Requirements for a quality measurement instrument for semantic standards[1]

Erwin Folmer
UNIVERSITY OF TWENTE
TNO INFORMATION AND COMMUNICATION TECHNOLOGY

Dennis Krukkert
TNO INFORMATION AND COMMUNICATION TECHNOLOGY

Paul Oude Luttighuis
NOVAY

Jos van Hillegersberg
UNIVERSITY OF TWENTE

Abstract

This study describes requirements for an instrument to measure the quality of semantic standards. A situational requirements engineering method was used, resulting in a goal-tree in which requirements are structured. This structure shows requirements related to the input of the instrument; stating that the instrument should be useful for a set of different semantic standards. It also shows that the instrument should be efficient and especially easy to use. Finally there a set of requirements related to the outcome of the instrument, stressing that a high quality outcome is important, including improvement suggestions. Based on this set of requirements a foundation for the design phase has been created.

1. Introduction

In the late 80's and early 90's, e-business was only available for large companies because of the costs of Value Added Networks (VAN) necessary for EDI. The introduction of XML and the Internet made e-business accessible for SME's. As a result, lots of XML based, semantic standards were developed. Semantic standards describe the syntax and semantics of messages that are exchanged, and are usually developed in a certain branch. Although these standards are usually developed with the best intentions, they often have quality issues like difficult to understand, multiple interpretations, etc. Previous studies (Folmer, Berends, Oude Luttighuis, & Van Hillegersberg, 2009; Folmer, Oude Luttighuis, & Van Hillegersberg, 2010) showed that improving the quality of these standards will result in better interoperability between (IT systems of) organisations, while on the other hand this topic qualifies as research gap. In order to improve the quality of semantic standards, an instrument is needed to measure the quality. Before developing such an instrument, it is necessary to determine its requirements, bringing us to the main research question in this paper:

What are requirements for an instrument to measure the quality of semantic standards?

The goal of this study is to answer this research question by performing a requirements study. We start by presenting the chosen research method, followed by details about the gathering process. We will present the resulting requirements, as well as our conclusions and directions for further research.

2. Research Method

For our purposes, we decided to embrace the notion of situational method engineering (Brinkkemper, 1998; Coulin, 2006), and assemble our requirements gathering process using fragments from three well-known requirements engineering methods: QFD, KAOS, and Volere.

QFD is a method for requirements elicitation and transformation of requirements into product design. It has been developed by Akao (Akao, 1990), based on the quality concepts of Deming. It is primarily used for designing physical products, but can also be used for IT products. Its best known aspect is the so-called House of Quality. But, QFD also includes a team-based iterative method for understanding the customer requirements. The House of Quality is a matrix with the "whats" and the "hows" are plotted on each of its axes. The "whats" represents the customer requirements and the "hows" represent the functional requirements for the system. In consecutive steps, the "hows" from the previous step are the "whats" for the next step. This gives a level-like structure to requirements, while maintaining the link with the customer requirements at the highest level.

The KAOS (Knowledge Acquisition in automated Specification) methodology has been designed at the University of Leuven (Louvain) in the early 1990s, and continues to be improved (Al-Subaie & Maibaum, 2006). KAOS is a so called "Goal Oriented Requirements Engineering (GORE)" (Van Lamsweerde, 2004) method. In a GORE method, a "goal-tree" is developed during the requirements engineering phase. All goals and requirements must contribute to a higher-level goal in the goal tree, and eventually leading to one pre-defined top-level goal. This property supports the requirements elicitation and selection process because one can find higher level goals by asking the "why" question, and lower level goals by asking the "how" question. It is a flexible method in the sense that it supports top-down, bottom-up and middle-out requirements gathering and also makes it possible to start by providing guidelines on how different requirements relate, and by relating requirements to a pre-defined, top-level goal.

On a different level there is the Volere Requirements Process Model (Robertson & Robertson, 1999), which is a process for gathering and testing requirements. An important pragmatic element of Volere is the Volere shell; a template to make sure you gather all information about a requirement, such as the history of the requirement, customer satisfaction, its rationale and fit criteria.

Although each of these three methods might have done the job sufficiently for our requirements study, neither one perfectly matched our situation. Therefore, we chose a combination of elements from each of the methods. The goal-tree approach of KAOS was selected for its ability to structure requirements, just as the reasoning approach (asking how and why questions). KAOS formal information modelling approach was not chosen, because it was too extensive for our purposes.

