Emerging Concepts and Trends in Collaborative Modeling: A Survey

Matthew Stephan
Dept. of Computer Science and Software Engineering, Miami University, 510 East High Street, Oxford Ohio, U.S.A.

Keywords: Model-Driven Engineering, Modeling for the Cloud, Collaborative Modeling, Collaborative Software Engineering.

Abstract: Just as in other engineering disciplines, software engineering is well suited to collaboration; Having different perspectives and diverse experiences strengthens engineering projects. Software modeling is a fundamental aspect of software engineering and is becoming increasingly collaborative. Collaborative modeling approaches are maturing and related research is growing significantly. While surveys exist on collaborative modeling tools and research, they are aimed at academics and can be verbose. In this article, we conduct a research survey intended to provide practitioners and researchers an accessible and abstract at-a-glance perspective of emerging trends and directions in collaborative modeling. We complete a systematic literature review, which we crosscheck with existing surveys. To explicate trends in the last five complete years and overall trends, we perform concept extraction and domain analysis by analyzing abstracts. We visualize these trends in word clouds and trend charts, and provide insights. We hope this article helps spread awareness of collaborative modeling trends and future directions, and educates practitioners and researchers.

1 INTRODUCTION

Software modeling is an integral part of software engineering. For complex software systems, Model-Driven Engineering (MDE) is one approach that helps deal with accidental complexity and allows designers to work with artifacts at an abstraction level closer to their problem domains (Mellor et al., 2003). MDE involves creating software models that act as the primary artifacts throughout all phases within the software engineering life cycle. Whether organizations are employing software models in formal MDE approaches or informally, such as in design decisions and inception prototyping, software modeling is increasingly prevalent in industry (Hutchinson et al., 2011).

Engineering, by its nature, is suited to collaboration as multiple minds with different perspectives can serve only to improve results. This holds true for collaborative modeling, where many software engineers work together to develop models. However, given the advent of Semantic Web technologies (McIlraith et al., 2001) and the increasing importance being placed on collaboration in general, new collaborative research projects are emerging. Thus, our goal is to provide an abstract overview of recent collaborative modeling research to identify temporal trends. While other surveys exist on collaborative modeling, ours provides a temporal perspective, discovers emerging concepts, and identifies trends. We present emerging concepts in collaborative modeling research from 1) the most recent complete 5 years (2012-2017) and 2) all time, and contrast these to explicate trends and directions. We additionally pick 2012 as a starting point because it was the first year of a dedicated workshop on the subject (Ober et al., 2012), indicating a new focus and importance placed on collaborative modeling by the research community.

We aim to answer the following questions we believe of use to practitioners and researchers,

1. What are the emerging concepts in collaborative modeling within the last 5 complete years and all time?
2. In what way has the focus on these concepts changed over time?

To provide a foundation and the data for answering these questions we perform a systematic literature review (SLR) of collaborative modeling. This includes following a specific, constrained, and repeatable search process. To ascertain trends, we analyze abstracts of papers to determine concept frequency and conduct domain analysis (Paraschiv et al., 2015). We visualize this analysis in the form of word clouds

Stephan, M.
Emerging Concepts and Trends in Collaborative Modeling: A Survey.
DOI: 10.5220/0007255502400247
In *Proceedings of the 7th International Conference on Model-Driven Engineering and Software Development (MODELSWARD 2019)*, pages 240-247
ISBN: 978-989-758-358-2

and trend charts to illustrate temporal content evolution (Cui et al., 2010). We then provide an overview of notable emerging trends and discuss these findings. We begin with a presentation of related surveys in Section 2, and follow with our survey protocol in Section 3. We summarize our data in Section 4 including prevalent topical words, quantitative data, and our word clouds. In Section 5, we discuss the changing trends and interpret the data to provide some of our insights and thoughts. We conclude the paper in Section 6.

