Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Constructing sustainability science: a reflexive analysis of the boundary problem to define a scientific field

Julien Vanhulst
Universidad Católica del Maule
Alejandro Espinosa-Rada
anespinosa@uc.cl
Pontificia Universidad Católica de Chile

Research Article

Keywords: field delineation, scientometrics, Sustainability Science, boundary specification problem, reflexivity, validation practices

Posted Date: July 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9751043/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Sustainability Science emerged as an interdisciplinary domain, but delineating its boundaries is challenging. Field mappings in emerging domains are not merely descriptive: they produce what they claim to map. Boundary choices condition how a field's growth and impact are interpreted, making Sustainability Science a critical test case for scientific field delineation. Such mappings appear in research assessments and institutional reports, yet the consequences of boundary choices remain underexamined. This study reviews 22 scientometrics mappings of Sustainability Science through structured content analysis, coding each study's boundary-setting strategy, validation practice, and epistemic stance.

Findings reveal three patterns: First, delineation strategies are overwhelmingly nominalist: 77% rely on keyword queries as primary entry, treating the field as an analyst-defined publication set often presented as if charting a pre-existing entity. Second, validation is weak: 14/22 studies (64%) score zero on validation checks across all four dimensions; only one study (CS.18) approaches precision/recall quantification. Third, boundary choice produces divergent portraits. Estimates of size, thematic composition, and geographic spread can vary by up to three orders of magnitude (Table 2; from ~ 200 to over 500,000 records) depending on whether boundaries rely on keywords, curated journals/categories, or citation-based cores.

These results support a reflexive, transparent, and multi-method approach to delineating sustainability science and comparable interdisciplinary fields. Comparative reporting of alternative boundaries, sensitivity analyses, and disclosure of assumptions improve reliability of scientometrics mappings and align with science-studies insights on the social construction of fields. The review provides grounded guidance for future mappings that make boundary decisions explicit and reproducible.

1. Introduction

Claims about the emergence and growth of Sustainability Science have multiplied over the last two decades. They appear in research assessments, institutional reports, and in science-of-science studies that map Sustainability Science as a new inter-transdisciplinar scientific “field” (Clark & Harley, 2019; Kajikawa et al., 2014; Schirone, 2024; Vanhulst et al., 2025). Such mappings (typically built from Web of Science or Scopus records) are more than descriptive: they produce what they claim to capture or map. By deciding which documents, journals or keywords are deemed to belong to the emerging field of Sustainability Science, scientometrics studies construct the very object of study they set out to analyze. In this sense, boundary setting is not a purely technical prelude but a constitutive step carrying implicit metatheoretical commitments. In what follows, we ask how scientometrics studies have delineated “Sustainability Science”, with what metatheoretical assumptions, and with what consequences for how the field is represented.

To explore these questions, we examined how scientometrics studies operationalize the delineation of Sustainability Science. The focus is on the boundary choices that precede mapping - search strings,

journal seed sets, subject categories, institutional filters, and author-based heuristics - and on the underlying stance these choices reveal about the ontology of fields. A recurrent pattern is the adoption of a nominalist posture (Laumann et al., 1983), treating Sustainability Science as a label for a network of publications constructed through analytical choices, while presenting results as if they disclosed a real pre-existing entity. The tension between these stances helps explain why different studies arrive at divergent size estimates, shifting topical profiles, and unstable geographies.

When delineating a research field through bibliometric approaches - such as local maps of science (Rafols et al., 2010) - scholars typically adopt one or a combination of several strategies (Zitt et al., 2019). A first approach relies on established subject categories or journal classifications (e.g. Leydesdorff & Rafols, 2009, who employed categories and journals to map global science; or Porter & Rafols, 2009, who mapped interdisciplinarity across six fields using journal data). A second strategy constructs bespoke search strings to capture relevant publications (e.g. Maltseva & Batagelj, 2019, used the query “social network*” to characterize the field of social network analysis, combining the results with core journals). A third approach begins with a seed set of core journals, affiliations, or authors and expands the dataset through citation- or link-based methods (e.g. Keuchenius et al., 2021, analyzed the diffusion of Granovetter’s “The Strength of Weak Ties” to trace different interpretations of the “weak ties” hypothesis). Finally, bottom-up strategies cluster publications via citation or co-word networks, after which domain expertise is used to identify relevant clusters (e.g. Lievrouw et al., 1987, combined quantitative and qualitative techniques to reveal an “invisible college” among biomedical scientists).

Each of these strategies involves trade-offs between recall and precision and carries implicit metatheoretical assumptions regarding the ontological status of the field - whether it is treated as a pre-existing entity or as an analytical construct. Because choices about data sources, thresholds, search terms, and inclusion criteria directly determine what is included or excluded, different operationalizations of the “same” field may produce divergent estimates of its size, structure, and thematic composition.

The contribution of this paper is threefold. First, it synthesizes and systematizes delineation strategies employed to carve out Sustainability Science from large bibliographic databases. Strategies are classified into keyword-based, journal- or category-based, scientific programs or institution-based, and hybrid approaches, and examined in relation to the boundary specification problem. Second, a metatheoretical lens is introduced to evaluate the implications of these strategies. Instead of asking which single boundary is “correct”, the analysis explores what each kind of boundary strategy affords and obscures. Third, recommendations are proposed for reflexive, multi-perspectival delineation, which involve comparing alternative strategies, reporting sensitivity to boundary choices, without presupposing that Sustainability Science is already a fully stabilized scientific field.

A prior question deserves explicit treatment before reviewing how delineation has been done: why delineate at all? The question is not rhetorical. Garry Brewer’s oft-cited observation that “the world has problems, but universities have departments” (Brewer, 1999, p. 328) applies with full force to sustainability science, whose defining ethos was precisely to defy disciplinary fences and work at the

problem–solution interface. From this vantage point, attempts to draw a boundary around the field risk reinscribing the very partitions it set out to dissolve, potentially suppressing the explanatory breadth that is sustainability science's comparative advantage. A second concern runs in the opposite direction: without some boundary, the field cannot be assessed, resourced, or held accountable; mappings that mean everything mean nothing. The tension, then, is not between delineating and not delineating but between different kinds of delineation and different purposes they serve. A research-manager who needs to allocate positions to "sustainability science" requires a different boundary than a historian of science who wants to trace the diffusion of ideas across disciplinary streams. Rather than treating delineation as a technical prerequisite with a single correct answer, we treat the choice of boundary strategy as itself a theoretically charged act whose implications vary by purpose and audience (see also Jerneck et al., 2011; Mahmoud et al., 2018). A related perspective in the literature holds that characterizing rather than delineating sustainability science may be more analytically productive: asking what kind of thing it is (problem-solving vs. critical, bridging vs. transformative; Cox, 1981; Fang et al. 2018; Jerneck et al., 2011; Mahmoud et al., 2018) rather than where its edges lie. This tension is productive rather than disabling: delineation in practice already encodes implicit answers to the characterization question. A keyword-based boundary treats the field as a discourse; a venue-based boundary treats it as an institution; a citation-core boundary treats it as a research program. Making those implicit characterizations explicit and comparative is, we argue, a form of reflexive characterization, one that respects the plurality of what sustainability science might be rather than forcing a single answer.

Empirically, the analysis builds on a systematic sample of scientometrics papers that identify Sustainability Science as instances of scientific field delineation practice. A PRISMA-style screening was complemented with a content analysis of methodological sections, coding for (i) units of delineation (documents, journals, authors, institutions, or hybrids), (ii) data sources and time spans, (iii) boundary criteria, and (iv) validation practices such as precision/recall checks, expert consultation, or triangulation. This approach enables comparison of strategies on their own terms and shows the trade-offs each embeds.

The paper is structured as follows. Section 2 develops the conceptual framing, revisiting classic treatments of the boundary specification problem studying scientific fields. Section 3 details the review protocol and coding scheme. Section 4 presents comparative findings on delineation strategies in the Sustainability Science literature. Section 5 discusses methodological and substantive implications. Section 6 concludes with reflections on prospects for more reflexive and plural approaches to field delineation.

2. Background and conceptual framing

2.1. Sociology of science and scientific fields

In the sociology of scientific networks (Bellotti & Espinosa-Rada, 2025), formal and informal groups of researchers working on similar problems have been described through a variety of concepts: invisible

colleges (Crane, 1972; de Solla Price, 1965; Zuccala, 2006), scientific communities (Hagstrom, 1965; Kornfeld & Hewitt, 1981), scientific fields (Whitley, 2000), and scientific specialties (Mulkay, 1979; Mullins, 1972). These constructs show that scientific are not naturally given entities but social formations that undergo phases of emergence, consolidation, and decline. While some succeed in establishing a coherent research program and institutional base, others do not achieve stable development (Mullins & Mullins, 1973).

Consolidation of a scientific field typically requires identifiable leaders, coherent groups sharing epistemological orientations, and durable collaborative networks (Mullins, 1972; Whitley, 2000). Merton (1973) showed how institutional norms such as peer review connect the social organization of science with its cognitive development. These perspectives establish that scientific networks are simultaneously social and cognitive constructs, historically contingent, and with boundaries that are never self-evident (Espinosa-Rada et al., 2024, 2025; White, 2011).

The operationalization of such networks through bibliometric traces (publications, citations, co-authorship) is what connects the sociology of science to the scientometrics practice we review here (Glänzel & Schubert, 2003; Morris & Van der Veer Martens, 2008; Small, 1973). In scientometrics, the notion of scientific field has been widely mobilized as a practical unit of analysis[1]. Scientific fields are often described as self-organizing networks of researchers who work on related topics, publish in overlapping sets of journals, and cite each other's work, thereby generating a cumulative body of knowledge. Common heuristics for delineating scientific fields include the use of keyword queries, journal lists, and subject categories, which carry important analytical and ontological implications.

