Participatory conceptual mapping with guardrailed large language models: An audit-ready DAPSI(W)R(M) workflow from workshop transcripts to scenario testing

Maik Neukirch[1*], Xavier Garcia[1], Elisa Berdalet[1], Aurelija Armoskaite[2], Giulia Bellon[3], Miriam von Thenen[3], and Erasmia Kastanidi[4]

[1]Institute of Marine Sciences (ICM-CSIC), Barcelona, Spain
[2]Latvian Institute of Aquatic Ecology, Riga, Latvia
[3]Leibniz Institute for Baltic Sea Research Warnemünde, Rostock, Germany
[4]Hellenic Centre for Marine Research (HCMR), Institute of Oceanography, Attiki, Greece

Abstract

Participatory conceptual mapping is widely used in sustainability research to represent stakeholders' knowledge, concerns, and perceptions of causal relationships within complex social-ecological systems. Yet translating workshop dialogue into reliable, auditable causal maps is labor-intensive and prone to interpretive drift. We present a four-step workflow combining guardrailed large language model (LLM) assistance with human oversight to transform workshop transcripts into DAPSI(W)R(M)-consistent conceptual maps, feedback structures, and qualitative scenarios. The workflow comprises (1) evidence-anchored extraction of nodes and directional links; (2) controlled merging of workshop maps into an identifier-stable representation; (3) identification of vicious cycles, focal points, and virtuous cycles using only the verified combined structure; and (4) scenario tracing through validated pathways. Guardrails include a no-invention rule, quotation-level evidence anchoring, alias tracking, stable identifiers, clean-slate re-entry, and human verification. Applied to two workshops on sustainable fisheries in Catalonia, Spain, the workflow consolidated 64 initial themes and 242 links into 43 themes and 243 verified relations. Three major vicious-cycle systems, several focal points, and twelve virtuous loops and chains informed ten stakeholder-style qualitative scenarios. The case study suggests that guardrailed LLM assistance can support synthesis of stakeholder-derived knowledge while keeping analytical decisions traceable. The workflow provides an auditable foundation for AI-supported participatory systems analysis and identifies future refinements including polarity checks, indicator integration, and empirically testable scenarios.

Keywords

participatory conceptual mapping; guardrails; large language models; fisheries; traceability

1. Introduction

Participatory conceptual mapping is a foundational tool in sustainability research and stakeholder-driven system analysis (Sedlacko et al., 2014; Hayward et al., 2020). It enables diverse groups, such as local communities, practitioners, and policy experts, to articulate complex social-ecological systems through shared diagrams that reveal drivers, pressures, feedbacks, and potential responses. In marine governance, frameworks such as the

Correspondence:
* Contact M. Neukirch at maik.neukirch@icm.csic.es

Cite this article as:
Neukirch, M., Garcia, X.,Berdalet, E., Armoskaite, A., Bellon, G., von Thenen, M., & Kastandidi, E. Participatory conceptual mapping with guardrailed large language models: An audit-ready DAPSI(W)R(M) workflow from workshop transcripts to scenario testing
Socio-Environmental Systems Modelling, vol. 8, 19051, 2026, doi:10.18174/sesmo.19051

This work is **licensed** under a Creative Commons Attribution-NonCommercial 4.0 International License.

Socio-Environmental Systems Modelling
An Open-Access Scholarly Journal
http://www.sesmo.org

DAPSI(W)R(M) concept (Elliott et al., 2017) make these representations especially valuable by structuring how human activities generate pressures, affect ecosystem states, and trigger impacts and responses. Following Elliott et al. (2017), we use DAPSI(W)R(M) to distinguish *drivers* and *activities* that generate *pressures*, their effects on *state* and *impacts* (on human *welfare)*, and the *responses (measures)* they motivate. Yet as participatory exercises scale up, translating rich, often messy workshop dialogue into reliable, auditable conceptual maps becomes increasingly demanding. Manual coding of transcripts is widely reported to be time-intensive and interpretively demanding; moreover, documentation of how stakeholder narratives are transformed into formal structures is often incomplete (Newberry & Carhart, 2024; Powell et al., 2024).

Recent advances in *large language models* (LLMs) and related *artificial intelligence* (AI) tools offer new opportunities to support this work. LLMs can process large volumes of workshop transcripts, assist in extracting candidate themes and causal links, and help identify feedback structures embedded in stakeholder discourse. Evaluations of text-to-map systems show that performance improves when models are required to cite supporting text and operate under retrieval-based grounding (Hosseinichimeh et al., 2024). Related work in bio-ontology points to the value of retrieval-based and evidence-anchored LLM use (Manda, 2025). Such models can also support the harmonization of maps from multiple workshops, surface potential overlaps or gaps between stakeholder groups, and assist in generating internally consistent scenario narratives. When used under explicit guardrails, such as no content invention, strict evidence anchoring, and human verification, these tools suggest not only efficiency gains but also more systematic treatment of stakeholder language and logic (Hosseinichimeh et al., 2024; White & Barbrook-Johnson, 2025; Sharma et al., 2025).

The emerging literature on AI-supported systems mapping has documented important building blocks of such workflows. Studies have shown that LLMs can extract cause-effect pairs from text, propose initial causal loop diagrams, and use semantic similarity to suggest merges between related concepts, provided that polarity checks and provenance tracking are enforced (Hosseinichimeh et al., 2024; White & Barbrook-Johnson, 2025). Methodological contributions include structured prompting to avoid content invention, explicit loop-closure queries, human-in-the-loop verification of links and signs, and versioned pipelines that store maps as stable, machine-readable tables (Reinholtz et al., 2025; Liu et al., 2025). Across this work, a consistent theme is that LLMs should function as amplifiers of human analysis rather than autonomous modelers, with stakeholders and researchers retaining interpretive authority (Sharma et al., 2025; Quan & Lee, 2025).

Despite these advances, several gaps remain. Much of the current work concentrates on the front end of the pipeline, especially the extraction of variables and pairwise causal relations from text into causal maps, including causal-loop-diagram-style feedback representations. Semi-automated causal loop diagram approaches tend to prioritize this extraction step over consolidation, cycle detection, and scenario evaluation (Newberry & Carhart, 2024). Work on manual causal mapping emphasizes careful coding and representation of causal claims (Powell et al., 2024), and research on LLM-supported causal loop diagrams primarily concentrates on extracting and validating links and loops from textual sources (Hosseinichimeh et al., 2024; Liu & Keith, 2025). In participatory conceptual-mapping processes that are further developed into cross-workshop synthesis and scenario exploration, analysts may also need to combine maps from multiple workshops while preserving stakeholder nuance, identify and audit vicious and virtuous cycles, including focal points, and trace how stakeholder-relevant interventions propagate through system structure. Emerging approaches to reproducible LLM-orchestrated pipelines for causal loop diagram construction (e.g., Reinholtz et al., 2025) address parts of this agenda, but there is still limited practical guidance on how to operationalize similar guardrails across an end-to-end workflow that spans extraction, merging, feedback-structure analysis, and scenario-based reasoning.

This manuscript addresses these gaps by presenting and demonstrating a four-step, audit-ready workflow for AI-assisted participatory conceptual mapping. The workflow (1) extracts themes and directional links from stakeholder workshop transcripts under a strict no-invention rule; (2) combines individual workshop maps into a single, identifier-stable representation (i.e., identifiers remain unchanged when labels are revised, and original workshop identifiers are preserved during cross-workshop merging) that reconciles shared conversation starters and DAPSI(W)R(M) framing with bottom-up stakeholder nuance; (3) analyses this combined map to identify vicious cycles, focal points, and stakeholder-anchored virtuous cycles; and (4) tests the resulting system through qualitative scenarios framed as stakeholder-style “what-if” questions. At each step, we formalize guardrails such as evidence anchoring, alias and glossary management, stable *identifiers* (IDs), step-wise reports, and clean-slate re-entry artefacts, so that every causal claim and analytical move remains traceable and revisable. These design

choices follow methodological guidance that emphasizes provenance-preserving causal-mapping workflows and version-controlled checkpoints (Powell et al., 2024; Reinholtz et al., 2025).

