Towards a Sound Adoption of Measurement and Metrology in Software Engineering

Luca Santillo, MSc (Agile Metrics)

UKSMA/COSMIC
International Conference on
Software Metrics & Estimating
London, October 27th / 28th 2011

UKSMA
United Kingdom
Software Metrics
Association

COSMIC
Common Software Measurement International Consortium

Agenda

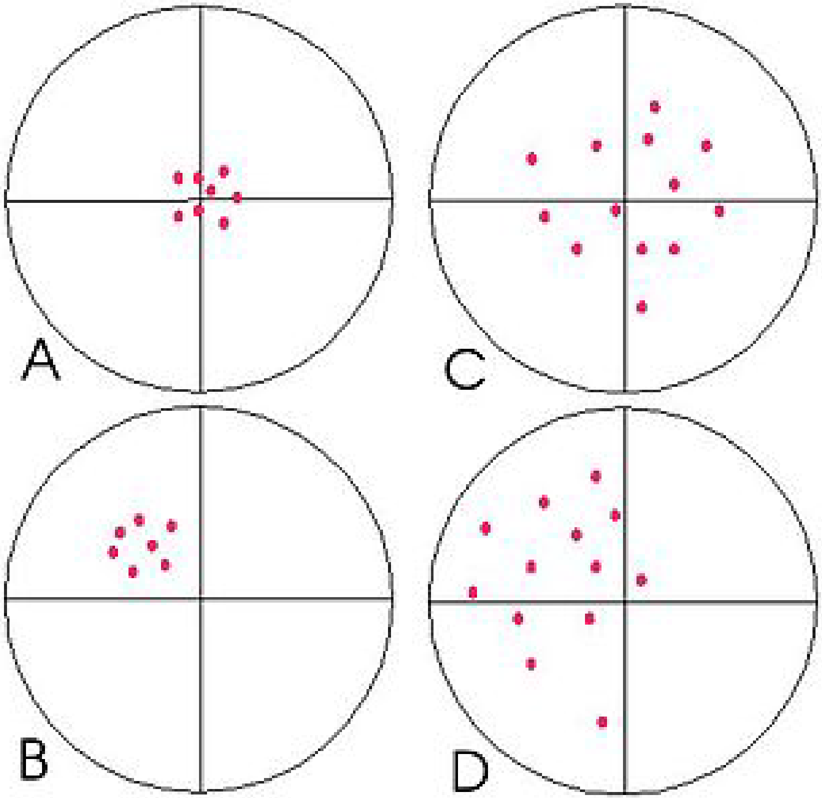
1. Why being so “accurate & precise” in SWEng?
2. International Standards on FSM
3. Metrology concepts
4. Software Measurement – State of the art
5. A topic for standardization – FSM Etalon(s)
6. Prospects ahead

Why being so “accurate & precise ” with Software Engineering & Standards?

- Accuracy and precision in hard science (eg. Physics) and applied science (eg. Engineering)
  - theoretical definitions of measurable quantities
  - adequate procedures to measure those quantities
  - long-term & continuous progress (scale of centuries)
  - uncertainty management & safety reasons
- Software measurement standards (etalons)
  - highly desirable, but not yet addressed adequately
    - to improve software estimation and to enable correct comparisons (eg. software projects benchmarking)

Accuracy vs. Precision

(quick source: Wikipedia)



A. Precise and accurate

B. Precise but not accurate

C. Accurate but imprecise

D. Not accurate nor precise

Measurement units – Real-world cases

- Boeing 767 “Gimli Glider” (1983). Ran out of fuel in mid-flight because of two mistakes in figuring the fuel supply of Air Canada's **first aircraft to use metric measurements**. Apparently, confusion both due to the simultaneous use of **metric & Imperial measures** as well as **mass & volume measures**.
- Korean Air cargo flight 6316, Shanghai-Seoul (Apr ‘99). Lost due to the crew confusing **tower instructions (in metres)** and **altimeter readings (in feet)**. 3 crew + 5 people on ground killed, 37 injured.
- NASA Mars Climate Orbiter (Sept ‘99). Crashed instead of entering orbit, due to miscommunications about the value of forces: different computer programs used **different units of measurement (newton versus pound force)**. Enormous amounts of effort, time, and money were wasted.

