CrowdRE in a Governmental Setting: Lessons from Two Case Studies

Jelle Wouters, Rob Janssen,
Bas van Hulst, John van Veenhuizen
Royal Netherlands Marechaussee
Postbus 20701, 2500 ES Den Haag, NL
Email: j.wouters.01@mindef.nl

Fabiano Dalpiaz, Sjaak Brinkkemper
Department of Information and Computing Sciences
Utrecht University, 3584 CC Utrecht, NL
Email: f.dalpiaz@uu.nl, s.brinkkemper@uu.nl

***Abstract*—Crowd-based Requirements Engineering, or CrowdRE, is a recent paradigm that promotes the active participation of a large number of stakeholders in RE. In CrowdRE, requirements elicitation can be crowdsourced by creating an online platform that allows stakeholders to formulate ideas regarding the specific product. Although some case studies on crowd-based elicitation exist, no conclusive evidence can be derived on the effectiveness of such techniques. In this paper, we study crowd-based elicitation within a large governmental organization. We conduct two case studies of CrowdRE within the Royal Netherlands Marechaussee. For this, we construct the *KMar-Crowd* method, which adapts CrowdRE ideas to the needs of governmental organizations. While one case compares the crowd-generated ideas with requirements elicited via interviews and similar techniques, the other measures the usefulness of the gathered ideas for a product for which no prior elicitation was conducted. The results of the case studies, which attracted larger crowds than in previous studies, indicate that CrowdRE can be successfully applied to engage the users of a software product in requirements elicitation. Contradicting earlier studies, the inclusion of gamification elements in the CrowdRE method did not increase the motivation to participate, possibly due to the nature of governmental organizations.**

I. Introduction

Crowd-based Requirements Engineering (CrowdRE) is an emerging paradigm that promotes the active involvement of a "crowd" of stakeholders, including current and potential users, of a software product [1]. This extends the reach of traditional RE approaches, which involve a selection of the stakeholders, toward the democratic participation of users in RE [2].

Elicitation is the phase that current CrowdRE research has studied more in depth [3]. Within CrowdRE, two approaches have been identified to elicit requirements, and other feedback, from the users[1] [1]: (i) *pull feedback* occurs when the crowd is asked to provide requirements and ideas through some platform; and (ii) *push feedback* consists of gathering requirements from usage statistics collected from the product under study.

We focus on elicitation via *pull feedback*. While existing case studies provide some evidence on CrowdRE in practice [4], [5], [6], [7], the volume of studies is by far too limited for organizations to assess the potential and pitfalls of adopting pull-based elicitation practices. In particular, the size of the crowd is one of the selling points of CrowdRE, and previous studies did not manage to create and maintain a large crowd.

Furthermore, none of these studies were executed in a governmental setting. In public and semi-public organizations, several factors are fundamentally different than in private organizations. For example, while private organizations are mainly evaluated based on their financial results, public organizations are evaluated based on their contribution to society. Managers in public organizations are also being held accountable in different ways than at private organizations [8].

Requirements elicitation and evolution is more constrained in governmental settings. For example, in the European Union, a formal tender procedure needs to be executed before a software application of significant size can be built by a supplier. After this tender procedure has been completed, no substantial changes to the requirements specification can be made, even when the work processes change. The difficulties of conducting RE in tender-based settings has been acknowledged by some scholars [9].

In this paper, we report on two case studies in which we apply crowd-based, pull-feedback elicitation in a governmental setting. To enable the research, a crowd-based elicitation web platform was constructed in-house. These two case studies have been executed at the Royal Netherlands Marechaussee (in this paper, based on the Dutch name: *KMar*; international abbreviation: RNLM), part of the Dutch Ministry of Defence. The main research question this paper will answer is *Can CrowdRE be used in a governmental setting to complement the requirements elicitation practices?* In particular, we make the following contributions:

- Building on previous CrowdRE research [4], we propose the tool-supported *KMar-Crowd* method for conducting pull-based, crowd-based requirements elicitation within governmental organizations.
- We compare the overlap between the requirements collected via *KMar-Crowd* with requirements for the same system gathered via traditional elicitation techniques such as interviews, task analysis, and introspection. We do so via the "S-Sys" case study, in which 135 participants posted 32 ideas and cast over 300 votes.
- We test the dynamics of a larger crowd, and we determine the usefulness of the ideas generated by the crowd,

[1]While CrowdRE promotes the involvement of *all* stakeholders, most studies, including this, focus on current and prospective *users*.

through a case study (385 participants, 78 ideas, over 500 votes) regarding system "V-Sys", for which elicitation had not yet been conducted.

We use the term *idea* to refer to the inputs from the crowd. While some of these ideas may be (almost) directly mapped to a *requirement* for the product, others have to be discussed and refined by the analysts prior to becoming requirements.

The rest of this paper is structured as follows. In Section II, we present the relevant background. Section III describes the research method. Section IV introduces the method used in the paper: the *KMar-Crowd*. The results of the case studies are discussed in Sections V and VI. Finally, we draw conclusions and sketch future directions in Section VII.

II. Background and Related Work

We first introduce the elementary background on CrowdRE in Section II-A. Then, we discuss previous case studies in Section II-B. Finally, we review alternative approaches to conducting CrowdRE in Section II-C.

A. Background on CrowdRE

CrowdRE was conceptualized by Groen *et al.* as an "umbrella term for all automated RE techniques, including crowdsourcing, text mining and data mining" [10] that can be utilized to actively involve a crowd of stakeholders, including users, in the RE process. Groen *et al.* propose a multi-method solution, where both quantitative (using mining techniques) and qualitative (using crowdsourcing) data is collected as a source of requirements.

Similar research was conducted independently by other groups. For example, Snijders *et al.* introduced the term *Crowd-centric requirements engineering* [11]. They formulated a clear reason for CrowdRE, arguing that: "users are seldom involved, despite the common agreement that doing it would result in better requirements elicitation and higher chances for project success" [11].

Hosseini *et al.* also studied crowdsourcing in requirements elicitation. They argued that, because of the fast-changing landscape of IT products, especially with the introduction of software-as-a-service and cloud products, the user groups of these products become more heterogeneous. Therefore, traditional requirements elicitation efforts might not be sustainable, but using crowdsourcing to gather requirements might be [12].

These research groups co-authored a landscape paper [1] that distinguishes two main approaches to CrowdRE: (i) pull feedback to elicit user feedback, e.g., using a crowdsourcing platform; and (ii) push feedback, i.e., collecting requirements by monitoring and analyzing contextual and usage data. Both of these information streams are gathered from a crowd of users who use a product, and they are analyzed by a development team which uses these data to further improve the product. While both parts of CrowdRE are equally important, this paper focuses on using a crowd to elicit user feedback.

B. Case Studies: CrowdRE via Elicitation Platforms

Several authors already performed case studies on CrowdRE. The first substantial case study of CrowdRE was the *Requirements Bazaar*, executed by Renzel *et al* [6]. The Requirements Bazaar is a web platform to support requirements elicitation by providing tools for co-creation and prioritization. Users were able to formulate ideas and to prioritize them.

Since one of the key elements of CrowdRE is to allow every user to participate, the quality of the gathered ideas is somewhat limited. To mitigate this, Menkveld *et al.* [5] built a platform that users to provide their ideas by expressing the key parts of a user story: the role, the action, and the benefit. Via a 5-week case study, they found that the use of user stories was perceived as useful by the end users and that the ideas were good enough to serve as requirements for the developers. As a limitation, only 13 users provided their opinion on the platform through the final survey.