We demonstrate the workflow using two participatory workshops on sustainable fisheries in Catalonia, Spain, which we refer to as the Costa Brava case study, and apply it to DAPSI(W)R(M)-consistent maps derived from stakeholder transcripts and to a set of qualitative scenario questions. The aims of the paper are threefold: (i) to specify a transparent, auditable workflow for LLM-assisted participatory conceptual mapping; (ii) to demonstrate, through the fisheries example, how the workflow retains traceability to stakeholder contributions while supporting feedback-focused diagnosis and scenario exploration; and (iii) to use the example to reflect on methodological strengths, limitations, and avenues for refinement. A distinctive feature of the workflow is that models are restricted to extracting information from transcripts and validated tables under a strict no-invention rule, and that each step restarts from human-verified CSV (*comma separated values*) artefacts ("clean-slate re-entry") rather than from previous chats, which helps curb both unsupported causal links and gradual drift across iterations. Section 2 presents the methods, including the workshops, transcript handling, AI guardrails, and the four-step workflow. Section 3 reports the results of applying the workflow to the Costa Brava case study. Section 4 discusses methodological and practical implications, and Section 5 concludes.

2. Methods

2.1 Case study and workshops

The empirical basis for the methods presented below consists of two participatory workshops on sustainable fisheries in Catalonia, Spain. Workshop 1 (Barcelona) included three academic stakeholders and four team members; Workshop 2 (Llançà) engaged ten fisheries-sector stakeholders and four team members. Both addressed the question, “How can sustainable fisheries be achieved by 2040 in Catalonia?” and began from the same eight whiteboard conversation starters (Water Quality, Human Consumption, Regulation, Temperature, Users, Pollution, Infrastructure, and Knowledge), which oriented discussion while preserving stakeholder discourse.

Both workshops began with these eight conversation starters written on the whiteboard. During the guided discussion, stakeholders were repeatedly prompted to explain how these themes interacted, what they resulted in, and what affected them. The facilitation team recorded this discussion directly on the whiteboard, progressively extending it into a collaborative conceptual map by adding stakeholder-defined themes and directional connections. The workshop audio recordings therefore capture the guided discussion from which the whiteboard conceptual maps were produced, although in Barcelona they also include a subsequent more open discussion.

In Barcelona, the workshop lasted approximately three hours and combined guided discussion with a more open, exploratory format. In Llançà, the workshop lasted similarly long but relied more heavily on guided discussion, followed by a collaborative questionnaire segment. This difference in format reflected learning across the process and the use of the first workshop to better understand which forms of facilitation most effectively elicited stakeholder reasoning. The two workshops were designed to capture contrasting stakeholder perspectives relevant to the case study, rather than to constitute a statistically representative sample.

2.2 Data sources, language, translation, and transcript preparation

The workflow relied on transcripts of workshop audio recordings, the collaborative whiteboard conceptual maps developed during the discussions, and workshop notes. Both workshops were conducted in Catalan, while manuscript reporting is in English. Where translation was required for reporting, original quotations were retained alongside their English renderings in the audit trail.

To prepare transcripts for evidence anchoring, raw dialogue was cleaned by reformulating fragmented speech into complete sentences, removing repetition and filler words, and standardizing grammar while preserving meaning and tone. The cleaned transcript was then segmented into coherent excerpts (approximately twenty seconds) to maintain local context and enable quotation-and-timestamp verification. The cleaned and segmented transcripts formed the basis for evidence-anchored node and link extraction. Evidence quotations

are taken from these cleaned transcripts rather than directly from unedited speech. The original audio recordings and raw transcripts are retained. Transcript quality varied between workshops because recording conditions differed: the Barcelona workshop relied on a single consumer-grade microphone and contained some inaudible sections, whereas the Llançà workshop used multiple recording devices and yielded more continuous coverage. These differences affected transcript preparation and verification effort, but not the procedural logic of the workflow. The change reflected our learning process with respect to the need for redundancy in robust audio data acquisition.

2.3 AI setup and general guardrails

We implement a four-step workflow (summarized in Figure 1) that transforms stakeholder workshop discussions into a stakeholder-faithful conceptual system for diagnostic mapping (Steps 1 and 2), feedback-structure analysis (Step 3), and scenario exploration (Step 4).

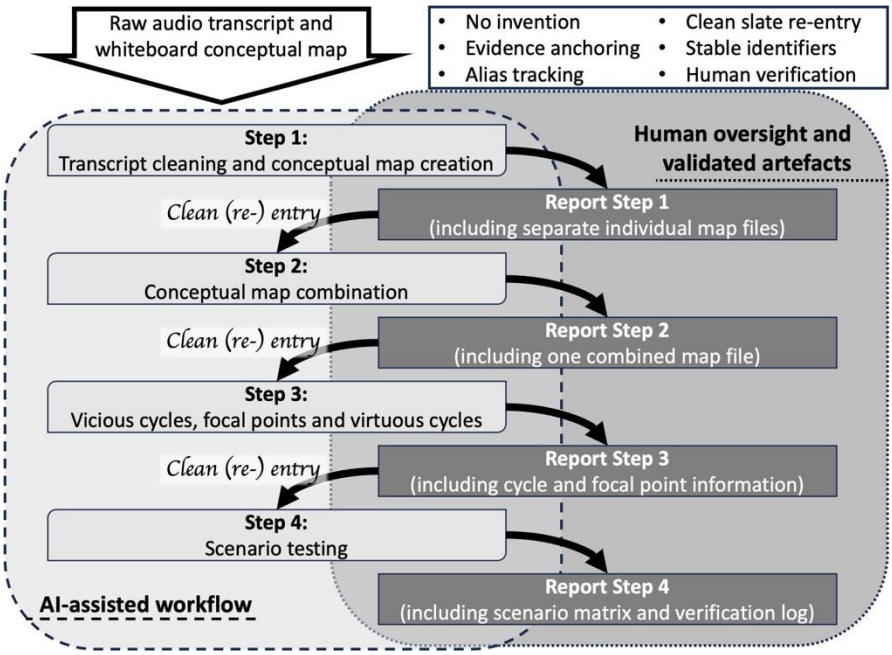


Figure 1: Audit-ready, guardrailed LLM-assisted participatory conceptual mapping workflow. Starting from workshop audio transcripts and whiteboard conceptual maps, the four-step workflow produces a progressively consolidated and auditable representation of stakeholder reasoning. Step 1 cleans and segments transcripts and extracts evidence-anchored nodes and directional links for individual workshop maps. Step 2 combines workshop maps into an ID-stable representation with alias handling and decision logging. Step 3 identifies focal points and vicious/virtuous feedback structures using only the verified combined map. Step 4 traces stakeholder-style scenarios through the validated structure to generate reviewable, decision-support narratives. Across all steps, the workflow is constrained by six guardrails: (i) no-invention, (ii) evidence anchoring, (iii) alias tracking, (iv) clean-slate re-entry, (v) stable identifiers, and (vi) human verification; and each step produces a reportable artefact for audit and reuse.

Two procedural principles guide the workflow. First, transcript-based map construction is distinguished from subsequent analysis of the finalized combined map: transcripts are used initially only to identify nodes and directional links, while the reported cycles, focal points, and scenarios are developed later using the verified combined structure; at later steps, transcripts are used only to verify consistency with stakeholder intent. Second, the workflow formalizes evidence anchoring and versioning: each causal statement is linked to a quotation from the cleaned transcript and its timestamp, each element receives a stable identifier from its first appearance, and each step ends with a short, audited report documenting decisions. The next step starts from these verified CSV artefacts and step-reports, rather than from raw chat or prompt history, creating an explicit reset point (“clean-slate re-entry”) that preserves validated content and reduces incremental drift (White & Barbrook-Johnson, 2025; Sharma et al., 2025; Reinholtz et al., 2025).

AI assistance was implemented through ChatGPT models available under business-plan access between March and October 2025. Multiple models were tested during workflow development, and several were used in producing the final results. Use focused on the provided project materials, including deep-search functions applied to those materials where broader exploration of the project corpus was needed, rather than on open-web search. No model-specific parameter tuning was applied beyond model selection and the activation of the deep-search function. Because multiple models were used across development and validation, exact model-level reproducibility is not claimed for this implementation.

Across all steps, LLM assistance is governed by six constraints: (i) no-invention of content; (ii) strict evidence anchoring with quotations and timestamps; (iii) alias tracking; (iv) stable identifiers; (v) step-reports with clean-slate re-entry artefacts; and (vi) human verification at all steps (cf. Powell et al., 2024; Manda, 2025). These constraints serve epistemic, social, and ethical purposes. Epistemically, evidence anchoring and clean-slate re-entry are intended to limit unsupported causal claims and interpretive drift. Socially, preserving original wording as aliases, workshop provenance, and related but distinct themes helps retain differences between stakeholder contributions during synthesis (Section 2.5). Ethical safeguards include consent for AI-assisted analysis, confidentiality provisions, and researcher accountability for analytical decisions (Section 2.8).