Reasons for methodological improvements

- Poor adherence to sound metrology procedures and requirements
  → frequent inconsistencies in real-world applications
  → limited or poor adoption of software measures in the industry
- Lack of "etalon(s)" ("measurement standards") definition
  - (so far, only an exploratory study proposed)
- Poor knowledge of uncertainty propagation in estimation models based on software measurement results
  → high-risk or unreliable estimates in real-world applications

Spin-off: PhD Research Goal(s)

Le génie pour l'industrie

- **Identify, test, and provide etalon(s) (for the selected FSM method)**
  - driving criteria: Reproducibility and Repeatability
- Provide a sound etalon design methodology
- Compare available (candidate) options available to serve as etalons ('Material' reference? Procedure? Tool? …?)
- Check sound metrology concepts and practices applied to software measurement standards and practices
- Clarify the 'nature' of software '(functional) size'
  - Identify possible improvements to FSM method(s) principles and definitions
- Seek and develop a consensus on the proposed etalon(s)

Example: ISO/IEC FSM-related Intl. Standards

- 14143 FSM series/TR (1-6)
  - Definition of concepts / Conformity evaluation / Reference model (→)
  - No specific method is recommended
- 19761 (latest: 2011): COSMIC
  (functional processes, data movements/groups)
- 20926 (latest: 2009): IFPUG
  (elementary processes, DET's or attributes, logical files' references)
- 20968 (latest: 2002): MkII (I/O attributes , file references)
- 24570 (latest: 2005): NESMA (ca. = IFPUG)
- 29881 (latest: 2010): FiSMA (more physical/technical aspects)

ISO/IEC 14143-4 (FSM) Reference Model

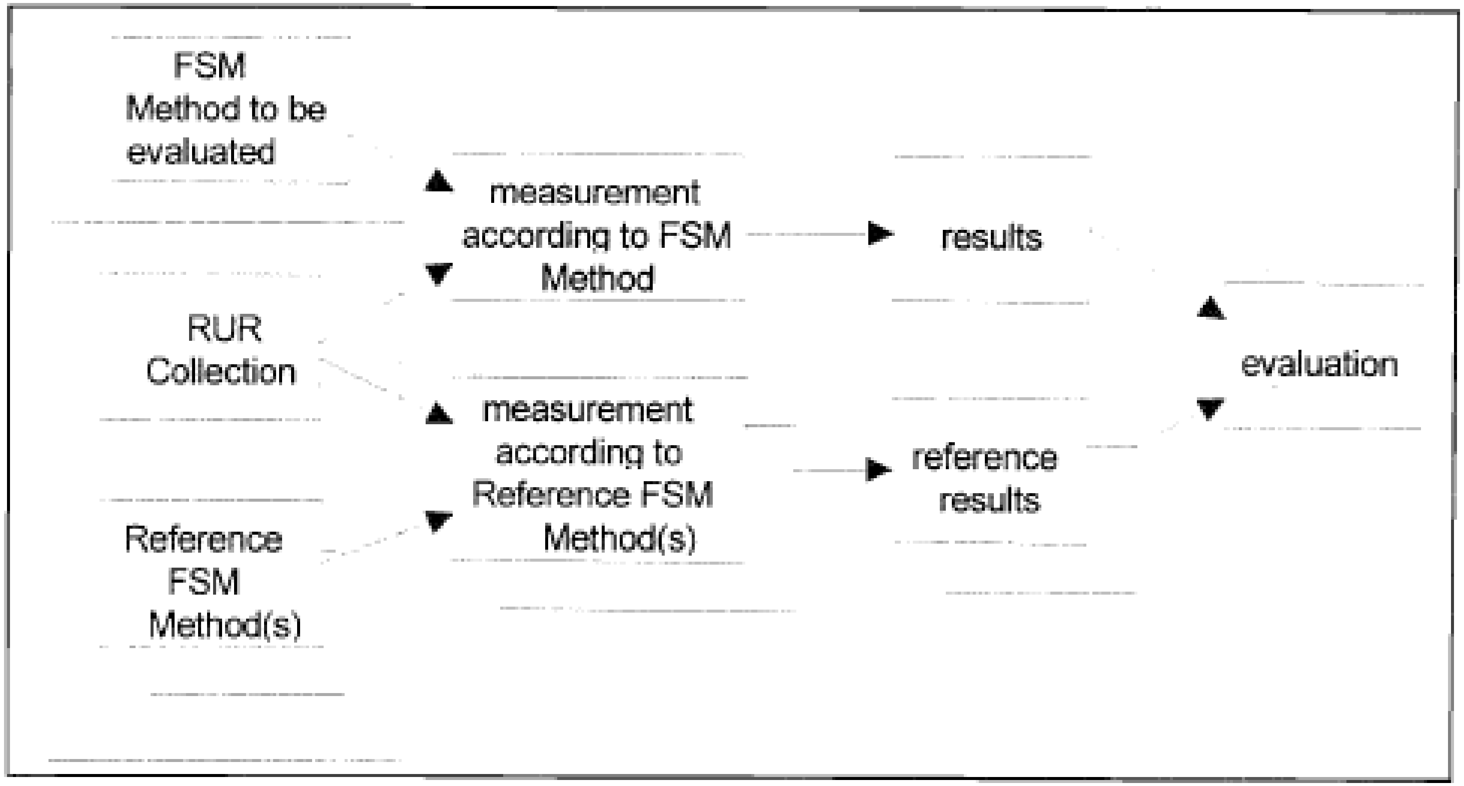


Figure 0.1: Use of RUR and Reference FSM Methods

- Set of Reference User Requirements (8 «business» + 14 realtime/control)
- Not implemented (no Reference FSM Method identified)

Metrology legal entities and references

Bureau International des Poids et Mesures (BIPM), established by Article 1 of the Convention du Mètre, in 1875.