From QFD, we took the customer approach and the workshops, as efficient ways to involve stakeholders. Since domain expertise is essential in requirements elicitation, we involved

potential end users of the instrument in our workshops. Participants have backgrounds in international standardisation initiatives and compliance testing. A two-step approach was chosen to improve the results and also not to ask too much time of participants. The workshops were lead by requirements gathering experts, who afterwards processed the outcome in a consistent and complete result. The House of Quality was not used, because it includes considerable amounts of physical product-related aspects, and the goal-tree from KAOS present a viable alternative intended for use in the domain of information systems.

Like QFD, Volere emphasizes end user involvement, the role of domain expertise and the use of workshops for elicitation. Our use of the KAOS goal tree ends with the identification of the requirements. From the Volere shell, we took additional attributes of the requirements, like "Fit Criterion", enabling us to express how compliance to a requirement can be tested preventing requirements from vagueness, and "Priority" useful when requirements compete, either because they conflict, or because implementation resources are scarce. Another important attribute from Volere "rationale" was not explicitly used, since it follows from the goal tree.

3. The process

Preparations were carried out for two workshops involving potential end users. Five domain experts participated in the first workshop, which was held in June 2009. Experience from semantic standards from different domains was included by these experts: temporary staffing (hr-XML, SETU), finance/e-invoicing (UBL), disaster management, education (IMS, Edustandaard) and healthcare (HL7, CEN/EN 13606, Nictiz, Vektis).

In the second workshop, also five experts contributed. This time, experience from technical Standard Development Organisations (SDO's) was involved: IEEE, 3GP, OMA, OPT, and ITU-T. Although these are not the main type of potential users, this session was extremely valuable. Those involved in technical standards have many years of experience, while expertise within semantic standards is relatively new. Semantic interoperability does not have the same rich history as technical interoperability.

The exact form of the instrument (e.g. software tool, method or book) was not determined prior to the workshops. We wanted the participant not to feel restricted beforehand. Also, the meaning of concepts like quality and semantic standard was left implicit. This turned out to work quit well, since interesting discussions started on details of definitions. Figure 1 was used as the starting point of the workshops. It shows the instrument as a black box converting input (standard) to output (report). It also suggest possible forms of the instrument, like some kind of handbook and/or tool. Different actors are shown that will be involved in using the instrument. The distinction between the principal and tester shows a possible differentiation in the person who commissions the use of the instrument and is selecting the measurements and the persons who is carrying out the measurements (tester).
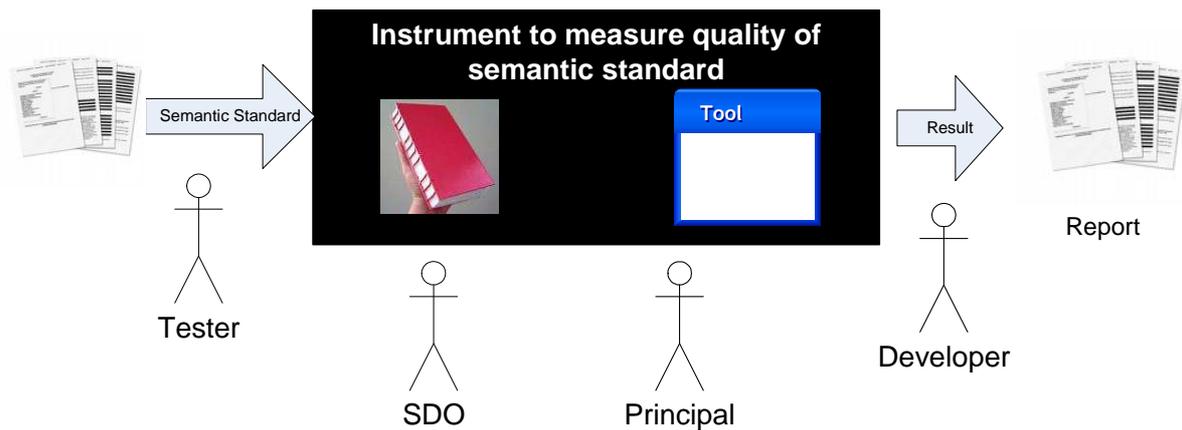
Figure 1 - Context diagram

During the workshop, participants were asked to think about, and write down the requirements and, after several minutes, present them to the other participants. With help of the requirements expert, the requirement was then added to the goal tree. This process was repeated several times within the workshop. This constitutes a bottom-up approach, starting with a set of initial requirements and expanding it by asking how and why questions.

