Software and Systems Modeling
https://doi.org/10.1007/s10270-025-01302-4

SPECIAL SECTION PAPER

Exploring modeling methods for information systems analysis and design: a data-driven retrospective

Iris Reinhartz-Berger[1] · Adir Solomon[1] · Jelena Zdravkovic[2] · John Krogstie[3] · Henderik A. Proper[4]

Received: 4 February 2025 / Revised: 26 April 2025 / Accepted: 22 May 2025

Abstract
Modeling for information systems (IS) analysis and design offers broad insights into the advances and challenges of enterprise, business process, software, and conceptual modeling. In celebration of its 30th edition, this paper presents a data-driven retrospective analysis of studies published at the *Exploring Modeling Methods for Systems Analysis and Development (EMMSAD)* working conference from 2005 to 2024. EMMSAD has long been a key venue for research on Information Systems (IS) Modeling, covering areas such as conceptual modeling, enterprise modeling, and model-driven engineering, as well as the evaluation of modeling techniques and tools. Using machine learning, specifically Dynamic Topic Modeling (DTM) with BERTopic, this study identifies recurring topics, emerging trends, and shifts in research focus within the IS modeling community. The findings highlight key areas of alignment between IS modeling and the broader modeling landscape, providing insights into the field's evolution and future research opportunities.

Keywords IS analysis and design · Dynamic Topic Modeling · BERTopic · Data-driven approach · EMMSAD

1 Introduction

Information Systems (IS) Modeling has long played a key role in IS analysis and design, providing structured approaches for understanding organizational contexts, analyzing business processes, and developing effective systems [27, 33]. By bridging theoretical contributions with practical implementation, modeling enhances organizational adaptability to technological change and emerging requirements. Over the years, the field has witnessed significant advancements, driven by innovations in modeling methodologies [14, 30], the integration of artificial intelligence and machine learning techniques into modeling processes [21], and the growing emphasis on developing responsible information systems that address concerns such as sustainability [7].

The *Exploring Modeling Methods for Systems Analysis and Development (EMMSAD)* working conference has become a prominent venue for the exchange of ideas and the advancement of the field. By fostering collaboration among researchers and practitioners, EMMSAD has consistently highlighted emerging trends, novel approaches, and practical applications in IS modeling.

Toward its 30th anniversary, this paper presents the design and findings of a systematic, data-driven retrospective analysis of the last two decades of research published at EMMSAD. This approach enables an objective and scalable examination of thematic developments over time, reducing potential biases inherent in expert-driven top-down analyses. By exploring recurring topics and their evolution over time, the study offers valuable insights into the historical development of IS modeling. This retrospective spans the period from 2005 to 2024, during which EMMSAD proceedings

Communicated by Dominik Bork and Arnon Sturm.

✉ Iris Reinhartz-Berger
iris@is.haifa.ac.il

Adir Solomon
asolomon@is.haifa.ac.il

Jelena Zdravkovic
jelenaz@dsv.su.se

John Krogstie
john.krogstie@ntnu.no

Henderik A. Proper
henderik.proper@tuwien.ac.at

1 University of Haifa, Haifa, Israel

2 Stockholm University, Stockholm, Sweden

3 Norwegian University of Science and Technology (NTNU), Trondheim, Norway

4 TU Wien, Vienna, Austria

Published online: 27 June 2025

were officially published with CEUR and subsequently with Springer. Over this period, a total of 289 papers were presented at the conference, reflecting the diverse and evolving interests of the IS modeling community.

Using advanced analytical techniques, particularly *Dynamic Topic Modeling (DTM)* [5] with *BERTopic* [16], the study identifies emerging topics and shifts in research focus. These trends are contextualized within the broader landscape of conceptual, software, and business process modeling. This multifaceted analysis aims to provide a comprehensive view of the field's progression. Specifically, the study addresses the following research questions:

Research Questions

RQ1: What are the recurring topics addressed in IS modeling, as reflected in EMMSAD papers?
RQ2: How have the topics evolved over the past two decades?
RQ3: How does the topical focus of IS modeling align with and differ from other related modeling communities?

The remainder of this paper is organized as follows. Section 2 provides a high-level overview of IS modeling, tracing its historical foundations and key milestones, while also highlighting the development of EMMSAD throughout this timeline. Section 3 outlines the research methodology, including data collection and analysis methods, while Sects. 4, 5, and 6 present the results addressing RQ1, RQ2, and RQ3 , respectively. Finally, Sect. 7 concludes with key findings and directions for future research.

2 Background

Modeling has a long history, which can be traced back to at least the first half of the twentieth century [24, 32]. Nevertheless, the foundations of modern (visual) IS modeling began to take shape in the 1960 s. This period witnessed significant theoretical advancements, notably the introduction of various levels of abstraction for information systems, as described in the book *"Theoretical Analysis of Information Systems"* [8, 20]. In this book, Langefors proposed a set of concepts central to modeling of IS, including the partitioning of the development life cycle into four important *method areas*:

1. Methods for the management and control of organizations,
2. Methods for the analysis and description of IS at an elementary, "problem-oriented" level (the "infological" area),
3. Methods for the design and analysis of computerized IS at "product-oriented" level (the "datalogical" area), and
4. Methods for the implementation of IS on computer hardware.

During the 1970 s, the two middle levels—analysis and design—were supported with visual notations such as Data Flow Diagrams (DFD) [15] and Entity-Relationship (ER) [10] diagrams. These techniques became central to the agenda of the IFIP WG 8.1—Design and Evaluation of Information Systems, established under the Technical Committee 8 (TC 8)—Information Systems of the International Federation for Information Processing (IFIP). In 1974, TC 8 was approved, and shortly afterward, the first two working groups were established: WG 8.1, focusing on the design and evaluation of IS, and WG 8.2, exploring the interaction between IS and organizations.

The 1980 s saw a proliferation of modeling notations alongside efforts to evaluate their quality, with particular emphasis on expressiveness. This period also laid the groundwork for further advancements in modeling methodologies. Based on contributions from central figures within IFIP WG 8.1, the CAiSE (Conference on Advanced Information Systems Engineering) conference was established in 1989, with its first edition held in Stockholm, Sweden.

In the 1990 s, modeling approaches expanded beyond behavioral, functional, and structural techniques to incorporate object-oriented analysis and design (OOAD), language-action approaches based on speech act theory, and goal-oriented methods. While the early 1990 s saw over 50 methods in OOAD, this diversity eventually converged into a standardized approach with the development of the Unified Modeling Language (UML) [6]. The decade also witnessed the emergence of Domain-Specific Modeling (DSM) tools supporting meta-modeling and method engineering [18, 19].

As the CAiSE conference matured and grew, it began accommodating workshops that targeted specialized research communities. REFSQ (Requirement Engineering: Foundation for Software Quality), which later evolved into a standalone conference, started as one of the first CAiSE's workshops in 1995. Similarly, the EMMSAD workshop was established in 1996, anchored in the core activities of IFIP WG8.1. A few years later, the Business Process Modeling, Development, and Support (BPMDS)[1] workshop, which shares some thematic overlap with EMMSAD, was established. Over time, both EMMSAD and BPMDS evolved into annual working conferences affiliated with IFIP WG8.1.

Entering the new millennium, IS modeling witnessed further standardization, for instance, Model-Driven Architecture (MDA) [23] for (UML-based) model-driven engineering, ArchiMate [1, 2] for enterprise (architecture) modeling,

[1] https://sites.google.com/view/bpmds/

and BPMN [31] for business process modeling. Over the last two decades, these advancements have influenced research presented at the EMMSAD working conference. This paper provides a comprehensive quantitative analysis of EMMSAD papers published since 2005, shedding light on the evolution of topics and trends in IS modeling during this period.

3 Research methodology

Figure 1 presents our underlying research methodology. We started with paper collection and preprocessing, where the text of each paper was extracted and cleaned. Next, BERTopic [16] was applied to generate a latent representation for each paper and perform Dynamic Topic Modeling [5]. This enabled us to identify and analyze temporal trends within the detected topics. Finally, we compared our results with recent findings from related conferences, providing key insights and highlighting potential future directions for IS modeling.

Below we elaborate on the data collection and preprocessing, topics and trends identification, and expert evaluation, as well as threats to validity.

3.1 Data collection and preprocessing

The dataset comprises a corpus of PDF files including all EMMSAD papers from the years 2005-2024, with each paper represented by a single PDF file. As an initial step, we created a PowerShell script to convert the PDF files into text files. During the conversion process, the script also removed metadata, including author names, institutional affiliations, and bibliographic references, to focus solely on the main content.

Each text file underwent a systematic preprocessing procedure. First, all letters were converted to lowercase to ensure uniformity and minimize variability in token representation. Next, common English stop-words were filtered using the NLTK library [4]. To further refine the text, a stop-word list specific to scientific publications was constructed to exclude terms irrelevant to the content, such as "figure," "proceedings," "table," and "springer."

