SLE as an Evolving Body of Knowledge: A 19-Year Comparative Analysis of Calls and Proceedings

Vadim Zaytsev
vadim@grammarware.net
Formal Methods and Tools, University of Twente
Enschede, Netherlands

Abstract

The Software Language Engineering (SLE) community has accumulated nearly two decades of content. This content comprises not only papers as such, but also other tangible artefacts like conference descriptions, calls for papers, tools, repositories. Yet, the field's self-description (how it positions itself to potential contributors) and its accumulated output (what the community actually produces) are rarely studied together in a reproducible, fine-grained way.

This paper offers a non-judgemental inventory of SLE's topical evolution by comparatively analysing 19 Calls for Papers (SLE 2008–2026), 6 Calls for Papers and Editorials of corresponding special issues, and 433 proceedings papers and keynotes (SLE 2008–2025), using a shared topic tagset aligned with SLEBoK-style knowledge areas.

We quantify which topics are continuously emphasised, which are episodically foregrounded, and where the calls and proceedings diverge and converge among themselves and over years. We quantitatively identify topics that can be seen as core, near-core and fringe core for SLE, and compare them to neighbouring conferences. We also propose some future directions for SLEBoK.

The dataset produced by this project is released for public inspection at https://slebok.github.io/cfpbok.

ACM Reference Format:
Vadim Zaytsev. 2026. SLE as an Evolving Body of Knowledge: A 19-Year Comparative Analysis of Calls and Proceedings. In *19th ACM SIGPLAN International Conference on Software Language Engineering (SLE '26), July 02–03, 2026, Rennes, France.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3806383.3815513

1 Introduction

Software Language Engineering (SLE) as a field studies the *engineering principles* [Combemale et al. 2017; Klint et al. 2005; Lämmel 2018; Wasowski and Berger 2023] of software languages: their design, implementation, and evolution. Since the first SLE conference in 2008, the community of SLE researchers has produced a sizeable body of research artefacts and methodological knowledge. As a form of guidance, the programme committees and the steering committee, representing the community, have repeatedly been (re)articulating a view of what counts as SLE through its Calls for Papers (CfPs). In parallel, community-wide curation efforts such as the SLE Body of Knowledge [SLEBoK 2013] were launched as a result of both reaching some critical mass and holding a series of focused events where these efforts were discussed and future directions contemplated.

This paper asks a simple but under-explored question: *what body of knowledge has SLE actually built so far, and how does it relate to what SLE has repeatedly asked for?* We approach this question as a comparative, reproducible meta-analysis of conference *requested inputs* (CfPs) and *tangible outputs* (proceedings). Our goal is **not** to rank or judge topics themselves, and not even to argue for the degree to which any specific topic is "interesting" and relevant to the field of SLE, or what should be the "right" direction for SLE to pursue further. The goal is to build an evidence-based inventory that can support: (i) possible SLEBoK curation (e.g., what to prioritise as "core" vs. "emerging" areas), (ii) future CfP design (e.g., how to phrase and structure topics of interest), and (iii) future meta-research (e.g., venue-to-venue comparisons, author migration, topic gaps).

Another important point is that we focus on SLE as a venue, not on "SLE-related work" in a broad sense. The analysis is centred on publicly available CfP pages[1] and proceedings listings for SLE 2008–2025, augmented with occasional special issues. Our analysis is structured to be extensible to adjacent (GPCE, PLDI, MoDELS, OOPSLA, ICSE) or predecessor (LDTA, ASF+SDF, ATEM, WAGA, GTTSE) venues, but we treat such extensions as future work.

Similarly importantly to note is that in this work we focus on *impersonal* view on SLE and SLEBoK, viewing both calls for papers as well as papers themselves to be *generated by the community.* An alternative or follow-up investigation could instead track individuals within that community, cross-checking topic shifts with these individuals drifting across venues. This paper is merely intended as an inventory and a measurement instrument for community self-understanding.

[1] https://www.sleconf.org

SLE '26, Rennes, France

Intended audience and practical use. The immediate audience of this inventory are possible SLEBoK contributors and maintainers looking for a frame of reference such as the distinction between core topics and emerging concerns. We also look at current and future steering committee and programme committee members and chairs who repeatedly revise the CfPs and may benefit from knowing which topics are already implicit in the community's output and which remain aspirational despite repeated invitations. Lastly, it is also meant for authors, especially newcomers, who can use the inventory as a map of how contributions have historically been positioned within SLE. The practical contribution is not meant as a prescription of what SLE should become, but a reusable evidence base for documentation, CfP design, and community meta-analyses.

2 Research Questions

We frame this study as a descriptive, longitudinal account of how SLE management articulates its scope (through CfPs) and how the SLE community realises that scope (through accepted papers). The analysis proceeds in four steps: (i) defining a topic taxonomy that is stable enough for cross-year comparison, (ii) using it to characterise topical change over time in both CfPs and proceedings, (iii) quantifying how calls and proceedings align, including whether shifts in calls are followed by shifts in what gets published, and (iv) distilling implications for the future evolution and curation of SLE as a body of knowledge.

RQ1 Taxonomy adequacy. What topics (knowledge elements) should an SLE-oriented taxonomy contain to meaningfully cover both CfP statements and published SLE papers, and which topical boundaries and groupings make the taxonomy coherent and usable for annotation?

RQ2 Topical evolution. How do these topics evolve over time in *both* calls and papers: which topics are persistent, which emerge or fade, and how does the overall topical diversity change across years?

RQ3 Call–paper alignment. Where do calls and papers systematically differ in topical prominence (topics requested more than they appear vs. topics published more than they are explicitly requested), and to what extent do changes in CfP emphasis predict or precede corresponding changes in proceedings topics (possibly with a lag)?

RQ4 Implications and future directions. What can we reasonably conclude from this inventory for the future consolidation of SLE as a body of knowledge, including implications for SLEBoK curation, for CfP design, and for follow-up meta-analyses (e.g., robustness of tagging, adjacent venues, and community dynamics)?

3 Background

The SLE conference was founded in 2008. At first, its publishing model was to compose the pre-proceedings before the event, and then have the post-proceedings, titled *Revised Selected Papers*, printed by Springer. The first change was in 2013, since that was the first year where *Proceedings*, also Springer LNCS, appeared prior to the conference. In 2015, together with acquiring the SIGPLAN status, SLE moved to ACM as the publisher, where it stayed at least till 2026. Special issues were done irregularly and at different journals.

In this paper we essentially analyse two corpora: the calls for papers, and the proceedings. The rest of this section provides details on both before we start the tagging process. Quantitative summary of them can be seen in Table 1.

3.1 SLE Calls for Papers Corpus

The corpus consists of:

- **18** texts of calls for papers for SLE 2008–2010 and SLE 2012–2026, collected from the respective conference web pages on `https://sleconf.org`, the "community website" migrated from original but lost domain `http://planet-sl.org`.
- **1** call for papers for SLE 2011 which had to be extracted from the archive of the EAPLS mailing list[2]. It is identical to the SLE 2011 call for papers preserved in other mailing lists and CfP aggregators.
- **1** call for papers for the special issue of SLE 2008, which was published as a separate item [Favre et al. 2007].
- **1** call for papers for the special issue of SLE 2017, which was a targeted email[3].
- **5** editorials for special issues [Burgueño et al. 2025; Combemale et al. 2015; Czarnecki and Hedin 2014; Favre et al. 2008; Pearce and Schultz 2020] as imperfect proxies of the editors' expectations.

3.2 SLE Proceedings Corpus

- We started by **including** all items from DBLP lists directly attributed to the SLE conference at https://dblp.org/db/conf/sle/. The 18 editions yielded 15–26 papers per edition, or **381** papers in total.
- At the previous step, we **excluded** the SLE Doctoral Symposium proceedings from 2012, ignoring **9** entries.
- We ensured that ITSLE — a track in the name only, having its own goals and proceedings volumes — is also **excluded**, which was already the case by virtue of how DBLP archived it.
- We checked that all prefaces and other front matter and back matter entries were **excluded**, and DBLP has already removed them, resulting in **0** adjustments.

[2]https://eapls.org/news/conferences/sle-2011-final-call-for-papers-297

[3]We include this one for the sake of completeness, but unfortunately this call for papers was not written around the content, and yielded no topic taggings.

- We made sure to **include** all keynotes. Some editions already included full papers of keynote speakers, others omitted them or gave them limited space, and on one occasion (SLE 2012) only one of two keynotes ended up with a written contribution. To guarantee fairness across the years, we tracked down all keynotes and included those for all years. This added 7 more entries (2012, 2014, 2015, 2019, 2021, 2 × 2023).
- We **included** 2 poster papers [Hansen et al. 2015; Zaytsev 2015]. Initially, SLE 2013 had accepted two submissions as posters, and the *Preface* in its proceedings mentions a "complementary volume" with poster papers. After a journey through JOT and CEUR, poster papers were published at arXiv, marked as belonging to "poster proceedings of SLE 2013" but mentioned neither on any page of the SLE website nor in DBLP.
- We tried to track down SLE special issues at various journals over the years to **include** their papers as well, and found 35 more:
  - 6 papers for SLE 2008 at TSE volume 35 issue 6;
  - 7 papers for SLE 2012 at SCP volume 96 part 4;
  - 6 papers for SLE 2013–2014 at COMLAN vol. 44A;
  - 2 papers for SLE 2017 at COLA volumes 50 and 57[4];
  - 3 papers for SLE 2018 at COLA volumes 56–59[5];
  - 11 papers for SLE 2024 at JSS volumes 210–213, 230.
- We have also **included** the pre-SLE special issue of IET Software with its 7 papers — technically it is an open call special issue of ATEM 2006, but the editorial note explicitly talks about SLE and refers to it.

Special issue papers may be extended versions of earlier conference papers (such relations were identified manually and explorable online), so we cannot treat the resulting corpus as a set of independent research ideas. However, we also do not collapse conference papers and their journal extensions into a single canonical item, because doing so would require paper-by-paper judgements about how much new material was added and how this affects assigned tags — which ultimately makes the corpus less reproducible. Instead, we keep the corpus explicit: regular conference proceedings and special-issue material are both tagged, but some tables and claims that are sensitive to this distinction will report on conference-only values as well. Special issues should be read as supplementary evidence about the SLE record, not as an independent sample of community output.

[4]The FlowSpec paper from volume 57 is marked as an "original article" in the journal's metadata and appears more than a year later than its "sibling" paper from "the same" special issue, but is clearly an extension of the FlowSpec paper from SLE 2017. Its acknowledgement section also explicitly thanks "the anonymous reviewers of SLE'17". Elsevier metadata also marks it as belonging to the special issue: https://www.sciencedirect.com/special-issue/10K8MQCSG5T — but ignores the CBS paper because that one was published under JLVC before the journals fully merged.

[5]Volume 56 had a section titled "Special Issue on GPCE&SLE18" with one GPCE paper. Volume 59 only included the editorial explaining the context of the papers which appeared in the preceding volumes.

Item	Value
CfPs analysed	20
Editorials analysed	5
Total CfP taggings	1039
Average taggings per CfP/editorial	41.56
Proceedings analysed	25
Proceedings papers tagged	433
Total paper taggings	1456
Average lax taggings per paper	3.37
Topic tagset size	38

Table 1. Dataset summary for our comparative analysis. We treat SLE 2026 as CfP-only (no proceedings yet at time of writing), and for the purposes of this table do not count the special issue of 2017 (which technically had a call for papers but there was no taggable content there), but do count the special issue of 2008 with its official call for papers, and we count other special issues as editorials/prefaces.

Our archives are possibly incomplete. For example, the opening slides for SLE 2017 in Vancouver[6] mentioned a COMLAN special issue for selected papers of SLE 2016 and SLE 2017. However, as far as we know, the actual special issue was covering only SLE 2017 papers in a close call via personal emails. To complicate matters further, one of the submissions [Mosses 2019] for this special issue was handled very fast, being the proper 10 page version of the keynote which only had a one page abstract [Mosses 2017]. There have been at least two other accepted submissions, one of which was lost by the manuscript submission malfunction when COMLAN was merged into JLVC which subsequently was renamed to COLA. So, the actual "special issue" of COMLAN has no official call for papers, no editorial, and two papers: one from February 2019 with a JLVC DOI, and one from April 2020 with a COLA DOI. This may seem like a concatenation of coincidences, and it is, but the story is included here as an example of challenges one faces when collecting information for the body of knowledge. In this particular case, we have double checked publication lists of all authors of SLE 2017 and found no other overlaps with COMLAN/COLA/JLVC within the feasible timeframe. Distribution of all papers in our dataset is presented in Table 2.

The primary data record for each CfP is its full text, usually defining the scope, listing bullet points describing topics of interest, and possibly defining paper categories; for each paper it is either its PDF (when available) or at least the title and abstract. We then add an *annotated layer* that assigns topics from a shared tagset to both CfP phrases and papers.