The result was a large amount of post-its, including redundant requirements, vague descriptions, general remarks, etc. Processing these involved selection and removal (redundant requirements, remarks), structuring within a tree, completing the goal-tree by adding requirements, and formulating the requirement. Then, the requirements were annotated with fit criteria and priorities.

4. The results

This section will present the goal tree gathered from the workshop sessions. It will start with the top-level goal, and the three level-two goals. In each of the following sections, one of the level-two goals is further decomposed. At the bottom of the goal tree (the leafs), requirements are specified.

4.1 Overview

The top-level goal of the instrument is to support semantic SDO's in developing high quality standards. The rationale for this goal is the general believe that higher quality standards will lead to improved interoperability. The term SDO is used throughout this paper, while others (including (Cargill & Bolin, 2007; Jakobs, 2009) restrict this term for use on formal organisations like ISO only, and use the term Standard Setting Organisation (SSO) or Standard Setting Body (SSB) for non-formal organisations. Since this distinction is not relevant for our purpose, we use the term SDO for all organisations involved in standards development and maintenance. Figure 2 shows the top-level goal, and the three level-two goals.
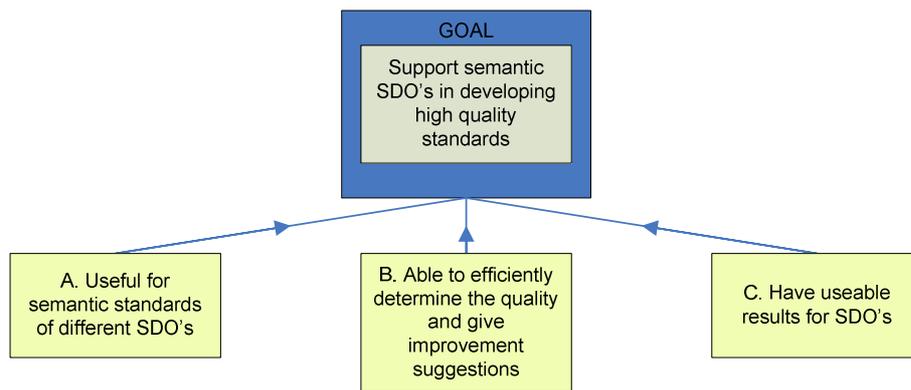
Figure 2 - Top of the goal tree

These three sub goals will be decomposed in the following sections. A detailed description of all requirements is presented in appendix 1.

4.2 Useful for semantic standards of different SDO's

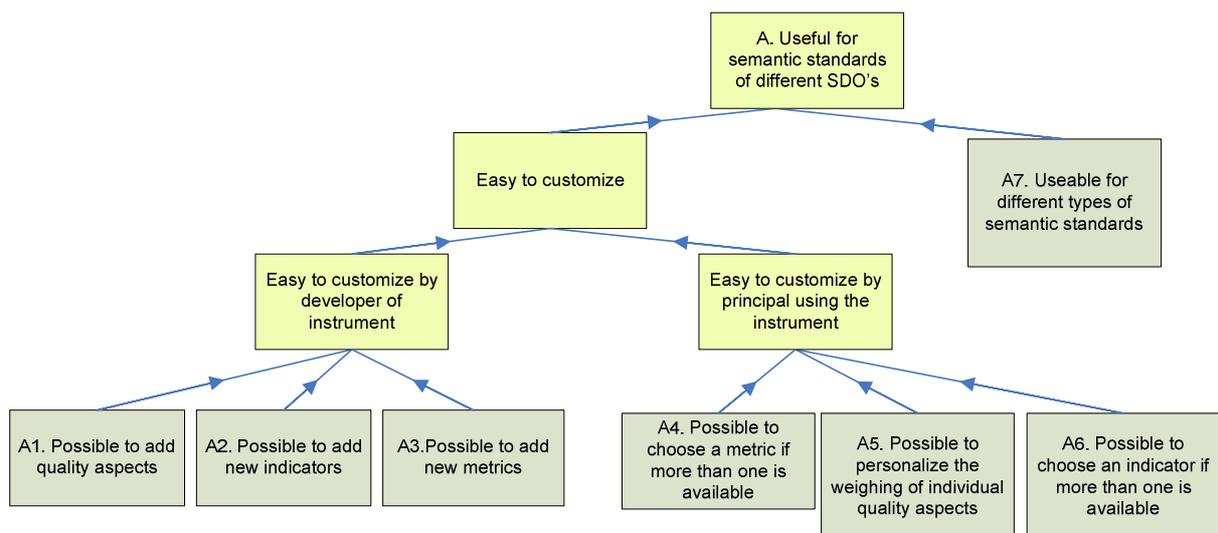Figure 3 gives an overview of all the sub-goals and requirements that need to be fulfilled for this level-two goal.