Today it is possible to identify a recognizable nucleus of researchers, specialized journals, and recurrent citation patterns, as well as dedicated training programs explicitly labeled as sustainability science. A defining feature of this scientific field is its interdisciplinary orientation: rather than consolidating within the boundaries of a single discipline, it brings together contributions from the natural and social sciences, engineering, and the humanities around shared problem framings. These features suggest that, despite its breadth and heterogeneity, sustainability science has acquired many of the characteristics of a scientific field in formation, making this case an appropriate analytical lens for our study.

2.2. Study of scientific field delineation in scientometrics

Building on these sociological insights, scientometrics research has developed strategies to empirically identify and analyze scientific fields. For scientometrics analysis, delineation is a foundational step: it determines which researchers, papers, or journals are included as part of the scientific field. This decision directly shapes the observed size, thematic profile, and network structure of the field, as well as the indicators derived from it (Glänzel & Schubert, 2003; Zitt, 2005). Scientific fields typically exhibit a core-periphery structure (Breiger, 1976; Moody, 2004), with a concentrated set of researchers and publications surrounded by a wide array of peripheral contributors, and a densely connected interior that citation-based filtering strategies aim to isolate (Waltman & van Eck, 2012; Zitt et al., 2019). They also

display overlap, since many researchers or documents belong partly to adjacent scientific fields, complicating clear demarcations (Morris & Van der Veer Martens, 2008, pp. 236–238).

These features make delineation more than a technical operation. Choices about keywords, journal lists, or subject categories produce different representations of the field, leading to divergent conclusions about its maturity, coherence, and institutional geography. Beyond methodological implications, these delineations also carry ontological stakes: by selecting who and what is included, scientometrics studies do not simply measure but actively shape the contours of a field.

2.3. The boundary specification dilemma

A central methodological challenge in network research is the boundary specification problem (Borgatti et al., 2024; Laumann et al., 1983; McLevey et al., 2023; Scott, 2017). It refers to the difficulty - or impossibility - of defining boundaries that are simultaneously exhaustive and selective: wide enough to capture all relevant actors and relations yet narrow enough to preserve coherence. As Lietz (2020) aptly notes, to delineate a field is to draw an “impossible boundary”.

The boundary problem is particularly acute for interdisciplinary fields, and sustainability science is a paradigmatic case (Clark & Harley, 2019; Jerneck et al., 2011). In established disciplines, boundaries are partly stabilized by shared training programs, canonical journals, and departmental structures. In interdisciplinary domains, these anchors are absent, distributed, or contested: researchers trained in ecology, engineering, economics, and political science converge on shared problem spaces without sharing a single disciplinary home. This means that any single delineation strategy (however carefully designed) will privilege some disciplinary threads over others, producing portraits of the field that are not merely incomplete but structurally biased toward the epistemic assumptions embedded in the chosen entry point. A keyword-based entry privileges lexically self-labeling work; a journal-based entry privileges institutionalized venues; a citation-based entry privileges the most-connected core. The boundary problem in interdisciplinary fields is not just about where to draw the line; it is about which kind of line to draw, each of which produces a fundamentally different field.

In their work on the boundary specification problem in network analysis, Laumann, Marsden and Prensky (1983) discuss the criteria of inclusion used to define the boundaries of a social network. They distinguish between two perspectives: on the one hand the realist approach, when boundaries are defined from the objective and subjective perspectives of the actors themselves. “Network is treated as social fact only in that it is consciously experienced as such by the actors composing it” (Laumann et al., 1983, pp. 20–21). On the other hand, the nominalist approach, when boundaries are defined analytically by the researcher: “Network closure has no ontological independent status. […] The perception of reality is assumed to be mediated by the conceptual apparatus of the analysis” (Laumann et al., 1983, p. 22).

In practice, every delineation involves trade-offs between inclusiveness and specificity. If boundaries are drawn too broadly, the resulting population risks becoming heterogeneous and noisy; if they are drawn too narrowly, important actors and connections are excluded. By relying on broad keywords,

scientometrics mappings may obscure these internal differentiations. For scientific fields, this dilemma is particularly acute given their dynamic character: they expand, fragment, and overlap with adjacent scientific domains over time.

Sustainability science exemplifies this condition. Widely recognized as an emerging scientific field (Clark & Harley, 2019; Schirone, 2024; Vanhulst et al., 2025), it nonetheless lacks fully institutionalized borders. It functions as a non-formal network of researchers linked by shared concerns about unsustainability and sustainability transitions, but without a single disciplinary or organizational anchor. This makes it a paradigmatic case for examining how different boundary choices (keywords, journal lists, citation thresholds, network filters) produce distinct portrayals of the field.

In this review, the central focus is on how researchers address the boundary specification problem when attempting to define sustainability science, and on the consequences of adopting one or another delineation strategy. There is an ongoing debate about whether sustainability science should be regarded as a mature field or rather as a proposal still under construction. As Fang et al. (2018, p. 1) put it, “sustainability science […] has become a well-recognized new field of study. However, the ‘identity’ of Sustainability Science remains unclear. […] Efforts are still needed to further consolidate its identity by developing a coherent and rigorous scientific core”.

As argued above, sustainability science can be understood as a network organized around shared problems, research interests, and communicative links, rather than around fully institutionalized borders. Scientometrics studies of this network are almost exclusively based on bibliometric traces of actors and their relationships, using only scientific articles as the unit of observation, and the typical delineation pattern is discursive/lexical capture, based on the use of keywords to select articles, sometimes combined simultaneously or sequentially with selection of specific venues, use of direct quotations, or some kind of threshold to refine the initial search. Reviewing these studies thus allows us to assess not only the representations they produce of sustainability science but also the performative effects of different boundary choices.

3. Data and Methods

3.1. Review protocol and data sources

The study follows the PRISMA 2020 protocol for article selection (Page et al., 2021), adapted to the objective of reviewing scientometrics approaches to the delineation of sustainability science.

Original search (WoS, title-restricted): the initial search was conducted in the Web of Science Core Collection in August 2025, using the following Advanced Search query: (TI=“Sustainability Science” AND AB=(Field AND Publication)) OR (TI=“Sustainability Science” AND AB=(bibliometric* OR scientometric*)). Document types were restricted to Article, Review Article, or Early Access. The search identified 27 records. After removing false positives through title/abstract screening, 11 records remained; one full

text was unobtainable, and two further records were excluded for focusing exclusively on a single journal. This yielded 8 seed articles.

Citation chasing: to complement the seed, citation chasing was conducted (both backward and forward), yielding 26 additional records. After applying the same eligibility criteria (scope, language, and document type), 14 were retained and 12 excluded. In total, 22 studies are included: 8 from the seed search and 14 from citation chasing (see PRISMA flow diagram, Fig. 1).

Broader search (OpenAlex, abstract-level): to quantify the recall gap created by our title-restricted query, we replicated the search strategy in OpenAlex using two complementary lenses. Lens A (title-restricted, analogous to the WoS query) returned 26 records. Lens B (abstract-level: “sustainability science” in the abstract AND a bibliometric/scientometrics term) returned 137 records. The combined, deduplicated set totals 139 records. This broader set was not subjected to full eligibility screening, as its purpose was to quantify the recall gap rather than to extend the corpus. The full OpenAlex search syntax is reported in Appendix A.

Reflexivity note on the corpus: the reliance on citation chasing to retrieve 14 of 22 studies (64% of the final corpus) is itself a methodological result. It demonstrates empirically that a narrow, title-anchored query misses most relevant work. We treat this not as a flaw to conceal but as evidence in favor of the paper’s central argument: that single-strategy nominalist delineations systematically under-recover relevant material. The OpenAlex validation reported in the preceding paragraph makes this gap independently legible: a broader, abstract-level search recovered 139 candidate records against the 8 retrieved by the title-restricted WoS seed, underscoring why citation chasing was indispensable as a complement.

Source: own elaboration based on (Page et al., 2021).

3.2. Content analysis and coding scheme

To examine how these studies delineate sustainability science, a structured coding scheme was developed (Krippendorff, 2018; Schreier, 2012). Each article was coded along the following dimensions (see Online Resource 1 for applied coding by articles):

- Corpus construction (sampling): data source(s); retrieval trigger (query strings/seed lists); inclusion criteria; thresholds; time window; resulting corpus size.
- Operational delineation design (inductive): operation(s) used (lexical/keyword; venue/category; relational/citation network; thresholds; program/institution anchors) and combination logic (single-step; sequential filters; union).
- Validation practices: we coded four non-exclusive forms: (1) Expert review (external vetting of strings/corpus), (2) Precision/Recall benchmarking (sampling tests), (3) Alternative boundary comparison (empirical overlap/robustness across keyword vs. venue/category vs. relational

criteria), and (4) Stability/robustness tests (algorithm/parameter sensitivity; temporal or cross-database robustness).

- Reflexivity on boundary choices (analyzed in Section 4.5): the extent to which studies critically reflect on their own boundary definitions (discussing limitations, arbitrariness, or alternatives) or, conversely, treat such criteria as self-evident. We also noted brief “Reflexivity Evidence” cues (verbatim/summary) to justify the level assigned.
- Ontological stances: How the study conceptualizes the field of sustainability science.
- Post-corpus analysis techniques: bibliometric and semantic tools applied after sampling (e.g., co-citation, co-authorship, bibliographic coupling, keyword co-occurrence, NLP clustering, geometric data analysis, systematic content analysis).
- Outputs reported: main results, such as field maps, collaboration structures, or thematic classifications.
- Claims about Sustainability Science status: the study’s explicit claims about sustainability science.