2 RELATED WORK

In this section, we discuss related secondary studies/surveys to this paper. While there are several studies that provide different perspectives on collaborative modeling (Renger et al., 2008; Franzago et al., 2017; Rocha et al., 2011; Portillo-Rodríguez et al., 2012), none of them consider trends nor temporal aspects. The most complete, recent, and relevant to ours is that of Micro Franzago et al. (Franzago et al., 2017) published in 2017. They classify existing techniques and discuss 48 primary studies spanning 19 years. Our survey has a different purpose and provides a different perspective, that being a temporal focus and comparison to provide an abstract at-a-glance view and discussion of trends. We were thus more inclusive than they were, as we will describe in our protocol. However, we did cross-reference their paper to ensure we did not miss any primary studies. As a result, our study includes their 48 primary studies and more.

3 SURVEY PROTOCOL

While our goal was to provide an abstract act-a-glance view of trends, we still attempted to follow established guidelines for conducting software engineering systematic literature reviews (Kitchenham and Charters, 2007). This helps ensure that others can repeat and reproduce our work.

3.1 Research Question

To help frame our research question, we use Petticrew and Roberts' PICOC criteria: (Petticrew and Roberts, 2008):

- **Population:**
  1) Practitioners interested in comparing and employing collaborative modeling who want a temporal perspective illustrating what direction research/tooling is heading.
  2) Researchers interested in developing new collaborative modeling approaches and seeing a) what areas are gaining the most traction, and b) if the trends are appropriate or if new topics should be addressed.
- **Intervention:** Techniques intended to facilitate coordination, communication, and collaboration (3C) among those employing software modeling and MDE.
- **Comparison:** Not applicable. To identify concepts and trends, we consider all techniques.
- **Outcomes:** Research and approaches that facilitate or support collaboration in MDE.
- **Context:** Published academic research.

This allowed us to form the research questions of this paper, which are a refinement of the more general question we phrased earlier:

1. What are the emerging concepts in all collaborative modeling research articles?
2. What are the emerging concepts in collaborative modeling research articles published from 2012 until 2017, inclusively?
3. What are the (changing) trends implied by contrasting these two sets of concepts?

3.2 Search Strategy

Our search strategy included all major online libraries relevant to this domain (Kitchenham and Charters, 2007). This included IEEExplore, ACM Digital Library, Google Scholar, and ScienceDirect. We additionally consider all proceedings from the "International Workshop on Model-Driven Engineering on and for the Cloud" and the "International Workshop on Collaborative Modelling in MDE". Lastly, we include/cross-check our list with all primary articles identified in the survey performed by Micro Franzago et al. (Franzago et al., 2017) to help ensure article coverage.

Our search string includes any entries containing the term "collaborative modeling" in the metadata. The search string further included articles with metadata containing the words "collaboration" or "collaborative" with the union of any of "MDE", "Model driven engineering", "software modeling", and "domain specific modeling language". We included both spellings of "modeling" and "modelling" in all terms, as recommended in the SLR guidelines (Kitchenham and Charters, 2007) when discussing word variations. While the exact query we

used varied depending on which library we were considering, it appeared roughly in the form of the following exact query from our IEEE library search, *("collaborative modelling" OR "collaborative modeling") OR (("Document Title":collaboration OR "Document Title":collaborative) AND ("Document Title":MDE OR "Document Title":"Model driven engineering" OR "Document Title":"software modeling" OR "Document Title":"software modelling" OR "Document Title":"domain specific modeling language" OR "Document Title":"domain specific modelling language"))*.

3.3 Study Selection Criteria

Our study selection criteria applied to the search results first ensures that each primary study describes collaborative software modeling tooling and research. Our initial criteria involves us including works from all years. From that list, we identified those published over the five year period of 2012 to 2017. We chose not to include the current year, 2018, as 1) the year is not complete, 2) it is possible that articles published earlier this year may not have been disseminated or included in libraries yet. Thus, in order to not make the overly-bold claim of including 2018 and present potentially skewed data, we do not include any articles from 2018. We excluded any non-English articles. We excluded any existing surveys, but considered their content when evaluating our search terms and results.

3.4 Study Selection Procedure

The author of this paper performed the study extraction and collection using the search terms. Inclusion/exclusion was performed through careful review of the abstract and paper.

