A STRUCTURED METHODOLOGY FOR MEASUREMENT DEVELOPMENT

Dario Petri(1), Luca Mari(2) and Paolo Carbone(3)

(1) University of Trento, 38123, Trento, Italy, e-mail: dario.petri@unitn.it

(2) Università Cattaneo - LIUC, 21053 Castellanza, Italy, e-mail: lmari@liuc.it

(3) University of Perugia, 06125 Perugia, Italy, e-mail: paolo.carbone@unipg.it

Abstract *– A conceptual framework is proposed in which measurement development is envisioned as a three-level hierarchically structured process constituted of stages, each one composed of activities performed through multiple tasks. The underlying assumptions are that measurement is a designed-on-purpose process and that the resources it requires should be justified on a pragmatic basis. The usefulness of the proposed methodology is then illustrated by applying it to an exemplary development of the measurement of the input-output transfer characteristic of an analog-to-digital converter and in the case of a research quality assessment. In this second case, the framework is used also to identify similarities and differences between this information gathering process and measurement.*

Keywords – measurement; conceptual framework; development methodology; modeling; design; uncertainty.

I. INTRODUCTION

The acquisition of information about physical quantities by means of sensors historically fostered the interpretation of measurement as a merely experimental activity: once a suitable physical transduction effect is identified and embedded into a device, it might seem that what remains to be done are only the technical operations needed for making the device interact in a controlled way with the object under measurement and then reading the outputs of the device itself. Thus, even though the role of measurement is considered important in many scientific and technical activities, providing a conceptual and procedural structure to measurement process, and on this basis a methodology for measurement development, is usually not perceived as a critical target by scientists and engineers, obviously with some notable exceptions (e.g., [1]–[4]). Conversely, the impressive results achieved in physical sciences and engineering thanks to measurement have been attracting scientists to the quest of measurement scales and measurement procedures for non-physical properties. This is particularly true in social sciences such as psychology [5] and economy [6], where the experimental acquisition of information is not performed by means of physical transducers, a situation sometimes called “weakly defined measurement” [7], or “measurement in soft systems” [8], or simply “soft measurement”. Opposite, physical quantities are sometimes known as “strongly defined” or “hard” properties.

Moreover, the availability of reliable information on the empirical world (in [9] this generic concept of reliability is developed in terms of object dependence and subject independence) is systematically required for the optimal management of complex processes or systems, thus soliciting the adoption of measurement processes in a multitude of human activities. In their turn, measuring systems are becoming more and more complex, also because of the increasing need to acquire information on both physical and non-physical properties in an integrated way [10].

This steadily evolving scenario involves several significant issues for measurement science and technology, not only at the operative level (for example, about the possible advantages of adopting the same concepts and procedures for measurement uncertainty evaluation in such different fields) but also, and primarily, in reference to the fundamentals of the discipline itself.

The axiomatic approach to measurement [11], [12] has provided some contributions to the construction of a shared conceptual basis for measurement, but precisely for its abstract nature it falls short in accounting for this increasing structural and procedural complexity, not to mention its inability to support complexity management. Complementarily, by inspiring to both the well-known plan-do-check-act cycle [13] employed to represent product development processes in manufacturing or software engineering [14], and the experience of the Authors with measurement of physical properties, this paper proposes a characterization of measurement development *from a pragmatic perspective*. The underlying assumption is that a better comprehension of the way measurement processes are developed can be achieved by means of a conceptual framework describing the operative structure of such processes. In fact, a framework eases discussions by setting a context of common language and shared concepts; it favors interpretation of evidence and assists researchers in attributing a meaning to subsequent findings; it may help understanding the potential weaknesses of existing measurement processes, also providing elements for discussions and improvements, and contributing to bridge the several bodies of knowledge that have been almost independently developed around the same target of measurability [15]. Accordingly, the proposed conceptual framework is aimed at giving operative guidelines to support a disciplined and methodologically sound development of measurement for physical properties, and plausibly at least in some cases for non-physical ones.

The paper is organized as follows. At first, in Section II, we introduce the conceptual framework by presenting, in a structured and operative way, the different stages and the various activities involved in the development of a measurement process. In Section III the illustrative application of such a development methodology to the description of activities performed to measure the dynamic transfer characteristic of an *Analog-to-Digital Converter* (ADC) further hints the soundness of the proposal. The framework is then used, in Section IV, to identify similarities and differences between a research quality assessment based on experimental data and a measurement process.

II. A CONCEPTUAL FRAMEWORK OF THE STRUCTURE OF MEASUREMENT

Measurement is a complex, model-based, goal-driven process [9], [16]. The multiple activities performed (sometimes implicitly) to accomplish a measurement can be effectively presented in a conceptual framework, based on the following assumptions:

- all empirical properties can be, in principle, measured by performing logically equivalent steps;
- models are unavoidable in measurement, and they are co-determined by the measurement goals.

Such a framework, synthesized in Fig. 1, interprets measurement as a three-level hierarchically structured process constituted of (i) *stages*, each one composed by (ii) *activities* performed through multiple (iii) *tasks*. For the sake of clarity, in the following each stage, activity, and task will be denoted by a unique identifier of the form {x}, {x.y}, and {x.y.z} respectively, where x identifies a stage, y an activity in the stage x, and z a task in the activity y. For example, {pln.gls.prd} is the task of purpose definition in the activity of goal setting in the stage of planning. Table 1 lists all stages, activities, and tasks with their identifiers.

A loose temporal sequence drives the execution of tasks (from top to bottom in Fig. 1), but the systematic presence of feedback paths (highlighted by the bidirectional path on the left side of the diagram) emphasizes the complexity of the whole process. In general, in any system, feedback is a source of structural complexity because it makes effects to become (delayed) causes. Conversely, waterfall models – strictly from planning to design to realization – are attractive for their simplicity, but prevent the improvement of previous activities and, sometimes, hinder the achievement of satisfactory results. In particular, the presence of feedback paths in the framework underlines the need, or at least the option, to perform successive steps of adjustment or refinement, until all planned requirements are met. This structural complexity is the basic justification for characterizing measurement development by means of a conceptual framework, instead of a step-by-step procedure.

In the first stage, *planning* {pln}, a priori knowledge, resources, and constraints are assumed, and on this basis the measurement goal is defined. This allows specifying the minimum acceptable level of delivered information about the measurand, formalized as the maximum acceptable measurement uncertainty in measurement results [9], [16], [17]. A model of the properties involved in measurement, i.e., the measurand and other relevant properties of the empirical context in which measurement is performed, must also be defined in this first stage. In addition, the measuring system must be designed.

Table 1. Identifiers of stages, activities, and tasks in the framework.

