Journey Towards Agility –
A Retro- and Prospective Review

Tim Dreesen
University of Cologne
dreesen@wiso.uni-koeln.de

Phil Diegmann
University of Cologne
diegmann@wiso.uni-koeln.de

Björn Binzer
University of Cologne
binzer@wiso.uni-koeln.de

Christoph Rosenkranz
University of Cologne
rosenkranz@wiso.uni-koeln.de

Abstract

After more than 15 years since the Agile Manifesto and extensive research on agile software development (ASD) for nearly three decades, a comprehensive body of knowledge is available and is constantly growing. ASD is considered an effective way for managing software development projects in environments characterized by rapidly changing requirements. This study aims to shed light on the existing knowledge on ASD by applying a structured literature review and computer aided analysis consisting of distinct text mining techniques. We analyzed a sample of 1,376 papers and provide results from articles among relevant information systems research as well as computer science conferences and journals. Based on our approach, we are able to (1) evaluate key articles and journals, (2) analyze the development of ASD research in the last three decades and, most importantly, (3) identify research foci of the past as well as gaps in our knowledge on ASD.

1. Introduction

Interest in *agile software development (ASD)* methodologies has increased in recent years in both research and industry [19, 28, 38]. Based upon the principles of the Agile Manifesto [10], different implementations, such as Scrum or eXtreme Programming (XP), have emerged and motivated a variety of research.

ASD has been applied to a wide range of projects: from small teams, situated in co-located offices [e.g., 16] to large scale, distributed, or outsourced projects [e.g., 51]. In this context, ASD methodologies and practices have been implemented successfully but also unsuccessfully [38]. Research also has

investigated the customization and configuration of agile approaches, the so-called method tailoring [e.g., 28, 36, 57]. Due to the wide variety of topics covered by ASD research, ranging from rather technical aspects [e.g., 9] to sociological or psychological factors [e.g., 43], and from an individual level to an organizational level [e.g., 62], a clear categorization of existing streams of research is difficult to recognize. Additional difficulties arise because the concept of ASD, its exact definition, and its applicability are debated [19].

Motivated by this, our study's objective is twofold. First, we ask which topics of ASD research have been explored in the past and are currently investigated. Second, we want to identify topics that are not covered in current research and therefore still remain non-existent in extant literature. Consequently, the central research questions guiding our study are: *(1) What research topics have been addressed within the last three decades by ASD research and (2) how do these topics differ in terms of available publications and their distribution over time?*

To answer our research questions, we conducted a structured and comparative literature review as described by the guidelines of Levy and Ellis [39] and Webster and Watson [58], followed by computer-aided topic modeling [5, 21] on the extant body of knowledge of ASD.

The remainder of this paper is structured as follows. We give an overview about related work, targeting research on the field of ASD. Next, we describe our research design being used for data collection and analysis. Following, we present and discuss our findings. Finally, we provide an outlook for and point out future research directions.

HICSS

2. Related Work and Background

2.1. Agile Software Development

In practice, approaches for developing software range from sequential approaches [49] to more cyclic, iterative approaches [12], that is ASD. During the last two decades, ASD methodologies such as eXtreme programming, rapid application development, or rapid prototyping complemented the iterative approach. Additionally, new management concepts associated with ASD, such as Scrum and Lean Software Management, have been proposed.

The four basic principles of the Agile Manifesto [10] can be found in most ASD methodologies. According to the Agile Manifesto, ASD should value individuals and interactions over processes and tools, working software over comprehensive documentation, customer collaboration over contract negotiation, and responding to change over following a plan [10]. Each of these principles have been subject to research in some sort: for instance, in regard to individuals and interactions, research has investigated the effects of communication in ASD teams [34], in regard to working software, extant literature investigated the influence of pair programming on software quality [9], in regard to customer collaboration, the funding process has been studied [17], and the ability to respond to change has been subject of studies as well [28].

Moreover, next to the methodologies themselves, extant research so far has studied individual or organizational phenomena, such as the use and effects of specific agile practices [9, 42], and effects regarding whole projects or organizations, such as the introduction of ASD methodologies to teams [e.g. 16]. Furthermore, the use of hybrid methodologies or the tailoring of agile methodologies to a team's specific needs is covered by extant research [36, 38, 57]. Literature investigating the success and failure of ASD mostly focusses on specific methodologies, such as Scrum or XP [31], or specific practices, for instance pair programming [17]. Extant research focusing on success and failure of ASD in general exists, but is rare [38].

2.2. Existing Literature Reviews

By conducting a structured literature review, we assessed the current state of research regarding summarizing and aggregating literature reviews. We searched for articles containing "literature" and "review" as well as synonyms for ASD (i.e., scrum, xp or kanban) in the title, abstract, or keywords. The search was limited to a timeframe up to and including August 2016 and the outlets of the "Senior Scholars' Basket of Journals" edited by the Association for Information Systems and top conferences. We finished the search process with a resulting set of 15 relevant papers, of which none did a historic-holistic approach, meaning each of the structured reviews does not necessarily considered all agile methodologies, an explicit focus on software development or a broader and up-to-date timeframe. Instead, they focused on a specific field of interest, such as software engineering for ubiquitous systems [e.g., 33], individual acceptance, tailoring, or use of agile methods and practices [e.g. 14, 35], general practices and challenges in agile requirements engineering [e.g., 35], or geographically distributed, large scale ASD and agility [7, 23].