Another challenge of CrowdRE is sustaining a motivated crowd. Snijders *et al.* [4] tried to increase the motivation of the crowd by introducing gamification elements into *REfine*, the CrowdRE platform they developed. Their case study led to 21 needs, 37 comments and 130 votes. Participants indicated that they were more engaged than in different requirements elicitation efforts. However, this study featured a small crowd.

The GARUSO platform [7] was built with the goal of involving stakeholders that are outside organizational reach. GARUSO went beyond *REfine* in terms of gamification, by offering a game-like experience to the participants, which was expected to engage them for a longer time. The researchers managed to involve 32 active stakeholders (out of over 700 participants who visited the platform) who contributed 56 ideas.

In 2019, Glinz gave an overview of the achievements, opportunities and pitfalls of CrowdRE [13], providing a compact overview of all published empirical studies. Among the remarkable findings, the case studies conducted so far all have a limited size (see also Table I). Our study contributes to the literature by providing two larger case studies, and by applying CrowdRE to the setting of governmental organizations.

C. Alternative Approaches to CrowdRE

We discuss the major strands of research within CrowdRE that go beyond pull feedback via elicitation platforms.

Alternative feedback channels have been explored, especially app stores [14], [15], [16], [17]. By writing reviews on an app store, the users of mobile apps can express their feedback without the necessity for the company who develops the software product to create a dedicated channel. The research in the field has focused mostly on the extraction of reviews that concern a particular feature [18]; the classification of user reviews among categories such as bugs and feature requests [14], [15], [16]; and the analysis of the reviews' sentiment [19]. While interesting, this channel is only applicable to apps that are made available publicly on an app store.

Researchers have also analyzed other feedback channels such as Twitter [20], [21] or online fora [22], [23], [24].

For Twitter, the difficulty consists of identifying valuable information within the massive amount of interactions that take place on such a broad channel. Online discussions in user forums are close to our research; an interesting future direction would be to apply the automated techniques developed for forums to the contents of a CrowdRE elicitation platform.

Another approach is the use of a crowd-work platform, where crowd workers are paid for the execution of RE-related tasks. This technique has been studied in the context of generating creative ideas [25], classifying app reviews according to software product qualities [26], and extracting requirements from privacy policies [27]. However, this approach is inadequate for our context, in which we aim to elicit ideas from prospective users of a software product that is not available to the general public.

III. Research Method

As stated above, the main research question is *Can CrowdRE be used in a governmental setting to complement the requirements elicitation practices?* To be able to answer this question, three main research goals are defined:

1) Determining the current requirements elicitation techniques in use at the governmental institution.
2) Developing a CrowdRE method that takes the specific nature of governmental organizations into account.
3) Validating the effectiveness of the CrowdRE method.

We mainly present results related to the second and third research goals. Learning about the current elicitation techniques was essential for us to devise the method and to interpret the results, but we will not provide a dedicated section that discusses those practices in this paper.

A. Canonical Action Research

Because of the highly practical nature of this research project, the method was developed using the principles of Canonical Action Research (CAR). According to Davison, Martinsons and Kock, CAR "is one of the more widely practised and reported forms of action research in the Information Systems literature" [28]. Two important principles are the *cyclical process model* and the *change through action*. To be able to achieve this, two case studies were conducted, with the results of the first case study being used to inform the second study, according to the principles of CAR.

The organization where we held the case studies is the Royal Netherlands Marechaussee (KMar), part of the Ministry of Defence of the Kingdom of the Netherlands. The KMar is a police force with military status, tasked with border protection, monitoring and protection duties and international and military police tasks. At the time of writing, around 7,000 employees are working for the KMar, of which around 5,000 are military employees. The KMar consists of a small strategic staff, a national shared service center, an in-house training facility, and 24 operational units (called 'brigades'), organized based on their location or strategic function.

The author who was tasked with gathering the data for the two case studies was embedded in the organization as an employee, which further strengthens one of the principles of CAR: *collaboration between researcher and client*.

B. Goals of the Case Studies

We execute two case studies following the research protocol that is presented below. In each case study, the online, tool-supported elicitation method *KMar-Crowd* is applied. More details on *KMar-Crowd* are provided in Section IV.

In the first case, *KMar-Crowd* is used to elicit ideas for an operational system (called "S-Sys" here) for which the requirements were gathered already in an earlier process, using traditional elicitation techniques: interviews, task analysis, and introspection. The main purpose of this case study is therefore to validate whether the *KMar-Crowd* will lead to similar requirements to those that were already gathered. The scale for the study is limited to a single brigade: the 478 employees of that brigade were invited to participate.

In the second case, *KMar-Crowd* is used to identify ideas for a product where no requirements were gathered for in an earlier process. The product, called "V-Sys" here, is going to replace another, out-dated product. The main goal of this case study is to assess whether *KMar-Crowd* can be scaled up to the whole organization and, while doing so, is still able to produce useful ideas for the analysts who will have to specify the requirements for the system to-be. Both case studies targeted operational employees, as they are the daily users of the systems for which requirements needed to be gathered. At the KMar, these employees are normally hardly involved in this process, even though they are the most important user group, as they need to carry out operational duties in the field and therefore simply do not have time to participate in this process. CrowdRE might be a good way to get them involved without pulling them from their operational duties.

To better compare our two case studies to earlier empirical research with CrowdRE elicitation platforms, Table I summarizes the gathered data from the two case studies and contrasts it to the studies of Menkveld *et al.* [5], Snijders *et al.* [4] and Kolpondinos and Glinz [7].

TABLE I
COMPARISON OF THE S-SYS AND V-SYS CASES WITH EARLIER STUDIES.

Measurement	S-Sys	V-Sys	Menkveld	REfine	GARUSO
Duration in days	33	56	35	35	92
Participants:					
- Invited	478	2,393	unk.	37	unk.
- Accessed	135	385	157	19	726
- Active	60	130	39	19	32
Ideas	32	78	57	21	56
Logins	240	623	247	unk.	unk.
Votes	316	531	89	130	160
Comments	28	78	14	37	unk.
Ideas / Accessed	0.24	0.20	0.36	1.11	0.08

In the table, we present three participant counts: *invited* is the number of people (possibly unknown) that were reached by an invitation to join the platform; *accessed* is the number of users who visited the platform at least once; and *active* considers participants who interacted actively, by posting an

idea, adding a comment, or expressing a vote. Note that the way of inviting participants differs per case study. In the study of Menkveld *et al.*, the invitation to participate was included in the product for which ideas were gathered. Snijders *et al.* targeted specific individuals, while Kolpondinos and Glinz used targeted advertising to recruit participants through organizational mailing lists. In our study, we used a combination of several mass emails to the organization and physical briefings executed by team leaders.

C. Research Protocol

Employees were invited to participate in case study via e-mail. In addition, a slide about the *KMar-Crowd* was included in the daily briefing for about four days. As there are three work shifts per day, the slide was therefore presented 12 times by an operational team leader. As soon as we reached the half-time mark for each study (the duration was set in advance), a reminder was sent to all employees and the briefing slide was re-introduced again for four days. At the end of the cycle, all employees who accessed *KMar-Crowd* at least once received a questionnaire. This questionnaire measured two core concepts contributing to the *effectiveness* of the CrowdRE method applied:

1) The appreciation of the way of working tested in *KMar-Crowd* using 7 self-developed Likert-scale questions;
2) The effect of the gamification elements in the platform on the motivation for users to participate.

