Toward a Metric Catalog for Large-Scale Agile Development

Toward a Metric Catalog for Large-Scale Agile Development

Completed Research

Pascal Philipp
Technical University of Munich
pascal.philipp@tum.de
Leon Menzel
Technical University of Munich
leon.menzel@tum.de

Franziska Tobisch
Technical University of Munich
franziska.tobisch@tum.de
Florian Matthes
Technical University of Munich
matthes@tum.de

Abstract

Nowadays, organizations use agile software development to remain competitive in their frequently changing business environment. Inspired by the success of agile methods on a small scale, organizations have started to apply them in larger contexts. However, the limited scalability of agile methods is a problem. Metrics can be a success factor for achieving agility at scale, thus adopting them is promising. Most scaling agile frameworks provide few recommendations regarding metrics. Likewise, research on metrics in large-scale agile development lacks concrete guidance for metrics or their organization-specific adoption. To fill this gap, we propose two artifacts. We present the design of a minimalistic metric management fact sheet (MMFS) for large-scale agile development to support practitioners in using metrics in their organization-specific development environment. Furthermore, the MMFS is the basis for our metric catalog documenting 196 metrics identified in an expert study to provide a comprehensive metric set for scaling agile environments.

Keywords

Agile software development, large-scale agile development, metrics.

Introduction

In today's highly competitive market environment, software companies must instill high flexibility and ensure short lead times in order to respond to volatile requirements (Petersen and Wohlin 2010). As a result, an increasing number of software companies have begun to use incremental and agile development methods (Petersen and Wohlin 2010). Agile methodologies are currently widely adopted (Maiden and Jones 2010). The limited scalability of agile methods was identified as a limitation (Cohen et al. 2004) in a comprehensive systematic review of empirical studies of agile software development (Dybå and Dingsøyr 2008). A problem contributing to this scalability issue lies in agile methods' original design adapted to contexts similar to the "agile sweet spot" (Kruchten 2013). Frequent deliveries, medium to low system criticality, a stable underlying architecture, and co-located teams of 5–12 members are all part of this sweet spot (Nord et al. 2014). Adopting agile methods in larger settings leads to challenges, such as coordinating multiple teams and stakeholders, building trust (Brown et al. 2013; Dikert et al. 2016; Uludağ et al. 2018), and communicating with concerned parties (Dikert et al. 2016; Uludağ et al. 2018). Although using agile methods in larger projects and organizations is more difficult (Dybå and Dingsøyr 2009), the benefits of using them on a small scale convinced organizations to scale them up (Dikert et al. 2016). Several large-scale agile practices and frameworks, such as the Scaled Agile Framework (SAFe) (Leffingwell 2018) and Large Enterprise Scrum (Larman and Vodde 2016), have been developed to help organizations apply agile methods in larger settings (Paasivaara 2017).

Metrics have played a role in software development for decades (Fenton and Neil 2000). Metric adoption appears to have potential for large-scale agile development as well. Introducing measurement in agile software development, according to a variety of scaling agile frameworks (Ambler and Lines 2012; Beedle 2018; Scaled Agile Inc. 2021; Thompson 2013) and research (Brown et al. 2013), is a success factor for achieving agility at scale. However, most scaling agile frameworks offer limited guidance related to metrics. Particularly, SAFe (Scaled Agile Inc. 2021) covers metrics to a greater extent and among different organizational levels. Similarly, much of the research on metrics in large-scale agile development lacks concrete recommendations for metrics or does not provide enough guidance and information about an organization-specific adoption. Against this backdrop and inspired by Monahov (2014), we propose the design of a minimalistic metric management fact sheet (MMFS) for the domain of large-scale agile development. It serves as the basis for documenting 196 metrics, which we identified in an expert study, in our metric catalog. Our study has two objectives. First, by proposing a minimalistic MMFS, we hope to provide practitioners with guidance on how to adopt metrics in a large-scale agile development environment that is unique to their organization. Second, we want to provide a comprehensive set of metrics for scaling agile development environments. To achieve our research goals, we deduce the following research questions that guide the solution development:

RQ 1: How can a generic and minimalistic MMFS for the domain of large-scale agile software development be designed?

RQ 2: How can a structured collection of metrics help organizations execute large-scale agile software projects?