We can therefore conclude that few summarizing or aggregating literature reviews on the field of ASD research exists and that those articles are oftentimes specialized and limited in scope. For instance, ASD has been included in a summary for information systems offshoring [53]. Other aggregating or summarizing literature focuses on the concept of agility itself [19], but only few provide an overview about existing studies [e.g., 24, 26]. In sum, a clear categorization of existing streams of research is difficult to recognize.

3. Research Method

3.1. First Phase: Structured Literature Review

The approach of a structured literature review is chosen because of its applicability to gain an overview of the field and extant research and help to identify research gaps [56]. The low number of review articles that are being published in the field further motivates the approach [48, 58]. Reviews are often a means to expose emerging issues to potential theoretical foundations, and because ASD itself is still a continuously emerging topic [24], this review aims at analyzing the extant research literature to summarize what has already been researched and what is left to be examined. To provide a comprehensive overview on current ASD topics and those topics that still have to be investigated, the existing literature is thoroughly examined, using a structured approach by following the guidelines of Levy and Ellis [39] and Webster and Watson [58].

Initially, our data collection process started by performing an extensive keyword search within leading journals. We set a focus on primarily high quality, peer-reviewed literature, published in journals of the "Senior Scholars' Basket of Journals"

and the AIS Toplist (including leading journals not only from IS but also Management and Computer Science). Additionally, we included articles of prominent conferences (e.g., HICSS, ICIS, ICSE). A complete set of all outlets is available from the authors on request. We defined a single search string for our keyword search to identify relevant articles in different databases: TIKEAB[1]:(software OR "information system") AND TIKEAB:(development OR engineering OR maintenance OR method* OR practic*) AND (TIKEAB:(agil* OR SCRUM OR XP OR "Extreme Programming" OR Kanban) NOT TIKE:(manufac*)) with TIKEAB searching in title ("TI"), keywords ("KE"), and abstract ("AB") and TIKE searching in title ("TI") and keywords ("KE").

As we aimed for an as broad and holistic overview as possible, we only applied minimal include and exclude criteria. We excluded those publications, which were either difficult to automatically analyze via text mining (e.g., non-English language or with no full text available) or which were not research-focused (e.g., an opinion or commentary). We decided to use a restriction for the publishing year of the articles, thus, articles that were published between January 1st, 1985 and December 31th, 2017 were included. January 1st, 1985 was chosen because the first article we found was from 1985 and all data was collected in August 2016, which is why we chose December 31st, 2017 as cap. Within the resulting set of papers, we further identified relevant articles for our project purpose ("in scope", i.e., investigating ASD) and dropped the others ("not in scope", i.e., not investigating ASD).

In total, after removing duplicates, our final set of articles consists of 678 articles matching our search indicators for ASD in journals and 698 articles in conference proceedings, totaling up to 1,376 articles. Further information concerning the distribution of results can be seen in Table 1.

Table 1. Distribution of results across top-five outlets of each type

Outlet	#
Conferences	
International Conference on Software Engineering	139
XP / Agile	132
Hawaii International Conference on System Sciences	98
International Conference on Global Software	52

Engineering	
Americas Conference on Information Systems	49
Journals	
IEEE Software	187
Journal of Systems and Software	99
Information and Software Technology	73
Computer	28
Communications of the ACM	23

3.2. Second Phase: Computer Aided Analysis

Following to the data collection, we analyzed all articles with the help of *Scikit-learn* [46], a computer-aided analysis and text mining tool. From within the Scikit-learn suite of machine learning tools, we specifically used *topic modelling* [5, 21], which uncovers topics shared by different articles. We use this technique to easily discover topics shared across research and therefore to help in answering our research questions. Research found text mining and especially topic modelling to be helpful in discovering hidden topics by classifying, summarizing, and clustering of text [41, 52] and topic trends over time [6]. This semi-automated approach is especially helpful in analyzing large amounts of text [41, 52].

In order to analyze the extracted data, we first had to convert the articles into a compatible format by extracting text where available or by applying optical character recognition where no text was directly accessible. Furthermore, we annotated the extracted text with additional information, such as author, year, title, and outlet to enable further reaching analysis.

Following the data preparation, we utilized *Latent Dirichlet Allocation* [LDA; 11] as implemented in Scikit-learn as a specific topic modeling approach. Within LDA, each document is seen as a mixture of different topics and each topic has certain probabilities of generating keywords. Keywords are allowed to occur in more than one topic. LDA has been used in various research studies [e.g. 18] and has been suggested as a suitable and helpful tool for research [21].

