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

INPUT MANAGEMENT FOR THE DCC

Benjamin Gloger *[a], *Lutz Doering* [a], *Siegfried Hackel* [a], *Justin Jagieniak* [a], *Gamze Söylev Öktem* [a]

[a] Physikalisch-Technische Bundesanstalt (PTB), Bundesallee 100, 38116 Braunschweig, Germany
Email address: benjamin.gloger@ptb.de, lutz.doering@ptb.de, siegfried.hackel@ptb.de, justin.jagieniak@ptb.de, gamze.soeylev-oektem@ptb.de
*Corresponding author: benjamin.gloger@ptb.de

Abstract – The aim of this publication is to break down what information is necessary to generate a DCC. It distinguishes between general and specific specifications and shows which information in addition to the DCC schema is necessary to generate a machine interpretable DCC.

***Keywords*: DCC, D-SI**

1. THE COMPONENTS OF THE ANALOGUE CALIBRATION CERTIFICATE

1.1 What is needed to create an analogue calibration certificate?

To create an analogue calibration certificate, two things are needed:

1. A set of rules is required according to which the calibration certificate is to be drawn up.
2. The content (e.g. administrative data, measurement results, ambient conditions) is needed to be documented in this calibration certificate.

The regulations are described in various norms and standards. The most important standard in this context is probably ISO/IEC 17025 [1].

The basic structure of the DCC is described in [2]. There are five essential regulations to be considered (Fig. 1):

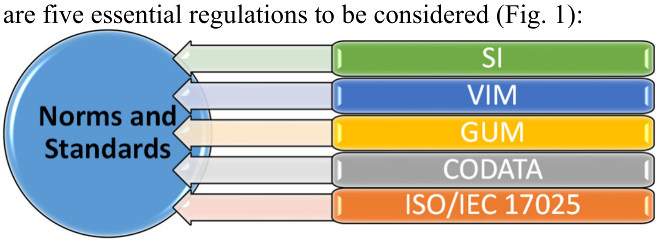

Fig. 1:Norms and Standards used in DCC

These are the D-SI-format [3] (Digital-SI; Digital-SI [4] https://gitlab1.ptb.de/d-ptb/d-si), the International Metrology Dictionary VIM [5] which deals with the vocabulary of metrology, the Guide for Uncertainty in Measurement GUM [6] which describes the declaration of measurement uncertainties, the Codata entries [7] and the standard ISO/IEC 17025 [1].

If you only try to create a template for calibration certificates with these general rules, you will find that you can describe the administrative data very well. For the measurement results, only a very rough framework can be offered. Without knowing the calibration object, it is impossible to include the specific set of rules. Next come the standards and guidelines that relate to the specific calibration object. For instance, a temperature sensor cannot be calibrated according to the specifications of a gauge block and vice versa. Therefore, these different types of calibrations need different templates, or a template that is very flexible, for their specific calibration data. In order to fill this flexible framework, all related standards must be known because they influence the structure of the calibration certificates.

Each calibration object represents a unique specimen. This object is sent by the client to the calibration laboratory of his choice. At the same time, the calibration laboratory receives a special document from the client: the DCR (Digital Calibration Request). The DCR contains the services requested by the client in connection with the calibration order. The results report, which is prepared by the calibration service provider, should then contain the services ordered.

Different analogue calibration certificates of the same calibration material that are prepared by different calibration laboratories can differ considerably from each other. This is the case although they are similar in content because they comply with the specifications. Somewhere in the different calibration certificates there is a table in which the measured values and the determined uncertainties are listed. However, it has not been necessary so far for the analogue certificates to be completely alike.

1.2 What is needed to use a calibration certificate?

The concretisation of this question is how the issued calibration certificate is to be used. In the case of an analogue calibration certificate, a person takes the calibration certificate in his hand, looks for the table with the values he is interested in and interprets them.