Related Work

Large-scale agile software development can refer to settings where several teams work together on a software product (Dingsøyr and Moe 2014; Dumitriu et al. 2019) or to organizations where agile methods and principles are implemented across a whole organization, including higher organizational levels (Dingsøyr and Moe 2014; Dumitriu et al. 2019). Subsequently, we present related work that presents structured collections of metrics, approaches to support metric adoption, and generic metrics documentation structures for large-scale agile development and related disciplines. Several authors investigated structured collections of metrics. Olszewska et al. (2016) propose a metrics model to quantitatively measure the impact of an agile transformation in a large software development organization. This metric model consists of a description structure for metrics which is based on the framework for evaluating metrics by Kaner et al. (2004) and the property-based software engineering measurement approach suggested by Briand et al. (1996). Bouwers et al. (2014) propose a catalog format to provide a consolidated overview of available software metrics. The format is designed to give a quick overview of each metric's current status while still providing enough information to make informed decisions about how to use it. Olsina et al. (2002) want to make it easier to understand and choose metrics. Therefore, the authors propose a sound and flexible mechanism for metric documentation and consultation. The authors present a catalog template for metrics based on a conceptual model (Olsina et al. 2002). The template is also used to document web development metrics (Olsina et al. 2002). For the domain of enterprise architecture management (EAM), Monahov (2014) proposes a metric catalog. Each metric in the catalog is documented with a description structure with a minimal number of description elements (Monahov 2014). A two-part navigation structure based on goals and concerns in EAM enables readers to identify relevant metrics efficiently (Monahov 2014). Some researchers examined approaches to support metric adoption. Hartmann and Dymond (2006) discuss aspects of appropriate agile metrics and propose tools to encourage discussion about the metrics' suitability in a particular context. The tools aim to establish measurements that are more congruent with the objectives of agile teamwork. The authors propose a checklist as a tool to improve understanding of the metrics' intent and use, as well as to mitigate the risk of misuse. The checklist, for example, covers the level of usage to indicate the intended usages for each metric at various levels of the organization. Olszewska et al. (2016) and Bouwers et al. (2014) build on the framework of Kaner et al. (2004) which aim to evaluate if proposed metrics do not conflict with construct validity. The framework consists of ten questions that address topics such as the metric's purpose and measurement instrument. Multiple authors studied generic metric documentation structures. The standard ISO/IEC/IEEE 15939:2017 (ISO 2017) provides an elaboration of the measurement process from ISO/IEC/IEEE 15288:2015 (ISO 2015) and ISO/IEC 12207:2008 (ISO 2008), which is applicable to management disciplines and systems and software engineering. The procedure is described using a model that specifies the process's required activities, how the measures and analysis results should be applied, and how to determine whether the analysis results are valid. In addition, a process is identified to assist in the definition of an appropriate set of measures for specific information needs. Moreover, the standard describes activities and tasks to identify, define, select, apply, and improve measurement. Furthermore, the standard defines commonly used metric terms. In the field of IT controlling, Kütz (2011) discusses metrics and metric models as a basis for controlling and management approaches. Additionally, a MMFS structure was proposed. Neely et al. (2002) discuss available measurement frameworks (e.g., The Balanced Scorecard (Kaplan and Norton 1992)), and introduce their own measurement framework, called the performance prism, to counteract the shortcomings of existing measurement concepts. Moreover, tools, techniques, and methodologies for assessing the various aspects of the performance prism are discussed. In addition, a ten-element generic metric description structure is presented. Popova and Sharpanskykh (2010) define a framework for modeling performance indicators and the relationships between them. The framework is part of a broader organization modeling framework that encompasses a performance-oriented, process-oriented, organization-oriented, and agent-oriented perspective. A description structure is presented in the performance-oriented perspective of the framework for modeling performance indicators. Parmenter (2015) proposes a model for simplifying Key Performance Indicators (KPIs) and avoiding common pitfalls for organizations. Part of the model are critical success factors that serve as guidance. Also, a description structure for KPIs consisting of ten elements is presented.

Step	Output towards the design of the artifacts
Problem identification and motivation	We recognized that extant research on large-scale agile development lacked guidance on how to adopt metrics and did not offer a comprehensive set of metrics applicable to scaling agile development contexts.
Objectives of a solution	Our first objective was the development of an MMFS to solve the problem of lacking guidance regarding metric adoption in scaling agile environments. This MMFS should be consistent with best practices offered by literature on structured collections of metrics, approaches to support metric adoption, and generic metrics documentation structures. Further, the MMFS should incorporate best practices and goals from practice. Our second objective was to develop a metric catalog to provide a comprehensive set of metrics for large-scale agile development. Each metric should be collected from practice and documented with the MMFS. Moreover, the metric catalog should provide navigational aids to select metrics efficiently.
Design and development	The design and development process of the MMFS is illustrated in Figure 1. We conducted an expert interview study to collect metrics, goals, and best practices for metric adoption in scaling agile environments. Moreover, we assessed which best practices from literature on structured collections of metrics (Bouwers et al. 2014; Monahov 2014; Olsina et al. 2002; Olszewska et al. 2016), approaches to support metric adoption (Hartmann and Dymond 2006; Kaner et al. 2004), and generic metrics documentation structures (ISO 2017; Kütz 2011; Neely et al. 2002; Parmenter 2015; Popova and Sharpanskykh 2010) are applicable to our MMFS. Finally, we combined the final set of best practices and goal categories into an MMFS. The final set of elements and their source of origin is illustrated in Figure 1. During the expert interview study, we collected 196 metrics. Subsequently, all metrics were documented with help of the MMFS and combined into a metric catalog providing navigational aids (e.g., Goal-Metric-Matrix) to find relevant metrics efficiently.
Demonstration	We demonstrated the resulting MMFS and metric catalog during an expert discussion and within two case organization as part of a systematic process for metrics adoption and long-term management.
Evaluation	We used three survey-based iterations to compare our design proposals of the MMFS and metric catalog with the objectives described above.
Communication	Communication is being done through this paper.

