**NISTIR 6025**

# Metrology for Information Technology (IT)

May 1997

Prepared for
MEL/ITL Management

by
MEL/ITL Task Group on Metrology for Information Technology (IT)

**Comment [COMMENT1]:** This page intentionally left blank

## Table of Contents

Preface.....

Introduction.....
Scope.....
Definitions .....

Establishing a Conceptual Basis for IT Metrology .....
Principles of Physical Metrology.....
Principles of IT Metrology .....

Methods of Testing for Digital IT Systems Quantities.....

Status and Opportunities for IT Metrology .....

Roles for NIST in IT Metrology .....

Conclusions.....

Annex A: References .....

Annex B: Glossary of Abbreviations.....

Annex C: Examples of Present IT Metrology at NIST.....

# Preface

In May 1996, NIST management requested a white paper on metrology for information technology (IT). A task group was formed to develop this white paper with representatives from the Manufacturing Engineering Laboratory (MEL), the Information Technology Laboratory (ITL), and Technology Services (TS). The task group members had a wide spectrum of experiences and perspectives on testing and measuring physical and IT quantities. The task group believed that its collective experience and knowledge were probably sufficient to investigate the underlying question of the nature of IT metrology. During the course of its work, the task group did not find any previous work addressing the overall subject of metrology for IT. The task group found it to be both exciting and challenging to possibly be first in what should be a continuing area of study.

After some spirited deliberations, the task group was able to reach consensus on its white paper. Also, as a result of its deliberations, the task group decided that this white paper should suggest possible answers rather than assert definitive conclusions. In this spirit, the white paper suggests: a scope and a conceptual basis for IT metrology; a taxonomy for IT methods of testing; status of IT testing and measurement; opportunities to advance IT metrology; overall roles for NIST; and recapitulates the importance of IT metrology to the U.S.

The task group is very appreciative of having had the opportunity to produce this white paper. The task group hopes that this white paper will provide food for thought for our intended audience: NIST management and technical staff and our colleagues elsewhere who are involved in various aspects of testing and measuring IT.

Task Group Members:

Lisa Carnahan (ITL)
Gary Carver (MEL)
Martha Gray (ITL)
Mike Hogan (ITL), Convener
Theodore Hopp (MEL)
Jeffrey Horlick (TS)
Gordon Lyon (ITL)
Elena Messina (MEL)

# Introduction

## Scope

The scope of this white paper is the testing or measuring of digital information technology (IT) systems attributes or properties; the use of digital IT systems in testing and measuring; and the underlying mathematical, computational, and statistical sciences used in testing and measuring. This paper suggests a conceptual basis for IT metrology; reviews IT testing methods, the status of IT metrology, and opportunities for advancing IT metrology; and notes possible roles for NIST.

One goal of this white paper is to apply the concepts of metrology to IT systems. Another goal is to relate measurements in IT to established concepts of traceability.

## Definitions

### Information Technology (IT)

Information Technology (IT) is a relatively recently coined term for referring to several industry sectors whose boundaries are increasingly fuzzy: computing, telecommunications, and entertainment. A generic, functional definition of IT is the storage, processing, transfer, display, management, organization, and retrieval of information. IT can be characterized as increasingly digital. IT systems are typically a blend of hardware and software. The hardware can be characterized as increasingly complex and difficult to manufacture. The software can be characterized as increasingly complex and difficult to develop while easy to replicate. Examples of IT systems are: computers, computer networks, telephones, telephone networks, televisions, and cable networks. IT systems are ubiquitous, impacting all businesses (manufacturing, health care, education, etc.) which means increasingly complex digital IT systems are everywhere and need to be tested for a variety of reasons.

The NIST Laboratory Mission is to promote the U.S. economy and public welfare through technical leadership and participation in the development of the nation's measurement and standards infrastructure. From this perspective, the NIST Information Technology Laboratory (ITL) has defined IT as:

> Information technology is the body of methods and tools by which communications and computing technologies are applied to acquire and transform data, and to present and disseminate information to increase the effectiveness of the modern enterprise.

### Metrology

The definition of the term "metrology" in the *International Vocabulary of Basic and General Terms in Metrology* (the VIM)[1] is:

> **metrology**
>
> science of measurement

The VIM further notes that metrology includes all aspects both theoretical and practical with reference to measurements, whatever their uncertainty, and in whatever fields of science or technology they occur.

Metrology for physical and chemical properties has advanced over the last 200 years, keeping pace with technology and industrial advancements. Metrology for IT systems is in its infancy. Measurement of IT system software consists of ascertaining or testing for logical/mathematical states or functionality in an IT system. IT system hardware is relatively easy to measure (except that complexity of VLSI causes its testing to remain incomplete, just like software), because it relies upon mature and sophisticated physical and chemical measurement science.

# Establishing a Conceptual Basis for IT Metrology

## Principles of Physical Metrology

In order to explain IT metrology, it is necessary to examine the logical basis of metrology. Many of the classical concepts of metrology have their roots in physics, but they have been successfully applied to other areas of science and technology. A model of the logical relationship between standards, measurement, and quantities is shown in Figure 1.

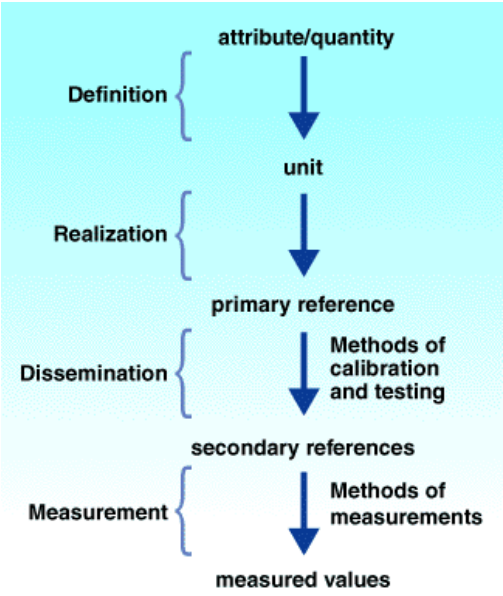


Figure 1 Logical relationship among metrology concepts for use in standardization in measurements.