However, the situation is completely different when a machine is to be tasked with interpreting a calibration certificate. The machine currently has no usable intelligence whatsoever to recognise contextual content! As early as 1950 the British logician, mathematician, cryptanalyst and computer scientist Alan Turing [8] formulated a test with which the "intelligence" of a machine can be determined, later named the Turing Test after him [9]. A form of the Turing Test is encountered more frequently by internet users when a picture with distorted numbers and letters appears on an internet page (Captcha [10]), the content of which must be entered into an extra input field. There are many publications on the subject of "artificial intelligence" [9], [11] and there is a "Strategy on Artificial Intelligence of the Federal Government of Germany [12]". All of them pursue the goal that a machine can grasp context-sensitive content. This "intelligence" is exclusively mainframe-based. Current, decentralized tool systems do not have the required capacities of a mainframe computer system. However, this is also not necessary if you use a vocabulary (wording) of subject-related definitions [13].

In order to use further advantages of the DCC, another aspect is crucial: the interpretation of the data. The goal is to be able to transfer this interpretation to a machine. To do this, it is necessary to keep the scope for interpretation as small as possible through the schema, the set of rules and the wording, without limiting the flexibility of the individual communities.

2 THE FOUR COMPONENTS OF A DIGITAL CALIBRATION CERTIFICATE (DCC)

Fig. 2 shows the four components of a DCC. The individual components are discussed in this section.

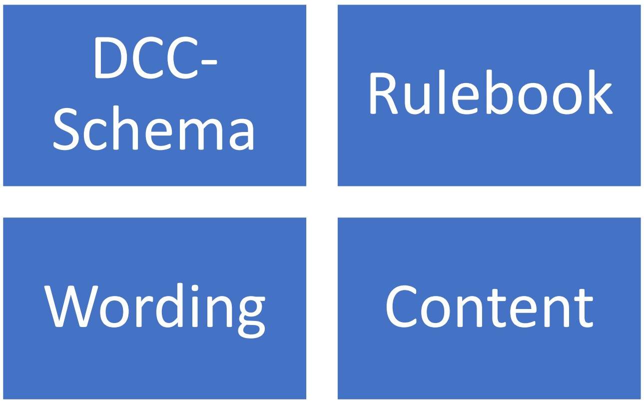

Fig. 2: The four components of a DCC

2.1 DCC Schema: the base of the xml featured DCC

In the analogue world, you could use a calibration certificate without using a machine with an integrated computer. The evaluation is done by humans. In the age of the 4th industrial revolution [14] communication between machines is becoming more and more important. In the process, the evaluation and interpretation are increasingly taken over by the machine. Therefore, there is a demand to digitise the information that was previously available in purely analogue form in calibration certificates in such a way that it can be interpreted by machines.

The most important component is the DCC schema. Where does this schema come from? We have already established that the ISO/IEC 17025 [1] standard plays a special role. It forms the basis for the accreditation of calibration laboratories and defines basic procedures in laboratory management, in particular content requirements. Within this standard, there are various specifications that relate to testing -and calibration certificates, particularly in section "7.8 Reporting of results". These specifications have been applied to the creation of an XML structure. With the help of such a schema, the preparation of a digital calibration certificate is made much easier.

The DCC-schema is described as an XML schema definition (ending ".xsd"). This schema definition describes the structure of a DCC-XML file without knowing its content. It is possible to specify different data types. It is established where a string or a numerical value (integer or floating-point value with a point as decimal separator) or an image can be inserted into the structure. But it is not determined which concrete number or image it actually is. You can think of it as a template that is placed over a sheet of paper. You can write something in every free space. But the person who created the template has no influence on what the user enters in the free fields, except the data type! The responsibility for this lies with the person who fills in the calibration certificate.

More information about the DCC-Schema can be found under https://www.ptb.de/dcc .

The names of the elements, which are already given in the schema, are chosen in such a way that they make sense in the context of calibration certificates in general, but not to cater to the specific wording of one actual calibration object. Considering that there are thousands of different calibration objects, this is not possible, or even desired or necessary. Only with this approach, it was possible to develop a structure that can include the majority of common calibration certificates.