[6]https://www.slideshare.net/slideshow/introduction-of-the-sle17-conference/81109686

Year	Proc.		Keyn.	Special issue		Co-location
(2007)				7	IET Software	
2008	20	LNCS	🟢🟢	6	IEEE TSE	MoDELS
2009	25	LNCS	🟢🟢			MoDELS, GPCE
2010	26	LNCS	🟢🟡			FOSD, GPCE
2011	22	LNCS	🟢🟡			GTTSE
2012	23	LNCS	🟢🔴	7	SCP	FOSD, GPCE
2013	22	LNCS	🟢	6	COMLAN	SPLASH
2014	20	LNCS	🔴			ASE, GPCE
2015	18	ACM	🔴			SPLASH
2016	24	ACM	🟡			SPLASH
2017	25	ACM	🟡	2	JLVC/COLA	SPLASH
2018	23	ACM	🟡	3	COLA	SPLASH
2019	21	ACM	🔴			SPLASH
2020	21	ACM	—			SPLASH
2021	16	ACM	🟡🔴			SPLASH
2022	24	ACM	🟡			SPLASH
2023	22	ACM	🔴🔴			SPLASH
2024	20	ACM	🟡	11	JSS	SPLASH
2025	18	ACM	🟡🟡			STAF
2026	1	ACM	🔴			STAF

Table 2. Distribution of **391** proceedings entries and **42** special issue papers over the years. Keynote markers distinguish full/extended written contributions (🟢, 3–18 pages), short proceedings abstracts (🟡, less than a page long), and abstracts recovered from conference websites (🔴). All SPLASH co-locations are also GPCE co-locations (as well as more weakly related venues like OOPSLA, DLS, DSLDI, REBLS, FOOL, etc). This means, together with MoDELS, ASE and FOSD years, that the only years when SLE did *not* co-locate with GPCE, were 2008, 2011, 2025 and 2026.

3.3 Annotation Protocol

We use a topic tagset designed to (i) capture recurring SLE themes at a useful level of granularity and (ii) support aggregation into higher-level knowledge areas aligned with SLEBoK-style structuring [SLEBoK 2013]. The tagset contains 38 topics grouped into six clusters (T1–T6) spanning language design/semantics, evolution/lifecycle, language kinds, tooling/transformations, evaluation/quality, application and experience. It will be explained in detail in the next section.

The corpus annotation is based on a structured codebook embedded into the dataset: CfP bullet points and paper entries are annotated with explicit tag markers. A single *primary* tag is identified per paper, with optional secondary tags when additional knowledge contributions are too prominent to ignore. We treat the strict and lax views as complementary: strict is closer to traditional single-label topic statistics; lax is closer to a BoK perspective where artefacts contribute to multiple knowledge elements.

The coding was performed by the sole paper author in a process that combined deductive and inductive coding. *Deductively*, the initial codebook was seeded from recurring SLE vocabulary, SLEBoK-style knowledge areas, and long-standing CfP phrases such as design, evolution, transformations, workbenches, testing, and formal methods. *Inductively*, the codebook was revised while annotating the corpus: candidate topics were split when they conflated observably different contributions (e.g., horizontal vs. vertical transformations), merged when they proved too broad or indistinguishable in practice (e.g., grammars/models as standalone topics), and removed when they could not be applied reliably from CfP text, titles, abstracts, or full papers. After stabilising the codebook, the corpus was revisited to make earlier taggings consistent with the final definitions. The full dataset is visualised and released at https://slebok.github.io/cfpbok to support inspection, replication, and future refinement.

4 Taxonomy: the Topic Tagset

To answer the first research question, we need to create a taxonomy with topics that are: (1) expressive enough to cover the vocabulary of SLE CfPs as well as the actual contributions to proceedings and special issues; (2) stable enough to support longitudinal comparisons without redefining categories for each year, and (3) usable enough that different annotators now and in the future can apply it consistently based on limited textual evidence. To clarify the last point: papers are mostly self-contained, but even then there are exceptions like keynote abstracts; the CfPs are much shorter and consist of bullet lists with intentionally concise items.

We derived the scheme below by iterating between recurring CfP phrasings and topic lists, and the recurring shapes of accepted contributions. In practice, CfPs are rich in *activity language* (design, evolution, deployment, evaluation), while papers often contribute *artefacts and techniques* (meta-languages, transformation languages, workbench components, runtimes), plus empirical or formal validation for those. This motivates and justifies a scheme that mixes "what is engineered" (e.g., see below for T3A Meta-languages, T3C DSLs) with "how it is engineered and assessed" (e.g., T5C Testing, T5E Empirical Evaluation).

CfP vocabulary changes over time, mostly growing by introducing new themes, but core SLE concerns keep reoccurring with varying phrasing which moves the emphasis or widens/narrows the intended scope. For example, "tools that support program refactoring" can become "mapping tools". To ensure stability, we normalised near-synonyms into single topics and kept the scheme coarse enough to absorb wording changes, while still fine-grained enough to expose meaningful differences. For instance, we kept separating T4B Horizontal Transformation from T4C Vertical Transformation rather than collapsing both into "transformation".

We have avoided creating tags for overly specific subfields that would not function without both full-text reading and deep domain expertise. Where ambiguity was common, we treated that as a boundary to document explicitly rather than as a reason to collapse topics. The process was highly iterative, and many initially proposed topics were either redefined as the tagging went on, or merged with others, or disappeared altogether.

For example, topic candidates such as "grammars" or "models" had to be dismissed, since both are terms with such broad interpretations within the SLE community, that tagging a paper with either was almost entirely devoid of meaning. The same happened to classic "concrete syntax" and "abstract syntax", since not only CfPs never mentioned those terms beyond the mapping context, but also virtually any paper presenting a software language, would include its concrete syntax in one way or another, and there were almost no papers that specifically focused on the design of the abstract syntax of a language (as opposed to using it as a means to quickly convey the set of language constructs). On the other side of the spectrum, it turned out to be insufficient to tag papers as doing "transformation", because there were clearly generative techniques papers performing these transformations "vertically", and they were distinct enough from "horizontal" ones with model-to-model mappings. In addition to those, some papers specifically contributed software languages to express transformations, or those which treated transformations in the context of evolution or maintenance activities.

Our final list includes **38** topics grouped into six families: software language definition and semantics (T1), software language evolution and governance (T2), language artefacts and representations (T3), toolchains and execution (T4), quality assurance and evaluation (T5), and application contexts and emerging directions (T6). For each topic, in the following subsections we provide a scope statement that captures what the topic *covers*, and occasionally add boundary notes indicating common confusions with neighbouring topics. These definitions are meant not only to collectively answer **RQ1**, but also to increase the replicability of this study by operationalising the topics into tagging guidance. Yet, this grouping is not claimed to be the only possible organisation; rather, it is a compromise that preserves interpretability while keeping categories mutually intelligible to annotators. For that reason, we intentionally separate *artefact kinds* (T3) from *tooling means* (T4) and from *validation lenses* (T5), while actual SLE papers frequently combine these dimensions.

4.1 T1: Defining a Software Language

T1A Design. Work whose primary contribution concerns *how* to design a software language: design principles, processes, challenges, trade-offs, notational choices, and domain modelling decisions that drive syntax and semantics. This includes design rationales, methods, and systematic approaches for constructing and bundling software language concepts. If the core novelty is a specific static or dynamic semantics technique, we use T1B/T1C and treat design as secondary.

T1B Static Semantics. Work focused on compile-time meaning: type systems, type inference rules, name resolution, scoping, symbol tables, constraint systems, static analyses for correctness, and static checking infrastructure. If the main contribution resides one linguistic layer higher (e.g., grammar well-formedness), we consider T5B instead.

T1C Behavioural Semantics. Work whose centre is runtime meaning: operational or denotational semantics, executable semantics, interpretable models, behavioural equivalence, semantic frameworks for execution, and techniques to define and compute behaviour. If the contribution is a concrete interpreter or a virtual machine rather than semantic definition techniques, we consider T4D.

T1D Composition. Work about decomposing and recombining languages or language components: language composition, embedding, coordination of language aspects with respect to syntax, semantics or tools, language integration in multi-language systems, and modular extension mechanisms.

T1E Reuse. Work where the core claim is reuse of software language artefacts: reusable grammars, reusable semantic components, reusable transformation libraries, reusable editors, and systematic reuse patterns across projects. If reuse was realised mainly as product-line variability, we treat T2C as primary.

4.2 T2: Software Languages over Time

T2A Lifecycle. Work that foregrounds end-to-end language governance: adoption, rollout, operations, deprecation, retirement, legacy language handling, organisational or process aspects of sustaining languages. If the technical core is in changing or maintaining the language, we lean towards using T2B.

T2B Evolution. Work about software language change and its consequences: language evolution, coupled/co-evolution of languages with tools or language definitions with language instances, migrations, refactorings across versions, maintenance techniques for language ecosystems. Process aspects often get secondary T2A, tracking changes by linking related artefacts can be secondary T5G.

T2C Variability. Work that treats families of software languages as first-class: variability modelling, software/language product lines, feature models for DSLs, configurable language components, systematic derivation of language variants. If variability is achieved primarily via modular composition mechanisms, T1D may be primary.

T2D Deployment. Work about putting languages into use across environments: deployment pipelines, portability across platforms, packaging, distribution, integration into build systems, IDEs or CI/CD, operational concerns such as language servers in toolchains. If the key novelty is compilation/code generation, we may prefer T4C.

T2E Documentation. Work where documentation or knowledge transfer is central: language documentation practices, literate specifications, pedagogical artefacts tightly coupled to a language, and documentation-driven engineering such as deriving software artefacts from docs/specs. Broader narratives around teaching or experience can be secondary T6A.

4.3 T3: Software Language Kinds

T3A Meta-languages. Work about languages used to define or manipulate software languages: grammar formalisms, metamodelling languages, attribute grammars, semantic specification languages, metaprogramming languages for language engineering, and extensions thereof. In some cases, when the contribution is both conceptual for the meta-language, and implementational for tool support, can be argued for either T3A or T4A as primary (e.g., SDF3 is T3A Meta-languages, but Spoofax is T4A Workbenches, yet both are of equal importance).

T3B Transformation Languages. Work whose object of study is a software language to express transformation itself: design, semantics, implementation of transformation DSLs, such as strategies, rewriting languages, query languages, and their expressiveness or correctness. If the main result is a specific transformation pipeline, we can use T4B/T4C.

T3C Domain-Specific Languages. Work that is primarily about domain-specific languages (DSLs) as artefacts: DSL design or implementation in a domain, comparative DSL case studies, DSL ecosystems, and domain-driven language engineering contributions. To some extent, almost all SLE papers use DSLs, so we avoid this tag when the DSL in question is merely a vehicle for demonstrating a software language engineering technique.

T3D General Purpose Languages. Work centred on general-purpose programming or modelling languages and their engineering: analyses of GPL design and semantics, tooling for GPLs, or leveraging GPL infrastructure in a way that advances language engineering understanding. If the key novelty is a meta-language or workbench, we prefer T3A/T4A.

T3E Ontologies. Work using knowledge graphs, ontologies, or semantic web representations for language engineering tasks: semantic integration, conceptual alignment, data inference, or reasoning over language artefacts with ontology-based formalisms. If the core contribution is requirements capture, T5A Requirements can be primary.

T3F Interfaces. Work about explicit interfaces around languages: APIs as software languages, embeddings as libraries, interoperation boundaries, structured access to external resources, and interfacing protocols that make languages accessible to other tools and services. The focus of such papers can also be on deployment or packaging, in which case T2D fits better, or on the IDE/tool integration infrastructure, in which case we should choose T4A.

4.4 T4: Software Language Tooling

T4A Workbenches. Work whose primary contribution is meta-tooling for building and using software languages: language workbenches as the most obvious example, but also all other meta-tools, editor and tool integration frameworks, infrastructures for generating or assembling language environments. While language analysis tools can fall into this category, the analyses of language definitions themselves belong in T5B, and transformation pipelines belong in T4B or T4C.

T4B Horizontal Transformation. Work where the central activity is translation or mapping at the same abstraction level: model-to-model, text-to-text, refactoring, migration, mappings, synchronisation, and conversions that preserve the level while changing representation.

T4C Vertical Transformation. Work about shifting abstraction levels downwards: compilation, code generation, model-to-text transformations, refinement, synthesis, lowering, as well as extraction, recovery, reverse engineering techniques that raise abstraction. If the key novelty is the transformation language itself rather than the pipeline or the implementation, we consider T3B.

T4D Interpretation. Work about executing languages via interpretation and runtime services: interpreters, virtual machines, runtime libraries, debugging and execution services, runtime architectures for language execution. At the same time, semantic definition techniques without a runtime emphasis fit better in T1C.