3.5 Study Quality Assessment Procedure

Our *inclusion criteria* were purposely less strict than the Franzago et al. survey, as we 1) did not require our articles to include study evaluation and 2) included Master's and Ph.D. theses. We argue these are relevant articles for our context and intention of considering trends and directions. That is, evaluation/validation and being thesis work should not be a reason to exclude articles when considering emerging topics and trends. Our *exclusion criteria* was any work not addressing methodologies and techniques for collaborative software modeling as we described in our PICOC intervention. Some examples of work we found using our query that we excluded includes primary studies on domain languages for generating collaborative software; secondary studies, which we described earlier; research dealing with hardware/circuity employing software instead of software development, and others.

3.6 Data Extraction Strategy and Synthesis

To discover emerging concepts in collaborative modeling, we decided to analyze articles' abstracts. We decided on abstracts as they are intended to cover an article's important concepts and keywords, and, as argued by Paraschiv et al., form a good basis and data for domain analysis (Paraschiv et al., 2015). We began by entering all the articles' metadata into our bibliography manager and verifying the entries manually. We then made a copy of that bibliography manager list, and updated that copy to include entries from 2012 until 2017 only. For both lists, we extracted out the abstracts into separate text files. We then passed these text files to word cloud generator software [1], which generates both word clouds and associated word counts. Word clouds are not perfect, however, so we took some additional steps (Harris, 2011). Our software allowed us to remove words from the word cloud and counts. To help eliminate noise, we filtered out context-establishing words such as "collaborative", "modeling", "software", "paper", "engineering", and others. We additionally filtered out non-conceptual words such as "can", "use", "work", "present", and others. We counted the amount of top words in distinct abstracts to present as data. For reproducibility, readers can find the bibliography files and abstracts, the list of included words and their counts, and the removed words and their counts on our associated web repository [2]. For the sake of brevity and paper-length constraints, we list all our primary studies on that web repository rather than in this paper.

3.7 Results of Protocol

We breakdown how many papers we found in each of the most prominent libraries in Table 1. Using our protocol, we discovered 58 articles published in 2012-2017, 45 articles published before 2012, and 103 articles overall. All of the metadata for these articles can be accessed on our associated web repository as further reading.

[1] https://www.wordclouds.com/

[2] https://sc.lib.miamioh.edu/handle/2374.MIA/6283

Table 1: Primary Studies Found in Main Libraries.

Library	# of Primary Studies
IEEE	137
ACM	157
Science Direct	45
Google Scholar	50

4 SUMMARY OF TREND DATA

We present the quantitative results of our findings here and discuss them in the next section. While word clouds may make it difficult to search for words of interest to readers, we present our complete lists as further reading on our associated web page.

4.1 Articles from 2012 - 2017

We provide a word cloud for collaborative modeling articles from 2012 until 2017 in Figure 1. Top topical words include “different”, “conflicts”, “control”, “management”, “detection”, “requirements”, “quality”, “stakeholders”, “conflict”, “version”. We illustrate their prevalence in Table 2. The list is sorted by the first column, which indicates the total number of word occurrences in all of the abstracts’ text. The second column illustrates the amount of distinct abstracts the word was found, and the third column indicates the percentage, rounded to two decimal places, of article abstracts the word appeared in out of the 58 articles.

4.2 All Articles

Readers can find a word cloud representing all the collaborative modeling articles until 2017 in Figure 2. Prevalent topical words found in the abstracts in all articles include “UML”, “different”, “control”, “management”, “version”, “integrated”, “merging”, “CASE”, “architecture”. We list these words in Table 3, sorted by the first column indicating the total number of instances of the words in all abstracts. The second column identifies the number of distinct articles with which the word appears, and the final column presents the percentage of all 103 article abstracts that contain the word.

5 CHANGING TRENDS

To discuss changing trends, we begin by contrasting the two word clouds and total counts. From a potential bias perspective, it is important to consider there are 45 articles before 2012 and 58 articles after. Thus, it is possible concepts in abstracts from 2012 until 2017 could be inflated. We could have chosen a different end year, however, we were interested in isolating recent, 5-year, research to provide a snapshot of the most recent complete half decade. The abstracts’ sizes present a minor bias/threat to validity as the abstracts we found ranged from 150-250 words, with the majority of abstracts being in the 250 upper limit range. We identify more threats to validity later in this paper.

