Noname manuscript No.
(will be inserted by the editor)

A Collaborative Method for Identification and Selection of Data Sources in Market-driven requirements engineering

Authors removed for blind review

Received: date / Accepted: date

Abstract Requirements engineering (RE) literature acknowledges the importance of early stakeholder identification. However, the challenge of identifying and selecting the right stakeholders and the potential of using other inanimate requirements sources for RE activities for market-driven products is not extensively addressed.

Market-driven products are influenced by a large number of stakeholders. Consulting all stakeholders directly is impractical, and companies utilize indirect data sources, e.g. documents and representatives of larger groups of stakeholders. However, without a systematic approach, companies often use easy to access or hard to ignore data sources for RE activities. As a consequence, companies waste resources on collecting irrelevant data or develop the product based on the sub-optimal information sources that may lead to missing market opportunities.

We propose a collaborative method to support identification and selection of the most relevant data sources for MDRE. The method consists of four steps and aims to build consensus between different perspectives in an organization and facilitates the identification of the most relevant data sources. We demonstrate the use of the method with two industrial case studies.

Our results show that the method can support identification and selection of the most relevant data sources for MDRE in two ways: (1) by providing systematic steps to identify and prioritize data sources for RE, and (2) by highlighting and resolving discrepancies between different perspectives in an organization.

1 Introduction

The success of software projects is determined by the degree of stakeholder satisfaction [1]. Requirements engineering (RE) aims to elicit stakeholder needs, constraints and wishes to support the rest of software engineering activities [26, 28]. Software companies operating in a Market-driven Requirements Engineering context (MDRE) [50] need to accurately identify many heterogeneous requirements sources, e.g. customers, end-users, prospects, partners, competitors, domain experts, internal engineering, business, marketing, sales representatives, external in-

Address(es) of author(s) should be given

vestors, regulators to name a few. Such a large and diverse group of stakeholders with different interests, inconsistent needs, and varying levels of commitment makes requirements engineering challenging [25, 6, 50] and greatly contributes to overloaded requirements engineering [50]. Therefore, timely filtering of requirements sources emerges as one of the crucial factors in MDRE. Digitalization and pervasive connectivity significantly contribute to a paradigm shift towards data-driven user-centered identification, prioritization, and management of software requirements [40].

The Software Engineering Body of Knowledge (SWEBOK [28]), the Software Engineering Institute (SEI [56]) and the Rational Unified Process (RUP [35]) acknowledge the importance of stakeholder identification and highlight that relevant stakeholders should be identified early in the product life-cycle process. At the same time, these reports seem to underestimate the challenges associated with it, especially in MDRE. The focus is primarily on categorizing [56, 55, 48, 57, 38] rather than exploring relationships and dynamic interactions, as only some authors emphasize interactions [54, 47].

Stakeholders in MDRE often are unaware of their roles, have difficulty articulating their needs towards a new product, or could be uninterested to collaborate [19]. Furthermore, requirements engineers may not be aware of all relevant stakeholders [65, 45]. Therefore, software requirements often cannot be elicited from stakeholders in a traditional sense, e.g., by arranging elicitation interviews. Instead, the stakeholders and their needs emerge from continuous interactions between the market, developers, and the product [3, 23]. Frequent releases and analyzing customer feedback are the essential practices to learn about stakeholders' needs [4, 18, 36].

Interactions among a large number of stakeholders produce a continuous flow of data, such as feature ideas, feedback, problem reports, requests for specific customizations, product usage data, market analysis reports and so on [18]. Documenting, analyzing, and prioritizing input from all these sources is impractical [4, 30]. As a consequence, companies often opt for the most natural path and consult with the most accessible sources of information [32, 44]. For example, by inventing requirements internally [30], and by responding to customer-specific feature requests [34]. However, when considering the needs of few stakeholders instead of a broader market, companies may miss the opportunities for growth and revenue that market-driven products can offer [44, 4]. Finally, although some consider multiple roles (viewpoints) in requirements prioritization [66], considering and balancing various data sources remains largely unexplored.

In this paper, we present a method for identification and selection of the most relevant data sources for MDRE. The method is aimed to a) support identification of relevant data sources, b) support ranking of data sources by their relevance, c) support prioritization of data sources in changing market needs and contextual factors.

The rest of this paper is structured as follows. Section 2 provides background and summarizes related work in the area, Section 3 describes the steps the research methodology, Section 4 presents the method and its steps, Sections 5.1 and 5.2 describes a static validation of the method, Section 8 concludes the paper.

2 Background and Related Work

RE is an integral part of a software product strategy definition that shapes the management of software products. Software Product Management (SPM) is a collaborative effort comprising of, e.g. marketing, sales, engineering, operations, and business development perspectives [31]. The roles involved in product development may differently perceive the importance of different data sources and stakeholders behind them. This contributes to the challenge of reaching a consensus within an organization of which data sources to select.

2.1 Data sources in MDRE

Bespoke requirements engineering typically elicits requirements from a limited number of customer representatives via interviews, focus groups, observations or similar methods [28]. In MDRE, companies are part of a loose network of customers, end-users, prospects, technology trends, partners, suppliers, other products, competitors and alike [50]. Requirements towards the product emerge from interactions between stakeholders and the product context [23, 29].

The main challenges are that there are too many potential requirements sources and not all are equally promising or important [30, 18]. The interactions with a large number of stakeholders produce a continuous flow of feature ideas, feedback, problem reports, requests for specific customizations, and so on [18]. Moreover, connectivity and social media make it very easy to submit feedback about the software products. The consequence is an inevitable evolution of MDRE into collaborative, data-driven and user-centered identification, prioritization and management of software requirements [40]. The emerging importance of implicit requirements feedback from product usage data calls for improved methods for filtering and synthesizing large amounts of user data.

Equally investing in documenting, analyzing, and prioritizing requirements from each data source is impractical [4, 30]. As a consequence, companies consult with the most accessible sources [32, 44], invent requirements internally [30], or respond to customer-specific feature requests [34, 33]. However, for new and innovative products candidate data sources and their relevance may be unknown, with the consequence that companies may miss growth opportunities that market-driven products can offer [44, 4].

2.2 Review of existing methods

A recent systematic review identifies a number of methods for stakeholder identification and quantification [27]. These methods propose stakeholder high level classes, e.g. into customers, development team, and business [9, 24, 44], and triage into mandatory, optional, and nice-to-have stakeholders [49]. However, these methods are not specifically designed for MDRE and lack support for identification and classification of inanimate data sources.

We perform a follow-up search based on the results from Hujainah et al. [27] and Bano et al. [12]. We use snowball sampling to identify exiting stakeholder

Table 1 Evaluation of existing stakeholder selection methods

Method summary	Support for different types of data sources	Support for collaboration	Support for adaptation
Barbar et al. [8] analyzes skill and level of interest of stakeholders	No	No	No
Razali and Anwar [49] considers mandatory, optional, and nice to have stakeholders and analyzes their knowledge, interest, and interpersonal skills.	No	Partial	No
Bendjenna et al. [15] considers stakeholder power, legitimacy, and urgency	No	No	No
Barbar et al. [10] focuses on classifying stakeholders based on their personal and professional characteristics.	No	No	No
Burnay [17] proposes a taxonomy of requirements elicitation sources comprising of people, organization, artefacts, processes, and environment.	Yes	-	No
Alexander [2] proposes a model for identifying stakeholders and surrogate requirements sources.	Yes	-	Yes
McManus [42] presents guiding questions to identify and classify relevant stakeholders	Partial	No	No
Ballejos and Montagna [11] proposes a model for stakeholder representation comprising of roles, interests, influences, power, and goals.	No	No	Yes
Lim et al. [39] proposes a method to identify and prioritize stakeholders using social networks analysis methods.	Yes	-	Yes

identification and classification methods. We evaluate identified methods for suitability for use in MDRE and crowd requirements engineering contexts [23, 40]. We set forth the following evaluation criteria:

1. Support for identification and classification of different types of data sources such as individual people, large groups of people, documents, artifacts, product usage patterns and alike [40, 2].
2. Support for collaboration between multiple analysts by capturing and illustrating different perspectives, in contrast to capturing only consensus view from the group [7].
3. Support for adaptation of the method for use in different contexts. Results of a method depends on how well it is suited for use in the given context [46].

We summarize our results in Table 1. We denote a method and criteria with “Yes” if the method description clearly addresses the criteria. We use “Partial” to denote potential method support for the criteria through an extension or adaptation, we use ”No” to denote a clear lack of support. Cases where a criterion is not applicable are denoted with “-”.

By reviewing the existing methods we observed several patterns. First, we observe that several methods, for example, Barbar et al. [8, 10], Razali and Anwar [49], and Bendjenna et al. [15], uses predefined generalized criteria for analyzing stakeholders. The proposed criteria, such as interest and communication skills, are relevant for human stakeholders and cannot be applied to artifacts. Thus, these methods are context specific and difficult to are tailor for use in a crowd RE setting.

Ballejos and Montagna [11] and Lim et al. [39] proposes methods for analyzing relationships between already known stakeholders. Both professional and personal relationships are relevant only for human stakeholders and cannot be applied to inanimate objects, e.g. product usage data.

None of the reviewed methods with partial exception of Razali and Anwar [49], support collaboration of multiple analysts. They suggest to extend the proposed method with prioritization techniques supporting collaboration, such as AHP [53].

We conclude that none of the reviewed methods is suitable for use identification and selection ofrelevant data sources for MDRE. Our work occupies this research gap.