The questionnaire can be found in our online appendix[2].

The research protocol described above was first tested in a small pilot study, to which 30 IT employees were invited. Once this pilot study was concluded, only one change was made to the platform. Initially, the reward system of the gamification element was based on real military ranks. Employees were, therefore, able to gain a higher (or lower) military rank then they had in daily life. Several testers indicated that this might result in a backlash of the users. Therefore, the ranking system was changed during the real case studies: employees were now able to collect stars.

IV. The *KMar-Crowd* Method

To allow using crowd-based elicitation in governmental environments, we propose the *KMar-Crowd* method that organizes the CrowdRE efforts into four phases, as shown in Fig. 1. During the first phase, a core team is created, which consists of people playing the role of requirements analysts and will oversee and manage the crowd. As soon as the core team formulates a question for the crowd, so to allow focused interaction through the platform, the crowd can be deployed by advertising the platform and its purpose. Upon invitation, in phase 2, participants are able to formulate ideas as well as to up/down-vote for or add comments to existing ideas. While phase 2 focuses on idea divergence, phase 3 focuses on convergence thinking [29]. The core team writes a short summary of all the ideas collected so far and writes some responses. This last step is important, as participants then see that something happens with their ideas. During the third phase, the crowd is still able to vote on ideas and discuss ideas, as the responses of the core team might lead to new discussion points. The activities in the second and third phase are undertaken in parallel. In the last phase (in which the activities are executed sequentially), the core team responds to ideas which they did not respond to yet, develops a timeline which describes to the crowd how to proceed next, invites highly-engaged crowd members to a focus group (giving them an additional sense of involvement), and finally executes the sprints to realize the timeline.

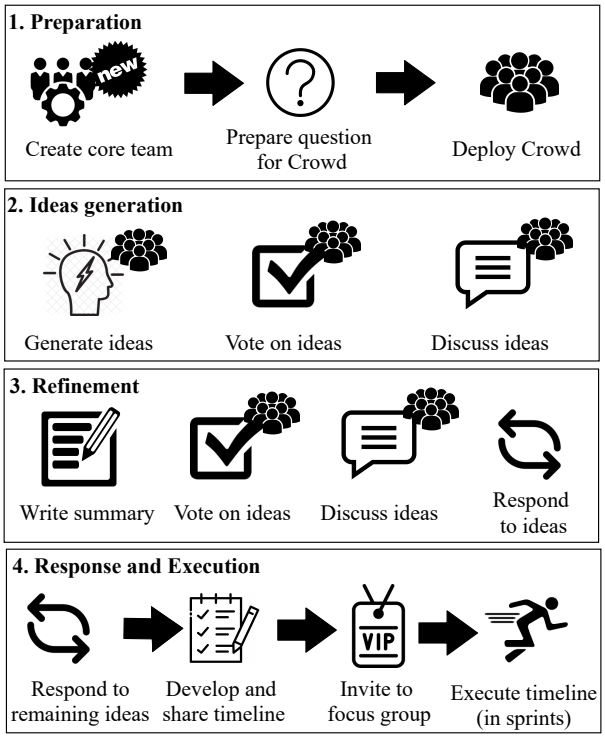

Fig. 1. The *KMar-Crowd* method for the application of CrowdRE in governmental organizations. The activities with the symbol are executed by the crowd, the others are performed by the core team. The ➡ symbol indicates sequential activities.

The *KMar-Crowd* method is inspired by the CCRE method proposed by Snijders *et al.* [12], but exhibits some essential differences. First, *KMar-Crowd* does only focus on initial requirements elicitation, and less on prioritization and negotiation, as these processes are complex in governmental organizations and cannot be fully executed by applying CrowdRE. Instead, prioritization and negotiation are included in the fourth phase, where a focus group is created with some highly-engaged crowd members. Another observable difference is the focus on the preparation of the CrowdRE process and the follow-up by the core team in the proposed method, as this is important to keep the crowd engaged over time. Thirdly, *KMar-Crowd* also introduces a summary written by the core team. By doing so, employees do not have to read all ideas,

[2]The online appendix includes the questionnaire and the spreadsheets we used to generate tables and charts: http://doi.org/10.5281/zenodo.5094295

but can easily see what is new and what they want to vote for. This reduces the chances of duplicate ideas and ensures that the workload for participants to stay up to speed is manageable once the number of ideas grows.

The case studies reported in this paper are executed through the use of a purpose-made CrowdRE platform, also called *KMar-Crowd*, which supports our method. This platform was designed and developed by the authors of this paper and is based on a WordPress site. It uses the lessons learned from earlier research: (i) it allows participants to express user stories via a simplified format [5]; and (ii) it includes gamification elements: points, badges, and a leaderboard [4].

As the platform is accessed through the internal KMar network, it incorporates Single-Sign-On, which makes it possible to retrieve the origin of participants and therefore offers seamless user management and login experience. When users open *KMar-Crowd* for the first time, they are presented an informed consent form and asked to either fill in their real names or a pseudonym in case they prefer to participate anonymously.

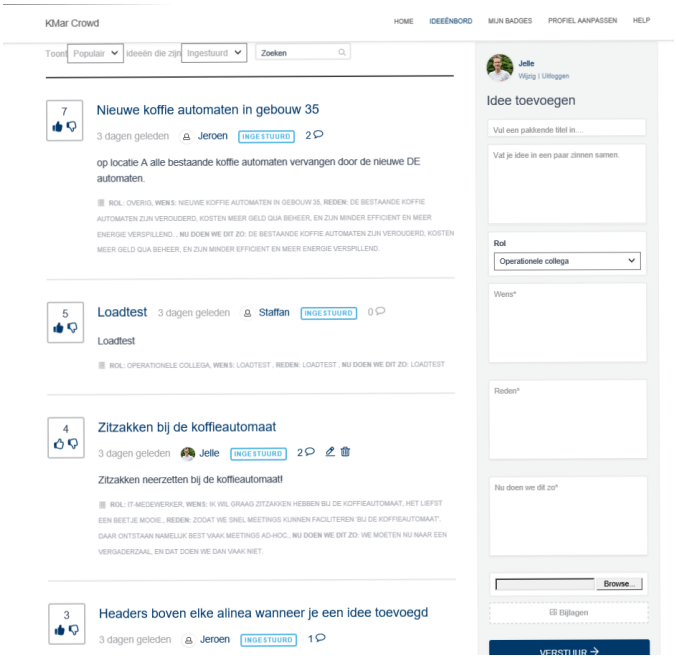

Fig. 2. Screenshot of the *KMar-Crowd* platform idea board (with dummy data, due to confidentiality reasons). On the left, the ideas together with a vote-up and vote-down button and the vote count are visible. On the right, a form is available to post new ideas in a simplified user-story format. Every visitor can enter a title, the user story text, and optionally a picture to illustrate the idea.

Via the platform, users can enter ideas (using the user story format as a template), add comments, and up-/down-vote existing ideas. For each of these actions, users gain points. When a certain amount of points of each category is collected, users are rewarded with stars. As a positive reinforcement, all users receive one star after logging into the platform. In our case studies, all the participants who collected two or more stars were eligible for a small prize that was assigned via a raffle. Fig. 2 shows a screenshot of the idea board of the *KMar-Crowd* while Fig. 3 shows the badges page of the platform. Table II shows the number of actions necessary for obtaining a certain number of stars. Note that, to advance to the next level of stars, all the action types (logging in, adding ideas, voting, commenting) need to be executed.