Member of **Joint Committee for Guides in Metrology (JCGM)** with Intl Electrotechnical Commission (**IEC**), Intl Federation of Clinical Chemistry and Laboratory Medicine (**IFCC**), Intl Organization for Standardization (**ISO**), Intl Union of Pure and Applied Chemistry (**IUPAC**), Intl Union of Pure and Applied Physics (**IUPAP**), Intl Laboratory Accreditation Cooperation (**ILAC**), Intl Organization of Legal Metrology (**OIML**).

Metre Convention Treaty: 50+ nations (+ 20 associated with the **General Conference on Weights and Measures**, meeting every 4 years).

- *Guide to the Expression of Uncertainty in Measurement* (GUM)
- *International Vocabulary of Basic and General Terms in Metrology* (VIM) (a common reference […] irrespective of the field of application)
- *International System of Units (SI) brochure* (reference)
- *Metrologia* (journal)

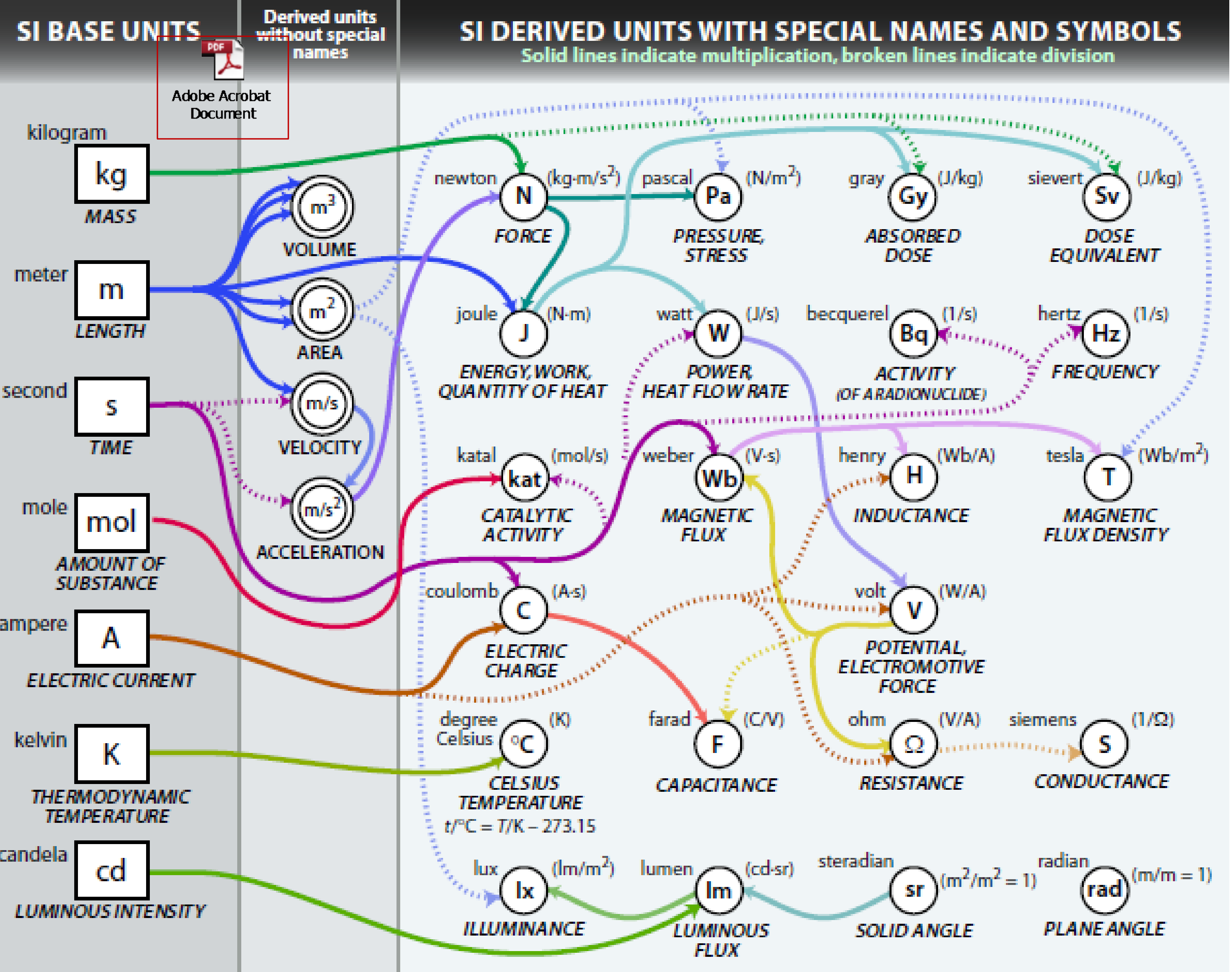
(… other publications)