A too high number of topics to extract might lead to an excessive number of meaningless topics and a too low number might constrain the results unnecessarily; thus, the number of topics to be extracted is the most crucial parameter of the analysis [21]. Therefore, we used four different algorithms [8, 15, 32, 45] aimed at evaluating the quality of topic models to decide which number of topics leads to the optimal topic model. After testing and evaluating different numbers of topics, we settled on 34 topics,

---

[1] Concatenation describing the focus of the keyword search, for example "TIKEAB" indicates searches within title ("TI"), keywords ("KE") and abstract ("AB")

as it provided differentiated topics. Of these topics, 8 topics were discarded, as they covered less than 0.5% of all tokens (i.e., text), resulting in a final set of 26 topics. Furthermore, we decided against the use of lemmatization or stemming to avoid misleading keywords (e.g., "agil" instead of "agility" or "agile"). We opted to use n-grams (i.e., creation of consecutive words such as "agile software development"; in this setting, we decided to use 3-grams) to reduce the number of words with identical meanings but different lexical representations. To further refine the results, we used a list of stop words, which consisted of frequently found words, which added no meaning, such as "et al." or "journal". A complete list of all stop words used within our analysis can be provided by the authors on request.

3.3. Third Phase: Coding

Following Saldaña [50] we applied different coding strategies as an exploratory problem-solving technique and to link our keywords to patterns, resulting in meaningful topic descriptions. At the core is the task of conceptualization, that is, "the process of grouping similar items according to some defined properties and giving the items a name that stands for that common link" [54]. As coding can be seen as cyclical [50], our coding process therefore can be distinguished between a first cycle coding and second cycle coding phase.

During the first cycle coding we started with "descriptive coding". Descriptive coding primarily leads to a categorized summary of the data's contents and builds the groundwork for second cycle coding and further analyses [61]. All authors independently and individually made use of descriptive coding and compared all resulting topics against each other by comparing the included keywords per topic. Based on the keywords, a summarizing phrase was suggested. In case of matching topic phrases, no further action was needed. In case of differing topic phrases, the reasoning for each phrase was compared and alternatives were discussed. Subsequently, descriptive coding for differing phrases was repeated and consensus was reached.

We then applied "pattern coding" as a second cycle coding method. Pattern coding is appropriate for the development of major themes from data [44, 50]. These codes are helpful for aggregating and grouping themes into a smaller number of sets, themes, or constructs [44]. Similar to first cycle coding, we then tried to group our descriptive codes into meaningful pattern codes – again first individually, followed by a discussion where needed.

Again, pattern coding was conducted twice until consensus was reached.

We completed the coding process within a final step, in which we did some post-coding activities such as fine-tuning of the wording and alphabetical order of the results. The outcome of the coding process is a final set of 26 topics and eight topic groups.

4. Results

Figure 1 displays the number of articles published per year, as well as the number of articles published each year in the Senior Scholars' Basket. In Figure 2 a further distinction between publications focusing on either computer science or information systems research is made. To get more into detail, Table 1 shows the number of papers found for each outlet with at least five publications. Conferences and journals are displayed separately, but each are ranked by the number of publications in descending order. Table 2 lists our identified topics, the topic groups, the keywords contained in each topic, and the rank in terms of frequency of the individual topics. As can be seen from Table 2, we identified several topic groups because of the different foci of the topics themselves: while some topics comprise more general information such as concepts, principles, or methodologies related to ASD (see "Agile Methodology & Practice Usage" or "General"), others focus on an organizational perspective and link agile principles such as flexibility or agility to different contexts (see "IT Capability & Agility"); still others focus on managerial implications (see "Business & Environmental Factors") or put emphasis on certain aspects such as social or team related aspects and requirements engineering (see "Teams & Team Management" or "Stakeholders & Requirements Engineering") or risks and success factors (see "Risk, Control & Success Factors in Agile"). Furthermore, we identified a topic group containing research regarding technological aspects (see "Technologies & Applications in Agile").
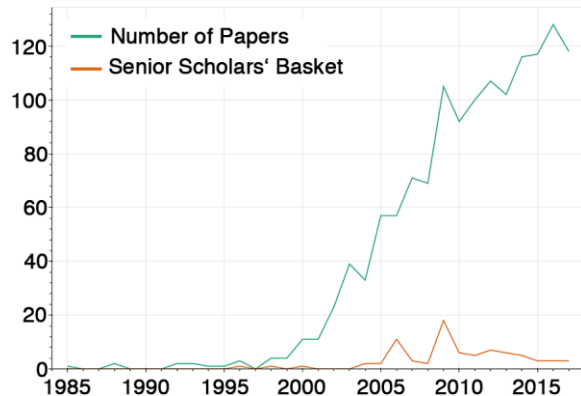
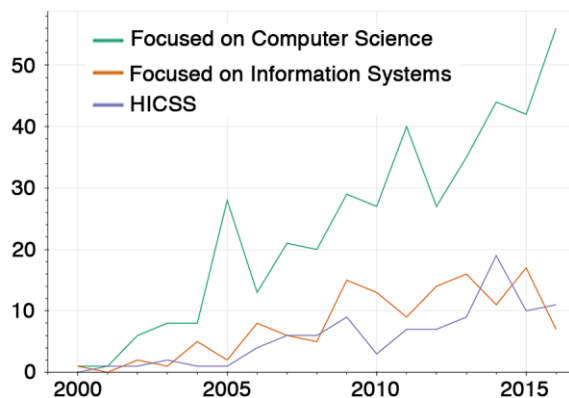**Figure 1. Number of papers published per year.**

Figure 2. Number of papers in computer science or information systems focused outlets and HICSS per year.

