Design Steps for developing Software Measurement Standard Etalons for ISO 19761 (COSMIC-FFP)

ADEL KHELIFI
Software Engineering Department
ALHOSN University
Abu Dhabi, P.O. BOX 38772
UAE
a.khelifi@alhosnu.ae

ALAIN ABRAN
Software Engineering Department
École de technologie supérieure
1100, Notre-Dame Ouest Montréal, Qc
CANADA
Alain.Abran@etsmtl.ca

Abstract: - Material measurement standard etalons are widely recognized as critical for accurate measurements in any field. The absence of standard etalons in software measurement is having a negative impact on software engineers when they come to use measurement results in decision-making. To verify measurement results and ensure unambiguous comparability across contexts, researchers in software measurement should design standard etalons and incorporate them into the design of every measure proposed. Since the design process for establishing standard etalons for software measures has not yet been investigated, this paper tackles this issue and illustrates the application of this process using ISO 19761: COSMIC-FFP.

Key-Words: Metrics, Software Measurement, ISO 19761, COSMIC-FFP, Standard etalon.

1 Introduction

Measurement in general is a mature science with a long tradition. In disciplines such as physics, chemistry and biology, it is a basic part of daily activities.

Measurement standards are designed to make life easier: for example, a liter is a well-known quantity around the world, and has exactly the same value in all countries. Similarly, from east to west and from north to south, the meter is the standard for length measurement: it is applied similarly everywhere and also has a single value.

According to the International Vocabulary of Basic and General Terms in Metrology [1] a standard etalon is “a material measure, measuring instrument, reference material or measuring system intended to define, realize, conserve or reproduce a unit or one or more values of a quantity to serve as a reference.” Consequently, it becomes relevant to develop, for both measurers and users of measurement results, a system of references made up of software measurement standards. Measurement standards are essential elements for an adequate metrological structure, in that they provide software engineers with a common reference and give them greater confidence in the measurement process. Indeed, standards facilitate the realization of measurement results on common bases.

It is difficult to develop measurement standard etalons. They are created through an iterative process in which each iteration represents an improvement over the previous ones, in terms of both accuracy and stability. Moreover, each iteration may span years, if not decades.

While it is difficult to determine the effect of measurements on software quality, it is clear that using standards of measurement would provide software measurers, developers and managers with much better indicators of that quality, as well as more time to react, and could reduce the number and seriousness of software failures. In the information technology domain, and more specifically in software engineering, concepts of units and etalons have seldom been used, and this is a symptom of the immaturity of the software measures themselves. Consequently, the field of software measurement is not yet mature enough to be recognized as having value in the daily practice of software development, nor for the purchase or sale of software products and packages.

Up to now, some characteristics of software have made it challenging to measure (see Figure 1):

1. It is an intangible product, and some doubt that metrology concepts are applicable.
2. It is an atypical product when compared to other industrial products, in that it varies greatly in terms of size, complexity, design techniques, test methods, applicability, etc.
3. There is little consensus on specific measures of software attributes, as illustrated by the scarcity of international standard measures for software attributes, such as software complexity and quality.

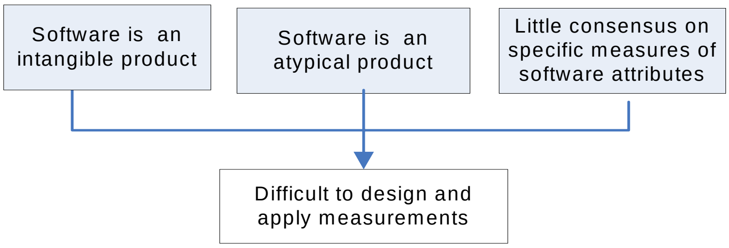
Software is an intangible product
Software is an atypical product
Little consensus on specific measures of software attributes
Difficult to design and apply measurements

Fig. 1. Challenges in the design of software measures

Because of these challenges, some have claimed that software "metrics" are somewhat unique, and, as such, cannot be constrained to meet all the metrological properties as defined in the ISO document on metrology [1]. However, the fact that there is currently no standard etalon for software does not imply that ones cannot be created; there is however, a lack of documented attempts to do so, and the lack of a methodology for doing so for software. In this paper, we postulate that it is feasible to create a standard etalon for software and that a methodology for doing so could be designed.