Some metrology concepts (source: VIM)

- **Quantity**: property of a phenomenon, body, or substance, where the property has a magnitude that can be expressed as a number and a reference
  - Note: A reference can be a measurement unit, a measurement procedure, a reference material, or a combination of such.
- **Kind of quantity**: aspect common to mutually comparable quantities
- **System of quantities**: set of quantities together with a set of noncontradictory equations relating those quantities
- **Base quantity**: quantity in a conventionally chosen subset of a given system of quantities, where no subset quantity can be expressed in terms of the others
- **Derived quantity**: quantity, in a system of quantities, defined in terms of the base quantities of that system

More metrology concepts and definitions

- **International System of Quantities (ISQ, or SI)**: system of quantities based on the seven base quantities: Length, Mass, Time, Electric Current, Thermodynamic Temperature, Amount Of Substance, and Luminous Intensity (see → diagram)
  - Q. Are all 7 base quantities– actually "BASE quantities"?
  - Q. Where does "(SW) FUNCTIONAL SIZE" might fit? And, is it to be considered a base or a derived quantity?
  - Q. What about "[development/other] EFFORT"?



Source: NIST, Physical Measurement Laboratory, http://physics.nist.gov/cuu/Units/SIdiagram.html

Further metrology concepts

- **Quantity dimension**: expression of the <u>dependence of a quantity on the base quantities</u> of a system of quantities as a product of powers of factors corresponding to the base quantities, <u>omitting any numerical factor</u>
  - Example. In the ISQ, the quantity dimension of force is denoted by: "dim F" = [F] = [L][M][T^{-2}]
  - Non-SI example. In SW Eng., "**<u>productivity</u>**" is usually taken as product amount (e.g. functional size) normalized to the effort spent (for development and/or other activities), therefore:
    [Productivity] = [Functional Size][$Effort^{-1}$]

And again, some metrology concepts…

- **Quantity of dimension one (or, dimensionless q.)**: quantity for which all the exponents of the factors corresponding to the base quantities in its quantity dimension are zero
    - "Dimensionless" since all exponents are zero in the symbolic representation of the dimension for such quantities.
    - The measurement units and values of quantities of dimension one are numbers, but such quantities <u>convey more information than a number</u>.
    - Some quantities of dimension one are defined as the <u>ratios of two quantities of the same kind</u>.
        - e.g. plane angle, refractive index, mass fraction.
    - <u>Numbers of entities</u> are quantities of dimension one.
        - e.g. number of turns in a coil, number of molecules in a given sample.

OK, the last one on metrology concepts

- **Measurement unit**: real scalar quantity, defined and adopted by convention, with which any other quantity *of the same kind* can be compared to express the ratio of the two quantities as a number
- **Submultiple of a unit**: measurement unit obtained by dividing a given measurement unit by an integer >1
    - Example. The millimetre is a decimal submultiple of the metre.
- **Measurement principle**: phenomenon serving as a basis of a measurement
    - Example. Thermoelectric effect applied to the measurement of temperature.
    - Note. The phenomenon can be of a physical, chemical, or biological nature.

Issues in 1st Generation FSM Methods

- **“same kind”**
  - Function Points “from” Logical Files <u>added</u> with Function Points “from” Elementary Processes
- **scale homogeneity**
  - I/O data elements added with (weighted) file references
- **scale correspondence with the measured phenomenon**
  - DET, RET, FTR ranges (ref. Hottentots’ “1, 2, many”)
  - Low / Average / High levels (ref. hotels stars rating)
- **oversimplification**
  - “one type fits all” (distribution profiling =/= measurement)
  - “one (given) size fits all” (averaged sizes =/= measurement)

I lied… THIS is the last one (on metrology concepts)

- **Measurement standard (Etalon)**: realization (sometimes: embodiment) of the definition of a given quantity, with stated quantity value and associated measurement uncertainty, used as a reference.
  - Examples. 1 kg mass measurement standard (with an associated uncertainty of 3 μg); Caesium frequency standard (with a relative uncertainty of 2×10^{-15}).
  - Note. A "realization of etc." can be provided by a measuring system, a material measure, or a reference material.
  - Note. "realization" used here in the most general meaning – 3 possible procedures:
    i) physical realization of the measurement unit,
    ii) "reproduction", by highly reproducible measurement standard based on a physical phenomenon (e.g. frequency-stabilized lasers for 1 m),
    iii) a material measure adopted as the meas. standard (e.g. the kg).