ID	stage	activity	task
{pln}	planning		
{pln.gls}		goal setting	
{pln.gls.prd}			purpose definition
{pln.gls.obs}			object under measurement specification
{pln.gls.gpd}			general property definition
{pln.gls.ctu}			choice of target uncertainty
{pln.mdl}		modeling	
{pln.mdl.mdf}			measurand definition
{pln.mdl.msm}			measuring system modeling
{pln.mdl.emd}			environment modeling
{pln.mdl.ipd}			influence properties definition
{pln.mdl.mid}			mutual interactions definition
{pln.mdl.uan}			uncertainty analysis
{pln.dsg}		design	
{pln.dsg.mpc}			measurement principle choice
{pln.dsg.mmd}			measurement method design
{pln.dsg.mpd}			measurement procedure design
{pln.dsg.mpl}			measurement planning
{exc}	execution		
{exc.stp}		setup	
{exc.stp.omd}			object under measurement detection and preparation
{exc.stp.msc}			measuring system calibration
{exc.stp.ssv}			scale setup and validation
{exc.stp.mss}			measuring system setup
{exc.daq}		data acquisition	
{exc.daq.mda}			measurand related data acquisition
{exc.daq.ipa}			influence properties data acquisition
{exc.ier}		information extraction and reporting	
{exc.ier.dap}			data analysis and processing
{exc.ier.vla}			value assignment
{exc.ier.unc}			uncertainty computation
{exc.ier.sir}			subsidiary information reporting
{int}	interpretation		
{int.dcs}		decision	
{int.dcs.cvl}			check for validation
{int.dcs.cga}			check for goal achievement
{int.lrn}		learning	
{int.lrn.lsl}			lessons learning

In the second stage, *execution* {exc}, the measuring system is calibrated, so to guarantee that the information it provides is traceable to a measurement standard, and the conditions for its proper interaction with the object under measurement are setup. These conditions possibly require some interventions on the object aimed at making the measurand, or a property dependent on it via a known functional relation, accessible. The measuring system is then put in interaction with the object under measurement and raw measurement data are acquired and processed to properly represent the measurement result. Finally the achieved information is expressed in a way appropriate for its presentation and communication.

In the third stage, *interpretation* {int}, the measurement result is exploited in fact-based decision making activities, usually performed to assess the achievement of the measurement goal. Measurement results can also be used to validate or improve the whole measurement process [16].

In the following subsections, these three stages are discussed with additional details.

A. Planning: Goal setting, Modeling and Design {pln}

Goal setting {pln.gls}

Typically, measurement is aimed at supporting decision-making. The methodology described in this paper suggests that the specific measurement purpose {pln.gls.prd} must be firstly established on the basis of the available a priori knowledge, while identifying and keeping into account the existing resources and constraints [21]. Hence, the key question to be asked before measuring is more "*why do we want to measure?*" rather than "*what do we want to measure?*". The initial definition of measurement goal is usually quite vague and several subsequent refinements, often suggested by the following stages of the process, are required before achieving a clear and operative purpose definition, thus involving various feedback paths in the measurement development. In the scientific literature concerning software measurements, where dealing with complex systems is more the rule than an exception, several frameworks have been proposed for the definition of the various goal components and the identification of properties to be measured accordingly, among them the Quality Function Deployment [22], the Goal Question Metric, and the Software Quality Metrics [21].

For sure, the class of possible objects under measurement (phenomena, bodies, or substances [17], but also pieces of software, individuals, research teams, industrial processes, business organizations, etc) has to be stated {pln.gls.obs}, because the general property (e.g., length, loudness, extroversion, quality) selected for measurement will be of one of such objects.

This activity also requires the definition of the general property itself {pln.gls.gpd}, possibly leading to different reference scales. Temperature is a classic example: it could be defined, e.g., on the basis of the order relation "hotter than", or a thermometric (interval) or a thermodynamic (ratio) scale.

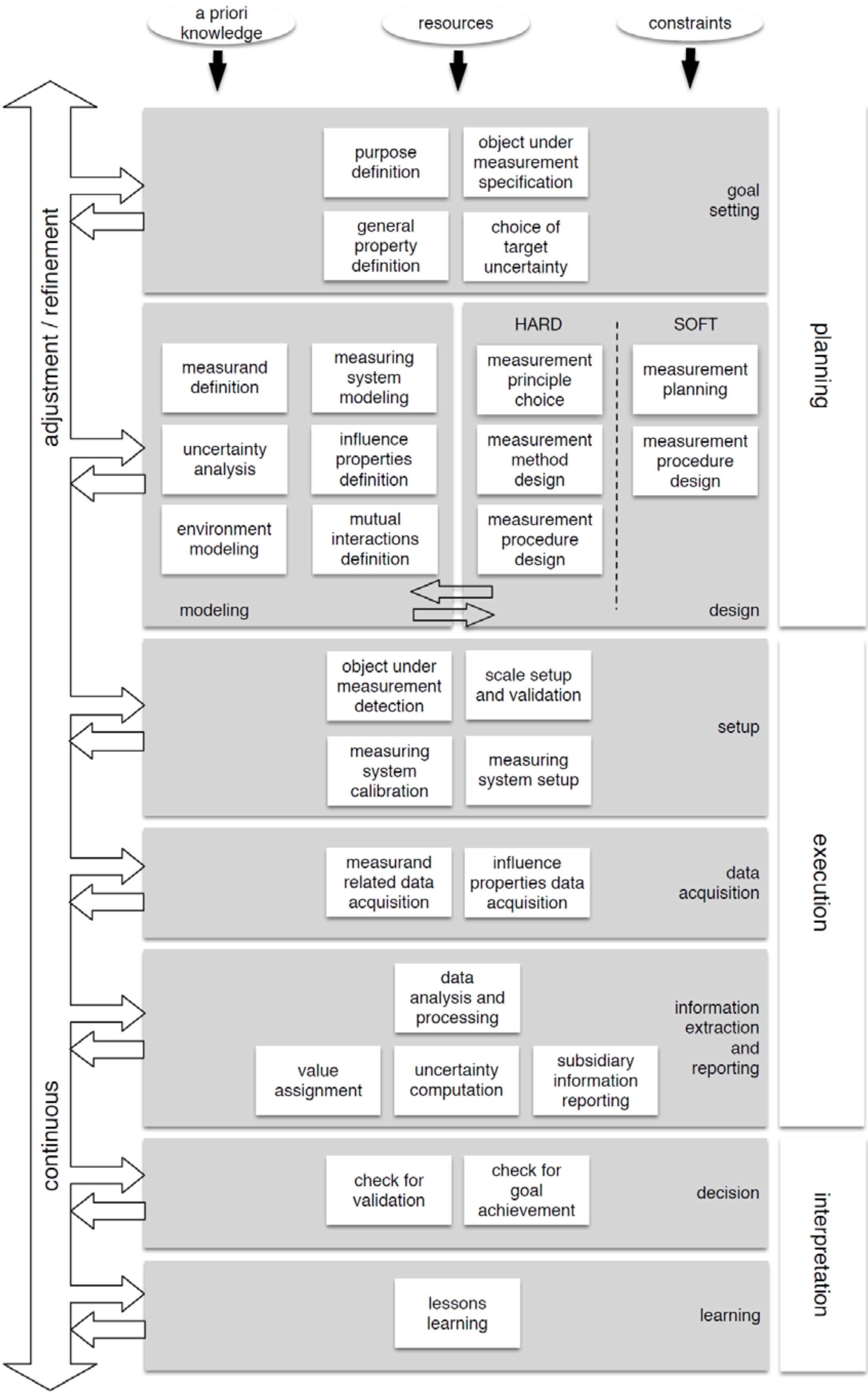

Fig.1. Diagram synthesizing the conceptual framework of the structure of measurement. White arrows represent feedback, whose presence highlights the complexity of the underlying information-intensive process.

The reference scale may be constructed not only directly, on the basis of empirical relations involving the considered property, but also indirectly, by deriving it from a known functional relation linking the considered property to other properties whose scales have been previously established, a usual situation in physical measurement, where modeling is supported by universally accepted theories. Opposite, when measuring soft properties critical issues arise since properties are often ill defined and models cannot be grounded on established theories. A practical approach to address this issue is presented in [23], where a tool is described that guides the user in the definition of the required property. Generally, all relevant stakeholders should be involved in the definition of the general property so to ensure that all meaningful underlying aspects are operatively taken into account.