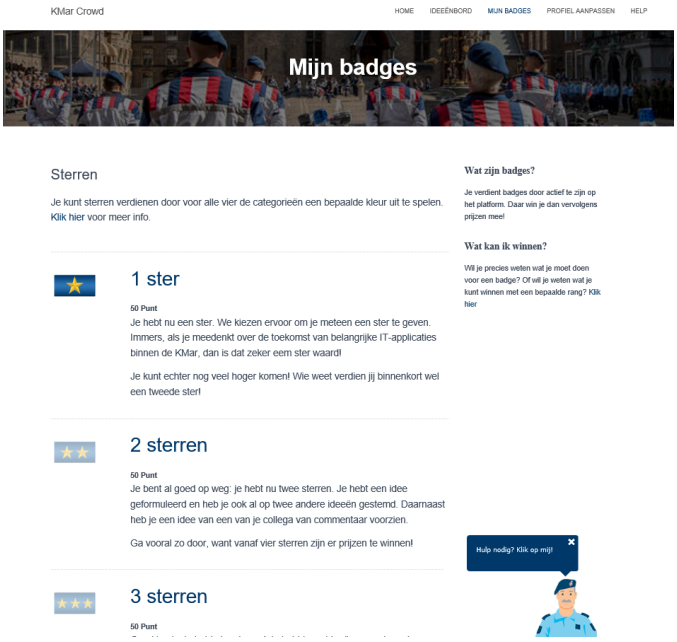

Fig. 3. Screenshot of the *KMar-Crowd* Platform badges overview. The title of the badge together with an image and some descriptive text can be seen. In this example, the user has one star. The right column gives a general explanation. The avatar in the bottom-right corner was available to provide some answers to predefined questions.

TABLE II
NUMBER OF ACTIONS NECESSARY TO GAIN STARS: ALL ACTION TYPES SHOULD BE EXECUTED TO ADVANCE AND GAIN A NEW STAR.

Stars	Logins	Ideas created	Ideas voted	Comments
1	Automatically obtained once logged in once			
2	2	1	2	1
3	5	2	5	2
4	10	5	10	5
5	20	10	20	10

To achieve this, we used the WordPress Content Management System as a basis, in combination with the IdeaPush and BadgeOS plugins. More plugins were used, but for the sake of brevity, these are omitted from the discussion. The core of the platform can be reproduced with these three components.

V. S-SYS: CAN THE *KMar-Crowd* PRODUCE SIMILAR RESULTS TO TRADITIONAL ELICITATION TECHNIQUES?

The goal of the S-Sys case study was to collect ideas for a yet-to-develop operational registration system. S-Sys, which will replace a legacy system, will allow to report on violations and offenses and to generate formal police reports. The elicitation process of this system was already completed using traditional methods such as interviews, task analysis and

introspection. Therefore, S-Sys served as a validation study for us to assess whether the deployment of the *KMar-Crowd* method would lead to comparable requirements, and to determine whether some interesting and important requirements were missed in the original requirements elicitation phase. In this case study, we did not analyze whether the crowd missed out requirements which were gathered earlier using traditional means, as we position *KMar-Crowd* as an addition to the requirements elicitation process, and not as a replacement for traditional methods.

The study was executed at one operational unit (a *brigade*) of the KMar. This specific brigade was selected as its employees used an analogous system, to be replaced, almost daily in their work. When we deployed our case study, the brigade consisted of 478 employees, all of whom were invited. The data collection phase took four weeks. The most important indicators of this study are presented in Table III.

TABLE III
USAGE INDICATORS FOR THE S-SYS CASE STUDY.

Indicator	Total
Unique users	135
Ideas	32
Votes	316
Login actions	240
Comments	28
Users upgraded to two stars	10
Users upgraded to three stars	1
Users upgraded to four stars	0
Page hits	1,554

The results of the first case study were measured along four dimensions: (i) user engagement; (ii) user origin; (iii) appreciation of the *KMar-Crowd* method and tool using the questionnaire introduced in Sec. III-C; and (iv) quality and usefulness of the crowd-generated ideas.

A. User Engagement

User engagement was measured using the indicators of Table III and was also plotted over the execution time of the case study, see Fig. 4. It can be seen how most of the interactions took place in the first few days, with a second spike originating after day 19, after a reminder was sent. It is also visible that a correlation exists between the time of registration and the other activities performed on the platform. This indicates that participants are most active on the platform at the moment of registration.

To be able to further interpret the usage statistics, we clustered users according to their participation level: (i) *none* if they only logged in without completing other actions; (ii) *passive interaction*, i.e., voting and/or commenting but not posting new ideas; and (iii) *active interaction*, i.e., posting ideas. Moreover, we clustered users according to visit frequency: (i) *onetime*, only one visit; (ii) *returning*, 2-4 times; and (iii) *power*, 5 or more times. Many users did only visit the platform once and did not interact in any form (56%). 44% of the users interacted in some form on the platform. More details on the interaction can be seen in Fig. 5.

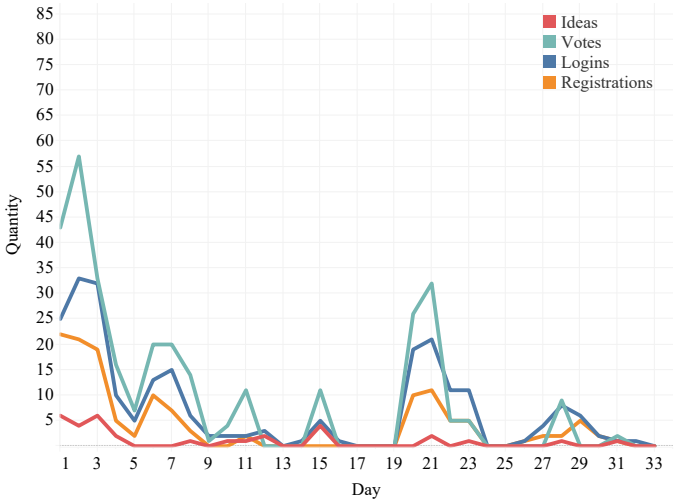

Fig. 4. Usage indicators for the S-Sys case study plotted over time.