The focus of this paper is the proposal of a design procedure for developing a standard etalon for a software Functional Size Measurement (FSM).

The structure of the paper is as follows: Section 2 presents related work in the design of measurement standards in general, and in FSM in particular. Section 3 presents a proposal for a design methodology for a software measurement standard etalon. Section 4 presents its application on ISO 19761 – COSMIC-FFP. Section 5 presents a discussion and identifies further research issues.

2 Related work in the design of a measurement standard-etalon

2.1 Primary reference material, calibration and testing

A measure is first defined in terms of its objectives, a meta-model of the entity to be measured and the characteristics of the attribute to be measured. This definition is then realized by means of a measurement unit and a corresponding scale (e.g. the assignment of numerical rules) [2]. Next, to ensure that measurements across a community are performed in a consistent manner, a base line is established as a primary reference.

This measure can be compared with the primary measurement reference by means of calibration and testing [3]. Calibration determines the performance characteristics of an instrument or the reference material. There are three main reasons for calibrating an instrument:

1. To ensure that the instrument readings are consistent with other measurements.
2. To determine the accuracy of the instrument readings.
3. To establish the reliability of the instrument, i.e. that it can be trusted.

Reference procedures can be defined as measurement or analysis procedures which are thoroughly characterized and proven to be under control, and intended for the quality assessment of other procedures for comparable tasks, the characterization of reference materials, including reference objects, or the determination of reference values.

2.2 Design issues for the measurement of the software concept-entity

In the software engineering literature, measurement concepts are often defined in vague ways. For example, the term "metric" has several definitions [2, 4] and the designers of software metrics have not yet embedded in their design the full set of measurement concepts that is embedded, and widely accepted, in the traditional field of metrology used extensively in the engineering disciplines. It has also been recognized by authors who have discussed frameworks for metrics validation that such frameworks are still incomplete [5, 6], with little theoretical basis and a lack of reference to metrology concepts and criteria. For instance, it has been observed that, in software engineering, most measurement proposals do not refer to any references (primary references or others), do not suggest any measuring instrument and do not design or adopt any measurement standard [7].

2.3 Software Functional Size Measures (FSM)

For illustrative purposes, a single type of software measures has been selected, that is, functional size. The key reason for this selection is that, of the numerous types of measures proposed for software, functional size measures (FSM) are currently the only ones to have developed a broad enough consensus to gain widespread recognition as international software measurement method standards..

FSM is "the approach to quantifying software in terms of the functionality it delivers to its users independently of the technical and quality aspects of its delivery. It provides a method of normalizing measures of productivity, speed of delivery, quality, etc. by providing a common measure of what is delivered which can be used to calculate unit values” [8].

The Functional User Requirements represent the user practices and procedures that the software must perform to fulfill the user’s needs. They exclude Quality Requirements and any Technical Requirements” [9].

ISO has developed a set of meta-standards with respect to FSM, that is, its ISO 14143 series, parts 1 to 6. In the specific domain of software FSM, four methods have been recognized as ISO international standards. Of these, only ISO 19761: COSMIC-FFP [10] specifically specifies and documents the concept of a size unit, and none of the four ISO-recognized FSM methods explicitly addresses the concept of a standard etalon.

In practice, the application of software functional measures requires knowledge in the specific software measurement method being used and sufficient experience in the interpretation of software artifacts. For instance, in the measurement process with the COSMIC-FFP method, the measurer must determine the following, from the available artifacts: software layers to be measured, software boundary, users, triggering events, functional processes, data groups and data movements. Should the documentation be complete and accurate, these measurement steps are easy. Unfortunately, in practice, the documentation is often incomplete, and, to measure software, the measurer has to supplement the information provided on some requirements which is either incomplete or ambiguous.