Furthermore, guarantee must be given that the information delivered by measurement suffices to effectively support the achievement of the intended goal. This constraint is specified as an upper limit to measurement uncertainty, i.e., the *target uncertainty* [17], [18], expressed in the chosen reference scale {pln.gls.ctu}: whenever measurement uncertainty is greater than target uncertainty, the information obtained in measurement is acknowledged not to be appropriate for the aimed decision-making.

The information derived during goal setting establishes the ground for the following activities of modeling and design.

<u>Modeling (of the measurement context) {pln.mdl}</u>

Decisions supported by measurements are about the defined general property as specified of an object, singled out with respect to other properties of the object itself and the surrounding environment. Thus, both the individual property intended to be measured, i.e., the measurand {pln.mdl.mdf} [17], the measuring system {pln.mdl.msm}, and the experimental environment {pln.mdl.emd} must clearly be described. To this aim, modeling activities provide a goal-oriented and operational description of all relevant entities (the object under measurement, the measurand, the measuring system and the environment) in the experimental setting which, as a whole, are called here the *measurement context*. Measurement context identification is then grounded on the available knowledge about the relations between the measurand and the other properties – of the object under measurement, the measuring system or the environment – which might influence the measurement result.

The measurand is the instance of interest of the general property defined during goal setting. The reference to intentions contained in its definition emphasizes the explicit link between the measurement goal and the modeling stage, primarily aimed at defining the measurand itself.

Any model provides just a partial description of the entity it is aimed at representing [21]. Thus, incomplete information about the measurement context is unavoidable. In particular, properties and interactions that are not considered (because we are not aware they can significantly affect raw measurement data or because we expect them to produce negligible effects on measurement data) limit the amount of information about the measurand conveyed by the measurement result. In other words, measurement uncertainty sources can be considered as originating in the modeling activities, that is, *before* the execution stage. In principle, these uncertainty sources are always present, regardless of the accuracy of the employed instrumentation. During the modeling activity, the different uncertainty sources are identified and the model to be used for the computation of their contribution to measurement uncertainty are defined {pln.mdl,uan}.

When defining the measurement context, different situations may occur depending on the amount of information that must be conveyed by the measurement result. For instance, the effect of all properties of the measurement context on raw measurement data may be negligible as compared to the target uncertainty, and therefore none of them needs to be modeled and controlled. At the opposite side, they must be measured and controlled in turn. The properties of the context unlike the measurand that significantly (with respect to target uncertainty) affect measurement data are called *influence properties* [17], [21]. They must be identified and modeled {pln.mdl.ipd}, together with their mutual interactions and their effect on measurement data {pln.mdl.mid}. Indeed, if one or more influence properties are not identified, and thus not included in the context model, the information provided by measurement might be wrong. Detecting these situations may help to improve the context model. Usually, influence properties are described in terms of mathematical variables and relationships, respectively, and represented either in deterministic or non-deterministic (such as stochastic or random-fuzzy) way.

Definitional uncertainty is the contribution to measurement uncertainty that originates from the finite amount of information available in the measurand definition, and includes also uncertainty originating in the description of the effects of the environment properties on the measurand. Thus, this uncertainty contribution constitutes a lower bound to measurement uncertainty: it makes no sense reducing any other contributions at a much lower level than definitional uncertainty itself.

Observe that the notion of definitional uncertainty is strictly related to two concepts that are crucial for modeling, i.e., *measurement significance* and *measurement validity*.

Through measurement significance the issue "*are we measuring the right property in the right way?*" is kept into account, which refers to whether the measurement result is expected providing useful information to support decisions needed for the achievement of the intended goal. Hence, measurement significance is related to the appropriate definition of the property intended to be measured in reference to the given goal, and it can be increased only through a better operational definition of the measurand.

Different from, but strictly related to, measurement significance is measurement validity. According to [17] and [22], a measurement process is validated by providing objective evidence that its result fulfills the intended goal with a suitable level of confidence. Hence, to be valid a measurement must be significant and the related uncertainty must be lower than the target uncertainty with a given confidence level. Therefore, the validity issue is "*are we measuring the right property, in the right way and with the right uncertainty?*", and it refers to whether a measurement provides enough information to drive to the achievement of the intended purpose. It should be also noted that, only the concept of measurement significance is discussed in most scientific literature [19], often under the name of measurement validity, since the notion of target uncertainty is usually neglected.

As it occurs for the measurand model, the finite amount of information about measuring system provided by the adopted models propagates to measurement results, thus contributing to measurement uncertainty. This contribution is called *instrumental uncertainty* and must be thoroughly modeled and controlled. A lower bound to instrumental uncertainty is the calibration uncertainty, that is the uncertainty derived from the calibration of the employed instrumentation. Indeed, any deviation between instrument operative conditions and calibration conditions increases instrumental uncertainty.

A third uncertainty component is the *interaction uncertainty*. It originates in the description of the interaction between the object under measurement and the measuring system. Hence, this component keeps into account the fact that the specific interaction might alter the status of the object under measurement, which may differ then from nominal conditions.

All the previous components of measurement uncertainty are physiological to measurement, that is, they are inherently present in any measurement. Their contribution on the measurement result can, and in our opinion should, be estimated before the execution stage.

In synthesis, modelers should guarantee that the expected measurement uncertainty is greater than the definitional uncertainty and lower than the chosen target uncertainty with a sufficiently high confidence level. In principle, uncertainty of the description of the measurement context – and then measurement uncertainty – can be always reduced by including additional levels of detail, as increasing the number of influence properties, improving the accuracy of mathematical relationships or narrowing the ranges of allowed values for influence properties. However, in this way the complexity of the context model increases, resulting in higher computational complexity of modeled relationships and more severe control requirements for influence properties. This choice generally imposes sophisticated and expensive instrumentation and measurement procedures, which in turn require skilled designers and experimenters. As a consequence, measurement costs increase and may not be justified by the achievement of the intended goal. In addition, the amount of obtained information may exceed the requirements specified in terms of target uncertainty, thus becoming partly useless, and therefore pragmatically unjustified in view of the resources required to acquire it. On the other hand, the amount of information available when using low complexity models may not be enough to effectively support decisions, i.e., measurement uncertainty may be greater than the specified target uncertainty.

It follows that, according to this framework, the construction of the most accurate context model is not the goal of the modeling actions. Indeed a *suitable* context model should pick only the aspects of the context that are significant for the measurement goal, by trading-off effective support to decision-making and affordable consumption of economic and technical resources.

Design (of the measuring system) {pln.dsg}

Although any measurement may be abstractly considered as directly referred to a measurement scale for the predefined general property, its execution is usually grounded on the design and the setup of a measuring system that suitably embeds (a subset of) such a scale and aims at guaranteeing the objectivity of the process and its results.