We illustrate some example comparisons in Figure 3 using a 100% stacked column chart. Specifically, we showcase words that were prominent on one list and not the other. We also include one word, “version”, which was about an equal split, as a middle baseline and interesting example. The darker shade represents the percentage of instances of each respective word we found in article abstracts from articles prior to the year 2012. The lighter shade represents the percentage of instances of the respective word found in article abstracts in articles from 2012-2017. Therefore, columns that are darker represent concepts that were more prevalent before 2012. Lighter columns represent concepts that emerged and gained prevalence in the years of 2012 to 2017.

5.1 Discussion

Unsurprisingly, we note that CASE model collaboration research was conducted exclusively prior to 2012, which can be explained by the evolution of CASE into object oriented approaches. We also note that UML specific research was much more predominate prior to 2012. Franzago et al. state that there is “a prominence of UML-based approaches” (Franzago et al., 2017), however, we take that one step further and note that “prominence” is outdated as we discovered the recent research focus is becoming more language agnostic. They also note that there are relatively few approaches that consider the interaction among synchronous and asynchronous collaboration. Our data further indicates this trend is likely to continue as the majority of both synchronous and asynchronous research occurred prior to 2012. Specifically, we found 20 instances of either topic prior to 2012 and only 5 instances of those topics in 2012-2017 articles.

On the other end of the spectrum, we observe a number of collaborative modeling research concepts emerging in the last five complete years. While the prevalence of “version” and “different” were roughly the same before and after 2012, it seems the notions of “conflicts” and “inconsistencies” have become more

Table 2: Top Words in Abstracts from Articles in 2012-2017.

Word	# of Total Instances	# of Articles Word is Found	% of 58 Articles
Different	23	19	33%
Conflicts	22	11	19%
Control	21	11	19%
Management	14	11	19%
Detection	13	6	10%
Requirements	12	8	14%
Quality	12	9	16%
Stakeholders	11	10	17%
Conflict	11	6	10%
Version	10	9	16%

Figure 1: Emerging Concepts in Abstracts from Articles Published 2012 to 2017.