The availability of a standard etalon for FSM would help improve the quality of FSM results on a practical level. Using a standard etalon can, therefore, help reduce the time spent in addressing inconsistency issues in measurement results.

2.4 Related work in FSM

2.4.1 Use of case studies as reference material

Up to now, individual ISO-recognized FSM communities have mostly developed case studies as reference material for training purposes, and these are very specific in terms of teaching some peculiarities of each FSM method; however, they are not generic enough to be used as reference material for calibration and testing purposes.

These case studies suffer from a number of limitations:

- there is no normalized input to their design process;
- they are limited in scope;

they have been drafted based on the judgments of experts within their own communities;

- they most often address only a limited number of measurement rules, sometimes in peculiar contexts.
- they cannot be used as generic reference material.

2.4.2 ISO work

The ISO has indirectly recognized the need for reference material through its provision of reference input material for measurement: indeed, ISO 14143-4 provides a set of Reference User Requirements (RURs) which were put together to provide FSM communities with material that could be used for convertibility studies across specific measurement methods. Such reference material could also be used to test some of the metrological properties of a specific measurement method, such as the accuracy, repeatability and reproducibility criteria quoted in ISO TR 14143-3.

However, ISO TR 14143-4 suffers from a number of important limitations: In its current state, it cannot be used to assess an FSM method against some standard reference points to determine whether or not it yields expected results in a given situation: in this standard, all the sets of RURs are described in a non-standardized textual format. There is, therefore, a great variation in the description of these RURs within a given set, and, of course, across sets.

In FSM, the process of software functional size measurement relies, generally, on its functional documentation. It has been illustrated in Nagano et al. [11] that the quality of the documentation has an impact on both the quality of the measurement results and on the effort required to carry out the measurements. For instance [12] has noted that the software documentation is often either incomplete or obsolete, and even sometimes erroneous. Moreover, this issue has not been addressed in ISO 14143-4, which leads to similar difficulties in measurement practice.

2.4.3 Related work on COSMIC-FFP

The topic of a standard etalon for ISO 19761 - COSMIC-FFP was initially discussed in [13] and initial drafts were published in [14]. A limitation of this pioneering work is that it is an individual effort and does not benefit from international recognition or worldwide diffusion. Official international recognition of a standard etalon for software measurement would be of practical interest to both industry and researchers. The work reported next builds on that in [14] and extends it to any FSM, and, by extension, potentially to any software size measure.

3 A design methodology for an FSM measurement standard etalon

The challenge is how to design a standard etalon for software which is not a material product. The generic process described below is based on the lessons learned from the preparation of case studies for training purposes and from work done to explore the design of an initial draft version of etalons for the COSMIC-FFP method, as well from the work reported in [14].

The design methodology proposed for developing a software measurement standard includes the following seven steps – see Figure 2.

1. Analysis and selection of candidate textual description of Functional User Requirements (FUR); the input is the literature survey of previous work on the design of a specific measurement method and available descriptions of FUR. However, these sets of FUR are often available in non-standardized textual format.

2. Identification and selection of quality criteria for the input to the measurement process. For FSM, the inputs are usually expressed in the form of textual descriptions of requirements, and related quality criteria are defined, for instance, in the IEEE standards on Specifications Requirements – IEEE 830. These quality criteria then become inputs to step 5.

3. Quality improvement of the set of FUR by transforming of the selected set of textual FURs into the selected specification language, and, in parallel, analysis of the quality of the requirements and correction of requirements defects (for instance, to remove ambiguities and inconsistencies in the requirements). The output of this step is then the FURs described in the selected notation specification language and which meet the specified quality criteria.

4. Selection or design of a template for presenting the measurement process and measurement results.

5. Initial measurement of the requirements documented in the adopted specification notation by an experienced measurer to produce an initial draft of measurement results using the adopted output format for the standard etalon.

6. Selection of a group of experts to review the initial measurement results; ideally, these measurement experts should be internationally recognized by industry for their specific FSM expertise; of course, it would add credibility if these experts were also active participants in the ISO standardization program on FSM.

7. Review by expert measurers of the initial measurement results and correction of either the inputs (the requirements themselves if they were incomplete or ambiguous) or of the outputs (the measurement results).

