D-SI in Short

Digital brochure on establishing the use of units in digitised communication

This brochure specifies principles for establishing the use of units in a digitalised communication of machine-readable data. These principles are the basis for all applications that transfer or require measurement data according to the specifications of the **Système International d'Unités (SI)** [1].

For the digital exchange of metrological data, it is essential to associate at least each numerical value to a corresponding unit. These two pieces of information enable a statement to be made about the value of a quantity that can be interpreted according to the SI unit system. Because of its indivisibility and fundamental importance, this form of representation is called **atomic representation**. One example is:

1 kg

Here, "1" corresponds to the numerical value and "kg" to the assigned SI unit of kilogram. Together, both pieces of information characterise a mass quantity.

In a digital network in which existing and new applications communicate with each other, even greater importance than before must be assigned to the SI. The ability of the SI to describe all physical processes using only seven base units leads to unprecedented clarity. This clarity is fundamental for the secure, harmonised and economical exchange of data in the emerging digitalisation of metrology.

It is important to distinguish between **human-to-human** and **machine-to-machine** interfaces. The specifications presented here primarily relate to an automated communication and are essential for communication between machines and algorithms operating in an innovative digitised value chain.

Against this background, a metadata model for the secure, unambiguous and unified exchange of metrological data was developed in the European EMPIR project 17IND02 [8] (Communication and validation of smart data in IoT-networks - SmartCom) involving partners all over the globe. This metadata model is referred to as the Digital-SI (D-SI) and is outlined conceptually in Figure 1.

The representation of measurement data using only the seven SI base units is fundamental to the development of the D-SI.

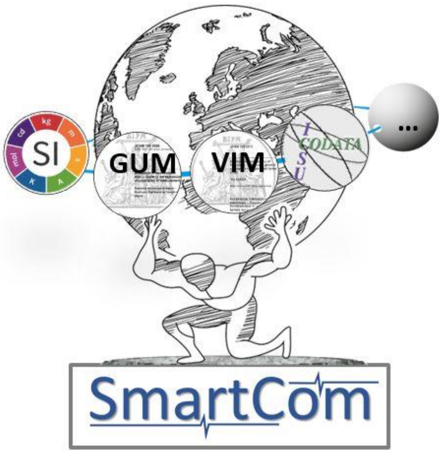

Figure 1 *SmartCom paradigm of D-SI metadata model based on fundamental metrological guides that carries a secure, unambiguous and unified exchange of metrological in a digitalised world.*

The metadata model presented here also allows the use of familiar units, for example, SI-derived units such as Newton and Pascal, and non-SI units like electronvolt.

As a limitation, however, non-SI units can only be accepted if the measured quantities are specified in parallel using SI base units or if the non-SI units are accepted for use with the SI by the BIPM [1]. It is expected that non-SI units will become less important in future machine communication of metrological data, as each of these units can easily be expressed by a combination of SI base units. There will be a **transition** as units that are familiarly used by humans are replaced by their equivalents expressed in terms of the seven SI base units. This transition will only affect the machine representation of units. It is not intended to change the well-established and internationally harmonised human-readable representation of units.

The following general conditions apply to the metadata model presented in this document:

- **Atom**: Atomic data are understood to contain at least one **number** (**numerical value)** and a corresponding **unit**. Optionally, **(measurement) uncertainty information**, a **label** and a **time stamp** can also be specified.
- **Uniqueness**: Units are expressed in **SI base units**. This representation enables a clear, error-

DOI: 10.5281/zenodo.3522074

resistant and economical exchange of digital metrological data.

- **Numbers**: Numbers (numerical values) are displayed in the **decimal system**. Exponential representation is permitted. The separator for decimal places shall exclusively be the dot. Data such as NaN and INF are not permitted as they have no metrological significance.
- **Character representation:** Characters are represented in **UTF-8 format** [5], allowing all official languages to be mapped.
- **Accuracy**: When exchanging data, the accuracy of numbers is of paramount importance. Hence, some non-SI units for angle and time are permitted. Using angle as an example, circular closure can be expressed as both 360 degrees (°) and 2 pi radians. Radian is the appropriate form from both the metrological and the mathematical point of view. However, since pi is an irrational number with its representation limited to finite precision, greater accuracy can be obtained using degrees to represent circular closure.
- **Representation:** The representation of metrological data is based on existing internationally recognised documents. The most important of these are the **BIPM SI brochure BIPM** [1], the **International Vocabulary of Metrology (VIM)** [2], the **Guide to the Expression of Uncertainty in Measurement (GUM)** [3] and the list of **fundamental physical constants (CODATA)** [4].
- **Deviating units**: SI-derived units or units not listed in the SI brochure [1] can be specified in parallel to the SI units. In the event of contradictions, information provided using SI units shall take precedence.
- **Responsibility:** The responsibility, e.g. for the metrological content and its application, lies with the user of the metadata model.
- **Metadata model:** This brochure provides a **metadata model** for the exchange of measurement data which helps to establish the use of units in digitalised communication. The model comprises **identifiers** and **simple data types** for mandatory and optional components designed for a reliable exchange of metrological data.