The concept of measurement system is traditionally defined focusing on the empirical component of the structure, e.g., "set of one or more measuring instruments and often other devices, including any reagent and supply, assembled and adapted to give information used to generate measured quantity values" [17] or "set of elements able to interact with the object under measurement and to produce, as a result of this interaction, an output, on the basis of which it is possible to assign values to the measurand (in agreement with a previously established reference scale)" [15]. Both these definitions assume that what measuring systems generate is not a measurement result but only an intermediate product, that must be then post-processed, plausibly by human beings. The current widespread options to introduce software components into hardware devices have broken this assumption. Indeed, if properly formalized, the measurement context model can be implemented and the process execution can be made fully automatic accordingly. Hence, an encompassing definition can be proposed by adapting the ones above: "a measuring system is a set of elements adapted so to allow interaction with the object under measurement and to produce, as a result of this interaction, a measurement result, or an intermediate output from which a measurement result can be obtained". Assuming that elements of measuring systems can be not only physical sensors, measurement standards or instruments but also, for instance, software programs or questionnaires, this definition seems to be general enough not to exclude any kind of measurement. At the same time it guarantees that the employed measuring system, however implemented, is subjected to basic metrological conditions, so ensuring measurement traceability and a proper control of instrumental uncertainty sources.

It is important to remark that, when measuring physical properties, the design activity implies establishing [17]:

- a *measurement principle*, i.e., a physical phenomenon at the basis of the measurement {pln.dsg.mpc};
- a *measurement method*, i.e., a generic description of the logical organization of operations required for a proper application of the adopted measurement principle {pln.dsg.mmd};
- a *measurement procedure*, i.e., a detailed description of all operational steps required to achieve the measurement result according to the chosen measurement method (for the sake of

conciseness, we assume here that suitable instrumentation, when needed, is available and does not need to be designed and implemented) {pln.dsg.mpd}.

Conversely, when dealing with soft measurements, no general theory involving the measurand usually exists. As a consequence, no degree of freedom in choosing among principles, methods, or procedures is allowed so that the measuring system is usually designed by firstly establishing a generic description of the operations to be performed {pln.dsg.mpl}, often called *measurement plan*, and then a detailed measurement procedure {pln.dsg.mpd} [24].

B. Execution: Set-Up, Data Acquisition, Information Extraction and Reporting {exc}

Before acquiring information about the measurand, some setup activities have to be performed {exc.stp}: the designed reference scale has to be properly implemented {exc.stp.ssv} and the designed measuring system must be properly assembled {exc.stp.msc} and calibrated in order to guarantee the traceability of measurement results. In measurement of physical properties, the outcome of calibration may be represented as a calibration diagram, i.e., the "graphical expression of the relation between indication and corresponding measurement result", which also provides information on instrumental uncertainty [17].

Moreover, to correctly acquire information on the measurand, the object under measurement has to be detected {exc.stp.omd} and sometimes properly prepared, as occurs for instance when dealing with chemical quantities. Also, and the measuring system must be properly setup with respect to both the object under measurement and the surrounding environment {exc.stp.mss}, for example by connecting the object under measurement and the adopted instruments, or by turning on in advance equipment to guarantee a thermal steady-state status, or by preparing and cleaning containers according to predefined procedures. Similarly, in soft measurements, forms to support a structured data collection must be arranged, test items must be calibrated and the environment where the interviews or the test will be conducted must be setup.

After this preliminary activity, data acquisition is performed according to the defined measurement procedure, and raw measurement data are obtained, on both the measurand {exc.daq.mda} and influence properties {pln.mdl.ipd}. The subsequent *information extraction and reporting* activity {exc.ier} is a collection of tasks aimed at extracting information about the measurand from raw measurement data and representing it as a measurement result having a presentation format suitable to answer the questions pertaining to the measurement goal. In this crucial activity, raw measurement data are processed according to the defined measurement context model {exc.ier.dap} and measurement uncertainty is computed {exc.ier.unc} by assigning a value to each component identified in the modeling activity and combining the obtained results according to the defined model [18]. To this aim, while the application of probability theory is recommended by international organizations [9], researchers have been proposing the formalization of uncertainty evaluation also by means of non-probabilistic frameworks, particularly in the context of the mathematical theory of evidence [25], the fuzzy set theory [26], [27], or the theory of random-fuzzy variables [28].

According to [17], a measurement result is a "set of quantity values being attributed to a measurand together with any other available relevant information." In particular, the measurement result may be reported as a single measured quantity value {exc.ier.vla} and its standard uncertainty, or an interval of values with the related coverage probability, possibly together with the limits of validity of the information provided, as related to the influence properties and their allowed ranges, and other possible subsidiary information {exc.ier.sir}. More generally, such information might be represented as a probability distribution (or a fuzzy subset, or a plausibility-credibility distribution, ...) over the set of values of the measured property [25]–[28].

C. Interpretation **{int}**

The interpretation stage can be implemented either on-line or off-line. It is aimed at exploiting the information provided by measurement to establish whether or not the obtained information is valid {int.dcs.cvl} and the defined goal is met {int.dcs.cga}. To ensure that interpretation is focused on the measurement goal and does not fluctuate depending on measurement result, the framework suggests that the interpretation plan should be defined during the design activity, i.e., before the experimental activities are performed. In particular, the decision criteria and thresholds used to interpret the measurement results should be clearly defined.

For sure, due to measurement uncertainty, the decision result is an uncertain event and the risk of wrong decision is not null. For example, when the conformance or non-conformance of a product or service has to be decided, consumer and producer risks must be specified and evaluated [29], [30]. This task can be formally performed by means of probability theory only when measurement uncertainty has been evaluated. This is reasonably the main pragmatic reason why the GUM [18] states that "when reporting the result of a measurement, it is obligatory that some quantitative indication of the quality of the result be given so that those who use it can assess its reliability".

Measurement uncertainty must also be taken into account in the decision rule. This issue is tackled in a standard related to geometrical property measurements [31]. Its basic assumption is that those who want to prove conformance or non-conformance are responsible for the effect of measurement uncertainty on decision results. Clearly, this assumption may be adopted in any fact-based decision about conformance involving hard or soft measurements.

Finally, the interpretation stage can be aimed at acknowledging the lessons learned during the development of the measurement process and at preserving the acquired knowledge for future use {int.lrn.lsl}, e.g., in the form of new procedures or best practices, thus addressing decision making activities about possible improvements or refinements in both the measurement process and the use of measurement results. The result of the interpretation stage may lead to update not only the previous measurement activities, but even to reconsider measurement goals.

III. AN APPLICATION OF THE CONCEPTUAL FRAMEWORK TO STRONGLY DEFINED PROPERTIES: MEASUREMENT OF THE INPUT-OUTPUT CHARACTERISTIC OF AN ADC

Analog-to-Digital Converters (ADCs) can be tested for several reasons both by producers and consumers. To help in testing activities, technical standards have been prepared that guide in the selection of applicable procedures for measuring the device characteristic [32]. This measurement task is complex because ADCs are mixed-signal devices, often manufactured by using state-of-the-art technologies to maximize conversion speed at given resolution, under strict constraints regarding power consumption. As a consequence, testing cost can become a significant part of the device production costs [33], also because adopted testing technologies must pair, or be even more performing than those used to produce devices under test. To highlight how the framework proposed in Fig. 1 can support the design and implementation of a specific measurement process, the problem of *measuring the input-output characteristic of an ADC with the purpose of assessing its conformance against given specifications* is described in the following. In the considered example, the measurand (i.e., the set of ADC threshold levels) is modeled as a deterministic vector quantity, while the ADC exciting signal is assumed to be dynamic, often a sine-wave.

A. Planning: Goal setting, Modeling and Design {pln}

Goal setting {pln.gls}

The definition of the measurement purpose {pln.gls.prd} is straightforward in this case, being the assessment of the input-output characteristic of an ADC against given specifications, relating to, e.g., equivalent number of bits or INL values [32], where the ADC input-output characteristic might be used for compensating ADC errors, correcting for gain and offset, or to qualify devices after production.