T4E Simulation. Work where simulation is the primary lens: simulators for modelling languages, discrete-event or continuous simulations tied to language semantics, and simulation-driven analysis workflows. Mentions of simulation languages like Simulink in early CfPs are also tagged with T4E.

T4F Visualisation. Work about visual representations and interaction: graphical modelling, visual syntaxes, diagramming, visual editors, visualisation-driven analytics of language artefacts, and human-facing rendering of language structures/behaviour. Usability studies of such tools are primarily T5H.

T4G AI-for-SLE. Work using AI/ML to support software language engineering: learning-assisted parsing or analysis, ML-guided transformation, LLM-based language tool support, mining language artefacts with AI, language workbenches with AI assistants. See also T6C for using SLE to engineer AI systems.

4.5 T5: Quality Assurance

T5A Requirements. Work that focuses on eliciting, capturing, analysing, or operationalising requirements relevant to software languages and software language-based systems: behavioural requirements, social or non-functional requirements, and systematic approaches linking requirements to

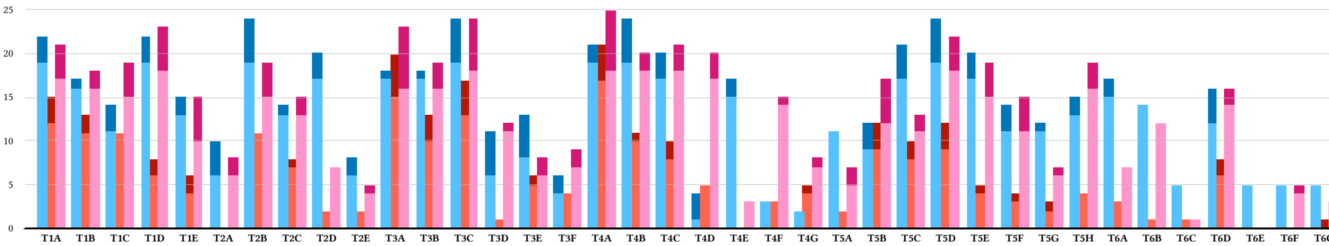


Figure 1. The summary of our coding: CfP prevalence of topics in blue, strict coding of proceedings in red, lax coding in pink, darker regions are special issue specific.

software language constructs or other elements of software language design.

T5B Language Description Analysis. Work analysing language specifications or definitions themselves: grammar well-formedness, specification completeness and consistency, meta-model quality checks, metrics application to language definitions, and static analysis of similar software language-defining artefacts.

T5C Testing. Work about testing software language implementations or language-defined systems: test generation for DSLs, model-based and property-based testing of software language artefacts, testing semantics, compilers, interpreters, and testing frameworks specific to software language engineering. Formal verification of language properties belongs to T5D.

T5D Formal Methods. Work applying formal verification and rigorous proof-based methods to software languages and software language artefacts: mechanised proofs, model checking, theorem proving, correctness of compilers, optimisations and transformations, formally verified semantics. Empirical performance benchmarking belongs to T5F.

T5E Empirical Evaluation. Work whose contribution is grounded in empirical observation: controlled experiments, quantitative and qualitative studies, corpus mining studies, evaluations comparing approaches, and methodological advances for evaluating software language engineering techniques. Usability-centred studies are typically primary T5H.

T5F Performance. Work focused on performance properties and optimisation: benchmarking of language tools, scalability studies, profiling and optimising transformations or runtimes, and in general performance engineering of software language infrastructure. If performance is only a supporting metric in an empirical evaluation, we consider T5E primary and T5F secondary.

T5G Traceability. Work establishing and exploiting trace links among language artefacts: traceability across models, grammars, code, transformations, versions, requirements, and derived artefacts; provenance and explainability of transformation chains. If the main goal is change management and co-evolution, T2B may be primary.

T5H Usability. Work where the main question is human use: readability of language notations, usability of abstract syntax, developer experience with language tools, interaction design for workbenches, cognitive dimensions of software language semantics, and user studies (see also T5E and T6A) focused on adoption and effectiveness.

4.6 T6: Software Language Context

T6A Experience Reports. Work that primarily reports practical experience, be it academic or industrial. Lessons learnt from building and using software languages, long-term maintenance experiences, pedagogical practice reports, and reflective case narratives. If the paper contributes a wider applicable technique while using experience in its use as validation, we tag that technique as primary.

T6B Industrial. Work whose centre is industrial application of SLE: industrial case studies, adoption in production settings, constraints and practices specific to industry, and reports emphasising industrial impact. User studies with industrial participants can be tagged primary with T5H Usability or T6A Experience Reports.

T6C SLE-for-AI. Work that uses SLE to engineer AI or ML systems and artefacts: DSLs for ML, language-based pipelines for AI, compilation and testing of ML systems, and language engineering as infrastructure for AI development. NB: using AI to improve SLE tools is T4G.

T6D Synergies. Work explicitly about cross-fertilisation between SLE and other (emerging) areas: conceptual bridges, shared methods, and agenda-setting pieces that position SLE relative to neighbouring communities. Papers specifically aiming to bridge different sides of SLE, can also be tagged with this.

T6E Quantum. Work about quantum software language engineering: quantum programming languages, modelling quantum processes, quantum DSLs, and language tooling specific to quantum computing contexts.

T6F Physical. Work about engineering software languages for physical systems and their digital counterparts: cyber-physical systems, internet of things, digital twins, robotics, any embedded contexts where modelling and execution interacts with the physical world.

	T1A Design	T1B Static Semantics	T1C Behavioural Semantics	T1D Composition	T1E Reuse	T2A Lifecycle	T2B Evolution	T2C Variability	T2D Deployment	T2E Documentation	T3A Meta-languages	T3B Transformation Languages	T3C Domain-Specific Languages	T3D General Purpose Languages	T3E Ontologies	T3F Interfaces
SLE 2007S	•		•	•	•	•	•	•		•			•	•	•	
SLE 2008	•			•		•	•		•	•	•		•	•	•	•
SLE 2008S	•	•		•		•	•		•	•			•	•	•	•
SLE 2009	•	•		•		•	•		•	•	•		•	•	•	•
SLE 2010	•	•		•		•	•			•	•	•	•		•	
SLE 2011	•	•		•		•	•			•	•	•	•		•	
SLE 2012	•	•		•		•	•		•	•	•	•	•	•	•	•
SLE 2012S															•	
SLE 2013	•	•		•		•	•		•	•	•	•	•	•	•	•
SLE 2014	•			•	•		•	•	•		•	•	•	•	•	
SLE 2014S			•			•	•		•				•	•	•	
SLE 2015	•			•	•		•	•	•		•	•	•	•	•	
SLE 2016	•	•	•	•	•		•	•	•			•	•			
SLE 2017	•	•	•	•	•		•	•	•			•	•			
SLE 2018	•	•	•	•	•		•	•	•		•	•	•			
SLE 2018S						•	•		•				•	•	•	
SLE 2019	•	•	•	•	•		•	•	•		•	•	•			
SLE 2020	•	•	•	•	•		•	•	•		•	•	•			
SLE 2021	•	•	•	•	•		•	•	•		•	•	•			
SLE 2022	•	•	•	•	•		•	•	•		•	•	•			
SLE 2023	•	•	•	•	•		•	•	•		•	•	•			
SLE 2024	•	•	•	•	•		•	•	•		•	•	•			
SLE 2024S	•		•	•	•		•				•	•	•	•		•
SLE 2025	•	•	•	•	•		•	•	•		•	•	•			
SLE 2026	•	•	•	•	•		•	•	•		•	•	•			
Prevalence	88%	68%	56%	88%	60%	40%	96%	56%	80%	32%	72%	72%	96%	44%	52%	24%
w/o S.I.s	100%	84%	58%	100%	68%	32%	100%	68%	89%	32%	89%	89%	100%	32%	42%	21%

	T4A Workbenches	T4B Horizontal Transformation	T4C Vertical Transformation	T4D Interpretation	T4E Simulation	T4F Visualisation	T4G AI-for-SLE	T5A Requirements	T5B Language Description Analysis	T5C Testing	T5D Formal Methods	T5E Empirical Evaluation	T5F Performance	T5G Traceability	T5H Usability	T6A Experience Reports	T6B Industrial	T6C SLE-for-AI	T6D Synergies	T6E Quantum	T6F Physical	T6G Socio-technical	Σ
SLE 2007S		•	•						•	•	•	•		•					•				19
SLE 2008	•	•	•					•	•	•	•												18
SLE 2008S	•	•	•	•					•	•													17
SLE 2009	•	•	•			•		•	•	•	•												20
SLE 2010	•	•						•	•	•	•	•				•			•				19
SLE 2011	•	•						•	•	•	•	•				•			•				19
SLE 2012	•	•	•		•	•		•	•	•	•	•			•	•	•		•				27
SLE 2012S											•								•				3
SLE 2013	•	•	•		•	•		•	•	•	•	•			•	•	•		•				27
SLE 2014	•	•	•		•				•		•	•					•		•				20
SLE 2014S	•	•		•	•				•	•	•	•	•		•	•			•				19
SLE 2015	•	•	•	•	•				•		•	•							•				20
SLE 2016	•	•	•		•					•	•	•	•	•	•	•	•						22
SLE 2017	•	•	•		•					•	•	•	•	•	•	•	•						22
SLE 2018	•	•	•		•					•	•	•	•	•	•	•	•						23
SLE 2018S		•			•						•		•			•			•				12
SLE 2019	•	•	•		•					•	•	•	•	•	•	•	•						23
SLE 2020	•	•	•		•					•	•	•	•	•	•	•	•		•				24
SLE 2021	•	•	•		•					•	•	•	•	•	•	•	•						23
SLE 2022	•	•	•		•			•		•	•	•	•	•	•	•	•	•	•	•	•	•	29
SLE 2023	•	•	•		•			•		•	•	•	•	•	•	•	•	•	•	•	•	•	29
SLE 2024	•	•	•		•			•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	30
SLE 2024S		•	•	•						•	•	•	•		•								18
SLE 2025	•	•	•		•		•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	30
SLE 2026	•	•	•		•		•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	30
Prevalence	84%	96%	80%	16%	68%	12%	8%	44%	48%	84%	96%	80%	56%	48%	60%	68%	56%	20%	64%	20%	20%	20%	
w/o S.I.s	100%	100%	89%	5%	79%	16%	11%	58%	47%	89%	100%	89%	58%	58%	68%	79%	74%	26%	63%	26%	26%	26%	

Table 3. Tagging all the CfPs (or proxies thereof): each column is a tag from section 4, each row is a conference "bundle". In the rightmost column, the sum is given as an indication of how broad the corresponding CfP is. In the lowest two rows, we included prevalence of each tags in all bundles, and in all proper conference bundles. SLE 2017S is excluded because its "call for papers" was an email without any taggable text. **Core** topics are highlighted in green, **near-core** in golden and **fringe core** in yellow.

T6G Socio-technical. Work foregrounding social, organisational, and human-system aspects intertwined with language engineering: socio-technical system modelling, governance and policy aspects encoded in languages, and human/organisational constraints treated as first-class concerns.

4.7 Resulting Adequacy

Our topic coding is summarised on Figure 1. It will be discussed in much more detail in the next section, but let us do a quick reflection to establish the adequacy of the chosen topics. This is a sanity check that the scheme is not biased towards proceedings-only distinctions or towards wishful thinking. From the summary we can see that every one of the **38** topics appears at least once in the CfP corpus (i.e., there exists at least one year which CfP explicitly supports mapping some CfP bullet to the topic).

In subsection 3.3 we mentioned that each paper from the proceedings and the special issues was tagged with one primary code and possibly some secondary ones. Under *strict* coding (considering only one primary tag per paper), **34**/**38** topics occur as a primary topic at least once; the remaining four (T2A Lifecycle, T4E Simulation, T6E Quantum, T6F Physical) are present in CfPs but never dominate as a paper's primary topic. Under the *lax* view (where each paper is viewed as a multi-topic entity, and all of its primary and secondary tags are considered), **37**/**38** topics occur at least once in proceedings tags; the only topic absent from proceedings in our current dataset is T6E Quantum, despite its appearance in recent CfPs. This pattern is informative rather than problematic: it indicates that the taxonomy is expressive enough to represent novel ambitions in CfPs as well as realised contributions in proceedings, and it motivates our later analysis of call–paper alignment and "under-realised" areas.