Following stop-word removal, the text was segmented into sentences and tokenized into individual words using the NLTK library. Tokens were evaluated to retain only alphabetic words, effectively removing any hidden characters introduced during the PDF-to-text conversion. To complete the cleaning process, special characters and numbers were eliminated using regular expressions.

3.2 Topics and trends identification

To identify topics in the preprocessed corpus, we leveraged BERTopic [16] which operates through the following main stages.

Paper Representation. BERTopic represented each paper using the Sentence-BERT [28] framework, which converts the text into latent vector embeddings. In particular, we employed the pre-trained language model "all-MiniLM-L6-v2."[2] In addition, we extended BERTopic's default implementation to include bi- and tri-grams, enabling the model to incorporate multi-word phrases into the topic detection process. In addition, to ensure the quality and relevance of the extracted topics, we filtered out terms that appeared in fewer than three papers (representing about 1% or less of the corpus). This step helped reduce noise caused by the repetition of terms within individual papers or by specific authors. Furthermore, we applied the BM25 term weighting method, which adjusts the importance of terms based on their frequency across documents [29].

Dimensionality Reduction and Clustering. To reduce the dimensionality of the paper embeddings, BERTopic leverages UMAP, which efficiently preserves both local and global features of high-dimensional data in a lower-dimensional space [22]. The reduced embeddings were then given as input to a clustering algorithm.

Given the potential hierarchical nature of the data, where paper topics may exhibit hierarchical relationships, we utilized agglomerative clustering [26], a hierarchical algorithm, to group document embeddings into topics. Agglomerative clustering starts by treating each paper as its own individual cluster and iteratively merges the closest clusters based on a specified distance metric, thereby forming a hierarchical tree of clusters. In our implementation, we used the *Euclidean distance* metric to compute distances between paper embeddings. For the linkage criterion, we adopted the *Ward linkage* method, which minimizes the variance within clusters during merging, making it particularly effective for creating compact and well-separated clusters. To determine the optimal number of clusters, we explored different distance threshold values. This allows for an adaptive clustering process that aligns with the intrinsic structure of the data.

Topic Representation. Using BERTopic, each paper was assigned to a single topic. While this approach simplifies the analysis, it may not fully capture the multidisciplinary nature of some research papers that span multiple topics. However, it remains effective for distinguishing between topics by emphasizing terms most representative for each topic. To better understand the distinctions between topics and the unique word distributions within each cluster, we employed a modified version of the TF-IDF measurement [17], referred

[2] https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

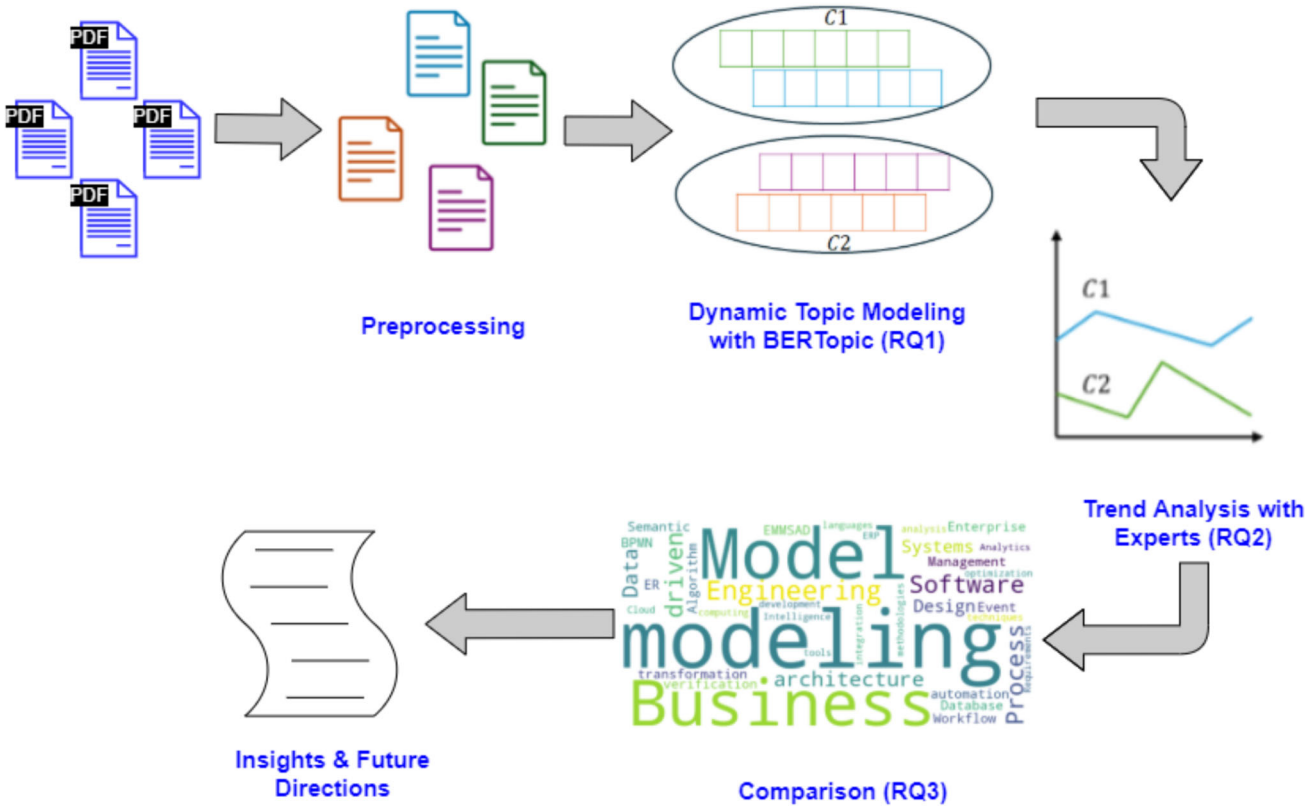


Fig. 1 An overview of the research methodology

to as *c-TF-IDF*. In this approach, all papers within a given cluster were treated as a single document. BERTopic then calculated the TF-IDF for each term and for each cluster, where the inverse document frequency (IDF) was determined based on all documents (clusters in this case). This method allows BERTopic to quantify the importance of each term within a cluster, resulting in topic-word distribution that characterizes the cluster.

Trend Analysis. Besides supporting global topic representation, BERTopic enables Dynamic Topic Modeling (DTM) [5] by recalculating term frequencies (TF) for documents within specific time periods. We applied this approach to ensure consistency in the overall topic structure while also capturing temporal variations. This dynamic modeling capability enabled us to monitor the evolution of topics over time, providing valuable insights into trends.

3.3 Expert evaluation

This step had two primary objectives: (1) to evaluate the quality of BERTopic's outcomes and refine its parameters and (2) to contextualize the topics and trends, providing the basis for insights into related fields and identifying patterns relevant to the evolution of the EMMSAD community.

We explored three different Euclidean distance thresholds—2, 3, and 4.5—which resulted in 26, 17, and 11 topics, respectively. For each detected topic, two headings were generated using ChatGPT−4.0: one based on the top ten c-TF-IDF-ranked terms associated with the topic's papers and the other based on abstracts of the papers. While these headings were shown to the experts, they did not solely determine the final topic headings. In particular, the experts considered the titles and abstracts of the associated papers, their keywords, and the hierarchical structure of the topics.

The expert evaluation was conducted by two of the authors, both of whom have extensive research experience with the EMMSAD community and have served as EMMSAD co-chairs for several years. Their domain expertise provided an informed basis for assessing topic coherence and relevance.

The experts focused their initial assessment on two main aspects. First, they evaluated the **alignment** between a sample of papers and the detected topics, assigning a score from 1 (poor match) to 5 (excellent match) for each paper across all three distance thresholds. Given the manual and detailed nature of this evaluation, a subset of ten papers was randomly sampled from different time periods (two papers per period): 2005–2008, 2009–2012, 2013–2016, 2017–2020, 2021–2024. Second, the experts examined the hierarchical

organization of the topics generated under each distance threshold. For each topic, they classified it into one of three categories: *Irrelevant*, *Relevant but Needs Improvement*, or *Relevant with No Need for Improvement*.

The initial evaluation results revealed that the distance threshold of 3, which generated 17 topics, achieved the highest scores across both assessments. This outcome aligns with expectations, as it strikes an optimal balance between granularity and coherence, effectively capturing meaningful topic distinctions without introducing excessive fragmentation or overgeneralization.

A second, thorough assessment was conducted on the results of the selected configuration (distance threshold of 3). Specifically, the same two authors divided the topics between themselves and wrote detailed comments on ***all*** papers assigned to these topics, as well as on the topics' variance, positioning within the entire dataset, and their evolution over time. The insights from this evaluation are mainly presented in Sects. 4 and 5, but also lead the discussion in Sect. 6.[3]

3.4 Threats to validity

Our methodology provides valuable insights into topics and trends. However, several potential threats to validity must be mentioned, along with the actions taken to mitigate them.