It follows that if three different people try to create a DCC based only on the DCC schema (dcc.xsd), three different DCCs will inevitably emerge. These DCCs then already fulfil a large part of the requirements that the schema specifies. However, they are neither uniform nor interpretable by machines, but most likely interpretable by humans.

2.2 Rulebooks

The sets of rules can be independent from and dependent on the calibration object.

2.2.1 Specifications independent of the calibration object

D-SI

The D-SI Format is used with every machine-interpretable numerical value in the DCC. The D-SI schema is available as a separate schema version that is integrated into the DCC schema. In the DCC (*.xml), there are separate elements for the D-SI, which cover the requirements of the SI when used.

VIM, GUM and CODATA

The vocabulary from the VIM, GUM and CODATA can be found in the DCC schema (dcc.xsd) as elements and in the DCC (*.xml) as element names. Furthermore, it should be used for the attributes in the DCC (*.xml) as shown in the current Good Practice (GP) examples [15]. Complete coverage of the VIM, GUM and CODATA vocabulary is not possible when using only the DCC schema. (See also section "2.3 Wording").

ISO/IEC 17025

The DCC schema (dcc.xsd) makes it possible to fulfil all the requirements of the standard. The aim in converting the standard into the schema is not to map the standard completely, but to offer a guided structure with which the content specifications of the standard can be implemented completely. This means that it is possible to create an xml file that is valid against the DCC schema (dcc.xsd), but does not comply with ISO/IEC 17025 [1].

2.2.2 Specifications dependent on the calibration object

Each calibration object has its own specifications. These are usually summarised in their own norms and standards. Thus, different standards sometimes result in different

specifications for the respective DCC (*. xml), which are not compatible among the calibration objects or sizes. For this reason, the information from specific norms and standards is not included in the DCC schema (dcc.xsd) in a restrictive manner. Therefore, the schema was designed openly in order to be able to create different DCCs (*. xml) complying to different standards.
When a new extension for the DCC schema (dcc.xsd) based on a requirement of a standard is proposed, it should be considered very carefully. Some requirements may be unnecessary from an information technology point of view, as they can lead to redundancies. [16]. For example, the specification of a formula and a graph is not always necessary because the graph is often only the representation of the information already present in a formula. Technically, it would be possible without any problems to leave the representation of the formula to the software, which is to represent the DCC in a human readable way.

Understandably, the current digitalisation drive was not yet foreseeable when many standards were drawn up. That is why these standards require the representation of the same information in different formats to ensure better usability of the calibration information on one sheet of paper.

2.3 Wording

The wording is divided into two categories:

1. Most of the general wording is in the schema file (dcc.xsd) in the elements. This wording has no connection to the content for which the DCC is to be created. It is deliberately kept neutral so that every common calibration certificate can be mapped with it.
2. The contextual wording is related to the calibration object. They are mostly terms that are used in the specific standards. These terms can be found in the DCC as element content or attributes. In order to achieve machine interpretability, this wording must be coordinated. The good practice examples are a first step in this direction.

Since the GP DCC examples claim to follow a cross-community structure, a general wording is to be created here. This wording should refer to the attributes "*Id*","*refId*" and the "*refType"*. In order to be able to assign the terms to specific dictionaries, the attributes are preceded by an abbreviation that indicates the namespace used. The basic namespace is named "*basic*" and serves to define a basic vocabulary across communities. The vocabulary found so far resulted from the transformation of analogue calibration certificates and can be found in the GP-Examples. Furthermore, each community can develop its own dictionary and use it under a namespace in order to make use of a community's own vocabulary in the DCC. This also allows to clearly identify terms from specific standards via a namespace, even if the term used here is described or defined differently in another standard. At present, basic terms are referred to as "*basic*" and community-dependent terms are referred to as "*gp*" for Good-Practice.