One can argue that a body-of-knowledge perspective is inherently multi-dimensional: a single artefact can contribute to, say, a transformation formalism, a meta-language

	T1A Design	T1B Static Semantics	T1C Behavioural Semantics	T1D Composition	T1E Reuse	T2A Lifecycle	T2B Evolution	T2C Variability	T2D Deployment	T2E Documentation	T3A Meta-languages	T3B Transformation Languages	T3C Domain-Specific Languages	T3D General Purpose Languages	T3E Ontologies	T3F Interfaces
SLE 2007S	1		0	3	1	1	1	1		0	4	2	1	0	1	
SLE 2008	2	1	1	4	3	0	4	1	0	0	1	3	3	0	0	0
SLE 2008S	2	0		2	1	0	0	1	0	0	3		3	0	0	1
SLE 2009	3	1		9	2	1	4	3	1	0	6	3	9	2	1	2
SLE 2010	5	0	3	5		0	6	3		1	3	2	6	1	2	
SLE 2011	4	2	3	5	3	1	3	1		0	4	4	4	1	2	4
SLE 2012	1	3	1	5	1	0	3	1	0	0	11	3	2	0	2	2
SLE 2012S	1		1	3		1					2		3		1	
SLE 2013	4	2	3	4	1	0	3	1	0	0	7	4	7	0	1	0
SLE 2014	3	1	4	6	0	1	0	1	0		4	5	5	1	0	
SLE 2014S			2	1	1	0	0		0		2	1	2	0	0	
SLE 2015	4	2	2	4	1		4	1	1		3	5	8	1	0	1
SLE 2016	4	5	3	4	0	1	3	0	1		1	3	7	2		1
SLE 2017	4	7	1	9	2		4	0	0	1	10	0	9	2	1	
SLE 2017S		1	1		1		1				2		1			
SLE 2018	2	4	0	6	0		0	1	0		3	2	2			
SLE 2018S						0	1		0		2		0	1	0	
SLE 2019	6	1	2	7	0	1	1	0	0		0	4	4			
SLE 2020	4	6	1	3	0	1	0	0	0		2	2	5			
SLE 2021	5	0	1	4	0		1	1	0		0	3	4	1		
SLE 2022	5	3	1	4	2		1	0	1		3	2	6	3		2
SLE 2023	0	3	3	6	1		3	1	1		4	2	6			
SLE 2024	4	3	0	2	0		4	1	2	1	3	2	6	3		2
SLE 2024S	2	1	1	2	1		2			1	2	3	2	0		1
SLE 2025	2	1	3	1	1		4	2	2	1	6	0	1	2		
Prevalence	84%	72%	76%	92%	60%	32%	76%	60%	28%	20%	92%	76%	96%	48%	32%	36%
w/o S.I.s	94%	89%	83%	100%	56%	33%	83%	72%	39%	22%	89%	89%	100%	61%	33%	39%
Papers	68	47	37	99	22	8	53	20	9	5	88	55	106	20	11	16

	T4A Workbenches	T4B Horizontal Transformation	T4C Vertical Transformation	T4D Interpretation	T4E Simulation	T4F Visualisation	T4G AI-for-SLE	T5A Requirements	T5B Language Description Analysis	T5C Testing	T5D Formal Methods	T5E Empirical Evaluation	T5F Performance	T5G Traceability	T5H Usability	T6A Experience Reports	T6B Industrial	T6C SLE-for-AI	T6D Synergies	T6E Quantum	T6F Physical	T6G Socio-technical	Σ
SLE 2007S	3	1	0					1	2	0	2	2		1					0				17
SLE 2008	8	7	2	1		1		0	4	1	2	3		2	1	2			3				23
SLE 2008S	2	0	1	0		1			1	0					1								12
SLE 2009	9	5	3	1		1		1	0	0	4				3				1				23
SLE 2010	8	6	3			2		0	5	0	3	5	2		2	0	1		1			1	23
SLE 2011	5	6	1	1		1		0	4	1	2	2	1	1		0	3		0				26
SLE 2012	11	3	1	2	0	1	1	0	5	1	5	0	2		3	0	1		1			1	26
SLE 2012S	6	1		1					2	1	3		1						1				15
SLE 2013	3	5	3	2	1	1		0	5	0	6	2			2	0	1		1				23
SLE 2014	11	2	2	4	0	3	1		4	2	4	1	2	1	5	1	0		0				24
SLE 2014S	4	0	1	1	0				1	0	2	2	1		1	0			1				15
SLE 2015	8	4	2	2	0	2			4	1	5	2	1		1		1		1		1		27
SLE 2016	8	4	3	7	1	3	1	1	3	2	2	0	1	2	3	1	2		1		1		30
SLE 2017	11	1	3	6	0	1		2	2	0	4	4	7	0	6	0	2		1				24
SLE 2017S	1																						7
SLE 2018	18	3	5	5	0					1	7	5	0	0	4	0	5		3				17
SLE 2018S	1	0	2		0						0	2	1			0			0				7
SLE 2019	15	2	3	2	0	1				0	4	5	0	0	5	0	3		1				18
SLE 2020	16	7	6	4	0					0	7	5	0	0	6	0	2		0				16
SLE 2021	13	2	3	4	0		1			0	3	4	0	0	4	0	1		1				18
SLE 2022	8	2	4	3	0	2	1	1	1	3	3	2	6	1	4	1	0	0	2	0	0	0	28
SLE 2023	8	3	3	4	0	2	2	0	1	1	2	1	2	2	3	1	1	1	1	0	1	0	28
SLE 2024	2	3	5	2	0	2	2	1	0	1	1	6	7	0	2	1	0	0	3	0	1	1	28
SLE 2024S	6	0	0	1			1	1	3	3	1	3	3		1						1		22
SLE 2025	6	1	6	3	1		0	0	1	2	3	1	2	0	0	2	0	0	0	0	0	0	23
Prevalence	100%	80%	84%	80%	12%	60%	32%	28%	68%	52%	88%	76%	60%	28%	76%	28%	48%	4%	64%	0%	20%	12%	
w/o S.I.s	100%	100%	100%	94%	17%	78%	39%	28%	67%	61%	100%	83%	61%	33%	89%	39%	67%	6%	78%	0%	22%	17%	
Papers	191	68	62	56	3	24	10	8	48	20	75	57	39	10	57	9	23	1	23	0	5	3	

Table 4. Number of papers in a year bundle (row) tagged with a specific code (column). Green numbers match the corresponding CfP; red zeroes represent topics requested in the CfP but not found in the proceedings; golden numbers are pleasant surprises — contributions on topics not present in the corresponding CfP. The data in this table corresponds to *lax* coding: the primary tag and secondary tags are considered together. The rightmost column contains the sum of topics, not of papers, because we want to see how broad each proceedings volume was. SLE 2026 is excluded because its papers are not yet available. **Core** topics are highlighted in green, **near-core** in golden and **fringe core** in yellow.

embedding, and a validation method simultaneously. The annotated paper lists in the corpus repeatedly exhibit such combinations (e.g., work that is primarily about a transformation language while also relying on formal methods and workbench infrastructure). Therefore, the taxonomy is designed to support both a strict "what is this paper mainly about?" view and a lax "what knowledge elements does this paper contribute to?" view, allowing the longitudinal analysis to bracket the effect of single-label versus multi-label modelling. As will be substantiated in subsection 5.2, the lax approach seems more promising for the body of knowledge viewpoint.

Taken together, this evidence supports the claim that the taxonomy is: (1) broad enough to cover the topical surface area expressed in CfPs, (2) aligned with the kinds of contributions that appear in proceedings (especially under lax multi-topic coding), and (3) structured around boundaries that are explicit enough to be applied consistently. The remaining weakness is not primarily the absence of topics, but the robustness of annotation: a future replication should involve independent coders, report agreement for the multi-label setting, and adjudicate boundary cases against the codebook and, where necessary, the full papers, possibly even involving the authors of the original papers.

5 Topical Evolution in Calls and Papers

To answer the second research question from section 2, we analyse topical change over time in *both* SLE CfPs and SLE proceedings. For each "bundle" (conference year or a special issue), we derive the set of *requested* topics mentioned in the CfP, as well as the set of *received* topics occurring in accepted papers. Since for some special issues we only have the editorials and not the actual calls for papers, sometimes we handle those two subsets apart at least in some parts of the narrative below. For proceedings, we report two complementary views: a *strict* (single-label) view where each paper contributes exactly one primary topic and a *lax* (multi-label) view where papers contribute multiple knowledge elements.

The results are visualised in Table 3 for the CfP coding and in Table 4 for the lax paper coding. In the following subsections, we interpret this data, characterising evolution via *persistence* (topics occurring in all or almost all years), *emergence* (first appearance year), *fade-out* (last appearance year), and *topical diversity* per year (the number of distinct topics present, optionally supplemented by distributional diversity).

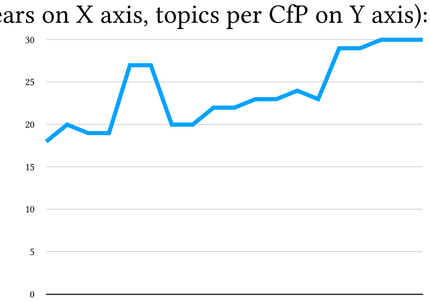
5.1 CfPs: Persistent Core, Expanding Perimeter

We focus on CfPs of SLE 2008–2026 and CfPs/editorials of special issues of 2007, 2008, 2012, 2013/2014, 2018, 2024, excluding the special issue of 2017 from the calculations in this section since its call was an email without any text markable for topical coverage. Across these **25** CfPs, we observe a remarkably stable **core** of seven topics that are requested in *every* conference year and almost every special issue year: T1A Design (absent from special issues of 2012, 2014, 2018), T1D Composition (same), T2B Evolution (from 2012), T3C DSLs (also from 2012), T4A Workbenches (from 2007, 2012, 2018, 2024), T4B Horizontal Transformation (from 2012), and T5D Formal Methods (from 2008). This persistent core effectively defines SLE's long-running self-description: the venue continuously positions itself around language engineering as an interplay between language definition and composition, change over time, dedicated tooling and transformations, and rigorous analysis.

To look beyond the core systematically, we compare topic prevalence across bundles. Among the **core** topics, prevalence in regular proceedings is 100%, but looking overall at all bundles, it ranges between **21–24**/**25** (84–96%). If we flip it around and look at other topics with overall prevalence upwards of 84%, we find another **near-core** topic: T5C Testing, absent from CfPs only of SLE 2014 and SLE 2015. Applying the same threshold in the opposite direction, we find the **fringe core** set of topics also appearing in all but 2 conferences' CfPs, which are: T2D Deployment (absent from SLE 2010–2011), T3A Meta-languages (from SLE 2016–2017), T3B Transformation Languages (from SLE 2008–2009), T4C Vertical Transformation (from SLE 2010–2011), T5C Testing (from SLE 2014–2015), and T5E Empirical Evaluation (from SLE 2008–2009). Taken together, the core, near-core and fringe core suggest that, rather than focusing on any single research style, SLE's CfPs consistently cover an end-to-end engineering story: definition ⟶ tooling ⟶ transformation/execution ⟶ validation ⟶ deployment/evolution. All the other **25** topics are less persistent in the CfP record; this does not make them less valuable, but it means that they function more as episodic, contextual, or emerging concerns under this corpus view.

In contrast, several topics show *explicit* fade-out in CfPs after the early years. Notably, T2A Lifecycle and T2E Documentation cease to be mentioned after 2013, and T3F Interfaces disappears after 2013 as well. T3D GPLs and T3E Ontologies persist slightly longer but are last explicitly requested in 2015 with a brief comeback in the editorial of the special issue of 2018 (which features no ontological papers!). Finally, T4F Visualisation and T4D Interpretation appear only sporadically and also stop being named explicitly after 2013–2015. Importantly, this is a statement about *CfP phrasing*, not necessarily about what the community publishes; we return to this mismatch below and in RQ3.

Emergence in CfPs. The most visible perimeter expansion occurs in the 2022 CfP, which introduces an explicit cluster of newer context topics: T6C SLE-for-AI, T6E Quantum, T6F Physical, and T6G Socio-technical. All CfPs before 2021 only vaguely hinted at applications and synergies without naming these topics explicitly. A second addition is T4G AI-for-SLE, first explicitly requested in 2025 and retained in 2026. Thus, the core has been stable since the founding of the conference in 2008 (or even before that, since the pre-SLE special issue mentions all core topics except T4A Workbenches). Yet, the CfPs of SLE conferences have been becoming broader in how they connect SLE to adjacent systems contexts and to AI both as an application domain and as a technique, and in general broader in terms of topics they cover (years on X axis, topics per CfP on Y axis):



Since we do not make any claims about the conceptual domain coverage of each topic, whatever definition we might give it, we cannot conclude that the SLE domain has actually been expanding over the last two decades, because the alternative explanation would be that the definition of the scope of SLE has just been becoming increasingly more refined.