The class of object under measurement {pln.gls.obs}, ADCs, is well known to experts. Moreover, in ADCs the relationship between the input analog signal and the output discrete values approximates a staircase transfer curve and the measurand is the set of voltages representing the ADC code transition levels. Thus, the general property to be measured {pln.gls.gpd} is a set of voltages; its definition and reference scale are both well established and do not need to be discussed here.

When qualifying devices, the target uncertainty for the measured transition levels {pln.gls.ctu} is related to consumer and producer risks involved by conformance assessments. Thus, constraints on these risks must be specified, together with a rule for deciding upon acceptance/rejection of devices. This, in turn, requires the choice of target uncertainty values for the considered transition voltages. As an example, if the absolute value of the INL is specified to be lower than 0.5 Least

Significant Bits (LSB), possible target uncertainty values in the measurement of INL must be in the order of 0.05 LSBs or less.

Modeling {pln.mdl}

ADCs are affected by other errors in addition to ideal quantization error. They can be either static or dynamic, depending on the rate of change of the digitized signal. Static errors, which include the quantization error, usually result from non-nominal spacing of the code transition levels. Dynamic errors are induced by the time variation of the input analog signal. These additional error sources include harmonic distortion arising in the analog input stages, signal-dependent variations in sampling times, dynamic effects in internal amplifiers and comparators, hysteresis phenomena and frequency-dependent variations in the code bin widths.

Regarding the definition of measurand {pln.mdl.mdf}, a code transition level is theoretically expressed as the value of the converter input voltage which causes half of the digital output codes to be greater than or equal to (and half less than) a given output code. However, it is not always possible to define a unique value for a particular code transition level [32], [34].

As for influence properties {pln.mdl.ipd}, ADC data sheets usually specify them and the related allowable ranges of variations. Typical examples are temperature, power supply voltages, clock frequency, and reference voltages. A proper function of the ADC must be assured when the influence factors fall within their specified ranges. Then, it is necessary to identify error sources and unwanted effects due to the employed instrumentation and algorithms. For instance, histogram methods can produce erroneous results if the ADC under test has output codes that are swapped with other codes or exhibits other types of non-monotonic behavior. To detect these situations and confirm that a non-monotonic behavior produces negligible effects, the analyzed converter is often also tested for Signal-to-Noise and Distortion (SINAD) performance [32]. Moreover, if the test signal is a full-scale sine-wave whose frequency is chosen large enough so that the ADC dynamic errors are significant, modelers must be aware that histogram testing is sensitive to some dynamic errors, while others will be averaged out [32], [34]. If given specifications are set for the INL at given signal input frequencies, then the transition levels are usually considered static DC values that are estimated using the Sinewave Histogram Test (SHT) or other techniques such as the Small Amplitude Triangular Waves (STW) [38]. In addition, the role of the measurement chain can be significant, as well as the modeling of the measuring system: the effect of disturbances along the whole measurement chain and the metrological characteristics of the adopted measurement instruments and signal sources must be taken into account. The empirical environment has also a significant role and a comprehensive mathematical model describing the contribution of the different error sources on the measured ADC transfer characteristic must suitably be defined [32], [34]: a laboratory in a production setting can offer a better control of influence properties than industrial laboratories. Also the definition of mutual interactions in the environments and between measuring instruments can be significant in terms of target uncertainty. Load effects and impedance matching are just two examples of effects due to mutual interactions between the device under test and the signal generators used to perform measurements.

Design {pln.dsg}

Different methods can be employed to determine ADC code transition levels, each one with pros and cons. Here, as measurement method {pln.dsg.mmd}, we consider the SHT [32], [34]–[37]. According to this approach, a histogram of code occurrences is generated in response to a pure sine-wave input signal with amplitude sufficient to slightly overdrive the ADC (i.e., the input signal range spans slightly more than the full-scale range of the ADC). Then, the obtained histogram allows the determination of the ADC transfer characteristic under the assumption that the ADC under test exhibits a monotonic behavior and is without hysteresis. To achieve the set 0.05 LSB target uncertainty all meaningful test design parameters must carefully be considered. This includes accuracy of the source signal generator, number of collected samples, amount of input-referred wide-band noise, type of processing algorithms, type of sampling (coherent/not coherent), choice of sampling over source frequencies, number of data records. All these parameters can influence the quality of the results and cannot be overlooked if 0.05 LSB is the target uncertainty [32], [34]-[37].

As an example, the performance of the employed measurement equipment must significantly exceed the expected performance of the ADC under test and at least one representative sample for every ADC code should be collected. In particular, coherent sampling allows the maximization of the number of distinct input phases that are sampled [32], [34].

Being able to design the SHT with this level of planned target uncertainty requires the usage of a complex model, identifiable using several design parameters. The description of this model is beyond the scope of this paper. Examples can be found in [34], [39], where all parameters of interest are linked to the desired target uncertainty and practical examples are fully carried out. The sensitivity of the target uncertainty to all SHT input parameters, such as number of samples or number of records, can be checked by using a software program implementing the measurement model. This will allow selection of the values of those parameters guaranteeing the achievement of the target uncertainty and pruning of the parameters not significantly influencing the test accuracy, when compared with the 0.05 LSB target uncertainty. Finally, all detailed information needed to actually perform the measurement can be put in a documented measurement procedure {pln.dsg.mpd}.

B. Execution: Set-Up, Data Acquisition, Information Extraction and Reporting {exc}

Once the planning phase has been carried out, actual measurement execution can take place. If the modeling and design phases were performed correctly, the execution phase becomes an instance of the designed test procedure. This requires some carefulness in choosing the proper instrumentation and adopting the correct measurement setup. With reference to the presented framework, the object under measurement, i.e., the ADC and its input and output pins can easily be detected {exc.stp.omd}. Also, the ADC manufacturer's recommendations must be followed when executing measurement. They can regard proper signal buffering and loading, input signal connections, transmission line matching, circuit layout patterns, power supply decoupling, and operating conditions. Edge characteristics for start-of-conversion pulse and clock must also be satisfied. To easily solve these problems, most manufacturers provide proper ADC evaluation boards. Thus, by following manufacturer's recommendations and the procedure defined in the planning phases {pln.dsg.mpd}, the instrumental chain is setup {exc.stp.mss} and calibration is performed {exc.stp.msc} when needed. Since transition voltages are to be measured, proper scales {exc.stp.ssv} and measuring systems must be used.

Next to the system setup, raw measurement data are taken, i.e., the ADC output codes are acquired {exc.daq.mda}. Hence, the obtained code frequencies of occurrence are processed by following the procedure defined in the planning stage {exc.ier.dap} and measurements of the ADC code transition levels are obtained. Also, various other ADC parameters, including integral and differential nonlinearity, missing codes, gain, and offset can then be measured from the values of the code transition levels.

Measurement uncertainty can be evaluated by using the mathematical model defined during the planning stage {exc.ier.unc} [32], [34]–[37], the values attributed to the test procedure design parameters (e.g., number of samples in a record) and the metrological characteristics of the adopted instruments. Subsidiary information can take the form of degrees of freedom in evaluating the standard uncertainty {exc.ier.sir}, chosen value for the coverage factor [18], or values for the influence properties. Finally, meaningful information is usually written in a documented report.