This figure shows the logical chain between a conceptualized property and the measured value of that property, within a system of standards and traceability. The following examines each of the components of Figure 1.

The term "standard," while perhaps unavoidable, must be used carefully. In English, it has two relevant meanings: as a specification (what is called "norme" in French) and as the reference realization of the unit of a quantity (what is called "étalon" in French). The VIM definition for the latter term is:

**(measurement) standard**

**étalon**

material measure, measuring instrument, reference material or measuring system intended to define, realize, conserve or reproduce a unit or one or more values of a quantity to serve as a reference

The two meanings are very different. For instance, the ASCII code is a standard in the first sense, but not in the second. Unfortunately, there is a tendency to use the term without regard to the sense in which it will be understood.

It is important to understand that Figure 1 is a diagram of logical relationships, not of chronological development. Historically, many (if not most) quantities began as qualitative comparisons (for example, "warmer" and "colder"), followed by the invention of a formally defined quantity (e.g., "temperature"), and finally with the development of units, scales, and a system of standards. IT is much more in the earlier part of this evolutionary process than are more mature fields such as physics or chemistry.

## Quantities

From the top of Figure 1, the VIM definition of the term "quantity" is:

**quantity**

attribute of a phenomenon, body or substance that may be distinguished qualitatively and determined quantitatively

This appears clear. However, it is necessary to examine the operative elements of this definition in order to apply it to IT. The first requirement is that it is necessary to deal with an attribute (of an IT system). In other words, there must be a specific, distinct property to measure.[2] It is critical to understand the impact of this seemingly obvious point. There are examples of "measurements" being done for which no quantity can be clearly identified (e.g., "flavor", "feel", "consumer confidence"). For these, it may be difficult to apply concepts of traceability and standards.

Not all qualitatively distinct attributes are subject to measurement, however. An attribute may be strictly qualitative (for example, whether a computer program is a word processor or a painting is beautiful). To be subject to measurement, it must be possible to determine an attribute quantitatively. A property is a quantity if it allows a linear ordering of systems according to that property.[3] In other words, a property $p$ is a quantity if one can always say of two systems possessing $p$ that the two are equal in $p$ or that one system is less than the other in $p$. Assigning numbers to properties is not enough. The numbers must be meaningful in terms of an ordering relationship among objects possessing that property. This requirement eliminates many taxonomic relationships from the possibility of quantitative treatment.

**Units and Scales**

The existence of a quantity is a necessary, but not a sufficient, requirement for the existence of a measurement. In order to make measurements, it is also necessary to be able to assign numbers to quantities. Ellis proposes the following definitions for a measurement:[4]

1. Measurement is the assignment of numerals to things according to any determinative, non-degenerate rule.
2. We have a scale of measurement if and only if we have such a rule.

This specification is quite open-ended, since the rule of assignment is arbitrary. For the measurement of a specific quantity, however, he adds additional requirements to the effect that the numerals obtained by measurement are consistent with the ordering determined by the quantity. Other authorities are more specific about the requirements of measurement. Their aim is to define measurement in a way that conforms to intuitive notions. To this end, the following requirements are usually put forth:[5]

1. There is a rule for assigning a distinguished value (usually zero) to the quantity;
2. There is a specified, reproducible state of objects for which a second, distinguished value (usually one) of the quantity should be assigned (that is, there should be a *unit*); and
3. There is a scale, of multiples and sub-multiples of the unit, for which there is a rule stating the empirical conditions under which two intervals between measured values are equal. (For example, a centimeter is the same interval of length everywhere along a ruler.)

There is, however, the possibility of another type of measurement.[6] For these measurements, the requirement of ordering can be replaced by a looser requirement of equality. This is supplemented by two additional rules: that of the unit (number 2 above) and a new requirement that quantities be additive. This means that when two objects possessing a quantity are combined (in a well-defined way), the combined object possesses the quantity in a magnitude that is the exact sum of the magnitudes of the quantity in the components. Thus, for instance, a combined object has a mass equal to the sum of the masses of its components. (Not all quantities are additive: when equal amounts of water at a given temperature are combined, the resultant water will not have a temperature that is the sum of the temperatures of the individual amounts.)

The VIM defines a value of a quantity as a "magnitude of a particular quantity generally expressed as a unit of measurement multiplied by a number." However, it allows the possibility that a quantity might not be expressible as a unit of measurement

multiplied by a number. In that event, it may be expressed by reference to a conventional reference scale and/or to a measurement procedure.

The process of defining quantities, units, and scales is one of establishing a consensus. Generally, there is a certain level of arbitrariness in this process, and other systems could have served equally well. This is certainly true of the SI system of units. Having said that, there is also a great deal of empirical truth constraining the development of a system. To be practicable, a system of quantities and units must be both internally consistent and consistent with reality as we experience it. Likewise, the starting point is never the unit; it is always necessary to start with a definition of the quantity to be measured. (Thus, for instance, saying that the "bit" is a unit of measure in IT is not valid without specifying what quantity is being measured. The bit, for instance, can be used to measure optical resolving power,[7,8] probably not what most computer scientists associate with the term.)

**Realization and References**

Definitions of quantity and unit are not enough to provide a means of measurement. Measurement is, in essence, the comparison of an object not to the unit of the quantity being measured, but to a physical realization of the unit. As stated by Ellis:[9]

> "The thing to be measured is matched, in respect to the quantity concerned, by a series of operations with the members of a set of standards, or their equivalents."