Metrology in Software Measurement
State of the art

- Software Measurement History
    - ‘Archeology’ to Modern era
- Etalons in Software Measurement
    - Preliminary research

Software Measurement ’70-’80

- ‘software metrics’
  - pragmatic approach, no metrology adherence
  - software quality discussed, rather then ‘size’
- ‘software science’ and ‘software physics’
  - attempt to define a ‘system of metrics’
  - serious definition flaws and lack of validation
- warnings on the risk of pragmatic approaches
- distinction between product and process metrics
- distinction between physical and logical attributes

Software Measurement ’90

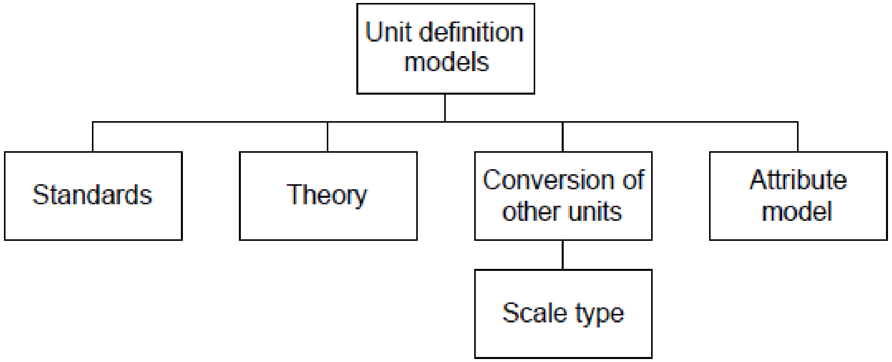
- increasing amount of proposals on the topics
- still pragmatic approaches (e.g. GQM)
- warnings on the importance of metrology basis
  - ‘empirical relations’, mapping between objects and representation, ‘direct’ vs. ‘indirect’, ‘scale types’, …
- standardization of concepts of Functional Size Measurement (ISO/IEC 14143 series)
  - definition of generic FSM concepts

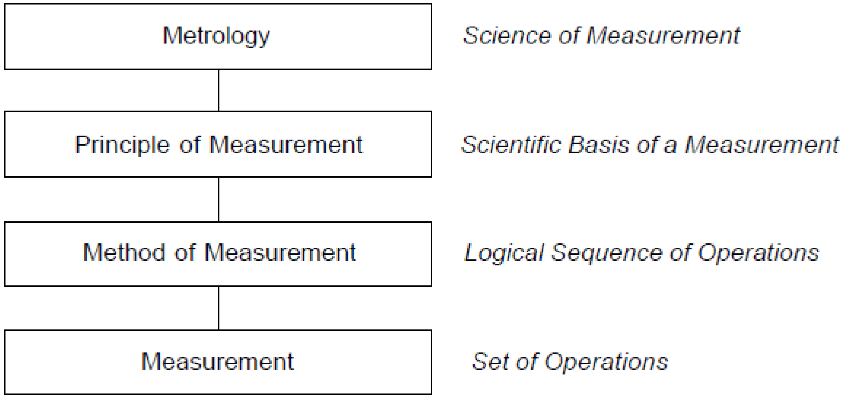
Software Measurement ’90-2000’s

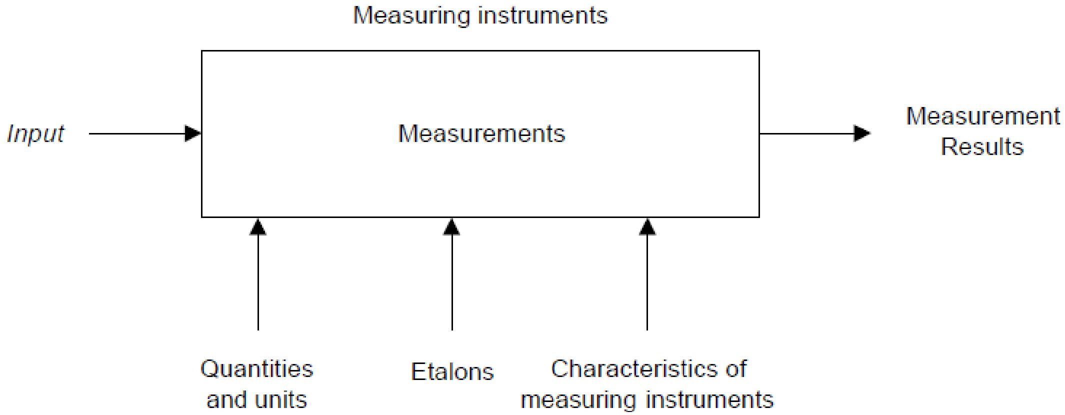
- parallel standardization
  - ‘software product quality’ (ISO/IEC 9126)
  - ‘software measurement process’ (ISO/IEC 15939)
    - no consideration of the concept of ‘etalon’
- warnings on the importance of metrology concepts
  - ‘unit of measurement’, ‘scale types’, ‘measurement validity’ and ‘reliability’, ‘measurement error’ or ‘uncertainty’, measurement ‘accuracy’ vs. ‘precision’, …
- standardization of concepts of Functional Size Measurement (ISO/IEC 14143 series)
  - Part 4: ‘reference model’ (catalogue of req’s samples)
  - COSMIC and other methods: 14143-compliant