2.3 Decision making scenarios

Software engineering is a collaborative activity and requires cooperation between multiple individuals [63]. Requirements engineering is a decision making activity to determine what functional or non-functional features/requirements are needed [7].

In general, there are two scenarios of decision making situations [41]:

The group-based approaches aim to achieve consensus by discussion and negotiation between the group members. This approach carries the risk that the final outcome is dictated by more powerful members of the group. The risk is particularly high in groups consisting of individuals from different levels of organizational hierarchy. Negotiations and discussions are possible only in relatively small groups, e.g. agile teams. Furthermore, individual views in the group could be conflicting to a degree that a consensus is impossible without compromising the final decision or rejecting conflicting decision items. Thus, group-based approaches are feasible in relatively small and rather aligned groups.

Individual-based approaches aim to support individuals in communicating their perspective and aggregating individual views into a final decision. Such approach treats each individual perspective as equal regardless of rank and negotiation power. However, individual based approaches suffer from the neglect of interactions, exchanges of ideas, and arguments that arise from discussions in the group.

As shown by Tindale et al. [58], the degree to which ideas, preferences, knowledge, etc., are shared in the group determine the acceptance of the final decision, and affects the quality and accuracy of the final decision. However, group discussions are not feasible in MDRE context with a large number of stakeholders, enriched by product-usage data. Thus, eliciting and aggregating individual perspectives is a more promising approach.

3 Research Methodology

We follow the design science research method (DSRM) by Wieringa [64]. DSRM describes the steps to design an artifact considering two perspectives: i) a specific real-life situation requiring a practical solution, and ii) an abstract, research view of the problem and solution. The method consists of 6 steps, problem identification and motivation, definition of objectives, design and development, demonstration, evaluation, and communication. In the following subsections we describe each step in detail.

3.1 Research objectives

In this section, we formulate the research problem with a tempalate propsed by Wieringa [64]. The objective of this study is to:

- Improve identification and selection of the most relevant data sources for MDRE
- with a systematic method
- ensuring transparency and building consensus between different perspectives
- to select the most relevant data sources for a given task.

To guide our research we formulate the following research questions:

RQ1: What are the needs towards a method for identification and selection of the most relevant data sources for MDRE?
Rationale: By answering this research formulate specific requirements towards the method.

RQ2: How a method can support identification and selection of the most relevant data sources for MDRE?
Rationale: By answering this research question we aim to propose a method design.

RQ3: What improvements to the method are needed for its use in industry?
Rationale: We aim to validate the method in an industrial setting to collect input for its further development.

The research questions are answered in an iterative process of the design science methodology [64]. We identify the specific challenges of selecting stakeholders and data sources, and review existing methods in Section 2. We outline specific objectives of the method in Section 3.3, the design steps are presented in Section 4. Objectives and steps of the validation are outlined in Section 3.5, results of the validation are presented in Section 5. We discuss our results in Section 7.

3.2 Step 1: Problem identification and motivation

This research is inspired by related work on innovative software-intensive product engineering in MDRE and known challenges in MDRE [30]. Several authors, e.g. Klotins et al. [34, 33], Jin et al. [29], and Groen et al. [23], identify requirements engineering, and specifically the identification and selection of data sources representing key stakeholders, as one of the key practices to ensure that the product offers the right set of features and is technically and commercially successful. The similar challenge of identifying the right sources and filtering out the most promising requirements makes MDRE difficult [30]. As summarized in Table 1 there exist a shortage of methods for identification and classification of data sources (both human and machine) in MDRE. In particular, there is a need for developing methods that can support different types of data sources, collaboration between multiple analysts by capturing and illustrating different perspectives and offer adaptation for various contexts. Identification of appropriate data sources is central to understanding the context in which the product is developed and operated [59].

3.3 Step 2: Objectives of the method

The main objective of the method is to support identification and selection of the most relevant data sources for MDRE, e.g. market analytics, people, product usage analytics, sensor data, and similar, for consideration in new, innovative, market-driven, software-intensive product engineering. We decompose the main objective into sub-goals:

1. Support for adaptation of the method for solving specific problems in different contexts
2. Support for identification and selection of the most relevant data sources
3. Support for different types of data sources such as individuals people, large groups, documents, artifacts, product usage data and alike
4. Support for collaboration and consensus building among multiple analysts but also capturing the individual viewpoints and disagreements
5. Transparency and understandability to comprehend the method and its output.

3.4 Step 3: Design of the method

The method was incrementally designed in a series of review rounds. At each round, the authors discussed design concerns, evaluated different potential solutions, and updated the method accordingly. We documented rationales behind our design choices and discussed potential alternative solutions, see Section 4.

3.5 Step 4: Demonstration

We perform a static validation of the method with the purpose to evaluate its usability and collect feedback for further development of the method [22]. The validation consists of the following steps:

1. We start by presenting the aims of our study and establishing a common vocabulary. Throughout the demonstration we avoid posing our own views, rather we elicit participants perspectives on data sources identification and selection, utilized practices, and challenges from their experience.
2. We ask the participants to describe a recent example where multiple data sources were selected and used in crafting ideas for further product development. We ask the participants to list all the data sources that were considered and, with hindsight, rank them according to their contribution to the outcome. The purpose of this step is to establish a baseline for comparison with the results from the method.

3. We continue by introducing the method step by step and ask participants to describe the trigger, list criteria, and candidate data sources that were relevant for the example. This step is done in a discussion format inquiring participants about motivation of selecting or ignoring certain data sources or criteria.
   The purpose of this step is to validate constructs of the method, their understandability, and practitioners ability to provide meaningful criteria and data sources.
4. We ask participants to provide scores for criteria, evaluations to data sources, and run the calculations to arrive to the final ranking. We use a spreadsheet for data collection.
   The purpose of this step is to arrive at ranking of data sources based on the method for comparison with the baseline.
5. Reflections on the results. At the end, we ask participants to reflect on the method and its results, compare it with the initial ranking, and provide their perspective on the differences.

We conduct two case studies and describe results in Section 5. Demonstrations are handled by the first and second authors with help of presentation materials and a semi-structured interview guide. Two practitioners were involved in each demonstration.

3.6 Steps 5-6: Evaluation and communication

Further evaluation and communication of the method is planned as continuation of this work. We aim to follow the technology transfer model by Gorschek et al. [22] and perform several industry feedback and method improvement rounds.

The design for evaluation is similar to what we outline in Step 4. We start by familiarizing the participants with the challenge of selecting data sources, then we establish a baseline, guide the participants through the method application, and help them to interpret the results. However, as the method matures we would like to both reduce researcher support for the use of the method and analyze new situations.

3.7 Threats to validity

We follow the guidelines by Runeson et al. [52] and discuss the four perspectives of validity threats.

3.7.1 Construct validity

Construct validity is concerned with establishing appropriate measures to observe the intended concepts. Key concepts of the method originates from related empirical work on innovative software-intensive product engineering and related work in MDRE and crowd RE. We conducted several review rounds to refine formulations and relationships of the concepts.

We further strengthen the construct validity during the case study when asking participants to discuss their view on practices and challenges associated with stakeholder and data sources selection.

A threat is introduced by us potentially influencing the participants views and responses. Knowing the objectives of our study, practitioners may inflate the challenges of data sources selection. Furthermore, participants may be reluctant to reveal their real opinions about the method because the authors were present. To mitigate this threat we avoid posing specific views, rather we introduce concepts neutrally and ask practitioners to reflect on their experience.

3.7.2 Internal validity

Internal validity is concerned with uncontrolled factors affecting causal relationships between concepts. The method infers that criteria are the only yardstick to rank data sources. There could be additional influences to ranking that do not fall under our definition of criteria. To minimize this threat, the method needs to be further validated and operationalized, as well as more work is needed towards understanding decision-making factors in stakeholder selection.

A potential limitation of the method is the need to avoid ties between the criteria. Having dependencies between criteria would impair rankings and the final results. We plan to explore different strategies for removing this limitation with further work.

3.7.3 External validity

External validity is concerned to what extent the results can be generalized outside the studied cases and remains the main limitation of our work. We cannot claim that the studied cases are representative to all companies operating in MDRE. Thus, the fitness of the method needs to be further validated. In particular, we plan to add support for the theoretical constructs and the usefulness of the method.

3.7.4 Reliability

concerns the degree of repeatability of the study. To support traceability and transparency, we described the rational and objective for each step of the research method and the case study. We also provide raw data and our calculations as supplemental material online [1] However, the method embodies ideas, knowledge, and interpretations that remains to some extent subjective.

4 The method

The proposed method is a structured approach to combine input from multiple analysts and produce a ranked list of the most appropriate data sources to analyze and consult in a given decision making situation. The data sources are ranked based on their attributes which are selected and estimated by the analysts. The method consists of four main steps, see Fig 1.

In *the Step 1*, the analysts identify a task and formulate a problem statement. The problem statement implies the need to decide and consult multiple stakeholders and data sources. In *the Step 2*, the analysts identify criteria that are relevant

[1] Link removed for blind review

for the task (problem) under decision on what data sources and stakeholders to consult. In *the Step 3*, the analysts evaluate the data sources according to the identified criteria. In *the Step 4*, the analysts analyze and interpret the results.