Including reflexivity as a distinct coding dimension allowed us to identify whether boundary-setting practices were naturalized (e.g., keyword queries or journal categories used without justification) or explicitly problematized as methodological and epistemological choices; we then used these ratings to contextualize how delineation choices shaped subsequent analyses (Miles et al., 2013). Coding was performed independently by both authors across all 22 studies before results were compared. To assess inter-coder reliability, a post-hoc check was conducted on a purposive sample of 8 studies representing the full range of boundary designs (one each of Types I→III→IV, I + II, I→III [×2], II [×2], I [×2]). Both coders applied the coding scheme independently to each study's without consulting the other's codes. Agreement was 100% across the four binary validation dimensions (Expert review, Precision/Recall, Alternative boundary comparison, Stability tests) and the delineation design dimension, which is determined mechanically from the retrieval operations described in each study's methods section. For Reflexivity Level and Ontological Stance, disagreements were resolved through structured discussion with reference to verbatim passages from the reviewed texts; codebook anchor examples guided contested cases (Gläser & Laudel, 2013; Lombard et al., 2002).

3.3. Analytic strategy

Coding proceeded iteratively. For each study, we first extracted and coded the *sampling logic* that determines inclusion in the corpus (data source(s), retrieval triggers such as query strings or seed lists, inclusion/exclusion filters, and thresholds), and we kept this strictly separate from *post-corpus analysis* (network, bibliometric, or semantic techniques applied after the dataset was assembled). This structured coding scheme is consistent with established practices in qualitative content analysis (Schreier, 2012), allowing systematic comparison across heterogeneous studies.

Building on these codings, we derived an inductive “operational typology of delineation designs” (see Table 1) based on (i) the main delineation operation(s) used to construct the corpus (lexical query-based; venue/category-based; relational/citation-based; reputational/impact thresholds; and program/report-

based anchors) and (ii) the combination logic across operations (single-step designs; sequential designs; and unions).

Table 1
Operational typology of delineation designs

Type Code	Parent operation (sampling)	What counts as inclusion (rule family)	Typical examples (from our review coding)
I	Lexical (query-based)	Records are included if they match specified terms/phrases in bibliographic fields (e.g., Title/Abstract/Keywords/Topic) via search strings.	Topic searches with “sustainab*”; phrase-based queries such as “sustainability science”; Boolean query blocks.
II	Venue/Category (source-based)	Records are included because they are published in selected journals or within journal categories (used as a proxy boundary).	Curated journal lists; restricting the corpus to journals in the WoS/JCR category “Green & Sustainable Science & Technology”.
III	Relational (network-based)	Records are included/excluded based on their position in a network defined by links (e.g., citations/co-authorship), using structural filters.	Retaining only papers in the Largest/Maximum Connected Component (LCC/MCC) of a direct-citation network; citation-based expansion from a seed set.
IV	Reputational thresholds (impact-based)	Records (or venues) are included only if they meet analyst-defined thresholds intended to capture authority/influence/impact.	Minimum citation counts; “top-N most cited per year”; journal-level cutoffs such as JCI/impact thresholds.
V	Programme report–based core publication lists	Records are included because they appear in pre-existing publication lists extracted from official programme documentation (e.g., annual reports / core project reports).	Core publications taken from IGBP/IHDP annual reports and core project reports, then matched/retrieved in WoS for analysis.

Source: own elaboration

Finally, we used the resulting operational typology to compare prevalence, combinations, and validation practices, and to interpret the epistemic implications of boundary-setting choices for subsequent sustainability science field portraits (as a diffuse domain vs. a more bounded field).

We use the nominalist/realist distinction as a heuristic lens (not as a classificatory system with clean boundaries) to interpret the epistemic commitments implicit in boundary-setting strategies. Several clarifications are warranted.

First, as Laumann, Marsden, and Prensky (1983) note in their original network-analysis context, a realist boundary requires that the network be “consciously experienced as such by the actors composing it.” In bibliometric practice, however, no strategy is purely analyst-constructed: venue-based filters aggregate thousands of actor decisions (authors choosing to submit to Sustainability Science, editors ratifying placement, indexers assigning categories); citation-based cores reflect communities of recognized interlocutors. In this sense, the nominalist/realist axis in bibliometrics is better understood as a degree of mediation between analyst decisions and actor decisions, rather than a binary distinction. What we call nominalist here (and what the reviewed studies overwhelmingly exemplify) is a strategy in which the analyst’s definitional logic (search string, venue list, citation threshold) is primary and actor recognition is latent and unverified.

Second, what follows from this heuristic reading is not that “all observed strategies are nominalist” is trivially true by definition (though it is by construction if the term is stretched to cover all analyst-mediated decisions). Rather, the key empirical result is that none of the 22 studies attempts to verify or surface the actor-recognition component embedded in their chosen proxies: no study surveys researchers about their felt membership in the field; none uses conference rosters or institutional self-designation to validate the analyst-constructed corpus. The nominalism finding is therefore not definitional; it concerns an absence of realist validation, regardless of how much actor information is latent in the proxies used.

4. Findings

4.1 Prevalence of delineation strategies

Across the 22 scientometric studies reviewed (2007–2025), boundary specifications are overwhelmingly nominalist and rely heavily on lexical entry points (see Fig. 2). Type I (Lexical/Query-based) is the sole or primary boundary in the vast majority of cases. Specifically, strict single-operation designs appear in 13 studies: nine rely exclusively on Type I (lexical queries), three on Type II (Venue/Category-based), and one on Type V (Program report-based).

Sequential and hybrid designs account for the remaining nine studies, most of which use Type I as the starting point (8/9). These combinations typically refine a broad lexical search with subsequent filters: three studies tighten the corpus using Type III (Relational/Network-based) criteria (e.g., the largest connected component); two apply Type IV (Reputational/Impact-based) thresholds; and one combines both relational and impact filters (I > III > IV). Additionally, two studies combine lexical and venue boundaries (Type I + Type II). Finally, only one study departs from the lexical entry point in a sequential design, starting with a venue-based list (Type II) and refining it with impact thresholds (Type IV).

Source: own elaboration

Two poles dominate entry into the field. First, keyword/query-based designs (Type I), meaning topic searches on sustainab* and related terms, remain the modal entry; they are scalable but can be

semantically diffuse (e.g. Asatani et al., 2020; Bettencourt & Kaur, 2011; Kajikawa et al., 2014; Yarime et al., 2010). Second, venue-based rules (Type II), such as curated journal lists or database categories, offer a more institutionalized gate (e.g. Blanco et al., 2025; Horcea-Milcu et al., 2020).

Beyond single-rule studies, many adopt hybrid designs. There are 13 single-rule, 7 sequential (AND), and 2 union (OR) designs. Recurrent patterns include: I→III in 3 studies (keyword entry, then a relational core such as the largest connected component or a citation-defined nucleus; e.g., Asatani et al., 2020; Kajikawa et al., 2014); I→IV in 2 studies (keyword entry refined by reputational thresholds); I + II in 2 studies (union of lexical and venue-defined sets; e.g., Blanco et al., 2025); II→IV in 1; and I→III→IV in 1 (Buter & Van Raan, 2012; see Fig. 2). In short, Type I is often the entry (77% of the sample), while III (structure) and IV (impact) are used to tighten inclusions.

Temporally, early studies tend to rely on single-rule I (e.g. Bettencourt & Kaur, 2011; Kajikawa et al., 2007). Over time, studies show more two-step and hybrid designs alongside curated multi-term queries (e.g. Clark & Harley, 2019; Horcea-Milcu et al., 2020) and explicit robustness checks (e.g. Blanco et al., 2025), signaling growing methodological pluralism (Fig. 3).

Source: own elaboration.

Overall, nominalist–lexical type boundaries remain the default; relational, venue and reputational-based boundaries complement it; and realist delineation is absent. This tilt toward analyst-defined boundaries makes what “counts” as Sustainability Science highly sensitive to search logic and category choices, a theme we unpack in Sections 4.2–4.4 and the Discussion.

4.2 Assumptions in Keyword-Based versus Journal/Category-Based Delineation

Studies typically follow two broad strategies: keywords (lexical capture) or journals/categories (institutional capture). Each encodes a distinct ontology of what Sustainability Science is, and thus yields different portraits of the field.

Keyword delineation treats Sustainability Science as a discursive domain signaled by terminology (e.g., sustainab* or more complex keywords queries), not by fixed institutions. It assumes authors share a vocabulary through which relevant work can be retrieved. Its main strengths are flexibility and inclusiveness: it can surface emergent, interdisciplinary threads outside any one venue (Hassan et al., 2013). The trade-offs are known: broad tokens are semantically diffuse (noise/false positives), and some relevant work won’t self-label with sustainability terms (false negatives). In short, keyword logic presumes a fluid, expanding discourse network and risks heterogeneity without careful query design.

Venue strategies proxy Sustainability Science by a stable institutional footprint, a set of journals or a subject category thought to represent the core (as used for example in Pandiella-Dominique et al., 2021). Others compile leading sustainability journals and analyze only those outlets (e.g., an 18-journal list; Schirone, 2024). Ontologically this reflects analyst-chosen metadata categories. The strengths are

precision and thematic focus; the costs are taxonomy bias and potential circularity (defining the field by the outlets that claim to represent it). Classic early work that restricted analysis to a few core journals (e.g., Sustainability Science journal itself; Kajikawa, 2008) illustrates the clarity gained and the risk of missing work outside that core. Category choices can also skew emphasis (e.g., toward technology/environment streams while under-representing social/policy research).

Keywords import a self-designation bias (only self-labeled items count); venues import an institutional-scope bias (the field is what selected outlets say it is). Several studies therefore adopt hybrid/iterative fixes. On the keyword side: use curated multi-token/Boolean strings vetted by experts to raise precision/recall (Clark & Harley, 2019; Hassan et al., 2013; Horcea-Milcu et al., 2020); avoid overly generic tokens; require co-occurrence with context terms. On the venue side: prefer data-driven journal sets (e.g., citation-derived cores; Schirone, 2024) or analyze category sub-clusters to expose internal diversity (Pandiella-Dominique et al., 2021). A practical hedge is to combine content and venue logics: for example, merge a keyword corpus with all items from Sustainability Science so that core-journal papers lacking explicit keywords are not lost (as implemented by Blanco et al. 2025 and Troullaki et al. 2021).