Metrology and Software Measurement

- researched started at ETS (Abran, Sellami)


Unit definition models
Standards
Theory
Conversion of other units
Attribute model
Scale type


Metrology — *Science of Measurement*
Principle of Measurement — *Scientific Basis of a Measurement*
Method of Measurement — *Logical Sequence of Operations*
Measurement — *Set of Operations*


Measuring instruments
Input
Measurements
Measurement Results
Quantities and units
Etalons
Characteristics of measuring instruments

Etalon Design in Software Measurement

- applied research
  - analysis of COSMIC measurement method under metrology concepts (Abran *et al*, 2004)
    - "close to the concept of etalons (for FSM methods) are the ‘case studies’ documented"
    - a combination is proposed of
    1. documented FURs (Functional User Requirements),
    2. FUR description in terms of Use Cases according to the RUP model, and
    3. functional size measurement results

- exploratory study (Khelifi, 2005) (→)

Exploratory study on COSMIC/RUP prototypes as etalons

- Khelifi focuses on
  - how to tackle the etalon design (process)
  - how to propose an initial version of etalons, for FSM

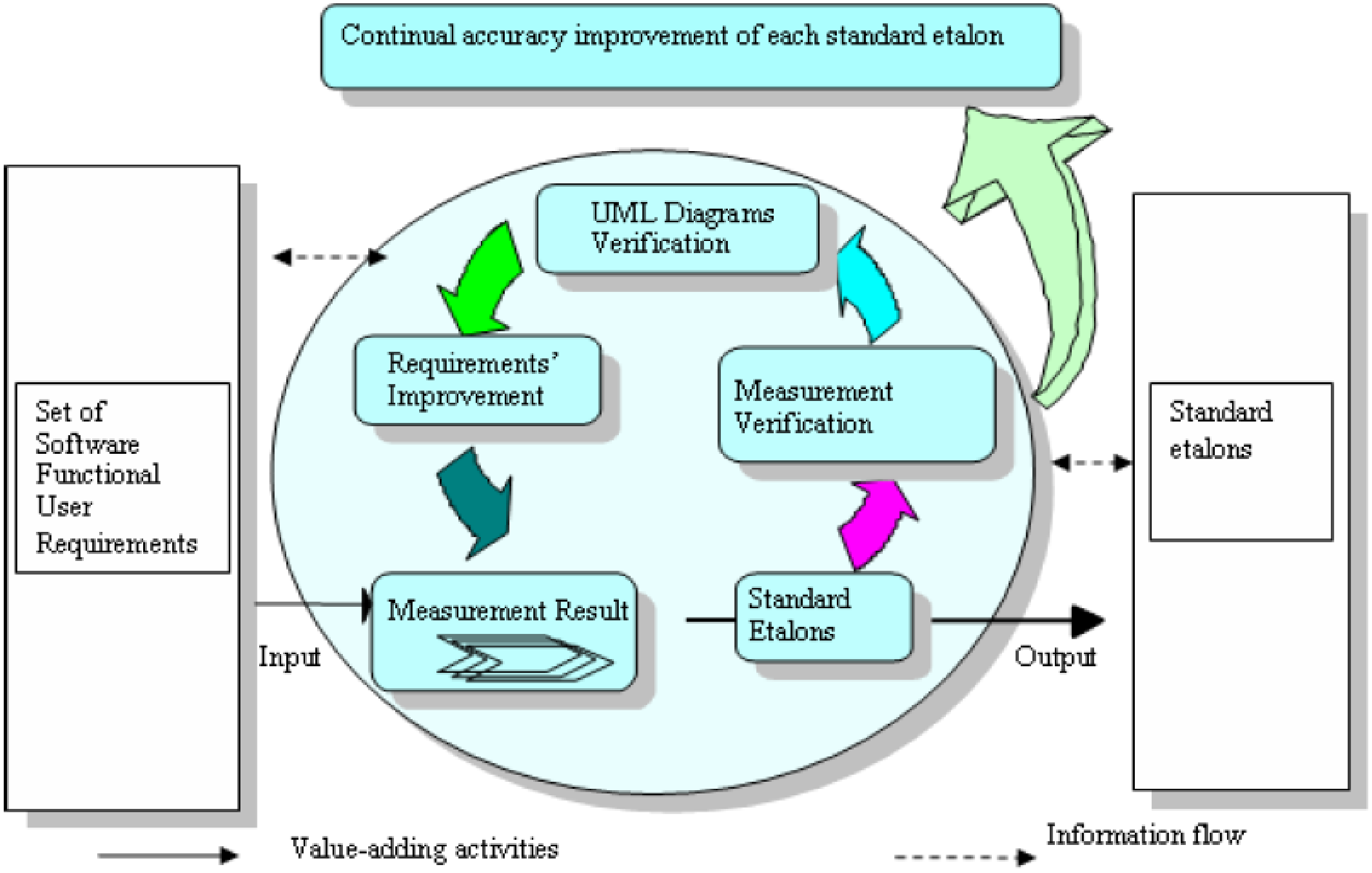
  (exploratory and qualitative research)
- methodology: 'liaison method'
  - '<u>linking of any measurement result to one or more recognized measurement references</u>'