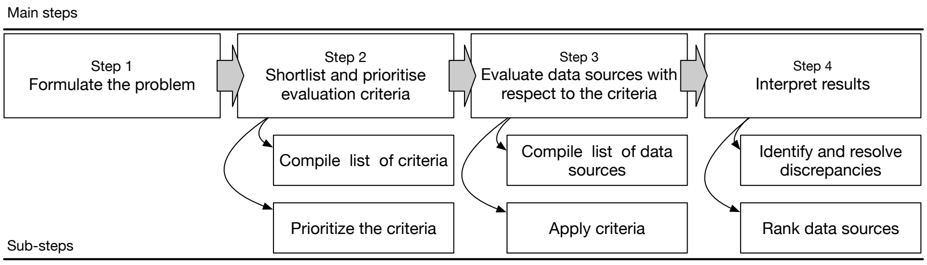

Fig. 1 Overview of the method. The method consists of 4 main steps with several sub-steps

4.1 Preconditions

We start the method design by identifying and outlining the core concepts, described below.

A trigger is circumstances leading up to the formulation of the problem statement. A trigger could be, for example, identification of a new market opportunity, a customer request for a specific feature, market pressure, or technological shift.

A problem statement is brief description of a non-trivial issue to be addressed or a situation to be improved. A problem statement describes the gap between the desired situation and the current, sub-optimal situation [5]. A problem statement should be quantifiable, e.g., number of lost customers, amount of potential revenue or saved resources.

Data sources may have different forms, for example, individual stakeholders (such as key customers, managers, or product engineers), the crowd [23] consisting of e.g. potential customers, opinion leaders, and users of competitor products, other organizations (such as competitors, partners, or suppliers), analytics (such as product usage data, analysis of customer data, or results from market analysis), industry standards, and applicable laws and regulations. Different data sources may offer different types of information and different perspectives on the decision.

We define a set of all data sources as:

$$D = \{d_1, d_2, \ldots, d_n\}.$$

Criteria are principles how data sources are selected and compared. The exact criteria depends on the problem to be addressed. For example, an internal engineer is easier to access than a customer. In turn, customers have more profound knowledge on their actual needs than engineers but may have a little insight in underlying technology constraints. Another example are the constructs used by [43] in his theory of stakeholder identification: power, legitimacy, urgency and salience.

We define a set of all potentially relevant criteria as:

$$C = \{c_1, c_2, \ldots, c_m\}.$$

An analyst is an individual who collaborates with other analysts on selecting the data sources and on the use of the method. We use the term analyst to refer to a range of people that could be involved in product decisions, such as, software engineers, product managers, financial officers, marketing, sales representatives, and alike. The analysts are involved in the decision making and are the primary stakeholders, that is, the analysts are the decision makers.

We define all analysts involved in the decision making as:

$$E = \{e_1, e_2, \ldots, e_k\}.$$

4.2 Step 1: Define the problem statement

The first step is to formulate the problem to be addressed. The problem formulation should contain the following information:

1. Description of the current situation.
2. Characterization of the desired situation.
3. Quantification of the difference between current and desired situation, e.g improve X by Y.
4. Preliminary candidate solutions to address the gap and reach the desired situation. The candidate solutions are to be evaluated and refined by the input from selected stakeholders.

4.3 Step 2: Shortlist and prioritize criteria

The second step is to identify and prioritize criteria for comparing potential data sources. This step has three sub-steps: 1) compile the initial list of criteria, 2) shortlist relevant criteria by voting, and 3) prioritize (weight) the criteria.

4.3.1 Compile a list of criteria

The analysts put together an initial list of criteria that will be used to compare different data sources. The aim is to identify all relevant characteristics of an optimal data source to resolve the problem defined in Step 1. The proper criteria are domain and problem specific and can be identified through, e.g. discussions and brainstorming.

In order to facilitate the identification of relevant criteria we adapt a taxonomy of requirements criteria proposed by Riegel et al. [51], see Table 2. We propose to use the high level criteria from the table as a starting point for the initial analysis and brainstorming to identify additional criteria.

To avoid inconsistencies in further prioritization and estimation, the analysts need to avoid ties between the criteria. For example, if criteria such as knowledge and domain knowledge appear together, analysts may have a difficulty to assign consistent scores [62]. A counter-strategy could be to identify potential ties early and break down higher level criteria into more fine-grained criteria, thus eliminating the ties.

Table 2 Example criteria of requirements to bootstrap data source criteria

Category	High level criteria	Examples of detailed criteria
Benefits in terms of knowledge	Level of knowledge	Domain, customer needs and wishes, product technologies or features, business processes, laws, regulations, standards
Benefits in terms of experience	Amount of experience	The product, similar products, specific market segments, domain, product technologies
Benefits in terms of financial value	Revenue	Customer life-time value, price per purchase
Costs	Associated costs	Access, establishing access, maintaining access
Penalties	Opportunity cost	Business, market, technical, financial, reputation, dissatisfaction of other stakeholders
Risks	Generic risk	Human factors, technical risks, implementation risks, volatility, business risks, time, budget, project scope, dependencies, reputation, legitimacy
Temporal context	Timing	Lead time, time to access, timeliness of data, frequency of access
Suitability for use	Ease of use	Ease of access, mutual trust, understandability, granularity, analyzability, accuracy, overhead
Behavioral	Suitability for collaboration	Availability, interest to contribute, commitment, volatility, trustworthiness, willingness to experiment, capacity volunteer resources for collaboration, power, leverage

4.3.2 Shortlist of relevant criteria

After the initial list of criteria is compiled, the analysts vote to shortlist more relevant and exclude less relevant criteria, thus reducing the effort in the next steps. Based on the votes and a cut-off threshold, relevant criteria are selected for further consideration.

Each analyst takes a list of all criteria and assigns a binary vote (relevant/not relevant) to each criterion. The votes produced by the different analysts on each criterion are summed. Criteria with votes above a given threshold are included for further consideration. For example, a majority vote can be used to decide which criteria to prioritize, i.e. we can apply a cut-off threshold of $k/2$, where k is the number of analysts.

Shortlisting may be relevant only if at least two analysts are involved. For example, if maximum two analysts are involved then a cutoff threshold of $k/2$ is 1. In this context criteria supported by both analysts should be considered or it may be decided to consider all criteria proposed by the analysts.

Assume that in this step a list of m candidate criteria and k analysts are involved in the voting. Each analyst e_i $(i = 1, 2, \ldots, k)$ casts a vote v_{ij} on criterion c_j $(j = 1, 2, \ldots, m)$, where $v_{ij} \in \{0, 1\}$, i.e. a binary vector $v_i = [v_{i1}, \ldots, v_{im}]$ presents the criteria votes for analyst i. The result is a $k \times m$ binary matrix V:

$$V = \begin{bmatrix} v_1 \\ \vdots \\ v_k \end{bmatrix} = \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & & \vdots \\ v_{k1} & \cdots & v_{km} \end{bmatrix}.$$

We further count votes V_j on each criterion j $(j = 1, 2, \ldots, m)$ by summing the values in each column of V, i.e.:

$$V_j = \sum_{i=1}^{k} v_{ij}.$$

The resulting set of criteria C' can be defined as:

$$C' = \{c_j | V_j > T\}, \text{where } C' \subseteq C \text{ and } T = k/2.$$

4.3.3 Prioritize the criteria

The purpose of this step is to prioritize the criteria by their relevance to the problem under decision. Each analyst assigns a score to each criterion assessing its relevance (importance). We propose to use an ordinal scale from 0 to 5, i.e., all scores in $\{0, 1, 2, 3, 4, 5\}$, where 0 indicates no relevance at all, 5 indicates the highest relevance, and the numbers 1, 2, 3, 4 represent intermediate values. In Table 3 we provide semantic meaning of each score. To arrive at the final ranking, the scores from each analyst are normalized.

Table 3 Semantic meaning of the measurement scale

Score	Meaning
0	Not relevant at all (to the given problem)
1	Marginally relevant
2	Somewhat relevant
3	Moderately relevant
4	Very relevant
5	Most relevant

Each analyst e_i $(i = 1, 2, \ldots, k)$ evaluates each criterion $c_j \in C'$ by assigning a score $w_{ij} \in \{0, 1, 2, 3, 4, 5\}$. The result is a vector $w_i = [w_{i1}, w_{i2}, .., w_{im'}]$ of scores produced for each analyst e_i $(i = 1, 2, \ldots, k)$, where m' is the cardinality of C' (the set of the selected relevant criteria).

We further calculate weights denoting the relative relevance (importance) of each criterion. Namely, the scores of each analyst e_i are normalized by dividing each value of vector w_i with the maximum given score by the analyst. Thus a vector $w'_i = [w'_{i1}, w'_{i2}, .., w'_{im'}]$ of weights for each involved analyst e_i $(i = 1, 2, \ldots, k)$ is generated, where

$$w'_{ij} = \frac{w_{ij}}{\max_{j=1}^{m'} w_{ij}}.$$

We average weights from all the analysts to calculate the overall weights, i.e., a vector $W = [W_1, W_2, .., W_{m'}]$, where the overall weight W_j of criterion j ($j = 1, 2, \ldots, m'$) is calculate as:

$$W_j = \frac{1}{k} \sum_{i=1}^{k} w'_{ij}.$$

The resulting vector W consists of the relative weights (importance) of the selected evaluation criteria.

The produced individual and overall weights can be studied and compared to better understand and interpret the problem under consideration. It could be useful to analyze whether there is a significant discrepancy in analysts' opinions about the criteria importance. Analysing discrepancies can help to improve transparency, quality and acceptance of the results [58].

4.4 Step 3: Evaluate data sources with respect to criteria

In the third step, the analysts compile a list of potential data sources and use the selected criteria to evaluate each data source. This step has two sub-steps: i) compile a list of data sources, and ii) apply criteria

4.4.1 Compile a list of data sources

The analysts put together a list of potential data sources. The list can be created by writing down already known data sources, brainstorming, or a combination of both. The aim is to identify a diverse set of potentially informative data sources for further consideration.