In sum, keyword-based delineation treats Sustainability Science as a dynamic discourse captured through evolving language, whereas journal/category-based delineation treats it as a consolidated institutional domain identifiable via established markers. The former prioritizes breadth but risks noise and under-labeling; the latter prioritizes focus but risks exclusion and circularity. Reflexive studies acknowledge these trade-offs, compare multiple boundaries, and make the affordances and blind spots of each approach explicit, an orientation taken up in the Discussion and operationalized through the multi-lens protocol and validation checks summarized in Box 1.

4.3 Validation practices (and their absence)

Validation varied widely across the 22 studies, and in many cases was absent. The reported approaches can be clustered into four types:

Expert review of queries/corpus: four studies (SS.4; CS.12; CS.18; CS.19) used expert vetting to assess keyword strings or retrieved sets, sometimes iteratively refining them, a step useful for improving precision and guarding against topical omissions.

External comparison to journal lists/classifications: seven studies compared results to established lists or database categories: four fully (SS.4; CS.11; CS.16; CS.18) and three partially (SS.1; SS.2; SS.3) as an external boundary check. Mismatches (missing widely recognized journals; inclusion of off-scope outlets) means delineation problems.

Precision/recall tests: only one study (CS.18) attempted a partial information-retrieval evaluation. The remaining 21/22 did not estimate retrieval accuracy, leaving precision (share of relevant items among retrieved) and recall (share of relevant items captured) unknown.

Robustness/sensitivity checks: ten studies reported some robustness analysis: two fully (SS.1; SS.7) and eight partially (SS.2; SS.3; SS.4; SS.5; SS.6; SS.8; CS.9; CS.13) by varying clustering algorithms, thresholds, databases, or time windows to test stability of structures and findings.

Seven studies reported none of the above validation strategies (CS.10; CS.14; CS.15; CS.17; CS.20; CS.21; CS.22). Their authors implicitly assumed that the initial delineation adequately represented the field, then proceeded without any external check on corpus quality or completeness. The field definition's rigor rested on the search strategy's persuasiveness: 'validation by assumption.' Applied strictly across all four coded dimensions, the figure rises to 14 of 22: the seven above, plus seven more that referenced robustness or validation in passing without formally employing any of the four procedures.

These limited validation practices have three practical consequences. First, heterogeneous validation contributes to result variability: alternative boundaries (rarely benchmarked) produce divergent maps and narratives. More fundamentally, there is no single ground-truth partition for scientific fields; even publication-level classification methods acknowledge the absence of a "golden standard" and call for evaluation with expert feedback (Waltman & van Eck, 2012). Second, absent precision/recall estimates, external comparisons, or sensitivity analyses, corpus legitimacy and completeness remain uncertain. Hybridizing lenses (texts, venues/categories, and citation relations) improves coverage and interpretability while keeping choices transparent and reproducible (Rafols et al., 2010; Zitt, 2005). Finally, repeated reuse of the same tokens or categories risks stabilizing a conventional image of the field. Information-retrieval queries and bibliometric mappings are complementary and are best used together in multi-step designs, not treated as definitive on their own (Leydesdorff & Probst, 2009; Waltman & van Eck, 2012; Zitt et al., 2019).

Overall, validation was often minimal or absent, indicating clear room for improvement. Systematic steps (expert feedback, external boundary checks, precision/recall sampling, and robustness analyses (see Box 1)) would markedly strengthen the reliability of field delineations.

4.4. What representations do different boundaries produce?

Boundary rules yield markedly different portraits of Sustainability Science. Each criterion (keywords, journals/categories, or citation/core-sets) constructs a distinct map, emphasizing some features while downplaying others.

Table 2
Corpus sizes across the 22 reviewed studies, by boundary type and time window

Study	Year	Boundary	Time window	Included corpus (primary/final)
CS.10	2008	II (venue)	2003–2007	191
CS.16	2019	V (programme)	1997–2013	398
CS.14	2016	I (keyword)	2013 only	253
CS.15	2018	I (keyword)	open	832
SS.4	2021	I + II	open–2020	1,502 (SS corpus only)
SS.1	2013	I→III→IV	1999–2008	10,594 (seed); 1,038 (HCKB core)
CS.17	2020	I→IV	open–2016	3,118 (combined SS + SES)
CS.22	2011	I→IV	1981–2008	7,800 (retrieved); 3,334 (≥ 2 cites)
CS.9	2007	I→III	open–2006	29,391 (retrieved); 9,973 (LCC)
SS.7	2025	I + II	1950–2021	2,076 (post-screen)
CS.19	2021	II (venue)	1968–2018	29,992
SS.8	2024	II→IV	2001–2021	71,871
CS.11	2012	I (keyword)	1996–2009	17,226
CS.12	2013	I (keyword)	2000–2010	N/R (sub-area counts only)
CS.13	2016	I (keyword)	1987–2015	59,926
CS.18	2020	I (keyword)	2000–2019	20,813
SS.2	2011	I (keyword)	1973–2009	≈ 20,376
SS.5	2010	I (keyword)	open	41,487
CS.20	2021	I (keyword)	1993–2017	8,765
SS.6	2021	II (category)	1994–2016	49,065
SS.3	2014	I→III	open–2013	89,908 (retrieved); 51,390 (LCC)
CS.21	2020	I→III	open–2019	514,480 (retrieved); 312,584 (LCC)

Note: “N/R” = not reported as a single aggregate. “LCC” = largest connected component (relational filter). Time windows reflect the study’s own stated start and end years; open = no explicit lower bound. Corpus sizes are as reported by the authors.

Table 2 makes the “order of magnitude” claim concrete. Within the keyword-only designs (Type I), corpus sizes span from 253 (CS.14, single year 2013) to 514,480 (CS.21, broad Scopus query with no lower year bound): a range of more than three orders of magnitude. Even restricting comparison to studies with

overlapping time windows and similar databases, the variation remains stark: SS.3 (I→III, 89,908 retrieved, 51,390 retained in LCC) versus SS.4 (I + II, 1,502 records) versus CS.10 (II venue-only, 191 records). These differences are not attributable to the passage of time alone: CS.21 (2020, 514,480 records) and CS.19 (2021, 29,992 records) overlap in time but differ by a factor of 17. The primary driver is boundary logic, not the growth of the literature. Direct comparison across identical time windows is not possible given the heterogeneity of study designs, but the magnitude of variation confirms that boundary choice (not the underlying phenomenon) is the dominant source of dispersion in how the field appears.

Query-based corpora

Topical queries generate large, heterogeneous sets that surge over time, spanning ecology, engineering, economics, and policy across many venues. Early global mappings depicted a fast-growing, varied domain rather than a tightly bounded field (e.g. Bettencourt & Kaur, 2011; Kajikawa et al., 2007). However, outputs may differ depending on the granularity of keyword queries. Single-token searches maximize size and scatter, whereas curated multi-term lists improve topical alignment, reduce semantic noise and keep coverage sufficiently wide to capture the principal thematic streams. In the resulting maps, curated sets typically exhibit fewer isolated clusters and higher intra-topic density than single-token corpora, without becoming overly narrow or collapsing onto a limited set of venues (Clark & Harley, 2019; Hassan et al., 2013; Horcea-Milcu et al., 2020).

Venues/category-based corpora

delimiting by outlets (e.g., Clarivate's Green & Sustainable Science & Technology) or curated "sustainability" journals yields smaller, more cohesive datasets that reflect indexing priorities. Results often skew toward environmental/technical streams, missing diffuse work outside selected venues (Pandiella-Dominique et al., 2021). Such classifications are known to be imperfect lenses for interdisciplinary domains (Leydesdorff & Probst, 2009; Rafols et al., 2010).

Citation/core-set approaches

Network-based filtering (e.g., largest connected component; highly cited cores) distills broad entries into tightly knit field cores (hundreds to low thousands), foregrounding sub-communities and knowledge bases while de-emphasizing peripheral or less-cited lines (e.g. Buter & Van Raan, 2012).

These divergent outputs affect how sustainability science is perceived: (i) Apparent size can shift by up to three orders of magnitude (Table 2). Broad keywords portray an expansive, rapidly growing domain; multi-term, venue, or core-set designs produce compact portraits. (ii) Disciplinary profile changes with boundaries. Inclusive keyword maps amplify cross-disciplinary breadth, whereas venue/core filters heighten cohesion and eco-technical emphasis. (iii) Geography and institutional representation also shift: broad keywords capture a wider global reach yet inherit database and language biases; venue

filters reflect regions and organizations dominating indexed outlets, under-representing activity elsewhere.

Taken together, these comparisons show that the image of sustainability science is highly contingent on boundary choice: query-based delineations portray a sprawling discourse; venue-based boundaries yield a smaller, more cohesive and institution-centered portrait; and citation/core-set approaches foreground a tightly knit field core. Rather than presuming a single “true” boundary, a plural and reflexive stance is warranted, one that makes explicit what each boundary affords and obscures and, where possible, compares alternative delineations side by side. This recognition sets up the discussion, where we examine the ontological and methodological stakes of boundary choice and outline practical guidelines for validation and transparent reporting.

4.5 Reflexivity and ontological stance: patterns across the corpus

Two additional coded dimensions (reflexivity on boundary choices and ontological stance) allow us to address questions the operational typology alone cannot answer: do studies that reflect more explicitly on their delineation choices also validate more rigorously? And does the ontological framing of the field covary with boundary design? Of the 22 studies, 5 were coded as High reflexivity (SS.7; CS.12; CS.17; CS.18; CS.19), 16 as Medium or Medium-High, and 1 as Low-Medium (SS.5). No study was coded as having no reflexivity; all at least acknowledge the existence of alternative approaches in passing, even if they do not act on that acknowledgment. High-reflexivity studies share a common feature: they explicitly discuss the boundary problem in their methods section, name the trade-offs of their chosen strategy, and in some cases compare results across alternative criteria (CS.17; CS.18).