The no-invention rule constrains what the AI may add or infer; it is not used to rank or filter transcript-supported stakeholder concerns. Accordingly, themes supported by the transcript can be retained even when they extend beyond the initial whiteboard map, while ambiguous meanings are left unspecified rather than resolved through model inference. DAPSI(W)R(M) coding allows more than one category per node according to its meaning in the transcript (Supplementary Material A.1). Nevertheless, coding and the selection of actionable feedback structures remain researcher-defined analytical choices, not neutral reproductions of stakeholder knowledge.

The definitive system of record is a set of structured tables (CSV files), which are interpreted rather than replaced by narrative text. Stable identifiers (Node-IDs, Link-IDs, and later Cycle-IDs) are assigned at first mention and retained across steps, including when labels change, while an alias glossary records alternative labels, keep/drop decisions, and brief rationales. Supplementary Material C illustrates how unchanged node identifiers allow the development of the Workshop 2 map to be traced despite changes in node names. AI-assisted similarity checks are used only to surface candidates for human review, while final merge/non-merge decisions remain with researchers (Hosseinichimeh et al., 2024; Reinholtz et al., 2025).

Each step concludes with checks against the source material for unsupported content, logical consistency, completeness, and whether link directions and any stated signs match the quoted context. These checks form part of an iterative, researcher-supervised process. As specified in Supplementary Material A, outputs are reviewed in short batches, with corrections recorded in the working tables and decisions documented in a log.

When the model invents content or reinterprets definitions, the prescribed response is to stop, clarify the relevant prompt instructions, and rerun the affected batch from verified inputs. The no-invention and evidence-anchoring commitments remain fixed, while their application through prompts and evidence-based interpretations can be refined. This permits correction of interpretations without relaxing the requirement that feedback and scenario analyses use only the verified structure.

2.4 Step 1: Cleaning transcripts and extracting nodes and links

Step 1 translates workshop dialogue into a structured representation of stakeholder reasoning. It produces a node list capturing stakeholder-defined concepts and a set of directional links describing how one concept influences another. Extraction is constrained to transcript-supported concepts and relations. Researcher-led refinement also made implicit bridging concepts explicit where the dialogue supported the decomposition of a complex relationship into intermediate steps (Section 3.1). Conceptual plausibility alone was not sufficient for inclusion.

In this implementation, polarity (i.e. whether one component has an increasing or decreasing effect on another component) was not assigned systematically. When stakeholder statements implied different effects in different excerpts, the sign was recorded as context-conditional; otherwise, it was left unspecified. Final sign assessment for links was deferred to Step 3, before candidate feedback structures were classified as vicious or virtuous.

Extraction proceeds in two passes. In the first pass, the AI identifies distinct conceptual nodes under a strict no-invention rule; uncertain meanings are left blank. Node records preserve participants' wording, include concise definitions, and attach representative quotations and timestamps plus a workshop tag, following evidence-anchored practices (White and Barbrook-Johnson, 2025). AI assistance is then used to flag potentially similar labels for an initial alias glossary; researchers confirm or reject merges. In the second pass, the AI identifies directional links between nodes and records source and target concepts, any sign indicated in the excerpt (where present), a short justification, and the supporting quotation and timestamp. Researchers review all candidate nodes and links and decide whether to include, modify, or remove them based on transcript evidence. Step 1 produces a node table, a link table, an initial alias log, and a step report. Prompt format examples are documented in Supplementary Material A.

2.5 Step 2: Combining individual maps into a single, ID-stable map

Step 2 produces a single combined map from two or more workshops by harmonizing labels, reconciling links, and preserving provenance without introducing structure not grounded in stakeholder statements. We apply controlled merging in three tiers: (i) exact or near-exact duplicates are combined; (ii) strong conceptual overlaps are merged only when analytic resolution can be preserved; and (iii) related but distinct themes remain separate. For borderline cases, we use a conservative "two of three" test across (a) definition equivalence, (b) transcript-anchor equivalence, and (c) indicator compatibility; if uncertainty remains, both terms are retained and their relationship is recorded in the alias glossary with a short rationale. AI tools are used only to surface candidate overlaps for human review; final merge/non-merge decisions remain with researchers (Hosseinichimeh et al., 2024; Reinholtz et al., 2025).

Because Step 1 does not create cross-workshop links, any cross-workshop relation proposed in Step 2 is treated as a blind link and screened explicitly for evidence. We assemble a cross-workshop node matrix and grade candidate relations as Tier 1 (explicitly evidenced in transcripts from all relevant workshops), Tier 2 (explicitly evidenced in one workshop and clearly implied as a continuation in the other), or Tier 3 (contextually plausible but weakly evidenced). These tiers assess evidence for the relationship in the dialogue, not whether both endpoint concepts were separately identified as nodes in each workshop. Only Tier 1 and Tier 2 relations that pass human review are added to the combined link table; Tier 3 items are logged but not drawn. Throughout, polarity is treated as context-conditional: different signs can occur in different excerpts and are logged separately; loop labelling is deferred to Step 3, where reviewers revisit quotations for any links that enter candidate cycles. Step 2 closes with a consistency scan (duplicate labels, unsupported relations, dangling nodes) and produces an updated alias glossary, decision log, and combined node and link tables.

2.6 Step 3: Analyzing the combined map: vicious cycles, focal points, and virtuous cycles

Step 3 works only on the combined, evidence-anchored map and does not alter its structure or vocabulary. Its purpose is to identify the main feedback structures represented by stakeholders: (1) vicious cycles (self-reinforcing problems) and, where closure is not evidenced, directional chains; (2) focal points where multiple vicious dynamics converge; and (3) virtuous cycles that stakeholders described as improvements acting on those same concentrations. Loops are prioritized and tagged to support later scenario testing, but chains remain permissible and are explicitly labelled. This sequencing maintains a strict boundary between stakeholder evidence and analyst inference (White & Barbrook-Johnson, 2025; Reinholtz et al., 2025).

Operationally, Step 3 restarts cycle identification "from scratch" on the combined map: any loops previously noted in separate workshop analyses are discarded, and each accepted loop or chain is assigned a Cycle-ID and recorded as an ordered set of Node-IDs and Link-IDs with supporting quotations. AI support is used to enumerate closed feedback loops using only existing nodes and links; when transcript evidence supports a directional sequence but not closure, we log a chain rather than forcing circularity. Candidate focal points are flagged by recurrence across vicious loops/chains and then checked against transcript excerpts, applying three criteria: stakeholder salience, systemic reach, and reversibility potential; and confirming focal points that satisfy at least two. Virtuous cycles are identified by tracing stakeholder-anchored improvement mechanisms in the same node-link structure, rather than by inverting problem loops, and are prioritized where they intersect established focal points. Verification proceeds on three levels. Evidential verification checks that each link is supported by at least one quotation and timestamp, and that directions (and any asserted sign) match the quoted context; structural verification checks closure and directional consistency; and conceptual verification requires that each

accepted loop includes at least one Response or otherwise actionable Pressure/State element so that dynamics remain connected to points of perceived leverage. A cycle-quality table was maintained as a verification record.

2.7 Step 4: Testing the combined map through scenario analysis

Step 4 stress-tests the combined, evidence-anchored map by running scenarios that begin from realistic stakeholder-style questions. Each scenario traces how effects propagate through already accepted nodes and links, as well as through the vicious and virtuous feedback structures identified in Step 3. It then reports expected short-, medium-, and long-horizon consequences along with potential risks or backfires and monitoring indicators that could corroborate or challenge the pathway interpretation. Scenario input is limited to the validated combined map plus the Step 3 cycle catalogue.

Scenarios are documented with a consistent specification: entry node identifiers, the cycles or chains expected to be engaged, a mechanism path recorded as identifiers only, effects by time horizon, plausible risks or backfires, and indicators tied to specific nodes or links. Standard notes on scope, human review, versioning, and researcher responsibility clarify that scenarios are decision-support artefacts rather than forecasts. The full specification template and scenario catalogue are provided in Supplementary Material B. To preserve fidelity to the validated structure, scenario narration mirrors established node labels, uses explicit increase/decrease language at each step, and restates any qualitative delays. In this case study, scenario prompts were formulated as qualitative, stakeholder-style questions tied to core themes in the map, such as water quality, infrastructure, knowledge, and governance.