The proper data sources are usually domain specific. However, we propose to use the high level data sources given in Table 4 as a starting point for the initial analysis and brainstorming to identify additional data sources. Note that stakeholders can be accessed directly, e.g., by consulting with engineers and indirectly, e.g., by analyzing user behavior through product usage patterns.

4.4.2 Quantify the data sources (apply criteria)

The analysts apply the selected criteria to evaluate each data source. Each analyst e_i ($i = 1, 2, \ldots, k$) evaluates each data source d_j ($j = 1, 2, \ldots, n$) on a scale from 0 to 5, where 0 denotes the lowest, and 5 the highest score. Higher scores are awarded to more favorable evaluations, e.g., lower cost and higher accuracy. In Table 5 we provide the semantic meaning of each score. Note that we consider only favorable (positive) scores. The scores of each analyst e_i for the data sources given in D on the set of criteria C' produces a $n \times m'$ matrix U_i of weights:

$$U_i = \begin{bmatrix} u^i_{11} & \cdots & u^i_{1m'} \\ \vdots & & \vdots \\ u^i_{n1} & \cdots & u^i_{nm'} \end{bmatrix},$$

Table 4 Example stakeholders and data sources

Category	High level examples	Detailed examples
Internal stakeholders	Engineers, Product managers, business stakeholders	Product engineers, architects, customer service representatives, sales representatives, managers, company executives
External stakeholders	Users	Premium customers, feemium customers, prospects, end users, partners, competitors, suppliers, lawmakers, regulators
Analytics	Product usage data	Telemetry data, user data, user behavior analysis
Reports	Market research	Market analysis, public surveys, trends, analysis of similar products
Environment	Domain knowledge	Domain experts, Technology standards, laws, regulations, industry conventions, opinion leaders

Table 5 Semantic meaning of the measurement scale

Score	Meaning
0	The data source does not favorably contribute to a criteria at all
1	The data source provide marginal favorable contribution
2	The data source provide somewhat favorable contribution to the criteria
3	The data source provide a moderately favorable contribution to the criteria
4	The data source provide a favorable contribution to the criteria
5	The data source most favorably contribute to a criteria

where vector

$$U_{i(j)} = \begin{bmatrix} u^i_{1j} \\ \vdots \\ u^i_{nj} \end{bmatrix}$$

represents the assessments of analyst e_i of the data sources w.r.t. criterion c_j. In that way k different matrices are obtained, i.e. one for each analyst e_i $(i = 1, 2, \dots, k)$. We initially normalize the analyst scores in each matrix U_i $(i = 1, 2, \dots, k)$ by calculating the relative importance of each data source w.r.t. each criterion, i.e. the value of each entry u^i_{lj} $(l = 1, 2, \dots, n$ and $j = 1, 2, \dots, m')$ of U_i is divided by the maximum value in the column the entry appears in. The overall scores of the data sources can also be calculated by averaging over all matrices $\{U_i\}_{i=1}^k$, i.e. a matrix U of overall weights can be obtained as follows:

$$U = \begin{bmatrix} u_{11} & \cdots & u_{1m'} \\ \vdots & & \vdots \\ u_{n1} & \cdots & u_{nm'} \end{bmatrix},$$

where $u_{lj} = 1/k \sum_{i=1}^k u^i_{lj}$ represents the overall assessment of data source d_l $(l = 1, 2, \dots n)$ w.r.t. criterion c_j $(j = 1, 2, \dots, m')$. Evidently, the vectors $U_{(j)}$ (columns of matrix U, $j = 1, 2, \dots, m'$) can be used to rank and compare the data sources separately for each criterion. Finally, the vector Y of overall data source ranks can be obtained by taking into account the overall criteria weights given in vector W, i.e.

$$Y = W \times U = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},$$

where $y_l = \sum_{j=1}^{m'} W_{lj} \cdot u_{lj}$ for each l $(l = 1, 2, \ldots, n)$. Notice that the resulting vector Y contains overall assessments denoting the importance of different data sources, respectively. These can be used to rank the data sources with respect to their relevance to the problem under decision.

In addition to the above, we can take into account the relative importance of different criteria defined by each analyst e_i (vector w'_i) by producing a weighted mean aggregation with the corresponding normalized matrix U_i, i.e. a vector y_i is generated for each analyst e_i $(i = 1, 2, \ldots, k)$ as follows:

$$y_i = w'_i \times U_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{in} \end{bmatrix},$$

where y_{ij} $(j = 1, 2, \ldots, n)$ is the overall relative importance of data source d_j w.r.t. analyst e_i. It is interesting to notice that in the above expression we can use W instead w'_i, i.e. $W \times U_i$. The calculated scores present the data sources' evaluation of analyst e_i who has taken into consideration the analysts' group criteria evaluation.

In this context we can calculate the vector Y of overall ranks (weights) of data sources w.r.t. the all involved analysts by calculating the average of the vectors y_i $(i = 1, 2, \ldots, k)$, i.e.

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},$$

where $y_j = 1/k \sum_{i=1}^{k} y_{ij}$ for each j $(j = 1, 2, \ldots, n)$.

4.5 Step 4: Interpret results

Interpretation of results consists of two steps: i) Identify and resolve discrepancies between analyst perspectives, ii) Rank and select data sources.

4.5.1 Identify and resolve discrepancies between perspectives

Breaking down the results and analyzing how different analysts have estimated criteria and data sources can help spotting discrepancies. Analysis of such discrepancies can be useful for better understanding of the problem under consideration, removing ambiguity and improving results of the method. We propose to visualize the vectors y_i $(i = 1, 2, \ldots, k)$ to identify discrepancies between the analysts (e.g., see Fig. 3).

The values in matrices U_i $(i = 1, 2, \ldots, k)$ show how different analysts have evaluated each data source according to each criterion. For example, vector $U_{i(j)}$ represents the assessments of analyst e_i of the data sources w.r.t. criterion c_j. The vectors produced by the different analysts for each single criterion can further be studied in order to understand the cause of differences in vectors y_i $(i = 1, 2, \ldots, k)$ containing the assessments of the data sources w.r.t. all the criteria (e.g., see Fig. 2, right side). In addition, vectors $U_{i(j)}$ $(i = 1, 2, \ldots, k)$ containing the individual evaluations of the analysts can be compared with the corresponding overall (group) assessments of the data sources given in vector $U_{(j)}$. This can provide an additional insight about the data sources and evaluation criteria.

We can also study in the same fashion the analysts' criteria evaluations represented by vectors w'_i $(i = 1, 2, \ldots, k)$ versus the overall weights given in W. In addition, we can compare the ranking of the data sources produced based on the analyst individual evaluations $(w'_i \times U_i)$ with the ranking he/she will generate by using the overall criteria weights $(W \times U_i)$. These can facilitate the analysis and better understanding of the decision scenario and further identification of the reasons of a discrepancy in the analysts' evaluations.

In this study, discrepancy is analysed by using fuzzy concepts like *negligible*, *moderate* and *severe*. For example, a discrepancy can be interpreted as *negligible* if the difference between two weights is below or equal to 0.3 (i.e. in the interval $[0, 0.3]$), *moderate* in case of $[0.3, 0.7]$, and *severe* when it is above 0.7 (i.e., in the range $[0.7, 1]$). These boundaries can vary for different contexts.

It is important to identify the cause of any discrepancies in order to improve and strengthen the results of the method. We identify the following causes for discrepancies:

1. *Mistakes.* A discrepancy can be caused by a wrongly entered number, miscalculation, or issue in a tool, etc. Such discrepancies can be easily removed by remedying the cause.
2. *Misunderstanding.* Different analysts may have different interpretation of certain criteria or data sources. Misunderstandings can be mitigated by initiating discussions and conducting of analysis of the analysts' estimates. As a result of these the analysts will arrive to shared understanding that will be followed by revising their estimates.
3. *Different perspective.* Discrepancies could be also caused by different and multimodal perspectives on the problem at hand. Genuine different perspectives are handled by the method.

Discrepancies caused by misunderstanding, misinterpretation and perspectives modalities can be addressed by breaking down higher level criteria and data sources into more specific terms. For example, a product could consist of many technologies, thus a criterion such as knowledge of product technology could be too broad to be used for ranking data sources. Breaking down such high level criteria into more specific can help arrive at more consistent evaluations.

4.5.2 Rank and select data sources

The weights in the vector Y indicate the overall relevance of each data source based on the evaluations of all the involved analysts. That is, higher ranking data sources are more relevant to the problem under consideration and should be consulted with higher priority, e.g., see Fig 2 and Fig 4. We suggest to interpret the results in parallel with analyzing discrepancies and exploring analysts' views on each data source and criterion. Thus, benefiting both from aggregated individual-based approach, and exchange of ideas and arguments of group-based approach [41].

We propose to interpret values in the vector Y using fuzzy concepts like *low*, *medium*, and *high*. For example, relevance of a data source can be interpreted as *low* if the value is below or equal to 0.3 (i.e. in the interval $[0, 0.3]$), *medium* in case of $(0.3, 0.7]$, and *high* when it is above 0.7 (i.e., in the range $(0.7, 1]$). These boundaries can be adjusted to suit different contexts.

5 Case studies

In this section, we describe the demonstration and application of the method in two industrial cases along with practitioners reflections on each component of the method, lessons learned from each step and the collected data.