Figure 3 - Useful for semantic standards of different SDO's

First, the instrument should be easy to customize. This is because SDO's differ in their approach and in the quality aspects they find important. Also, the instrument should be useable for different types of semantic standards.

Regarding the customizability of the instrument, it is important that several roles involved in using the instrument can perform the customization. For the designer of the instrument, it is important that new elements (quality aspects, indicators, and metrics) can easily be added to the instrument. For the principal using the instrument, it is important that he can easily choose between different elements (e.g. measurements) of the instrument (if more are available), and that he can customize his "view" on quality by easily changing the weight factors.

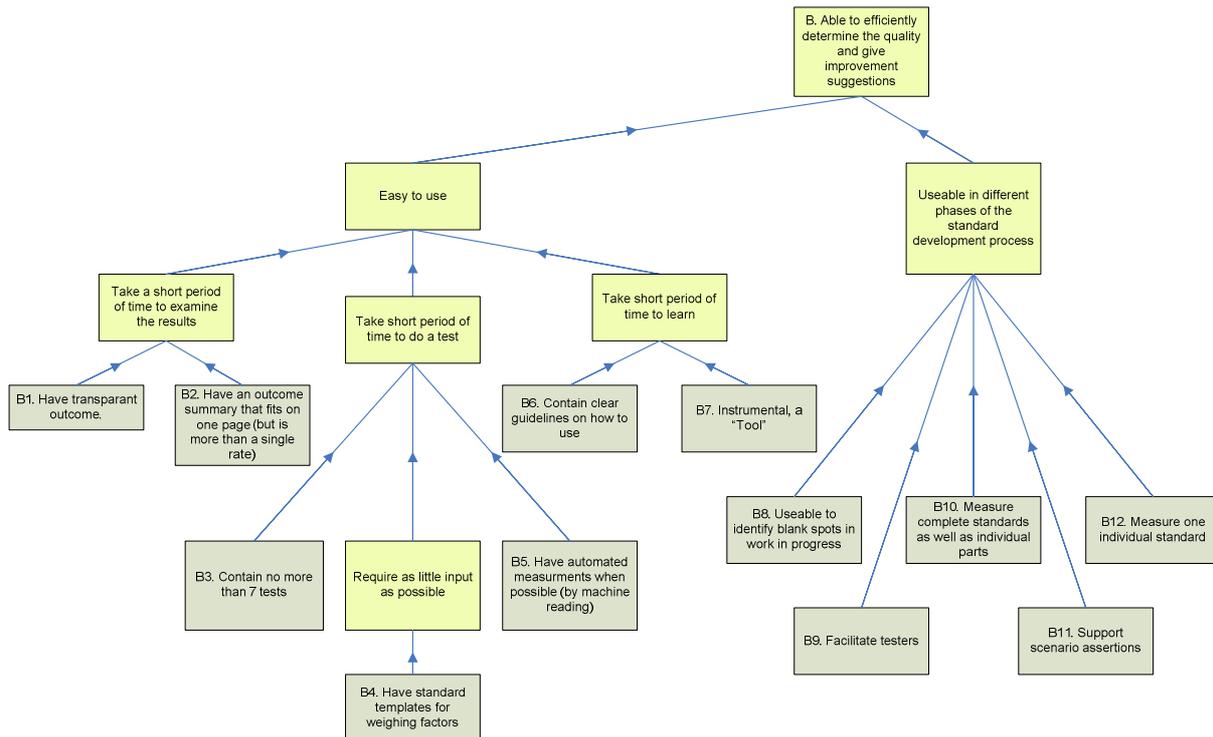## 4.3 Efficiently determine the quality and improvement suggestions

Figure 4 - Efficiently determine the quality and improvement suggestions
Figure 4 shows the decomposition of this level-two goal.

During the workshops, ease of use mainly focused on the time required to execute certain activities. A distinction was made between the time for learning how to use the instrument (short learning curve), the time for executing a test, and the time for interpreting the results. Requirements were specified for all three aspects. In order to reduce the time taken for execution, the instrument should require as little input as possible.

In order to make the instrument useable in different phases of the standard development process, a number of requirements have to be met. These requirements focus mainly on the input that has to be provided to the instrument, as well as some functional aspects of the instrument.