The VIM defines a number of types of standards. There is usually one, distinguished standard:

> **primary standard**
>
> standard that is designated or widely acknowledged as having the highest metrological qualities and whose value is accepted without reference to other standards of the same quantity

The realization of a unit usually takes the form first of a *primary standard*. This is a physical object or phenomenon deemed to embody the unit of the quantity in question. In the SI system, only the unit of mass (the kilogram), is defined in terms of an artifact. All other units are defined in terms of scientific principles and the realization of the unit is a technological challenge.

Secondary standards are standards whose values are assigned by comparison with a primary standard of the same quantity. Secondary standards are used when it is impractical for all measurements to be made by direct comparison to the primary standard.

## Measured Values

A measured value is the numerical result obtained from the application of a measurement method to an object, possessing a quantity. One characteristic of a measured value of interest to the task group is traceability. Much of trade requires traceable measurements. The VIM definition is:

> **traceability**
>
> property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties

This definition is intended to be applied within a system of measurements that conforms to Figure 1. A challenge facing NIST is to apply the definition of traceability to assessments of IT product characteristics. It is necessary to either put into place a metrology system that is consistent with the existing structure, or to extend the structure to include IT products.

## Number, Counting, and Probability

It is worth briefly examining the logical status of counting and of probability in the philosophy of metrology. Historically, some questions have been posed about counting and probability which are somewhat ironical since so many physical measurements are based upon these concepts.

The process of counting poses difficulties for philosophers: is counting objects a measurement procedure? In one sense, it seems to be. Certainly, number is a quantity in the sense that it satisfies the previous definitions of a quantity. What seems lacking is the arbitrariness of a scale of measurement; there seems nothing which corresponds to choosing a unit. As Ellis states, “If we must speak of counting as a measuring procedure, it is unique among all measuring procedures.”

Carnap claims that measurement “goes beyond” counting in that it gives values that can be expressed by irrational numbers, hence enabling the application of calculus and other powerful mathematical tools. However, many physical phenomena (such as charge) are in essence discrete. Despite their discrete nature, advanced mathematical tools are used to analyze quantitative relationships among them, measuring them, and treating measured values as having uncertainty. If discrete quantities are essentially different from continuous ones, the logical basis of the distinction has not been clearly put forth.

Probability presents different, but equally serious challenges to philosophers of measurement. Is the assessment of probability a measurement? In the sense of probability as “relative frequency” or as “subjective probability” there seems to be agreement that this is indeed measurement, since the outcome depends on the actual

state of the world. However, probability is understood in another sense: as "degree of confirmation."

Carnap[10] claims that the term probability is ambiguous, involving two distinct kinds (which may be called empirical and logical). More importantly, he claims that assessment of logical probability is not measurement. Ellis, however, argues that the distinction between kinds of probability is based on reasoning that can be applied to every other quantity concept. His conclusion is that, just as the distinction between empirical and logical temperature, length, etc. are unimportant, so is the distinction between empirical and logical probability. All such assessments should be considered measurements.

## Principles of IT Metrology

After reviewing the logical relationships between metrology concepts illustrated in Figure 1, the task group believes that these concepts and the concept of traceability apply to metrology for IT. However, it is important to recognize two aspects which delineate or distinguish IT metrology from physical metrology. First, useful IT quantities are not realizable solely by use of a physical dimensioning system; such as SI. [*]Secondly, existing methods for calculating expressions of uncertainty in physical metrology can not be easily or always applied in IT metrology.

There appears to be no recognized, established dimensioning system or quantities relevant to IT metrology. Of the seven base units in SI, only the "second" for time, appears essential for IT metrology. Possibly, the only other base unit necessary for IT metrology is the "bit" for information. There is no equivalent in IT metrology to the ISO 1000 (and ISO 31) for SI in physical metrology. Possibly developing such an equivalent would be useful, maybe not. One advantage in IT metrology appears to be that, whatever base and derived units are used, the technological challenge posed in realizing SI units does not exist. In other words anyone can define and establish a "bit" of information without use of a measurement device. Possibly all that is needed to define the quantity of information is reference to a classic work, such as *Mathematical Theory of Communication* by Shannon and Weaver.[15] Such work preceded the present, dramatic deployment of digital IT systems but still may sufficiently characterize information as a quantity and bit as a unit of measure.

The VIM definition of traceability requires evaluation of uncertainty. For IT metrology, uncertainty can be difficult to define, much less to quantify. Statistical methods of treating repeatability and accuracy in physical metrology don't clearly apply to the many logical measurements associated with IT. When test results are represented by pass/fail instead of quantitative results or when test results can not exhaustively test to an IT standard (i.e., number of possible tests are too large to economically or quickly

---

[*]SI units of measure are very useful and well established for measuring many physical quantities.[11] However, some physical quantities are more usefully measured in non-SI units, such as a hardness scale,[12] pH,[13] and Richter scale.[14] In fact, the SI specifically states that it does not treat conventional scales, results of conventional tests, currencies, nor information content. Here conventional tests mean such measurements as of pH which are carried out under a convention different from SI.

complete), it appears that methods for establishing a level of confidence are more useful for establishing traceability in IT metrology.

Figure 2 illustrates and compares the concepts of measuring physical quantities and measuring digital information technology systems quantities. Figure 2 includes and expands upon the metrological concepts illustrated by Figure 1. The concept of **definition** from Figure 1 maps into the **specification** row in Figure 2. The concepts of **realization**, **dissemination**, and **measurement** from Figure 1 map into the **methods of testing** row in Figure 2. Figure 2 adds a third row for **commercial products** to illustrate how commercial products depend upon measurements.

Therefore, the three rows in Figure 2 are intended to show how specifications, which may employ physical or digital information systems quantities, are implemented correctly in commercial products by use of appropriate methods of testing. The three columns in Figure 2 (from left to right) are intended to show how specifications, methods of testing, and commercial products can become increasingly complex. The conformance of implementations (commercial products) with respect to the specification may be established through traceability calculations or level of confidence assertions.