Construct Validity. Construct validity concerns whether the study accurately measures what it intends to measure. A key challenge in our study is determining the optimal number of topics. This process relied on manual analysis conducted by experts, who evaluated a small subset of papers. While their expertise ensured that the selected topics were contextually relevant and meaningful, this approach introduced a degree of subjectivity. To mitigate this, papers were randomly selected from various time periods to create a representative sample. Additionally, each expert worked independently to reduce potential biases, and the analysis was further validated using two distinct scoring measures: alignment and organization.

Another construct validity threat arises from assigning each paper to a single topic. Some papers may span multiple themes, leading to potential misclassification. To address this, we incorporated hierarchical clustering within BERTopic, enabling the exploration of relationships between topics. This approach allows for a more nuanced representation of topic structures, capturing broader or nested associations between research areas. Nevertheless, during the final analysis stage of paper assignment, we observed that while some topics had a perfect match between assigned papers and their thematic focus, others exhibited partial alignment. The results are presented and discussed in detail in Sect. 4.

Internal Validity. Internal validity refers to the question whether the results are influenced by confounding factors. One limitation in our study is the reliance on EMMSAD proceedings as the only data source. While this ensures a focused scope on modeling for IS analysis and design, it excludes relevant studies published in other venues and particularly in journals. This selection bias may influence topic distributions and trends. However, since our study is motivated by the 30th anniversary of EMMSAD, our primary objective was to analyze its evolution over time, making it necessary to focus exclusively on papers from this venue. Nevertheless, this limitation should be considered when interpreting the generalizability of the findings, as broader modeling trends might not be fully captured.

External Validity. External validity pertains to the generalizability of the findings beyond the specific dataset. While our topic modeling approach effectively captures trends within EMMSAD, its applicability to other modeling communities remains unclear. To bridge this gap, we conducted a comparative analysis in response to RQ3, comparing our findings with those of Michael et al. [25], who analyzed three leading modeling conferences (see Sect. 6).

Conclusion Validity. Conclusion validity concerns the reliability and robustness of the results. The primary risk here is the potential misinterpretation of trends due to the limited dataset size. Since some topics have fewer papers, changes in their prevalence might be due to short-term variations rather than long-term trends. To address this, we examined the entire accessible data and verified topic assignments using both automated analysis and expert review.

Another aspect affecting conclusion validity is the hierarchical structure of topics. While hierarchical clustering enhances topic organization, it may introduce complexities in interpretation, especially when topics evolve over time. To mitigate this, we discuss the hierarchical relationships and provide qualitative insights alongside the quantitative topic distributions.

4 Discovered topics and their hierarchy (results for RQ1)

This section presents the 17 topics identified by our method. Figure 2 presents the topics in a tree structure, showing the number of papers assigned to each topic. The topics are numbered in descending order, starting with 0, according to the number of papers classified under each. This numbering reflects the relative prevalence of topics within the dataset.

In the following, we discuss the content and scope of these topics which we thematically grouped into four groups: Enterprise and Business Process Modeling, Requirements

[3] The evaluation notes can be found at https://zenodo.org/records/15489799.

and Software Modeling, Information and Data Modeling, and Foundational and Methodological Aspects. These groupings are based on our informed interpretation of the topic content and their alignment with EMMSAD tracks over the past decade.

4.1 Enterprise and business process modeling

Enterprise (Architecture) Modeling and Meta-Modeling (Topic 1) brings together contributions from both enterprise architecture (EA) modeling and enterprise modeling (EM) in general. In an EA context, models need to focus on providing a holistic view of the organization's current and desired future state of affairs, while EM in general also includes model-based representation and analysis of specific organizational elements, such as goals and business processes. The contributions on this topic emphasize the foundational role of meta-modeling and its application in EA modeling and EM in general. At a more detailed level, some contributions focus on strategy- and goal-driven modeling and architecture modeling, while several others highlight the suitability of the capability concept for facilitating modeling and transformation process, as well as for creating organizational maps as an efficient EA (model-based) artifact type. Several contributions address methods and modeling languages, including enhancements to existing modeling languages such as ArchiMate. In particular, there is a strong focus on modeling networked enterprises, including digital business ecosystems, and addressing interoperability concerns within this scope. Collaborative, interactive, and decentralized modeling approaches, with an emphasis on the role of stakeholders, are also presented. Finally, some contributions address modeling for maturity, transparency, and organizations oriented toward software production.

Business and Enterprise Engineering (Topic 5) merges two closely related disciplines: *business engineering*, which focuses on designing and improving business components such as strategies and value models, and *enterprise engineering*, which addresses the enterprise as a holistic system encompassing business, IT, and stakeholders. The included papers specifically explore the modeling of business strategies for aligning internal processes to IT infrastructure, modeling business value exchanges and agreements, business rules, and data-driven business services. They also propose methods for designing complex business environments, situation-specific business model development methods, and best practices for developing patterns for innovation.

Within the enterprise engineering scope, the contributions cover participatory enterprise modeling, EA management, enhancing customer experience, as well as best practices for analyzing EM approaches, including task patterns, roles, and supporting tools. In addition, some papers discuss the evaluation of EM with stakeholders and practical usage. Specifically, two papers propose the integration of design thinking into the process of EM and EA management and into the design of remote collaborative environments.

This topic share some commonalities with *Topic 1: Enterprise (Architecture) Modeling and Meta-Modeling*, as shown in Fig. 2. However, the topics differ in focus—while Topic 1 primarily concerns the creation of models, this topic takes a broader engineering (and architecting) perspective, emphasizing the strategic re-engineering of organization's processes, services, and IT systems.

Business Process Modeling (Topic 6), an important specialization of EM, encompasses contributions focusing on various aspects of business process modeling, including analytical evaluations, methodical improvements, and practical applications. Several works enhance existing standards, such as BPMN, e.g., an analytical evaluation using the Semiotic Quality Framework, translation mechanisms between BPMN and BPEL, and an approach extracting concepts from serialized BPMN models. Several innovative frameworks and approaches are proposed, such as a Business Process Metadata Model (BPMM) to facilitate model interpretation and navigation, a meta-model integrating business rules with BPMN adaptation patterns, and a two-dimensional framework for capturing design-time variability. Other contributions include the integration of Markov theory in operational environments, context-aware reasoning and configuration, expanding Process-Goal Alignment with simulation mechanisms, and leveraging large language models (LLMs) for automated process model generation from textual descriptions. A systematic literature review further examines the quality of BPM formalisms.

Business Rule and Regulatory Compliance (Topic 16), another important specialization of EM, is addressed by contributions that focus on modeling, analyzing, and operationalizing rules and constraints to ensure adherence to regulatory requirements. Contributions include modeling of static deontic rules, characterizing business rules for practical IS applications, and proposing an alternative framework for capturing business constraints that can be embedded into IS. Semantic vocabularies are utilized to map regulations to enterprise operational details, enabling actionable compliance measures. Additionally, advancements in AI are demonstrated in the automatic creation of legal Goal-oriented Requirements Language (GRL) models from textual regulations. These studies highlight the interplay between regulatory compliance and business rule modeling, offering practical and theoretical insights to enhance compliance processes in enterprise environments.

4.2 Requirements and software modeling

Model-driven Software Development (Topic 2) encompasses contributions related to Model-Driven Architecture