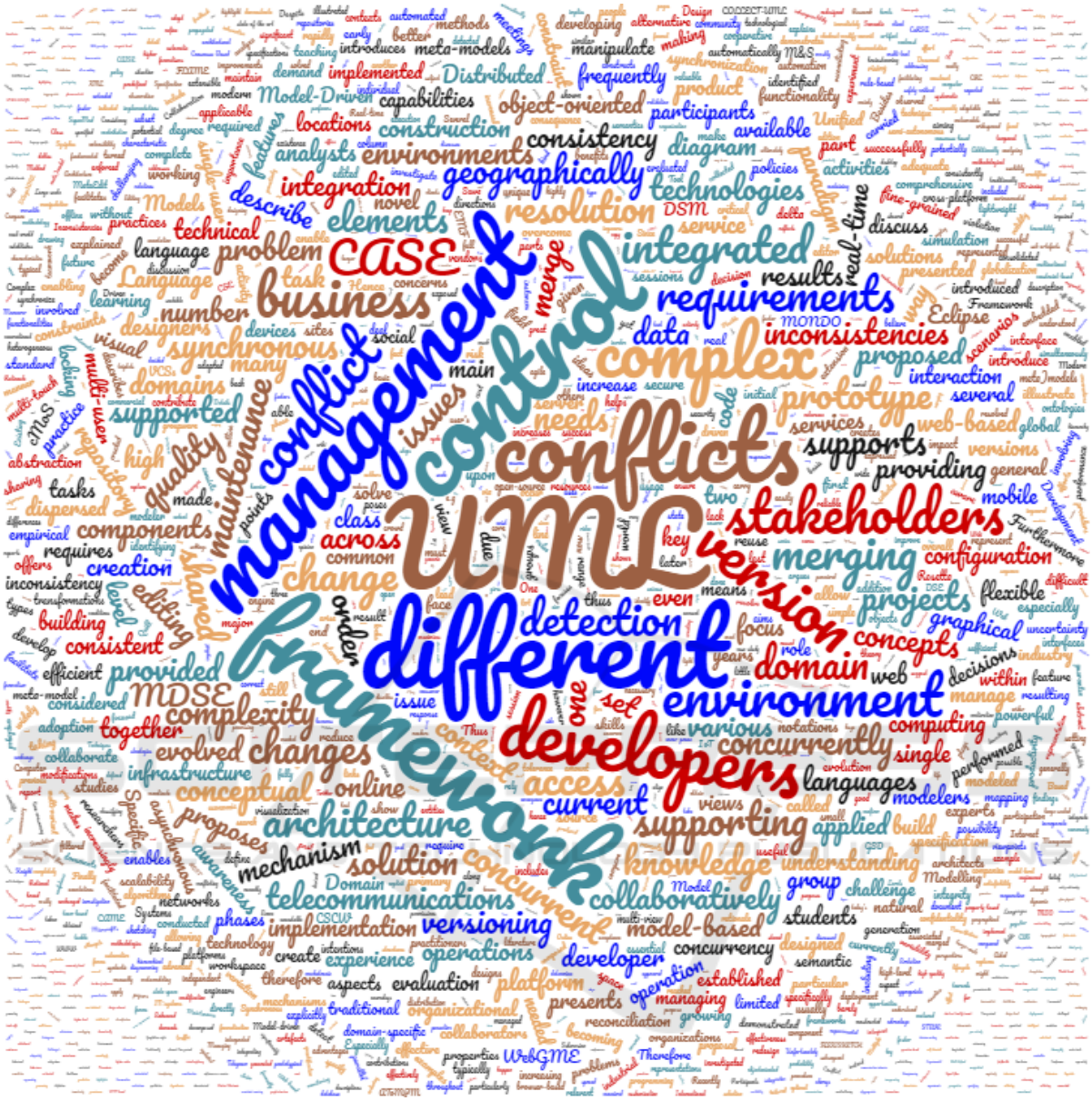

Figure 2: Emerging Concepts in Abstracts from All Articles.

Table 3: Top Words in Abstracts from All Articles.

Word	# of Total Instances	# of Articles Word is Found	% of 103 Articles
UML	42	21	20%
Different	39	30	29%
Control	38	22	21%
Management	35	21	20%
Conflicts	34	17	17%
Version	23	16	16%
Integrated	17	12	12%
Merging	17	10	10%
Case	17	6	6%
Architecture	16	11	11%