Table 1. Design Science Research Steps According to Peffers et al. (2007)

Research Methodology

We developed two artifacts for the domain of large-scale agile development using the design science methodology proposed by Peffers et al. (2007): (i) a structured documentation template for metrics in the form of an MMFS, and (ii) a metric catalog, documenting metrics and goals found in the industry using the previously developed MMFS. We chose the design science research approach since it is well-known in information systems research and can be used to solve important organizational problems by developing and evaluating purposeful IT artifacts (Hevner et al. 2004). The design of the artifacts is partly based on the insights of an expert study. We describe the expert study's research methodology in another publication (Philipp et al. 2022). Table 1 summarizes each step of the design science approach.

Results

MMFS

In this section, we answer RQ1: *How can a generic and minimalistic metric management fact sheet (MMFS) for the domain of large-scale agile software development be designed?* Therefore, we present the design process of the MMFS (see Figure 1) and propose a minimalistic and generic structure to document metrics and support their organization-specific implementation. By searching current software development literature, we were able to identify description structures and their associated elements during design and development. Following that, we evaluated each identified description element's suitability for the domain of large-scale agile development. Inspired by Monahov (2014), we provide an overview of all elements from literature, including their origin (see Figure 1). The elements without a checkmark are unique to our design and stem from insights gained in the expert interviews. In total, we used ten elements in our MMFS. Our MMFS is the most similar to Monahov's (2014) design because both structures share the most elements. The calculation rule and title are the most proposed elements among all sources. Like Monahov (2014), we structured the elements into two categories. We declared elements as general elements if they were "independent from the context of a given organization" (Monahov 2014). We consider elements that "describe the configuration of the metric in a given organization" to be organization-specific elements (Monahov 2014). The findings of our expert interview study confirm the importance of organization-specific elements, as organizations adopted similar metrics but used different configurations in terms of tools, responsible persons, terminology, and metric target values on multiple occasions. Further, organization-specific elements allow keeping information confidential (e.g., employee names).

Elements of the MMFS

Like Monahov (2014), we only included "a minimal number of description elements, which are required to ensure a comprehensive metric documentation as a starting point for an organization-specific metric implementation." Subsequently, we provide a description and a rationale for each element included. We included the following seven elements as general elements:

(1) Title: The title is written in natural language and provides a concise and precise metric summary. It aids in the search for metrics. Excluding the title can make it more difficult for metric stakeholders to understand a metric quickly and easily (Monahov 2014).

(2) Description: The purpose and meaning of the metric are explained in detail in this description. As a result, it contributes to a common understanding of a metric's motivation. Dismissing the description obstructs the involved stakeholders' shared understanding of the motivation, expected benefits, and assumptions (Monahov 2014).

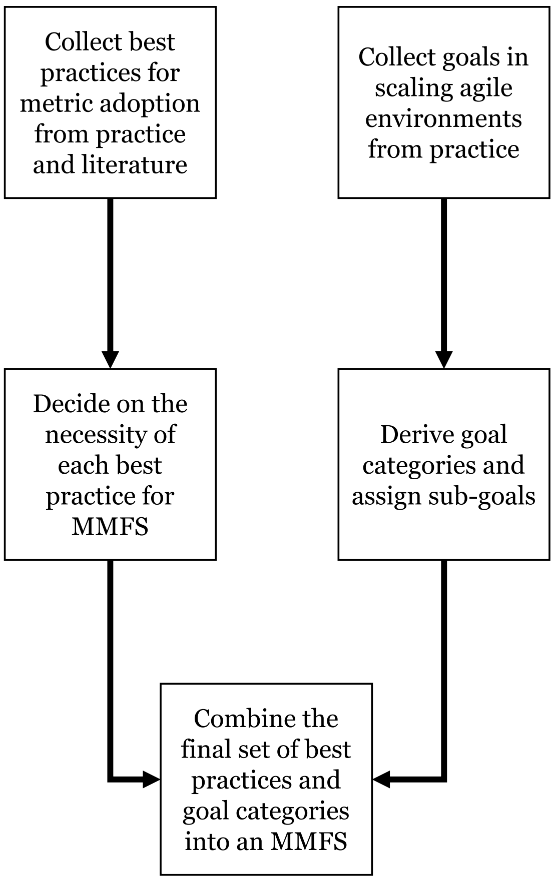
(3) Calculation rule: The calculation rule should include all relevant variables and document how a metric is calculated. A calculation rule can be expressed in natural or formal language. Both approaches are valid, according to Hartmann and Dymond (2006) and Bouwers et al. (2014). We take Monahov's (2014) approach and formulate the calculation rule in natural language while keeping it aligned with its formal definition. Thus, an easy understanding of the metric without losing information is possible, in our opinion. Furthermore, our interview research revealed that ignoring the calculation rule can cause issues because some tools require it.

