Software Metrics & Software Metrology

Alain Abran

Chapter 14
Design of Standard Etalons: The Next Frontier in Software Measurement

Agenda

This chapter covers:

- An introduction to the concepts of measurement standard etalon
- Calibration and testing: reference material and uncertainty
- Related work in software functional size measurement
- A (draft) methodology to design a software measurement standard etalon

Agenda

This chapter covers:

- **An introduction to the concepts of measurement standard etalon**
- Related work in software functional size measurement
- A (draft) methodology to design a software measurement standard etalon

Introduction - Measurement standard etalon

- Measurement standards are designed to make life easier. In disciplines such as physics, chemistry, and biology, the reference to a measurement standard etalon is critical to ensure the correctness and consistency of measurement results across contexts and countries.

- According to the International Vocabulary of Basic and General Terms in Metrology a **measurement standard etalon** is:

  - **'The realization of the definition of a given quantity, with stated quantity value and associated measurement uncertainty, used as a reference'** [VIM 2007]**.**

Introduction - Measurement standard etalon

- The [VIM 2007] provides a number of examples which illustrates a number of ways to implement a measurement standard etalon. It is noted in particular in the VIM the following:
  - A “realization of the definition of a given quantity” can be provided by a **measuring system**, a **material measure**, or a reference material.
  - A measurement standard is frequently used as a reference in establishing **measured quantity values** and associated measurement uncertainties for other quantities of the same **kind**, thereby establishing **metrological traceability** through **calibration** of other measurement standards, **measuring instruments**, or measuring systems.

Introduction - Measurement standard etalon

- In sciences, as well as in many economic activities, it is most relevant to develop, for both measurers and users of measurement results, a system of references made up of software measurement standards.

- Measurement standard etalons are essential elements of an adequate metrological structure:
  - they provide measurement users with a common reference and give them greater confidence in the measurement process.
  - Indeed, standards facilitate the realization of measurement results on common bases.

Introduction - Measurement standard etalon

- Using a standard etalon can improve competitiveness by reducing the cost of both manufacturing and market transactions:
  - A producer does not need to reinvent the specifications or performance criteria incorporated in the standard, and can therefore concentrate resources elsewhere.

- Furthermore, a standard etalon can contribute to the propagation of innovations, and consequently enhance the economic benefit to be derived from them.

Introduction - Measurement standard etalon

- 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.
  - **Each iteration may span years, if not decades**

- Ideally, the design of standard etalons 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

Agenda

This chapter covers:

- An introduction to the concepts of measurement standard etalon
- **Calibration and testing: reference material and uncertainty**
- Related work in software functional size measurement
- A (draft) methodology to design a software measurement standard etalon

Calibration and testing: reference material and uncertainty

- A measurement method 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 assignment of numerical rules.
- Next, to ensure that measurements across a community are performed in a consistent manner, a baseline should be established as a primary reference.

Calibration and testing: reference material and uncertainty

- Any specific measurement results using such a measurement method (or typically a measuring instrument in traditional science) can be compared with the primary measurement reference by means of calibration and testing:
  - Calibration determines the performance characteristics of a measurement instrument or the reference material.
- There are 3 main reasons for calibrating a measuring instrument:
  1. to ensure that the instrument readings are consistent with other measurements,
  2. to determine the accuracy of the instrument readings, and
  3. to establish the reliability of the instrument, i.e. that it can be trusted.

Calibration and testing: reference material and uncertainty

- Reference procedures can be defined as testing, 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.

Calibration and testing: reference material and uncertainty

- The uncertainty of the results of a reference procedure must be adequately estimated and appropriate for the intended use.
- According to this definition, reference procedures can be used to:
  - verify other measurement or test procedures used for a similar task, and determine the level of uncertainty associated with them,
  - determine reference values for the properties of materials which can be compiled in handbooks or databases, or
  - determine reference values which are embodied in reference material or a reference object.
- Uncertainty is a quantitative measure of the quality of a measurement result enabling the measurement results to be compared with other results, references, specifications, or standards.