4 The design methodology for a COSMIC-FFP measurement standard etalon

This design methodology for developing an FSM standard etalon is a generalization of the steps carried out in [14]. Of course, the modeling of these steps has been further refined. Its specific instantiation for COSMIC-FFP is documented next.

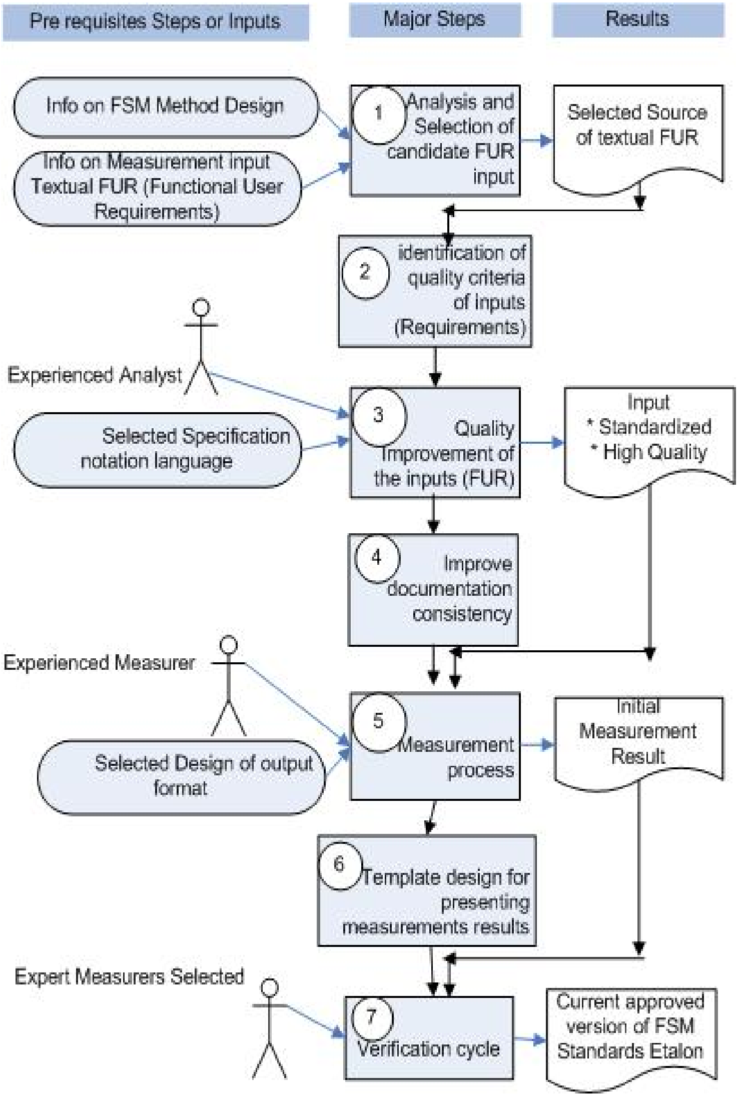
Pre requisites Steps or Inputs
Major Steps
Results
Info on FSM Method Design
Info on Measurement input Textual FUR (Functional User Requirements)
1 Analysis and Selection of candidate FUR input
Selected Source of textual FUR
2 identification of quality criteria of inputs (Requirements)
Experienced Analyst
Selected Specification notation language
3 Quality Improvement of the inputs (FUR)
Input * Standardized * High Quality
4 Improve documentation consistency
Experienced Measurer
Selected Design of output format
5 Measurement process
Initial Measurement Result
6 Template design for presenting measurements results
Expert Measurers Selected
7 Verification cycle
Current approved version of FSM Standards Etalon

Fig. 2. A design methodology to develop a software measurement standard

1. This step includes the prerequisites to beginning the process of designing a standard etalon for COSMIC-FFP. In this specific instance, it consists of the output of the literature survey of previous work on the design lessons learned from COSMIC-

FFP case studies, as well as on the identification of a set of candidate inputs for measurements. In this specific instance, the ISO work on FSM was selected (that is, ISO TR 14143-4 2000 − Reference User Requirements (RURs) [15]), since it contains an inventory of textual descriptions of requirements collected for measurement purposes.