Among the 5 high-reflexivity studies, 4 employed at least one formal validation strategy: CS.18 (expert review, precision/recall assessment, and alternative boundary comparison), CS.19 (expert consultation on journal selection), SS.7 (cross-database replication in Scopus), and CS.12 (iterative expert review of keyword lists). Among the remaining 17 studies, only 3 did (18%). CS.17, despite being the most explicit in acknowledging boundary arbitrariness, used no formal validation procedure, relying instead on qualitative transparency. This pattern holds across the corpus, yet the absolute level of validation remains low in both groups: no study covers more than 2 of the 4 validation dimensions, and 14 of 22 studies (64%) score zero across all four. Articulating awareness of boundary arbitrariness does not consistently translate into procedural safeguards, a gap that mirrors the paper's broader finding about nominalism.

Furthermore, no clear improvement in validation practices is visible over the 18-year span of the corpus: studies from 2019–2025 are not noticeably more rigorous than those from 2007–2013. A modest shift in reflexivity is evident: more recent studies tend to acknowledge the boundary problem more explicitly in their methods sections. This growing self-awareness has not, however, translated into more procedural safeguards. The divergence between rising reflexivity and stagnant validation is the most consequential pattern in the corpus: awareness of the problem is not producing the checks that would address it.

Regarding the ontological stances coded across the corpus, we can identify three dominant framings: (1) “field as discourse/subdomains emerging from citation relations”, characteristic of keyword-based and relational designs; (2) “field with identifiable core venues and actionable framework”, characteristic of venue-based designs; and (3) “field as interdisciplinary integration”, characteristic of multi-pillar keyword designs (CS.11; CS.20). None of the 22 studies explicitly foregrounds the constructivist-performative view (that the mapping itself constitutes the field) even though this is the paper’s central theoretical claim. This absence is itself a finding: the studies reviewed either naturalize their boundary choices (reifying the corpus as the field) or acknowledge them in passing, but do not develop them as metatheoretical commitments.

The token-stabilization thesis advanced in Section 4.2, that repeated reuse of the same tokens and categories can stabilize a conventional image of the field, can be partially tested: of the 17 studies where Type I forms the primary (or co-primary) entry, 15 (88%) use some morphological variant of “sustainab*” as the core retrieval token. Citation analysis of this subset shows that several studies reference earlier keyword-based mappings (notably Kajikawa et al., 2007; Bettencourt & Kaur, 2011) as the source of their query strings, inheriting the same semantic frame rather than critically rederiving it. This genealogical reuse (treating prior search strings as validated anchors rather than provisional choices) is the mechanism through which the nominalist image of the field becomes self-reinforcing.

5. Discussion

5.1. Re-visiting the boundary problem and metatheoretical trade-offs

Three research questions organized this review. First, how have scientometrics studies delineated sustainability science, with what strategies and what metatheoretical assumptions? Second, what are the empirical consequences of these choices for how the field appears? Third, what would a reflexive, multi-perspectival alternative look like in practice? Our findings answer each in turn. On the first: keyword-based, analyst-defined strategies dominate (77% as primary entry), with limited recourse to actor-recognized anchors and no study adopting an explicitly realist boundary procedure. On the second: corpus sizes vary by up to three orders of magnitude (Table 2) and the disciplinary, geographic, and thematic portraits of the field shift systematically with boundary type (Section 4.4). On the third: the multi-lens protocol and minimal checklist in Box 1 operationalize a reflexive alternative that makes trade-offs explicit without imposing a single correct boundary.

Delimiting what constitutes “Sustainability Science” is not a trivial task; it epitomizes the classic boundary specification problem of social networks in science studies (Laumann, Marsden & Prensky, 1983). In scientometrics, delineation is a foundational step: it fixes who and what enter the scientific field and thereby shapes the observed size, topical profile, and network structure, as well as any indicators derived from them (Glänzel & Schubert, 2003; Zitt, 2005). As our review shows, the contours of Sustainability Science remain blurred and contested, an expansive field in search of its own coherence

and direction, with thematic boundaries still under construction (Vanhulst et al., 2025). This indeterminacy enables openness to new issues but also hinders consolidation of a shared agenda and a coherent identity for the field. Our findings make this performative dynamic empirically visible: the same underlying literature produces a small, institutionally coherent field (venue designs, ≈ 200–2,000 records) or a vast, heterogeneous discourse (keyword designs, up to 514,000 + records) depending solely on the analyst's entry strategy (Table 2).

Our review documents how this dilemma is handled in practice for Sustainability Science. The decisive boundary rules are overwhelmingly nominalist, especially query-based (Type I), for example, constructing corpora using keywords such as sustainab* in titles/abstracts/keywords. Source-based rules (Type II), like curated journal lists or database categories, and network-based rules (Type III), like keeping only the largest component or a citation-defined core, appear as tightening steps. Realist entries or filters are absent.

Each stance involves trade-offs. Strategies used are scalable, reproducible, and align with the affordances of bibliographic databases; but they risk reifying an analyst-constructed set as if it were the field. The use of some realist strategies promises aligning boundaries with community recognition (in principle, a realist stance treats Sustainability Science as a "collectively shared" social entity), but they can be resource-intensive, partial (e.g., privileging well-institutionalized segments), and difficult to implement for an evolving, multi-/inter-/transdisciplinary field. Given Sustainability Science's ongoing institutionalization and breadth, neither stance alone is fully satisfactory.

What follows, then, is a reflexive proposal: make boundary choices explicit; acknowledge their metatheoretical commitments (nominalist vs. realist); and inspect how conclusions vary under plausible alternatives. This reframing treat delineation not as a technical prelude but as a constitutive, theory-laden step that requires transparency.

5.2. Searching alternative approaches: potential of multi-perspectival and realist delineation

Building on these findings, we argue for a multi-perspectival approach for delineation that explicitly combines complementary lenses rather than relying on a single rule. Over time, designs in our corpus have moved from one-shot, query-based boundaries (Type I) toward hybrid logics, most often sequential and occasionally hybrid (see Asatani et al., 2020; Bettencourt & Kaur, 2011; Kajikawa et al., 2014; Yarime et al., 2010). Conceptually, this plural strategy operationalizes the idea of "multiple foci" (Feld, 1981), treating Sustainability Science simultaneously as a discourse (lexical/keywords), an institutional footprint (venues/categories), and a networked structure (relations/impacts). In practice, comparing outcomes across these lenses does not seek a single "correct" boundary; rather, it makes visible the trade-offs among inclusiveness, coherence, and thematic emphasis, turning boundary setting into a reflexive analytic step where differences in size, thematic profiles, and topology are reported and interpreted rather than hidden.

We propose three steps most studies can implement without excessive burden:

- Broad capture (discursive/lexical lens): use a carefully designed keyword strategy (or a topic model-assisted seed[2]), with explicit fields (title/abstracts/keywords) and time window.
- Core identification (structural or institutional lens): Apply either a relational filter (e.g., retain the largest direct-citation component or a k-core; or select clusters above a cohesion threshold) or a venue filter (e.g., retain items in a curated sustainability journal set / relevant subject categories).
- Compare & report: quantify size, thematic, and geographic differences between (1) and (2); note what is dropped/retained and why. Include a short Boundary Statement (stance, foci, sequence rule, thresholds).

Considering realist approaches, a full-field self-ascription survey is unrealistic at the scale of Sustainability Science. A more tractable route is to leverage actor-recognized anchors that already exist in the community. One option is to start from institutional memberships, mapping the outputs of researchers affiliated with well-recognized Sustainability Science centers and programs (e.g., the Stockholm Resilience Centre, the Earth Institute) and with global initiatives such as Future Earth or PECS (see for example Balvanera et al., 2017). In parallel, one can draw on conference rosters from reputable Sustainability Science meetings or associations, and on assessment bodies by compiling contributor lists from processes closely aligned with sustainability science problem-spaces (e.g., IPBES/IPCC working groups most relevant to sustainability science). Finally, instead of large-N surveys, targeted expert panels (mini-Delphi) can adjudicate edge cases: small, purposive groups review samples of top-cited clusters or borderline keyword hits to flag false positives/negatives and surface missing venues or concepts. Together, these anchors fold community recognition into an otherwise nominalist pipeline while staying within the validation repertoire we coded (expert review and face validity, and they make the contrast between analyst-defined and actor-recognized boundaries empirically visible without imposing prohibitive data-collection burdens.

As shown, Sustainability Science is multi-, inter-, and transdisciplinary and still in the process of institutionalization (Vanhulst et al., 2025). Its thematic sprawl and the absence of a single disciplinary anchor are constitutive rather than anomalous features of the field in formation. Calls within the field urge integration coupled with reflexivity, linking knowledge to action while cultivating plural epistemic practices (Clark & Harley, 2019; Kates et al., 2001; Spangenberg, 2011). A plural, reflexive delineation fits with that ethos: it accepts diversity while striving for coherence and makes trade-offs visible rather than tacit; the concise protocol summarized in Box 1. A clear, minimal checklist for reflexive delineation operationalizes this stance for analysts.