**Figure 2**

## Measuring Physical Quantities

[length, mass, time, electric current, thermodynamic temperature, luminous intensity, pH, hardness, Richter Scale, ...]

| | **Units** | **Standards** | **Applied Uses/Practices** |
|---|---|---|---|
| **Definition and Specification** | ISO 1000 [meter, ...] | ISO 261, ISO 262, ISO 724, ISO 965 [metric screw threads], ISO 7, ISO 228 [pipe threads] | NFPA 70 [national electrical code] |
| **Methods of Testing** | primary reference [atomic clock, cesium laser], standard reference material, standard reference data, calibration | primary reference [standard reference thread, scratch standard, gage], calibration, conformance testing | inspection, calibration, reference material, reference data, conformance testing, interoperability testing |
| **Commercial Products** | measurement instrument [laser interferometer, tape measure] | building components [pipe, nut, bolt, screw] | structure [building, bridge] |

**Figure 2 - continued**

# Measuring Digital Information Technology Systems Quantities

[time, information, mathematical operations, ...]

| | Units | Standards | Applied Uses/Practices |
|---|---|---|---|
| **Definition and Specification** | ISO 2382 [bit, byte, word, error, fault, ...], ISO 1000 [second,...] | ISO 646 [ASCII], ISO 2382 [floating point rep], ISO/IEC 9899 [C] | ISO 10303 [STEP], IETF RFC 1610 [TCP/IP], ISO 9945-1 [POSIX] |
| **Methods of Testing** | calibration, conformance testing | conformance testing, interoperability testing, reference data, reference implementation | inspection, conformance testing, interoperability testing, reference data, reference implementation |
| **Commercial Products** | performance analyzer, logic tester | C compiler, printer, monitor, microprocessor | operating system, networking software, router, computer assisted manufacturing device |

In an effort to develop a taxonomy for methods of testing, the following key definitions in Figure 3 were collected. Where definitions could not be found, the task group developed its own definition. From Figure 3, the task group has developed a taxonomy of testing or measuring:

- calibration
  - reference material
- inspection
- reference data
- conformance testing
  - reference implementation
- interoperability testing
  - reference implementation

## Key Definitions

| Term | Definition | Source |
|---|---|---|
| **calibration** | Set of operations that establish, under specified conditions, the relationship between values of quantities indicated by a measuring instrument or measuring system, or values represented by a material measure or a *reference material*, and the corresponding values realized by standards | VIM |
| **conformity** | Fulfilment by a product, process or service of specified requirements. | ISO/IEC - Guide 2 |
| **conformity evaluation** | Systematic examination of the extent to which a product, process or service fulfills specified requirements. | ISO/IEC - Guide 2 |
| **conformity testing** | *Conformity evaluation* by means of testing | ISO/IEC - Guide 2 |
| **inspection** | *Conformity evaluation* by observation and judgement accompanied as appropriate by measurement, *testing* or gauging. | ISO/IEC - Guide 2 |
| **interoperability testing** | The testing of one implementation (product, system) with another to establish that they can work together properly. | Task Group |
| **means of testing** | Hardware and/or software, and the procedures for its use, including the executable test suite itself, used to carry out the testing required. | ISO/IEC 9646-1 |
| **measurement** | Set of operations having the object of determining a value of a quantity. | VIM |
| **reference data** | In physical metrology, reference data is quantitative information, related to a measurable physical or chemical property of a substance or system of substances of | Task Group |

| | | |
|---|---|---|
| | known composition and structure, which is critically evaluated as to its reliability.<br><br>In information technology, reference data is any data used as a standard of evaluation for various attributes of performance. | |
| **reference implementation** | Implementation whose attributes and behavior are sufficiently defined by standard(s), tested by certifiable test method(s), and traceable to standard(s) that the implementation may be used for the assessment of a measurement method or the assignment of test method values. | Task Group |
| **reference material** | Material or substance one or more of whose property values are sufficiently homogeneous and well established to be used for the *calibration* of an apparatus, the assessment of a measurement method, or for assigning values to materials. | VIM |
| **test** | Technical operation that consists of the determination of one or more characteristics of a given product, process or service according to a specified procedure. | ISO/IEC - Guide 2 |
| **testing** | Action of carrying out one or more *tests*. | ISO/IEC - Guide 2 |
| **traceability** | Property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties. | VIM |

**Figure 3**

All of these methods of testing or measuring (calibration, inspection, reference data, conformance testing, interoperability testing) are applicable to either physical or digital IT systems metrology. Many of the terms in Figure 3 are defined in basic metrology or conformity assessment documents (VIM[1], ISO/IEC Guide 2[16]). Somewhat surprisingly, the task group was unable to find suitable existing definitions

for interoperability testing, reference data, and reference implementation. Suitable definitions for these testing methods were developed by the task group in order to allow for a complete discussion about all of the methods of testing presently being used for digital IT systems quantities.

It is interesting to note that the VIM defines **measurement** but not **test** or **testing** and that the ISO/IEC Guide 2 defines **test** and **testing** but not **measurement**. To the task group, **measurement** and **testing** appear to be defined so that these terms are either conceptually equivalent or, at least, very close to equivalent. Therefore "testing and measurement" are often combined in this white paper not to delineate but to emphasize their rough equivalence. The task group also acknowledges that, in some fields, a distinction between these terms is made by considering testing to be a measurement together with a comparison to a specification.

# Methods of Testing for Digital IT Systems Quantities

Of the five methods of testing identified in the previous section -- calibration, conformance testing, interoperability testing, reference data, and inspection -- all but calibration are in widespread use as methods for testing for digital IT systems quantities. Conformance and interoperability testing often make use of the concept of reference implementations.

The following provides a brief review and status on methods of testing for digital IT systems quantities.

**Calibration**

The concept of calibration is well understood in the physical metrology community. Calibration means that the measurement of the value of the properties is related to measurements on primary standards usually provided by the primary national laboratory. The relation is called traceability.

The purpose of calibration and traceability is to ensure that all measurements are made with the same sized units of measurement to the appropriate level of uncertainty so that the results are reliably comparable from time to time and place to place.