2.8 Human verification, ethics, and reproducibility

Building on the step-to-step handoffs above, the case study demonstrates how provenance, human oversight, and versioning can be operationalized in AI-supported conceptual mapping. First, nodes, links, cycles, and scenario pathways all carry quotations, timestamps, and workshop tags; merge and alias decisions are recorded with rationales. This ensures that every causal claim remains open to inspection and revision.

Second, human oversight is pervasive and role-specific. AI is used where tasks are repetitive but well-constrained: extracting candidate nodes and links, surfacing merge candidates, enumerating feedback loops consistent with the combined map, and drafting scenario narratives from defined entry themes and pathways. Researchers retain authority over all decisions: admitting or rejecting nodes and links, deciding when to merge or split, designating cycles as “vicious” or “virtuous”, and editing or discarding AI-drafted text that does not accurately reflect stakeholder intent. Stakeholders contributed workshop themes and causal relationships, and local collaborators who participated in the workshops co-drafted scenario questions (Sections 2.1 and 3.4), while post-workshop analytical decisions remained researcher-led. This arrangement should be distinguished from stakeholder co-design or approval of the AI guardrails. Verification here means checking correspondence with recorded stakeholder material and structural consistency; it does not establish stakeholder consensus or independently validate causal relationships. This division of labor aligns with recommendations that AI should amplify but not replace human interpretive judgement in causal mapping (Sharma et al., 2025; White & Barbrook-Johnson, 2025; Hosseinichimeh et al., 2024).

Third, reproducibility is supported by step-reports and ID-stable CSV files for nodes, links, cycles, alias glossaries, and scenarios. Because each step produces a compact, archivable package that can be re-used as input, independent teams could, in principle, re-run a single step under alternative assumptions without altering the underlying workshop material. This structure also makes it feasible to publish not only the final maps but also the intermediate artefacts that document how they were constructed, subject to consent, confidentiality provisions, and data-governance limits. Comparable recommendations appear in Reinholtz et al. (2025) regarding version-controlled, checkpointed causal mapping pipelines. Translation handling forms part of this reproducibility logic as well: where translated quotations are reported in the manuscript, the original-language quotations remain part of the auditable record described in Section 2.2. Ethical and privacy safeguards apply throughout: participants provide consent for AI-assisted analysis; transcripts are handled under agreed confidentiality provisions; and where translation is required, original quotations are retained alongside their renderings.

3. Results

This section reports the outputs of applying the four-step workflow to the Costa Brava case study. To preserve the distinction between methods and results, the subsection headings mirror the four workflow steps, but the text below reports only case-study outputs.

3.1 Step 1: Cleaning transcripts and extracting nodes and links

Step 1 produced evidence-anchored node and link tables for each workshop, retaining original wording with timestamped quotations for traceability. For Workshop 1, the eight whiteboard conversation starters served as seed nodes and the guided and open discussions introduced additional themes. One duplicated label ("Reduce Human Consumption") was judged equivalent to "Human Consumption" and merged after review. Before cross-workshop merging, preliminary application of the Step 3 feedback-analysis approach to the Workshop 1 map identified complex links that appeared to assume intermediate concepts. Targeted transcript re-reading supported three such bridging nodes. Researchers made these concepts explicit to represent intermediate steps implicit in the dialogue but not separately named by stakeholders. Table 1 summarizes seed nodes, discussion-generated themes, verified merges, and transcript-confirmed additions. The final Workshop 1 dataset comprises 30 themes and 86 directional links.

Table 1: Decision log for Workshop 1 summarizing seed nodes, discussion-generated themes, verified merges, and transcript-confirmed additions during Step 1 extraction.

Decision type	Description	Node IDs	Justification
Seed nodes	Eight conversation starters placed on whiteboard as conversation anchors	1-8	Provided common starting points for discussion
Add themes from discussion	Sixteen new themes introduced during stakeholder dialogue	9-23	Emergent content generated by participants
Merge	Theme "Reduce Human Consumption" merged with "Human Consumption"	2	Duplicate meaning identified during review
Add themes from transcript analysis	Four new themes added through AI assisted extraction	24-27	Explicitly supported by transcript evidence
Add transcript-supported bridging themes after targeted re-reading	Three bridging themes added before merging, following preliminary feedback analysis of the Workshop 1 map and transcript re-reading	28-30	Researcher-interpreted intermediate concepts implicitly supported by the dialogue

Workshop 2 followed the same procedure. Transcript analysis expanded the initial map by adding themes discussed but not written on the whiteboard, and two consolidation rounds removed redundancies and harmonized terminology. The full merge rationale is documented in Supplementary Material C. After these refinements, the Workshop 2 map contained 34 distinct themes. Link extraction and consolidation followed the same logic in both workshops: additional relations discussed but not drawn were included only when supported by transcript evidence, and redundant links created by node merges were removed or combined, yielding 156 validated relations in Workshop 2.

Taken together, the two workshops show that Step 1 yields context-rich, evidence-anchored conceptual maps with an explicit audit trail. Alias decisions and merges are recorded (Table 1 for Workshop 1 and Supplementary Material C for Workshop 2), and node- and link-level provenance makes each element traceable to stakeholder quotations. The resulting maps differ in scale but were produced through the same extraction logic. In both workshops, the evidence-anchored node and link tables produced in Step 1 provide only the starting empirical material for Step 2 merging. The combined, ID-stable map produced in Step 2 then serves as the direct structural basis for Step 3 feedback-structure analysis and, together with the Step 3 cycle catalogue, for Step 4 scenario testing.

3.2 Step 2: Combining individual maps into a single, ID-stable map

The two Step 1 maps contained 64 original nodes (30 from Workshop 1 and 34 from Workshop 2). Applying the merge criteria consolidated these into 43 combined themes, merging close duplicates and strong overlaps while retaining related but distinct items where scope or stakeholder specificity differed. Each combined node preserves its original workshop indices for traceability, and links are re-expressed using the combined node indices. Duplicates are collapsed and redundancies created by node merges are removed, yielding a stable combined link table. A small set of Tier 1 and Tier 2 cross-workshop relations were added under the blind-link screening protocol; each accepted relation carries matched quotations and timestamps, while weaker candidates were retained only in the decision log.

Step 2 therefore establishes the ID-stable combined map. Nodes receive a new combined index (1–43) while preserving original indices, and links receive new Link-IDs (1–243) with evidence-anchored justifications. In substantive terms, the combined map integrates shared framing (e.g., water quality, governance, economic conditions) while keeping distinctive workshop contributions intact (e.g., ecological detail emphasized in Workshop 1 and socioeconomic/cultural dimensions emphasized in Workshop 2), reducing redundancy without erasing stakeholder specificity. Representative examples of node merges, non-merges, and link decisions are summarized in Table 2, and the overall node-set evolution is shown in Figure 2.

3.3 Step 3: Analyzing the combined map: vicious cycles, focal points, and virtuous cycles

Applying Step 3 to the combined map yielded a restricted set of coherent vicious feedback structures that capture widely discussed stakeholder concerns rather than an exhaustive enumeration of all mathematically possible loops. One recurrent degradation pathway (Figure 3), for example, links Pollution (Chemical & Litter) to degraded Water Quality, intensified Eutrophication, reduced Primary Production, and Reduced Catch/Stock Decline, with declining stocks undermining Fisher Livelihood Security and prompting diversification pressures that can feed back into higher Pollution through tourism-related activities and infrastructure-enabled marine-use pressures. Within this example degradation pathway, three focal points recur: Pollution (Chemical & Litter), Water Quality, and Fisher Livelihood Security.

Virtuous cycles were then traced as stakeholder-described improvement pathways acting on these focal points while remaining strictly within the existing node-link backbone. In the Figure 3 example, one integrative virtuous loop begins with Local Fishing Cooperatives and Local & Traditional Expertise, strengthens Adaptive & Inclusive Governance and improves the Regulation & Policy Framework. These governance improvements connect to more effective Pollution Management & Reduction, feeding back by mitigating Pollution and indirectly supporting improved Water Quality and more stable livelihoods. Figure 3 provides a simplified representation of one such vicious and virtuous feedback structure, while the full overlays are provided in Supplementary Material D (Figures D1 and D2). All identified loops, chains, focal points, and improvement mechanisms remain traceable to the combined map and to transcript quotations, providing an auditable bridge from stakeholder discourse to scenario testing in Step 4.