5.2 Proceedings: Stable Breadth

Turning to proceedings (SLE 2008–2025), the perceived topical evolution depends strongly on whether we look at primary topics only or at multi-topic (BoK-like) contributions. This time, we include the special issue of 2017, but exclude the proceedings of 2026 since they are not yet available. So, the total number of bundles for computations in this section is still 25, just for different reasons.

Strict (primary-topic) view. Under strict coding, only one topic appears as a primary topic *every year*: T4A Workbenches (which in our interpretation covered any meta-tools and meta-environments). This does not mean SLE is "only" about workbenches or meta-tooling; rather, it indicates that

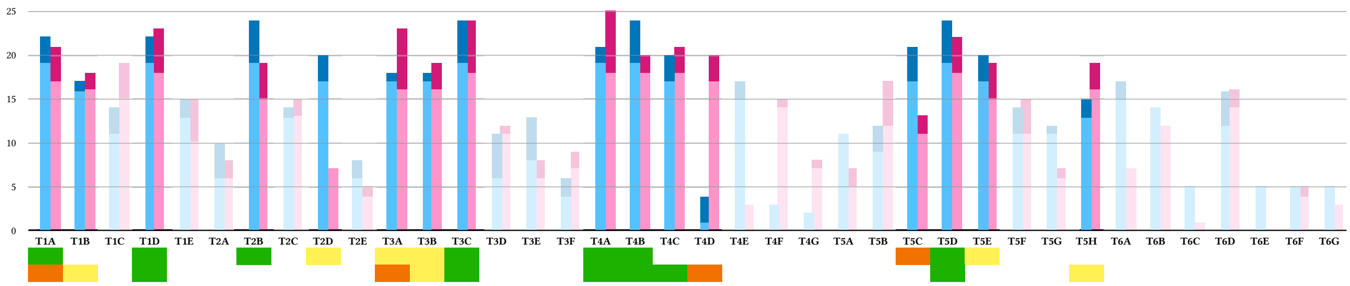


Figure 2. A revised summary of our coding from Figure 1: strict proceedings coding data is removed, bars for non-core topics dimmed out, core status added on the bottom. The higher line represents CfP coding, the lower line is about proceedings coding. Green labels **core** topics, golden are for **near-core** and yellow for **fringe core**. Darker regions on the bars are special issue specific.

when one is forced to choose a single primary label, workbench/tooling contributions repeatedly dominate the centre of gravity of yearly proceedings. Other frequently recurring primary topics include T3A Meta-languages and T3C DSLs (each present as a primary topic in 14/18 years), and T1A Design and T1B Static Semantics (13/18 years). In contrast, some topics are rarely or never primary in proceedings (e.g., T2A Lifecycle and T4E Simulation never appear as primary topics), even though they may still appear as secondary contributions.

Lax (multi-topic) view. Under the lax view, proceedings exhibit a substantially more stable and broad topical footprint. Six **core** topics occur *in every conference year* as (primary or secondary) contributions: T1D Composition, T3C DSLs, T4A Workbenches, T4B Horizontal Transformation, T4C Vertical Transformation, and T5D Formal Methods. This set is strikingly close to the CfP core (with T4C present here instead of T2B, which is still highly prevalent). It supports a BoK interpretation in which SLE papers routinely contribute to multiple knowledge elements even when their headline framing is "about" a particular tool or language artefact.

Let us compute adjacent topics like we did for CfPs. Out of the core topics, T2B Evolution has the lowest prevalence across all bundles (including special issues), which is 20/25 (80%). If we turn this around and look back at topics which have the same or higher overall coverage, we find a few topics we can see as **near-core** since they are occasionally missing from regular proceedings. These are: T1A Design (not in SLE 2023), T3A Meta-languages (not in SLE 2019 and 2021), and T4D Interpretation (not in SLE 2010). The lowest proceedings coverage is of T3A Meta-languages with 16/18 (89%), and flipping that back, we get our **fringe core** set: T1B Static Semantics (not in SLE 2010 and 2021), T3B Transformation Languages (not in SLE 2017 and 2025), and T5H Usability (not in SLE 2011 and 2025).

Emergence and fade-out in proceedings. Compared to CfPs, proceedings show fewer sharp phase transitions. Still, there are notable signals. T3E Ontologies is the clearest fade-out topic in proceedings, with its last occurrence (even as a secondary tag) in 2017, aligning with the topic's disappearance from CfPs by 2015. T6C SLE-for-AI is the clearest genuinely new proceedings topic in the dataset, first appearing in 2023, shortly after being introduced in the 2022 CfP. Other "new in CfPs" topics such as T6E Quantum have not yet materialised in proceedings at all through 2025.

5.3 Topical Diversity Evolution

We quantify topical diversity in two complementary ways: *set diversity* (the number of distinct topics present per year) and *distributional diversity* (how evenly topic counts are distributed within a year).

Set diversity. CfP topic breadth increases over time. Early SLE CfPs typically mention about 18–20 distinct topics per year (e.g., 18 in 2008; 20 in 2009). After 2020, CfPs broaden substantially, reaching 29–30 distinct topics per year in 2022–2026. Averaged by era, CfPs mention 21.7 distinct topics per year in 2008–2013 and 2014–2019, but 27.5 topics per year in 2020–2025. This expansion reflects an increasing tendency of CfPs to provide comprehensive (and sometimes catalogue-like) topic lists rather than a small set of illustrative examples.

Proceedings diversity is more stable. Under strict (primary-only) coding, proceedings exhibit about 11–17 distinct primary topics per year (mean 14.2 in 2008–2013; 10.8 in 2014–2019; 12.3 in 2020–2025). Under the lax (multi-topic) view, proceedings consistently cover roughly 23–28 topics per year, remaining close to 23–24 topics on average across all eras. Hence, while CfPs broaden over time, the proceedings' *BoK-level* topical footprint is comparatively stable, and the apparent narrowing in some periods is largely an artefact of forcing a single primary label per paper.

Distributional diversity. Using Shannon entropy over topic counts as a coarse distributional proxy, CfP diversity increases from roughly 4.0 bits in 2008 to a peak around 4.69 bits in 2024, consistent with the increased breadth and more even spread across topics in recent calls. Proceedings

Topic	Core?	Years			Taggings		
		CfP	Strict	Lax	CfP	Strict	Lax
T4A Workbenches		21	**21**	**25**	46	**54**	**191**
T3C Domain-Specific Languages		**24**	17	24	50	35	106
T1A Design		22	15	21	61	35	68
T3A Meta-languages		18	20	23	32	32	88
T5D Formal Methods		**24**	12	22	69	16	75
T2B Evolution		**24**	11	19	**84**	19	53
T1D Composition		22	8	23	52	20	99
T4B Horizontal Transformation		**24**	11	20	64	15	68
T4C Vertical Transformation		20	10	21	64	17	62
T1B Static Semantics		17	13	18	23	25	47
T3B Transformation Languages		18	13	19	19	23	55
T5C Testing		21	10	13	32	13	20
T5E Empirical Evaluation		20	5	19	30	8	57
T4D Interpretation		**4**	5	20	**4**	12	56
T5H Usability		15	4	19	16	5	57
T2D Deployment		20	**2**	**7**	23	**2**	**9**

Table 5. A candidate list of core topics of SLE, based on most frequent topics in CfPs and proceedings of SLE. The colours of core/near-core/fringe core are the same as in all previous tables. Per topic, in each row, we report quantitatively first: the number of years where that topic was tagged; the number of years where that topic was a primary tag for at least one paper; and the number of years where there was at least one paper tagged with it; and then: the number of taggings in all CfP; the number of primary taggings in all proceedings; and finally the number of all taggings of that topic in all proceedings. Maximum values in each column and minimum values in each column are highlighted in red and green, respectively. The order of rows is determined by the collective rank computed over all six columns.

entropy is more variable under strict coding (reflecting year-to-year differences in how dominant the main primary topics are), but remains comparatively stable under lax coding (roughly 4.1–4.5 bits), again indicating that multi-topic contributions maintain a broad knowledge footprint even when one topic dominates primary labels.

These metrics are used comparatively within our corpus, not as absolute quality measures. A higher number of distinct topics or higher entropy does not mean that a CfP or proceedings volume is "better"; it means that topic coverage is broader or *more evenly distributed* under our taxonomy. Since the maximum entropy over **38** equally represented topics would be $\log_2 38 \approx 5.25$ bits, the observed values around 4.0–4.7 bits indicate broad but not uniform coverage. The relevant signal is therefore the relative change over time and the difference between strict and lax coding, not whether a single entropy value is good or bad.

5.4 Summary and Inventory

Across 19 years and 26 bundles of CfPs and proceedings, we find a consistent picture: SLE's topical *core* has been relatively stable since 2008, and both calls and papers continuously cover key knowledge elements around language design, modularity, evolution, workbenches, transformations, and rigorous methods. At the same time, CfPs have broadened substantially since 2022, explicitly incorporating socio-technical and AI-adjacent contexts, whereas proceedings show slower uptake. We have seen clear evidence only for T6C SLE-for-AI by 2023 and no proceedings evidence yet for T6E Quantum through 2025. Finally, the multi-topic (BoK-like) view reveals stable and broad topical coverage in proceedings across eras, suggesting that SLE's body of knowledge grows primarily through cross-cutting contributions rather than through cleanly separated topical subcommunities.

Two examples illustrate requested topics that have not (yet?) received sufficient resonance in accepted papers. T6G Socio-technical was in all conference CfPs since 2022 (but not mentioned in the editorial of the special issue of 2024). Yet, there was only one paper from SLE 2010 on naming conventions, one keynote from SLE 2012 and one panel from 2024. Similarly for T6F Physical, there was one paper from SLE 2015 which wanted IoT for its concurrent nature, one with a comparable motivation from the special issue of SLE 2024, one paper from SLE 2016 which used CPS to motivate units as a language concept, the SLE 2023 keynote on digital twins, one embedded systems paper from SLE 2023. T6F Physical was also requested in CfPs from SLE 2022–2026, but not mentioned in the editorial of the special issue of 2024.

In Figure 2, we refine Figure 1 to have a more prominent visualisation of core topics we have identified in the previous sections. A further refinement yet is the overview given in Table 5. There, we list the most frequent proceedings topics under strict (primary) tagging, along with their multi-tag totals and CfP mention counts. A first observation is that T4A Workbenches is the most frequent primary topic and appears every year in the proceedings corpus. In contrast, topics such as T2B Evolution and T5D Formal Methods are heavily emphasised in CfPs but are less dominant as *primary* tags, while being more visible when adopting the multi-tag (BoK-like) view.

Not all topics have the same behaviour through the years: as we might recall, Table 3 had some sparse areas. Focusing purely on the conference CfPs, we summarise the persistence of all topics in Table 6 and also list the size of the longest streak as well, since topic presence in CfPs can be interrupted. For instance, T4F Visualisation was introduced in 2009, then disappeared for two years, then re-emerged in 2012, stayed for 2013, and was removed afterwards. We could perform the same visualisation for proceedings presence, but it makes less sense to do so: CfPs have a natural inheritance relation across the years, since when writing the new call, programme chairs first always engage with the call of the preceding year. Proceedings, on the other hand, are more "accidental" and individual in the sense that each year is not directly related to other years, and there are many confounding factors like paper rejections, deadline positioning, alternative venues, research trends, authors migrating across communities, etc.

As one can see, proceedings contributions do not always match the calls for papers topic-wise (more on this in the next section), but the majority of *topics* (**23**/**38**) seem to be present from the very first conference and certainly since SLE 2010 (**33**/**38** topics). Notable latecomers are T4G AI-for-SLE (2012), T4E Simulation (2013), T6F Physical (2015), T6C SLE-for-AI (2023), as well as the never-present T6E Quantum.

Despite the already noted continuous growth of CfPs in terms of topic coverage, there are some topics that were initially prominent in CfPs, but later "retired". Not taking special issues into account (whose editorials are written *a posteriori* and thus directly influenced by papers), mostly this concerns T2A Lifecycle, T2E Documentation, T3F Interfaces and T4F Visualisation (gone since 2014); as well as T3D GPLs, T3E Ontologies and T4D Interpretation (gone since 2016). Some topics, however, make a comeback, which can be fleeting (T5B Language Description Analysis was gone since 2016 but reappeared once in 2024) or persistent (T5A Requirements was gone in 2014–2021 but is consistently included since 2022).

Having experimented with strict single-topic tagging which assigns each paper into one topic category, we have to conclude that the lax tagging allowing multiple topics to be associated to the same paper, was both easier to do and produced more sensible results. Since this leads to an annotated dataset with 2–4 topics tagged per paper, we can also consider topic co-occurrences. A network graph visualising those, is depicted in Figure 3. We take top ten of the most popular co-occurrences and write them out explicitly in Table 7: almost all of them are combining T4A Workbenches (which covers any other metatools) with aspects like modularity, usability, or verification. There are only two exceptions: a combination of T1D Composition with T3C DSLs (modular/extensible/embedded software languages) and a combination of T3B Transformation Languages with T4B Horizontal Transformation (model transformation languages).