The definition of traceability is the ability to relate individual measurement results through an unbroken chain of comparisons leading to one or more of the following sources: national primary standards, intrinsic standards, commercial standards, ratios, and comparison to a widely used standard which is clearly specified and mutually agreeable to all parties concerned.

In the open systems subcommunity of IT, ISO/IEC TR13233[17] states "Since measurement traceability and calibration are not generally directly relevant to software

and protocol testing, the title of clause 9 in this interpretation has been changed to 'Validation and traceability'." This report concludes that validation is to software and protocol test tools as calibration is to measurement equipment.

**Conformance Testing**

The IT method of testing with the greatest amount of experience, widespread use, and development of methodology is conformance testing of digital IT systems. Testing methodologies have been developed for operating system interfaces[18], computer graphics[19], document interchange formats[20], computer networks[21], and programming language processors[22]. Additionally, about fifteen years ago, IT standards developers began to realize that standards for digital IT systems were becoming quite complex and dependent upon both physical metrology and non-physical metrology. Consequently, assessing conformity of hardware/software implementations is now on inherently complex and somewhat ambiguous process. There are only a very few documents which address such conformity issues[23,24].

Most of the testing methodology documents cited above use the same concepts, if not the same nomenclature. IT standards are almost always developed and specified in a natural language, English, which is inherently ambiguous. Sometimes the specifications are originally developed or translated into a more unambiguous language called a formal description technique (FDT). Since the specifications in IT standards are often very complex, as well as ambiguous, most testing methodology documents require the development of a set of test case scenarios (e.g., abstract test suites, test assertions, test cases) which must be tested. The standards developing activity usually develops the standard, the FDT specification, the testing methodology, and the test case scenarios. Executable test code which tests the test case scenarios is developed by one or more organizations which may result in more than one conformance testing product being available. However, if a rigorous testing methodology document has been adhered to, it should be possible to establish whether each conformance testing product is a quality product and an equivalent product. Sometimes an executable test code and the particular hardware/software platform it runs on become accepted as a reference implementation for conformance testing. It should be noted that, on occasion, a widely successful commercial IT product becomes both the defacto standard and the reference implementation against which other commercial products are measured.

In IT, an example of a primary standard might be a reference implementation of a function (assuming that such an implementation is a measurement standard to begin with). It is possible to have multiple primary standards (or, depending on one's viewpoint, no primary standard). For instance, a reference implementation of an algorithm may be running on two (nominally identical) machines. This raises issues because the behavior of the two running systems may differ; mechanisms must be established for intercomparison of primary standards.

### Interoperability Testing

No interoperability testing methodologies have been established comparable to existing conformance testing methodologies. Interoperability testing usually takes one of three approaches to ascertaining the interoperability of implementations (i.e., commercial products). The first is to test all pairs of products. Typically an IT market can be very competitive with many products and it can quickly become too time consuming and expensive to test all of the combinations. This leads to the second approach of testing only part of the combinations and assuming the untested combinations will also interwork. The third approach is to establish a reference implementation and test all products against the reference implementation.

### Reference Data

The use of reference data is very important in both physical and IT metrology. When the task group could not find any existing definition for reference data, the task group turned to NIST experts for suggestions, and as a result, Figure 3 has separate definitions for reference data as applied to physical and IT metrology. For IT, reference data is used to measure various aspects of performance of digital IT systems.

### Inspection

Inspection, as a method of testing, is a concept that applies equally well to either physical or IT metrology. There has been at least one attempt to document an inspection methodology for one area of IT, the evaluation of software products.[25]

Inspection of complex structures, for instance buildings, in physical metrology has a legacy of many decades of experience. While inspection of digital IT systems is a relatively new area compared to building inspections, there is one advantage in IT metrology. In the area of software products, each copy of a product can reasonably be assumed to be identical and inspection of one copy is therefore sufficient to know something about all copies.

The pass/fail decision based on inspection is usually more subjective than objective. This forces two necessary conditions. The first condition is that the inspector (the person performing the inspection) is qualified to make a subjective decision. The second condition is that the surrounding environment be as defined and consistent with similar inspections as possible. For example, to determine that an application produces a correct color for viewing an inspection could be performed. The conditions that would be defined for the inspection could be the room lighting, the hardware/software platform of the application, the monitor type used for the inspection, and the expertise of the inspector.

## Status and Opportunities for IT Metrology

The state of IT metrology is best illustrated by comparing it to the state of physical metrology. Many of the definitions and general terms for metrology[1], standardization[16], and requirements for calibration and testing laboratories (ISO/IEC Guide 25)[26] apply equally well to physical and IT metrology. IT metrology has some concepts and terms for which no well established definitions exist (e.g., reference data, interoperability testing, reference implementation). Also, some IT testers believe that the requirements in ISO/IEC Guide 25 for calibration and testing laboratories require extensive interpretation for IT testing and have spent considerable time and resources in developing such an interpretation.[17] Other IT testers believe that ISO/IEC Guide 25 is sufficient, without extensive interpretation, for IT testing.

For physical metrology there are at least several decades of papers refining metrological concepts such as traceability.[27, 28, 29, 30] There is no comparable literature for determining the level of confidence in IT test results which might serve the same purpose as establishing traceability in physical metrology. NIST staff members have been major participants in the advancement of physical metrology.

The IT equivalent of physical measurement uncertainty may be straightforward or, for more complex software, a genuine frontier for IT metrology. Three examples can illustrate the spectrum of difficulty in dealing with uncertainty in software measurements. In the first case, a software standard may be unambiguous and the combinations/permutations to be tested are finite and possible to exhaustively test (e.g., 128 characters in seven bit ASCII). In the second case, a software standard may be unambiguous (e.g., an encryption algorithm such as DES) and the combinations/permutations to be tested are very large and not feasible/possible to exhaustively test (e.g., DES has more than 10**36 possible tests). In the third case, a software standard may be somewhat ambiguous (e.g., the syntax and semantics for a programming language, such as C) and the combinations/permutations to be tested are very large and not feasible/possible to exhaustively test (e.g., possible C code is infinite). In the above first case, uncertainty is clearly more measurable than the above third case.