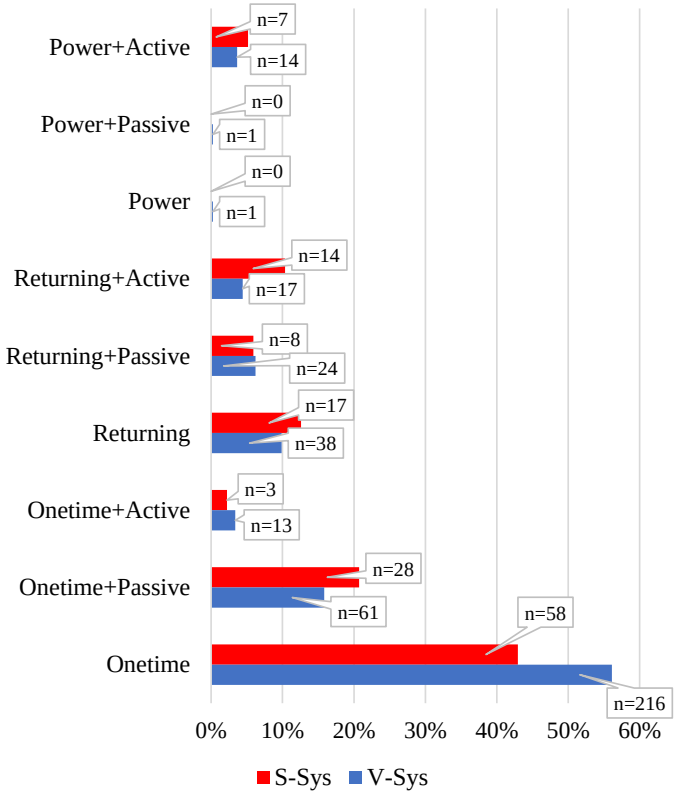

Fig. 5. Spread of users based on their activity rate during the two case studies. The chart compares the cases in percentages on the x-axis, with callouts showing the absolute numbers per user engagement category.

B. User Origin

Next, the user origin was measured, in order to assess whether the *KMar-Crowd* was able to connect to the actual users of the software under study. In a military context, it is easy to identify who is using the software by looking at the rank of the participant. Table IV shows the user distribution across ranks and the proportion of ideas, votes, and logins

attributed to these rank groups. The 'non-targeted employee' user type indicates stakeholders who participated in the crowd-based elicitation but were not part of the targeted user group.

TABLE IV
ACTIVITY PER USER TYPE IN THE S-SYS CASE STUDY (N=135).

Origin	% of total	Per user activity		
		Ideas	Votes	Logins
Operational employee	58.52%	0.23	2.89	1.84
Middle management	8.15%	0.82	4.00	2.55
Non-targeted employee	33.34%	0.11	0.88	1.55

In absolute terms, operational employees were more present on the platform and posted more ideas. However, per employee, middle management was more active: operational employees posted 0.23 ideas per user, while middle management posted 0.82 ideas per user. We can therefore imply that middle management was most active on the platform, in line with the expectations, as the military culture is structured around rank. Middle management also voted more per user: 4 votes per user, as opposed to 2.89 votes for operational employees. Therefore, we conclude that the operational employees—a user category that would seldom be included using traditional elicitation methods—were reached and that they delivered substantial input, but the participation per user was lower than that of middle management.

C. Appreciation of the KMar-Crowd Method and Tool

The questionnaire included in the online appendix was used to measure the appreciation of the *KMar-Crowd* method and of the support tool. The questionnaire was shared with all participants. We received 29 (21.5%) responses to the way of working questions, 28 (20.7%) of which also including the gamification questions: one respondent quit the survey after the first set of questions. All respondents were evenly distributed amongst the different user groups discussed above. A summary of the results of this questionnaire is visualized in the diverging stacked-bar chart of Fig. 6, with the questions split according two main areas: the appreciation of the *way of working* with *KMar-Crowd* and the opinion on the *gamification*.

The way of working was rated positively by the respondents: employees indicated that they appreciated the fact that they were asked to participate (Mdn. = agree) and they were enthusiastic to think about IT development in this way (Mdn. = agree) although they are a bit sceptic that their ideas would actually be used (Mdn. = neutral). Employees were less enthusiastic about the gamification elements, indicating that they were not motivated by the gamification element (Mdn. = disagree). When asked whether they would have interacted less without the gamification element, respondents gave a neutral response (Mdn. = neutral).

D. Usefulness of the Ideas

Finally, the usefulness of the gathered ideas was measured, by comparing them with the requirements already gathered using more traditional means. This analysis was executed by the two requirements engineers who gathered these requirements. They judged the ideas on three factors: (i) their classification according to the KANO model [30] to determine their innovative nature; (ii) whether they had been gathered earlier through traditional techniques; and (iii) whether the idea is complete enough to define a requirement for development, or, on the other hand, the idea would need additional investigation.

The results of this analysis are presented in Table V. The requirements engineers indicated that of the 32 ideas, 2 were unrelated to the goal of the elicitation process and therefore excluded. 19 of these 30 ideas were identified in an earlier stage, 5 were partly identified in an earlier stage, and 6 were completely new. Furthermore, when evaluating the ideas according to the KANO model, 13 of them were must-be requirements, 10 were one-dimensional (i.e., detrimental if not implemented, useful when implemented), and 7 were attractive qualities, i.e., delighters. This shows that the CrowdRE activities could contribute to enriching the requirements, although many of the inputs were already identified earlier. Finally, 2/3 of the ideas were missing important details prior to their use for development: this is not surprising, since involving the crowd of users amounts to allowing people with no RE experience to participate.

TABLE V
USEFULNESS OF THE IDEAS IN THE S-SYS CASE STUDY, ASSESSED BY THE ANALYSTS WHO CONDUCTED ELICITATION WITHOUT *KMar-Crowd*.

Measurement	Value	# Ideas
KANO model	Must-be	13
	One-dimensional	10
	Attractive	7
Gathered earlier	Completely	19
	Partly	6
	Not at all	5
Complete for dev teams	Yes	11
	No	19

An example of an idea which needs to be analysed further is about push notifications, as the requirements engineers could not identify what information would be handy to provide in the push notification, and when the push notification needs to be send. Another example is the idea of adding a certain data field. Including this field would require employees to fill it in, even though the data can be automatically distilled from other data fields. Adding this extra field might therefore lead to lower efficiency. The idea also needs to be investigated further, as the request does show that certain information is not easily accessible right now.

VI. V-SYS: DOES A SCALE-UP OF THE *KMar-Crowd* LEAD TO USABLE RESULTS?

The second case study focuses on an application that had to be built yet, and no requirements elicitation efforts were made earlier. This to-be application, V-Sys, is a replacement for another operational registration system, which is widely used by several operational brigades. V-Sys is intended for reporting on all border-related processes, such as requests for asylum or visa. Unlike the S-Sys case, every operational