4.4 Have useable results for SDO's

Besides providing quality scores for a standard, the instrument should also provide the user with suggestions for adjusting the standard so a higher quality can be achieved. An instrument for determining the quality of standards should, of course, have a high quality output itself. Figure 5 shows the goals and requirement that have to be met in order to have useable results.
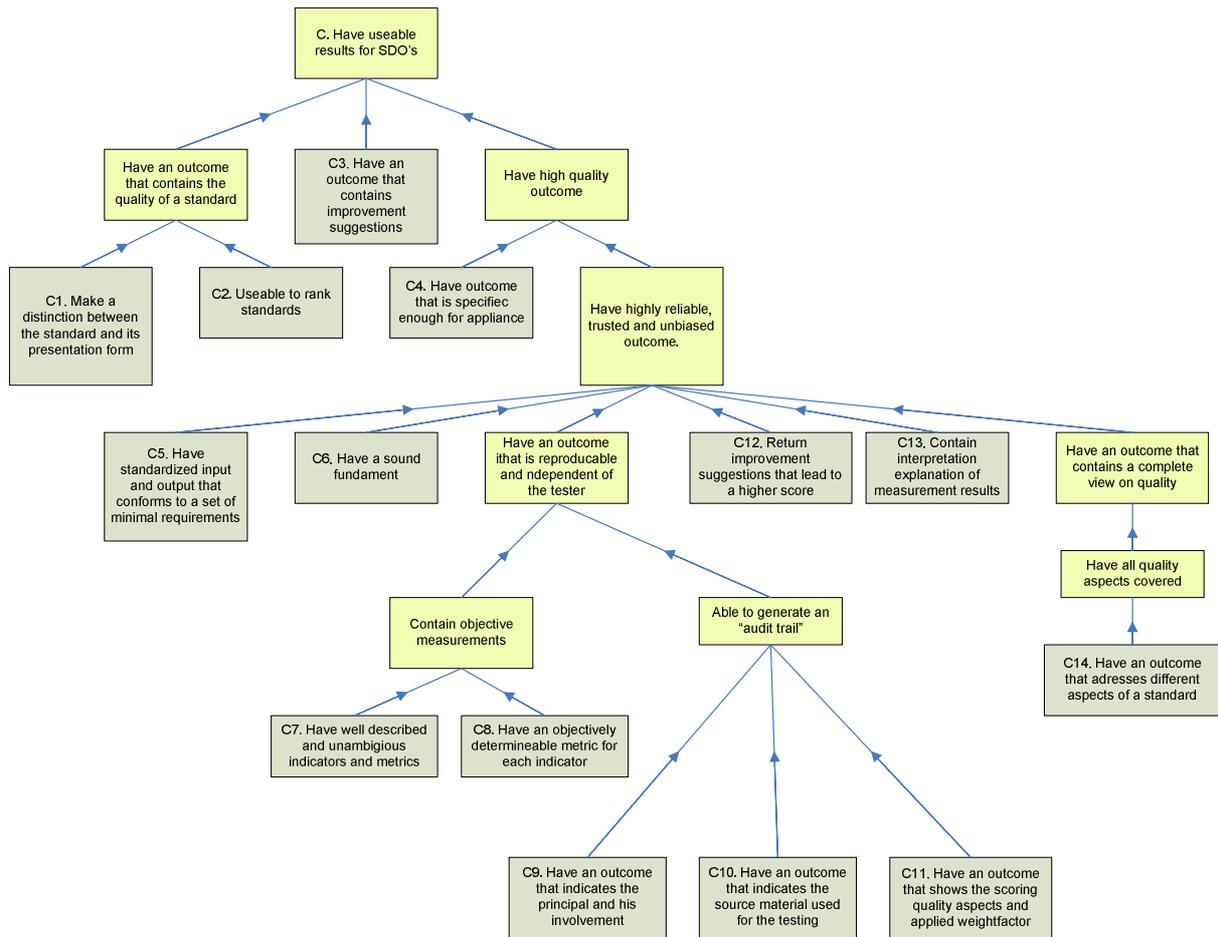
Figure 5 - Useable results for SDO's

In order for the outcome of the instrument to be of high quality, it should be reliable, trusted, and unbiased. Besides, the outcome has to be reproducible en independent of the tester. This can be achieved by generating an audit trail, and having objective measures.

Also, the instrument should enable a complete view on quality, meaning that all quality aspects can be covered.

4.5. General observations and discussion

During the workshops, we focussed on gathering requirements for the quality instrument. Nonetheless, we received several suggestions for specific quality aspects. These quality aspects will not be used in this phase of our research, but is an interesting "by catch" for usage in a later stadium.