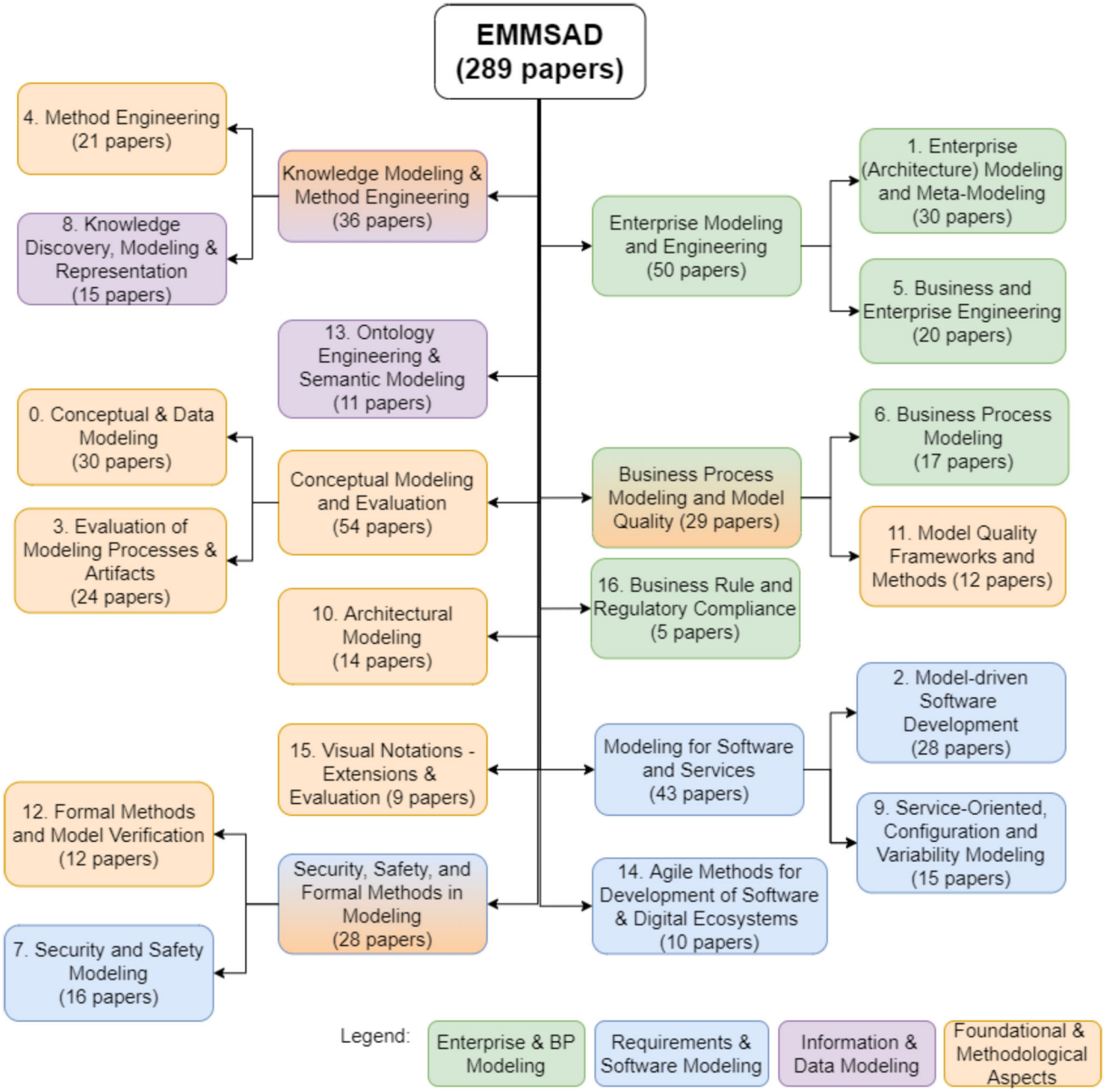


Fig. 2 Generated Hierarchy of Topics

(MDA), Model-Driven Development (MDD), and Model-Driven Engineering (MDE), focusing on leveraging models to guide various software development processes. In particular, the contributions include frameworks, methods, meta-models, and tools to enhance different aspects of software engineering, such as a multi-view system modeling approach, meta-models for architectural design rules and variability modeling, and methods for generating or extending MDA models. In addition, the topic spans advances in conceptual modeling, goal-oriented modeling, traceability, usability, and system architecture reconstruction in the context of MDA/MDD/MDE. These contributions collectively demonstrate the versatility and applicability of model-driven approaches in addressing challenges across various software development domains. It is important to note that some contributions overlap with adjacent topics, addressing issues such as quality assurance, Software Product Line Engineering (SPLE), ontologies, or conceptual modeling in general.

Service-Oriented and Configuration and Variability Modeling (Topic 9) integrates the two related disciplines of *Service-Oriented* and *Configuration and Variability Modeling*. Both of these topics are also related to *Topic 2:*

Model-driven Software Development, as shown in Fig. 2. Contributions to service-oriented modeling include frameworks and methods for designing value-added services, formalizing the redesign of services for multi-channel access, modeling services using contracts, and investigating costs in Process-Aware Information Systems (PAIS) engineering through simulation models. Other works introduce ITIL-based frameworks for service IT management and standard components for federated identity management.

In configuration and variability modeling, the contributions address modeling data variability in software product lines, analyzing feature-oriented versus UML-based approaches, and supporting dynamic variability related to both operational and quality-of-service aspects. Formal methods, such as a framework for automatic variability analysis and an Abstract State Machine model for monitoring services addressing system failures, provide theoretical advancements.

Security and Safety Modeling (Topic 7) also captures contributions from two strongly related disciplines: *Security Modeling*, which focuses on protecting systems from external threats like hacking and malware, and *Safety Modeling*, which addresses the protection of sensitive information and internal, accidental issues, such as system failures, to prevent harm caused by these failures. The contributions emphasize the importance of addressing security and safety concerns early in system development and advocate for the integration of their modeling into mainstream analysis and design processes. Notably, some contributions propose threat modeling and hazard analysis by means of stochastic simulation models, requirements specification using security and risk-aware modeling languages, or integration into Domain-Specific Languages (DSLs). MDE is also explored as a means to design security requirements, while some papers address security modeling across different stages of the Software Development Life Cycle (SLDC), including requirements prioritization and prototyping.

Specific applications of security modeling are also discussed, including its use in multi-agent systems and cloud environments. Beyond the primary focus on security and safety, two papers concern MDD, independently of security or privacy concerns, and one paper investigates the use of chatbots to facilitate querying domain-specific models.

Agile Methods for Development of Software and Digital Ecosystems (Topic 14) focuses on methods for IS development that emphasize adaptability, user collaboration, continuous delivery over non-rigid planning, and simplicity. Specific contributions explore various aspects of agile methods, including how the values and principles of the Agile Manifesto align with the needs of globally distributed software development; refinements of agile approaches for requirements engineering; requirements development to logical architecture and requirements change handling; and integration of agile methods with EA for continuous delivery, and with participative EM.

Some contributions also advocate the use of low-code/no-code platforms, which facilitate the creation of smart, extendable software while supporting diverse modeling formats (e.g., UML, chatbot interfaces). This includes the application of agile methods and low-code/no-code tools for developing software requirements within ecosystems and prototyping digital platforms, such as those based on blockchain.

4.3 Information and data modeling

Knowledge Discovery, Modeling, and Representation (Topic 8) consolidates contributions related to the capture and use of both explicit knowledge (such as documentation and processes) and tacit knowledge (e.g., from user behavior or logs) to enhance models and methods for IS development. On the foundational side, some studies focus on incorporating experience into modeling tasks, understanding the rationale behind methods-in-use, transforming human mental models into machine learning models through conceptual models, and creating know-how maps to guide technology adoption and development decisions. Another group of papers addresses the leveraging of domain knowledge, including the creation of domain-specific labels in process models for better human comprehension, conceptual modeling of "soft domain information," the development of system analysis pipelines, managing large knowledge maps, and building domain-specific action recognition classifiers. The remaining papers explore a user-centric approach to knowledge discovery, with studies on review-based user profiling, discovering user behavior in systems, mapping user visualization requirements, and analyzing how people interact with visual conceptual models.

Ontology Engineering and Semantic Modeling (Topic 13) encompasses the methodological aspects of the ontology development life cycle, as well as the methods for representing the meaning of data elements within their contextual use. Specifically, some contributions propose ontologies for IS development, incorporating contextual aspects, and some focus on enhancing existing reference ontologies. Several studies explore improvements in ontology creation, particularly through the use of various NLP techniques and automation. A group of contributions evaluates ontology-building methods for optimization between expressivity and computational tractability. Furthermore, some research suggests innovative uses of ontologies, such as for validating and improving different types of conceptual and enterprise models; developing methods for reusing ontologies across application domains; and representing and visualizing ontologies using techniques such as knowledge graphs. A smaller subset of contributions specifically addresses the modeling of semantics, including the semantic annotation of

process templates using domain-specific ontologies, and the automation of such annotation processes.

4.4 Foundational and methodological aspects

In this subsection, we discuss topics that are foundational for IS development, as well as those that span across the topics as discussed above, particularly focusing on methodological aspects.

Conceptual and Data Modeling (Topic 0) aggregates the contributions that focus on the semantic representation of the real-world entities, data structures, and their implementation in specific languages and technologies. Concretely, some contributions analyze conceptual modeling languages, such as ER, Object-Role Modeling (ORM) and UML, evaluating their individual and comparative capabilities, as well as possible extensions for enhancing entity, relationship, and object modeling. Other contributions explore the integration of ontologies into conceptual modeling, aiming to adhere to semiotic principles and improve design heuristics. Further concerns identified within this topic include automated model analysis, evaluation, model evolution and change, and maintenance. A group of contributions specifically address data modeling for database conceptualization and querying, with an emphasis on NoSQL approaches. Finally, additional contributions emphasize modeling aspects for information search and information mapping. Notably, some contributions overlap with adjacent topics, addressing issues related to information and data modeling.

Method Engineering (ME) (Topic 4) brings together the contributions concerning the design of customized methods for IS development. They focus on conceptualizing, combining and visualizing modular, reusable components known as method fragments or chunks. These fragments can be tailored to contextual, project-specific requirements (situational method engineering). Several papers focus on the application of ME in specific development disciplines, such as software product line engineering, multi-agent systems, requirements engineering, IS development with commercial off-the-shelf (COTS) solutions, and SDLC. Other contributions address the management aspects of ME, such as adoption, maintenance, and reuse, as well as the guidance for managing complex method tasks through procedures, visualizations, and other aids.