C. Interpretation {int}

To achieve the goal of assessing conformance or not conformance, a decision is taken whether the ADC can be qualified or not qualified {int.dcs}. Acceptance or rejection may be based on a specified bound, such as 0.5 LSB for the absolute value of the INL, set in the goal phase. To verify if this specification is met by the measured ADC, the uncertainty of 0.05 LSB must also be taken into account. To this aim, a suitable *coverage factor* and its corresponding confidence level must first be set according to the method described in [18]. Then, to prove conformance or not conformance the 0.5 LSB bound must be reduced by as much as the allowed error, given by the product of the 0.05 LSB uncertainty and the coverage factor [31]. This approach produces a new bound, lower than the original 0.5 LSB specification that will result in a most probable validation of the measurement procedure and in the achievement of the measurement goal, set at the beginning. Being this approach based on a probabilistic framework, risks of wrong decisions cannot be avoided entirely.

Validation and improvement {int.dcs.cvl} of the test procedure can always be carried out as supporting processes during off-line laboratory activities. This may require comparing results under different measurement setups, improvements in the model relating all relevant quantities and increase in the performance of the employed instrumentation. The very last phase aimed at the analysis of the “lesson learned” {int.lrn.lsl} can include increase in the awareness of the limits in the designed procedure and information useful for successive refinements of the same measurement on future occasions.

IV. AN APPLICATION OF THE CONCEPTUAL FRAMEWORK TO WEAKLY DEFINED PROPERTIES: ASSESSMENT OF SCIENTIFIC RESEARCH QUALITY

Knowledge development and scientific research are unanimously recognized as main drivers for the socioeconomic growth of nations. Intellectual assets are acknowledged as essential keys to value creation and increase in national gross domestic products [40]. To ensure the effectiveness of their investments, nations are increasingly resorting to research quality assessment programs, mainly aimed at ranking research institutions and funds recipients.

While the overall goal is extremely clear in its intentions, the assessment process has proved to be hard to be implemented and the achieved results are often criticized. Unfortunately, analysis of the structured activities that ensure the reliability of information provided by measurement has been largely ignored. As a result, sources that limit the quantity of achievable information are quite vaguely known and it is very hard to provide a sound evaluation of result uncertainty. Indeed, the concept ‘measurement’ is rightly avoided in these cases and the process is customarily denoted by terms like “assessment”, “evaluation”, or “appraisal”.

In order to stress the usefulness of the proposed methodology in the development of information gathering processes and the identification of similarities and differences between measurement and assessment of soft properties, in this Section the framework presented in Fig.1 is applied to analyze the activities performed when the property intended to be measured is research quality. In particular, potential sources of uncertainty in the process outcomes are highlighted.

A. Planning: Goal setting, Modeling and Design {pln}

<u>Goal setting {pln.gls}</u>

When dealing with research quality, various purposes can apply {pln.gls.prd}. For instance, information is gathered and analyzed to fund projects, to promote or hire people, to check the appropriate usage of resources, to rank organizations, to assess the societal impact of research outcomes. Purposes often depend on the parties expressing the needs: while governments are interested in assessing for accountability and funding reasons, research institutions may be interested for promotion or hiring reasons.

Purposes depend also on the class of objects that manifest the property of interest {pln.gls.obs}, as journal papers, research projects, research teams or whole research institutions. The description of the class should be sharp and complete enough to allow deciding whether a given object belongs to it or not. For example, possible class of objects are research papers in a given subject category, published in a specified set of journals, within a specified time interval.

Moreover, research assessments usually consider single research items. A research item represents a codified form of knowledge and can take different forms depending on communications habits of the considered research domain, such as: scientific journal paper, book, book chapter, letter, seminar, conference talk, poster, software product, prototype, patent and so on. Assessment procedures may consider a single item or composite sets of items depending on the object under assessment.

Normally, different general properties apply to objects of different classes {pln.gls.gpd}. For instance, the quality of the technical content can be a property of a scientific paper while the amount of funds collected in a given year can be a property charactering the quality of a research group.

According to the proposed framework, target uncertainty should be settled {pln.gls.ctu}. This is when the experimenter may realize that an important attribute of the measurement result may not easily be chosen. In fact, when assessing research quality, setting of target uncertainty is rarely explicitly accomplished even though it can implicitly be settled through the number of ordered levels chosen to express the provided results. For instance, while it might be considered reasonable to express the quality of a scientific paper using a scale based on 3 or 5 different levels, it is plausible that 10 levels convey a too high resolution, thus inducing wrong interpretations of the provided information.

Not addressing explicitly uncertainty issues is a major difference between assessment and measurement. This lack of information about uncertainty does not formally allow the evaluation of decisional risks when using assessment results in decision making processes.

Modeling {pln.mdl}

Regarding the general property definition, modeling activities should operatively define both concepts of scientific research and quality {pln.mdl.mdf} [41]. This latter concept is commonly intended as “the degree to which a set of interrelated characteristics fulfills requirements” [22]. In fact, quality has many distinct views and aspects and all the relevant ones should be identified, modeled and assessed. This is normally done by operatively defining properties (often called indicators) that act as approximations (proxies) of specific aspects, but that cannot convey information on the property as a whole. However, research quality is such a broad, multifaceted, and even fuzzy, concept that some of its determinants can be hardly captured by any set of operatively defined properties.

As for the different views of research quality, typically, a distinction is made between:

- *Conformance* of a research item to the standards implicitly defined by the reference community of scientists (national or international) over a specified period of time. For instance, different aspects to be considered for a bibliographic product include: originality, relevance (or potential influence) for the development of knowledge in the given field, methodological rigor, clarity, writing style, robustness of data and evidences on which conclusions are drawn.

- *Influence* on the scientific community, intended as effects produced by the considered research item on the development of knowledge in the given field. For a bibliographic product, it is typically assessed by means of the number of citations over a given time period. Conformance is usually a necessary but not a sufficient condition to guarantee the influence of an item. For instance, if a paper appears in a qualified scientific journal, it can be assumed that it satisfies the quality standards adopted by the scientific community, but not necessarily it will have an influence.
- *Impact* on society, i.e., the research returns on investments for the society, including the capability of solving societal problems, e.g., by promoting occupational and economic growth. Hence, unlike conformance and influence, impact requires the adoption of complex and uncertain cause–effect models.

The environment in which the assessment takes place has usually a relevant effect on the achieved result {pln.mdl.emd}. For example, when prestigious scientists and their research products are assessed by peers, there is the risk of biased judgments known as the Matthew effect ("the rich get richer and the poor get poorer") [42].

Various measurement methods can be applied when considering research quality, with peer review processes and bibliometric data analyzes playing a major role. Outcomes of peer review processes depend on competency of reviewers. In addition they can be subjected to distortions due, for example, to opportunistic behaviors induced by conflicts of interest or to prejudices about specific characteristics of reviewed researchers, such as sex, personal reputation, affiliation to prestigious institution.

As far as bibliometric data are concerned, it is often assumed that citation indexes are less influenced by subjectivity or opportunistic behaviors than peer review processes. However, citations are used not only to recognize the importance of a research outcome, but also to stigmatize low quality research approaches or inconsistencies in the results. Also, high quality outcomes, but in a very narrow field or related to frontier problems, may have low influence because outside the disciplinary mainstream, to which higher influence research items usually belong. Moreover, even citations may be affected by opportunistic behaviors.

