IMEKO TC6 International Conference on Metrology and Digital Transformation
September 19 – September 21, 2022, Berlin, Germany

REPRESENTING METROLOGICAL TRACEABILITY IN DIGITAL SYSTEMS

Blair D. Hall [a,†,*] *and Ryan M. White* [b,‡]

[a]Measurement Standards Laboratory of New Zealand
[†]ORCID: `0000-0002-4249-6863`
[b]National Research Council, Canada
[‡]ORCID: `0000-0003-3589-5900`
[*]Corresponding author. Email: `blair.hall@measurement.govt.nz`

Abstract – When metrological traceability is considered in the context of modern digital systems, a model of the staged collaborative nature of traceable measurements emerges. The key elements of this model are instances of external factors that influence the measurement, and intermediate measurement stage results. Digital implementation of the uncertainty calculation method described in the *Guide to the expression of uncertainty in measurement* (GUM) requires unique identification of individual influence factors and intermediate results. Systems using the model presented here can implement more rigorous GUM-compliance and provide more valuable data than is currently possible to consumers of traceable measurements.

Keywords: Metrological traceability, measurement uncertainty, calibration, quality assurance, digital transformation

1. INTRODUCTION

When referring to the modern origins of metrological traceability, a paper by Belanger at the National Bureau of Standards is usually cited: *Traceability: an evolving concept* [1]. This work pre-dates the *Guide to the expression of uncertainty in measurement* (GUM) [2], the *International Vocabulary of Metrology* (VIM) [3], and the International Committee for Weights and Measures (CIPM) Mutual Recognition Arrangement (MRA) [4], which now underpin implementations of traceability. Sophisticated national and international networks deliver traceable measurement services wherever they are needed in society. Collectively, this is called the measurement quality infrastructure.

The international metrology community is now embarking on a digital transformation of measurement quality infrastructures. The myriads of processes currently carried out by skilled individuals will be digitalised. This will allow efficiencies and new services to be realised; and one day it may allow machines to take over much of the work involved. But to do this, the essence of each process must be clear: digital systems require a rigorously logical description of tasks. Foundational concepts must be understood and digital formats that capture and represent these concepts must be developed and widely accepted. To this end, the CIPM, which directs metrological activities carried out by parties to the Metre Convention, has established a task group on the "Digital SI" to look for uniform and unambiguous formats for information exchange (108th meeting of the CIPM, Decision CIPM/108-28, Oct 2019).[1]

Metrological traceability is an important foundational concept. There is ample documentation on how skilled people can implement traceability but it is not yet clear how this notion will be handled in the digital world. Existing workflows often appear complicated, and may require a significant amount of tacit knowledge to be executed correctly. The purpose of this article is to focus attention on where efforts to digitalise traceability should begin. We highlight a mathematical, and therefore logical, structure that provides a conceptual foundation for organising and managing additional information about a measurement.

In the next section we discuss where and how value is created by traceable measurements. We argue that this occurs at the end of a traceability chain where information about the physical world will inform decisions. In section 3, we discuss the structure of traceability chains and consider the meaning of calibration in the context of traceability. In section 4, we look at how the essence of a traceability chain may be captured in a form suitable for digitalisation. We conclude with some final remarks in section 5.

2. END USERS BENEFIT FROM TRACEABILITY

It is important to understand where value is added by the provision of traceable measurement services and what it is that is valuable. Substantial public and private funds go into quality infrastructures, but why?

Value is realised at the end of a chain. Starting with the definitions for units of measure, national metrology institutes (NMIs) realise standards and, using these, calibrate instruments and artefacts for 2nd-tier calibration laboratories. Those laboratories, in turn, provide calibration and testing services to their customers. The chain continues to grow until an ‘end user’ finally makes a measurement that, instead of being passed on, is used to inform some sort of decision. It is this final decision-making that is valuable: in society, critical decisions informed by measurement are expected to be reliable. However, only traceable measurements can provide the high level of confidence required by

---

[1] Both authors are involved in activities that support the Digital SI task group. However, the opinions expressed here are of the authors alone.

end users. So, traceability addresses society's need for reliability and trust: that aeroplanes will fly, that commercial measurements are fair, that radiation levels are within safety limits, etc, etc. The trustworthiness of traceable measurements is worth paying for.[2]

Belanger emphasised that traceability is intended to ensure measurement results are of adequate accuracy [1]. But adequate for what exactly? Those who participate in the upstream stages of a measurement know nothing about an end-user's requirements. Belanger discussed this, but did not give a succinct answer. It is now possible to do so, by focusing on the need for decision-making informed by measurement at the end of a traceability chain.