Beyond method engineering, three papers concern method design for business process discovery and improvement. Two contributions propose reusable process mining methods, while the third introduces a method for analyzing data inaccuracy in process models. However, none of these studies explicitly refer to method fragmentation.

Architectural Modeling (Topic 10) covers a diverse set of contributions concerned with modeling in various architectural contexts, including EA, MDA, and agent-based systems, alongside considerations of requirements engineering and in particular Non-Functional Requirements (NFRs) such as trust and ethics. Most papers contribute directly to advancing architectural methodologies and frameworks. Examples of contributions include clarifying the conceptual landscape of EA, the design Architectural Modeling languages such as ArchiMate, devising Design and Engineering Methodology for Organizations (DEMO) guidelines and process patterns, and developing frameworks for assessing model-based transformations between requirements and architecture. Other works address specialized domains, such as designing accountability architectures for socio-technical systems, valuing smart grid architectures from a business perspective, and formulating requirements for digital twin development in industrial contexts. Ethical and socially responsible software development is explored through methods that adapt Corporate Social Responsibility principles, integrate NFRs like trust and privacy, and employ ontology-driven requirements analysis.

Evaluation of Modeling Processes and Artifacts (Topic 3) encapsulates a broad spectrum of contributions that emphasize the evaluation of modeling processes and artifacts, concerning quality, including comprehension, and usability. The evaluation is mainly done through experiments and case studies. Some papers have theoretical contributions, such as evaluation frameworks on conceptual modeling and design science. The empirical studies explore diverse evaluation themes, e.g., the influence of cognitive style, task self-efficacy, and domain knowledge on modeling quality; the effectiveness of modeling approaches for trained versus untrained individuals; and the effects of feedback-enabled tools and multi-perspective modeling on comprehension and learning.

Our data-driven method grouped this topic with *Topic 0: Conceptual and Data Modeling* as some contributions within it evaluate proposed modeling approaches.

Model Quality Frameworks and Methods (Topic 11) brings together contributions that assess the extent to which conceptual, data, process, or other models fulfill their intended purpose and meet stakeholder needs. The majority of contributions focus on the application, adaptation, or extension of some generic frameworks to specific needs. An example of such a framework is SEQUAL, which encompasses as main dimensions syntactic, semantic, and pragmatic model quality concerns. Several studies specialize SEQUAL for particular model types (e.g., conceptual models) or business domains, or refine it to address additional aspects, such as data quality, knowledge state transitions, and rule-based process modeling. Other contributions explore model quality assessment using the Goal-Question-Metric (GQM) framework or defining metrics for domain-specific models and their application. These studies cover diverse

application domains, including healthcare and clinical environments, the energy market, and cartography.

The remaining contributions propose novel approaches for evaluating the quality of enterprise models, such as goal or value models, by applying formal properties or pattern-based assessments. As evident from the discussion above and the hierarchy illustrated in Fig. 2, this topic shares to some extent common terms with *Topic 6: Business Process Modeling*.

Formal Methods and Model Verification (Topic 12) encompasses research contributions primarily focused on evaluation, verification, and Model-Driven Architecture (MDA). Differently from *Topic 2: Model-driven Software Development* and *Topic 3: Evaluation of Modeling Processes And Artifacts*, the papers in this topic mainly utilize formal approaches. Contributions include formalizing and verifying UML-based models, such as categorizing ontological rules for UML conceptual modeling, integrating UML activity diagrams with temporal logic for precise semantics, and using graphical UML views to improve the readability and validation of formal B specifications. Yet, some papers do use empirical techniques to examine the comprehensibility of UML diagrams, the cognitive support provided by UML to developers, and user preferences.

The formal verification advancements include using Abstract State Machines (ASM) predicates for model validation and employing static analysis techniques for executable domain-specific modeling languages (DSMLs). Comparative analyses, such as those between constraint languages (like OCL), Java, and natural language, further explore the interplay of formal and informal modeling approaches.

Visual Notations - Extensions and Evaluation (Topic 15) explores advancements in visual notations, their extensions, and the assessment of their efficacy in modeling contexts. Contributions include the development of a graphical modeling notation based on colored Petri nets for evaluating the performance and costs of mobile applications, as well as the design of VCL, a novel visual, formal and modular language that exceeds UML in expressing abstract specifications. Extensions to established notations are also explored, such as adding visual elements to BPMN diagrams to capture location and equipment details in mobile work processes. Other contributions investigate visual representation techniques, comparing color with black-and-white pattern fills or symbolic icons, and evaluating meaningful variability in visual notations. A systematic approach for applying Physics of Notation (PoN) to the design of concrete syntax is complemented by studies on operationalizing PoN and separating its mathematical and user-specific aspects.

Summary of Results for RQ1

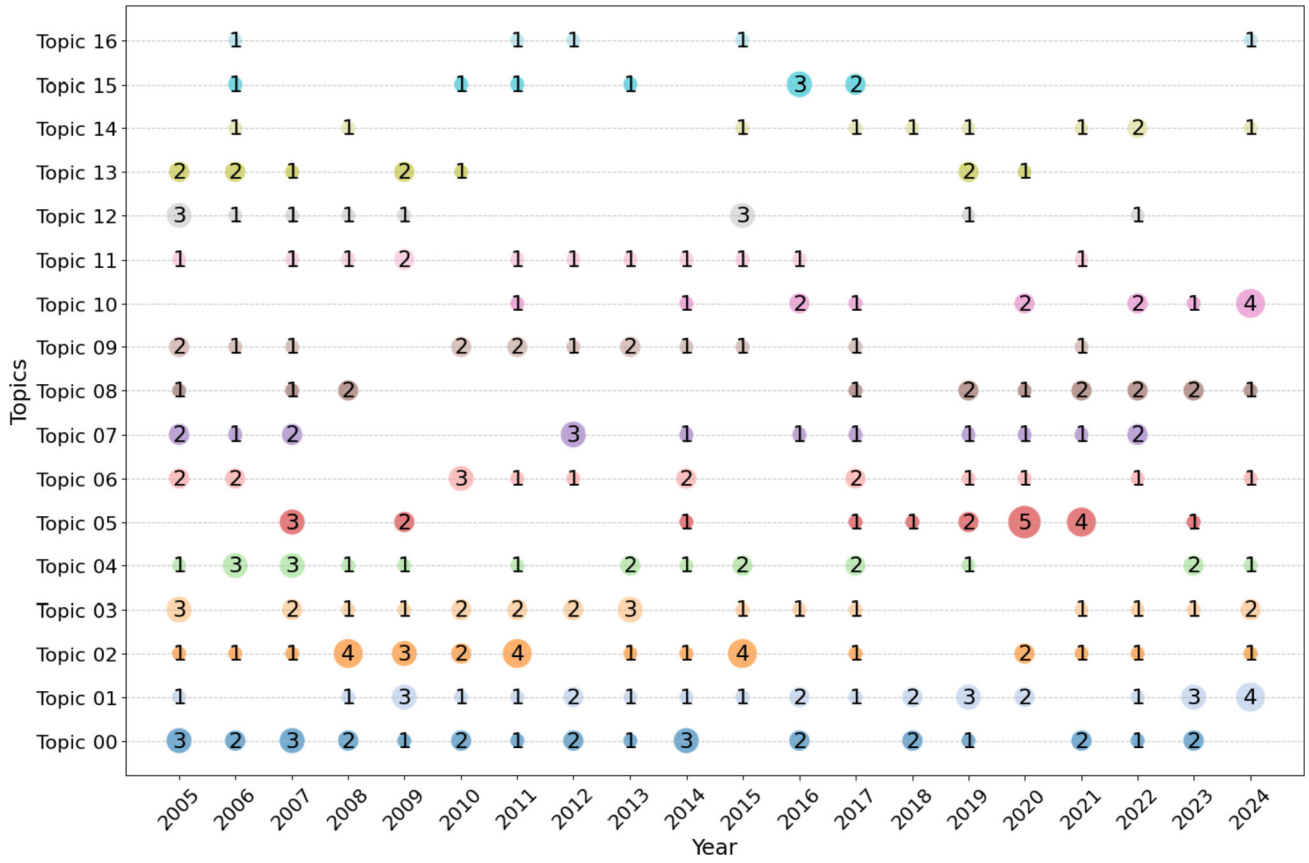
To summarize, our data-driven approach identified **17 topics**, which are visually represented in Fig. 2 in a tree structure, alongside the number of contributions for each topic. These topics are organized into four key thematic categories: **Enterprise and Business Process Modeling**, **Requirements and Software Modeling**, **Information and Data Modeling**, and **Methodological Aspects**.