(4) Information model: We document and visualize the information model underlying a metric using a UML class diagram, as suggested by Monahov (2014). The information models aid in the identification of each data item required for the calculation as well as the comprehension of the relationships between the data items. Furthermore, the information model allows for the mapping of each metric to a context and terminology specific to the organization. The data items and their relationships cannot be visualized without the information model (Monahov 2014).

(5) Code: The code functions as a metric identifier. It facilitates the quick retrieval of a metric and ensures an unambiguous use by all stakeholders (Monahov 2014). The removal of the code from the MMFS makes retrieving metrics more difficult (Monahov 2014).

(6) Goals: We assign each metric to at least one management goal within the large-scale agile development environment, similar to Monahov (2014). Each metric contributes to the achievement of the goals. There are five goal categories and 27 sub-categories that we identified in our interview study. By excluding goals from the MMFS, the link between goals and metrics is interrupted, which is a critical aspect of ensuring the efficiency of metrics (Basili et al. 1994).

(7) Organizational level: Each metric is linked to the organizational level on which it is used. According to Hartmann and Dymond (2006), the organizational level is particularly important in an agile context because it can indicate the metrics usage level's limits. In addition, the organizational level helps in the retrieval of metrics. Metrics and related objectives are often linked to organizational levels in the domain of large-scale agile development (Korpivaara et al. 2021; Stettina and Schoemaker 2018). We connected each metric to the team, program, portfolio, or enterprise level in alignment with Korpivaara et al. (2021) and Stettina and Schoemaker (2018). Removing the organizational level from the MMFS would make it impossible for stakeholders to link a metric to its organizational usage level directly and unambiguously.



MMFS element	Neely et al. [61]	Popova and Sharpanskykh [104]	Parmenter [62]	Kütz [131]	ISO/IEC/IEEE 15939:2017(E) [130]	Hartmann and Dymond [63]	Monahov [129]	Bouwers et al. [64]	Olsina et al. [128]	Kaner et al. [106]	Olszewska et al. [90]
General elements											
Title	✗	✗		✗	✗	✗	✗	✗	✗		✗
Description			✗	✗			✗		✗		
Calculation rule	✗	✗		✗	✗	✗	✗	✗	✗	✗	✗
Information model							✗				
Code			✗				✗				
Goals							✗				
Organizational level						✗				✗	✗
Organization-specific elements											
Organizational instance			✗								
Data mapping											
Data item							✗				
Mapped name							✗				
Data source	✗			✗			✗				
Data contact				✗			✗				
Properties											
Measurement tool									✗	✗	✗
Calculation type											
Measurement frequency/event	✗		✗	✗			✗		✗		
Monitoring frequency/event	✗										
Reporting frequency/event											
Owner				✗					✗		
Definition responsible(s)											
Implementation responsible(s)											
Achievement responsible(s)	✗			✗			✗				
Monitoring responsible(s)											
Target value(s)	✗		✗	✗			✗	✗	✗		
Planned value(s)	✗			✗			✗	✗	✗		
Tolerance value(s)		✗		✗	✗		✗	✗	✗		
Escalation rule(s)				✗	✗		✗				
Interested stakeholder						✗					

Figure 1. Visualization of the MMFS Design Process (left) and Final MMFS Elements (right)

We included the following three elements as organization-specific elements:

(8) Organizational instance: Our expert interviews revealed that most organizational environments have a variety of organizational structures (i.e., multiple teams, programs, or portfolios). Metrics may be implemented multiple times within a single organization. As a result, the organizational instance aims to document the organizational unit of each metric, such as a single team implementing a specific metric. It would be more difficult to identify the metrics' affiliation if the organizational instance was removed from the MMFS.

(9) Data mapping: This element aids in the organization-specific adoption of metrics. Organizations use different terminology for the data items of the information model, according to our interview study. As a result, we took Monahov's (2014) approach and mapped each data item in the information model to its organization-specific context. The internal term or concept is linked to a class, attribute, or relationship by

the mapped name. In addition, each data item is associated with the appropriate data source in order to document all data sources required for metric calculation. The data contact responsible can also be specified. This role is in charge of ensuring data quality and availability. The element's removal from the MMFS would result in issues with organization-specific terminology mapping, documentation of required data sources, and data contacts (Monahov 2014).

(10) Properties: To better embed the metric into its context, we propose a property element with 15 organization-specific attributes. All tools used during the metric adoption should be documented using the first attribute measurement tool. We recommend specifying the calculation type, as suggested by Olsina et al. (2002), to indicate whether a metric is calculated manually or automatically. Our expert study showed such a specification is reasonable since some tools require manual actions, whereas others are fully automated. Moreover, we recommend using an attribute for measurement frequency or event to understand the metrics underlying time series and bound it to agile events such as the sprint or program planning. The frequency or event of reporting is another recommended attribute. This attribute encourages bringing in collected data on a regular basis in the form of a report and sharing it with relevant stakeholders. The evidence from our interviews supports the definition of events for monitoring and reporting. These routines are often carried out during recurring agile events. Furthermore, we advocate defining responsibilities for activities related to metric adoption. A metric owner is responsible for ensuring sufficient resources for adopting the metric such as budget or personal. The owner is a receiver of reporting information. Four additional roles can be defined to clarify responsibilities regarding the metric definition, implementation, monitoring, and achievement. The metric achievement responsible is in charge of ensuring that the defined metric target values are met on time. One person can have multiple roles. For each responsibility, we recommend naming only one person. The remaining attributes, the target value(s), planned value(s), tolerance value(s), can be used to know when to follow the escalation rule. The planned values indicate intermediate goals at particular points in time. The escalation rule documents the responsible person's action plan when tolerance values are reached. Interested stakeholders, the final attribute, can include additional metric stakeholders. By ignoring the properties element, multiple critical elements required for successfully implementing a metric are left undocumented (Monahov 2014).