Box 1. A clear, minimal checklist for reflexive delineation

- State your stance & foci: declare nominalist/realist posture and which foci you use (query/sources/relations/impact/programs; acknowledge multiple foci when combined).
- Publish retrieval assets: verbatim queries (fields and timeframe), seed lists, venue lists, thresholds, and any embedding-based expansion terms.
- Precision/recall spot-checks: hand-code small samples to estimate false positives/negatives; report protocol and estimates.
- Expert review (targeted): a mini-panel validates queries/seed/sources/programs cases and flags omissions; record what changed.
- Alternative-boundary comparison: show how key findings change under a second boundary.
- Robustness: vary clustering/thresholds/sources; test temporal windows; replicate on a second database.
- Face validity: Check coherence with the field's history and known landmarks (e.g., expected hubs, journals, programs).
- Scope conditions: explicitly note coverage limits (database scope, languages), and known biases.

5.3 Limitations of the review and its inferences

Our synthesis inherits several constraints. Coverage: because the studies we analyzed are largely WoS/Scopus-based, our conclusions reflect that universe and under-represent non-indexed outlets, non-English scholarship, and grey literature. Language: English-language queries and metadata shape both how fields are defined and what is retrieved. Taxonomy dependency: venue/category filters import database design choices and the biases of curated journal lists. Temporal cut-offs: we review studies up to the present window; new techniques (e.g., more sophisticated semantic models) may spread, changing practice. Finally, no ground-truth partition exists: fields lack canonical boundaries. Our remedy is plural delineation with explicit reporting, not a search for a definitive border. Sample size: the corpus of 22 studies limits the generalizability of any pattern we describe. All findings should be treated as exploratory and warrant replication in larger corpora.

Conclusion

Delineating Sustainability Science is both an analytical and a constitutive act. The review of 22 scientometrics mappings shows a marked reliance on nominalist strategies, especially keywords-based retrievals, accompanied by growing, though still uneven, experimentation with hybrid designs. Validation practices remain limited overall, and alternative boundary rules often yield markedly different portraits of the field in size, thematic emphasis, disciplinary tilt, and geography distribution. These patterns confirm that boundary choices do not simply reflect Sustainability Science; they actively produce it, making explicit, reflexive practice indispensable.

Three implications follow. First, pluralization of boundary-setting is preferable to singular definitions. Sustainability Science is multi-, inter-, and transdisciplinary and still institutionalizing; no single criterion can capture its semantic breadth, evolving institutions, and relational structures. A multi-lens protocol (beginning with inclusive discursive capture, then refining by structural cohesion or venue attributes, and comparing outcomes) offers a practical route to balance recall and precision while keeping trade-offs

visible. The minimal checklist in Box 1 translates this stance into actionable steps, emphasizing publication of retrieval assets, alternative-boundary comparison, precision/recall sampling where feasible, targeted expert review, and robustness analyses across algorithms, thresholds, time windows, and databases.

Second, validation should be treated as a core component rather than an optional add-on. Absent explicit checks, corpus legitimacy and completeness remain uncertain, and results become difficult to compare across studies. Simple, auditable improvements are readily available: expert vetting of queries and edge cases; precision/recall spot-checks on small, hand-coded samples; sensitivity tests that vary clustering methods and parameters; temporal robustness via alternative windows; and cross-database replication. For the semantic layer, gold-standard seeding (e.g., foundational papers, flagship outlets, or prominent centers) and embedding-assisted expansion can reduce noise and under-coverage without sacrificing transparency, provided seeds and expansion terms are published.

Third, the integration of realist anchors (institutional rosters, conference participant lists, expert panels) with nominalist pipelines remains underexplored in sustainability science mapping. The most productive next step is not to choose between nominalist and realist strategies, but to make their divergence empirically legible: applying common boundary packages across the same time window and reporting variance as an epistemic resource rather than a methodological limitation.

In sum, field delineation in Sustainability Science should be approached as reflexive, comparative, and validated, with mixed methods that integrate discursive, institutional, and relational lenses, clear disclosure of assumptions, and systematic checks. A final self-reflexive note is warranted: this review is itself a delineation. We assembled a corpus of 22 studies through a title-anchored initial query and a citation-chasing supplement, strategies subject to precisely the limitations we document. A broader, abstract-level search (see the OpenAlex supplement described in Section 3.1) retrieved additional candidate studies, confirming that our corpus under-represents work that maps aspects of sustainability science without using the exact phrase in the title. We report this not to weaken our conclusions (the 22 reviewed studies support them consistently), but to illustrate the kind of reflexive self-accounting that could productively become standard practice in scientometric field mapping. The gap between a narrow title-based corpus and a broader abstract-based one is itself evidence for the paper's central claim: how you look determines what you find.

Declarations

The authors have no relevant financial or non-financial interests to disclose.

Ethics approval not applicable (the study is based on secondary data).

The dataset (coding scheme and list of reviewed studies) is available as Supplementary Material as Excel file with Id, full reference, year, DOI, and the variables used for coding: data source, unit of observation, data extraction, query string or seed, subject category used (if applicable), journal list used

(if applicable), time window start and end, corpus size, inclusion thresholds, sequence type (none, sequential, union), boundary sequence, boundary steps notes, expert review (validation), precision recall (validation), comparison with alternative boundary (validation), stability tests (validation), reflexivity level and evidence, ontological stance, main analytical techniques, main findings, claims about sustainability science status, Limitations recognized, notes for findings 4.1. – 4.4., and coding notes. The codebook is included in the Excel file.

Acknowledgements

This research was funded by the National Agency of Research and Development of Chile (ANID), Fondecyt Program, grant number 1220560. Julien Vanhulst and Alejandro Espinosa-Rada conceived the idea of the article, analyzed the data and wrote the manuscript. Julien Vanhulst designed methodology and collected the data. All authors gave final approval for publication.

Funding/Declarations

This work was supported by the National Agency of Research and Development of Chile (ANID), Fondecyt Program, grant number 1220560. Authors also thank the supercomputing infrastructure of the NLHPC (ECM-02, Chile).

References

1. Asatani, K., Takeda, H., Yamano, H., & Sakata, I. (2020). Scientific Attention to Sustainability and SDGs: Meta-Analysis of Academic Papers. Energies, 13(4), Article 4. https://doi.org/10.3390/en13040975
2. Balvanera, P., Daw, T. M., Gardner, T. A., Martín-López, B., Norström, A. V., Speranza, C. I., Spierenburg, M., Bennett, E. M., Farfan, M., Hamann, M., Kittinger, J. N., Luthe, T., Maass, M., Peterson, G. D., & Perez-Verdin, G. (2017). Key features for more successful place-based sustainability research on social-ecological systems: A Programme on Ecosystem Change and Society (PECS) perspective. Ecology and Society, 22(1). https://www.jstor.org/stable/26270056
3. Bellotti, E., & Espinosa-Rada, A. (2025). Scientific networks. In Handbook of Culture and Social Networks (pp. 154–167). Edward Elgar Publishing. https://www.elgaronline.com/edcollchap/book/9781803928784/book-part-9781803928784-18.xml
4. Bettencourt, L. M. A., & Kaur, J. (2011). Evolution and structure of sustainability science. Proceedings of the National Academy of Sciences, 108(49), 19540–19545. https://doi.org/10.1073/pnas.1102712108
5. Blanco, J., Moreau, C., Carrière, S. M., Fache, E., Cué Rio, M., Calatayud, F., Castella, J.-C., Le Meur, P.-Y., Coudel, É., Hervé, D., Méral, P., & Therville, C. (2025). Gaps and overlaps between sustainability science and the environmental humanities. Sustainability Science, 20(2), 581–596. https://doi.org/10.1007/s11625-024-01596-1

6. Borgatti, S. P., Everett, M. G., Johnson, J. C., & Agneessens, F. (2024). Analyzing Social Networks. SAGE Publications.
7. Breiger, R. L. (1976). Career Attributes and Network Structure: A Blockmodel Study of a Biomedical Research Specialty. American Sociological Review, 41(1), 117–135. https://doi.org/10.2307/2094376
8. Brewer, G. D. (1999). The challenges of interdisciplinarity. Policy Sciences, 32(4), 327–337. https://doi.org/10.1023/A:1004706019826
9. Buter, R. K., & Van Raan, A. F. J. (2012). Identification and analysis of the highly cited knowledge base of sustainability science. Sustainability Science, 8(2), 253–267. https://doi.org/10.1007/s11625-012-0185-1
10. Clark, W. C., & Harley, A. G. (2019). Sustainability Science: Toward a Synthesis. Annual Review of Environment and Resources, 45(1), 331–386. https://doi.org/10.1146/annurev-environ-012420-043621
11. Cox, R. W. (1981). Social forces, states and world orders: Beyond international relations theory. Millennium, 10(2), 126–155. https://doi.org/10.1177/03058298810100020501
12. Crane, D. (1972). Invisible Colleges: Diffusion of Knowledge in Scientific Communities. University of Chicago Press.
13. de Solla Price, D. J. (1965). Networks of Scientific Papers. Science, 149(3683), 510–515. https://doi.org/10.1126/science.149.3683.510
14. Espinosa-Rada, A., Bellotti, E., Everett, M. G., & Stadtfeld, C. (2024). Co-evolution of a socio-cognitive scientific network: A case study of citation dynamics among astronomers. Social Networks, 78, 92–108. https://doi.org/10.1016/j.socnet.2023.11.008
15. Espinosa-Rada, A., Lerner, J., & Fritz, C. (2025). Socio-cognitive networks between researchers: Investigating scientific dualities with the group-oriented relational hyperevent model. Social Networks, 83, 1–13. https://doi.org/10.1016/j.socnet.2025.04.005
16. Fang, X., Zhou, B., Tu, X., Ma, Q., & Wu, J. (2018). “What Kind of a Science is Sustainability Science?” An Evidence-Based Reexamination. Sustainability, 10(5), 1478. https://doi.org/10.3390/su10051478
17. Feld, S. L. (1981). The Focused Organization of Social Ties. American Journal of Sociology, 86(5), 1015–1035. https://doi.org/10.1086/227352
18. Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. Scientometrics, 56(3), 357–367. https://doi.org/10.1023/A:1022378804087
19. Gläser, J., & Laudel, G. (2013). Life With and Without Coding: Two Methods for Early-Stage Data Analysis in Qualitative Research Aiming at Causal Explanations. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 14(2). https://doi.org/10.17169/fqs-14.2.1886
20. Hagstrom, W. O. (1965). The Scientific Community. Basic Books.