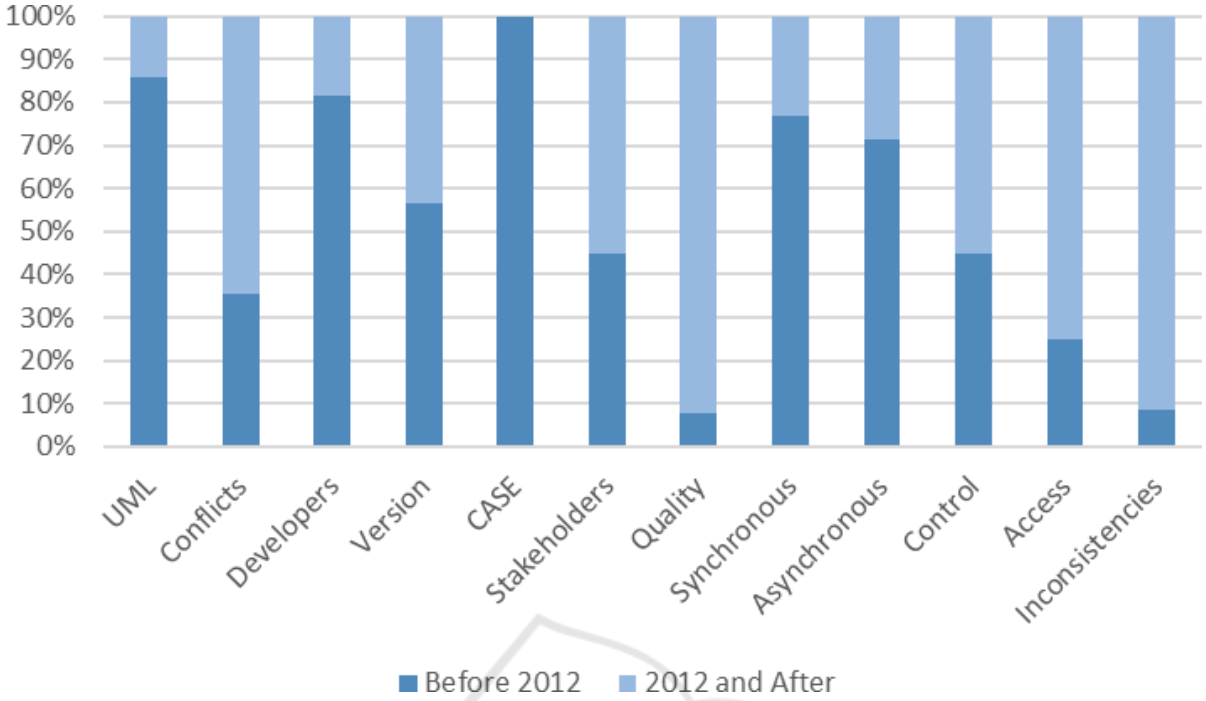

Figure 3: A 100% Stacked Column Chart of Abstract Word Occurrence Before 2012 and After.

of a focus in recent years. This can potentially be explained by the need for conflict resolution arising from the advent of the more advanced systems being built by organizations and the increasingly intricate and necessary collaborations experienced by modelers. It also coincides with the increase in model comparison research (Stephan and Cordy, 2012; Stephan and Cordy, 2013), which involves detecting differences and conflicts among a set of models. Our trend chart also demonstrates an increase in the need for assessing and establishing "quality" in collaborative modeling. This corresponds with the increase in research and industrial desire for quality evaluation in software modeling in general (Mohagheghi and Aagedal, 2007; Giraldo et al., 2016). Security, an aspect of quality, demonstrates the same trend, as "access" and "control" concepts are more prevalent in the post 2011 research. This is not surprising given the high security focus that many organizations are employing due to very public software security breaches in the media (Sen, 2018). Another trending concept, not shown in our chart, is "telecommunications". All research referencing telecommunications was found to be published exclusively in 2012 or later. An interesting observation is the shift of focus from "developers" to "stakeholders". While stakeholders is an all-encompassing term that potentially includes developers, this may also reflect the stakeholder/customer driven Agile software development methodology growth (Abrahamsson et al., 2002) where methods and processes are focused on all stakeholders, rather than just the developers.

5.1.1 Limitations and Threats to Validity

One aspect we did not consider necessary was to employ a taxonomy to help characterize/categorize and index the results. While Franzago's study included one (Franzago et al., 2017), it made more sense for their purposes of a mapping study intended to identify "white spots" to warrant further research. That is, our study was more reflective, and theirs was more prescriptive. Secondly, we limited our paper to word clouds based on abstracts, which limited the paper to a very abstract level. That was our intention, but that does carry with it some potential threats to our trend data. While we contend it is useful for those at the managerial level and practitioners, it could be less helpful for the latter as our data is abstract. It may be a technically implemented survey, but we conducted it in an original, and temporally-focused, manner and believe it has value for the reader and collaborative modeling community. At the very least, readers can 1) access our primary study metadata on our repository as a good source of information, 2) benefit from an abstract view of trends and contrast of different time periods. Lastly, we acknowledged that we purposely left out 2018 data for the reasons we discussed in our study selection section.

6 CONCLUSION

Based on the sheer number of articles we found in our systematic review from 2012 to 2017 compared to all articles, it is clear collaborative modeling research is very topical and on the rise. This is additionally supported by the emergence of dedicated workshops such as the "International Workshop on Collaborative Modelling in MDE" and explicit sessions dedicated to modeling at conferences. We learned that, while the importance of model version control has remained consistent, there has been a relatively recent influx and focus on detecting and dealing with conflicts and inconsistencies. Additionally, the quality of models created through collaborative modeling, and security access in the collaboration process are becoming more paramount. It is likely that the research results from this relatively recent focus will permeate into collaborative tooling, more so than it already has.