5.1 Case I - Supply Chain Digitalization

Stockfiller AB offers a software-intensive service digitalizing the supply chain between sellers (farmers, food producers, and food importers) and buyers (restaurants and grocery stores) of food products in Sweden. The main value created by the service is more efficient and transparent supply chain that historically is based on personal contacts and manual processes. The monetization model of the service is based on the volume of goods traded in the platform. The service manages about 10 000 sellers and buyers. The company is in the market for 5 years. The main objective of the company is to secure the local market before expanding to nearby regions.

Two managers of the company, chief of operations and head of sales, participated in a workshop. The workshop follow the steps described in Section 3.5. We start by eliciting practitioners experience with selecting data sources and a specific example that can be used as an input to the method. We further guide the participants through the steps of our method. Each step of the workshop is detailed in further subsections.

Reflections on the challenge: The participants reflected that they used to elicit feature ideas from potential customers and implement them in the service. However, the quality of such ideas was low due to lack of understanding about the service and potentially sub-optimal processes at the customer side. Many new features suggested by some customers were rejected by others.

According to the participants, customers are often not capable to articulate their actual needs, thus demonstrating a prototype and eliciting feedback on that works much better than interviews. The company uses knowledgeable and experienced customers as a filter to vet new feature ideas, both internal and provided by other customers. The respondents also identified product vision (their view of how the supply chain should be organized), as a key tool to shape the product and gauge customer requests for new features.

Table 6 Before presenting our method, we asked participants from Case I to evaluate the relevance to the given problem of each data source. The participants were asked to assign a score 0-5, where denotes 0 the least and 5 the highest relevance, to each used data source. We compare these baseline results with the results from our method to evaluate its usefulness.

Data source	Scores		Normalized scores		Aggregated
	Analyst 1	Analyst 2	Analyst 1	Analyst 2	score
Key customers	4	5	0.800	0.675	0.738
Similar Products	3	3	0.600	0.405	0.503
Internal engineers	5	1	1.000	0.135	0.568
New Customers	0	4	0.000	0.540	0.270
Technologies	2	2	0.400	0.270	0.335
Technology trends	0	3	0.000	0.405	0.203
Business Needs	5	5	1.000	0.675	0.838
Engineering concerns	1	4	0.200	0.540	0.370
Laws & regulations	0	1	0.000	0.135	0.066
Industry standards	2	5	0.400	0.675	0.538
Product vision	5	5	1.000	0.675	0.838
Average customer	0	2	0.000	0.270	0.135

Triggers: The company discovered that a substantial amount of sellers have poor sales performance on the service. Lagging sellers present missed revenue, since the monetization model is based on the traded volume. Principals of the company had proposed to explore the matter.

Analysts: The principals of the company are the main product decision makers. Chief of operations, head of product, and head of sales regularly meet and discuss opportunities for new product features among other concerns. In the demonstration, only chief of operations and head of sales were present.

Problem statement: The principals set forth an objective to support sellers by providing insights on their sales performance. Initially, there were no insights for sellers to assess their performance and spot opportunities for improvement. As a consequence, the company was losing potential revenue. The ideal scenario would be that the service provide useful information for sellers to monitor their performance, and suggest adjustments to boost their trade volumes.

We used the example to discuss what data sources were useful in hindsight, i.e. to establish a baseline, see Table 6. The participants provided their estimates on a scale 0-5, where 0 denotes the lowest, and 5 the highest relevance. We normalize their scores to arrive at scale of $[0, 1]$. The aggregated score is an arithmetic mean of the standardized scores.

Criteria: The participants immediately identified good relationship and knowledge as the primary criteria to select stakeholders for collaboration. The participants also identified knowledge and experience of the product, accessibility, value per purchase, life time value, mutual trust, openness, capacity and eagerness to contribute, e.g by piloting experimental features, as relevant criteria. There was no need to shortlist the criteria because of only two participants.

In Table 7 we summarize the criteria and participants (Analyst 1 and Analyst 2) estimates on their importance, standardized scores, and final aggregated score. The participant scores are normalized to arrive to scale $[0, 1]$, the aggregated score is the arithmetic mean of normalized scored.

Table 7 Participant estimates on criteria importance from Case I. The participants were asked to evaluate each criterion on a scale 0-5,where 0 denotes lowest and 5 the highest importance to the given problem. We also show normalized scores and the final aggregated score.

Criterion	Scores		Normalized scores		Aggregated
	Analyst 1	Analyst 2	Analyst 1	Analyst 2	score
Knowledge of the product	3	4	0.600	0.661	0.630
Accessibility	2	5	0.400	0.826	0.613
Trust	5	3	1.000	0.496	0.748
Price per purchase	3	4	0.600	0.661	0.630
Life-time value	4	4	0.800	0.661	0.730
Capacity to contribute	2	3	0.400	0.496	0.448

Further discussions revealed that some criteria could be broken down to be more specific, and suit both stakeholders and data sources. For example, accessibility could be interpreted as the ease of access (for meetings, observations, or reading), and understandability (of the actual needs and descriptions).

Data sources: Participants acknowledged using multiple inputs in their decisions. They listed experienced and knowledgeable customers (key customers), similar products, internal engineers, new customers, technologies (used design templates, patterns, frameworks), technology trends (such as new available technologies), business needs, engineering concerns, laws, and industry standards as relevant data sources. In Fig. 2 (left side) we show the original baseline estimates on data source relevance by both analysts.

Application of the method: The participants had no difficulty understanding the rationale behind assigning the scores to criteria and evaluate the data sources. However, there were some discussions on the exact definitions of data sources and criteria. For example, the product vision could have no lifetime value at all, if it is considered as a document. However, the vision could have the highest possible life-time value as realizing the vision is the purpose of the company.

Interpretation of the results:

We summarize the results in Table 8. Rows in the table are ordered by initial baseline ranking of data sources. We also show the ranking after applying the method and the difference between baseline and final ranking. A positive difference denotes that a data sources has gained relevance compared to the baseline, a negative difference denotes a reduced relevance.

Relevance column shows final and aggregated relevance scores and their interpretation. In the last column, we quantify the disagreement between analysts views.

We observe several tendencies when looking at the results, see also Fig. 2. First, results from the method shows mostly negligible differences between analysts views on how relevant each source are. A moderate disagreement concern the relevance of new customers as a source of data. Examining this discrepancy, we found that the root cause is a disagreement on how much value per purchase new customers deliver (see Fig. 3). Analyst 1 wished to emphasize that new customers directly contributes to the bottom line of the company. However, the other analyst expressed a view that new customers contribute relatively little compared to established customers.

Table 8 Aggregated results and their interpretation from Case I. The rows are ordered by the baseline estimates. We also show the resulting ranking from the method and the difference in ranking.

Data source	Ranking			Relevance		Disagreement	
	Baseline	Method	Difference	Score	Interpretation	Score	Interpretation
Product vision	1	4	-3	0.400	medium	0.200	negligible
Business Needs	2	5	-3	0.396	medium	0.210	negligible
Key customers	3	3	0	0.475	medium	0.031	negligible
Internal engineers	4	6	-2	0.266	low	0.146	negligible
Industry standards	5	9	-2	0.180	low	0.028	negligible
Similar Products	6	11	-5	0.118	low	0.100	negligible
Engineering concerns	7	7	0	0.243	low	0.000	negligible
Technologies	8	12	-4	0.106	low	0.138	negligible
New Customers	9	1	+8	0.717	high	0.567	moderate
Technology trends	10	8	+2	0.181	low	0.096	negligible
Average customer	11	2	+9	0.482	medium	0.000	negligible
Laws & regulations	12	10	+2	0.121	low	0.026	negligible

Secondly, comparing the original estimates and results from the method, see Fig. 2, we observe that the method arrives substantially more consistent results between both analysts. Such results demonstrate that the method is useful for consensus building among the analysts. In addition, it enables to conduct a multi-layer analysis of the analysts' assessments and in that way it further facilitates the interpretation and better understanding of the selected criteria, the used data sources and the connections between them.

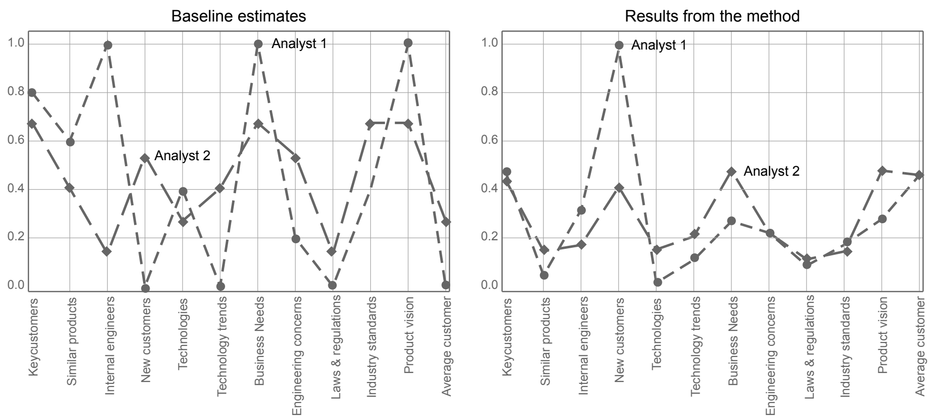

Fig. 2 Results from the Case I. Figure on the left show baseline estimates, i.e. respondent estimates without the method. Figure on the right show results from applying the method. Y-axis denotes the relative importance of data sources.