Metric Catalog[1]

The second research question is: *How can a structured collection of metrics help organizations execute large-scale agile software projects?* As a result, we propose a metric catalog design to aid in the establishment of metrics in large-scale agile software development. Matthes et al. (2012) influenced the catalog structure. Our catalog begins with a brief introduction concentrating on the motivation for designing it, its purpose, and the intended target group (i.e., practitioners and researchers). Finally, we provide navigation support to ensure that relevant metrics are retrieved quickly. Three different navigation support elements make up the navigation structure. The first element, navigation based on large-scale agile software development goals, uses a Goal-Metric-Matrix like Monahov (2014). The second element, navigation based on the metric's organizational level of usage, is connected to the organizational level element of the designed MMFS. The final navigation element, navigation through related metrics, can be used to discover metrics based on relationships. The core of the catalog is the collection of 196 metrics. Each metric is documented using the MMFS, except that for each metric, additionally all related metrics are listed and linked (see Figure 1).

Demonstration and Evaluation

During an initial demonstration, the artifacts were presented to three experts. All experts confirmed the overall idea for both artifacts, the clear structure and design of the MMFS, and the importance of connecting metrics with goals. Thereafter, we conducted three iterations of demonstrations and survey-based evaluations. The first two iterations for the MMFS and metric catalog included expert evaluations since multiple experts participated in the assessment (Peffers et al. 2012). We used the first iteration to receive feedback during a demonstration which was incorporated before performing the second iteration. In the second iteration we assessed if potential users and essential stakeholders perceive the solution artifacts as valuable. In addition to the first two iterations, we performed a third iteration for the metric catalog where

[1] Metric Catalog for Large-Scale Agile Development: https://bit.ly/39bEJRk

we applied it in two case organizations as part of a systematic process for metric adoption and long-term management. Thus, the evaluation in the third iteration was an action research evaluation (Peffers et al. 2012). In total, 14 experts participated in the MMFS evaluation and 24 experts in the metric catalog evaluation.

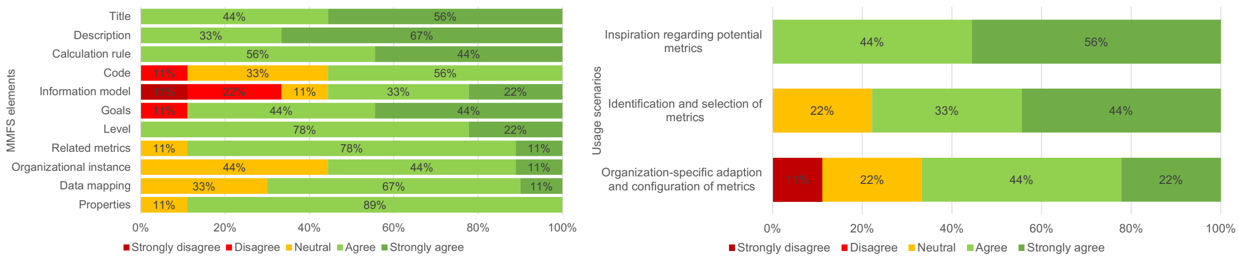


Figure 2. Evaluation of the MMF Elements and the Catalog's Usefulness in Different Usage Scenarios

(1) During the first demonstration, the artifacts were presented to 5 experts. Four recommendations were made by the experts. First, one expert suggested reformulating the calculation rule in mathematical form because it is easier to prove and relate to other metrics or assumptions this way. Because users' understanding of the mathematical definition is dependent on prior mathematical knowledge, we did not incorporate this suggestion for improvement. Second, one expert suggested extending the five goal categories we initially suggested for the goal element since it did not provide sufficient granularity for a reasonable connection between metrics and goals. We incorporated the feedback by extending the five goal categories with 27 sub-goals. Third, one expert proposed to include "a range for every metric showing you if the metric is useful or not useful." Again, we did not incorporate this feedback due to a lack of required data. Forth, including the purpose of each metric, was the final improvement suggestion of one expert. We did not add an extra element for this requirement because the second improvement suggestion (adding the 27 sub-goals) already includes information about the metric's purpose.