Topic	CfP Years	Streak	First paper
T1A Design	2008 – 2026	**19** years	2008
T1B Static Sem.	2009 – 2026	11 years	2008
T1C Behavioural Sem.	2008 – 2026	11 years	2008
T1D Composition	2008 – 2026	**19** years	2008
T1E Reuse	2014 – 2026	13 years	2008
T2A Lifecycle	2008 – 2013	6 years	2009
T2B Evolution	2008 – 2026	**19** years	2008
T2C Variability	2014 – 2026	13 years	2008
T2D Deployment	2008 – 2026	15 years	2009
T2E Documentation	2008 – 2013	6 years	2010
T3A Meta-languages	2008 – 2026	9 years	2008
T3B Transf. Languages	2008 – 2010	17 years	2008
T3C DSLs	2008 – 2026	**19** years	2008
T3D GPLs	2008 – 2015	4 years	2009
T3E Ontologies	2008 – 2015	8 years	2009
T3F Interfaces	2008 – 2013	**2** years	2009
T4A Workbenches	2008 – 2026	**19** years	2008
T4B Horizontal	2008 – 2026	**19** years	2008
T4C Vertical	2008 – 2026	15 years	2008
T4D Interpretation	2015 – 2015	**1** year	2008
T4E Simulation	2012 – 2026	15 years	2013
T4F Visualisation	2009 – 2013	**2** years	2008
T4G AI-for-SLE	2025 – 2026	**2** years	2012
T5A Requirements	2008 – 2026	6 years	2009
T5B Lang. Desc. Analysis	2008 – 2024	8 years	2008
T5C Testing	2008 – 2026	11 years	2008
T5D Formal Methods	2008 – 2026	**19** years	2008
T5E Empirical	2010 – 2026	17 years	2008
T5F Performance	2015 – 2026	11 years	2010
T5G Traceability	2015 – 2026	11 years	2008
T5H Usability	2012 – 2026	11 years	2008
T6A Experience Reports	2010 – 2026	11 years	2008
T6B Industrial	2012 – 2026	11 years	2010
T6C SLE-for-AI	2022 – 2026	5 years	2023
T6D Synergies	2010 – 2026	6 years	2008
T6E Quantum	2022 – 2026	5 years	—
T6F Physical	2022 – 2026	5 years	2015
T6G Socio-technical	2022 – 2026	5 years	2010

Table 6. Presence and persistence of topics in CfPs of pure conferences through: per topic, we list the (possibly interrupted) range of years this topic was actively mentioned; and the length of the longest streak for the topic to stay in CfPs.

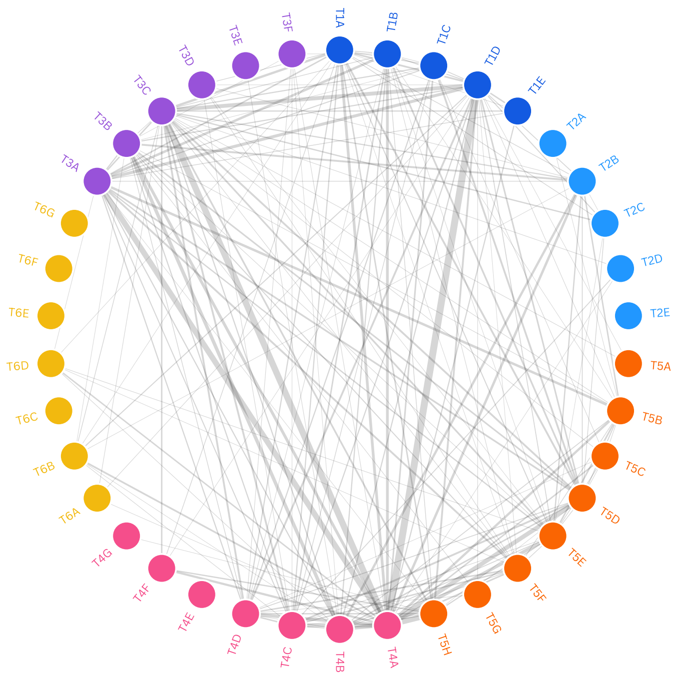


Figure 3. Radial network graph of topic co-occurrence. Relations linking topics co-occurring only once or twice are considered accidental and removed from the visualisation.

6 Call–paper Alignment and Responsiveness

The third research question from section 2 asks whether SLE's calls (what the community *asks for*) and its proceedings (what the community *publishes*) exhibit systematic topical mismatches, and whether shifts in calls are reflected in subsequent proceedings (possibly with a lag). We analyse alignment at two complementary levels: (i) *set-level* alignment of topics per year (which topics appear at all), and (ii) *prominence-level* alignment (which topics are disproportionately emphasised in calls vs. proceedings). For each year y, let R_y be the set of topics requested in the CfP, and let P_y be the set of topics present in accepted papers. Similarly, for each topic t, let R_t be the set of years that this topic was tagged in the CfP, and let P_t be the set of years which had accepted papers tagged with this topic. For proceedings we again distinguish a *strict* view (only primary topic per paper) and a *lax* view (multi-topic, BoK-like coding).

For Figure 4, we condense and visualise the observations one might make by staring for some substantial time at Table 4. The green part of each horizontal bar are the topics represented both in the CfP and in the proceedings of the corresponding year; the red part are the CfP requests that have not been answered by papers; and the golden part are "unsolicited" topics for which there are papers despite the lack of explicit mentions about that topic in the CfP.

6.1 Set-level Alignment

A simple, interpretable alignment score is the Jaccard index:

$$J_y = \frac{|R_y \cap P_y|}{|R_y \cup P_y|}$$

Intuitively, $J_y = 0$ would mean that $R_y \cap P_y = \varnothing$ and there is no overlap between the CfP topics and the proceedings topics at all, while $J_y = 1$ would mean that the overlap is perfect and $R_y = P_y$. Across all 25 bundles, mean set overlap is moderate under strict coding ($\overline{J} \approx 0.31$) and substantially higher under lax coding ($\overline{J} \approx 0.49$). Having observed that some special issues have no call-paper overlap under strict coding and very low ones under lax coding, we can remove them from the computation and focus on 18 conference editions for SLE 2008–2025, which brings Jaccard indices to $\overline{J} \approx 0.36$ for strict coding and $\overline{J} \approx 0.56$ for lax coding. Technically this indicates that about half of the union of requested and realised topics overlaps under the chosen coding view.

The strict view is constrained by design: each paper contributes only one primary topic, so $|P_y|$ is naturally much smaller than $|R_y|$: $2 \leqslant |P_y| \leqslant 11$ for special issues and $16 \leqslant |P_y| \leqslant 26$ for regular proceedings, while $13 \leqslant |R_y| \leqslant 31$ with a notable exception of one special issue with $R_{2012S} = 4$. Per year, CfPs request on average $\overline{|R_y|} \approx 22.72$ topics ($\overline{|R_y|} \approx 23.72$ without special issues), while strict proceedings cover only $\overline{|P_y|} \approx 10.32$ primary topics ($\overline{|P_y|} \approx 12.44$ without special issues). Nevertheless, the intersection still contains about $|R_y \cap P_y| \approx 7.64$ topics on average ($|R_y \cap P_y| \approx 9.5$ without special issues). Under lax (multi-topic) coding, $|P_y|$ matches CfP breadth ($\overline{|P_y|} \approx 20.52$ or $\overline{|P_y|} \approx 23.61$ with and without special issues, respectively), and the intersection grows to $|R_y \cap P_y| \approx 14.2$ topics on average ($|R_y \cap P_y| \approx 16.89$ without special issues).

This gap between strict and lax alignment is itself a significant finding: *many* topics that are "present" in proceedings are rarely the *primary* theme of a paper, yet they occur persistently as knowledge elements (methods, validation angles, contexts) that travel with other contributions. In other words, CfPs and proceedings align better when SLE is viewed as a *body of knowledge* (multi-dimensional contributions) rather than as a collection of single-topic papers.

6.2 Systematic Mismatches in Topical Prominence

Set overlap does not reveal *how much* a topic is emphasised. To identify systematic prominence mismatches, we compare (i) CfP emphasis via total topic mentions in CfPs over time, and (ii) proceedings emphasis via the number of papers tagged with that topic (strict: as primary; lax: as any tag). Two stable patterns emerge.

Topics over-emphasised in calls (requested more than they appear). Even under lax coding, several topics are consistently requested yet appear in relatively few proceedings years. Table 8 (left) lists the largest such mismatches by "years requested" vs. "years observed in proceedings". To select those systematically, we use normalised difference:

$$\delta_t = \frac{|R_t| - |P_t|}{|R_t| + |P_t|}$$

Simply looking at the difference between $|R_t|$ and $|P_t|$ was insufficient since for instance the difference between five years requested and zero (five fewer) published for T6E Quantum is drastic in practice, while the same difference of five for T2B Evolution which happened to be lacking from five proceedings volumes, but in general still accumulated 50+ papers over 19 bundles, is almost entirely meaningless.

The strongest examples are T4E Simulation, T2D Deployment and T6A Experience Reports: all three are repeatedly requested across the years, yet they occur only sporadically in proceedings taggings. Finally, T6C SLE-for-AI and T6E Quantum are prototypical "aspirational" topics: introduced in recent CfPs but not (yet) realised in proceedings up to 2025 (T6C was used to tag one of the SLE 2023 keynotes and T6E was not used even once at all).

Topics under-emphasised in calls (published more than explicitly requested). Conversely, several topics are

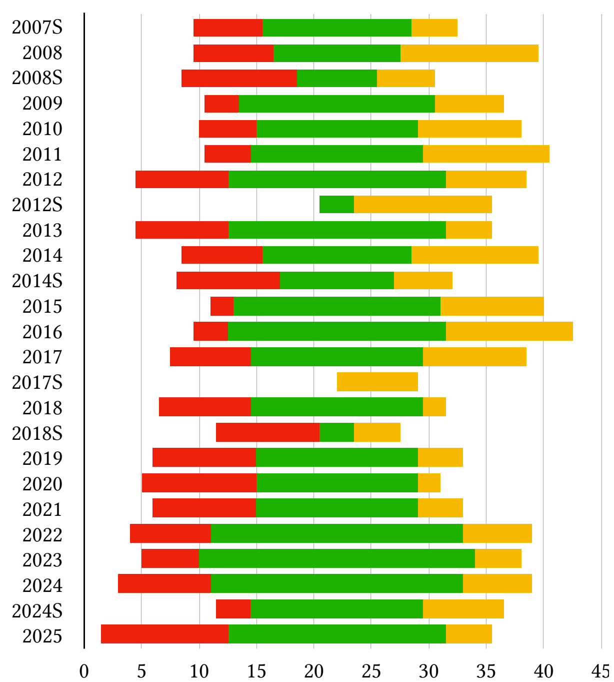


Figure 4. Call–paper alignment over the bundles: green shows overlap, red colour designates topics requested by the CfP but not delivered by any accepted paper even as a secondary tag, and golden colour shows topics covered by papers of that year but not the corresponding CfP. In other words, red+green are CfP topics, green+golden are proceedings topics.

published broadly while being mentioned rarely (or late) in calls, which we collect on the right side of Table 8. The clearest example is T4D Interpretation: it is explicitly requested in only one conference call (SLE 2015) and was marked in three more special issues (as "execution platforms" in 2008, "selected at runtime" in 2014 and "platform targeting" in 2024), yet appears in almost all proceedings years under lax coding. When this tag was created, we expected to find more mentions of runtimes in the CfPs. The same pattern holds for T4F Visualisation and (to a lesser extent) T5H Usability: these appear as recurring knowledge elements in papers even when they are not foregrounded in CfP topic lists. Finally, T4G AI-for-SLE demonstrates that papers can lead calls: it appears in proceedings before it becomes an explicit CfP focus (see below).

Two interpretive points matter for a SLEBoK framing. First, many "under-emphasised" topics are not *missing* from the CfP because they are *out of scope*, but because they are treated as *implicit scope*. These might be assumed background methods or cross-cutting concerns such as interpretation as a software language implementation technique, or visualisation as part of tool support. Second, many "over-emphasised" topics appear to be *desired capacity* rather than *realised centre of gravity*. This is especially true for topics that require sustained infrastructure, long-lived industrial collaborations, or shared benchmarks (e.g., deployment at scale, simulation as a distinct research theme, or applying language engineering to quantum computing).

6.3 Responsiveness and Lag

Do calls actually lead or follow? CfPs may either *steer* the community (topics appear after being highlighted) or *follow* the community (topics are added after papers already appear). We find evidence of both dynamics.