This brochure describes the fundamental atomic type for real quantities and physical constants.

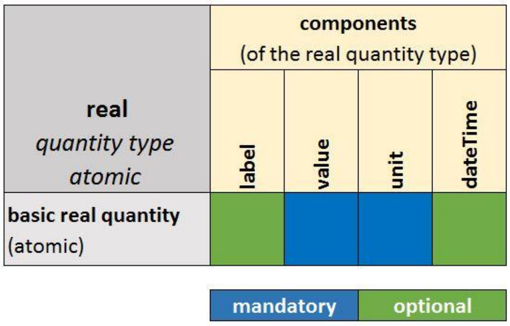

Figure 2 *Metadata model for real quantity values*

Data model for real measurement quantities

The real quantity value must consist at least of a numerical value and an assigned unit. The following information is permitted:

- **Component "value":** The numerical value (number) of the quantity must be provided. It must be written as a decimal floating-point number, which consists of
  - an optional leading sign "+" or "-",
  - followed by a mantissa with the allowed digits "0" to "9" and the dot (".") as the only permitted decimal separator,
  - followed by an optional integer exponent starting with separator "e" or "E" and optional signs "+" or "-".

  Examples: 1.0, 1.3e-2, 0, -0, 0e0, 2E-1

- **Component "unit":** The unit must be a UTF-8 compatible string. Reliable machine readability is only guaranteed if the SI units from the BIPM SI brochure [1] are used. A machine-readable syntax for writing SI units is used for this purpose. Units are written to string according to the following principles:
  - The full unit names and SI prefix names from the BIPM SI brochure are used to define unique identifiers of components that comprise a unit.
  - A basic unit term consists of a unit identifier with an optional prefix identifier and an optional exponent for powers of the unit (operator "tothe{<integer>}").
  - The basic unit term starts with a backslash "\" and all of its components for prefix, unit identifier and exponent are separated by backslashes.
  - A combined unit that is represented by the multiplication of several units is indicated by the concatenation of basic unit terms.

Examples:

- $\frac{\mathrm{m}}{\mathrm{s}}$ → \metre\second\tothe{-1}
- kg → \kilogram
- mg → \milli\gram

- **Component “label”:** A label can be assigned to the metrological data of the atomic quantity by that element. As with the unit, it must be a UTF-8 compatible string. The label can provide information on the kind of measurement and/or the name of the quantity.

- **Component “dateTime”:** This element is used to assign a time stamp for the measurement to the quantity value. All time stamps must be recorded in legal local time with a difference (offset) to the Universal Coordinated World Time (UTC). The extended UTC format from ISO 8601 [7] is used for implementation. As shown in the example below, it consists of a date followed by a time followed by an offset of the time-zone to UTC time.

  Example: 2018-09-05T16:15:03.09-00:01

An XML [6] implementation of the D-SI data model for real quantities is available online. The latest version is provided by PTB [9] and describes further features of the D-SI metadata model, for example, extended real quantities with expanded measurement uncertainty and coverage intervals, and multivariate quantities (including complex quantities).

Metadata model for fundamental constants

Fundamental physical constants from the CODATA list [4] and mathematical constants such as pi have always played a significant role for many measurements. The use of constants became even more substantial with the redefinition of the SI system of units on the pure basis of natural constants that came into force on 20 May 2019 [1]. Therefore, the D-SI metadata model allows for the exchange of quantities representing constants. The corresponding data element is shown in Figure 3.

The **“constant” quantity** data model has the same basic structure as the atomic **“real”** quantity. This comprises the mandatory statement of a **“value”** component together with the associated **“unit”** component.

The **“label”** component is used to identify the constant. The **“dataTime”** component is of great importance for the definition of fundamental physical constants.