Agenda

This chapter covers:

- An introduction to the concepts of measurement standard etalon
- Calibration and testing: reference material and uncertainty
- **Related work in software functional size measurement**
- A (draft) methodology to design a software measurement standard etalon

Related Work in Software Measurement

- **Related issues in software measurement**
  - 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 software measurement.
  - Up to now, some characteristics of software have made it challenging to measure:
    - 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.
    - There is little consensus on specific measures of software attributes, as illustrated by the scarcity of international measurement standards for those attributes, such as software complexity and quality.

Related Work in Software Measurement

- Because of these challenges, some have claimed that software “metrics” are to some extent unique, and, as such, cannot be constrained to meet all the metrological properties as defined in the ISO document on metrology.
  - However, the fact that there is currently no standard etalon for software does not imply that one cannot be created.
  - There is, indeed, a lack of documented attempts to do so, and a lack of a methodology for doing so for software.

- In this chapter, it is postulated that it is feasible to create a standard etalon for software and that a methodology for doing so can be designed.

Related Work in Software Measurement

- If measurement reference material in the form of standard etalons were available to software practitioners, it could:
  - be used as a common baseline for measurement,
  - offer a point of reference for software measurers to verify their measurement results and their ability to measure the same reference material, and
  - allow measurers to use the related reference concept, and thus to speak at the same level.
- Material measurement standard etalons are widely recognized as critical for accurate and repeatable measurements in any field.
  - However, none of the designers of software measures has yet included any of them in their design.

Related Work in Software Measurement

- To verify measurement results and ensure unambiguous comparability across contexts and measurers, researchers in software measurement should design standard etalons and incorporate them into the design of every measure proposed.

> **Lack of standard etalons in software measurement**
> In software engineering, most measurement proposals:
> – do not refer to any references (primary or other),
> – do not suggest any measuring instrument, and
> – do not design or adopt any measurement standard etalon.

- The absence of standard etalons in software measurement is most probably having a negative impact on software developers and managers when they come to use measurement results in decision making.

Related Work in Software Measurement

- While it is difficult to determine the effect of measurements on software quality, for instance, 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.

- Since the design process for establishing standard etalons for software measures has not yet been investigated, this chapter tackles that issue and illustrates the application of such a process for one of the international standards for software functional size measurement, ISO 19761: COSMIC.

Related Work in Software Measurement

- The focus of this chapter is on a proposal of a design procedure for developing a standard etalon for a software Functional Size Measurement (FSM) [Khelifi 2004, 2005].

  - The motivation for proposing an initial software measurement standard for functional size is the need for a traceable and widely recognizable standard etalon in software measurement, as exists for measurement in other human endeavors.

Related Work in Software Measurement

- **ISO pioneered work on FSM reference material**

  Of the 100+ proposals to measure various attributes of software, only in functional size measurement has there emerged a broad enough consensus to lead to software measurement methods as international standards.

ISO 14143 Meta-standard for Functional Size Measurement Methods

Part 1: Definition of Concepts

Part 2: Conformity Evaluation of Software Size Measurement Methods

Part 3: Verification of Functional Size Measurement Methods

Part 4: Reference Model

Part 5: Determination of Functional Domains for use with Functional Size Measurement

Part 6: Guide for Use of ISO 14143 Series and Related International Standards

Parts 1 and 2 are recognized International Standards.

Parts 3 to 6 are Technical Reports only, published by the ISO.

Related ISO FSM International Standards

– ISO 19761: COSMIC

– ISO 20926: Unadjusted Function Points

– ISO 20968: Mk II

– ISO 24570: NESMA

– ISO 29881: FISMA

Related Work in Software Measurement

- Of these 5 specific FSM methods recognized by the ISO:
  - none explicitly addresses the concept of a standard etalon;
  - only COSMIC specifically specifies and documents the concept of a measurement unit for size.
- The ISO community has directly recognized the need for reference material on FSM:
  - ISO TR 14143-4 provides a set of Reference User Requirements (RUR), which were put together to provide FSM communities with material intended to be used for convertibility studies across specific measurement methods.
  - Such reference material in ISO TR 14143-4 was intended to 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.