2. Since the input to this step contains multiple sets of requirements, one specific set was selected as the basis for the work reported here, which was RUR B9 − Valve Control System (from ISO 14143-4).

3. In ISO TR 14143-4, all the sets of RURs are described in a non-standardized textual format. There is, therefore, great variation in the description of these RURs within this specific B9 set. This is typical of most inputs for the measurement of the functional size of software, in particular when the measurements are taken early in the software life cycle. As a result, it is necessary to verify the quality and completeness of these requirements. The quality criteria selected as prerequisites were selected from the IEEE standard on software requirements, that is, IEEE 830.

4. In this step, a specification language is selected as an input, and the selected set of textual FURs is transformed into a specification language. To improve the consistency of the documentation to be used as input to the FSM, the decision was made to adopt the UML notation for this research, such as use cases and sequence diagrams for the software to be measured. The UML Use Case diagram is a tool for representing the entire functionality of a system; a sequence diagram is a structured representation of software behavior as a series of sequential steps over time. Developing such diagrams can improve the comprehension of software functions and provide the measurer with more consistent and precise documentation as input to his measurement process. This allows the measurer to have his measurement inputs documented in a consistent manner, which in turn allows him greater transparency in the intermediate steps of the measuring process and more repeatable results.

5. An analyst with expertise in UML notation carried out this step, which consisted of analyzing the textual description of the requirements and their transformation into UML notation, and, within this process, the correction of defects (for instance, to remove ambiguities and inconsistencies in the requirements).

6. The next prerequisite for a major step is the selection or design of a template for presenting the measurement process and measurement results: since there had already been documented case studies for COSMIC-FFP, these were reviewed and tailored for the purpose of documenting the intermediate steps of the measurement process, as well for the outcome in terms of measurement results.

7. The initial measurement was performed, by an experienced measurer, of the requirements documented in the adopted specification notation to produce an initial draft of measurement results.

8. In this step, a group of experts was selected to review the initial measurement results; ideally, these measurement experts should be internationally recognized by industry for their specific FSM expertise; of course, it would add credibility if these experts were also active participants in the ISO standardization program on FSM. Ideally, the design of standards is an activity which must be undertaken at the international level by groups of experts from several countries in order to obtain a broad consensus. The ISO organization represents the most adequate framework for this type of activity. The selection of experts for the draft COSMIC-FFP standard etalon was made through the Software Engineering Research Laboratory contacts. It included international experts in software measurement within the COSMIC group, a group of international volunteer experts in software measurement. Some of these experts were also members of WG12, an ISO working group specializing in software FSM. However, this work was not done in an official context, and the credibility of the measurement outcomes is derived from their individual expertise, and not from an official international process recognized by national institutions.

9. This step constituted a review of the initial measurement results and correction, even of the requirements themselves if they were incomplete or ambiguous. The final output was then the currently approved version of a standard etalon for COSMIC-FFP. It is to be noted that, for traceability purposes, the output in software measurement must include both the inputs and the outputs of the measurement process for establishing the standard etalon.

In summary, the end-result of the design of a standard etalon for software FSM with the COSMIC-FFP method consists of a detailed report using a template documenting both the inputs and the outputs of the measurement process on a set of software FURs.

The iterative verification process is highly relevant at the international level; in practice, this verification process will go through an iterative cycle.

5 Discussion

The development of a standard etalon for software measurement could have a far-reaching impact: for instance, many standard etalons, such as the meter standard or kilogram standard etalon, contribute in the management of many aspects of our daily life.

In this paper, we presented a methodology for developing a standard etalon for software measurement and illustrated it using ISO 19761 – COSMIC-FFP. The application of the COSMIC-FFP measurement method by experts in software FURs generates the measurement results. It is the consensus among measurement result experts that defines the quality of a standard etalon for the result, and the verification of every part of the standard etalons by recognized experts and COSMIC members provides the standard etalon with greater accuracy. The addition by measurers or software engineers of UML diagrams, use cases and sequence diagrams, and their verification by UML developers, further enhances the software functionalities by providing greater understandability, accuracy and completeness. This allows the measurers to re-analyze the measurement results and make other improvements if necessary.