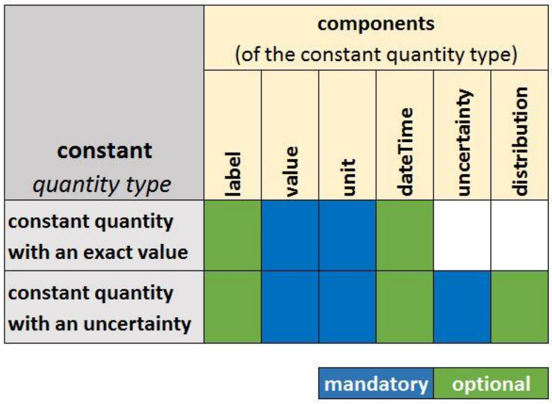

Figure 3 *Metadata model for quantities representing fundamental physical and mathematical constants.*

The current values of these fundamental constants are regularly published by CODATA. The **“dateTime”** element allows identification of the release of CODATA to which the constant is associated.

The constants are divided into those that have an exact numerical value and those that have an uncertainty attributed to the numerical value. The uncertainty value is implemented as the component **“uncertainty”** in the constant data model. Its value refers to the unit element that is also valid for the value element.

The value of the **“uncertainty”** component shall provide the **standard deviation for values from CODATA** [4]. In addition, the distribution of the uncertainty, if known, can be provided in the component **“distribution”**

Mathematical constants play a special role. Numbers like pi or the Euler number are irrational numbers. They require rounding to a finite number of decimal digits for machine representation. Rounding leads to a rounding error that shall be expressed by a rectangular probability distribution that contains the true value of the constant with 100% probability. The rectangular distribution shall be centred at the numerical value that is provided by the element value. The component **“uncertainty”** shall give the **standard uncertainty value for the rectangular probability distribution**.

Summary

A central challenge in the field of digitalisation is the reliability, uniqueness, security and objective evaluability of transmitted metrological information. The foundations for addressing this challenge are provided in the approach outlined in this brochure.

The presented metadata model will be the starting point for reliable and trustworthy new technologies and services for the largely digitalised world of tomorrow, in which metrological information will be available to all at any place and at any time.

Basically, there are two approaches to use the presented metadata format. It can either be transferred into existing data exchange formats or form the basis for establishing the use of units in new exchange formats.

In the future, networked sensors will record all aspects of the production process and make them available to a comprehensive quality management system. With these complete data sets, the performance of systems and processes can then be captured effectively and efficiently, allowing data analytic methods to provide information on optimised system performance. This activity leads to reduced downtime, less waste, significant improvement in quality, and ultimately greater economic success.

E-government concepts, which will be established in many industrialised countries in the near future, will benefit from the proposed uniform presentation of results, as will universities, research institutes and industry. Smart Products, Smart Logistic, Smart Grids, Smart Mobility and Smart Health will find a common, globally coordinated communication standard and so will be able to be reliably integrated into everyday life.

The electronic transmission of measurement results for quantities that are not specified by a combination of a number and a unit of measurement, but are based, for example, on a reference material or a measurement method, is also part of further development.

References

[1] BIPM Brochure, *The international System of Units (SI) 9th edition 2019*, https://www.bipm.org/en/publications/si-brochure/ (accessed August 2019)

[2] JCGM 200:2012 *International Vocabulary of Metrology – Basic and general concepts and associated terms (VIM)*, 3rd edition, 2008 version with minor corrections, 2012

[3] JCGM 100:2008 *Evaluation of Measurement Data – Guide to the Expression of uncertainty in Measurement*, GUM 1995 with minor corrections, first edition, September 2008

[4] CODATA, *Internationally recommended 2018 values of the Fundamental Physical Constants,* https://physics.nist.gov/cuu/Constants/index.html (accessed October 2019)

[5] RFC 3629 *UTF-8, A transformation format of ISO 10464*, https://tools.ietf.org/html/rfc3629 (accessed October 2019)

[6] World Wide Web Consortium *Extensible Markup Language (XML)*; version 1.0; fifth edition https://www.w3.org/TR/xml/ (accessed October 2019)

[7] ISO 8601:2004 Data elements and interchange formats – Information interchange – Representation of dates and times, 3rd edition, December 2004

[8] EMPIR 17IND02 SmartCom, *Webpages*, https://www.ptb.de/empir2018/smartcom/home/ (accessed October 2019)

[9] EMPIR 17IND02 SmartCom, *D-SI XML reference implementation v1.3.0*, https://zenodo.org/record/3366902#.XabORWbgpaQ (accessed October 2019)

Acknowledgement

We thank all partners and stakeholders from the SmartCom project consortium for their valuable contribution to the development of the D-SI data model for establishing the use of units and the exchange of metrological data.

The EMPIR initiative is co-funded by the European Union's Horizon 2020 research and innovation programme and the EMPIR Participating States

Authors

Physikalisch-Technische Bundesanstalt (DE):
Daniel Hutzschenreuter
Frank Härtig
Thomas Wiedenhöfer

National Physical Laboratory (UK):
Ian Smith
Clifford Brown

Contact

smartcom@ptb.de