For all contents (elements, attributes and specifications in *refType*), the lower camel case notation is used in the DCC. This means that a) words are grouped together, b) the initial letter is written in lower case and c) new word beginnings within the composition are capitalised. Examples of this are "*coreData*", "*respAuthority*" or "*nominalValue*".

```
<dcc:quantity refType="basic_referenceValue">
  <dcc:name>
    <dcc:content lang="de">Bezugswert</dcc:content>
    <dcc:content lang="en">Reference value</dcc:content>
  </dcc:name>
  <si:hybrid>
    <si:realListXMLList>
      <si:valueXMLList>306.248 373.121 448.253 523.319 593.154</si:valueXMLList>
      <si:unitXMLList>\kelvin</si:unitXMLList>
    </si:realListXMLList>
    <si:realListXMLList>
      <si:valueXMLList>33.098 99.971 175.103 250.169 320.004</si:valueXMLList>
      <si:unitXMLList>\degreecelsius</si:unitXMLList>
    </si:realListXMLList>
  </si:hybrid>
</dcc:quantity>
```

Fig. 3: Sampling extracted from Temperature Good Practice simplified [17] (dcc:quantity for the reference values)

Fig. 3 shows a part of the good practice temperature. The attribute "basic_referenceValue" was attached to the element dcc:quantity. The values "basic_referenceValue", "basic_measuredValue" and "basic_measurementError" were identified as cross-community terms for the attribute "*refType"* and included in the "basic_" vocabulary. These refTypes can be used as entry points into the DCC when interpreting it.

XPath (XML Path Language) [18] is a language that is used for addressing parts of an XML document. XPath expressions can be used to search for a specific element in the DCC. With contextual wording of the attributes, one can address different elements. Fig. 4 shows an XPath example. With the following XPath expression, one can select all "*dcc:quantity"* elements that have a *"refType"* attribute whose content is "basic_referenceValue".

```
//dcc:quantity[@refType = 'basic_referenceValue']
```

Fig. 4: XPath Example

A large part of the general wording can be addressed directly without any problems. As an example, we see in Fig. 55 *"dcc:uniqueIdentifier"* element which contains a unique identifier for the DCC in question. This element can be reached in every DCC with the following XPath expression "//dcc:uniqueIdentifier".

```
<dcc:administrativeData>
  …
  <dcc:coreData>
    …
    <dcc:uniqueIdentifier>GP_DCC_temperature_minimal_1.2</dcc:uniqueIdentifier>
    …
  </dcc:coreData>
  …
</dcc:administrativeData>
```

Fig. 5: Sampling extracted from Temperature Good Practice simplified (dcc:uniqueIdentifier)

This is because the element name dcc:uniqueIdentifier is already integrated in the schema and may only occur once in each DCC. It applies for every element that comes forth only once in a DCC. If this is not possible due to multiple elements with the same name, as for example with the measurement results, they receive a refType in the namespace "basic_". If the information is dependent on the calibration object, it is either given a refType with its own namespace "xxx_" or described in an element, sometimes both.

Each piece of information that is recorded as a measured value or that results from the evaluation is to be entered in an XML element. The evaluation does not lead to a change of attributes.

3. Opportunities for community participation

With version 3.1, the DCC schema has reached a stage where it offers a possibility to create Digital Calibration Certificates with a high degree of self-similarity. As described in the text above, the schema offers at the same time enough degrees of freedom to map any calibration certificates. It is important to describe these degrees of freedom so that the resulting XML calibration certificates are not only machine-readable but also machine interpretable. The way to achieve this is via the GP-examples. It is planned to generate GP-DCCs in different communities. This will determine which terms and structures can later be used across communities. At the same time, terms and structures are found that are only needed in the corresponding community, whereby the GP can be seen as a template for further DCCs.

ACKNOWLEDGEMENTS

The Project GEMIMEG-II is carried out with the Grant reference of GEMIMEG 01 MT20001E from BMWK (Federal Ministry for Economic Affairs and Climate Action).