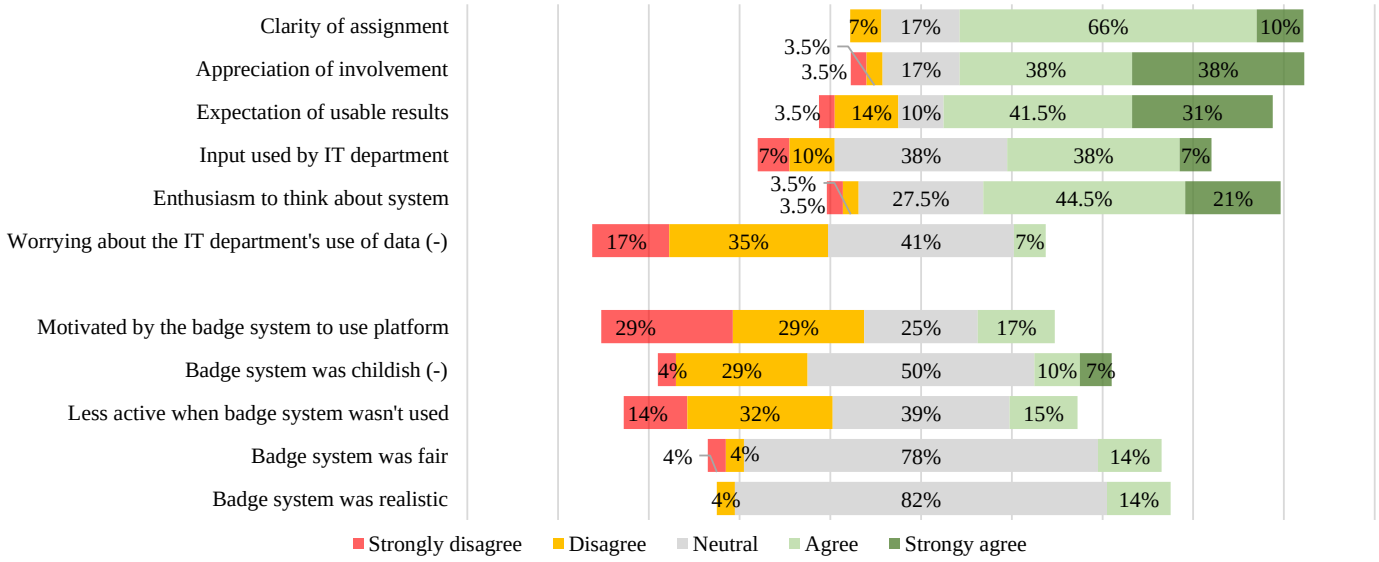

Fig. 6. Results of the questionnaire held in the S-Sys case study, showing *way of working* (N=29) in the upper half and *gamification* (N=28) in the lower half. The answers to the questions with a (-) label were inverted, due to the way these questions were formulated in the questionnaire.

brigade in the organization was asked to participate by e-mailing the commander of that brigade. All brigades where the application was widely used gave permission to send out the e-mail and add the briefing slide. In total, 2,393 employees spread over 13 brigades were asked to participate. As not all brigade commanders responded at the same time and due to time constraints, the roll-out was executed gradually. The total duration of this case study was 8 weeks. After three weeks, all employees received a reminder e-mail, and the briefing slide was reintroduced for two days. After the period of eight weeks was finished, employees received the same questionnaire we used for the first case study.

The most important indicators are presented in Table VI. In the following, we discuss the results according to the four dimensions considered for S-Sys.

TABLE VI
USAGE INDICATORS FOR THE V-SYS CASE STUDY.

Indicator	Total
Unique users	385
Ideas	78
Votes	531
Login actions	623
Comments	78
Users upgraded to two stars	11
Users upgraded to three stars	5
Users upgraded to four stars	2
Users upgraded to five starts	0
Page hits	3,940

A. User Engagement

Fig. 7 summarizes user engagement for the V-Sys case. While two peaks were visible in the S-Sys case (see Fig. 4), employees were invited more gradually in the V-Sys case, once their brigade commander gave consent. Because of this, the activity on the *KMar-Crowd* platform was more spread out over time. In total, 385 participants used the platform, which is 15.8% of the total invited employees. This is a bit lower than in the S-Sys case (28.25%), since the larger scale of this case study made it harder for the researcher to pay attention to the participation of all brigades. As for S-Sys, a correlation between registrations and the other activities is visible.

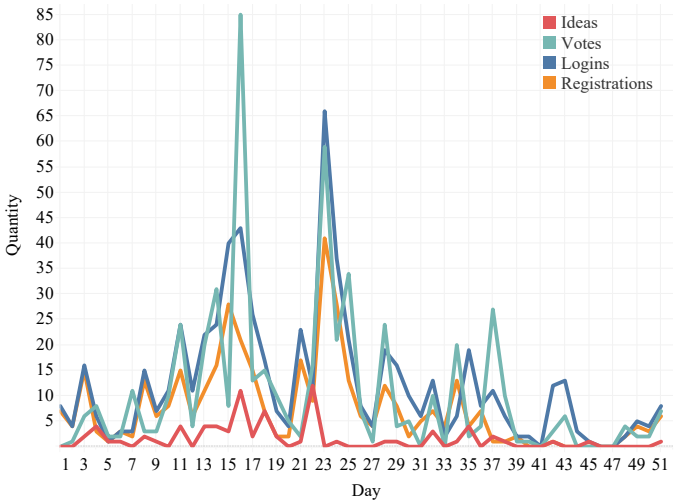

Fig. 7. Usage indicators for the V-Sys case study plotted over time.

As in the first case study, users were clustered by their participation level. The results are visible through the blue bars of Fig. 5 above. While 33.77% of the participants interacted in some form with the platform, 66.23% of the participants did only visit the platform without adding ideas, voting, or commenting. Although the total number of interactions were higher during the V-Sys case study, the interaction was somewhat lower, as visible by comparing the blue bars of Fig. 5 against the red bars. One possible reason for this is the

principle of *social loafing*, which is defined as "the tendency for individuals to expend less effort when working collectively than when working alone" [31].

B. User Origin

Since the second case study was executed across different brigades, no distinction was possible between operational employees, middle management and upper management, as each brigade was structured differently. Therefore, an analysis based on rank was made. Generally, soldiers[3], corporals and non-commissioned officers (NCOs) can be interpreted as operational employees, subaltern officers can be seen as middle management and head-officers can be seen as upper management. The *KMar-Crowd* was particularly aimed towards soldiers, corporals and NCOs.

TABLE VII
ACTIVITY PER USER RANK IN THE V-SYS CASE STUDY (N=385).

Origin	% of total	Per user activity		
		Ideas	Votes	Logins
Soldiers	0.00%	-	-	-
Corporals	3.64%	0.36	3.43	2.43
NCOs	76.62%	0.22	1.41	1.63
Sub-altern officers	11.43%	0.10	1.45	1.68
Head officers	2.34%	0.00	0.00	1.22
Citizens	5.97%	0.09	0.13	1.30

Table VII summarizes activity sorted by rank. It can clearly be observed that most participants were NCOs. This could be expected: users of the application to be developed are primarily NCOs. There were not many corporals amongst the participants, but the corporals who did participate were very active: the corporal group submitted the most ideas and votes per user. No soldiers were active on the platform. Although it is not clear why soldiers did not participate, one possible reason might be that they do not have enough experience to give meaningful input. Another reason might be that starting employees with a low rank might not feel confident enough to share their opinion.

C. Appreciation of the Method and Tool

As during the first case study, a questionnaire was held amongst all participants. 96 (24.9%) participants finished the way of working part, 88 of which also filled in the gamification part (22.86%). The results of this questionnaire are visualized in Fig. 8. Generally, the participants agreed with those of the first case study. The way of working was appreciated, as participants indicated that they liked being involved in this way (Mdn. = agree) and that they expected that a usable result would be produced by using the *KMar-Crowd* (Mdn. = agree). These results are in line with the first case study, and the difference in the spread of these questions is within 5% across the case studies. The gamification element was, again, not appreciated by the participants: participants indicated that it did not motivate them to use the platform (Mdn. = disagree) and that they would not have been less active when the badge system was not used (Mdn. = disagree). In this case study, this opinion was more polarized than during the first case study.

[3]We use 'Soldiers' to indicate NATO ranks OR-1 and 2, 'Corporals' for NATO ranks OR-3 and 4, 'NCOs' for NATO ranks OR-5 up to 9, 'Sub-altern officers' for NATO ranks OF-1 and 2, and 'Head officers' for NATO ranks OF-3 up to 5. NATO ranks are from the STANAG 2116 standard.