Table 2. Identified topics, groups and ranks

Group		Topic	Rank
Agile Methodology & Practice Usage	1	Lean	2
	2	Large-Scale ASD	5
	3	Agile Architecture & Design	7
	4	Scrum	13
	5	Tests & Test-Driven-Development	14
	6	Pair Programming	17
	7	Extreme Programming	20
	8	Documentation	23
Business & Environmental Factors	9	Open Source	18
	10	Business, Transformation, Rules	25
	11	Technical Debt	26
General	12	Theory in ASD	1
IT-Capability & Agility	13	Supply Chain, Agility, Capabilities	9
Risk, Control, & Success Factors in Agile	14	Risk Management, Outsourcing, Project Management	8
	15	Effort Estimation, Success Metrics	10
	16	Control	19
Stakeholders & Requirements Engineering	17	User Participation & Design	4
	18	Requirements Engineering & Stakeholder Management	6
	19	Roles in ASD	11
	20	Requirements, Interdependencies, Prioritization	22
Teams & Team Management	21	Teams & Kanban	3
	22	Teaching and Learning Agile	12
	23	Communication in Distributed ASD	16
	24	Decision Making in ASD	21
Technologies & Applications in Agile	25	Cloud, Services, Security	15
	26	Big Data	24

4.1. Research Foci Over the Last Decades

Although at first glance our topics presented in Table 2 seem to randomly comprise a lot of different and wide spread themes, further investigation and analysis of our results reveal distinct and meaningful patterns. The resulting topics, consisting of specific keywords, are overlapping but each one of them has its "raison d'être", as they represent themes that have been addressed in ASD research within the last decades.

As can be seen from Table 2, the first topic group, "Agile Methodology & Practice Usage", summarizes the "basics" of ASD. The keywords are centered around ASD methods, concepts, practices, management, and tasks. The second topic "Business & Environmental Factors" deals mainly with distinct business contexts such as open source while topic three contains a more general, theoretical perspective on ASD. The fourth topic group, "IT Capability & Agility", relates to a broader view on agile, namely organizational agility and IT capabilities. Similarly, "Risk, Control & Success Factors in Agile" entails risk assessment, quality and success factors, as well as control related content. "Stakeholders & Requirements Engineering" entails topics centered around different stakeholders, the process of requirements engineering, and generally speaking the

involvement of users in the software development process. The topic group "Teams, & Team Management" is focused more on project management activities involving the team on a more abstract level. The last topic group "Technologies & Applications in Agile" relates to some technical and application-oriented facets, namely cloud technologies, security, and big data in ASD.
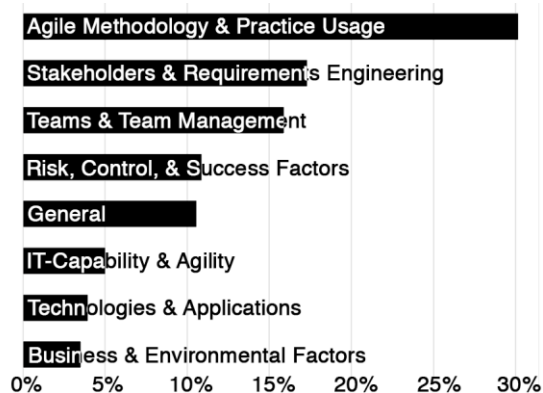

Figure 3. Topic Group Distribution

Looking at the rankings of the topics and overall distribution (see Figure 3), one sees that ASD methodologies have been covered most (32.35%). While "Teams & Team Management" appears to be covered well (15.87%), actual team interaction (i.e., "Teaching and Learning Agile", "Communication in Distributed ASD", and "Decision Making in ASD") has been covered less so (3.72%, 2.79%, and 0.95% respectively) and most of the distribution stems from "Teams & Kanban" (8.41%). The ranking of topics of each group serves as a proxy in their distribution (ranks 3, 12, 16, and 21 for this example).

4.2. Key Outlets and Articles

Based on the number of publications per outlet displayed in Table 1, we clearly see that the computer science-oriented conferences (e.g., ICSE, XP/Agile) dominate the information systems oriented conferences with nearly three times the number of publications (ICSE: 139 vs AMCIS: 49). The most prestigious information systems conference, the International Conference on Information Systems (ICIS), shows up second to last with 22 publications. This might hint at the orientation of extant ASD research being more technical and less managerial, social, or interdisciplinary (see also Figure 2).

Regarding the journal-based publications, the field is dominated by IEEE Software with 187 publications, followed by the Journal of Systems and Software (99) and Information and Software Technology (73). The most published-in journal of the Senior Scholars' Basket is the European Journal of Information Systems with 21 publications, ranked sixth, tied with IEEE Transactions on Software Engineering.