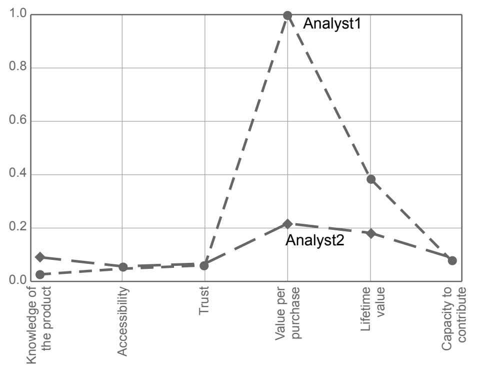

Fig. 3 Discrepancies in analyst estimates evaluating the relevance of new customers from Case I. Y-axis denote normalized estimates on how much data from new customers contribute to each criterion. Observe the disagreement between analysts on how much value per purchase new customers deliver.

Lessons learned: Feedback from practitioners added support to our hypothesis on the need for a structured method for stakeholder identification. The participants were already familiar with all the concepts and their relationships, thus we did not identify any shortcomings in the construct validity.

The discussions revealed that the main area for improvement could be more refined guidelines for formulating criteria that are applicable to both human stakeholders and other types of data sources.

Application of the method helped participants to reflect on their views on what are the most relevant sources of input and why. Specifically, how use of certain data sources connect to short and long term revenue.

5.2 Case II - Construction Equipment

The second case is a company developing and manufacturing construction and mining machinery. The machines are software-intensive and most new features concern updates in the software. The company uses their global partner network to sell and service their machines, and employs over 13 000 people.

Two managers participated in the workshop. The managers are involved and oversee research and development of new product features. We follow the steps described in Section 3.5, and started by presenting the method and then collected the data using a spreadsheet. We did follow-up interviews with each participant to discuss and interpret the results from the method.

Reflections on the challenge: The participants noted that the organization recognises the challenge and had developed an internal framework for gathering information concerning future product development. However, the application of the framework varies per each use.

Example case: The company introduced connectivity to construction machines. The connectivity provides telemetry on machine performance, error reports, maintenance needs, and operator behavior among other data.

Triggers: The opportunity to benefit from the recent IoT development is considered as a trigger.

Analysts: The company has a dedicated product planning group for product development. In addition, they involve key accounts and dealers to product decisions.

Problem statement: A substantial part of company's offering is maintenance of the machines to ensure efficient operation and to reduce downtime. However, there is limited feedback to product development why a machine breaks down and what usage patterns led to it. IoT technology delivers real-time data on how machines are operated, thus enabling the company to actively prevent breakdowns and proactively improve the machines. This reduces maintenance costs and improves availability, leading to overall improvement of service quality and reduced operational costs.

Table 9 We asked participants from Case II to evaluate the relevance to the given problem of each data source, before presenting our method. The participants were asked to assign a score 0-5, where denotes 0 the least and 5 the highest relevance, to each used data source. We compare these baseline results with the results from our method to evaluate its usefulness.

Data source	Scores		Normalized scores		Aggregated
	Analyst 1	Analyst 2	Analyst 1	Analyst 2	score
Key accounts (customers)	5	5	1.000	0.769	0.885
Customer clinics	2	1	0.400	0.154	0.277
Laws & regulations	4	5	0.800	0.769	0.785
Dealers	3	3	0.600	0.462	0.531
Operators (end users)	1	2	0.200	0.308	0.254
Product planning group	1	5	0.200	0.769	0.485
Regional partners	4	5	0.800	0.769	0.785

We asked the participants (Analyst 1 and Analyst 2), with hindsight, to list and rank all utilized data sources. Their estimates are shown in Table 9. The participants provided their estimates on a scale 0-5, where 0 denotes the lowest, and 5 the highest relevance. We normalize their scores to arrive at scale of $[0, 1]$. The aggregated score is an arithmetic mean of the standardized scores. The estimates are provided on a scale 0 - 5, where 0 denotes the lowest, and 5 the highest contribution to the final solution.

Criteria: The participants needed guidance for selecting criteria, thus they adopted the high level criteria from Table 2. We show the exact criteria and participants estimates on their relevance to the problem, normalized scores, and final aggregated score in Table 10. The participant scores are normalized to arrive to scale $[0, 1]$, the aggregated score is the arithmetic mean of standardized scores.

Data sources: Participants reflected that the key source for ideas are dealers and key accounts. Dealers and regional partners work closely with end-customers and have the first hand knowledge of their actual needs. Furthermore, key accounts and dealers provide substantial revenue for the company, thus there is an financial incentive to prioritize their input. Other sources of input are industry standards, regulations, engineering, direct feedback from customers, workshops with machine operators, and engineers. We show the initial baseline ranking of data sources in Fig. 2, left side.

Table 10 Participant estimates on criteria importance from Case II. The participatns were asked to evaluate each criterion on a scale 0-5,where 0 denotes lowest and 5 the highest importance to the given problem. We also show normalized scores and the final aggregated score.

Criterion	Scores		Normalized scores		Aggregated
	Analyst 1	Analyst 2	Analyst 1	Analyst 2	score
Knowledge	4	5	0.752	1.00	0.876
Amount of experience	4	4	0.752	0.800	0.776
Revenue	5	4	0.939	0.800	0.870
Cost	5	4	0.939	0.800	0.870
Opportunity cost	4	2	0.752	0.400	0.576
Risk	3	5	0.564	1.00	0.782
Timing	3	3	0.564	0.600	0.582
Ease of use	2	3	0.376	0.600	0.488
Suitability for collaboration	3	1	0.564	0.200	0.382

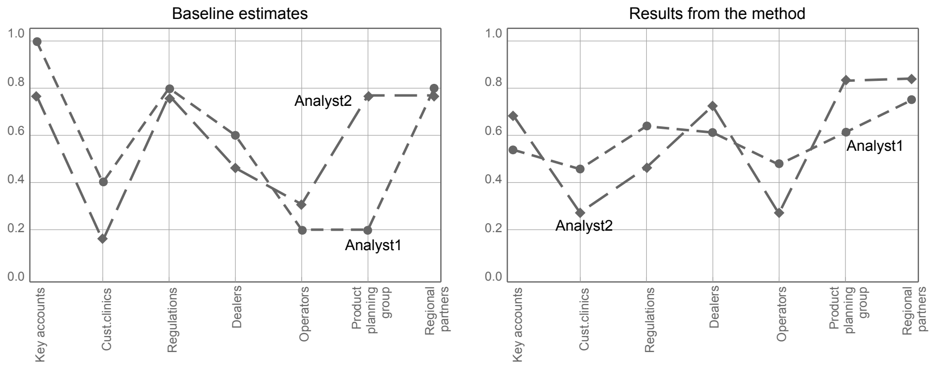

Fig. 4 Results from the Case II. Figure on the left show baseline estimates, i.e. respondent estimates without the method. Figure on the right show results from applying the method. Y-axis denotes the relative importance of data sources.

Application of the method: Method application run smooth and with only minor confusion. The first issue was to identify a problem to be studied, but after initial discussion the participants quickly agreed. The effort needed to get them acquainted was moderate and most time was dedicated towards individual estimated of the importance of each data source against the criteria (about 45 minutes on average for participant). Some additional effort was needed during the prioritization to explain the meaning of particular criteria for a particular data source.

Interpretation of the results:

We summarize the results in Table 11. Rows in the table are ordered by initial baseline ranking of data sources. We also show the ranking after applying the method and the difference between baseline and final ranking. A positive difference denotes that a data sources has gained relevance compared to the baseline, a negative difference denotes a reduced relevance.

The baseline estimates (Fig. 4, left side) show a discrepancy between the analysts regarding importance of the product planning group. However, application of the method (Fig. 4, right side) helps to arrive at more aligned results. The key

Table 11 Aggregated results and their interpretation from Case II. The rows are ordered by the baseline estimates. We also show the resulting ranking from the method and the difference in ranking.

Data source	Ranking			Relevance		Disagreement	
	Baseline	Method	Difference	Score	Interpretation	Score	Interpretation
Key accounts	1	4	-3	0.769	high	0.144	negligible
Regional partners	2	1	+1	0.954	high	0.092	negligible
Regulations	3	5	-2	0.710	high	0.175	negligible
Dealers	4	3	+1	0.828	high	0.113	negligible
Product planning group	5	2	+3	0.881	high	0.219	negligible
Customer clinics	6	7	-1	0.524	medium	0.185	negligible
Operators	7	6	+1	0.532	medium	0.207	negligible

accounts were estimated (without the method) to be the primary data source. However, the method suggests that input from regional partners, dealers, and product planning group should be prioritized.

Lessons learned: Participants reflections during the demonstration supported our earlier hypothesis that product decisions are often opinion based. Application of the method builds consensus and helps to reduce discrepancies.

We observed a need to accommodate a scenario where distributed analysts work independently and the method combines their input. This can be achieved by more stringent guidelines on how to select criteria and data sources for evaluation, and tool support to facilitate the method. Our participants saw great value in eliciting the criteria and consensus building among not only them, but also a larger group of stakeholders involved in decision making. In their opinion, that would greatly improve the selection of data sources.

6 Potential improvements and limitations

One of the central areas for improvements is providing better support for identifying relevant criteria and candidate data sources. Using overly broad criteria and data sources reduces usefulness of the results. However, identification and evaluations of many very specific terms may require excessive effort. To alleviate this shortcoming, we propose to iteratively apply the method by increasing the level of detail with each iteration. We suggest to start by high level data sources and criteria, e.g., ones given in Tables 2 and 4. Then gradually eliminate less relevant criteria and detail the relevant ones, to arrive at specific data instances (e.g., Customer 1 and Customer 2) from high level classes (e.g., key customers).