Recently, there has been several contribution on computers systems in metrology and the need for an empirical science for the performance of algorithms.[31, 32, 33, 34] Again, NIST staff members have contributed to this literature which is of potential value to advancing both physical and IT metrology.

There is a large amount of literature on IT metrics and measurement. A recent search on a major search engine on the web netted over 150 thousand entries on “software + metric”. Most of this literature discusses applying existing metrics for quality, size, complexity, or performance and refining these measures. There is very little discussion

on fundamental measurement strategies for IT. The task group knows of no journals devoted to IT metrology as there are for physical metrology (e.g., *CAL LAB The International Journal of Metrology*). There are newsletters,[35] journals, and books on software engineering and testing techniques which include discussions of metrics and measurements. At least one standard for software measurement is being developed.[36] There are also conference, symposia,[37] and ongoing research[38] in the area. Most of these publications and activities have occurred in the last thirty years since the IT field is fairly young.

**Opportunities**

From the literature reviewed and discussions held by the task group it is apparent that there are numerous areas with opportunities to advance the state of IT metrology. Some areas are already being worked upon by industry. Other areas have seen relatively little study and development to date. In no particular order, the task group suggests the following are areas with opportunities for advancing IT metrology:

1. Level of confidence in test results - Today, the quality of an information technology product or component is assured without rigorous metrics for the confidence factor. For instance, commercial producers of software may use a combination of the following to decide that a product is "good enough" to release:

   - a sufficient percentage of test cases run successfully
   - executing a test suite while running a code coverage analyzer to gather statistics about what code has been exercised
   - classification of defects into different severity categories, and analysis of numbers and trends within each category
   - beta testing: allowing real users to run a product for a certain period of time and reporting problems; analyzing the severity and trends for reported problems
   - analyzing the number of reported problems in a period of time; when the number stabilizes or is below a certain threshold for a period of time, it is considered "good enough".

   Although code coverage and trend analysis are initial steps towards a more rigorous definition of certainty of a product's quality, there is still much work that is needed in defining the mathematical foundations and methods for assessing the uncertainty in quality determinations.

   IT metrology would profit from the development of an equivalent set of concepts to calibration, traceability, and uncertainty which are so important in physical metrology. Where uncertainty is calculated by statistical methods for physical test results, the level of confidence can be calculated. Being able to analytically derive a level of confidence for IT test results would advance IT metrology.

2. Interoperability testing - If implementation A and implementation B interwork and if implementation B and implementation C interwork, what are the prospects of implementations A and C interworking?

3. Automatic generation of test code - Developing test code for IT conformance testing can be more time consuming and more expensive than developing the standard or a product which implements the standard. There are several efforts in specifying more formally the standard or specification and generating test code from this formalization. One example is the Assertion Definition Language (ADL) effort managed by X/Open, with funding from MITI based on ongoing research at Sun.[39, 40, 41] There is other ongoing research based on modeling, finite state machines, combinatorial logic, and other formal languages such as Z.

4. Need for IT dimensioning or description system(s) - The general concept of fundamental and derived units for IT metrology has been raised in this paper. Is there a need to expand upon this concept?

   A general vocabulary needs to be developed to describe components which comprise information systems. This entails developing a rich, standardized terminology to capture the functionality and capabilities of a software component, in addition to the interface specifications. This could be considered analogous to the situation one sees currently in the microelectronics hardware world, where a circuit designer chooses chips and chip sets for a board design based upon published specifications detailing performance characteristics. This is possible for hardware systems because specifications exist that comprehensively define the performance of hardware components.

   The definition of these formal specifications in a standardized, rigorous way will enable designers and systems integrators to select software components with confidence regarding the component's capabilities and how it will integrate into the system being built. Furthermore, automated composition of systems based on specifications will be possible once these types of definitions exist and are widely deployed in a certifiable way.

5. Software metrics - The need to more rigorously measure and test software as it is developed is being explored by industry. As software products become increasing complex, sound software metrics will be needed.

6. Algorithm testing - As researchers develop new algorithms, some means of measuring the performance of these algorithms for comparison purposes is needed. There exist some measures of performance today, such as Whetstones, Dhrystones, etc. which are benchmarking programs targeted at specific aspects of a computer's capabilities. A more general capability for establishing the performance of

algorithms in a similar fashion should be developed. For example, planning or scheduling algorithms could be run against standard datasets or scenarios (artifacts?). There are several challenges, including: determination of a theoretical foundation for measuring the performance of algorithms, and means of ensuring that implementation-dependent performance results are meaningful.

## Roles for NIST in IT Metrology

The task group developed Figure 2 to illustrate a conceptual basis for physical and IT metrology. Figure 2 also serves as a framework for discussing NIST's roles. As a key national measurement laboratory for U.S. industry, the task group believes NIST already serves in many measurement roles for all three columns in Figure 2 for measuring both physical quantities and digital IT systems quantities.

For the testing of digital IT systems, NIST has been very active in the testing of complex specifications. In this area (i.e., the right side of Figure 2) NIST has a successful history of providing key testing support. For physical metrology, NIST clearly has provided key measurement support for fundamental to complex specifications (i.e., from left to right side in Figure 2). There is also a substantive history of work by NIST in the mathematical, computational, and statistical sciences which support all of the columns in Figure 2. In other words, NIST's roles in metrology (past, present, and future) are, appropriately, the entire matrix of Figure 2.

It should be noted that NIST's IT metrology mandate will always be bounded by available resources. For instance, if the IT industry were to look to NIST for assistance in developing all of its conformance testing needs, the associated development costs could overwhelm the entire NIST measurement budget. NIST will have to continue to prioritize its program of work in IT metrology as part of its overall metrology program in support of U.S. industry.