Calls remaining unanswered (infinite/unknown positive lag). Some topics are requested but remain unrealised in the observable data up till 2025, illustrating that "call steering" is not guaranteed even when a topic is made explicit. For example, T6C SLE-for-AI is introduced explicitly in the 2022 CfP, it appears in proceedings only once as a keynote of SLE 2023, yielding an approximately one-year lag from call emphasis to proceedings uptake. Similarly, T6E Quantum remains an unrealised dream also since SLE 2022.

Calls steering proceedings (positive lag). In a less drastic way, there are several topics with an uptake time of several years. For instance, T4E Simulation was introduced to CfPs at SLE 2012 which tangentially mentioned Simulink, and was present in most years in some way. However, the first simulation paper on micro-machinations saw light in SLE 2013, the second one was a tool demo of GEMOC at SLE 2016, and the third and last at this moment was at SLE 2025 on live model checking. At the same time, T4E Simulation appeared in all CfPs of regular conferences in SLE 2013–2026, as well as in

T4A Workbenches	and	T1D Composition	49
T4A Workbenches	and	T3C DSLs	44
T4A Workbenches	and	T3A Meta-languages	39
T4A Workbenches	and	T5H Usability	38
T4A Workbenches	and	T5D Formal Methods	30
T1D Composition	and	T3C DSLs	27
T4A Workbenches	and	T4C Vertical Transformation	26
T3B Transformation Languages	and	T4B Horizontal Transformation	24
T4A Workbenches	and	T5E Empirical Evaluation	23
T4A Workbenches	and	T4D Interpretation	22

Table 7. Top 10 of most popular topic co-occurrences

some special issues. A bit more luck followed topics such as T2D Deployment which after the initial uptake became a common guest in SLE proceedings.

Calls giving up on steering (positive lag followed by retirement). T2A Lifecycle was featured prominently in the CfPs since the pre-SLE special issue of 2007, but disappeared since SLE 2013 (editorials of special issues of 2014 and 2018 have some vague reference to it, but not regular CfPs). Yet, there were papers on T2A Lifecycle at SLE 2014 (island parsing), 2016 (legacy), 2019 (keynote) and 2020 (legacy). Similarly, T2E Documentation was unanswered for two regular calls and two special issue calls, until we have seen a language documentation paper in 2010; SLE 2014 removed the topic from its CfP, but SLE 2017 had another unsolicited paper on T2E Documentation; after a long pause, more followed in SLE 2024–2025 and the special issue of SLE 2024. With somewhat varying prevalence, the same happened to T3D GPLs, T3E Ontologies and T3F Interfaces.

Proceedings leading calls (negative lag). Conversely, T4G AI-for-SLE demonstrates that proceedings can precede CfP emphasis. This topic appears in proceedings as early as 2021–2024 with one or two papers per year tagged under lax coding, while it is only explicitly requested in CfPs starting in 2025. Under our operationalisation, ML-based methods also count as AI. Notably, the first year in which the CfP strongly emphasises AI-for-SLE (2025; multiple mentions) does *not* coincide with an immediate increase in 2025 proceedings (no T4G-tagged papers in our 2025 proceedings data), which is consistent with the practical timeline. CfP framing and community output can be misaligned within the same year due to submission cycles, review selection, and the time required for new work to mature.

Lagged association between CfP emphasis and proceedings counts. Beyond discrete "new topic" events, it is also interesting to see whether fluctuations in CfP emphasis correlate with later fluctuations in proceedings presence. Across topics with sufficient variation (i.e., not always-on or always-off), we observe that the strongest positive associations often occur with a one- to two-year lag. For example, T5B Language Description Analysis exhibits a notably strong lagged association (best at around two years), and T3C DSLs shows a clearer association at about one year than at zero lag. These patterns are consistent with the idea that CfPs can influence the medium-term composition of submissions (and hence proceedings), while also reflecting that some topics require longer lead time to turn into publishable contributions. At the same time, we treat these lagged correlations as *indicative* rather than causal: the SLE community is small and interconnected, CfPs evolve based on programme committees and broader trends, and proceedings reflect both supply (submissions) and demand (selection).

6.4 SLE Core at Other Venues

As a lightweight sanity check, we also asked whether the SLE core identified above is merely generic PL/MDE vocabulary. We therefore sampled the last five years of regular proceedings from PLDI, MoDELS, and CC, obtained title/abstract pairs via OpenAlex and CrossRef, and applied an automatically generated keyword matcher derived from the topic definitions in section 4. This procedure is intentionally weaker than the manual tagging used for SLE, so we treat it only as indicative.

The biggest observable difference was when we looked at the **core** topics derived from the CfPs of SLE. Those topics cover **39%** of all SLE papers but only **32%** of MoDELS papers, **25%** of PLDI papers and **15%** of CC papers. The difference persisted in **near-core** (plus core) topics, with only 2%–5% rise for each conference, but almost disappeared for **fringe core** (plus near-core plus core) where (manually tagged) SLE papers scored **61%** while (approximately auto-tagged) CC, PLDI and MoDELS papers were in the **52%**–**56%** range. For the SLE proceedings derived cores, the difference among the conferences were not that noticeable (within 3%–6% for **core**), but the gap increased with each step, reaching 11%–17% for **near-core** and 13%–20% for **fringe core** coverage.

If we take the "SLE core" in a broad sense as a union of *all* topics from Table 5, then those core topics would collectively

Requested more than published					Published more than requested				
Topic	CfP years	Proc. years	CfP taggings	Papers	Topic	CfP years	Proc. years	CfP taggings	Papers
T6E Quantum	5	0	5	0	T4D Interpretation	4	20	4	56
T4E Simulation	17	3	19	3	T4F Visualisation	3	15	3	24
T6C SLE-for-AI	5	1	15	1	T4G AI-for-SLE	2	8	12	10
T2D Deployment	20	7	23	9	T3F Interfaces	6	9	10	16
T6A Experience Reports	17	7	23	9	T5B Lang. Desc. Analysis	12	17	20	48

Table 8. Largest call–paper mismatches under lax (multi-topic) coding, measured as years requested in CfPs/editorials vs. years observed in proceedings and special issues. We select based on the normalised difference for $\delta_t \geqslant 0.4$ on the left and $\delta_t \leqslant -0.17$ on the right.

cover **73%** of all SLE papers and about **61%** of all papers of MoDELS, PLDI, and CC. However, if we were to base any conclusion of this data (for instance, if this little experiment can be redone in the future with a better degree of reliability), it could serve as a reason to exclude CfP-tagged **fringe core** topics from such a unified SLE core set, but include proceedings-derived **fringe core** topics.

We conclude that SLE-core topics are certainly visible elsewhere, but their concentration and organisation differ. A proper finding-level comparison would require applying the same call–paper methodology to those venues, which we leave for future work.

6.5 Concluding remarks

Overall, SLE's calls and proceedings show moderate-to-strong topical alignment when we treat topics as BoK knowledge elements (multi-topic coding), but only moderate alignment under a single-topic view. Systematic mismatches are informative rather than pathological: they reveal (i) which topics are "implicitly assumed" in the community's output despite being rarely requested, (ii) which topics are persistently desired but remain under-realised, and (iii) when CfPs appear to follow the community versus when they appear to shape it over time.

Jaccard indices hover around 0.5 occasionally reaching 0.6 or higher (0.73 for SLE 2023) and have an increasing trend over the conference years, but this increase can hardly be attributed to executive decisions like joining SIGPLAN or co-locating with a particular conference. Jaccard indices of special issues vary wildly from 0.0 to 0.6, which is probably explainable with a paradigm choice: if the chairs decide to use the editorial as a platform to signal what SLE is about, it gets lower; if they stick to describing precisely what are the topics covered by the papers included in the special issue, it gets higher.

7 Implications for SLEBoK and future directions

The fourth research question from section 2 asks what can reasonably be concluded from this inventory for the future consolidation of SLE as a body of knowledge. In this paper, we use the term *curation* in a deliberately non-normative sense. We do *not* mean policing the community, filtering out topics, or prescribing what SLE ought to be. Instead, curation means maintaining a usable, evidence-based description of what the community has already built, how that knowledge is structured, and where the corpus suggests stable strengths, recurring absences, and emerging directions. In that sense, curation is closer to documentation and synthesis than to governance.

Three implications follow from our results. **First**, SLE exhibits a stable identity core that is visible in both calls and papers: language design, modularity, evolution, DSLs, workbenches, transformations, and formal methods recur often enough to form a defensible backbone of the field. **Second**, many knowledge elements function as *cross-cutting* concerns rather than stand-alone paper topics: under a multi-topic view, proceedings align with calls much better than under strict single-label coding, which suggests that SLE knowledge is accumulated through combinations of methods, artefacts, and validation lenses rather than through sharply separated subfields. **Third**, recent CfPs broaden the perimeter of SLE by explicitly naming AI-related, socio-technical, and cyberphysical contexts, but that broadening is only partially reflected in proceedings so far.

For SLEBoK, this does not imply that **core** (persistent) topics are more legitimate than others, nor that non-core topics should be filtered out or neglected. The distinction is a maintenance device: SLEBoK was intended to organise concepts, literature, tools, use cases, and teaching material for the field [Combemale et al. 2018]; our tags are not proposed as a replacement for that structure, but as an empirical index that can help decide how entries should be curated. Topics with persistent presence in both CfPs and proceedings are good candidates for mature entries with definitions, canonical literature, tools, and teaching pointers. Cross-cutting topics that frequently co-occur with others, such as interpretation, visualisation, or usability, are better represented through links across entries rather than as isolated silos. Frontier topics, such as T6E Quantum or some AI/application-context topics, should be marked as emerging or open-challenge areas, where SLEBoK can collect early examples without presenting them as already consolidated practice.

The co-occurrence analysis reinforces this maintenance-oriented reading. Frequent combinations such as workbenches with modularity, DSLs, formal methods, usability, or compilation, indicate that SLE contributions are often built as pipelines rather than as single-method results. Accordingly, SLEBoK entries would benefit from recording not only topic-specific literature, but also typical neighbouring topics and common integration paths. This would make the BoK more faithful to the structure of the actual corpus.

The same inventory also suggests practical implications for future CfPs. The long-term growth of topic lists has improved inclusiveness, but it also makes the call less informative as a signal of priorities. A better balance would be to keep a stable background scope while reserving explicit emphasis for a small number of highlighted topics each year. At the same time, topics that are consistently realised in papers but rarely named in CfPs, such as runtime libraries or visualisation, may deserve more explicit mention, especially for the benefit of newcomers. Conversely, topics that remain under-realised despite repeated mention, such as simulation, deployment, or experience reports, are unlikely to be strengthened by wording changes alone; they may need dedicated contribution formats, or even invited SLEBoK-style synthesis papers that lower the barrier to entry. Clearer evaluation criteria may also help, although our evidence is survival-biased by being necessarily limited to accepted papers; the topic distribution of rejected submissions may differ.

Finally, this paper also points to a modest follow-up agenda. The most immediate next step is robustness: the topic taxonomy from section 4 appears adequate when examined from several angles, but the reliability of its *application* should be stress-tested through double-coding and adjudication. A second next step is scope: some call–paper gaps may indicate not lack of activity, but activity displaced (misdirected?) to neighbouring venues. The apparent fade-out of *T3E Ontologies* is a concrete example where venue migration is at least as plausible as topical disappearance. A third next step is community dynamics: the present paper deliberately treats SLE impersonally, but author flows, recurring topic carriers, and venue migration could explain part of the temporal behaviour observed here. Taken together, these directions suggest that the most useful future role of this paper is not to settle what SLE is, but to provide a reproducible baseline from which SLEBoK can be maintained as a living, evidence-based inventory of the field.

The main practical lesson is that SLEBoK should not be organised as a flat topic list: the evidence supports a layered view in which persistent topics provide stable entries, cross-cutting topics provide links between entries, and frontier topics are documented as emerging/open-challenge areas. This layered view also makes the non-normative role of the inventory explicit: it helps decide how knowledge should be documented, cross-referenced, or marked as emerging, without deciding which topics the community should pursue.

8 Threats to validity

Construct validity. Our taxonomy aims to capture *topics*, but topic assignment is inevitably mediated by surface evidence: CfP bullet items are short by design, and some proceedings entries like keynotes offer only titles and abstracts. This creates a risk that tags reflect the wording of the corpus rather than the full substance of the underlying work. We have mitigated this in two ways. First, the taxonomy was iteratively derived from both CfPs and papers, which reduces the risk of encoding only one side of the comparison. Throughout the paper, we kept investigating both sides of the tagging. Second, we have maintained both a strict and a lax view of proceedings tagging; the gap between them is itself informative, because many SLE contributions are genuinely multi-topic and/or inter-disciplinary.