1. Measurement specification.
2. Measurement interpretation compared to the system of reference.
3. Liaison with the system – binds a measurement result to a reference and gives the measurement the corresponding name.
4. Addition of a new reference. (If a measurement is far from all the references defined previously, we add it as a new reference)

Exploratory study on COSMIC/RUP prototypes as etalons

- Iterative process
  - 8 available case studies measured with the COSMIC method, v2.2 (2003)
  - added with UML diagrams and distributed to meas. experts for verification
  - assigned a final size value, and published as an initial set of reference material
  - results discussed – 3 types of verification (individual, SME, and MPC verification)
- Led to the definition of an (iterative) etalon design process:
  1. Defining the expert intervention purpose.
  2. Selecting the experts.
  3. Developing the proposed etalons.
  4. Collecting the measurement results over the etalons from the experts.
  5. Comparing the measurement results.
  6. Discussing the measurement results' quality.
  7. Issuing the (first) version of the etalons.

Exploratory study on COSMIC/RUP prototypes as etalons



Practical issues with Khelifi's proposal

- Particular cases studies (uncovered measurements → add more case studies to the etalons set)
  - Possibly reaching a too-large set of etalons for practical purposes, with uncertainty on their mutual correct ratio
- The order of magnitude is higher than that of the Base Functional Component aimed to be measured
  - 1 m first defined as the 40-millionth part of the Earth circumference (with related issues), then evolved up to the length of the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second (note: light speed *c* is a cosmological constant, and a natural phenomenon)

Possible solutions (alternatives to be investigated)

- Ad interim refinement of the top-down approach (such as Khelifi's proposal)
  - Identify paradigmatic cases (<u>FUR patterns</u>, eventually per domain) at the level of the <u>functional processes</u>
  - Measure the FUR patterns according to current standard method(s)
  - Further measurements might be carried out with the method practices (count data movements) and independently validated by counting «how many / how many times the FUR Patterns are present in the case being measured»
  - Provide improvements (refinements, not redefinition) to the current method(s) to reduce ambiguities and variability among measurers and measures