Another important notice is that quality is situational and time-dependent. This means that quality statements may change over time. It also implies that aspects of the problem environment should be part of quality.

Another valuable contribution was the suggestion of the following requirement: The instrument should indicate the value of the standard for: 1. Investment, 2. Solution/Cost reduction, 3. Commercial (Patents). Although interesting, we think it does not support the highest goal in our goal tree. The commercial value of a standard seems irrelevant for the highest goal related to achieving interoperability. This requirement might lead to an interesting but different (complementary?) instrument for example a kind of adoption

measurement instrument that can be used by individual organisations to determine whether or not to invest the adoption or development of a standard.

Finally, people can hardly think of requirements without thinking of possible solutions. In time requirements gathering processes gain focus as they proceed, but have to end as well avoiding designing solutions during requirements engineering. In our case, the scope was set by having a short presentation about the problem domain within each workshop. We stopped the requirements engineering process after two rounds of workshops and engineering the results of the workshops.

5. Reflection on requirements

The second workshop was held with experts having a technical background, who usually are involved in technical, telecom-related standards. Although the scope of the instrument we are going to develop focuses on semantic standards, we got the impression, based on the workshops, that an instrument that focuses on technical standards might have similar requirements. This may imply that the instrument that will be developed based on these requirements might be useful for other type of standards as well.

On the other hand, we also found that all requirements engineering methods we examined assume that a product (physical item or software) is going to be produced. In our case, we have a more abstract concept "instrument", without having chosen the exact representation yet. This may have resulted in requirements that are abstract as well. One drawback of abstract requirements is that it is hard to determine whether we have a complete set of requirements. This makes it even more important to not only test whether the instrument fulfils the requirements, but also whether it presents a solution to its users.

The lightweight situated requirements engineering method worked quite well and produced useful requirements. The result is a set of structured requirements presenting a rich set of information. We did notice however that a lot of functional requirements were identified, and only very few non-functional. A possible explanation is again the abstract notion of the instrument and possible abstract requirements.

In both workshops, the experts made a distinction between a standard (consisting of a set of agreements, but quite an abstract concept), and the representation of the standard, for example a paper document. One standard (for example the GSM standard) can have multiple representation forms, for example in different languages. Both the standard (as an abstract concept) itself as well as the representation form have quality aspects. It may even occur that the standard itself has a good quality, while one of the representation forms has a poor quality. This poses an interesting problem: how does one measure an abstract concept? Also, if the quality of a standard is measured using one of the representation forms, how can one distinguish between the quality of the standard itself, and the quality of the representation?

We already concluded that the instrument might be useful for multiple types of standards. It would have been interesting to compare our results with other studies regarding requirements for quality instruments. Unfortunately, to our knowledge, very little research has been done on requirements for instruments that can be used to measure quality, which makes comparisons hard.

6. Conclusion

Using a situational method combining fragments of QFD, KAOS, and Volere requirements engineering methods, we constructed a set of requirements for a quality measurement instrument for semantic standards, and structured them in a goal tree.

The top goal "To support semantic SDO's in developing high quality standards" has been decomposed into three level-two goals, which have been further decomposed:

- usefulness for different semantic SDO's
- efficient to use
- and useable results

Overall we can conclude that the presented set of requirements do contribute to our knowledge about the desires from standardization practitioners regarding an instrument for quality measurement.

The next step would be to start developing an instrument based on the requirement as stated in this study.

7. References

Akao, Y. (Ed.). (1990). *Quality Function Deployment: Integrating Customer Requirements into Product Design*: Productivity Press.

Al-Subaie, H., & Maibaum, T. (2006). *Evaluating the effectiveness of a goal-oriented requirements engineering method.* Paper presented at the Fourth International Workshops on Comparative Evaluation in Requirements Engineering, CERE'06, St. Paul, MN.

Brinkkemper, S. (1998). Assembly techniques for method engineering. *Advanced Information Systems Engineering, 1413*, 381-400.

Cargill, C., & Bolin, S. (2007). Standardization: a failing paradigm. In S. Greenstein & V. Stango (Eds.), *Standards and Public Policy* (pp. 296-328). Cambridge: Cambridge University Press.

Coulin, C. (2006). A situational method engineering approach to requirements elicitation workshops in the software development process. *Software process improvement and practice, 11*(5), 451.