Internal validity. Tagging is interpretive, especially for boundary cases such as semantics versus runtime, transformation languages versus transformations done within a workbench, or empirical evaluation versus usability. The codebook in section 4 makes these boundaries explicit, but some judgement calls remain unavoidable. Comparability across years is also imperfect because CfP phrasing tends to change over time, and special issues are too heterogeneous: some provide actual calls, while others only editorials written *after* paper selection. A further consequence is that special issues can over-represent topics already present in the corresponding conference; this is why we separate regular conference-only summaries from all-bundle summaries whenever the distinction could affect the interpretation. Special issues are still useful as supplementary evidence and as a part of supporting statistics.

External validity. This paper deliberately studied SLE as a venue, not all SLE-related research wherever it appears, even though it is known that essential SLE work is also published at venues of higher rank and broader scope. Some topics that are rare in SLE proceedings may also be better represented in neighbouring communities, and some parts of the broader SLE knowledge base are not captured here at all, including books, workshops, tool repositories, non-publishing events with prominent industrial presence, and internal technical reports of community practice. Similarly, our archives of special issues and calls are extensive but may still be incomplete. The results should therefore be read as an evidence-based inventory of the published SLE record, not as an exhaustive description of all *SLE knowledge* circulating in the community.

Metric choice. Our main alignment metric is the Jaccard index, which is transparent and easy to interpret. However, it ignores topic intensity and treats all present topics equally. Likewise, prevalence and diversity measures are descriptive rather than causal. They are appropriate for the inventory goal of this paper, but they do not by themselves explain why topic shifts occur and do not identify causes.

9 Related work

This paper contributes to venue-level meta-analysis in software engineering and programming languages research. There are many comparable research initiatives aimed at questioning community health of conferences [Vasilescu et al. 2014], their diversity [Narayanan et al. 2023], engaging in the journal vs conferences debate [Briand 2024], as well as simply seeking insights and opportunities [Zaytsev 2017]. Methodologically, it is also close to work that studies scientific corpora through topic discovery and temporal change, including probabilistic topic models such as LDA [Blei et al. 2003; Griffiths and Steyvers 2004] and more recent embedding-based approaches such as BERTopic [Grootendorst 2022]. Unlike unsupervised topic models, our main goal is not to discover latent topics automatically, but to construct a stable, interpretable taxonomy that can be applied symmetrically to both calls and papers and reused as SLEBoK-oriented annotation guidance.

Even though these results are related in motivation, unfortunately most are not directly comparable at the level of our main findings. Studies of conference health, diversity, or publication models typically analyse authorship, participation, acceptance, or publication practices, whereas our central object is the relation between a venue's self-description (CfPs) and its realised topical output (papers). As a result, they do not usually report quantities analogous to our stable topical core, call–paper Jaccard overlap, or requested-only/received-only topic gaps. The closest comparison point within this paper is therefore the lightweight experiment in subsection 6.4. SLE-core topics are also visible in CC, PLDI, and MoDELS, but the concentration and organisation differ, suggesting that the identified SLE core is something more than just a generic PL or MDE profile. A full finding-level comparison would require applying the same call–paper methodology to other venues, which, as any significant extension, we leave as future work.

The paper also relates to literature on annotation reliability and corpus coding methodology. In particular, the interpretive nature of multi-label tagging makes inter-annotator agreement an important future step, and standard guidance from annotation research applies directly here [Artstein and Poesio 2008]. More broadly, our work is positioned as a descriptive inventory and science-mapping exercise for one research venue, with the distinctive feature that it analyses not only the published output but also the venue's self-description through its CfPs.

There are many books [Bettini 2016; Combemale et al. 2017; Dearle 2010; Fowler 2011; Hölldobler et al. 2021; Kleppe 2008; Lämmel 2018; Voelter et al. 2013; Wasowski and Berger 2023] and seminar reports [Combemale et al. 2018] with very condensed SLE knowledge, which were left out of our scope for this project despite missing out on the obvious dual benefit. On one hand, these books also contribute to the body of knowledge of software language engineering, and some have even been presented at SLE in some form. On the other hand, they usually engage actively with existing SLE material, which can be useful when polishing the taxonomy or looking for SLE work in unlikely venues. It was, however, unrealistically overambitious to cover both SLE papers and SLE books in the same endeavour, so we left this part for future work.

10 Conclusion

In this paper, we presented a comparative, tag-based meta-analysis of 19 SLE CfPs, 18 years of SLE proceedings, and associated special-issue material, all in order to inventory SLE as an evolving body of knowledge. The main picture is consistent across the analyses: SLE has a stable topical core, its proceedings are broader and more BoK-like under multi-topic coding than under a single-label view, and its calls have recently expanded faster than its realised proceedings output. This makes the conference simultaneously stable in identity and elastic at its perimeter.

The contribution of the paper is not to prescribe what SLE should be or become, but to make its accumulated record more legible: which topics are persistent, which are cross-cutting, which remain aspirational, and where calls and papers diverge, realign over time, or stay in disagreement. That inventory is meant to be directly useful for SLEBoK, for future CfP design, and for follow-up studies on robustness, neighbouring venues, and community dynamics.

At the same time, the paper captures only the explicit and published trace of the community. Like any venue-level inventory, it misses tacit knowledge, informal influence, tool use that was never written up, and the broader "you had to be there" dimension of a long-running research community. The ultimate goal therefore should be not only to extend the dataset, but also to turn it into a lightweight, maintainable SLEBoK pipeline for periodically refreshing the evidence base on which such inventories rest. The resulting artefact is therefore not merely supplementary material for this paper, but a starting point for maintaining the inventory as the conference continues to evolve.

The project website, https://slebok.github.io/cfpbok, is intended to remain online as living documentation, to be updated whenever new information surfaces in the future. A snapshot synchronised with the final version of this paper, is also preserved as a Zenodo artefact [Zaytsev 2026].

References

Ron Artstein and Massimo Poesio. 2008. Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics* 34, 4 (2008), 555–596. doi:10.1162/coli.07-034-r2

Lorenzo Bettini. 2016. *Implementing Domain-Specific Languages with Xtext and Xtend* (2nd ed.). Packt Publishing.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3 (2003), 993–1022. doi:10.1162/jmlr.2003.3.4-5.993

Lionel Briand. 2024. Conferences in Software Engineering: Reflections after 30 Years. ACM SIGSOFT Blog, https://sigsoft.medium.com/conferences-in-software-engineering-reflections-after-30-years-ca525d98c584.

Lola Burgueño, Walter Cazzola, and Dimitris Kolovos. 2025. Editorial to the Special Issue on Software Language Engineering. *Journal of Systems and Software* 230 (2025), 112498. doi:10.1016/j.jss.2025.112498

Benoît Combemale, Robert W. France, Jean-Marc Jézéquel, Bernhard Rumpe, Jim Steel, and Didier Vojtisek. 2017. *Engineering Modeling Languages: Turning Domain Knowledge into Tools*. CRC Press. doi:10.1201/b21841

Benoît Combemale, Ralf Lämmel, and Eric Van Wyk. 2018. SLEBOK: The Software Language Engineering Body of Knowledge (Dagstuhl Seminar 17342). *Dagstuhl Reports* 7, 8 (2018), 45–54. doi:10.4230/DagRep.7.8.45

Benoit Combemale, David J. Pearce, Richard F. Paige, and Eric Van Wyk. 2015. Special Issue on the 6th and 7th International Conferences on Software Language Engineering (SLE 2013 and SLE 2014). *Computer Languages, Systems & Structures* 44 (2015), 1–2. doi:10.1016/j.cl.2015.08.009 Special issue on the 6th and 7th International Conference on Software Language Engineering (SLE 2013 and SLE 2014).

Krzysztof Czarnecki and Görel Hedin. 2014. Software Language Engineering (SLE'12). *Science of Computer Programming* 96 (2014), 395. doi:10.1016/j.scico.2014.09.004 Selected Papers from the Fifth International Conference on Software Language Engineering (SLE 2012).

Fergal Dearle. 2010. *Groovy for Domain-Specific Languages* (1st ed.). Packt Publishing.

J.-M. Favre, D. Gašević, R. Lämmel, and A. Winter. 2007. Call-for-Papers: Special Issue on Software Language Engineering. *IEEE Transactions on Software Engineering* 33, 12 (2007), 891–891. doi:10.1109/TSE.2007.70760

J.-M. Favre, D. Gašević, R. Lämmel, and A. Winter. 2008. Software Language Engineering. *IET Software* 2 (2008), 161–164. Issue 3. doi:10.1049/iet-sen:20089010

Martin Fowler. 2011. *Domain-Specific Languages*. Addison-Wesley. http://vig.pearsoned.com/store/product/1,1207,store-12521_isbn-0321712943,00.html

Thomas L. Griffiths and Mark Steyvers. 2004. Finding Scientific Topics. *Proceedings of the National Academy of Sciences* 101, Suppl. 1 (2004), 5228–5235. doi:10.1073/pnas.0307752101

Maarten Grootendorst. 2022. BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure. *arXiv preprint arXiv:2203.05794* (2022). doi:10.48550/arXiv.2203.05794

Conner Hansen, Eugene Syriani, and Levi Lucio. 2015. Towards Controlling Refinements of Statecharts. arXiv:1503.07266 [cs.SE] In Poster Proceedings of Sixth Conference on Software Language Engineering (SLE) 2013.

Katrin Hölldobler, Oliver Kautz, and Bernhard Rumpe. 2021. *MontiCore Language Workbench and Library Handbook: Edition 2021*. Aachener Informatik-Berichte, Software Engineering, Vol. 48. Shaker. http://www.monticore.de/handbook.pdf

Anneke Kleppe. 2008. *Software Language Engineering: Creating Domain-Specific Languages Using Metamodels* (1 ed.). Addison-Wesley Professional.

Paul Klint, Ralf Lämmel, and Chris Verhoef. 2005. Toward an Engineering Discipline for Grammarware. *ACM Transactions on Software Engineering Methodology (ToSEM)* 14, 3 (2005), 331–380. doi:10.1145/1072997.1073000

Ralf Lämmel. 2018. *Software Languages: Syntax, Semantics, and Metaprogramming*. Springer. doi:10.1007/978-3-319-90800-7

Peter D. Mosses. 2017. Engineering Meta-Languages for Specifying Software Languages. In *Proceedings of the 10th ACM SIGPLAN International Conference on Software Language Engineering* (Vancouver, BC, Canada) *(SLE 2017)*. Association for Computing Machinery, New York, NY, USA, 1. doi:10.1145/3136014.3148041

Peter D. Mosses. 2019. Software Meta-Language Engineering and CBS. *Journal of Computer Languages* 50 (2019), 39–48. doi:10.1016/J.JVLC.2018.11.003

Aditya Shankar Narayanan, Dheeraj Vagavolu, Nancy A Day, and Meiyappan Nagappan. 2023. Diversity in Software Engineering Conferences and Journals. arXiv:2310.16132 [cs.SE] https://arxiv.org/abs/2310.16132

David Pearce and Ulrik Pagh Schultz. 2020. Editorial for GPCE&SLE2018 Special Issue. *Journal of Computer Languages* 59 (2020), 100988. doi:10.1016/j.cola.2020.100988

SLEBoK. 2013. Software Language Engineering Body of Knowledge. https://slebok.github.io.

Bogdan Vasilescu, Alexander Serebrenik, Tom Mens, Mark G. J. van den Brand, and Ekaterina Pek. 2014. How healthy are software engineering conferences? *Science of Computer Programming* 89 (2014), 251–272. doi:10.1016/J.SCICO.2014.01.016

Markus Voelter, Sebastian Benz, Christian Dietrich, Birgit Engelmann, Mats Helander, Lennart C. L. Kats, Eelco Visser, and Guido Wachsmuth. 2013. *DSL Engineering — Designing, Implementing and Using Domain-Specific Languages*. dslbook.org. http://www.dslbook.org

Andrzej Wasowski and Thorsten Berger. 2023. *Domain-Specific Languages — Effective Modeling, Automation, and Reuse*. Springer. doi:10.1007/978-3-031-23669-3

Vadim Zaytsev. 2015. Guided Grammar Convergence. arXiv:1503.08476 [cs.SE] In Poster Proceedings of Sixth Conference on Software Language Engineering (SLE) 2013.

Vadim Zaytsev. 2017. BibSLEIGH: Bibliography of Software (Language) Engineering in Generated Hypertext. In *Post-proceedings of the Eighth Seminar in Series on Advanced Techniques and Tools for Software Evolution (SATToSE 2015) (CEUR Workshop Proceedings, Vol. 1820)*, Anya Helene Bagge, Tom Mens, and Haidar Osman (Eds.). CEUR-WS.org, 54–64. http://ceur-ws.org/Vol-1820/paper-06.pdf

Vadim Zaytsev. 2026. CfPBoK. Zenodo. doi:10.5281/zenodo.19917567