In measurement, the quantity of interest (the measurand) can never be determined exactly. Sometimes an instrument may appear to satisfy the end-user's need for accuracy, by producing results that differ by a negligible amount from the measurand. However, that is a misleading impression: as Belanger explains, it is the measurement (process) *per se* that determines accuracy; traceability should not be attributed to the instruments or artefacts used [1].

Although measurement cannot determine the measurand, Y, an estimate, $y \approx Y$, is obtained. The following simple relation then holds:

$$Y = y - E_y \ , \tag{1}$$

where E_y is called the measurement error. Should the accuracy of a measurement be deemed adequate, there must be some sense in which E_y is considered small enough to satisfy a specific need. Yet, like Y, E_y cannot be determined exactly; at best, a probabilistic description of the likely values for E_y may be known. Usually the best estimate of E_y is zero and the extent to which E_y may differ from zero is expressed in a statement of uncertainty.

Now consider a simple decision that requires knowledge of Y. Suppose action must be taken if Y exceeds a threshold $y_{\max}$. A measured value y will be the best available estimate. However, experimental error may be such that $y \geq y_{\max}$ when Y is actually less than $y_{\max}$ or, alternatively, that $y < y_{\max}$ when Y is greater than $y_{\max}$. These situations lead to undesirable decision outcomes (a false-positive or a false-negative). To design a decision process with acceptably low rates of undesirable outcomes, probabilistic information about E_y will be useful.[3] Statistical inference and decision theory are not considered here, but note that: 1) at the end of a traceability chain, measurement results inform decisions, and 2) the reliability of outcomes depends on the accuracy of data about y and E_y.

3. TRACEABILITY AND CALIBRATION

Belanger's preferred definition of traceability includes a statement (slightly abridged here) that measurements are only traceable

> ... if scientifically rigorous evidence is produced on a continuous basis to show that the [measurement produces] results for which the total measurement uncertainty [relative to national standards] is quantified.

That is, the probabilistic behaviour of E_y must be captured rigorously and objectively in a statement of uncertainty to implement traceability. However, the staged nature of measurement along a chain is not apparent in this definition. The current definition of metrological traceability does, however, refer to a chain of calibrations [3, §2.41]:

> property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty.

Belanger also considered a definition with a chain of calibrations but did not favour it, because the characteristics of measuring instruments and calibration standards were emphasised, and also the definition did not mention uncertainty. We agree that the term calibration may be misunderstood, but the notion of an unbroken chain of measurement stages is fundamental to traceability.

Returning to a description of measurement, the GUM introduces a mathematical measurement model, or function, for the measurand

$$Y = f(X_1, X_2, \cdots, X_l) \ , \tag{2}$$

where the inputs $X_1, X_2, \cdots, X_l$ are external factors that influence the outcome of a measurement.[4] Like Belanger's preferred definition, this gives no sense of the stages involved. However, GUM notation may be modified slightly to include stages. Any function $f(\cdot)$ can be decomposed into an arbitrary sequence of intermediate steps, $h = 1, \cdots, m$, each described by a particular function

$$Y_h = f_h(\Lambda_h) \ . \tag{3}$$

The set of inputs to step h, here denoted Λ_h, may include any previous outputs, $Y_1, \cdots, Y_{h-1}$ and any inputs $X_1, \cdots X_l$. When the sequence of functions is evaluated, the final step yields $Y = Y_m(\Lambda_m)$.

Figure 1 shows a four-stage measurement, which could be interpreted as follows. Stages 1 and 2 calibrate a pair of references. Those references, Y_1 and Y_2, are used to determine an instrument calibration factor, Y_3, in stage 3

---

[2] Some types of decision that may be familiar to readers are: conformity testing, proficiency testing and hypothesis testing. However, the discussion need not be restricted to these.

[3] For instance, a simple guard band could lower the threshold to $y < y_{\max} - a$, with information about the distribution of E_y used to set a.

[4] Among these inputs will be terms representing nuisance factors that perturb the outcomes of measurement, such as Johnson noise in a resistor.

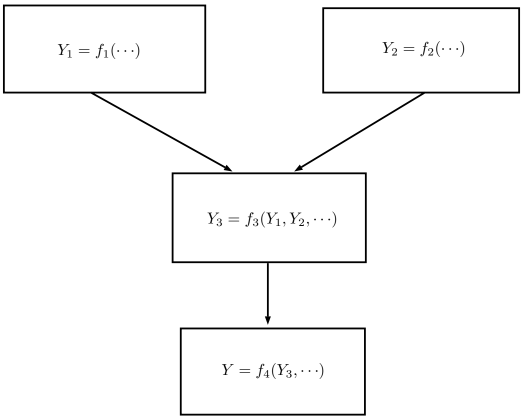

Fig. 1: A staged measurement. The unspecified function arguments '$\dots$' represent external influence quantities (any of the model inputs $X_1, \cdots, X_l$).

then, in stage 4, the instrument measures Y. The sequence of steps is equivalent to a model

$$f(X_1, \cdots) \equiv f_4(\, f_3(\, f_1(\cdots), f_2(\cdots), \cdots), \cdots) \,, \quad (4)$$

where the arguments shown as '$\dots$' on the right may be any external influence quantities, $X_1, \cdots, X_l$.