## Conclusions

IT metrology is a valid branch of metrology. The task group started with this as an assumption and ended with this as a belief. IT metrology differs from physical metrology in several ways including; the SI dimensioning system is not as relevant; less analytical methods exist to quantify uncertainty; and the area is relatively new compared to physical metrology. All of this means that IT metrology has its own unique set of challenges, opportunities, and priorities.

IT and IT metrology will be a key to U.S. competitiveness and international commerce in the twenty-first century. Advancing IT metrology and supporting specific priority IT testing and measurement needs of U.S. industry should be key goals for NIST. This paper has attempted to propose concepts, provide information, and pose questions which might help to establish a frame of reference for NIST staff and management as

they consider how to advance IT metrology and support U.S. industry's IT testing and measurement needs.

# Annex A: References


1. *International Vocabulary of Basic and General Terms in Metrology*. International Organization for Standardization: Geneva. 1993.

2. Stephan Korner, "Classification Theory," *Encyclopedia Britannica: Macropaedia*, 15th ed., 1977. According to Korner, we organize our understanding of the world in three ways: objects and their attributes; objects and their parts; and relationships between distinct classes of objects.

3. Brian Ellis, *Basic Concepts of Measurement*. Cambridge University Press: Cambridge, England. 1966.

4. Ellis, op. cit., p. 41.

5. Rudolf Carnap, *Philosophical Foundations of Physics*. Basic Books: New York. 1996.

6. Karel Berka, *Measurement Its Concepts, Theories and Problems*. D. Reidel Publishing: Dordrecht, Holland. 1983.

7. Donald M. MacKay, *Information, Mechanism, and Meaning*. The M.I.T. Press: Cambridge. 1969.

8. Tom Stonier, *Information and the Internal Structure of the Universe*. Springer-Verlag: New York. 1990.

9. Ellis, op. cit., p. 155.

10. Rudolf L. Carnap, *Logical Foundations of Probability*. University of Chicago Press: Chicago.1950.

11. ISO 1000:1992, SI Units and Recommendations for the Use of Their Multiples and of Certain other Units (ISO 31: 1992, Quantities and Units).

12. ISO 6508: 1986, Metallic materials - Hardness test - Rockwell test (scales A- B- C- D- E- F- G- H- K).

13. OIML Publication R54, pH Scale for Aqueous Solutions, 1981.

14. C.F. Richter, *Elementary Seismology*, W.H. Freeman & Co., San Francisco, 1958.

15. C.E. Shannon and W. Weaver, *Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.

16. ISO/IEC Guide 2:1996, Standardization and related activities - General vocabulary.

17. ISO/IEC TR13233:1995, Information Technology - Interpretation of Accreditation Requirements in ISO/IEC Guide 25 - Accreditation of Information Technology and Telecommunications Testing Laboratories for Software and Protocol Testing Services.

18. ISO/IEC 14515, Information Technology - Programming languages, their environments, and system software interfaces - Portable Operating System Interface (POSIX) - Test methods for measuring compliance to POSIX. (multiple part standard)

19. ISO/IEC 10641:1993, Information Technology - Computer graphics and image processing - Conformance testing of implementations of graphics standards.

20. ISO/IEC TR 10183, Information Technology - Text and Office Systems - Office Document Architecture (ODA) and interchange format - Technical Report on ISO 8613 implementation testing. (multiple part technical report)

21. ISO/IEC 9646, Information Technology - Open Systems Interconnection - Conformance testing methodology and framework. (multiple part standard)

22. ISO TR 9547:1988, Programming Language processors - Test methods - Guidelines for their development and acceptability.

23. ECMA TR/18, The Meaning of Conformance to Standards, June 1983.

24. ISO/IEC TR 10034:1990, Guidelines for the preparation of conformity clauses in programming language standards.

25. ISO/IEC 14598, Information Technology - Evaluation of software product. (multiple part standard)

26. ISO/IEC Guide 25:1990, General requirements for the competence of calibration and testing laboratories.

27. W.A. Wildhack, *Draft Proposal for a Policy on Traceability for IBS, NCSL Workshop on Measurement Agreement*, January 1962.

28. John A. Simpson, *Foundations of Metrology, Journal of Research of NBS*, January 1981.

29. Ernest L. Garner and Stanley D. Raspberry, *What's New in Traceability, Journal of Testing and Evaluation*, November 1993.

30. Charles D. Ehrlich and Stanley D. Raspberry, *Metrological Timeliness in Traceability, Measurement Science Conference*, for presentation at a January 1997 conference.

31. Theodore H. Hopp, *Computational Metrology*, Manufacturing Review, December 1993.

32. Computer Systems in Metrology, Recommended Practice RP-13, National Conference of Standards Laboratories, February 1996.

33. Theodore H. Hopp and Mark S. Levenson, *Performance Measures for Geometric Fitting in the NIST Algorithm Testing and Evaluation Program for Coordinate Measurement Systems*, Journal of Research of the National Institute of Standards and Technology, September-October 1995.

34. J.N. Hooker, *Needed: An Empirical Science of Algorithms*, Operations Research, March-April 1994.

35. Testing Techniques, *A Newsletter Devoted to the Technology of Software Testing*, Software Research Inc.

36. ISO/IEC 14143, Information Technology, Software Measurement. (multiple part standard)

37. Metrics '97, Fourth International Symposium on Software Metrics, (to be held November 1997).

38. Martha M. Gray, *Applicability of Metrics to Large Scale Infrastructure*, (To be published).

39. Sriram Sankar and Roger Hayes, *Specifying and Testing Software Components using ADL*.

40. Shane P. McCarron, *The API Definition Language Project - A Brief Introduction*, X/Open Company Ltd., July 1993.

41. Joseph L. Hungate and Martha M. Gray, *Automated Testing Technologies Workshop*, Section 3 of Conference Report, Journal of Research of the National Institute of Standards and Technology, November-December 1995.