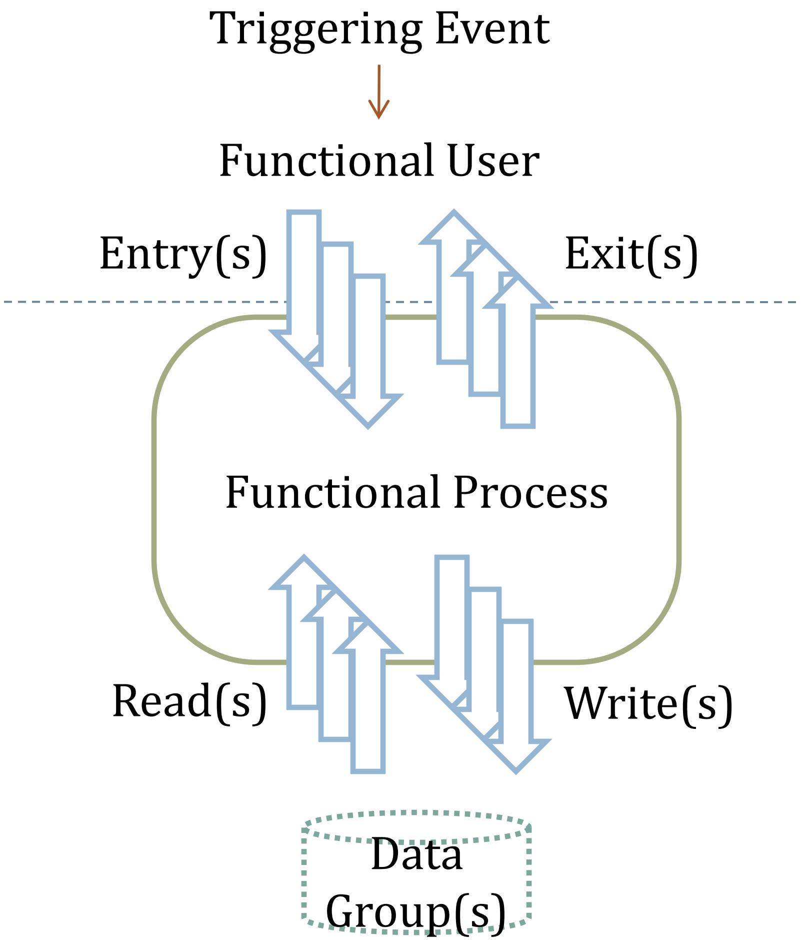
Lower scale alternatives

- Since the Data Movement (of a single Data Group) is the BFC in the COSMIC method, why can't we consider it as the order of magnitude of the desired etalon?
  - It would be most likely «self defined», not easily realized as a «concrete» reference for most practical purposes
  - Therefore, alternatives might come from <u>bottom-up</u> approaches such as
    - Counting amounts of data attributes per data movements, then averaging («the standard data movement is that DM that comprises N (fixed given number of) Data Attributes»
  - Drawbacks: defining and etalon on the basis of an average means relying on the law of large numbers (not necessarily true for any single case, and for «small» cases)
  - Data Manipulation is also averaged over Data Movements

COSMIC FSM Method

Base Functional Components & SW Generic Model

- Definition, Principles, and Rules exist to map FUR's onto **Functional Processes**
- BFC's: **Data Movements** (1-1 with **Data Groups**)
  - **Data Manipulation** is not measured separately (considered to be associated with Data Movements)
  - Data Groups are composed by **Data Attributes**


Triggering Event

Functional User

Entry(s)

Exit(s)

Functional Process

Read(s)

Write(s)

Data Group(s)

Prospects ahead 1/2

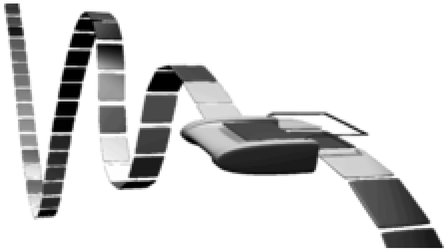
- Bottom-up <u>definition</u> of the quantity being measured (and therefore its corresponding *etalon(s)*)
  - Data Movements carry *INFORMATION* (each data attribute of a Data Group being moved is an information parcel)
  - *INFORMATION* <u>units</u> (***bit***, or ***nat*** [sometimes *nit*, or *nepit*]) are defined in Information Theory (eg. for Shannon entropy)
  - *INFORMATION*(or entropy) might be either <u>derived</u> by some other existing base quantities <u>OR</u> proposed for addition as a conventional, <u>new base quantity</u>
    - Data manipulations might be regarded as a special type of (inner) data movement (to be investigated)

Prospects ahead 2/2

- Functional size for industrial usage might be derived from the «bottom» quantities by means of averaging at a higher scale level (ie. tolerating larger uncertainties where acceptable)
  - The same way we require *ms*-scale when recording racing athletes on the 100 m running specialty, but we accept non-atomic time counting when evaluating queue waiting times in *minutes*, and even larger uncertainty when evaluating larger timescales such as human *ages*, historical *periods*, cosmological *eras*.
  - First, need to identify (and possibly standardize) the acceptable tolerance thresholds

Conclusions

- Measurement processes requires precision and accuracy
    - We can simplify complex methods for practical usage, the opposite would be impossible (to inject desired complexity into oversimplified methods)
- Re “etalons”: designing etalons is not a one-shot & perfect-at-first activity
    - Ad interim solutions, progressively refined, however useful
    - Methods and practices will improve in parallel
- Bottom level: we do have concepts and theorems useful for tackling the base «information» level (Information Theory) since decades
    - e.g. ANY algorithmic processing can expressed by means of a (universal) **Turing machine** (with a given set of rules), with a **minimal amount of steps** that would mimic the Entry-Read-Write-Exit sequence of any real functional process



Thank you for your attention!

Questions / Remarks / Feedbacks?

Luca Santillo, MSc

UKSMA/COSMIC
International Conference on
Software Metrics & Estimating
London, October 27th / 28th 2011

UKSMA United Kingdom Software Metrics Association

COSMIC Common Software Measurement International Consortium