Evaluating uncertainty in a staged measurement has been covered in an earlier paper [5] and is also discussed at this conference [6], allowing us to elide those details. Calibration, on the other hand, deserves further attention.

White *et al.* give a clear pragmatic account of what, in the context of traceable measurement, is intended by calibration [7]. They also give an excellent account of some common misconceptions associated with traceability and calibration. The authors explain that processes are needed to monitor measurement performance continuously, to establish evidence for reliability over time. This evidence supports the assumption of an enduring accuracy in calibration results, and resonates with Belanger's definition of traceability, as well as the current one.

White *et al.* frame their discussion in terms of two stages: a provider of calibration services and a consumer, which is a common view reinforced by the way quality infrastructures currently operate. However, a two-stage view of measurement is too narrowly focused. As already explained, the entire chain is important because all upstream influences can affect the final stage. Moreover, it may not be immediately obvious when the end of a chain has been reached. Suppose, for instance, that two apparently end-user measurements are subsequently compared as part of a simple decision process. Traceability should not stop at the individual measurement results; it should, for example, be possible to evaluate the uncertainty in a difference, or ratio, of results used to inform a decision.[5] Apparently small additional stages can dramatically affect a traceability 'chain' if there are common influences in intermediate results. The current definition of traceability falls short here, because only calibrations in a simple chain are recognised; the complexity that arises in practice is not acknowledged.

4. DIGITALISATION OF TRACEABILITY

A continuous chain of measurements and calibrations provides a conceptual framework for modelling traceability and organising metadata. There are two sets of elements that constitute a chain: stage outputs, $Y_1, \cdots, Y_m$, associated with with equation (3); and model inputs, $X_1, \cdots, X_l$, associated with equation (2). Only stage outputs are shown in figure 1; however, external inputs could be connected to the stages and shown with a different shape of node.

Another important point is that the measurement in figure 1 produces a unique estimate of Y. Repeating the measurement changes the accompanying measurement error, E_y in equation (1). That is why the data model discussed in [6] requires unique digital identifiers for each instance of the elements X_i and Y_h. Appropriate use of identifiers will capture the structure of related stages and account for systematic effects when the uncertainty is evaluated [5].

Two approaches to uncertainty evaluation should be considered, shown schematically in figures 2 and 3. Figure 2 depicts a GUM calculation applied to a staged measurement model. This is preferred because it satisfies rigorous GUM requirements for transferability and internal consistency [2, §0.4] [5]. However, conventional reporting uses a simpler approach, shown in figure 3. In these figures, the indices of stage outputs are shown along the left-hand side and indices of model inputs are along the top. Horizontal lines represent the stages and a dot is placed on a line when a component of uncertainty is evaluated for an influence factor (the pattern of dots is just an example).

The two approaches are clearly different. In figure 2, external influences give rise to components of uncertainty that are propagated between stages. In contrast, the conventional approach does not propagate these individual influences. Instead, the combined uncertainty of one stage becomes a new independent component of uncertainty in the next stage. These calculations will only produce the same result if all influence factors are independent [5].

To satisfy the purpose of traceability—to deliver scientifically objective data about measurement accuracy—a conceptual model must incorporate the staged nature of measurement and explicitly recognise the stage outputs (the Y_h) and external influences (the X_i) as elements. This may seem unnecessary to some readers, because current practice does not do this. However, that practice has evolved out of pragmatic compromises in paper-based quality infrastructures. Digital technology removes those constraints and of-

---

[5] Often the influence of significant common factors can be attenuated or even eliminated in a comparison of results. One of the benefits offered by digitalisation is the possibility to account for such situations correctly, because at present insufficient information is available.

