Opin Environ Sustain 26–27:17–25. https://doi.org/10.1016/j.cosust.2016.12.005

Toivonen T, Heikinheimo V, Fink C, et al (2019) Social media data for conservation science: A
    methodological overview. Biol Conserv 233:298–315.
    https://doi.org/10.1016/J.BIOCON.2019.01.023

UBC Library (2022) Grey Literature for Health Sciences. In: Res. Guid. Univ. Br. Columbia.
    https://guides.library.ubc.ca/greylitforhealth#s-lg-box-6256248. Accessed 21 Aug 2022

West JD, Bergstrom CT (2021) Misinformation in and about science. Proc Natl Acad Sci U S A 118:.
    https://doi.org/10.1073/PNAS.1912444117

Wyborn C, Datta A, Montana J, et al (2019) Co-Producing Sustainability: Reordering the Governance
    of Science, Policy, and Practice. https://doi.org/101146/annurev-environ-101718-033103
    44:319–346. https://doi.org/10.1146/ANNUREV-ENVIRON-101718-033103

Yousefi-Nooraie R, Shakiba B, Mortaz-Hejri S (2006) Country development and manuscript
    selection bias: A review of published studies. BMC Med Res Methodol 6:1–5.
    https://doi.org/10.1186/1471-2288-6-37

Zurba M, Petriello MA, Madge C, et al (2021) Learning from knowledge co-production research and
    practice in the twenty-first century: global lessons and what they mean for collaborative research
    in Nunatsiavut. Sustain Sci 2021 172 17:449–467. https://doi.org/10.1007/S11625-021-00996-X