Furthermore, research quality assessments may push researchers at improving their score rather than contribute to the advance of knowledge: as it happens in the domain of physical properties, observing an object may be a source of undesired interaction effect. For instance, if only papers in journal papers are assessed, researchers could be discouraged to participate at scientific congresses thus reducing, in turn, the fundamental role on the advancement of knowledge of public discussion and reasoning about research outcomes. This is a well-known phenomenon occurring when human beings are involved as objects of measurement. Indeed, their adaption capability to stated measurement goals (often known as "Hawthorne effect") forces a reaction that can modify the properties under measurement up to a point such that measurement may be exploited to drive people to a desired behavior instead of to acquire information.

<u>Design {pln.dsg}</u>

Only the research quality aspects that are internal to the scientific community, i.e., conformance and influence, are considered in the following for the sake of conciseness.

As for conformance, once the object undergoing assessment and the different properties that contribute to its conformance are defined, a comparison between these properties and the standards adopted in the scientific community has to be performed. Unlike hard measurement, in this case the standards, i.e., the reference scale, are not explicitly and operatively defined. Their knowledge only pertains to members of the given scientific community, who also have the capability of identifying the relevant elements of a research product and comparing them to standards. It is assumed that this capability has been acquired by scientists as a long-term result of the development of their own research. According to this approach, conformance assessment can be carried out only by members of the scientific community (the peers), using reviewers' panels. The involvement of persons acting as measuring systems poses fundamental issues about the reliability of the achieved information and suitable criteria need be followed in the selection of reviewers' panels, e.g.: involve reviewers with expertise in the considered subject, avoid conflict of interest and preventing interactions among reviewers to ensure independent reviews.

Conversely, influence of a research product is normally assessed by using citation indexes. Indeed they are considered good proxies of this quality aspect. Sometimes, other types of credits are employed such as award or invited talks. Details concerning data acquisition processes may play a significant role and their comprehensive analysis is always required. For instance, the list of publications containing valid citations must defined and decision about inclusion of self–citations must be taken.

When performing a retrospective assessment (e.g., a department research quality assessment) two approaches can be adopted: ad hoc peer review or information available as outcome of previous peer review processes done for different purposes (e.g., bibliometric data analysis).

Both approaches, somewhat improperly called "qualitative" and "quantitative" respectively, have strengths and weaknesses [43]. Thus, joint peer-review-based and indicator-based assessments often provide the best solution {pln.dsg.mmd}.

Finally, it is worth remarking that a single value composite property defined as combination of different indicators is often used when research quality assessment outcomes are used for ranking purposes. However, if the relative importance of the different aspects underlying the considered research quality concept is not well understood, this approach should be avoided [43].

As for the reference scale design, research quality assessments are normally performed on ordinal scales. Because of unavoidable (definitional, instrumental, and interaction) uncertainties sources above discussed, often difficult to estimate, only a very limited number of scale elements can guarantee trusted results. The characteristics of each element of the scale should be clearly specified during the design phase. For instance, when peer-review assessments are concerned, an evaluation form should be predefined so helping the reviewers in detecting the aspects to be assessed and standardizing their assessment results.

Whether peer-review-based or indicator-based processes have to be performed or not, clear and operative assessment procedures need to be provided as a result of the design stage {pln.dsg.mpd}.

B. Execution: Set-Up, Data Acquisition, Information Extraction and Reporting **{exc}**

If the modeling and design activities have properly been managed, the execution stage does not usually imply complex activities, and can often be performed even automatically in the case of indicator-based assessments. During this stage, previously planned data processing procedures or decision rules might be applied to compose contrasting reviews or to exclude inconsistent results. Overall, data collection {exc.daq.mda} tasks may involve a large set of tools and stakeholders. During the information extraction and reporting activity, all collected data are processed and prepared for public disclosure {exc.ier.dap}.

At present, there is no evidence of any assessment program providing estimates of result uncertainties {exc.ier.sir}. Nevertheless, uncertain outputs may be due to different sources highlighted above, among which definitional uncertainty, uncertain inputs, environment effects or variability associated to peer–review judgments.

C. Interpretation **{int}**

This last stage includes activities aimed at verifying the achievement of the predefined purposes, possibly together with the validation or the improvement of the whole process {int.dcs.cvl}. According to outcomes, a report could be written to highlight the strengths and weaknesses of the overall assessment procedure. Also, assessment programs are modified over time to account for lessons learned and for changes in the measurement environment {int.lrn.lsl}. It is also worth noticing that mature assessment programs are based on pilot studies carried out to test procedures and to validate processes.

Concluding, this Section highlighted that the proposed framework may be an enabling tool for reasoning about research quality assessment, focusing on questions to answer to achieve a measurable property. Indeed, as any measurement of physical properties, assessing research quality implies the construction of conceptual models prior the execution of any experimental activity. However, research quality assessment differs from measurement in at least three aspects:

uncertainty not explicitly evaluated, reference scale implicitly defined, comparison with standards performed in a subjective way.

V. CONCLUSIONS

In its nature of experimental information gathering process, measurement can be characterized by means of a meta-model building upon the functional dependence of measurement output (the measurement result) on measurement inputs (the measurand, the properties realized by the measurement standards, the influence properties, and the several sources of knowledge exploited in the definition of a measurement model). According to this approach and building on authors' experience about measurements of physical properties, this paper has proposed and discussed a conceptual framework describing the development of measurement as a process hierarchically structured in terms of stages composed by activities performed through multiple tasks. This is the analytical basis on which a systemic interpretation of measurement development is grounded, as emphasized by the presence of continuous adjustment / refinement feedback in the process. The framework offers structured guidelines to plan, execute, and interpret the results of an information gathering process whose result reliability can be quantified by a suitable parameter, typically called *measurement uncertainty*. How this generally and effectively applies also to weakly defined measurements is a subject of future explorations, that should keep into account not only the differences in the internal structure of soft measurements but also the ways experimental knowledge acquisition itself is differently intended there.

The framework can be also intended as a tool for guiding a meta-process aimed at transforming generic information gathering processes into measurements, thus making the framework as an operational tool for an at least partial characterization of the very concept of measurement. For example, as the framework highlights, a process that does not include a stage of uncertainty evaluation should not be considered a measurement. Also, the framework shows that models are unavoidable in measurement, either stated explicitly – because of technical protocols, mandatory clauses, contractual requirements, best practices, etc – or left implicit.

The framework shows that, at least in the engineering domain the following conditions are required for an information gathering process to be called a measurement: a clear purpose definition; a clear and validated measurement scale; a valid model; the capability to estimate measurement uncertainty; an experimental outcome; the traceability of results. If any of these conditions is not fulfilled, a process can hardly be considered a measurement. The fact that several cases of measurements, particularly in everyday life situations, do not explicitly include all these components is not an argument against the framework, but it only shows how reliably the principled complexity of measurement can be operatively tackled by means of socially accepted tacit assumptions, such as leaving measurement uncertainty implicit whenever target uncertainty is sufficiently greater than instrumental uncertainty. The presence in the framework of the stages of planning and interpretation emphasizes that execution is a necessary but not sufficient stage of measurement, thus making it concrete the claim that generally “data do not speak by themselves”

We are aware of the limits and weaknesses of this framework. For example, when dealing with multivariate measurands, multiple criteria decision-making activities are required. Then, not only the mathematical apparatus grows in complexity (e.g., measurement uncertainty formalized by multidimensional probability distributions or covariance matrices), but further dimensions of feedback among the several process activities have to be taken into account, thus making the proper management of the whole process a highly challenging activity. An extended version of this development methodology might include guidelines for managing measurements in such encompassing scenarios. The lack of an explicit, pragmatic guidance about which development tasks should be prioritized and which can be omitted in any given practical situation by recurring to default options is a major missing point of the framework, which in a future version might be dealt with by means of complementary materials to support the implementation, such as templates, best practices, and examples, if not even actual measurement development procedures.