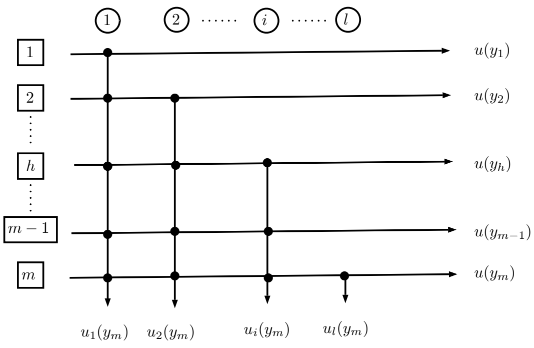

Fig. 2: Schematic representation of a rigorous GUM uncertainty calculation for a traceability chain. Boxed indices indicate stages (the Y_h) and circled indices indicate external influences (the X_i). A dot represents a component of uncertainty (terms for final-stage components are shown along the bottom). Combined standard uncertainties (the $u(y_h)$) are on the right.

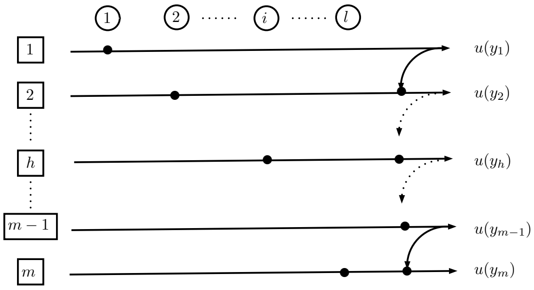

Fig. 3: Conventional reporting practice only passes the combined uncertainty from one stage to the next. This leads to a different calculation of uncertainty than the model shown in Fig. 2.

fers obvious potential benefits. For instance, in a particular measurement, the critical factors that limit accuracy at the end of a chain could be identified, even if those factors arise during a prior stage. Such information would enable global optimisation of traceable measurement services.

5. FINAL REMARKS

This work has discussed how the concept of traceability could be modelled in digital systems. Focusing on fundamental aspects, we conclude that details about the contribution to uncertainty from measurement errors in staged measurements is a foundational concept that should be supported. This addresses the need to deliver scientifically rigorous information about the accuracy of a measurement. The resulting model is compatible with current practice but allows digital systems to evolve towards more rigorous GUM-compliance, which would deliver superior data handling and support improved machine-based inference and decision-making at the end of traceability chains.

Our analysis underlines the uniqueness of certain model elements. So, a satisfactory means of generating and managing unique digital identifiers, with associated data, will be needed to represent these elements. We do not speculate how that could be done but note that satisfying the FAIR principles should not be difficult to achieve. We also anticipate, for example, that stage identifiers could provide access to supplementary information about processes during a stage, such as quality assurance needed to satisfy documentary requirements for metrological traceability.

Finally, the reader may be concerned that *people* at the end of traceability chains would find additional data a burden. We agree. Indeed, it is surely for this reason that paper-based workflows have not implemented the GUM guidelines. Nevertheless, traceable measurement is clearly valued *per se*. So, when digital technology can sweep aside perceived difficulties and handle the details, users will surely applaud the improvement in accuracy and reliability afforded by better infrastructure.

ACKNOWLEDGEMENTS

BDH thanks P. Saunders and V. Bubanja for reviewing the manuscript and D. R. White for useful discussions. This work was partly funded by the New Zealand government.

REFERENCES

[1] B. C. Belanger, “Traceability: an evolving concept,” *ASTM Standardization News*, vol. 8, no. 1, pp. 22–28, 1980.

[2] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML, *Evaluation of measurement data – Guide to the expression of uncertainty in measurement JCGM 100:2008 (GUM 1995 with minor corrections)*, 1st ed. Paris, Sèvres: BIPM Joint Committee for Guides in Metrology, 2008.

[3] BIPM, IEC, IFCC, ILAC, IUPAC, IUPAP, ISO, and OIML, *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)*, 3rd ed. Paris, Sèvres: BIPM Joint Committee for Guides in Metrology, 2012.

[4] T. Quinn, “The metre convention and world-wide comparability of measurement results,” *Accred. Qual. Assur.*, vol. 9, pp. 533–538, 2004.

[5] B. D. Hall and D. R. White, “Digital representation of measurement uncertainty for metrological traceability,” in *Advanced Mathematical and Computational Tools in Metrology and Testing XII.* Singapore: World Scientific, 2021, vol. 90, pp. 262–272.

[6] B. D. Hall, “The Uncertain Number: a data model for measurement,” in *IMEKO TC6 International Conference on Metrology and Digital Transformation.* IMEKO, September 19–21 2022, *submitted.*

[7] D. R. White, K. Jones, J. V. Nicholas, and C. M. Sutton, “Designing Calibrations in a Quality Environment,” in *AIM-TEC 94: Third Australasian Instrumentation and Measurement Conference; Test and Evaluation in the Asia-Pacific Region.* Institution of Engineers, Australia, 1994, pp. 197–222, (accessed 20 March 2022). [Online]: https://www.researchgate.net/publication/359337431_Designing_calibrations_in_a_quality_environment.