5 Evolution of topics over time (results for RQ2)

We further analyzed the evolution of the presented topics over time. Figure 3 presents the results. We identified five patterns of development within the topics, though we remain cautious in our interpretation due to the relatively small number of papers in each category.

Mainstay topics: These had a consistent presence over the years, with papers published regularly since 2005 and new contributions almost every year. The stable topics include: *Topic 0: Conceptual and Data Modeling*, *Topic 1: Enterprise (Architecture) Modeling and Meta-Modeling*, which started a little bit later in 2008, likely due to the establishment of standard modeling notations such as ArchiMate, and *Topic 2: Model-driven Software Development*. These three topics also have the largest number of papers and span across the three main areas of IS modeling, namely conceptual, enterprise, and software modeling, respectively.

Sunset topics: These topics were once popular, but have lost interest to some extent in recent years. The topics in this category are: *Topic 9: Service-Oriented, Configuration and Variability Modeling: Service-oriented computing*, which experiences less focus as a topic area, and *Topic 11: Model Quality Frameworks and Methods*, *Topic 12: Formal Methods and Model Verification*, *Topic 13: Ontology Modeling and Semantic Modeling*, and *Topic 15: Visual notations - Extensions and Evaluation*. The decline in interest in these topics could be attributed to different factors, such as industry evolution, shifts in research focus, adoption of more modern practices and technologies, and growing interdisciplinary connections between fields, which have led to the evolution of these topics into new ones with broader or different foci. Alternatively, some of these topics may have reached a level of maturity or standardization that reduces the perceived need for further research; this is potentially the case, for instance, with *Topic 15: Visual notations - Extensions and Evaluation*. A future citation-based analysis could offer further insights into whether these topics remain influential despite reduced publication activity.



00. Conceptual & Data Modeling
01. Enterprise (Architecture) Modeling and Meta-Modeling
02. Model-driven Software Development
03. Evaluation of Modeling Processes & Artifacts
04. Method Engineering
05. Business and Enterprise Engineering
06. Business Process Modeling
07. Security and Safety Modeling
08. Knowledge Discovery, Modeling & Representation
09. Service-Oriented, Configuration and Variability Modeling
10. Architecture Modeling
11. Model Quality Frameworks and Methods
12. Formal Methods and Model Verification
13. Ontology Engineering & Semantic Modeling
14. Agile Methods for Development of Software & Digital Ecosystems
15. Visual Notations – Extensions & Evaluation
16. Business Rules or Regulatory Compliance

Fig. 3 Evolution of topics over time

Frontier topics: These topics were not widely popular 15 to 20 years ago but have gained increased interest over the last decade. They include: *Topic 5: Business and Enterprise Engineering*, *Topic 10: Architectural Modeling*, and *Topic 14: Agile Methods for Development of Software and Digital Ecosystems* (where the term "digital ecosystems" has only gained prominence in the last decade). The growing interest in these topics can be attributed to the increasing demand for business agility, the rising complexity of systems, the need for greater interconnectivity, and the emphasis on continuous value delivery.

Phoenix topics: These are topics that were once popular, and then fell out of fashion, but have recently experienced a resurgence. The topics in this category are: *Topic 3: Evaluation of Modeling Processes and Artifacts*, *Topic 4: Method Engineering*, *Topic 7: Security and Safety Modeling* (an area which is increasingly important in the IT field in general), and *Topic 8: Knowledge Discovery, Modeling and Representation*. These topics are making a return for various reasons, primarily driven by the increasing complexity of systems. However, the dominant factor seems to be the emergence of AI, which not only enables new evaluation

methods but also introduces new requirements, such as the need for explainability and the emergence of new types of threats. For example, *Topic 7: Security and Safety Modeling* has regained relevance in the context of threat modeling and domain-specific languages (DSLs) for security and safety, while *Topic 4: Method Engineering* has been revived through advancements in mining (particularly process mining) and automatic code generation.

Peripheral topics: These topics have attracted only occasional contributions and include *Topic 6: Business Process Modeling* and *Topic 16: Business rule and Regulatory compliance*. The modest but steady presence of these topics is likely not due to a general decline in relevance, but rather may be influenced by the concurrent presence of the BPMDS working conference, which addresses similar themes alongside EMMSAD and may attract related submissions.

Summary of Results for RQ2

To summarize, we identified five patterns of topic evolution over time: **Mainstay topics** (consistent presence), **Sunset topics** (declining interest in recent years), **Frontier topics** (increased interest in the past decade), **Phoenix topics** (resurgence after a decline), and **Peripheral topics** (occasional contributions).

6 EMMSAD relations to related fields (results for RQ3)

To address RQ3 and examine the relationships between IS modeling and other modeling communities, we used the work by Michael et al. [25]. Their study provides a bibliometric analysis of three major conferences in modeling: ER (International Conference on Conceptual Modeling, held annually since 1979), MODELS (International Conference on Model-Driven Engineering Languages and Systems, established in 1998 as the UML conference and renamed in 2005), and BPM (International Conference on Business Process Management, held annually since 2003). These conferences represent distinct but interconnected modeling areas: ER focuses on data and conceptual modeling, MODELS on software and model-driven engineering, and BPM on process modeling and workflow management. In addition to the bibliometric data, a survey was distributed via mailing lists such as IS-WORLD and DB-World, gathering 153 responses (128 researchers and 25 industry practitioners. Expert interviews were also conducted with three founders of these conferences: Peter P.S. Chen (ER), Jean Bézivin (MODELS), and Wil van der Aalst (BPM).

Table 1 presents the topics identified in the study by [25] and their mapping to our topics. These topics were derived through *manual coding* of responses from free-text fields in the survey. The table includes areas with at least three mentions overall.

It is important to clarify that Michael et al.'s study focuses on identifying perceived connections and "exciting" topics within a broad modeling community, rather than analyzing actual research trends. While our comparison is not intended as a direct alignment of research trends, it offers a view on how IS modeling topics relate to those seen as promising ("exciting") by the modeling community. Moreover, it is important to highlight two specific differences between the studies: (i) Our study adopts a data-driven approach, whereas Michael et al. employed a survey methodology involving expert input and manual coding, and (ii) our research focuses on modeling for IS analysis and design, as reflected in the papers accepted to the EMMSAD conference, while Michael et al.'s study covers a broader scope, addressing various modeling communities, including those related to conceptual, software, and business process modeling.

Our comparison shows clear overlap in many areas. In particular, **Conceptual Modeling** from Michael et al. closely maps to *Topic 0: Conceptual and Data Modeling*. However, our study also includes contributions related to **Data Modeling**, which is identified as a separate topic in [25].

Other prominent matches include **Model-Driven Engineering**, which aligns with *Topic 2: Model-driven Software Development*; **Ontologies**, corresponding to *Topic 13: Ontology Engineering and Semantic Modeling*; **Process Modeling**, which shows significant overlap with *Topic 6: Business Process Modeling*; **Enterprise Architecture**, mapped to *Topic 1: Enterprise (Architecture) Modeling and Meta-Modeling*, and to some extent also to *Topic 5: Business and Enterprise Engineering*; and **Agility**, which aligns with *Topic 14: Agile Methods for Development of Software and Digital Ecosystems*.

The most frequently mentioned topic in Michael et al.'s study, **Artificial Intelligence (AI)**, did not emerge as a separate topic in our study. However, contributions that use AI for modeling or modeling for AI, including generative AI and LLM, appear across different topics, such as *Topic 6: Business Process Modeling*, *Topic 7: Security and Safety Modeling*, *Topic 10: Architectural Modeling*, and *Topic 16: Business Rule and Regulatory Compliance*. Most AI-related papers in EMMSAD employ symbolic AI combined with other techniques. Nevertheless, we expect that the growing interest in machine learning and LLM will significantly increase the number of submissions on this topic in the near future. In this context, our method also did not identify separate topics for *Automation* and *Flexible Modeling*.