21. Hassan, S.-U., Haddawy, P., & Zhu, J. (2013). A bibliometric study of the world's research activity in sustainable development and its sub-areas using scientific literature. Scientometrics, 99(2), 549–579. https://doi.org/10.1007/s11192-013-1193-3
22. Horcea-Milcu, A.-I., Martin-Lopez, B., Lam, D. P. M., & Lang, D. J. (2020). Research pathways to foster transformation: Linking sustainability science and social-ecological systems research. Ecology & Society, 25(1), 1–30.
23. Jerneck, A., Olsson, L., Ness, B., Anderberg, S., Baier, M., Clark, E., Hickler, T., Hornborg, A., Kronsell, A., Lövbrand, E., & Persson, J. (2011). Structuring sustainability science. Sustainability Science, 6(1), 69–82. https://doi.org/10.1007/s11625-010-0117-x
24. Kajikawa, Y. (2008). Research core and framework of sustainability science. Sustainability Science, 3(2), 215–239. https://doi.org/10.1007/s11625-008-0053-1
25. Kajikawa, Y., Ohno, J., Takeda, Y., Matsushima, K., & Komiyama, H. (2007). Creating an academic landscape of sustainability science: An analysis of the citation network. Sustainability Science, 2(2), 221–231. https://doi.org/10.1007/s11625-007-0027-8
26. Kajikawa, Y., Tacoa, F., & Yamaguchi, K. (2014). Sustainability science: The changing landscape of sustainability research. Sustainability Science, 9(4), 431–438. https://doi.org/10.1007/s11625-014-0244-x
27. Kates, R. W., Clark, W. C., Corell, R., Hall, J. M., Jaeger, C. C., Lowe, I., McCarthy, J. J., Schellnhuber, H. J., Bolin, B., Dickson, N. M., Faucheux, S., Gallopin, G. C., Grübler, A., Huntley, B., Jäger, J., Jodha, N. S., Kasperson, R. E., Mabogunje, A., Matson, P., … Svedin, U. (2001). Sustainability Science. Science, 292(5517), 641–642. https://doi.org/10.1126/science.1059386
28. Keuchenius, A., Törnberg, P., & Uitermark, J. (2021). Adoption and adaptation: A computational case study of the spread of Granovetter's weak ties hypothesis. Social Networks, 66, 10–25. https://doi.org/10.1016/j.socnet.2021.01.001
29. Kornfeld, W. A., & Hewitt, C. E. (1981). The Scientific Community Metaphor. IEEE Transactions on Systems, Man, and Cybernetics, 11(1), 24–33. IEEE Transactions on Systems, Man, and Cybernetics. https://doi.org/10.1109/TSMC.1981.4308575
30. Krippendorff, K. (2018). Content Analysis: An Introduction to Its Methodology. SAGE Publications.
31. Laumann, E. O., Marsden, P. V., & Prensky, D. (1983). The boundary specification problem in network analysis. In R. Burt & M. Minor (Eds.), Applied network analysis (pp. 18–33). SAGE Publications.
32. Leydesdorff, L., & Probst, C. (2009). The delineation of an interdisciplinary specialty in terms of a journal set: The case of communication studies. Journal of the American Society for Information Science and Technology, 60(8), 1709–1718. https://doi.org/10.1002/asi.21052
33. Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. Journal of the American Society for Information Science and Technology, 60(2), 348–362. https://doi.org/10.1002/asi.20967
34. Lietz, H. (2020). Drawing impossible boundaries: Field delineation of Social Network Science. Scientometrics, 125(3), 2841–2876. https://doi.org/10.1007/s11192-020-03527-0

35. Lievrouw, L. A., Rogers, E. M., Lowe, C. U., & Nadel, E. (1987). Triangulation as a research strategy for identifying invisible colleges among biomedical scientists. Social Networks, 9(3), 217–248. https://doi.org/10.1016/0378-8733(87)90021-9
36. Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. Human Communication Research, 28(4), 587–604. https://doi.org/10.1111/j.1468-2958.2002.tb00826.x
37. Mahmoud, Y., Sousa-Poza, A., Wiek, A., & Lang, D. (2018). At the nexus of problem-solving and critical research. Ecology and Society, 23(4), 46. https://doi.org/10.5751/ES-10607-230446
38. Maltseva, D., & Batagelj, V. (2019). Social network analysis as a field of invasions: Bibliographic approach to study SNA development. Scientometrics, 121(2), 1085–1128. https://doi.org/10.1007/s11192-019-03193-x
39. McLevey, J., Scott, J., & Carrington, P. J. (2023). The Sage Handbook of Social Network Analysis. SAGE Publications.
40. Merton, R. K. (1973). The Sociology of Science: Theoretical and Empirical Investigations. University of Chicago Press.
41. Miles, M. B., Huberman, A. M., & Saldana, J. (2013). Qualitative Data Analysis: A Methods Sourcebook. SAGE Publications.
42. Moody, J. (2004). The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999. American Sociological Review, 69(2), 213–238. https://doi.org/10.1177/000312240406900204
43. Morris, S. A., & Van der Veer Martens, B. (2008). Mapping research specialties. Annual Review of Information Science and Technology, 42(1), 213–295. https://doi.org/10.1002/aris.2008.1440420113
44. Mulkay, M. J. (1979). Science and the Sociology of Knowledge. G. Allen & Unwin.
45. Mullins, N. C. (1972). The development of a scientific specialty: The phage group and the origins of molecular biology. Minerva, 10(1), 51–82. https://doi.org/10.1007/BF01881390
46. Mullins, N. C., & Mullins, C. J. (1973). Theories and theory groups in contemporary American sociology. Harper & Row.
47. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. Systematic Reviews, 10(1), 89. https://doi.org/10.1186/s13643-021-01626-4
48. Pandiella-Dominique, A., Bautista-Puig, N., & Filippo, D. D. (2021). Detecting trends in sustainability publications: Research development and dynamics in “Green and sustainable science and technology” category. International Journal of Innovation and Sustainable Development. (world). https://www.inderscienceonline.com/doi/pdf/10.1504/IJISD.2021.118375

49. Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. Scientometrics, 81(3), 719–745. https://doi.org/10.1007/s11192-008-2197-2
50. Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. Journal of the American Society for Information Science and Technology, 61(9), 1871–1887. https://doi.org/10.1002/asi.21368
51. Schirone, M. (2024). The formation of a field: Sustainability science and its leading journals. Scientometrics, 129(1), 401–429. https://doi.org/10.1007/s11192-023-04877-1
52. Schreier, M. (2012). Qualitative Content Analysis in Practice. SAGE Publications Ltd. https://doi.org/10.4135/9781529682571
53. Scott, J. (2017). Social Network Analysis. SAGE.
54. Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for Information Science and Technology, (24), 265–269.
55. Spangenberg, J. H. (2011). Sustainability science: A review, an analysis and some empirical lessons. Environmental Conservation, 38(3), 275–287. https://doi.org/10.1017/S0376892911000270
56. Vanhulst, J., Padilla-Navarro, P., Espinosa-Rada, A., Cantillan, R., Velásquez, R., & González Tapia, K. (2025). The reflexive process in sustainability science: A short critical review. Environmental Science & Policy, 171, 104145. https://doi.org/10.1016/j.envsci.2025.104145
57. Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. Journal of the American Society for Information Science and Technology, 63(12), 2378–2392. https://doi.org/10.1002/asi.22748
58. White, H. D. (2011). Scientific and Scholarly Networks. In J. Scott & P. J. Carrington (Eds.), The SAGE Handbook of Social Network Analysis (pp. 271–285). SAGE.
59. Whitley, R. (2000). The Intellectual and Social Organization of the Sciences. Oxford University Press.
60. Yarime, M., Takeda, Y., & Kajikawa, Y. (2010). Towards institutional analysis of sustainability science: A quantitative examination of the patterns of research collaboration. Sustainability Science, 5(1), 115–125. https://doi.org/10.1007/s11625-009-0090-4
61. Zitt, M. (2005). Facing Diversity of Science: A Challenge for Bibliometric Indicators. Measurement: Interdisciplinary Research and Perspectives, 3(1), 38–49. https://doi.org/10.1207/s15366359mea0301_6
62. Zitt, M., Lelu, A., Cadot, M., & Cabanac, G. (2019). Bibliometric Delineation of Scientific Fields. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), Springer Handbook of Science and Technology Indicators (pp. 25–68). Springer Verlag.
63. Zuccala, A. (2006). Modeling the invisible college. Journal of the American Society for Information Science and Technology, 57(2), 152–168. https://doi.org/10.1002/asi.20256

Footnotes

1. However, scientific fields are not confined to scientometrics studies; their exploration and further development have also been a focus of social network research employing multiple methods, including bibliometrics, surveys, documentary analysis, and qualitative approaches (Chubin, 1976; Lievrouw et al., 1987; Mullins, 1972).
2. For a stronger semantic basis for step 1, two simple upgrades are feasible and transparent (even if they weren't widely used in our sample):
   Gold-standard seeding: curate a compact seed (e.g., seminal Sustainability Science papers; a short list of flagship SS journals; core centers/programs) and estimate precision/recall of your keyword design against that seed.
   Semantic expansion: learn word embeddings (or use contextual embeddings) on the retrieved corpus to expand/refine the keyword list (e.g., add high-similarity terms) and rerun retrieval. Keep the process auditable by publishing the seed list and expansion terms.
   Neither addition makes the protocol unwieldy, and both directly address the known semantic diffuseness of naïve *sustainab** queries.