Folmer, E., Berends, W., Oude Luttighuis, P., & Van Hillegersberg, J. (2009). *Top IS research on quality of transaction standards, a structured literature review to identify a research gap.* Paper presented at the 6th International Conference on Standardization and Innovation in Information Technology, Tokyo, Japan.

Folmer, E., Oude Luttighuis, P., & Van Hillegersberg, J. (2010). Quality of semantic standards needs improvement; A survey amongst 34 semantic standards. *To be published.*

Jakobs, K. (2009). Perceived Relation between ICT Standards' Sources and their Success in the Market. In K. Jakobs (Ed.), *Information Communication Technology Standardization for E-Business Sectors; Integrating Supply and Demand Factors* (pp. 65-80). Hershey: Information Science Reference.

Robertson, S., & Robertson, J. (1999). *Mastering the Requirements Process*. Harlow: Addison-Wesley.

Van Lamsweerde, A. (2001). *Goal-Oriented Requirements Engineering: A Guided Tour.* Paper presented at the 5th IEEE International Symposium on Requirements Engineering, Toronto.

Van Lamsweerde, A. (2004). *Goal-Oriented Requirements: A Roundtrip from Research to Practice.* Paper presented at the 12th International Requirements Engineering Conference, Kyoto.

Appendix 1 - Set of requirements

The Volere method emphasizes the rationale and fit criterion of the requirement (Robertson & Robertson, 1999). All of the requirements found in the requirement engineering phase are described in the table below. For each requirement the following information, based on Volere, is provided:
- Number: the number of the requirement, matches the numbers used in the figures in chapter 4
- Short name: a short description of the requirement, matches the names used in the figures in chapter 4
- Description: a detailed description of the requirement
- Fit criterion: the criteria to determine whether the requirement is fulfilled
- Priority: the priority of the requirement, used for choosing between conflicting requirements and when time limitations prevent implementing all the requirements

Number	Short name	Description (what?)	Fit criterion	Priority
A1	Possible to add quality aspects	The instrument should be flexible to add new quality aspects.	The end user should be able to add aspects without dependency on the instrument designer.	Medium
A2	Possible to add new indicators	The instrument should be flexible to add new indicators for existing quality aspects.	The end user should be able to add indicators without dependency on the instrument designer.	Medium
A3	Possible to add new metrics	The instrument should be flexible to add new metrics to measure existing indicators.	The end user should be able to add metrics without dependency on the instrument designer.	Medium
A4	Possible to choose a metric if more than one is available	The user should have the possibility to chose a metric if more than one is available for measuring an indicator. Depending on the preferences of the user, he could select a rigid but time-consuming metric or a less rigid but ease to determine metric.	The instrument should present the user a choice if more than one metric is available.	High
A5	Possible to personalize the weighing of individual quality aspects	The overall quality of a standard is determined by combining all the individual quality criteria. However, different users may have different opinions to which quality criteria are important. The instrument should allow users to personalize the weighing for all the individual criteria.	The user must be able to personalize the weighing factors himself, without the help of the designer of the instrument.	High
A6	Possible to choose an indicator if more than one is available	The user should have the possibility to choose an indicator if more than one is available for a given quality attribute. Depending on the preferences of the user, he could select a better but time-consuming indicator, or a lesser but easy to determine indicator.	The instrument should present the user a choice if more than one indicator is available.	High
A7	Useable for different types of semantic standards	Semantic standard may vary in content en format, but the instrument should be useable for all semantic standards.	The instrument should be useable for all standards presented on www.remlof.eu.	High
B1	Have transparent outcome	The outcome of the instrument should provide insight on how the results are determined. To do this, the instrument must relate quality aspects to attributes of the standard.	The outcome of the instrument should contain all applied metrics and weighing factors. For all metrics that require human interpretation, an explanation must be provided.	Medium
B2	Have an outcome summary that fits on one page (but is more than a single rate)	In order to be useable by the user of the instrument, the outcome summary should contain no more than one page.	Summary of outcome maximum of one page A4 size using font size 10.	High
B3	Contain no more than 7	Contain no more than 7 tests	The number of tests	low