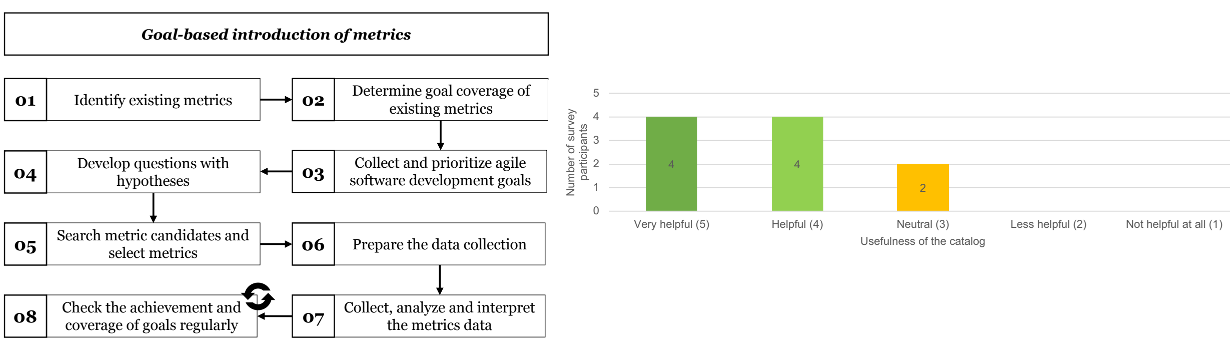


Figure 3. Metric Adoption Process (left) and Evaluation of the Catalog's Usefulness (right)

(2) During the second demonstration, the artifacts were presented to 9 experts. We evaluated whether the MMFS and metric catalog are valuable enough to be used in practice after incorporating feedback from the first iteration of our survey-based evaluation, or if they require further modification. First, we evaluated the relevance and usefulness of each element of the MMFS (see Figure 2). The survey responses show that each element is considered relevant and valuable. Only vis-à-vis the code, information model, and goals did some experts exhibit disagreement. Therefore, we decided not to apply changes to any MMFS elements. Subsequently, we introduced the experts to three possible usage scenarios of the catalog. The usage scenarios are inspiration in terms of potential metrics, identification and selection of metrics, and organization-specific adaption and configuration of metrics. Then, the experts were asked to evaluate the usefulness of each of the three usage scenarios. Even if most experts can imagine applying the catalog for all three scenarios, using the catalog for inspiration, identification and selection were the scenarios with the most agreeing answers. Therefore, we chose inspiration regarding potential metrics and identification and selection of metrics as usage scenarios for the catalog's third iteration of demonstration and survey-based evaluation.

(3) We used the catalog as part of our process for metric adoption within two case organizations (see Figure 3). The process aims at a goal-based introduction of metrics. Within the process the metric catalog was used in two process steps. In Step 2, teams determined the goal coverage based on previously identified metrics. Therefore, they conducted bottom-up searches within the catalog. Moreover, teams applied the catalog's Goal-Metric-Matrix top-down in Step 5 to compile a list of metric candidates and then select the metrics that best matched the goals and questions. We received the feedback that some relevant goals (e.g., improving collaboration) and metrics (e.g., security-related story points per sprint) are not included in the catalog. Both improvement suggestions were incorporated into the catalog. The results show that most survey participants experienced the catalog as very helpful or helpful (see Figure 3).

Discussion

Our research yielded two major conclusions. First, our study shows that a metric catalog for the domain of large-scale agile software development can aid practitioners in metric adoption. As a result, the catalog can serve as a source of ideas for potential metrics. It also makes metric identification, selection, and organization-specific configuration and adaption a lot easier. The helpfulness of the MMFS for the organization-specific configuration and adoption was also a finding of Monahov (2014) in a case study in the domain of enterprise architecture management. Second, we contend that the MMFS can facilitate metric adoption by addressing some of the challenges and success factors identified in our expert interview study. In total, we identified eleven challenges and 22 success factors (Philipp et al. 2022). The information model and data mapping elements can help overcome data collection challenges. Some experts experienced a lack of metric usefulness, which can be overcome by the description element highlighting the metric motivation and purpose. Practitioners can tackle metric calculation challenges with the calculation rule. All elements counteract a lack of metric understanding. The alignment of metrics to higher-level goals helps overcome the challenge of focusing on local instead of a global optimum. Each element of the MMFS supports the avoidance of metric definition challenges. Due to its organization-specific elements, the MMFS considers the success factor of a context-dependent metric adoption. Further, it incorporates the success factor of understanding the metric purpose. The organization-specific elements help achieve the success factor of managing the interplay of the metric and its environment. Furthermore, employing the MMFS simplifies metric adoption, contributing to the success factor of keeping metric adoption simple. The MMFS helps implementing the success factor standardization of metric adoption by providing a structure for each metric to be documented in the same format. Moreover, the goal element helps achieve the success factor of ensuring goal orientation of each metric. Furthermore, all elements contribute to the success factors of providing documentation. The title, description, and calculation rule all support accomplishing the success factor of understanding what the metric is measuring. With the responsibility attributes, the properties element achieves the success factor of assigning clear responsibilities. Finally, the attributes target, planned, and tolerance values ensure the success factor of realistic expectation management. Regarding the design science research, the limited number of experts participating in the evaluation is a limitation. Further, we created both artifacts based on the insights from a limited number of expert interviews and scientific studies. Additionally, only two case organizations from the same industry applied the catalog. We evaluated the threats to the validity of our expert interviews in another publication (Philipp et al. 2022).