Figures

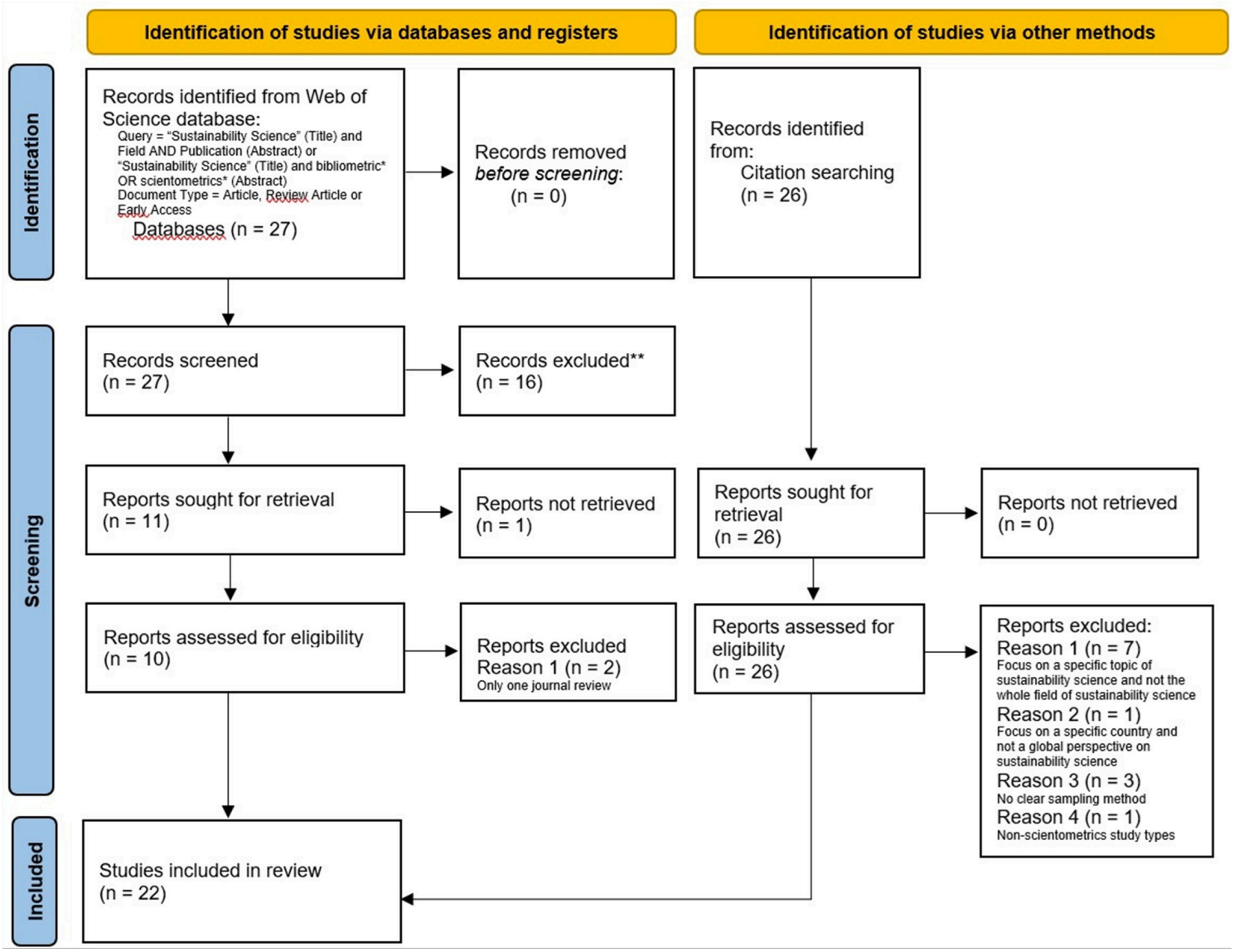


Figure 1

PRISMA 2020 flow diagram

Source: own elaboration based on (Page et al., 2021).

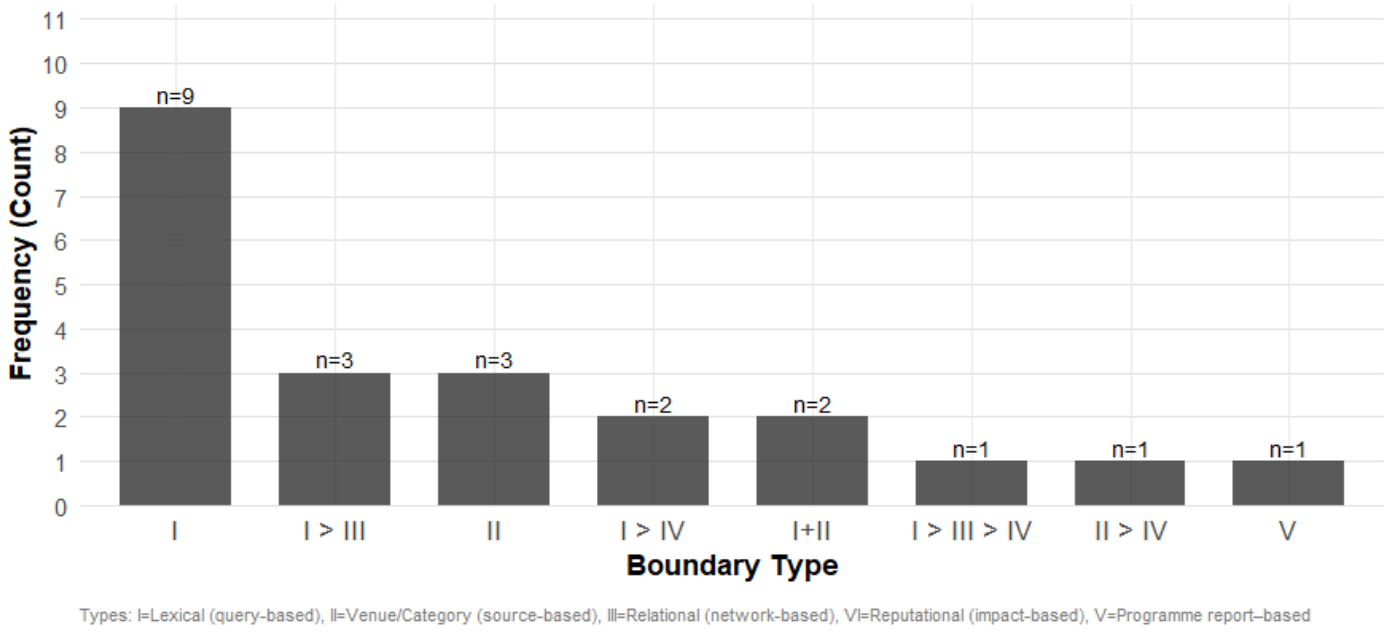


Figure 2

Distribution of Boundary Types (I-V)

Source: own elaboration

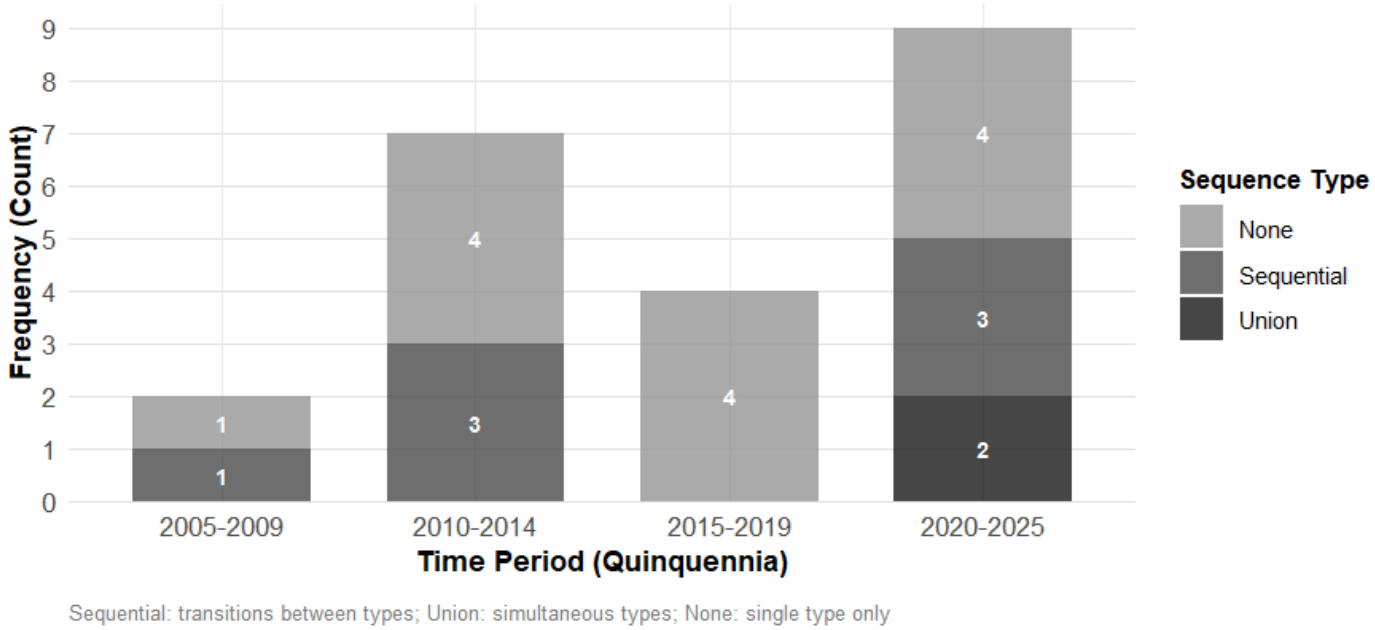


Figure 3

Temporal evolution of boundary combination logic by quinquennia (N=22)

Source: own elaboration.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.xlsx
- AppendixA.docx