Looking at the most published-in outlets over time, one can identify different trends. While some outlets have been publishing ASD research early on (e.g., IEEE Software, Computer, ICSE, or HICSS), some started out later (e.g., ECIS, Journal of Systems and Software, or Information and Software Technology). While IEEE Software has been early on a very important outlet for ASD research, it has lost steadily since 2010 – but an upward trend started in 2016.

Looking at more recent publication statistics, especially XP/Agile, HICSS, and the Journal of Systems and Software appear to be the most up-and-coming outlets for ASD research. The trend for PACIS and ICIS appears to be declining.

Furthermore, topic modeling allows for identifying those papers, which cover each topic the most. It is important to note that "most covering" does not mean that these articles are the most influential or most important ones for this topic but rather are covering the topic most precisely in terms of the LDA model. We see that some topics are driven by the same authors repeatedly (e.g., "Effort Estimation, Success Metric" by Abrahamsson or "Pair Programming" by Balijepally), or that some authors are involved in different topics (e.g., Conboy in "Lean" and "Communication in Distributed ASD").

5. Discussion

5.1. Trends

By further investigation of our timeline regarding the distribution of published articles (see Figure 1, Figure 2), we recognize several interesting findings. First, ASD seems to strongly draw the interest of the research community starting around the year 2000, spiking at around 2003. Since then, there is a significant increasing slope of the graph, indicating that more articles have been published in the following years. Popular works published within this year are for example Williams and Cockburn's article "Agile Software Development: It's about Feedback and Change" [60] and of course the "Agile Manifesto" [10]. All publications have in common that they deal with the topic of ASD from a methodology perspective, putting emphasis on concepts, principles, or detailed information

concerning a specific approach. Some other articles published in the year 2003 deal with the topic of "virtual teams" [27]. This is not surprising, since the concept of virtual teams is seen as an important antecedent for "doing agile" in organizations [13, 25].

Second, we identified a peak in our timeline between 2008 and 2009. One explanation for this may be the call for papers for special issue themes, such as "flexible and distributed ISD" in Information Systems Research (ISR) journal [29] or previous works, which inspired further research, such as Larman's "Agile and Iterative Development: A Manager's Guide" [37] or Poppendieck and Poppendieck's "Lean Software Development: An Agile Toolkit" [47]. The ISR special issue was intended to build on the success of a previous special section of Communications of the ACM [2] and mini-track at HICSS in 2006 [3]. Ågerfalk and Fitzgerald argued that "it became clear from these efforts that as a very active emerging area of research, there was an imminent need for a forum that allowed for the development and dissemination of full-research papers of the highest quality" [4]. Similarly, a special issue of the European Journal of Information Systems was published in 2009 [1]. It aimed at improving the understanding of various phenomena in ASD.

Consolidating this description of the trend in publications of ASD research, we suggest that ASD, while being a highly important topic to practice [55], and despite a high and still growing number of publications, still lacks coverage in the top journals of information systems research as both curves drift further apart over time (see Figure 1).
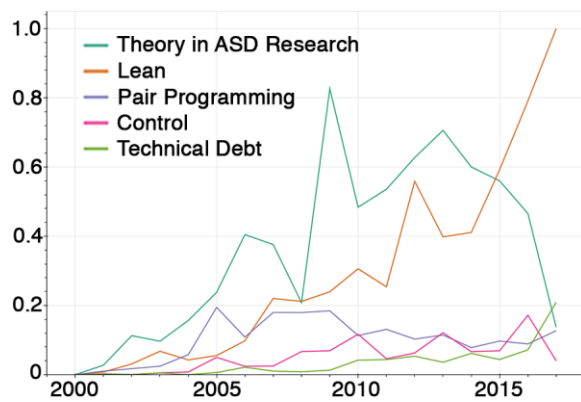
5.2. Implications

Figure 4. (Normalized) Distribution of selected topics over time (excerpt)

Combining the outlined descriptions and looking at the evolution of topics present in research (see Figure 4), we found the majority (approximately 75%) of all mentioned topics gaining popularity over time. On closer examination of the data, however, we found topic-specific differences with regard to the respective trend development. First of all, the topic "Theory in ASD Research" is overall losing traction since its highest peak in 2009 and a smaller spike in 2013. Compared and in contrast to this trend, the topic "Lean" is overall showing a positive trend in topic distribution over time. While, from a trend development perspective, both topics, "Theory in ASD Research" and "Lean", are very much alike, we see a notable turning point in 2015, where for the first time "Lean" became distributed wider than theory-related topics. Moreover, the trend development of "Lean" represents by far the steepest slope compared to all other topics since 2015, indicating that this topic is not yet saturated but currently is the most discussed topic, with only temporary drops in its ascend to the top. This is interesting, since Dingsøyr, Nerur, Balijepally and Moe [24] made this assumption in 2012: "A growing interest is evident at agile conferences on identifying ways to combine principles of lean development with software development" [24 p. 1218]. Besides these examples for either strongly increasing and decreasing topic trend developments, we also found topics which have developed almost constantly over time: "Pair Programming" and "Control in ASD" are good examples. A striking feature of the latter topic is the peak in 2016, which can be explained by the extensive literature review and the call for further research on the topic of control by Wiener et al. [59]. Both topics are generally less often discussed but show a comparatively non-volatile behavior over time.