	tests	(compare with car testing: city, snow, dessert, test track, long ride, etc).	performed by the instrument should be no more than 7 for one single standard.	
B4	Have standard templates for weighing factors	The instrument should have "standard" templates for users who do not wish to tailor the weighing factors to their own need.	It must be able to use the instrument without spending any time on determining weighing factors.	Medium
B5	Have automated measurements when possible (by machine reading)	To make the instrument as easy as possible to use, the instrument should perform automated measurements when possible.	All metrics that can be determined by machine reading should require no human interaction.	high
B6	Contain clear guidelines on how to use	The instrument should be easy to use, and therefore contain clear guidelines.	A guidelines document should be available.	high
B7	Instrumental, a "Tool"	The instrument should be practical useful by being implemented as tool.	All parts of the instrument should be covered by physical or software products.	High
B8	Useable to identify blank spots in work in progress	The instrument should not only be useable for determining the quality of finished standards, but also give improvement suggestions when used on work in progress	In the results the blank spots are presented.	low
B9	Facilitate testers	The instrument should be useable by testers that are implementing a (draft) standard.	The instrument should make it possible to test parts of the standard.	low
B10	Measure complete standards as well as individual parts	The user should be able to use the instrument not only on a complete standard, but also on parts of the standard.	When combining test of individual parts of a standard, 90% of the standards should have less than 10% deviation from the testing of the complete standard.	medium
B11	Support scenario assertions	The instrument should support scenario assertions, "what if...".	It should be possible to use at least two scenarios (minimum and maximum).	low
B12	Measure one individual standard	The instrument should take one individual standard (or a part) as input.	The instrument should never require a second standard to be used.	Medium
C1	Make a distinction between the standard and its presentation form	One can distinguish a standard (set of agreements) and the presentation (usually a document). Some standards are presented in different forms (e.g. different languages). The instrument should give insight in whether a quality measurement is done on the representation (document) or the standard.	For each measurement and attribute it must be clear whether the standard or the presentation was subject of investigation.	Low
C2	Useable to rank standards	The outcome of the instrument should be useable to compare and rank two or more standards (this standard is better than that standard).	The outcome of the instrument should also include one score on the scale of 1 to 10 (latter is better).	Medium
C3	Have an outcome that contains improvement suggestions	The instrument should not only return the quality of the standard, but also suggestions to improve the standard.	For each standard that has a quality score less then 10, the instrument should return at least one improvement suggestion.	Medium
C4	Have outcome that is specific enough for appliance	The improvement suggestions should be specific enough for the user to apply on the standard, without having to consult an experienced user.	When improvement suggestions are processed by 5 independent users, 4 out of 5 should make the same changes to the standard.	Medium
C5	Have standardized input and output that conforms to a set of minimal requirements	In order to process standards in a comparable way, the input should conform to a minimum set of requirements. When conforming to the minimum set of requirements, the output should also conform to a set of requirements.	For 5 standards that comply to the input requirements, at least 4 of the outcomes comply to the minimum set of requirements.	Medium
C6	Have a sound fundament	The result of the instrument should not be easy to devaluate, therefore the instrument should have a sound, theoretical fundament.	The model behind the instrument should be supported by at least one scientific theory.	High
C7	Have well described and unambiguous indicators and metrics	The indicators and metrics shall be well described and unambiguous.	When asking users to explain the indicators and metrics, 4 out of 5 users give the same explanation for at least 90 percent of the indicators and	High

			metrics.	
C8	Have an objectively determinable metric for each indicator	Each indicator has at least one metric that can be determined objectively.	When 5 independent users test a standard, at least 90 percent of the metrics shall score within a 10 percent margin.	High
C9	Have an outcome that indicates the principal and his involvement	In order to determine the objectivity of the outcome of the instrument, the principal and his role in the standard (development process) should be known.	The outcome should always include the principal and his involvement.	Medium
C10	Have an outcome that indicates the source material used for the testing	The outcome of the instrument should always indicate all the source material (documents) that is used for the testing.	The outcome should always indicate the source material.	Medium
C11	Have an outcome that shows the scoring quality aspects and applied weight factor	The outcome of the instrument should provide insight on how the results are determined. To do this, the instrument must relate quality aspects to attributes of the standard.	The outcome of the instrument should contain all applied metrics and weighing factors. For all metrics that require human interpretation, an explanation must be provided. (Similar to B1)	Medium
C12	Return improvement suggestions that lead to a higher score	After processing the improvement suggestions given by the instrument, testing the standard should lead to a higher score.	When using the instrument on a standard that is not yet finished, a second test after applying the improvement suggestions should return a higher score.	High
C13	Contain interpretation explanation of measurement results	The outcome of the instrument should be easy to interpret, and therefore contain an explanation of the results.	Each of the score of an quality attribute should contain an explanation.	Medium
C14	Have an outcome that addresses different aspects of a standard	In order to give a complete view on the quality of a standard, all quality aspects that are important to the user should be measured.	Each attribute that has a weighing factor that is larger than 0, should be assessed during the testing.	Medium