On the whole, the strategic perspective we are aimed at in the evolutionary path of the framework is to contribute to the acknowledgment that complex measurements, as more and more required by our society, have to be developed according to systemic principles.

REFERENCES

[1] P.H. Sydenham, Structured understanding of the measurement process; Part 1: Holistic view of the measurement system, Measurement, 3, 115–120, 1985; Part 2: Development and implementation of a measurement process algorithm, Measurement, 3, 161–168, 1985.

[2] F. Abdullah, L. Finkelstein, S.H. Khan, W.J. Hill, Modeling in measurement and instrumentation – An overview, Measurement, 14, 41–53, 1994.

[3] P.H. Sydenham, Relationship between measurement, knowledge and advancement, Measurement, 34, 3–16, 2003.

[4] F. Crenna, R.C. Michelini, G.B. Rossi: Hierarchical intelligent measurement set-up for characterising dynamical systems, Measurement, 21, 3, 91-106, 1997.

[5] J. Michell, Measurement in psychology – Critical history of a methodological concept, Cambridge University Press, 1999.

[6] M. Boumans, Measurement in economics – A handbook, Academic Press, 2007.

[7] L. Finkelstein, Widely, strongly and weakly defined measurement, Measurement, 34, 39–48, 2003.

[8] L. Finkelstein, Problems of measurement in soft systems, Measurement, 38, 267–274, 2005.

[9] L. Mari, P. Carbone, D. Petri, Measurement fundamentals: A pragmatic view, IEEE Trans. Instrum. Meas., 61, 8, 2107–2115, 2012.

[10] BIPM, Evolving needs for metrology in trade, industry and society and the role of the BIPM, International Bureau of Weights and Measures, 2007.

[11] D. Krantz, R.D. Luce, P. Suppes, A. Tversky, Foundations of measurement, Academic Press, 1971, 1989, 1990.

[12] L. Narens, Abstract measurement theory, MIT Press, 1985.

[13] ISO 9001:2008, Quality management systems: Requirements, International Standardization Organization, 2008.

[14] S.H. Kan, Metrics and models in software quality engineering, 2nd edition, Pearson education, 2003.

[15] G.B. Rossi, Measurability, Measurement , 40, 545–562, 2007.

[16] P. Carbone, L. Buglione, L. Mari, D. Petri, A comparison between foundations of metrology and software measurement, IEEE Trans. Instrum. Meas., 57, 2, 235–241, 2008.

[17] JGCM 200:2012, International Vocabulary of Metrology – Basic and general concepts and associated terms (VIM), Joint Committee for Guides in Metrology, 2012 (2008 edition with minor corrections), http://www.bipm.org/en/publications/guides/vim.html.

[18] JCGM 100:2008, Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM, originally published in 1993), Joint Committee for Guides in Metrology, 2008, http://www.bipm.org/en/publications/guides/gum.html.

[19] N.E. Fenton, S.L. Pfleeger, Software metrics: A rigorous and practical approach, PWS Publishing, 1996.

[20] Y. Akao, QFD: Quality Function Deployment – Integrating customer requirements into product design, Yoji Akao Editor, 2004.

[21] A. Giordani, L. Mari, Measurement, models, and uncertainty, IEEE Trans. Instrum. Meas., 61, 8, 2144–2152, 2012.

[22] UNI EN ISO 9000:2005, Quality management systems – Fundamentals and vocabulary, 2005.

[23] D.D. Harris, P.H. Sydenham, PRO-MINDS: Development of a software tool to support the measurement system designer, Measurement, 15, 1, 1-14, 1995.

[24] ISO/IEC 14598-1:1999, Information Technology – Software Product Evaluation – Part 1: General overview, International Standardization Organization, 1999.

[25] A. Ferrero, S. Salicone, Fully comprehensive mathematical approach to the expression of uncertainty in measurement, IEEE Trans. Instrum. Meas., 55, 3, 706–712, 2006.

[26] G. Mauris, L. Berrah, L. Foulloy, A. Haurat, Fuzzy handling of measurement errors in instrumentation, IEEE Trans. Instrum. Meas., 49, 1, 89–93, 2000.

[27] G. Mauris, V. Lasserre, L. Foulloy, Fuzzy modeling of measurement data acquired from physical sensors, IEEE Trans. Instrum. Meas., 49, 6, 1201–1205, 2000.

[28] A. Ferrero, S. Salicone, The random-fuzzy variables: A new approach to the expression of uncertainty in measurement, IEEE Trans. Instrum. Meas., 53, 5, 1370–1377, 2004.

[29] JCGM 106:2012, Evaluation of measurement data – The role of measurement uncertainty in conformity assessment, Joint Committee for Guides in Metrology, 2012, http://www.bipm.org/utils/common/documents/jcgm/JCGM_106_2012_E.pdf.

[30] OIML, The role of measurement uncertainty in conformity assessment decisions in legal metrology, 2009, http://www.oiml.org/download/cds/tc3_sc5_p2_1cd_uncertainty.pdf.

[31] ISO 14253-1:1998, Geometrical Product Specifications (GPS) – Inspection by measurement of workpieces and measuring equipment – Part 1: Decision rules for proving conformance or non-conformance with specifications, International Standardization Organization, 1998.

[32] IEEE 1241, IEEE Standard for terminology and test methods for analog-to-digital converters, 2000.

[33] K. Parthasarathy, T. Kuyel, D. Price, L. Jin, D. Chen, R. Geiger, BIST and production testing of ADCs using imprecise stimulus, ACM Trans. on Design Automation of Electronic Systems, 8, 4, 522–545, 2003.

[34] J. Blair, Histogram measurement of ADC nonlinearities using sine waves, IEEE Trans. Instrum. Meas., 43, 3, 373–383, 1994.

[35] P. Carbone, E. Nunzi, D. Petri, Statistical efficiency of ADC sine wave histogram test, IEEE Trans. Instrum. Meas., 51, 4, 849–852, 2002.

[36] P. Carbone, D. Petri, Noise sensitivity of the ADC histogram test, IEEE Trans. Instrum. Meas., 47, 4, 1001–1004, 1998.

[37] F. Stefani, D. Macii, A. Moschitta, P. Carbone, D. Petri, Simple and time-effective procedure for ADC INL estimation, IEEE Trans. Instrum. Meas., 55, 4, 1383–1389, 2006.

[38] P. Carbone, S. Kiaei, F. Xu (eds), Design, modeling and testing of data converters, Springer, 2014, ISSN 1860-4862.

[39] F. Alegria, P. Arpaia, A. C. Serra, P. Daponte, ADC histogram test by triangular small-waves, proc. Instrumentation and Measurement Technology Conference, 2001. Budapest 2001, 1690-1695.

[40] Innovation and growth rationale for an innovation strategy – Organization for economic co–operation and development, 2007. http://www.oecd.org.

[41] Research Excellence Framework REF 2014, http://www.ref.ac.uk.

[42] R. K. Merton, "The Matthew Effect in Science," Science, vol. 159, pp. 3810, Jan. 5, 1968.

[43] Research Quality Framework: Assessing the quality and impact of research in Australia. Quality metrics, Sept. 2006.