While we were concerned with what research was being completed and published, a future goal of both practitioners and researchers must be to better communicate and work together more so that the focus and direction of research is coming from practitioners. In the majority of the research we encountered, it appears that researchers were tackling important problems, but there was no evidence nor support that these were the areas most interesting and desirably to practitioners. The general interests of practitioners and researchers align in theory, and thus should be more connected. It is our hope that this article's at-a-glance overview of emerging trends in collaborative modeling research may inspire practitioners to reach out to researchers to let them know if these trends correspond to their interests or if there is something else researchers should be focusing on to better help serve the software modeling community.

REFERENCES

Abrahamsson, P., Salo, O., Ronkainen, J., and Warsta, J. (2002). Agile software development methods: Review and analysis. Technical Report VTT Publications 478, VTT Technical Research Centre of Finland.

Cui, W., Wu, Y., Liu, S., Wei, F., Zhou, M. X., and Qu, H. (2010). Context preserving dynamic word cloud visualization. In *2010 IEEE Pacific Visualization Symposium (PacificVis)*, pages 121–128.

Franzago, M., Di Ruscio, D., Malavolta, I., and Muccini, H. (2017). Collaborative model-driven software engineering: a classification framework and a research map. *IEEE Transactions on Software Engineering*.

Giraldo, F. D., Espana, S., Pastor, O., and Giraldo, W. J. (2016). Considerations about quality in model-driven engineering. *Software Quality Journal*, pages 1–66.

Harris, J. (2011). Word clouds considered harmful. *Nieman Journalism Lab*.

Hutchinson, J., Rouncefield, M., and Whittle, J. (2011). Model-driven engineering practices in industry. In *Proceedings of the 33rd International Conference on Software Engineering*, pages 633–642. ACM.

Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report Technical Report EBSE-2007-01, School of Computer Science and Mathematics, Keele University.

McIlraith, S. A., Son, T. C., and Zeng, H. (2001). Semantic web services. *IEEE intelligent systems*, 16(2):46–53.

Mellor, S. J., Clark, T., and Futagami, T. (2003). Model-driven development: guest editors' introduction. *IEEE software*, 20(5):14–18.

Mohagheghi, P. and Aagedal, J. (2007). Evaluating quality in model-driven engineering. In *International Workshop on Modeling in Software Engineering (MISE'07: ICSE Workshop 2007)*, pages 1–6.

Ober, I., Gokhale, A., Georg, G., Felderer, M., Bruel, J.-M., and Lugato, D. (2012). Mdhpcl 2012 workshop summary. In *Proceedings of the 1st International Workshop on Model-Driven Engineering for High Performance and CLoud Computing*, MDHPCL '12, pages 1:1–1:2, New York, NY, USA. ACM.

Paraschiv, I. C., Dascalu, M., Trausan-Matu, S., and Dessus, P. (2015). A semantic approach to analyze scientific paper abstracts. In *The International Scientific Conference eLearning and Software for Education*, volume 1, pages 3–9.

Petticrew, M. and Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons.

Portillo-Rodríguez, J., Vizcaíno, A., Piattini, M., and Beecham, S. (2012). Tools used in global software engineering: A systematic mapping review. *Information and Software Technology*, 54(7):663–685.

Renger, M., Kolfschoten, G. L., and De Vreede, G.-J. (2008). Challenges in collaborative modelling: a literature review and research agenda. *International Journal of Simulation and Process Modelling*, 4(3-4):248–263.

Rocha, R. G., Costa, C., Rodrigues, C., Ribeiro de Azevedo, R., Junior, I. H., Meira, S., and Prikladnicki, R. (2011). Collaboration models in distributed software development: A systematic review. *CLEI Electronic Journal*, 14(2):1–2.

Sen, R. (2018). Challenges to cybersecurity: Current state of affairs. *Communications of the Association for Information Systems*, 43(1):2.

Stephan, M. and Cordy, J. R. (2012). A Survey of Methods and Applications of Model Comparison. Technical Report 2011-582, Queen's University. Revision 3.

Stephan, M. and Cordy, J. R. (2013). A Survey of Model Comparison Approaches and Applications. In *International Conference on Model-Driven Engineering and Software Development (Modelsward)*, pages 265–277. SCITEPRESS.