3.4 Step 4: Testing the combined map through scenario analysis

In this implementation, Step 4 was constrained to the combined conceptual map and underlying transcripts: no new links were added and no external datasets were introduced. Ten stakeholder-style questions were co-drafted with local collaborators (project partners who participated in the workshops) using workshop notes and recurring transcript themes as starting points. For each question, first-, second-, and third-order pathways were traced through the combined map and interpreted through the vicious and virtuous feedback structures identified in Step 3, recording which existing virtuous mechanisms support desirable trajectories and which vicious mechanisms risk reactivation.

Table 2: Summary of node and link decisions during Step 2 consolidation of the two workshop maps. This table documents the decision logic applied when merging the Workshop 1 (W1) and Workshop 2 (W2) node sets into a single combined map. For each decision type, it lists representative examples, affected node IDs before and after consolidation (merged node IDs shown as C_i), and the justification based on Step 2 criteria (semantic equivalence, functional overlap, stakeholder specificity, or transcript-verified causal consistency). The table also records decisions on link consolidation and the graded inclusion of cross-workshop relations (Tier 1–3), ensuring transparent provenance and auditability of the final combined-map architecture.

Decision type	Description	Affected Node IDs (W1 / W2 → Combined)	Justification (per Step 2 criteria)
Merge: Exact duplicate	Merge identical or near-identical wording across workshops	W1: *Water Quality*, W2: *Water Quality* → C1: *Water Quality*; W1: *Human Consumption*, W2: *Human Consumption* → C2: *Human Consumption Patterns*; W1: *Regulation & Policy*, W2: *Regulation & Policy* → C3: *Regulation & Policy Framework*	Equivalent scope and phrasing; no evidence of distinct conceptual boundaries in transcripts.
Merge: Strong overlap	Consolidate conceptually similar mechanisms with minor wording differences	W1: *Infrastructure*, W2: *Infrastructure (Ports, Energy, Logistics)* → C7: *Infrastructure (Ports, Coastal & Energy);* W1: *Knowledge & Scientific Data*, W2: *Knowledge & Scientific Capacity* → C8: *Knowledge & Scientific Data*; W1: *Local Competitiveness*, W2: *Fair Market Pricing / Economic Lobbies* → C14: *Local Competitiveness, Fair Markets & Economic Lobbies*	Overlapping functions and causal roles; transcripts show interchangeable use or directly coupled mechanisms.
Retain as distinct: Related-but-distinct	Keep more granular academic items separate from umbrella nodes	W1: *Nutrients (N & P)* → C18; W1: *Water Transparency* → C21; W1: *Shellfish Health* → C23	Granular ecological indicators explicitly requested by academics; not referenced independently in Workshop 2.
Retain as distinct: Stakeholder-specific	Keep themes exclusive to one group without cross-workshop counterpart	W2: *Decline in Fishers (Workforce)* → C34; W2: *Local Seafood & Gastronomy* → C33; W2: *Immigrant Participation in Fisheries* → C31	No conceptual analogue in Workshop 1; transcripts show stakeholder-specific concerns.
Link consolidation	Remove or collapse redundant relations after node merges	All links referencing merged nodes (e.g. W1 link 14, W2 link 52) re-indexed to C-nodes	Consistent with Step 2 rule: no duplicated edges once node consolidation alters sources/targets.
Cross-workshop relation: Tier 1 (strong evidence)	Add cross-workshop causal link with explicit statements in both transcripts	C34 (*Decline in Fishers*) → C26 (*Small-scale Fishery*); C16 (*Marine Protected Areas*) → C37 (*Shared Marine Space Management*); C25 (*Primary Production*) → C27 (*Reduced Catch / Stock Decline*)	Both workshops provide explicit causal phrasing describing the same mechanism.
Cross-workshop relation: Tier 2 (moderate evidence)	Add link supported explicitly in one workshop and implicitly in the other	C20 (*Eutrophication*) → C29 *(Species & Ecosystem Changes);* C25 *(Primary Production)* → C27 (*Reduced Catch*)	One workshop provides explicit direction; the other offers consistent contextual implications.
Cross-workshop relation: Rejected (Tier 3)	Proposed link not added to combined map	Several candidates (e.g. *Pathogens → Market Demand*)	Evidence present in only one workshop or phrasing too indirect to meet inclusion criteria; logged for transparency.

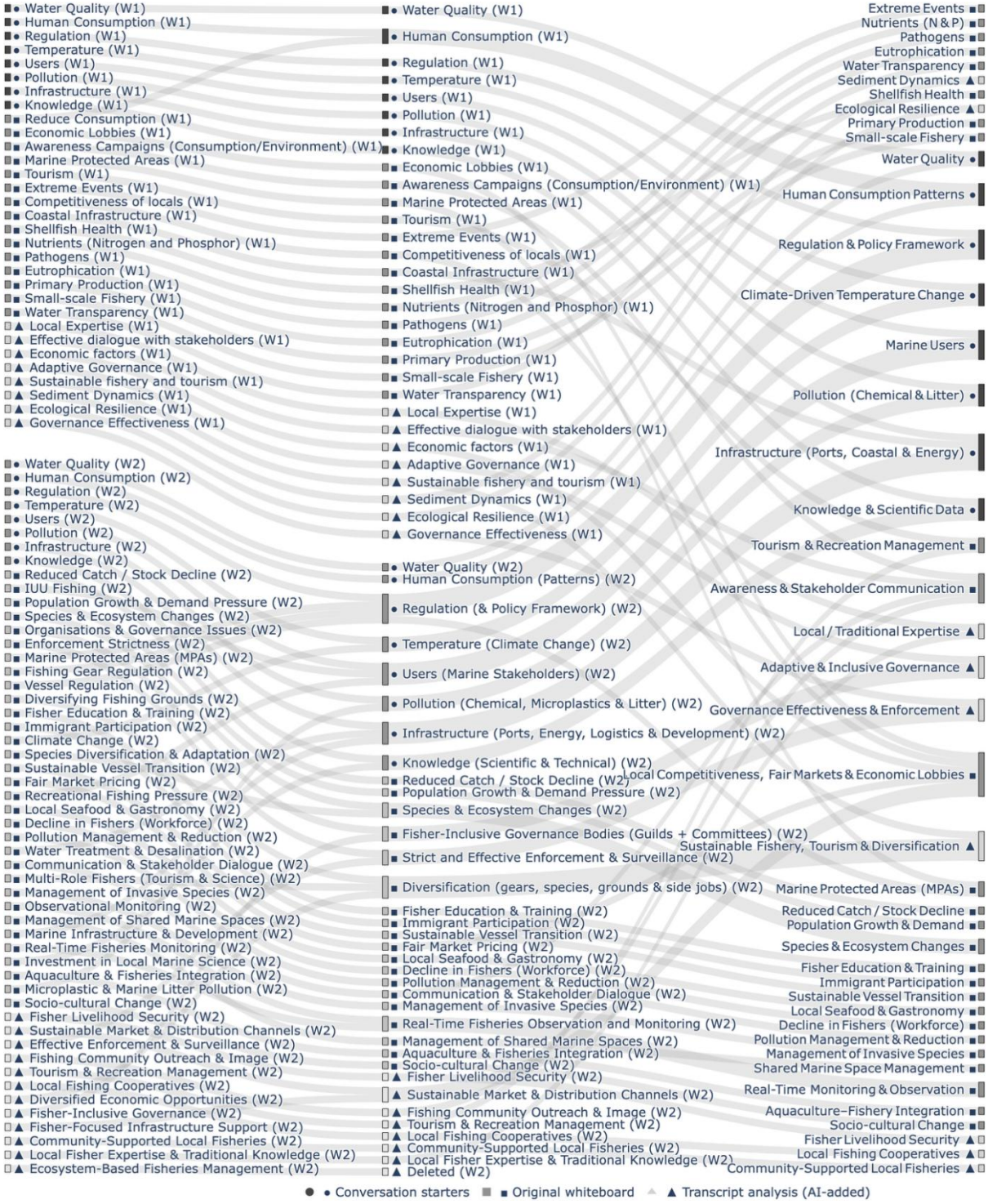


Figure 2: Evolution of workshop node sets and their consolidation into the combined map. The left column shows the original and transcript-derived nodes from Workshop 1 (W1; darker shades, top) and Workshop 2 (W2; lighter shades, bottom); the middle column shows the per-workshop condensed nodes after within-workshop merging; and the right column shows the final combined-map nodes. Node greyscale and symbol indicate whether nodes originate from conversation starters, original whiteboard contributions, or transcript-based additions (see legend), while vertical shading distinguishes the two workshops. Link thickness reflects the number of source nodes contributing to each merge step, so thicker bands indicate stronger many-to-one consolidation.