Several topics have a wide, yet partial, overlap. These include areas such as **Domain-Specific Languages (DSL)**, which aligns to some extent with our *Topic 15: Visual Notations - Extensions and Evaluation*; **Human Factors** and

Table 1 Mapping between Michael et al.'s Topics and ours. The numbers in parentheses indicate the number of mentions as the most exciting topics in Michael et al.'s work and the number of papers in our study

Topic by Michael et al. [25]	Our Topic
AI (21 mentions)	
Conceptual Modeling (12 mentions)	Conceptual and Data Modeling (30 papers)
DSL (10 mentions)	Visual Notations—Extensions and Evaluation (9 papers)
Data Modeling (9 mentions)	Conceptual and Data Modeling (30 papers)
Model-Driven Engineering (9 mentions)	Model-driven Software Development (28 papers)
Ontologies (9 mentions)	Ontology Engineering and Semantic Modeling (11 papers)
Process Modeling (9 mentions)	Business Process Modeling (17 papers)
Automation (7 mentions)	
Flexible Modeling (7 mentions)	
Model Integration (7 mentions)	
Human Factors (6 mentions)	Evaluation of Modeling Processes and Artifacts (24 papers)
Bioinformatics (5 mentions)	
Low Code (5 mentions)	
Model Transformations (5 mentions)	
Modeling Tools (5 mentions)	
NoSQL (5 mentions)	Conceptual and Data Modeling (30 papers)
Agility (4 mentions)	Agile Methods for Development of Software and Digital Ecosystems (10 papers)
Big Data (4 mentions)	
Enterprise Architecture (4 mentions)	Enterprise (Architecture) Modeling and Meta-Modeling (30 papers) Business and Enterprise Engineering (20 papers)
Knowledge Graphs (4 mentions)	Knowledge Discovery, Modeling, and Representation (15 papers)
Language Engineering (4 mentions)	
Modeling Theory (4 mentions)	
Security (4 mentions)	Security and Safety Modeling (16 papers)
Usability (4 mentions)	Evaluation of Modeling Processes And Artifacts (24 papers)
Industry 4.0 (3 mentions)	
Model Processing (3 mentions)	
Modeling Education (3 mentions)	
Multi-Level Modeling (3 mentions)	
Simulation (3 mentions)	
Testing (3 mentions)	
	Method Engineering (21 papers)
	Service-Oriented, Configuration and Variability Modeling (15 papers)
	Architectural Modeling (14 papers)
	Model Quality Frameworks and Methods (12 papers)
	Formal Methods and Model Verification (12 papers)
	Business Rule and Regulatory Compliance (5 papers)

Usability, both covered within our *Topic 3: Evaluation of Modeling Processes and Artifacts*; and **Knowledge Graphs**, which has overlaps with *Topic 8: Knowledge Discovery, Modeling, and Representation* in our categorization. Additionally, **Security** from Michael et al.'s work maps to *Security and Safety Modeling* in our study. Although these topics appear in both studies, their focus and the specific nature of the contributions may differ, with our study emphasizing aspects relevant to IS analysis and design.

Several topics exhibit a wide yet partial overlap between the two studies. These include **Domain-Specific Languages (DSL)**, which aligns to some extent with our *Topic 15: Visual Notations - Extensions and Evaluation*. Similarly, **Human Factors** and **Usability** are both addressed within contributions categorized under *Topic 3: Evaluation of Modeling Processes and Artifacts*. **Knowledge Graphs** shares overlaps with our *Topic 8: Knowledge Discovery, Modeling, and Representation*, and **Security** from Michael et al.'s work cor-

responds to our *Security and Safety Modeling*. While these topics appear in both studies, their specific focus and the nature of the contributions differ, with our study placing particular emphasis on aspects that are more directly relevant to IS analysis and design.

We further identified topics that do not have natural counterparts in our study. These topics can be divided into two categories: those from the broader modeling community and technology-specific topics. The first group includes **Model Integration**, **Model Transformations**, **Modeling Tools**, **Language Engineering**, **Modeling Theory**, **Model Processing**, **Modeling Education**, **Multi-Level Modeling**, **Simulation**, and **Testing**. The second group, which includes technology-specific topics, consists of **Low Code**, **NoSQL**, **Big Data**, and **Industry 4.0**. Although these topics did not emerge as distinct categories in our study, they have been mentioned sporadically in EMMSAD over the years. For instance, **NoSQL** can be loosely mapped to our *Topic 0: Conceptual and Data Modeling - Foundations and Languages*, as the only two papers explicitly mentioning NoSQL in their abstracts were categorized under this topic. Big data on the other hand was mentioned in contributions related to *Topic 5: Business and Enterprise Engineering*, *Topic 8: Knowledge Discovery, Modeling, and Representation*, and *Topic 13: Ontology Engineering and Semantic Modeling*. Additionally, **Bioinformatics** was not represented in our dataset, likely due to its focus on a specific application domain of modeling.

Lastly, several of our topics were not identified by Michael et al., yet they are considered fundamental to modeling for IS analysis and design. These include *Topic 4: Method Engineering*, *Topic 9: Service-Oriented, Configuration and Variability Modeling*, *Topic 10: Architectural Modeling*, *Topic 11: Model Quality Frameworks and Methods*, *Topic 12: Formal Methods and Model Verification*, and *Topic 16: Business Rule and Regulatory Compliance*. It is worth noting that some of these topics resulted in relatively low numbers of papers also within the EMMSAD corpus.

Summary of Results for RQ3

To summarize, our comparison with Michael et al.'s study [25] reveals **strong alignment** in areas such as Conceptual Modeling, Model-Driven Engineering, Ontologies, and Process Modeling. However, several topics from the broader modeling community, as well as **technology-specific topics**, did not have natural, distinct counterparts in our study and were included in papers categorized across different topics. Additionally, we identified topics **essential to IS analysis and design**, such as Method Engineering.

7 Conclusions

This study offers a retrospective analysis of the evolution of information systems (IS) modeling through the lens of the EMMSAD working conference over the past two decades. It identifies 17 key recurring topics, organized into four thematic categories: Enterprise and Business Process Modeling, Requirements and Software Modeling, Information and Data Modeling, and Methodological Aspects. The analysis also highlights five patterns of topic evolution: Mainstay, Sunset, Frontier, Phoenix, and Peripheral topics. A comparison with broader modeling communities reveals strong alignment in areas like Conceptual Modeling, Model-Driven Engineering, Ontologies, and Process Modeling, while also identifying gaps in foundational modeling areas that are spread across various topics.

Overall, our study highlights the dynamic nature of IS modeling, emphasizing that while models remain essential representations of real-world systems in IS analysis and design, they must continuously adapt to new assumptions and technological advancements. These adaptations include modeling not only physical but also digital realities, facilitating transitions between the two—e.g., through the growing focus on Digital Twins (DT [12]), as evidenced in *Topic 10: Architectural Modeling* of our analysis. Other emerging technologies such as Extended Reality (XR [9]) (including metaverse initiatives), Distributed Ledger (DL [11]), Quantum Computing (QC [3]), and IoT/5 G [13] become more and more relevant to the modeling community, and we assume that their integration will shape future modeling practices and research directions. Additionally, models are no longer produced and consumed solely by humans, but increasingly by digital actors, reflecting the expanding role of AI and machine learning in IS. Conceptual modeling is also shifting beyond traditional organizational boundaries, occurring in diverse contexts and applications. Given these transformations, close collaboration between researchers and practitioners is imperative to ensure that modeling approaches remain relevant and effective. EMMSAD, with its focused and interdisciplinary coverage, is well-positioned to facilitate this collaboration, bridging gaps across specialized modeling domains and fostering the integration of emerging technologies into IS modeling practices.

Funding Open access funding provided by University of Haifa.