D. Usefulness of the Ideas

For V-Sys, since no earlier requirements elicitation work was conducted, we could not reuse all the same measurements we employed for S-Sys. While we kept the KANO model classification, we introduced new measurements. We asked a pool of 4 requirements engineers to judge whether the ideas as collected by the *KMar-Crowd* would be sufficiently detailed for a minimum viable product (MVP) as well as for a complete and correct implementation of the requirement. The difference between these two can best be explained by whether a requirement is fulfilled completely: in a MVP, some faults in the implementation of the requirement may still be present (e.g., some business rules are not correctly implemented), while in the final product, the requirement should be implemented completely and to the satisfaction of the end user. Finally, the requirements engineers were also asked whether the idea could be classified as a user story, or as an epic. Table VIII summarizes the results.

TABLE VIII
V-SYS: USEFULNESS OF THE IDEAS, ASSESSED BY A POOL OF ANALYSTS.

Measurement	Value	% Ideas
KANO model	Must-be	50.6
	One-dimensional	36.7
	Attractive	12.7
Enough for MVP		59.5
Enough for product		27.8
Granularity	Epic	40.5
	User Story	54.4
	Not applicable	5.1

Of the gathered 85 ideas (some inputs were split for this analysis since they contained multiple ideas), 6 ideas were immediately dismissed, mostly because their implementation would be unfeasible due to legal reasons. Of the remaining 79 ideas, 59.5% were specific enough to implement in an MVP. Only 27.8% of the ideas were specific enough to implement in the final product. The results are, however, promising, since we are considering ideas provided by people who do not possess expertise in requirements engineering. Regarding granularity, 40.5% of the ideas were classified as epic, 54.4% were classified as a user story, and one (5.1%) idea regarded stakeholder identification.

Half (50.6%) of the ideas were classified as must-be. Examples of such ideas are a more elaborate implementation of a decision support system, or a responsive design of certain forms. 36.7% of the ideas were classified as one-dimensional. Examples are ideas about API-integration with other organizations and other systems within the KMar. 12.7% of the ideas were classified as attractive. For example, the registration of certain new variables or API-integration with less used systems.

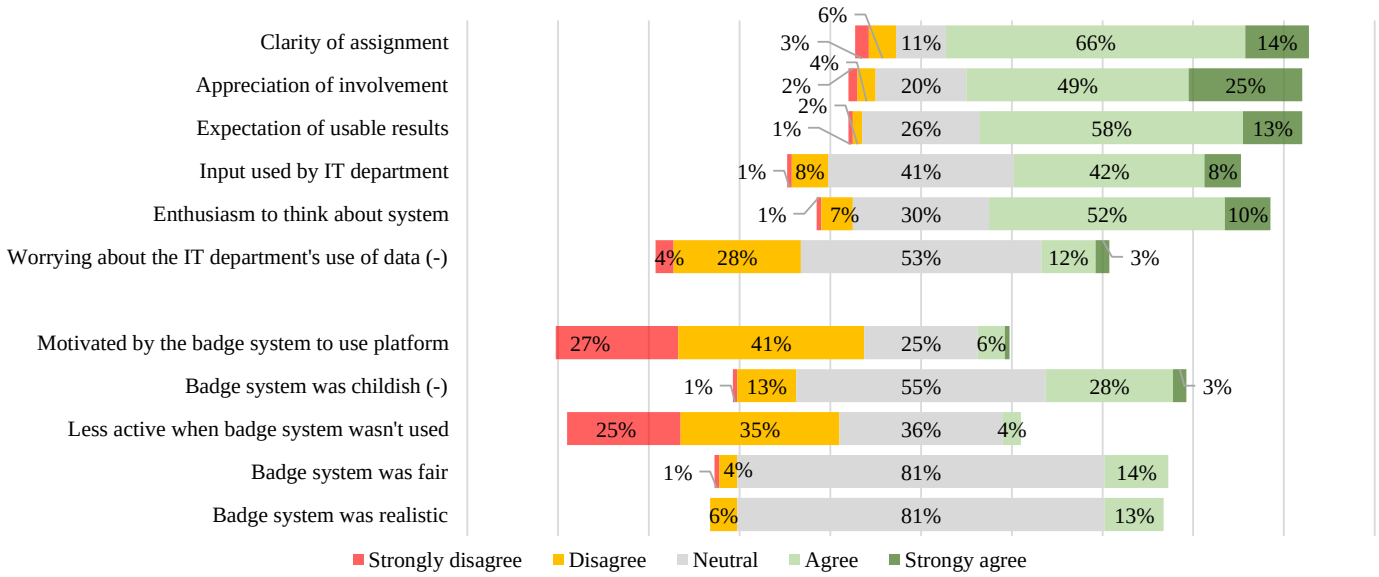

Fig. 8. Result of the questionnaire held in the V-Sys case study. For the way of working, N=96, while for the gamification, N=88.

VII. Conclusions and Future Directions

Both case studies indicate that CrowdRE can successfully reach and engage operational employees, regardless of their position in the organization. The crowds that emerged during our studies are larger than those in previous CrowdRE research, both in terms of number of participants as well as of their participation (e.g., number of ideas and votes). The collected ideas can be used by the requirements engineers to identify the themes they need to focus on during the rest of the elicitation process, and to perform stakeholder identification.

Governmental organizations are complex: in addition to the wishes of the primary users of the applications, the interests and opinions of many other stakeholders need to be taken into account, including partner organizations, the legislator, the parliament, political organizations, and the public opinion. Given this organizational context, CrowdRE should be seen as a good first step to gather initial ideas from the primary users, but its is not a complete replacement of the requirements elicitation and prioritization process.

The gamification elements tested in this study (the classic points, badges, and leaderboard triad, or PBL) did not have any significant effect on the motivation of participants to use the platform. Generally, participants were sufficiently motivated by the opportunity to actively contribute to making a software product, and therefore their work processes, better.

In Dutch governmental organizations, the PRINCE2 project management approach [32] is widely adopted in the software requirements engineering process. This poses challenges for the embedding of crowd ideas: the CrowdRE efforts shall be planned wisely so that the inputs from the crowd can quickly feed into the software development. This is necessary to sustain the engagement of crowd members, who want to see their ideas considered and responded to quickly.

A. Limitations and Research Opportunities

This study assumes that the Royal Netherlands Marechaussee is a good representation of a governmental organization. However, each governmental organization has its own rules, practices and culture; thus, generalization has limited power. Moreover, the study did not encompass a full development cycle, as process of developing an software product takes extensive time. A longitudinal study within governmental organizations might give an additional insight whether the application of CrowdRE can be successful, and on the role of CrowdRE practices after the initial elicitation.

Additionally, the case studies in this research focused on applications which would eventually replace other, out-of-date applications. The application of CrowdRE for a brand new application might lead to different conclusions, as employees would not have an example to base their ideas on. The use of CrowdRE in the context of the evolution or continuous development of an already implemented software product might also show different results. Therefore, future research needs to be conducted to investigate these alternative settings, both in governmental organizations and elsewhere.

Finally, we designed *KMar-Crowd* for governmental organizations who are still using waterfall-like methods to develop software products, often due to legal aspects such as the identification of contractors for the development via a tender. Further research is needed to embed *KMar-Crowd* with agile development practices, yet yielding the ability to control the software development process in the complex environments where governmental organizations operate in.

Acknowledgment

The authors would like to thank the participants in our two case studies, as well as the members of the RE-Lab at Utrecht University for the discussions around the research.