Regarding the overall coverage of different topics, the distribution over different outlets (see Table 1) the rankings of the topics (see Table 2), topic group distributions (see Figure 3), and the distributions over time (see Figure 4), we derive conclusions over gaps in the extant literature. The top three topics are about "Theory", "Lean" and "Teams & Kanban", indicating an emphasis on distinct methodology usage and team management in literature. Nearly all topic groups have at least one topic in the top 10, indicating some degree of coverage, with the only exceptions being the topic groups "Business & Environmental Factors" and "Technologies and Applications". As can be seen in Figure 4, "Technical Debt", as part of "Business & Environmental Factors", has only started to increase in coverage over the last two years, indicating an

upward trend. In general, topics focusing on the above mentioned aspects of ASD are found first at rank 15, indicating a gap in current research. With regard to the latter the lack is not surprising since ASD is a socio-technical process rather than a purely technical one [40] and consequently, the major problems of ASD projects are less technological as more sociological in nature [22]. While other topics might touch on social aspects as well (e.g., "Communication in Distributed ASD" or "Teaching & Learning Agile"), these aspects are far less pronounced and of a more ancillary nature in these topics. Contrary to the fact that these topics are themselves of ancillary nature due to their low ranking and distribution and that these aspects appear to be peripheral matter to extant research, research acknowledges the importance of a not only technical but also social focus of ASD [20, 43].

In line with Dingsøyr, Nerur, Balijepally and Moe [24] we observe a trend of increasing quantity and quality of ASD research and that some subfields (i.e., topics) in ASD research are more mature or saturated than others. Both, the findings from Dingsøyr, Nerur, Balijepally and Moe [24] and the "top 10 burning questions" [30] are reflected in our results: "Lean", "Effort Estimation, Success Metrics", "Agile Architecture & Design", or "Large-Scale ASD" are important topics, while "Pair Programming and "Extreme Programming" are becoming comparatively less important. Furthermore, Freudenberg and Sharp [30] point out that sociological studies are important but currently mostly of peripheral appearance, which is clearly still the case and echoed by our results – a chance for ASD researchers.

To encourage ASD research to close these gaps, we propose the following research agenda. First, technologies and applications (see Topic Group "Technologies & Applications") as well as tool support (see Topic Group "IT Capability & Agility" and related topics) should be investigated further. The low rankings of the specific topics (see Table 2) and the low overall distribution (see Figure 3) paint a clear picture of an underrepresented research area. Studies on the effects of the use of tools such as versioning systems or coding tools would be valuable, as issues relating to, for instance, communication [e.g., 34] could be improved with improved understanding of the role of tools in ASD. Second, the "social" aspect of "socio-technical systems" needs to be embraced more by researchers. Similar to the first point of our research agenda, our data shows clearly a lack of research on this aspect of ASD, as no single topic group focuses on social aspect. For example, studies on the effects of agile

SD on control (see Topic Group "Risk, Control, & Success Factors in Agile") or team-related issues (see Topic Group "Teams & Team Management") such as team composition or team diversity, could complement existing similar information systems research streams and answer calls for research [e.g., 38, 59]. Third, we encourage ASD researchers to increase the amount of self-reflecting and reviewing literature. By reflecting upon the current stage of ASD research, gaps become more apparent and by replicating extant research, trust in existing findings can be improved. We believe that the ASD research community specifically and the IS community in general would benefit greatly from extensive research on these three main points of our proposed research agenda.

It should be noted, however, that our discussion is based on the results of this topic modeling and not on statements of different authors. Therefore, our statements are of speculative nature and only backed by exemplary reasoning.

6. Conclusion and Outlook

Within this paper, we identified research topics on ASD covered by relevant journals and prestigious conferences. Our findings provide an overview of topics, which attracted the attention of the research community dealing with ASD methodologies over the last three decades.

Based on the topic modeling conducted on this data set, we demonstrated that computer-aided topic clustering can help to outline the current state of ASD research. With the help of computer-aided analysis, we were able to process large amounts of data and uncover topics within these texts. Further processing of this data and the results, as well as qualitative analysis, helped us gain deeper insights into the history of ASD research and uncover the topics in our body of knowledge regarding ASD research. Further, we waged an outlook into the future of ASD research by identifying less covered topics and looking for gaps in the topics covered by extant research. This might help other scholars in identifying new avenues and further extends the scientific community's knowledge about ASD.

We are confident that our study and results provide an appropriate degree of generalizability, completeness, and replicability. We described our procedure and sources to ensure replicability, while generalizability and completeness go together for this study. Due to the comprehensive literature basis provided by our structured literature review and the help of a computer-aided analysis, we are able to process extant research at large and discover topics.

This research design facilitates generalizability and completeness.