References

1. Arbab, F., Bekius, S.F., Bonsangue, M.M., Bosma, H., Campschroer, J., Cuvelier, M.J., Fennema, P., Groenewegen, L., Hoppenbrouwers, S.J.B.A., Iacob, M.E., Janssen, W.P.M., Jonkers, H., Krukkert, D., Lankhorst, M.M., Penders, P.G.M., Proper, H.A., Slagter, R.J., Stam, A.W., Steen, M.W.A., Wieringa, R.J., de Boer, F.S., ter Doest, H.W.L., van Buuren, R., van Eck, P.A.T., van Leeuwen, D., van der Torre, L.W.N., Veldhuijzen van Zanten, G.E.: Enterprise Architecture at Work: Modeling, Communication and Analysis. The Enterprise Engineering Series, Springer, Berlin, Germany, 4th edn (2017). https://doi.org/10.1007/978-3-662-53933-0
2. Band, I., Ellefsen, T., Estrem, B., Iacob, M.E., Jonkers, H., Lankhorst, M.M., Nilsen, D., Proper, H.A., Quartel, D.A.C., Thorn, S.: ArchiMate 3.0 Specification. Van Haren, Zaltbommel, The Netherlands (2016)
3. Beisel, M., Barzen, J., Leymann, F., Stiliadou, L., Weder, B.: Observability for quantum workflows in heterogeneous multi-cloud environments. In: Guizzardi, G., Santoro, F., Mouratidis, H., Soffer, P. (eds.) Advanced Information Systems Engineering, pp. 612–627. Springer Nature Switzerland, Cham (2024)
4. Bird, S., Klein, E., Loper, E.: Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, Inc (2009)
5. Blei, D.M., Lafferty, J.D.: Dynamic topic models. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 113–120 (2006)
6. Booch, G., Rumbaugh, J., Jacobson, I.: The Unified Modeling Language: User Guide. Addison-Wesley (2005)
7. Bork, D., David, I., España, S., Guizzardi, G., Proper, H.A., Reinhartz-Berger, I.: The role of modeling in the analysis and design of sustainable systems: a panel report. Commun. Assoc. Inf. Syst. **54**(1), 41 (2024)
8. Bubenko Jr, J.A.: From information algebra to enterprise modelling and ontologies-a historical perspective on modelling for information systems. In: Conceptual Modelling in Information Systems Engineering, pp 1–18. Springer (2007)
9. Campos-López, R., Guerra, E., de Lara, J., Colantoni, A., Garmendia, A.: Model-driven engineering for augmented reality. J. Object Technol. **22**(2), 1–15 (2023)
10. Chen, P.: The entity-relationship model: towards a unified view of data. ACM Trans. Database Syst. **1**(1), 9–36 (1976)
11. Curty, S., Fill, H.G.: SmartCML: a visual modeling language to enhance the comprehensibility of smart contract implementations. In: Paja, E., Zdravkovic, J., Kavakli, E., Stirna, J. (eds.) The Practice of Enterprise Modeling, pp. 87–104. Springer Nature Switzerland, Cham (2025)
12. David, I., Bork, D.: Infonomics of autonomous digital twins. In: Guizzardi, G., Santoro, F., Mouratidis, H., Soffer, P. (eds.) Advanced Information Systems Engineering, pp. 563–578. Springer Nature Switzerland, Cham (2024)
13. De Luzi, F., Leotta, F., Marrella, A., Mecella, M.: On the interplay between business process management and internet-of-things. Bus. Inf. Syst. Eng. (2024). https://doi.org/10.1007/s12599-024-00859-6
14. Di Ruscio, D., Kolovos, D., de Lara, J., Pierantonio, A., Tisi, M., Wimmer, M.: Low-code development and model-driven engineering: Two sides of the same coin? Softw. Syst. Model. **21**(2), 437–446 (2022)
15. Gane, C., Sarson, T.: Structured Systems Analysis: Tools and Techniques. Prentice Hall (1976)
16. Grootendorst, M.: BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794 (2022)
17. Joachims, T., et al.: A probabilistic analysis of the rocchio algorithm with tfidf for text categorization. In: ICML. vol. 97, pp. 143–151. Citeseer (1997)
18. Kelly, S., Tolvanen, J.: Domain-Specific Modeling-Enabling Full Code Generation. Wiley (2008)
19. Kelly, S., Lyytinen, K., Rossi, M.: MetaEdit+ A fully configurable multi-user and multi-tool CASE and CAME environment. In: Constantopoulos, P., Mylopoulos, J., Vassiliou, Y. (eds.) Advanced Information Systems Engineering, pp. 1–21. Springer, Berlin Heidelberg, Berlin, Heidelberg (1996)
20. Langefors, B.: Theoretical analysis of information systems. Studentliteraturen (1967)
21. Marcén, A.C., Iglesias, A., Lapeña, R., Pérez, F., Cetina, C.: A systematic literature review of model-driven engineering using machine learning. IEEE Trans. Softw. Eng. (2024)
22. McInnes, L., Healy, J., Melville, J.: Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426 (2018)
23. Mellor, S.J.: MDA Distilled: Principles of Model-driven Architecture. Addison-Wesley (2004)
24. Mendling, J.: Business process modeling in the 1920s and 1930s as reflected in fritz nordsieck's PhD thesis. Enterprise Modelling and Information Systems Architectures, **16** (2021). https://doi.org/10.18417/emisa.16.6.
25. Michael, J., Bork, D., Wimmer, M., Mayr, H.C.: Quo vadis modeling? Findings of a community survey, an ad-hoc bibliometric analysis, and expert interviews on data, process, and software modeling. Softw. Syst. Model. **23**(1), 7–28 (2024)
26. Murtagh, F., Contreras, P.: Algorithms for hierarchical clustering: an overview. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. **2**(1), 86–97 (2012)
27. Mylopoulos, J.: Conceptual modelling and telos. Conceptual modelling, databases, and CASE: an integrated view of information system development pp. 49–68 (1992)
28. Reimers, N.: Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084 (2019)
29. Robertson, S., Zaragoza, H.: The probabilistic relevance framework: BM25 and beyond. Found. Trends Inf. Retr. **3**(4), 333–389 (2009). https://doi.org/10.1561/1500000019
30. Schmidt, D.C., et al.: Model-driven engineering. Comput.-IEEE Comput. Soc. **39**(2), 25 (2006)
31. Silver, B.: BPMN Method and Style. Cody-Cassidy Press (2009)
32. ...van der Aalst, W.M.P., Becker, J., Bichler, M., Buhl, H.U., Dibbern, J., Frank, U., Hasenkamp, U., Heinzl, A., Hinz, O., Hui, K.L., Jarke, M., Karagiannis, D., Kliewer, N., König, W., Mendling, J., Mertens, P., Rossi, M., Voß, S., Weinhardt, C., Winter, R., Zdravkovic, J.: Views on the past, present, and future of business and information systems engineering. Bus. Inf. Syst. Eng. **60**(6), 443–477 (2018). https://doi.org/10.1007/s12599-018-0561-1
33. Wand, Y., Weber, R.: An ontological model of an information system. IEEE Trans. Softw. Eng. **16**(11), 1282–1292 (1990)

Iris Reinhartz-Berger is an associate professor in the Department of Information Systems at the University of Haifa, Israel. She holds a BSc in Computer Science and Applied Mathematics, and an MSc and PhD in Information Management Engineering, all from the Technion—Israel Institute of Technology. Her research interests include conceptual modeling, domain analysis, model-driven engineering, and systems development processes. Her recent work focuses on the impact of emerging technologies, particularly large language models, on software engineering practice and modeling. She co-chaired the EMMSAD conference from 2017 to 2022 and currently serves on its steering committee. She is an Associate Editor for several journals, including Software and System Modeling (SoSyM), Data and Knowledge Engineering, Requirements Engineering, and Business and Information Systems Engineering. She has also held program chair and other organizational roles at different international conferences, including CAiSE, SPLC, ER, and MODELS.

Adir Solomon has been an Assistant Professor (Lecturer) since 2023 in the Department of Information Systems at the University of Haifa. He received his PhD in 2022 from Ben-Gurion University of the Negev and completed a postdoctoral fellowship (2022–2023) at KU Leuven's Research Center for Information Systems Engineering. His research focuses on user modeling, recommender systems, and natural language processing. Dr. Solomon uses machine learning and deep learning techniques to enhance personalization, improve communication in healthcare, and mitigate biases in AI systems, with a particular emphasis on morphologically rich languages such as Hebrew. He has published in leading journals and conferences and actively contributes to the research community through editorial and program committee roles.

Jelena Zdravkovic is a Professor at the Department of Computer and Systems Sciences (DSV) at Stockholm University. She earned a PhD in Computer and Systems Sciences at The Royal Institute of Technology (KTH) in 2006. Jelena has published over 100 refereed papers on the topic of IS engineering in international conferences and scientific journals. Her current research activities include data-driven requirements engineering with a focus on digital ecosystems and digital twins. Jelena has participated in several national and international projects related to requirements for IS, and model-driven IS development. Jelena was on the editorial board of the BISE journal for about 10 years and is currently associate editor for the RE Journal, and guest editor for the SoSyM and DKE journals; as well as she is serving as a regular reviewer for them, and others, including Elsevier's Journal of Systems and Software, the Information and Software Technology Journal, and IEEE Computing. She has served as the program chair and is on the program board of several international conferences, including CAiSE, ER, RCIS, PoEM, and as a program committee member in a number of other conferences and workshops in the IS engineering discipline.

John Krogstie (1967) holds a PhD (1995) and a MSc (1991) in information systems from the Norwegian University of Science and Technology (NTNU), where he is currently a full professor in information systems at the computer science department (IDI). At IDI, he was Department Head 2017–2021. Director NTNU Digital 2022–2023 and Center for Sustainable ICT—NTNU 2022. John Krogstie is the Norwegian representative and previously Vice-Chair for IFIP TC8 and was chair of IFIP WG 8.1 on information system design and evaluations (2010–2015). He was co-arranger of EMMSAD 2001–2014 and is currently part of the steering committee. His research interests are information systems development, information systems engineering, quality of models and modeling languages, neuro-conceptualization, eGovernment, sustainable smart cities, and sustainable digitalization in general. He has published more than 400 refereed papers in journals, books, and archival proceedings since 1991.

Henderik A. Proper Erik for friends is Full Professor in Enterprise and Process Engineering in the Business Informatics Group at the TU Wien. Erik has a mixed background, covering a variety of roles in both academia and industry. His core research drive is the development of "theories that work," i.e., research that leads to results that have both scientific rigor and practical relevance. His general research interest concerns the foundations and applications of domain modeling, in particular in the context of enterprises. Over the past 25 years, he has applied this research drive and general research interest toward the further development of the field of enterprise engineering, and enterprise modeling in particular. Presently, Erik is chair of the IFIP 8.1 working group, while also being the representative for the Netherlands in IFIP's TC8 technical committee. He is also the Stellvertretender Sprecher of the EMISA working group of the German Computer Science Society (Gesellschaft für Informatik).