Conclusion

In this study, we designed and used an MMFS to create a metric catalog for large-scale agile development that includes 196 metrics and their goals. By investigating related work and conducting 23 experts interviews in thirteen organizations, we were able to inform the design choices for both artifacts. We showed that practitioners could use the catalog to get inspired regarding potential metrics and ease the identification, selection, organization-specific metric adoption and configuration. Furthermore, the MMFS can assist in overcoming challenges that arise during metric adoption. In the future, new metrics and goals should be identified and added to the metric catalog. Moreover, as an additional navigation element, a stakeholder-based metric identification could extend the catalog. We recommend testing the catalog in further case organizations to receive more suggestions for improvement. Moreover, we plan to implement a prototypical web application of the metric catalog to ease the management of metrics and respective goals.

REFERENCES

Ambler, S.W., and Lines, M. 2012. *Disciplined Agile Delivery: A Practitioner's Guide to Agile Software Delivery in the Enterprise*. IBM Press.

Basili, V.R., Caldiera, G., and Rombach, H.D. 1994. "The Goal Question Metric Approach," in: *Encyclopedia of software engineering*. pp. 528-532.

Beedle, M. 2018. "Enterprise Scrum Definition - Agility of the 21st Century." Retrieved 16. February, 2022, from http://static1.1.sqspcdn.com/static/f/608893/27795406/1515443164947/Enterprise+Scrum+Definition.pdf?token=bHxw30Cxx%2BfiftPIBq50%2F8RLYWU%253

Bouwers, E., Deursen, A.v., and Visser, J. 2014. "Towards a Catalog Format for Software Metrics," *Proceedings of the 5th International Workshop on Emerging Trends in Software Metrics (WETSoM)*, Hyderabad, India: ACM, pp. 44-47.

Briand, L.C., Morasca, S., and Basili, V.R. 1996. "Property-Based Software Engineering Measurement," *IEEE Transactions on Software Engineering* (22:1), pp. 68-86.

Brown, A.W., Ambler, S., and Royce, W. 2013. "Agility at Scale: Economic Governance, Measured Improvement, and Disciplined Delivery," *Proceedings of the 35th International Conference on Software Engineering (ICSE)*, San Francisco, CA, USA: IEEE, pp. 873-881.

Cohen, D., Lindvall, M., and Costa, P. 2004. "An Introduction to Agile Methods," in *Advances in Computers, Advances in Software Engineering,* M.V. Zelkowitz (ed.). Amsterdam: Elsevier, pp. 1-66.

Dikert, K., Paasivaara, M., and Lassenius, C. 2016. "Challenges and Success Factors for Large-Scale Agile Transformations: A Systematic Literature Review," *Journal of Systems and Software* (119), pp. 87-108.

Dingsøyr, T., and Moe, N.B. 2014. "Towards Principles of Large-Scale Agile Development," *Proceedings of the 15th International Conference on Agile Software Development (XP 2014),* T. Dingsøyr, N.B. Moe, R. Tonelli, S. Counsell, C. Gencel and K. Petersen (eds.), Rome, Italy: Springer, Cham, pp. 1-8.

Dumitriu, F., Meșniță, G., and Radu, L.-D. 2019. "Challenges and Solutions of Applying Large-Scale Agile at Organizational Level," *Informatica Economica* (23:3), pp. 61-71.

Dybå, T., and Dingsøyr, T. 2008. "Empirical Studies of Agile Software Development: A Systematic Review," *Information and Software Technology* (50:9-10), pp. 833-859.

Dybå, T., and Dingsøyr, T. 2009. "What Do We Know About Agile Software Development?," *IEEE Software* (26:5), pp. 6-9.

Fenton, N.E., and Neil, M. 2000. "Software Metrics: Roadmap," *Proceedings of the Conference on the Future of Software Engineering*, Limerick, Ireland: ACM, pp. 357-370.

Hartmann, D., and Dymond, R. 2006. "Appropriate Agile Measurement: Using Metrics and Diagnostics to Deliver Business Value," *Proceedings of the AGILE 2006 (AGILE'06),* Minneapolis, MN, USA: IEEE, pp. pp. 6 pp.-134.

Hevner, A.R., March, S.T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75-105.

ISO. 2008. "Iso/Iec International Standard - Systems and Software Engineering - Software Life Cycle Processes," in: *ISO/IEC 12207:2008*.

ISO. 2015. "Iso/Iec/Ieee International Standard - Systems and Software Engineering - System Life Cycle Processes," in: *ISO/IEC/IEEE 15288:2015*.

ISO. 2017. "Iso/Iec/Ieee International Standard - Systems and Software Engineering - Measurement Process " in: *ISO/IEC/IEEE 15939:2017*. pp. 1-49.

Kaner, C., Member, S., and Bond, W.P. 2004. "Software Engineering Metrics: What Do They Measure and How Do We Know?," *Proceedings of the 10th International Symposium on Software Metrics*, Chicago, Illinois: IEEE.