# Annex B: Glossary of Abbreviations

| | |
|---|---|
| ADL: | Assertion Definition Language |
| AP: | Application Protocol |
| ASCII: | American Standard Code for Information Interchange |
| ATEP-CMS: | Algorithm Testing and Evaluation Program - Coordinate Measuring System. |
| ATS: | Abstract Test Suite |
| DES: | Data Encryption Standard |
| DSA: | Digital Signature Algorithm |
| DSS: | Digital Signature Standard |
| DSSVS: | Digital Signature Standard Validation System |
| FDT: | Formal Description Technique |
| IEC: | International Electrotechnical Commission |
| IETF: | Internet Engineering Task Force |
| ISO: | International Organization for Standardization |
| IT: | Information Technology |
| ITI: | Industrial Technology Institute |
| ITL: | Information Technology Laboratory (NIST) |
| MEL: | Manufacturing Engineering Laboratory (NIST) |
| MITI: | Ministry of International Trade and Industry |
| NFPA: | National Fire Protection Association |
| NIST: | National Institute of Standards and Technology |
| pH: | The negative logarithm of the hydrogen ion concentration in solution. |
| POSIX: | Portable Operating System Interface |
| RFC: | Request For Comments |

| | |
|---|---|
| SHS: | Secure Hash Standard |
| SI: | International System of Units (the modern metric system) |
| STEP: | Standard for the Exchange of Product Model Data |
| TCP/IP: | Transmission Control Protocol/Internet Protocol |
| TS: | Technology Services (NIST) |
| VIM: | International Vocabulary of Basic and General Terms in Metrology |
| VLSI: | Very Large Scale Integration |

# Annex C: Examples of Present IT Metrology at NIST

The following examples helped the task group to sort through and understand the basic testing concepts behind the ongoing IT testing activities at NIST. Therefore, they are listed here as illustrative examples and not as a representative sampling or as a complete summary of present IT testing activities at NIST.

**Case 1: Testing DES, DSS, SHA Implementations**

NIST has developed conformance tests for FIPS 186, Digital Signature Standard and FIPS 180-1, Secure Hash Standard. The tests, called the DSS Validation System (DSSVS) are described in DRAFT Digital Signature Standard (DSS) and Secure Hash Standard (SHS): Requirements and Procedures.

The SHS is used for calculating a message digest that can be used with the DSS. The calculation transforms any message of length 264 bits to a 160-bit output. Since the outputs of each SHA transformation becomes the inputs of the next SHA transformation, the final message digest is a function of each bit of the message. Any change to a message in transit will, with a very high probability, results in a different message digest. Using black box test methods the DSSVS tests for conformance to the SHS using three tests: messages of varying length, selected long messages, and pseudo randomly generated messages.

FIPS 186 specifies a DSA for generating and verifying digital signatures on data that has been condensed into a message digest using the SHA. The digital signature itself is a pair of large numbers that are computed on data using the DSA and a set of parameters such that it can be used to verify the identity of message's claimed sender and the integrity of the message itself. Signature generation makes use of the private key, which is a large number, to generate the digital signature. Signature verification makes use of a public key that is related to the private key used to generate the signature. The DSSVS uses black box test methods for conformance to the DSS in three areas: prime number generation, generation of public/private key pair, and signature generation/verification.

**Case 2: Algorithm Testing and Evaluation Program for Coordinate Measuring Systems (ATEP-CMS)**

NIST is now offering a new Special Test Service, the Algorithm Testing and Evaluation Program for Coordinate Measuring Systems (ATEP-CMS). This new Special Test Service is offered under the Office of Measurement Services Calibration Program.

ATEP-CMS evaluates the performance of data analysis software used in coordinate measuring systems (CMSs). Tested software is treated as a filter that transforms point coordinate data into feature parameters according to a defined transfer function. NIST evaluates the accuracy of the filter under conditions typical of those found in industrial practice. NIST independently compares the output of the software under test to predetermined corresponding reference values. NIST uses orthogonal-distance least squares algorithms and supports the following geometry types: circle, line, plane, sphere, cylinder, cone, and torus.

In the Special Tests, the reported measurement uncertainty is determined by the effects of computational roundoff and convergence settings used to generate the reference fits, the propagation of these effects through the comparison algorithms, and sampling uncertainty due to the number of data sets used to perform the test.

**Case 3: STEP Conformance Testing**

STEP is an international standard (ISO 10303) designed to let companies effectively exchange engineering information both internally and with their customers and suppliers. Experience with complex standards has shown that vendor claims of compliance with a standard are not reliable. For this reason, the STEP standard provides testing methods and tools support the objective measurement of software implementations that will ultimately aid in achieving conferment and interoperable systems.

STEP is implemented through a series of standard specifications called Application Protocols (APs). For each AP, an Abstract Test Suite (ATS) is developed that contains test purposes generated from the AP, verdict criteria and input specifications. The ATS is realized into an executable test case by testing labs that will be used to quantify the conformance of an implementation under test.

NIST has teamed with Industrial Technology Institute (ITI) to provide a means by which STEP products can be objectively measured against the standard. This is being done by developing a set of value-added software tools for use by vendors during product development. These tools must be extensible to accommodate the expanding series of STEP Application Protocols. This is being accomplished by a modular system with two elements: a test system which integrates various testing tools and administers the actual tests, and a set of tools for generating a test suites for each AP which are used in the testing process. This unique approach offers many advantages over traditional conformance testing. Conformance testing is generally challenged by U.S. vendors as not being cost effective. Under this approach, vendors can gain confidence that their product can successfully pass testing, they have access to the tools to improve the quality of their products, and they gain from the expanded market that user confidence in a tested product brings. The same tools can also be employed by

end-users to assess the ability of these products to interoperate in an industrial context, further expanding the market for standards-based products.

These tools are being used in the development of early pilot implementations of the standard.