REFERENCES

[1] „DIN EN ISO/IEC 17025:2018-03, Allgemeine Anforderungen an die Kompetenz von Prüf- und Kalibrierlaboratorien (ISO/IEC_17025:2017); Deutsche und Englische Fassung EN_ISO/IEC_17025:2017“, Beuth Verlag GmbH. Zugegriffen: 24. September 2021. [Online]. Verfügbar unter: https://www.beuth.de/de/-/-/278030106

[2] „Digital Calibration Certificate - DCC“, 14. April 2021. https://www.ptb.de/cms/en/research-development/into-the-future-with-metrology/the-challenges-of-digital-transformation/kernziel1einheitlichkeitim/digital-calibration-certificate-dcc.html (zugegriffen 23. April 2021).

[3] D. Hutzschenreuter *u. a.*, „SmartCom Digital System of Units (D-SI) Guide for the use of the metadata-format used in metrology for the easy-to-use, safe, harmonised and unambiguous digital transfer of metrological data“, Nov. 2019, doi: 10.5281/zenodo.3522631.

[4] D. Hutzschenreuter *u. a.*, „SmartCom Digital-SI (D-SI) XML exchange format for metrological data version 2.0.0“, Juli 2021, doi: 10.5281/zenodo.4709001.

[5] „VIM definitions with informative annotations“. https://jcgm.bipm.org/vim/en/index.html (zugegriffen 7. Februar 2021).

[6] BIPM, „Evaluation of measurement data — Guide to the expression of uncertainty in measurement“. 2008. Zugegriffen: 9. März 2022. [Online]. Verfügbar unter: https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6?version=1.9&t=1641292658931&download=true

[7] „Fundamental Physical Constants“, *CODATA, The Committee on Data for Science and Technology*. https://codata.org/initiatives/data-science-and-stewardship/fundamental-physical-constants/ (zugegriffen 15. April 2022).

[8] „Alan Turing“, *Wikipedia*. 9. Januar 2022. Zugegriffen: 28. März 2022. [Online]. Verfügbar unter: https://de.wikipedia.org/w/index.php?title=Alan_Turing&oldid=218951197

[9] „Turing test“, *Wikipedia*. 20. März 2022. Zugegriffen: 28. März 2022. [Online]. Verfügbar unter: https://en.wikipedia.org/w/index.php?title=Turing_test&oldid=1078269420

[10] „Captcha“, *Wikipedia*. 3. Januar 2022. Zugegriffen: 30. März 2022. [Online]. Verfügbar unter: https://de.wikipedia.org/w/index.php?title=Captcha&oldid=218753212

[11] „AG1_Whitepaper_280619.pdf“. https://www.tu-darmstadt.de/media/daa_responsives_design/03_forschung_medien/forschungskompetenz/ki_forschung_1/AG1_Whitepaper_280619.pdf (zugegriffen 19. August 2020).

[12] „Home - KI Strategie“. https://www.ki-strategie-deutschland.de/home.html (zugegriffen 30. März 2022).

[13] „Künstliche Intelligenz I Ohne Normen und Standards geht es nicht“. https://www.din.de/de/forschung-und-innovation/themen/kuenstliche-intelligenz/normungsroadmap-ki (zugegriffen 19. August 2020).

[14] „Plattform Industrie 4.0“. https://www.plattform-i40.de/PI40/Navigation/DE/Home/home.html (zugegriffen 16. März 2021).

[15] „DCC - GoodPractice · GitLab“, *GitLab*. https://gitlab.com/ptb/dcc/dcc-goodpractice (zugegriffen 28. Juli 2022).

[16] „Redundancy (linguistics)“, *Wikipedia*. 20. März 2022. Zugegriffen: 16. April 2022. [Online]. Verfügbar unter: https://en.wikipedia.org/w/index.php?title=Redundancy_(linguistics)&oldid=1078188100

[17] „DCC - GoodPractice / Temperature Simplified · GitLab“, *GitLab*. https://gitlab.com/ptb/dcc/dcc-goodpractice/temperature (zugegriffen 28. Juli 2022).

[18] „XPath Tutorial“. https://www.w3schools.com/xml/xpath_intro.asp (zugegriffen 28. Juli 2022).