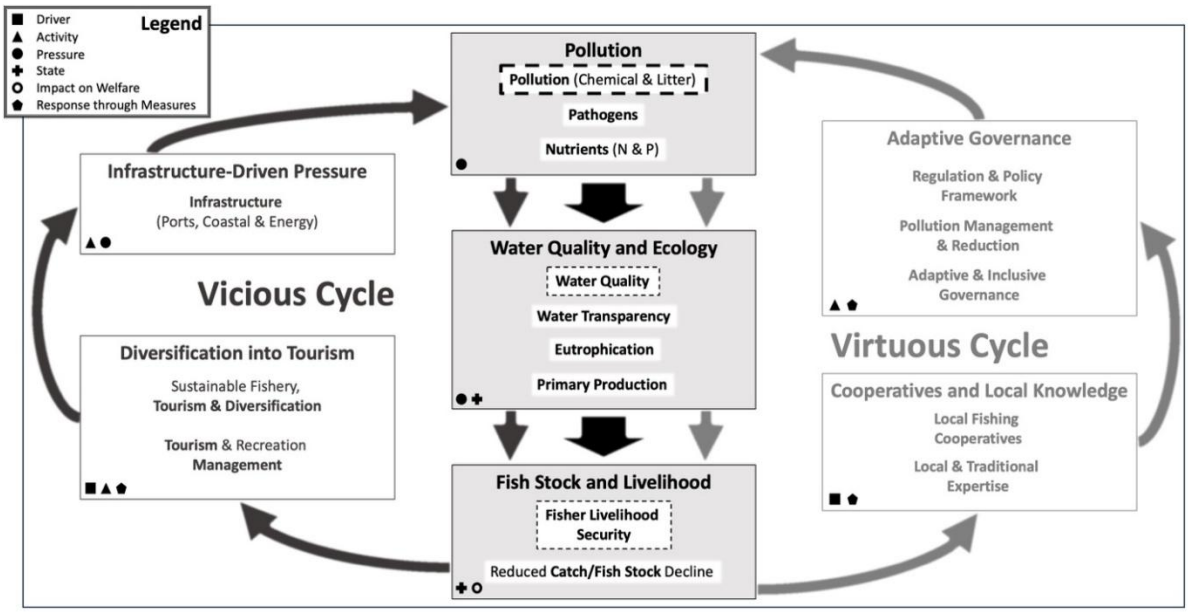


Figure 3: Representative vicious and virtuous feedback structures derived from the combined conceptual map. The diagram shows one representative vicious cycle (left, dark grey) and one corresponding virtuous cycle (right, light grey) identified in Step 3. Both cycles are simplified and grouped into seven thematic boxes, using only nodes present in the combined map, to highlight their structural logic and points of convergence. In the central backbone, Pollution, Water Quality & Ecology, and Fish Stock & Livelihood form the core causal progression shared by both dynamics. The vicious cycle depicts how diversification into tourism, marine-use pressure, and infrastructure-driven pressures reinforce a downward ecological and socioeconomic spiral, ultimately feeding back into increased Pollution. The virtuous cycle shows how stronger local cooperatives, local and traditional expertise, adaptive governance, regulatory improvements, and pollution-management responses feed back into Pollution mitigation and indirectly support ecological recovery and livelihood stability. Symbolic DAPSI(W)R(M) markers indicate the categories to which each grouped element belongs. The main focal point targeted by the virtuous cycle, Pollution (Chemical & Litter), is highlighted by a thick dashed border, while the secondary focal points, Water Quality and Fisher Livelihood Security, are more subtly marked.

Table B1 (Supplementary Material B) summarizes the ten scenarios, their main leverage points, dominant positive and negative trajectories, and key guardrails. Here, scenario guardrails refer to conditions identified in the map-based analysis as potentially supporting desirable trajectories, rather than to the constraints governing AI-assisted analysis. All model-assisted drafts were checked and edited by the authors, and scenario texts were re-read against the combined map and the Step 3 cycle catalogue to confirm that every causal statement corresponds to an existing link or chain and that no new themes or relations were introduced. Substantively, the ten scenarios converge on recurring guardrails around access and space rules, pollution management and water quality, cooperative and market architecture, and knowledge and monitoring. These themes recur across different stakeholder questions and are explicitly represented in the combined map.

Taken together, Step 4 converts the validated map into a bounded set of reviewable scenario narratives rather than a predictive model, while showing that diverse stakeholder questions repeatedly activate a limited set of common leverage areas already visible in the combined map.

4. Discussion

4.1 Methodological implications

4.1.1 Interactions among steps

The Costa Brava case study shows that the four-step workflow functions as a serial, provenance-preserving sequence of audited handoffs rather than as a single continuous model session. Each step starts from the verified artefacts produced by the previous step, rather than from accumulated model exchanges, and applies the corresponding procedure described in Section 2. Step 1 produces evidence-anchored node and link tables from

stakeholder dialogue; Step 2 consolidates these into a combined, ID-stable map; Step 3 derives feedback structures and focal points using only that validated combined structure; and Step 4 traces scenarios through the validated map and the Step 3 cycle catalogue. Taken together, the workflow implements a controlled movement from transcript evidence to consolidated structure, feedback-structure analysis, and finally scenario interpretation.

Across Steps 1 to 4, the workflow therefore enforces a provenance-preserving handoff: extraction is evidence-anchored and non-inventive, consolidation remains ID-stable and auditable, feedback structures are derived without adding edges, and scenarios are constrained to tracing only validated pathways. Scenario narratives therefore function as bounded tests of coherence and implication rather than as free-standing stories. DAPSI(W)R(M) coding supports interpretation by clarifying how pressures, states, impacts, and responses appear within the identified feedback structures (Elliott et al., 2017).

Across the steps, “clean-slate re-entry” artefacts proved important. Each new round of AI assistance was seeded only with the versioned step reports and CSV files, not with previous model exchanges. Corrections and clarifications, for example in cycle descriptions or scenario drafts, could thus be implemented by rerunning only the affected step or steps rather than the entire workflow. Cumulative drift was contained by always returning to the last audited artefacts.

4.1.2 Limitations and points of friction

The Costa Brava case study also reveals limitations that are directly tied to methodological choices. These concern transcript dependence, reliability checking, polarity and delay interpretation, the scenario output quality, and the roles of participant composition and analyst judgement.

Most importantly, the workflow is transcript-dependent. All structure originates in workshop dialogue; where the conversations were thin, for example on certain ecological or economic mechanisms, the maps remained sparse. The workflow deliberately did not allow the AI to fill gaps with plausible but unsupported links. This preserves stakeholder fidelity but leaves blind spots that only additional engagement or complementary data can address. More generally, the resulting structures reflect not only what was said, but also who was in the room and how analysts judged equivalence, leverage, and narrative coherence. Stakeholder composition and modeler judgement shape the maps alongside the guardrails intended to reduce unsupported inference.

We did not calculate formal agreement statistics. However, hybrid human-AI approaches illustrate how sampling and comparing human and model coding can reveal drift and guide prompt refinement. Liu and Sun (2025) report that iterative prompting and validation improved alignment between model outputs and expert interpretation in stakeholder-interview analysis, while Liu et al. (2025) synthesize prompt-design practices that support more consistent, auditable outputs. Such procedures could inform future reliability assessments within this workflow.

The current implementation relies on manual interpretation of link polarity and delays, where delay denotes qualitative time-lags implied in stakeholder statements (e.g., immediate vs. seasonal vs. multi-year effects). While the methods anticipate automated extraction of signs and temporal cues, in practice these features were inferred and checked by hand when identifying vicious and virtuous cycles and drafting scenarios. This limits scalability and leaves timing assumptions less systematic than they could be, even though it maintains the no-invention constraint. Two refinements are feasible: first, introducing a “four-check” procedure to test sign consistency (Hosseinichimeh et al., 2024); and second, improving recall of longer loops by combining an algorithmic scan (e.g., depth-first search of the adjacency list) with an LLM-based explanation pass constrained to existing edges (Hosseinichimeh et al., 2024).

While scenario work sketched connections to observable indicators and hinted at falsifiable outcomes, no quantitative validation was performed. Scenario narratives therefore remain qualitative readings of the map rather than tested predictions. This is consistent with the scope of the paper but marks a clear boundary of what the present workflow can claim.