Future research might expand on this research by adding more outlets or updating the conclusions based on more recent publications to further extend the applicability and generalizability of our findings. We also call for replication of our study to improve the confidence in our results and our conclusions. A continued effort in keeping track of the developments in ASD research might help in keeping researchers focused and aware of trends, topics, and gaps.

7. References

[1] Abrahamsson, P., et al.: 'Lots done, more to do': the current state of agile systems development research. European Journal of Information Systems 18, (2009), pp. 281-284

[2] Ågerfalk, P.J., Fitzgerald, B.: Flexible and distributed software processes: old petunias in new bowls. Communications of the ACM, pp. 27-34. Citeseer (2006)

[3] Ågerfalk, P.J., Fitzgerald, B.: Understanding flexible and distributed software development processes. Proc. Thirty-Ninth Hawaii Internat. Conf. System Sciences, pp. 1. IEEE Computer Society Press Washington, DC (2006)

[4] Ågerfalk, P.J., et al.: Flexible and Distributed Information Systems Development: State of the Art and Research Challenges. Information Systems Research 20, (2009), pp. 317-329

[5] Aggarwal, C.C., Zhai, C.: Mining text data. Springer Science & Business Media (2012)

[6] Alghamdi, R., Alfalqi, K.: A Survey of Topic Modeling in Text Mining. International Journal of Advanced Computer Science and Applications (IJACSA) 6, (2015), pp. 147-153

[7] Alzoubi, Y.I., et al.: Empirical studies of geographically distributed agile development communication challenges: A systematic review. Information & Management 53, (2016), pp. 22-37

[8] Arun, R., et al.: On finding the natural number of topics with latent dirichlet allocation: Some observations. Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 391-402. Springer (2010)

[9] Balijepally, V., et al.: ARE TWO HEADS BETTER THAN ONE FOR SOFTWARE DEVELOPMENT? THE PRODUCTIVITY PARADOX OF PAIR PROGRAMMING. MIS Quarterly 33, (2009), pp. 91-119

[10] Beck, K., et al.: Manifesto for Agile Software Development. (2001) http://www.agilemanifesto.org/

[11] Blei, D.M., et al.: Latent dirichlet allocation. Journal of machine Learning research 3, (2003), pp. 993-1022

[12] Boehm, B.W.: A Spiral Model of Software Development and Enhancement. IEEE Computer 21, (1988), pp. 61-72

[13] Bowen, S., Maurer, F.: Process support and knowledge management for virtual teams doing agile software development. Proceedings 26th Annual International Computer Software and Applications, pp. 1118-1120 (2002)

[14] Campanelli, A.S., Parreiras, F.S.: Agile methods tailoring - A systematic literature review. The Journal of Systems and Software 110, (2015), pp. 85-100

[15] Cao, J., et al.: A density-based method for adaptive LDA model selection. Neurocomputing 72, (2009), pp. 1775-1781

[16] Cao, L., et al.: A framework for adapting agile development methodologies. European Journal of Information Systems 18, (2009), pp. 332-343

[17] Cao, L., et al.: Adapting funding processes for agile IT projects: an empirical investigation. European Journal of Information Systems 22, (2013), pp. 191-205

[18] Chen, C., et al.: Detecting the association of health problems in consumer-level medical text. Journal of Information Science 44, (2016), pp. 3-14

[19] Conboy, K.: Agility from First Principles: Reconstructing the Concept of Agility in Information Systems Development. Information Systems Research 20, (2009), pp. 329-355

[20] Conboy, K., et al.: People over Process: Key Challenges in Agile Development. IEEE Software 28, (2011), pp. 48-48

[21] Debortoli, S., et al.: Text Mining For Information Systems Researchers: An Annotated Topic Modeling Tutorial. Communications of the Association for Information Systems (CAIS) 39, (2016), pp. 110-135

[22] DeMarco, T., Lister, T.: Peopleware: productive projects and teams. Dorset House Publishing Co., Inc. New York, NY, USA (1987)

[23] Dikert, K., et al.: Challenges and success factors for large-scale agile transformations: A systematic literature review. Journal of Systems and Software 119, (2016), pp. 87-108

[24] Dingsøyr, T., et al.: A decade of agile methodologies: Towards explaining agile software development. Journal of Systems and Software 85, (2012), pp. 1213-1221

[25] Domino, M., et al.: Applying agile software development processes to global virtual teams: a study of communication modalities. Proceedings of the 2002 ACM SIGCPR conference on Computer personnel research, pp. 76-78. ACM, Kristiansand, Norway (2002)

[26] Dybå, T., Dingsøyr, T.: Empirical studies of agile software development: A systematic review. Information and software technology 50, (2008), pp. 833-859

[27] Edwards, H.K., Sridhar, V.: Analysis of the effectiveness of global virtual teams in software engineering projects. HICSS 2003. IEEE (2003)

[28] Fitzgerald, B., et al.: Customising agile methods to software practices at Intel Shannon. European Journal of Information Systems 15, (2006), pp. 200-213

[29] Fitzgerald, B., et al.: Call for papers - Information Systems Research Special Issue: Flexible and Distributed Information Systems Development. Information Systems Research 17, (2006), pp. 323-325