Kaplan, R.S., and Norton, D.P. 1992. "The Balanced Scorecard: Measures That Drive Performance," *Harvard Business Review* (83:7), pp. 71-79.

Korpivaara, I., Tuunanen, T., and Seppänen, V. 2021. "Performance Measurement in Scaled Agile Organizations," *Proceedings of the 54th Annual Hawaii International Conference on System Sciences (HICSS)*, Kauaii, Hawaii, USA: University of Hawai'i at Manoa, pp. 6912-6921.

Kruchten, P. 2013. "Contextualizing Agile Software Development," *Journal of Software: Evolution and Process* (25:4), pp. 351-361.

Kütz, M. 2011. *Kennzahlen in Der It: Werkzeuge Für Controlling Und Management*. Heidelberg: dpunkt. verlag.

Larman, C., and Vodde, B. 2016. *Large-Scale Scrum: More with Less*. Boston, MA: Addison-Wesley Professional.

Leffingwell, D. 2018. *Safe 4.5 Reference Guide: Scaled Agile Framework for Lean Enterprises*. Addison-Wesley Professional.

Maiden, N., and Jones, S. 2010. "Agile Requirements Can We Have Our Cake and Eat It Too?," *IEEE Software* (27:3), pp. 87-88.

Matthes, F., Monahov, I., Schneider, A., and Schulz, C. 2012. "Eam Kpi Catalog V 1.0," Technische Universität München, Munich, Germany.

Monahov, I. 2014. "Integrated Software Support for Quantitative Models in the Domain of Enterprise Architecture Management." Technische Universität München.

Neely, A.D., Adams, C., and Kennerley, M. 2002. *The Performance Prism: The Scorecard for Measuring and Managing Business Success*. Prentice Hall Financial Times London.

Nord, R.L., Ozkaya, I., and Kruchten, P. 2014. "Agile in Distress: Architecture to the Rescue," *Proceedings of the 15th International Conference on Agile Software Development (XP 2014),* T. Dingsøyr, N.B. Moe, R. Tonelli, S. Counsell, C. Gencel and K. Petersen (eds.), Rome, Italy: Springer, Cham, pp. 43-57.

Olsina, L., Lafuente, G., and Pastor, O. 2002. "Towards a Reusable Repository for Web Metrics," *Journal of Web Engineering* (1:1), pp. 61-73.

Olszewska, M., Heidenberg, J., Weijola, M., Mikkonen, K., and Porres, I. 2016. "Quantitatively Measuring a Large-Scale Agile Transformation," *Journal of Systems and Software* (117), pp. 258-273.

Paasivaara, M. 2017. "Adopting Safe to Scale Agile in a Globally Distributed Organization," *Proceedings of the 12th International Conference on Global Software Engineering (ICGSE)*, Buenos Aires, Argentina: IEEE, pp. 36-40.

Parmenter, D. 2015. *Key Performance Indicators: Developing, Implementing, and Using Winning Kpis*. John Wiley & Sons.

Peffers, K., Rothenberger, M., Tuunanen, T., and Vaezi, R. 2012. "Design Science Research Evaluation," *Proceedings of the International Conference on Design Science Research in Information Systems*, Las Vegas, NV, USA: Springer, pp. 398-410.

Peffers, K., Tuunanen, T., Rothenberger, M.A., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* (24:3), pp. 45-77.

Petersen, K., and Wohlin, C. 2010. "The Effect of Moving from a Plan-Driven to an Incremental Software Development Approach with Agile Practices," *Empirical Software Engineering* (15:6), pp. 654-693.

Philipp, P., Tobisch, F., and Matthes, F. 2022. "Challenges and Success Factors for Metrics in Large-Scale Agile Development," in: *28th Americas Conference on Information Systems*. Minneapolis, Minnesota: AIS.

Popova, V., and Sharpanskykh, A. 2010. "Modeling Organizational Performance Indicators," *Information systems* (35:4), pp. 505-527.

Scaled Agile Inc. 2021. "Metrics." Retrieved 29. January, 2022, from https://www.scaledagileframework.com/metrics/

Stettina, C.J., and Schoemaker, L. 2018. "Reporting in Agile Portfolio Management: Routines, Metrics and Artefacts to Maintain an Effective Oversight," *Proceedings of the 19th International Conference on Agile Software Development (XP 2018),* J. Garbajosa, X. Wang and A. Aguiar (eds.), Porto, Protugal: Springer, Cham, pp. 199-215.

Thompson, K. 2013. "Recipes for Agile Governance in the Enterprise." Retrieved 16. February, 2022, from https://www.cprime.com/wp-content/uploads/woocommerce_uploads/2013/07/RAGE-Final-cPrime1.pdf

Uludağ, Ö., Kleehaus, M., Caprano, C., and Matthes, F. 2018. "Identifying and Structuring Challenges in Large-Scale Agile Development Based on a Structured Literature Review," *Proceedings of the 22nd International Enterprise Distributed Object Computing Conference (EDOC)*, Stockholm, Sweden: IEEE, pp. 191-197.