We use ordinal scale from 0 to 5 in order to facilitate analysts in the task of ranking the criteria [37]. Alternative and more sophisticated ranking methods might also be used, for example, ones based on pairwise comparison [14] and large preference relation [20]. However, evaluations in ordinal scale are more intuitive, faster, and require relatively fewer steps than the other two methods.

Another limitation is the use of arithmetic mean, which is well-known to be sensitive to extreme values. In general, the choice of aggregation operator is critical since some aggregation operators can lead to a significant loss of information since their values can be greatly influenced by extreme scores (e.g., the arithmetic mean), while others are penalizing too much for low-scoring outliers (e.g., the geometric mean and the harmonic mean) [16]. A possible solution to the described problem is to use different aggregation operators in order to find some trade-off between their conflicting behaviour, e.g., hybrid aggregation operators proposed in Tsiporkova et al. [60, 61].

A potential scalability issue emerges when inputs from many analysts, a large number of criteria or data sources need to be considered. A large number of criteria (C) and data sources (D) require each analyst to provide $C+D\times C$ inputs. However, this limitation remains theoretical until full-scale validation of the method.

7 Discussion

We structure our discussion around the research questions. We discuss the method with our initial objectives stated in Section 3.3 and lessons learned from the workshops.

7.1 RQ1: What are the needs towards a method supporting selection of data sources for MDRE?

We identify and formulate specific needs towards the new method in Sections 2 and 3.3. In this section, we revisit the needs and discuss to what extent our method addresses each specific requirement.

The review of related work in Section 2.2 reveals that most stakeholder selection practices consider requirements engineering as a fixed activity at the beginning of a project. Stakeholders are selected based on generalized characteristics, such as importance and influence on the project [9, 10, 11]. Such an approach ignores the continuous nature of MDRE.

Discussions in the workshops confirmed that requirements engineering in a market-driven context is a continuous series of micro-decisions. The decisions pertaining to how to respond to market fluctuations, customer requests, new business objectives, to name a few. Different kinds of data may be needed to support each decision. Thus, the selection of data sources needs to be specific to each situation (*Objective 1*).

Our proposed method is flexible and context-independent. We specifically avoid coupling the method with any specific context, criteria, or data sources. Instead, we provide instructions on how to identify inputs for the method from its context of use. Thus, the method can be tailored for use in a wide range of situations. However, we exemplify some criteria and data sources to describe the method.

In a crowd requirements engineering context, analysts may benefit from a wide range of data sources, both within companies' control and beyond [23, 2]. Identification of new and unrealized sources is a crucial step in selecting an optimal set of data sources (*Objective 2*).

Our method addresses this objective by detailing the context of use and identifying data sources through a series of steps, namely describing the problem and setting forth the evaluation criteria of the sources. Such steps help to understand the context of use, thus supporting the identification of unrealized data sources.

An alternative approach could be to develop a taxonomy of data sources for crowd requirements engineering. However, at this stage of our research, we specifically avoid prescribing any specific sources to mitigate threats to construct validity.

Literature characterizes stakeholders in MDRE and crowd contexts with limited motivation to participate in product engineering and unable to articulate their actual needs. Thus, directly consulting specific stakeholder groups may not be worth the investment. In turn, stakeholder needs could be elicited indirectly, for example, by examining documents, artifacts, and data generated by stakeholders interacting with the product. The method should be able to handle a broad range of data sources to support the selection of data sources in such a context (*Objective 3*).

There are no inherent limitations on what types of data sources the method can handle. The difficulty could be to formulate evaluation criteria to suit both individuals and inanimate data sources. Formulation of criteria and data sources can be done iteratively. The initial set of criteria can help to understand the needs of data sources, thus supporting the identification of new data sources. Knowledge of what data sources need to be evaluated can help to arrive at more generalized criteria.

Requirements engineering is a collaborative activity, and product decisions could be cross-cutting and require consideration of multiple perspectives [21, 13]. Aggregating multiple perspectives and especially reaching consensus between opposing views is challenging. The method should support collaboration between multiple analysts (*Objective 4*).

The method has built-in support for collaboration and consensus development among multiple analysts. We use the arithmetic mean to combine views from multiple analysts. To explore different viewpoints, we propose to estimate the discrepancies and to use plots visualizing the differences.

A review of discrepancies and visualizing immediate results with plots contribute to the transparency and understandability of the results produced by the method (*Objective 5*).

An important concern is to what extent the benefits of using a method outweigh the resources needed to use it. Using our method requires several analysts to meet, set forth relevant criteria and candidate data sources, run the method, and discuss the results. We argue that such meetings already take place, and the use of the method adds structure to the meetings.

In both workshops, practitioners mentioned a committee that regularly meets and steers product decisions. Thus, using the method to support such meetings would require minimal additional effort. Not utilizing any structured approach could lead to unproductive discussions and suboptimal decisions. At a minimum, a suboptimal decision requires an immediate extra effort to correct it. However, it could also lead to wasted development effort and missed market opportunities. Therefore, potential benefits from using the method substantially outweigh the additional effort.

7.2 RQ2: How to support selection of data sources for MDRE?

Related work on decision-making scenarios identifies two main approaches. Group-based approaches use discussions and negotiations to arrive at the final decisions. However, such unstructured discussions could be unproductive, and the outcome could be skewed towards the views of more powerful group members. Individual-based approaches aggregate views from individuals regardless of their power. However, such an approach substantially limits the exchange of ideas and arguments in the group [53, 58].

Our method proposes to collect individual preferences in multiple steps and at each step, analyze the discrepancies between the preferences. In this way, we mitigate the adverse effects of group-based approaches and encourage focused discussions on specific disagreements.

7.3 RQ3: What improvements are needed to use the method in industry?

To identify opportunities for improvements, we summarize participant feedback and our lessons learned from the workshops. In both workshops, we identified the need for additional guidance for identifying and interpreting the evaluation criteria. Even though we exemplify some criteria in the method description, the selection of criteria was a difficulty. The support for criteria selection could be provided with, for example, a taxonomy of criteria and their descriptions.

In both workshops, the practitioners were able to quickly list relevant data sources based on their experience and our examples. However, the identification of candidate data sources could be supported further by a more extensive taxonomy.

For the workshops, we implemented the method using spreadsheets and a simple script. During the application of the method, participants were guided and supported by researchers. However, we identify the need for a robust tool guiding analysts through the steps of the method and minimizing the need for guidance from researchers.

8 Conclusions and further work

In this paper, we have proposed and demonstrated a group-based decision support method for the selection of data sources in MDRE. The method comprises of systematic steps to collaboratively identify candidate data sources, evaluate them according to agreed criteria, and to produce a ranking of data sources by their relevance to the decision. We have paid particular attention to consensus building and provided means to analyze and understand the final results.

We have demonstrated the method on two case studies where we evaluated its usability and gather data for further refinement of the method. We learned that the method helps to build consensus between analysts and to arrive at a more consistent view of the importance of individual data sources. The systematic steps of the method enable analysts to analyze and remedy their disagreements.

Based on the results from demonstrations, we believe that our method can improve the selection of the data sources for MDRE by highlighting and helping to resolve discrepancies in the analyst's views. Thus, improving data quality for following requirements engineering activities and improving the overall quality of the final product.

Further work focuses on additional extensive validations and improvement rounds. We plan to: a) develop a tool to facilitate the use of the method and data collection on its application, and b) extend our support towards criteria and data sources identification.

Acknowledgements The authors would like to thank participants of the demonstrations for their time dedicated to this study.

We would like to acknowledge that this work was supported by the KKS foundation through the S.E.R.T. Research Profile project at Blekinge Institute of Technology.

References

1. Agarwal N, Rathod U (2006) Defining 'success' for software projects: An exploratory revelation. International journal of project management 24(4):358–370
2. Alexander IF (2005) A taxonomy of stakeholders: Human roles in system development. International Journal of Technology and Human Interaction (IJTHI) 1(1):23–59
3. Alspaugh TA, Scacchi W (2013) Ongoing software development without classical requirements. In: Requirements Engineering Conference (RE), 2013 21st IEEE International, IEEE, pp 165–174
4. Alves C, Pereira S, Castro J (2006) A study in market-driven requirements engineering. Workshop em Engenharia de Requisitos WER06 pp 2–3
5. Annamalai N, Kamaruddin S, Abdul Azid I, Yeoh T (2013) Importance of problem statement in solving industry problems. In: Applied Mechanics and Materials, Trans Tech Publ, vol 421, pp 857–863
6. Anwar F, Razali R (2016) Stakeholders selection model for software requirements elicitation. American Journal of Applied Sciences 13(6):726–738
7. Aurum A, Wohlin C (2003) The fundamental nature of requirements engineering activities as a decision-making process. Information and Software Technology 45(14):945–954
8. Babar MI, Ghazali M, Jawawi DN (2014) A bi-metric and fuzzy c-means based intelligent stakeholder quantification system for value-based software. In: SoMeT, pp 295–309
9. Babar MI, Ghazali M, Jawawi DN, Elsafi A (2014) Stakeholder management in value-based software development: systematic review. IET Software 8(5):219–231
10. Babar MI, Ghazali M, Jawawi DN, Zaheer KB (2015) Stakemeter: Value-based stakeholder identification and quantification framework for value-based software systems. PloS one 10(3)
11. Ballejos LC, Montagna JM (2011) Modeling stakeholders for information systems design processes. Requirements engineering 16(4):281–296
12. Bano M, Zowghi D, Ikram N (2014) Systematic reviews in requirements engineering: A tertiary study. In: 2014 IEEE 4th International Workshop on Empirical Requirements Engineering (EmpiRE), IEEE, pp 9–16