REFERENCES

[1] E. C. Groen, N. Seyff, R. Ali, F. Dalpiaz, J. Doerr, E. Guzman, M. Hosseini, J. Marco, M. Oriol, A. Perini *et al.*, "The crowd in requirements engineering: The landscape and challenges," *IEEE software*, vol. 34, no. 2, pp. 44–52, 2017.

[2] T. Johann and W. Maalej, "Democratic mass participation of users in requirements engineering?" in *Proc. of the International Requirements Engineering Conference (RE)*. IEEE, 2015, pp. 256–261.

[3] J. A. Khan, L. Liu, L. Wen, and R. Ali, "Crowd intelligence in requirements engineering: Current status and future directions," in *Proc. of the International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ)*, ser. LNCS, vol. 11412, 2019, pp. 245–261.

[4] R. Snijders, F. Dalpiaz, S. Brinkkemper, M. Hosseini, R. Ali, and A. Ozum, "REfine: A gamified platform for participatory requirements engineering," in *Proc. of the International Workshop on Crowd-Based Requirements Engineering (CrowdRE)*. IEEE, 2015, pp. 1–6.

[5] A. Menkveld, S. Brinkkemper, and F. Dalpiaz, "User story writing in crowd requirements engineering: The case of a web application for sports tournament planning," in *Proc. of the International Workshop on Crowd-Based Requirements Engineering (CrowdRE)*. IEEE, 2019, pp. 174–179.

[6] D. Renzel, M. Behrendt, R. Klamma, and M. Jarke, "Requirements bazaar: Social requirements engineering for community-driven innovation," in *Proc. of the International Requirements Engineering Conference (RE)*. IEEE, 2013, pp. 326–327.

[7] M. Z. Kolpondinos and M. Glinz, "GARUSO: A gamification approach for involving stakeholders outside organizational reach in requirements engineering," *Requirements Engineering*, vol. 25, pp. 185–212, 2020.

[8] S. Bretschneider, "Management information systems in public and private organizations: An empirical test," *Public Administration Review*, pp. 536–545, 1990.

[9] S. Lauesen and J. P. Vium, "Communication gaps in a tender process," *Requirements Engineering*, vol. 10, no. 4, pp. 247–261, 2005.

[10] E. C. Groen, J. Doerr, and S. Adam, "Towards crowd-based requirements engineering a research preview," in *Proc. of the International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ)*. Springer, 2015, pp. 247–253.

[11] R. Snijders, F. Dalpiaz, M. Hosseini, A. Shahri, and R. Ali, "Crowd-centric requirements engineering," in *Proc. of the International Workshop on Crowdsourcing and Gamification in the Cloud (CGCloud 2014)*. IEEE, 2014.

[12] M. Hosseini, K. Phalp, J. Taylor, and R. Ali, "Towards crowdsourcing for requirements engineering," in *Joint Proceedings of REFSQ-2014 Workshops, Doctoral Symposium, Empirical Track, and Posters*, ser. CEUR Workshop Proceedings, vol. 1138. CEUR-WS.org, 2014, pp. 82–101.

[13] M. Glinz, "CrowdRE: Achievements, opportunities and pitfalls," in *Proc. of the International Workshop on Crowd-Based Requirements Engineering (CrowdRE)*. IEEE, 2019, pp. 172–173.

[14] D. Pagano and W. Maalej, "User feedback in the appstore: An empirical study," in *Proc. of the International Requirements Engineering Conference (RE)*. IEEE, 2013, pp. 125–134.

[15] W. Maalej, Z. Kurtanović, H. Nabil, and C. Stanik, "On the automatic classification of app reviews," *Requirements Engineering*, vol. 21, no. 3, pp. 311–331, 2016.

[16] S. Panichella, A. Di Sorbo, E. Guzman, C. A. Visaggio, G. Canfora, and H. C. Gall, "How can i improve my app? classifying user reviews for software maintenance and evolution," in *Proc. of the International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, 2015, pp. 281–290.

[17] F. Dalpiaz and M. Parente, "RE-SWOT: From User Feedback to Requirements via Competitor Analysis," in *Proc. of the International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ)*, ser. LNCS, vol. 11412, 2019.

[18] T. Johann, C. Stanik, W. Maalej *et al.*, "Safe: A simple approach for feature extraction from app descriptions and app reviews," in *2017 IEEE 25th International Requirements Engineering Conference (RE)*. IEEE, 2017, pp. 21–30.

[19] E. Guzman and W. Maalej, "How do users like this feature? a fine grained sentiment analysis of app reviews," in *Proc. of the International Requirements Engineering Conference (RE)*. IEEE, 2014, pp. 153–162.

[20] G. Williams and A. Mahmoud, "Mining twitter feeds for software user requirements," in *Proc. of the International Requirements Engineering Conference (RE)*. IEEE, 2017, pp. 1–10.

[21] E. Guzman, R. Alkadhi, and N. Seyff, "A needle in a haystack: What do twitter users say about software?" in *Proc. of the International Requirements Engineering Conference (RE)*. IEEE, 2016, pp. 96–105.

[22] I. Morales-Ramirez, F. M. Kifetew, and A. Perini, "Analysis of online discussions in support of requirements discovery," in *Proc. of the International Conference on Advanced Information Systems Engineering (CAiSE)*, ser. LNCS, vol. 10253. Springer, 2017, pp. 159–174.

[23] G. M. Kanchev, P. K. Murukannaiah, A. K. Chopra, and P. Sawyer, "Canary: Extracting requirements-related information from online discussions," in *Proc. of the International Requirements Engineering Conference (RE)*. IEEE, 2017, pp. 31–40.

[24] J. Tizard, H. Wang, L. Yohannes, and K. Blincoe, "Can a conversation paint a picture? mining requirements in software forums," in *Proc. of the International Requirements Engineering Conference (RE)*. IEEE, 2019, pp. 17–27.

[25] P. K. Murukannaiah, N. Ajmeri, and M. P. Singh, "Acquiring creative requirements from the crowd: Understanding the influences of personality and creative potential in Crowd RE," in *Proc. of the International Requirements Engineering Conference (RE)*. IEEE, 2016, pp. 176–185.

[26] M. van Vliet, E. C. Groen, F. Dalpiaz, and S. Brinkkemper, "Identifying and classifying user requirements in online feedback via crowdsourcing," in *Proc. of the International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ)*, ser. LNCS, vol. 12045. Springer, 2020, pp. 143–159.

[27] T. D. Breaux and F. Schaub, "Scaling requirements extraction to the crowd: Experiments with privacy policies," in *Proc. of the International Requirements Engineering Conference (RE)*. IEEE, 2014, pp. 163–172.

[28] R. Davison, M. G. Martinsons, and N. Kock, "Principles of canonical action research," *Information Systems Journal*, vol. 14, no. 1, pp. 65–86, 2004.

[29] A. Cropley, "In praise of convergent thinking," *Creativity Research Journal*, vol. 18, no. 3, pp. 391–404, 2006.

[30] C. Berger, R. Blauth, and D. Boger, "Kano's methods for understanding customer-defind quality," *Center for Quality of Management Journal*, 1993.

[31] S. J. Karau and K. D. Williams, "Social loafing: A meta-analytic review and theoretical integration." *Journal of Personality and Social Psychology*, vol. 65, no. 4, p. 681, 1993.

[32] C. Bentley, *Prince2: A practical handbook*. Routledge, 2010.