Finally, the approach assumes a relatively high level of methodological literacy among facilitators. Steps such as interpreting DAPSI(W)R(M) roles, deciding which feedback configurations count as substantively meaningful,

and editing AI drafts to retain stakeholder voice all require judgement and experience. Without this, the risk of overinterpreting model suggestions and reproducing modeler bias could increase.

4.1.3 Human-AI collaboration and stakeholder fidelity

This study illustrates three methodological safeguards intended to support fidelity to documented stakeholder contributions. Their use does not by itself establish complete representation or eliminate interpretive bias, and reliance on recorded statements may under-represent tacit or poorly articulated knowledge. First, evidence anchoring ensures that every AI-proposed element is backed by quotations; unmatched suggestions are simply dropped. Second, alias tracking and stable identifiers make it easier to track how stakeholder phrases evolve as themes are merged and maps combined. At the same time, they ensure that original labels remain recoverable throughout. Third, restriction to documented structure in Steps 3 and 4 prevents model-driven embellishment: cycles and scenarios must be explicable entirely in terms of already accepted nodes and links.

Under these conditions, the final vicious and virtuous feedback structures and scenario narratives still read as elaborations of stakeholder concerns about how future sustainable fisheries depend on water quality, governance, tourism, markets, and livelihoods, rather than as abstract model artefacts. The Costa Brava case study therefore illustrates how researcher-supervised AI assistance can organize stakeholder-derived knowledge while making analytical decisions traceable. The clean-slate re-entry design further prevents the model from building on its own earlier suggestions: each round of assistance operates only on the audited CSVs and step reports, which makes unsupported or drifted content easy to detect and discard.

Further participant review could focus on selected labels, merges, or pathway interpretations, with researchers translating the resulting feedback into documented analytical and prompt revisions where needed. Any new evidence and resulting changes would need to be recorded before repeating the affected analytical steps. The scope and feasibility of such review should be agreed with participants rather than assuming that they can scrutinize the complete synthesis (Section 4.2).

Taken together, these limitations are tractable and largely reflect deliberate guardrails, especially evidence anchoring, no-invention, and clean-slate re-entry. The refinements listed below translate them into concrete, testable development steps.

4.2 Practical implications, transferability, and planned refinements

The Costa Brava case illustrates how specific workflow mechanisms support the representation and traceability of stakeholder contributions. Evidence-anchored extraction captured themes and relations discussed but not recorded on the whiteboards (Section 3.1). Conservative merging preserved distinct contributions: Nutrients and Water Transparency remained separate from Water Quality, while stakeholder-specific concerns such as Decline in Fishers (Workforce) were retained without requiring a cross-workshop counterpart (Section 3.2; Table 2). These choices connect the inclusion of recorded knowledge and recognition of different perspectives to concrete mapping decisions. Original wording, workshop provenance, and documented merge decisions make the resulting interpretations traceable, although this does not establish stakeholder endorsement of the synthesis. In Step 4, ten stakeholder-style questions co-drafted with local collaborators were explored through existing map pathways (Section 3.4; Supplementary Material B), connecting scenario analysis to workshop concerns and producing narratives for further scrutiny rather than forecasts.

The transferable element is this documented sequence of evidence-anchored extraction, controlled synthesis, and bounded scenario reasoning, not the Costa Brava map or its substantive conclusions. Separating the steps and retaining versioned artefacts could allow organizations to adopt individual stages while preserving their evidential basis. Application elsewhere would require usable stakeholder records, appropriate permissions, and expertise to review coding and interpretation. The suitability of DAPSI(W)R(M), the merging criteria, and the opportunities and resources for stakeholder review would need to be assessed in each setting. Transfer to domains such as urban planning, climate change adaptation, or health systems therefore remains a proposition requiring adaptation and empirical testing.

In practice, resource demands remain uneven across steps. In this implementation, AI assisted extraction, while merging, polarity checking, and scenario verification still required intensive human judgement. We did not

quantify time use systematically in this implementation, so the manuscript does not claim measured efficiency gains. Our experience was, however, consistent with the emerging literature in suggesting that AI support is most helpful in the earlier, more repetitive stages of transcript scanning and first-pass extraction, whereas the more interpretive tasks of merging, verification, and scenario analysis remain comparably labor-intensive (Sharma et al., 2025; White and Barbrook-Johnson, 2025; Hosseinichimeh et al., 2024; Reinholtz et al., 2025). Any efficiency gains would need to be assessed alongside the time and expertise required for human review.

At the same time, the case highlights tangible avenues for refinement. Automatic extraction of polarity and delay information from textual cues, followed by structured human checking, could make cycle identification and scenario timing more systematic and scalable. More explicit integration of indicators and falsifiable outcomes would help connect conceptual maps to monitoring systems (Rowe et al., 2025), enabling scenario claims to become empirically testable where suitable administrative or ecological data exist. Future evaluation could assess the usefulness of the outputs for stakeholder discussion as well as their analytical coherence. A structured expert or stakeholder questionnaire could examine cycles and scenarios in terms of coherence, realism, actionability, and perceived representation of stakeholder perspectives, using the existing step-reports as common reference points. Participants could also identify omissions, contest interpretations, and propose revisions, connecting evaluation to their ability to influence how their knowledge is represented. This would align with recommendations by Powell et al. (2024) and Hosseinichimeh et al. (2024). Finally, stronger checkpointing mechanisms, such as retaining verified outputs and recording hashes of node and link tables and decision logs, could support reuse of accepted artefacts and detection of changes between versions, strengthening traceability and reproducibility in line with Reinholtz et al. (2025).

Overall, the Costa Brava case study demonstrates how stakeholder-derived knowledge can be carried through a traceable sequence from workshop records to a combined map and qualitative scenarios. These outputs provide structured material for subsequent discussion, revision, and application. The evaluation reported here concerns correspondence with source material and internal consistency; effects on stakeholder learning, engagement, or decisions, and usefulness in other settings, are not established by this implementation.

5. Conclusion

This paper presents an audit-ready workflow for guardrailed LLM-assisted participatory conceptual mapping that moves systematically from workshop transcripts to scenario testing while retaining links to recorded stakeholder contributions. The core contribution is not a single map, but a stepwise, auditable workflow in which transcript-based map construction is distinguished from subsequent analysis of the finalized combined map: stakeholder transcripts are used first to extract nodes and directional links, and later only to verify that subsequent analyses remain consistent with stakeholder intent. Across all steps, the workflow operationalizes explicit guardrails: no-invention, quotation-level evidence anchoring, alias tracking, stable identifiers, step-wise reporting, clean-slate re-entry from audited CSV artefacts, and human verification, so that every causal claim remains traceable, auditable, and revisable.

Methodologically, the workflow yields a linked set of artefacts that support cumulative analysis while limiting structural drift. Step 1 generates evidence-anchored node and link tables from transcript excerpts. Step 2 consolidates these into a controlled, ID-stable combined map that preserves provenance across workshops. Step 3 derives a documented catalogue of vicious cycles, focal points, and virtuous cycles strictly from the verified combined structure. Step 4 then frames qualitative scenarios as stakeholder-style “what-if” questions and traces them through validated pathways. By treating intermediate artefacts as the system of record, rather than narrative prose or chat history, the workflow makes it possible to rerun or revise individual steps without silently propagating earlier errors, and to inspect how each consolidation or interpretive step was justified.

The Costa Brava case study illustrates feasibility and the kinds of outputs the workflow can support in practice. In Step 1, two participatory workshops generated 64 initial themes and 242 directional links, which were consolidated in Step 2 into 43 combined themes and 243 verified relations through controlled merging and evidence-anchored screening. Using only this combined, DAPSI(W)R(M)-consistent structure, Step 3 identified three major vicious-cycle systems, several focal points where reinforcing dynamics converge, and twelve stakeholder-anchored virtuous loops and chains. Step 4 then traced ten qualitative scenarios through the same

validated map and cycle catalogue, producing structured narratives that should be read as reviewable hypotheses rather than empirically tested forecasts.

Several boundaries of the current implementation also clarify priorities for refinement. Because the workflow is transcript-dependent and enforces a strict no-invention rule, it does not fill gaps where stakeholder dialogue is sparse; those blind spots can only be addressed through additional engagement or complementary evidence. In addition, polarity and timing cues were handled conservatively and checked manually when needed for cycle labelling and scenario tracing, and no formal agreement statistics were computed for human-AI coding alignment. Future development can therefore focus on more systematic, guardrailed extraction and checking of polarity and delay cues; tighter integration of indicators and scenario "test hooks" to support empirically falsifiable outcomes where appropriate data exist; structured evaluation instruments for expert or stakeholder review of cycles and scenarios; and stronger reproducibility mechanisms such as idempotent checkpointing over versioned tables and decision logs.