13. Barczak G, Lassk F, Mulki J (2010) Antecedents of team creativity: An examination of team emotional intelligence, team trust and collaborative culture. Creativity and Innovation Management 19(4):332–345
14. Barzilai J (1997) Deriving weights from pairwise comparison matrices. Journal of the operational research society 48(12):1226–1232
15. Bendjenna H, Charre PJ, Eddine Zarour N (2012) Using multi-criteria analysis to prioritize stakeholders. Journal of Systems and Information Technology 14(3):264–280
16. Bullen PS, Mitrinovic DS, Vasic M (2013) Means and their Inequalities, vol 31. Springer Science & Business Media
17. Burnay C (2016) Are stakeholders the only source of information for requirements engineers? toward a taxonomy of elicitation information sources. ACM Transactions on Management Information Systems (TMIS) 7(3):8
18. Dahlstedt ÅG, Karlsson L, Persson A, Natt och Dag J, Regnell B (2003) Market-Driven Requirements Engineering Processes for Software Products - a Report on Current Practices. In: International Workshop on COTS and Product Software, RECOTS 2003
19. Davis CJ, Fuller RM, Tremblay MC, Berndt DJ (2006) Communication challenges in requirements elicitation and the use of the repertory grid technique. Journal of Computer Information Systems 46(5):78–86
20. Fodor J, Orlovski S, Perny P, Roubens M (1998) The Use of Fuzzy Preference Models in Multiple Criteria Choice, Ranking and Sorting, Springer US, Boston, MA, pp 69–101. DOI 10.1007/978-1-4615-5645-9_3
21. Gorschek T, Wohlin C (2006) Requirements abstraction model. Requirements Engineering 11(1):79–101
22. Gorschek T, Wohlin C, Garre P, Larrson S (2006) A model for technology transfer in practice. Learning and Leading with Technology 30(4):88–95
23. Groen EC, Seyff N, Ali R, Dalpiaz F, Doerr J, Guzman E, Hosseini M, Marco J, Oriol M, Perini A, et al. (2017) The crowd in requirements engineering: The landscape and challenges. IEEE software 34(2):44–52
24. Gu Q, Parkin M, Lago P (2011) A taxonomy of service engineering stakeholder types. In: European Conference on a Service-Based Internet, Springer, pp 206–219
25. Hamka F, Bouwman H, De Reuver M, Kroesen M (2014) Mobile customer segmentation based on smartphone measurement. Telematics and Informatics 31(2):220–227
26. Hofmann HF, Lehner F (2001) Requirements engineering as a success factor in software projects. IEEE software (4):58–66
27. Hujainah F, Baka RBA, Al-Haimi B, Abdulgabber MA (2018) Stakeholder quantification and prioritisation research: A systematic literature review. Information and Software Technology
28. IEEE Computer Society (2014) Guide to the Software Engineering Body of Knowledge (SWEBOK(R)): Version 3.0, 3rd edn. IEEE Computer Society Press, Los Alamitos, CA, USA
29. Jin J, Liu Y, Ji P, Liu H (2016) Understanding big consumer opinion data for market-driven product design. International Journal of Production Research 54(10):3019–3041
30. Karlsson L, Dahlstedt ÅG, Regnell B, och Dag JN, Persson A (2007) Requirements engineering challenges in market-driven software development–an inter-

view study with practitioners. Information and Software technology 49(6):588–604
31. Kittlaus HB, Fricker SA (2015) Software Product Management. Springer
32. Klotins E, Unterkalmsteiner M, Gorschek T (2017) Software Engineering Anti-patterns in start-ups. In review by IEEE Software
33. Klotins E, Unterkalmsteiner M, Chatzipetrou P, Gorschek T, Prikladniki R, Tripathi N, Pompermaier L (2019) A progression model of software engineering goals, challenges, and practices in start-ups. IEEE Transactions on Software Engineering
34. Klotins E, Unterkalmsteiner M, Gorschek T (2019) Software engineering in start-up companies: An analysis of 88 experience reports. Empirical Software Engineering 24(1):68–102
35. Kruchten P (2000) The Rational Unified Process: An Introduction, Second Edition, 2nd edn. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA
36. La Rocca A, Moscatelli P, Perna A, Snehota I (2016) Customer involvement in new product development in b2b: The role of sales. Industrial Marketing Management 58:45–57
37. Labovitz S (1970) The assignment of numbers to rank order categories. American sociological review pp 515–524
38. Lauesen S (2001) Software Requirements: Styles and Techniques, 1st edn. Pearson Education
39. Lim SL, Quercia D, Finkelstein A (2010) Stakenet: using social networks to analyse the stakeholders of large-scale software projects. In: Proceedings of the 32Nd ACM/IEEE International Conference on Software Engineering-Volume 1, ACM, pp 295–304
40. Maalej W, Nayebi M, Johann T, Ruhe G (2016) Toward data-driven requirements engineering. IEEE Software 33(1):48–54
41. Matsatsinis NF, Samaras AP (2001) Mcda and preference disaggregation in group decision support systems. European Journal of Operational Research 130(2):414–429
42. McManus J (2004) A stakeholder perspective within software engineering projects. In: 2004 IEEE International Engineering Management Conference (IEEE Cat. No. 04CH37574), IEEE, vol 2, pp 880–884
43. Mitchell RK, Agle BR, Wood DJ (1997) Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. The Academy of Management Review 22(4):853–886
44. Pacheco C, Garcia I (2012) A systematic literature review of stakeholder identification methods in requirements elicitation. Journal of Systems and Software 85(9):2171–2181
45. Paternoster N, Giardino C, Unterkalmsteiner M, Gorschek T, Abrahamsson P (2014) Software development in startup companies: A systematic mapping study. Information and Software Technology 56(10):1200–1218
46. Petersen K, Wohlin C (2009) Context in industiral software engineering research. ... Symposium on Empirical Software Engineering ...
47. Preiss O, Wegmann A (2001) Stakeholder discovery and classification based on systems science principles. In: Proceedings of the Second Asia-Pacific Conference on Quality Software, IEEE Computer Society, Washington, DC, USA, APAQS '01, pp 194–

48. Pressman RS (2001) Software Engineering: A Practitioner's Approach, 5th edn. McGraw-Hill Higher Education
49. Razali R, Anwar F (2011) Selecting the right stakeholders for requirements elicitation: a systematic approach. Journal of Theoretical and Applied Information Technology 33(2):250–257
50. Regnell B, Brinkkemper S (2005) Market-Driven Requirements Engineering for Software Products, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 287–308. DOI 10.1007/3-540-28244-0_13, URL https://doi.org/10.1007/3-540-28244-0_13
51. Riegel N, Doerr J (2015) A systematic literature review of requirements prioritization criteria. In: International Working Conference on Requirements Engineering: Foundation for Software Quality, Springer, pp 300–317
52. Runeson P, Höst M, Rainer A, Regnell B (2012) Case study research in software engineering. John Wiley & Sons, Inc.
53. Saaty TL (1988) What is the analytic hierarchy process? In: Mathematical models for decision support, Springer, pp 109–121
54. Sharp H, Finkelsteiin A, Galal G (1999) Stakeholder identification in the requirements engineering process. In: 10th International Workshop on Database & Expert Systems Applications, IEEE Computer Society, Washington, DC, USA, DEXA '99, pp 387–
55. Singh R (1996) International standard iso/iec 12207 software life cycle processes. Software Process: Improvement and Practice 2(1):35–50
56. Software Engineering Institute (2006) CMMI for Development, Version 1.2. Tech. rep., Software Engineering Institute, Carnegie Mellon University
57. Sommerville I (2010) Software Engineering, 9th edn. Addison-Wesley Publishing Company, USA
58. Tindale RS, Kameda T (2000) 'social sharedness' as a unifying theme for information processing in groups. Group Processes & Intergroup Relations 3(2):123–140
59. Tovar E, Pacheco C (2006) Stakeholder identification in requirements engineering: Comparison of methods. In: Proc. of Software Engineering Applications (SEA)
60. Tsiporkova E, Boeva V (2004) Nonparametric recursive aggregation process. Kybernetika, Journal of the Czech Society for Cybernetics and Information Sciences 40:51–70
61. Tsiporkova E, Boeva V (2006) Multi-step ranking of alternatives in a multi-criteria and multi-expert decision making environment. Information Sciences 176:2673–2697
62. Vigna S (2015) A weighted correlation index for rankings with ties. In: 24th international conference on World Wide Web, pp 1166–1176
63. Whitehead J (2007) Collaboration in software engineering: A roadmap. In: Future of Software Engineering (FOSE'07), IEEE, pp 214–225
64. Wieringa RJ (2014) Design science methodology for information systems and software engineering. Springer
65. Wnuk K (2017) Involving relevant stakeholders into the decision process about software components. Proceedings - 2017 IEEE International Conference on Software Architecture Workshops, ICSAW 2017: Side Track Proceedings pp 129–132, DOI 10.1109/ICSAW.2017.68

66. Xiaoqing Liu, Veera CS, Sun Y, Noguchi K, Kyoya Y (2004) Priority assessment of software requirements from multiple perspectives. In: Proceedings of the 28th Annual International Computer Software and Applications Conference, 2004. COMPSAC 2004., pp 410–415 vol.1, DOI 10.1109/CMPSAC.2004.1342872