[30] Freudenberg, S., Sharp, H.: The top 10 burning research questions from practitioners. Ieee Software 27, (2010), pp. 8-9

[31] Fruhling, A., de Vreede, G.-J.: Field Experiences with eXtreme Programming: Developing an Emergency Response System. Journal of Management Information Systems 22, (2006), pp. 39-69

[32] Griffiths, T.L., Steyvers, M.: Finding scientific topics. Proceedings of the National academy of Sciences 101, (2004), pp. 5228-5235

[33] Guinea, A.S., et al.: A systematic review on the engineering of software for ubiquitous systems. The Journal of Systems and Software 118, (2016), pp. 251-276

[34] Hummel, M., et al.: The Role of Communication in Agile Systems Development. BUSINESS AND INFORMATION SYSTEMS ENGINEERING 5, (2013), pp. 343-355

[35] Inayat, I., et al.: A systematic literature review on agile requirements engineering practices and challenges. Computers in human behavior 51, (2015), pp. 915-929

[36] Karlsson, F., Ågerfalk, P.: Exploring agile values in method configuration. European Journal of Information Systems 18, (2009), pp. 300-316

[37] Larman, C.: Agile and Iterative Development: A Manager's Guide. Pearson Education (2003)

[38] Lee, G., Xia, W.: TOWARD AGILE: AN INTEGRATED ANALYSIS OF QUANTITATIVE AND QUALITATIVE FIELD DATA ON SOFTWARE DEVELOPMENT AGILITY. MIS Quarterly 34, (2010), pp. 87-115

[39] Levy, Y., Ellis, T.J.: A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research. Informing Science 9, (2006), pp. 181-212

[40] Lyytinen, K., et al.: Attention shaping and software risk—a categorical analysis of four classical risk management approaches. Information Systems Research 9, (1998), pp. 233-255

[41] Maowen, W., et al.: Text topic mining based on LDA and co-occurrence theory. Computer Science & Education (ICCSE), 2012 7th International Conference on, pp. 525-528. IEEE (2012)

[42] Maruping, L.M., et al.: Role of collective ownership and coding standards in coordinating expertise in software project teams. European Journal of Information Systems 18, (2009), pp. 355-371

[43] Maruping, L.M., et al.: Folding under pressure or rising to the occasion? Perceived time pressure and the moderating role of team temporal leadership. Academy of Management Journal 58, (2015), pp. 1313-1333

[44] Miles, M.B., Huberman, A.M.: Qualitative data analysis : an expanded sourcebook. Sage Publications, Thousand Oaks (1994)

[45] Mimno, D., et al.: Optimizing semantic coherence in topic models. Proceedings of the conference on empirical methods in natural language processing, pp. 262-272. Association for Computational Linguistics (2011)

[46] Pedregosa, F., et al.: Scikit-learn: Machine learning in Python. Journal of machine learning research 12, (2011), pp. 2825-2830

[47] Poppendieck, M., Poppendieck, T.: Lean Software Development: An Agile Toolkit. Addison-Wesley Longman, Amsterdam (2003)

[48] Rowe, F.: What literature review is not: diversity, boundaries and recommendations. European Journal of Information Systems 23, (2014), pp. 241-255

[49] Royce, W.W.: Managing the Development of Large Software Systems: Concepts and Techniques. Proceedings of WesCon, (1970)

[50] Saldaña, J.: The coding manual for qualitative researchers. SAGE, London (2016)

[51] Sarker, S., Sarker, S.: Exploring Agility in Distributed Information Systems Development Teams: An Interpretive Study in an Offshoring Context. Information Systems Research 20, (2009), pp. 440-462

[52] Srivastava, A.N., Sahami, M.: Text mining: Classification, clustering, and applications. Chapman and Hall/CRC (2009)

[53] Strasser, A., Westner, M.: Information systems offshoring: results of a systematic literature review. Journal of Information Technology Management 26, (2015), pp. 70-142

[54] Strauss, A., Corbin, J.: Basics of qualitative research. Sage, London (1998)

[55] Version One: 12th annual state of agile report. Version One (2018)

[56] Vom Brocke, J., et al.: Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. Communications of the association for information systems 37, (2015), pp. 205-224

[57] Wang, X., et al.: Assimilation of agile practices in use. Information Systems Journal 22, (2012), pp. 435-456

[58] Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a literature review. MIS Quarterly 26, (2002), pp. xiii-xxiii

[59] Wiener, M., et al.: Control configuration and control enactment in information systems projects: review and expanded theoretical framework. MIS Quarterly 40, (2016), pp. 741-789

[60] Williams, L., Cockburn, A.: Agile Software Development: It's about Feedback and Change. Computer 36, (2003), pp. 39-44

[61] Wolcott, H.F.: Transforming qualitative data : description, analysis, and interpretation. Sage, Thousand Oaks, Ca. (1994)

[62] Zheng, Y., et al.: Collective agility, paradox and organizational improvisation: the development of a particle physics grid. Information Systems Journal 21, (2011), pp. 303-334