Overall, the workflow offers a transparent foundation for responsible LLM use in participatory systems analysis: models are treated as constrained assistants operating on evidence-anchored artefacts, while the analysis remains grounded in stakeholder contributions and researchers retain responsibility for post-workshop interpretation and verification. By combining quotation-level provenance, stable identifiers, and clean-slate re-entry between steps, the approach supports traceable movement from unstructured dialogue to consolidated maps, feedback-structure analysis, and scenario reasoning in a form that can be audited, revised, and adapted to other participatory domains.

Acknowledgements

The SEADITO project has received funding from the European Union's Horizon Europe research and innovation program under grant agreement No 101157243, and has the institutional support of the Severo Ochoa Centre of Excellence accreditation (CEX2024-001494-S funded by AEI 10.13039/501100011033) to the ICM-CSIC. We warmly thank all participating stakeholders, fishers, local cooperatives, scientists, and policy representatives, for generously sharing their time, insights, and experience during the workshops and discussions. Their voluntary engagement, given entirely in their free time, made this collaborative mapping and scenario work possible and deeply grounded in real-world knowledge. We also thank our case study partners Mireia Esteve Xifra and Judit Pagès Montroig of Associació Grup D'Acció Local Pesquer Costa Brava (GALP Costa Brava) for their collaboration and support throughout the case study.

Author contribution

Maik Neukirch: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review and Editing, Visualization.
Xavier Garcia: Conceptualization, Validation, Investigation, Resources, Data Curation, Writing - Review and Editing, Supervision, Project administration, Funding acquisition.
Elisa Berdalet: Conceptualization, Validation, Investigation, Resources, Data Curation, Writing - Review and Editing, Supervision, Project administration, Funding acquisition.
Aurelija Armoskaite: Conceptualization, Methodology, Writing - Review and Editing, Visualization.
Giulia Bellon: Conceptualization, Methodology, Writing - Review and Editing, Visualization.
Miriam von Thenen: Conceptualization, Methodology, Writing - Review and Editing, Visualization, Project administration, Funding acquisition.
Erasmia Kastanidi: Conceptualization, Methodology, Software, Validation, Formal Analysis, Writing - Review and Editing, Visualization, Project administration, Funding acquisition.

Data availability statement

The primary contribution of this paper is the four-step methodological workflow, documented in Section 2 and Supplementary Material A, including its prompt templates, output formats, guardrails, and human-verification procedures. These materials are provided with the article to support application of the workflow to other

suitable stakeholder datasets. The raw workshop data underlying this study, including audio recordings, transcripts, whiteboard captures, and workshop notes, are not publicly available because they contain identifiable and confidential participant information and were collected under conditions that do not permit open public dissemination. The processed case-study materials are not currently publicly available, pending establishment of the project-level data-sharing arrangements on Zenodo. Their public release is planned as part of a unified collection of SEADITO project outputs, to keep related data and documentation together rather than distribute them through isolated deposits. Release will be limited to materials that can be shared consistently with participant consent and confidentiality provisions.

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors used OpenAI's ChatGPT to assist with drafting support, language revision, and the organization of manuscript sections during the revision process. All text produced with AI assistance was reviewed, revised, and, where necessary, rewritten by the authors. All substantive interpretations, methodological decisions, verification of claims, and final editorial decisions remained the responsibility of the authors. The use of ChatGPT as part of the research workflow itself is described separately in the Methods section. The authors take full responsibility for the content of the published article.

Supplementary Material

The Supplementary Material for this article can be found online at https://sesmo.org/article/view/19051/18510.

References

Elliott, M., Burdon, D., Atkins, J. P., Borja, Á., Cormier, R., de Jonge, V. N., & Turner, R. K. (2017). "And DPSIR begat DAPSI(W)R(M)!" – A unifying framework for marine environmental management. Marine Pollution Bulletin, 118(1-2), 27-40. https://doi.org/10.1016/j.marpolbul.2017.03.049

Hayward, J., Morton, S., Johnstone, M., Creighton, D., & Allender, S. (2020). Tools and analytic techniques to synthesise community knowledge in CBPR using computer-mediated participatory system modelling. npj Digital Medicine, 3, Article 22. https://doi.org/10.1038/s41746-020-0230-x

Hosseinichimeh, N., Majumdar, A., Williams, R., & Ghaffarzadegan, N. (2024). From text to map: A system dynamics bot for constructing causal loop diagrams. System Dynamics Review, 40(3), e1782. https://doi.org/10.1002/sdr.1782

Liu, N.-Y. G., & Keith, D. R. (2024). Leveraging Large Language Models for Automated Causal Loop Diagram Generation: Enhancing System Dynamics Modeling through Curated Prompting Techniques. In Proceedings of the 42nd International System Dynamics Conference. Extended abstract / Short paper. PDF available at: https://proceedings.systemdynamics.org/2024/papers/P1229.pdf

Liu, N.-Y. G., & Keith, D. R. (2025). Leveraging Large Language Models for Automated Causal Loop Diagram Generation: Enhancing System Dynamics Modeling through Curated Prompting Techniques. arXiv preprint. https://doi.org/10.48550/arXiv.2503.21798

Liu, A., & Sun, M. (2025). From Voices to Validity: Leveraging Large Language Models (LLMs) for Textual Analysis of Policy Stakeholder Interviews. AERA Open. https://doi.org/10.1177/23328584251374595

Liu, K., Yigitcanlar, T., Browne, W., & Fu, Y. (2025). Prompts for Planning AI Integration: Effective Prompt Design for Large Language Models in Support of Sustainable Urban Development [Preprint]. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5323915

Manda, P. (2025). Large language models in bio-ontology research: A review. Bioengineering, 12(11), 1260. https://doi.org/10.3390/bioengineering12111260

Newberry, P., & Carhart, N. (2024). Constructing causal loop diagrams from large interview data sets. System Dynamics Review, 40(1), e1745. https://doi.org/10.1002/sdr.1745

Phatak, A., Agrawal, A., Mago, V. K., & Giabbanelli, P. J. (2024). Narrating Causal Graphs with Large Language Models. In Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS-57) (pp. 7530-7539). PDF available at: https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/ac123860-28e7-4b38-9ee4-88961e979b69/content

Powell, S., Copestake, J., & Remnant, F. (2024). Causal mapping for evaluators. Evaluation, 30(1), 100-119. https://doi.org/10.1177/13563890231196601

Quan, S. J., & Lee, S. (2025). Enhancing participatory planning with ChatGPT-assisted planning support systems: A hypothetical case study in Seoul. International Journal of Urban Sciences, 29(1), 89-122. https://doi.org/10.1080/12265934.2025.2462823

Reinholtz, K., Shahroudi, K. E., & Lawrence, S. (2025). LLM powered, expert refined causal loop diagramming via pipeline algebra. Systems, 13(9), 784. https://doi.org/10.3390/systems13090784

Rowe, E. C., Norton, L. R., McVittie, A., Watson, C. A., Wagner, M., Waterton, C., & Smith, L. G. (2025). Using fuzzy cognitive mapping to assess the sustainability impacts of transitioning to pasture-fed production in the UK beef sector. Renewable Agriculture and Food Systems, 40, e1. https://doi.org/10.1017/S1742170524000206

Sedlacko, M., Martinuzzi, A., Røpke, I., Videira, N., & Antunes, P. (2014). Participatory systems mapping for sustainable consumption: Discussion of a method promoting systemic insights. Ecological Economics, 106, 33-43. https://doi.org/10.1016/j.ecolecon.2014.07.002

Sharma, A., & Wallace, J. R. (2025). DeTAILS: Deep Thematic Analysis with Iterative LLM Support. In Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25). ACM. https://doi.org/10.1145/3719160.3735657

Sharma, A., Cochrane, K., & Wallace, J. R. (2025). DeTAILS: Deep Thematic Analysis with Iterative LLM Support [Preprint]. arXiv preprint. https://arxiv.org/abs/2510.17575v2

White, J., & Barbrook-Johnson, P. (2025). Guidance on using large language models to extract cause and effect pairs from texts for systems mapping. CECAN & INET Oxford. CECAN report available at www.cecan.ac.uk/resources