Related Work in Software Measurement

- In practice, however, ISO TR 14143-4 suffers from a number of important limitations, and 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 14143-4, all the sets of Reference User Requirements (RURs) are <u>described in a non standardized textual format</u>.
    - There is, therefore, great variation in the description of these RURs within a given set, and, of course, across sets.
    - In particular, none has been reviewed for quality control.

Related Work in Software Measurement

- Therefore, trial use of reference material in 14143-4 by both experts and beginners has highlighted a number of ambiguities and a lack of completeness, leading to different interpretations of these ambiguous functional requirements, and, of course, to various measurement results:
  - This, of course, defeats the purpose of the publication of this document, and the ISO should return it to the drawing board.

Observations from trial usages of ISO 14143-4

- Distinct measurers produce different measurement results when they need to make assumptions, and these will often vary from one person to another based, in particular, on their work experience.
- Of course, in the presence of incomplete or ambiguous requirements, distinct developers would also produce distinct software designs and related software implementations.

Related Work in Software Measurement

- The measurement of software functional size generally relies on the functional documentation of the software to be measured:
  - It has been illustrated in [Nagano 2001] that the quality of the documentation has an impact on both the quality of the measurement results and the effort required to perform the measurements.
  - Several researchers and practitioners have also noted that the software documentation is often either incomplete or obsolete, and even sometimes erroneous.

Related Work in Software Measurement

- This issue has not been addressed in ISO 14143-4, which leads to similar difficulties in measurement practice, where the application of software functional measurement requires:
  - knowledge of 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 method, the measurer must determine the following, from the available artifacts:
  - software layers to be measured,
  - software boundary,
  - functional users,
  - triggering events,
  - functional processes,
  - data groups, and
  - data movements.

Related Work in Software Measurement

- 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 addressing inconsistency issues in measurement results.

Related Work in Software Measurement

- **Case studies as reference material in FSM**

  Up to now, the ISO-recognized FSM communities have developed case studies as reference material for training purposes.

  - These case studies are very specific in terms of teaching some of the peculiarities of individual FSM methods.
  - However, they are not generic enough to be used as reference material for calibration and testing purposes.

Related Work in Software Measurement

These case studies suffer from a number of limitations:

- there is no normalized input to their design process;
- they have been drafted based on the judgments of experts within their own communities;
- they are limited in scope;
- they most often address only a limited number of measurement rules, sometimes in peculiar contexts.
- they cannot be used as generic reference material.

Related Work in Software Measurement

- ***Related work on COSMIC***
  - The topic of a standard etalon for ISO 19761 − COSMIC was initially discussed in [Khelifi 2004,2005].
    - A limitation of this pioneering work is that it is an individual effort and does not benefit from international recognition or worldwide dissemination.
  - Official international recognition of a standard etalon for software measurement would be of practical interest to both industry and researchers.
  - The next section builds on the drafts of a standard etalon in Khelifi [2005] and extends its use to any FSM, and, by extension, potentially to any software size measurement method.

Agenda

This chapter covers:

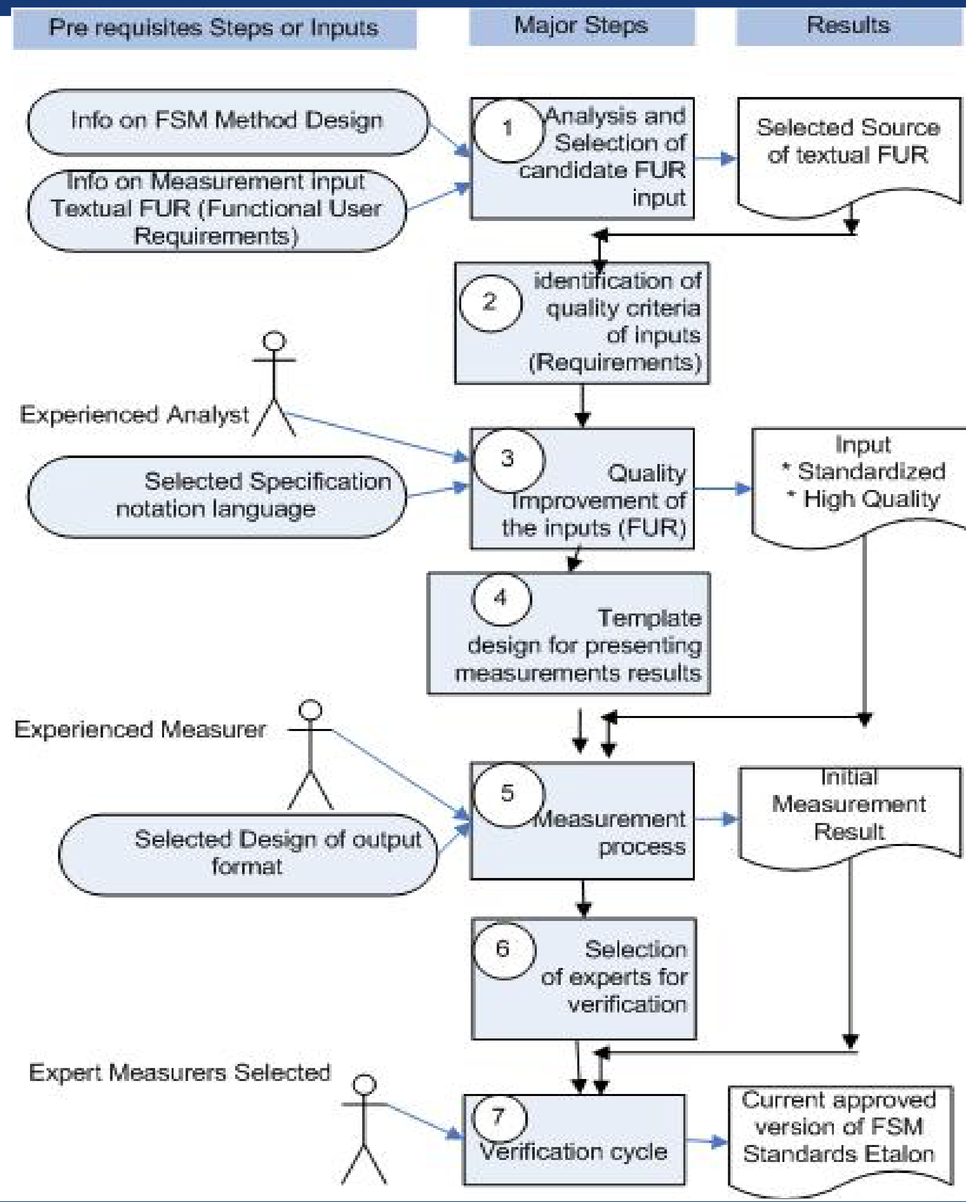
- An introduction to the concepts of measurement standard etalon
- Calibration and testing: reference material and uncertainty
- Related work in software functional size measurement
- **A (draft) methodology to design a software measurement standard etalon**

A (Draft) Methodology to Design an FSM Standard Etalon

- The challenge is how to design a standard etalon for software which is not a material product.
  - The generic process described next 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 method.
- This generic process includes the following seven steps – see next slide.
  1. Analysis and selection of textual descriptions of candidate 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.
     - It has been noted that these sets of FUR are often available in non standardized textual format.

A (Draft) Methodology to Design an FSM Standard Etalon

A Methodology to Design a Software Measurement Standard etalon



A (Draft) Methodology to Design an FSM Standard Etalon

2. Identification and selection of quality criteria for the inputs 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 3.
3. Quality improvement to the set of FUR by transforming the selected set of textual FUR into the selected specification language, and, in parallel, analysis of the quality of the requirements and correction of requirement defects (for instance, to remove ambiguities and inconsistencies in the requirements).
   - The output of this step is then FUR described in the selected notation specification language which meet the specified quality criteria.
4. Selection or design of an etalon template for presenting the measurement process and measurement results.
   - This step can be skip if a template already exists.

A (Draft) Methodology to Design an FSM Standard Etalon

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

Summary

This chapter has covered:

- An introduction to the concepts of measurement standard etalon
- Calibration and testing: reference material and uncertainty
- Related work in software functional size measurement
- A (draft) methodology to design a software measurement standard etalon

  - See also in the 'Advanced Readings' section an example of an initial draft of a measurement standard etalon for COSMIC – ISO 19761