Meanwhile, it is important that the software measurement community comes to appreciate that the development of a standard for the measurement of software may take many decades. For instance, it took two centuries for the definition of the meter to become established.

In conclusion, we, as designers of software measures, must learn how to build standards for software and accept that, as for any other standard etalons in physical sciences, initial software standard etalons will require improvements over time to provide the software engineering community with progressively more accurate standard etalons.

References

[1] International Vocabulary of Basic and General Terms in Metrology, International Organization for Standardization, Geneva, 1993.

[2] A. Abran, J.-P. Jacquet, "A Structured Analysis of the New ISO Standard on Functional Size Measurement-Definition of Concepts", 4th IEEE International Symposium and Forum on Software Engineering Standards, 1999, pp. 230-241.

[3] P. Howarth, “Metrology – in Short”, 2nd ed. Vol. 1. Lyngby: EUROMET, 2004. EUROMET project no. 595. 18 Sept. 2004, www.euromet.org/docs/pubs/docs/metrology_in_short_2nd_edition_may_2004.pdf?.

[4] J. Bieman, N. Fenton, D. Gustafson, A. Melton, L. Ott., “Fundamental issues in software measurement”, in Melton, A. ed. Software Measurement: Understanding Software Engineering, International Thomson Publishing (ITP), 1996, 39-74.

[5] B. Kitchenham, S.L. Pfleeger, N. Fenton, "Towards a Framework for Software Measurement Validation", IEEE Transactions on Software Engineering, vol. 21, no. 12, 1995, pp. 929 -944.

[6] A. Sellami, A. Abran, "The contribution of metrology concepts to understanding and clarifying a proposed framework for software measurement validation", 13th International Workshop on Software Measurement, Montréal (Canada), Shaker Verlag, Sept. 23-25, 2003, pp. 18-40.

[7] A. Abran, A. Sellami, W. Suryn, “Metrology, Measurement and Metrics in Software Engineering”, METRICS 2003, Sydney, Australia, September 2003, IEEE CS, Los Alamitos, 2003.

[8] UKSMA, “Functional Size Measurement Methods”,2005, www.uksma.co.uk/?action=35&what=35&type=0.

[9] ISO/IEC 14143-1: 1998. Software Engineering – Functional size measurement – Part 1: Definitions of Concepts, International Organization for Standardization – ISO, Geneva, 1998.

[10] ISO/IEC 19761:2003, Software Engineering – COSMIC-FFP – A Functional Size Measurement Method, International Organization for Standardization – ISO, Geneva, 2003.

[11] S.-I. Nagano, K.-I. Mase, Y. Watanabe, T. Watahiki, S. Nishiyama, “Validation of Application Results of COSMIC-FFP to Switching Systems,” ACOSM2001, 17 Nov., 2001, 6 Jan., 2004, www.gelog.etsmtl.ca/publications/pdf/343.pdf.

[12] J.-M. Desharnais, A. Abran, “Applying a Functional Measurement Method: Cognitive Issues,” 11th International Workshop on Software Measurement (Montréal, Québec, 2001), Shaker Verlag, 2001, pp. 26-50.

[13] A. Khelifi, A. Abran, L. Buglione: “A System of References for Software Measurements with ISO 19761 (COSMIC-FFP),” 14th International Workshop on Software Measurement - IWSM 2004, Berlin, Germany, November 3-5, 2004

[14] A. Khelifi, “A Set of References for Software Measurement with ISO 19761 (COSMIC-FFP): an Exploratory Study”, PhD thesis, École de Technologie Supérieure, Montréal, 2005, p. 495.

[15] ISO/IEC TR 14143-4: 2000 Information Technology – Software Measurement – Functional Size Measurement – Part 4: Reference Model, International Organization for Standardization – ISO, Geneva, 2000.