



# A MATHMET Quality Management System for data, software, and guidelines

Keith Lines<sup>1</sup>, Jean-Laurent Hippolyte<sup>1</sup>, Indhu George<sup>1</sup>, Peter Harris<sup>1</sup>

<sup>1</sup> National Physical Laboratory, Hampton Road, Teddington TW11 0LW, UK

## ABSTRACT

The European Metrology Network for Mathematics and Statistics (MATHMET) is creating a Quality Management System (QMS) to ensure that research outputs in the forms of data, software and guidelines are fit-for-purpose, achieve a sufficient level of quality, and are consistent with the aims of National Measurement Institutes to provide quality-assured and trusted outputs. The essential components of the QMS for all three forms of research output are discussed. On-line, interactive risk assessment tools that guide a user through the process of assigning an integrity level for the research outputs of data and software, are described. Examples of case studies that have been used to demonstrate the QMS are indicated.

**Section:** RESEARCH PAPER

**Keywords:** Quality Management System; MATHMET; data; software; guideline

**Citation:** Keith Lines, Jean-Laurent Hippolyte, Indhu George, Peter Harris, A MATHMET Quality Management System for data, software, and guidelines, Acta IMEKO, vol. 11, no. 4, article 8, December 2022, identifier: IMEKO-ACTA-11 (2022)-04-08

**Section Editor:** Eric Benoit, Université Savoie Mont Blanc, France

**Received** July 18, 2022; **In final form** December 2, 2022; **Published** December 2022

**Copyright:** This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The project 18NET05 MATHMET has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

**Corresponding author:** Keith Lines, e-mail: [keith.lines@npl.co.uk](mailto:keith.lines@npl.co.uk)

## 1. INTRODUCTION

The European Metrology Network for Mathematics and Statistics (MATHMET) [1] has been established to help bring a collaborative approach to addressing the needs of measurement scientists for expertise in applied mathematics, statistics, and computational tools. Ever more accurate, consistent, and traceable measurements will be vital to meeting challenges such as climate monitoring, clean energy, modern-day health care and sustainability. These measurements are often underpinned by new and increasingly complex mathematical and statistical techniques that are reliant on data sets and software.

Fit for purpose data sets, software, and guidelines to meet the requirements of the National Measurement Institutes (NMIs) and other stakeholder organisations and individuals, that will both draw on and contribute to MATHMET, will be vital to MATHMET's success. An outline of a MATHMET Quality Management System (QMS) for research outputs in the form of data, software, and guidelines was presented at the Mathematical and Statistical Methods for Metrology virtual workshop 2021 (MSMM 2021) [2]. Feedback from delegates helped confirm that

the ISO process-based approach taken, and described below, was appropriate.

This paper outlines the current version of the QMS, which has benefited from the feedback from MSMM 2021 and the input of other MATHMET members and is organised as follows. In section 2 the essential components of the QMS for all three research outputs are described, as well as on-line risk assessment tools that guide a user through the process of assigning an integrity level for the research outputs of data and software. In section 3, some examples of case studies that are being used to refine and demonstrate the QMS are indicated. The lessons learned from these case studies will be reported separately to this paper. Finally, conclusions are given in section 4.

## 2. COMPONENTS OF THE QMS

### 2.1. Background

The QMS follows a process-based approach as defined in ISO 9001:2015 [3] and related standards. This approach incorporates the "Plan-Do-Check-Act" (PDCA) cycle and risk-based thinking. Over a million organisations are certified to ISO 9001,

and NPL has held ISO 9001 certification for more than 25 years. NPL has also successfully applied TickITplus certification [4] (and its predecessor scheme TickIT), which builds on ISO 9001, to its software development activities. The experience gained from applying these schemes, plus the large number of ISO and IEEE standards and related documents supporting them, strongly implies that this approach is the right one to take.

## 2.2. Data and software

A quality management plan is key to the QMS components for data and software. This plan lists the quality management activities needed for a particular dataset or piece of software. These activities follow a typical life cycle from requirements capture to design and development, verification, and validation through to maintenance. Review is an important activity that is carried out throughout the life cycle.

As noted in section 1, risk assessment is a key element of the QMS, and risk is quantified using a value called an integrity level. The integrity level is a number between 1 and 4, where 1 indicates the lowest level of risk (for example, prototypes of software for internal use within an organisation) and 4 indicates the highest level (for example, software that is safety critical). The concept is analogous to, but should not be confused with, the safety integrity level of IEC 61508 [5].

The integrity level is used to decide the quality management activities and interventions to be listed on the plan. Higher integrity levels have a greater associated risk and therefore need more activities and interventions (for example, review by a third-party, independent of the team that developed the data or software).

For software, the QMS can include established quality procedures and templates from the MATHMET members.

The development of metrology software is usually a much smaller scale exercise than would normally be addressed using such a QMS. There are no large teams of developers to manage, a typical team may consist of no more than one or two people. The software may have a small number of highly specialist users, rather than an app distributed to many thousands of users with varying levels of technical expertise.

However, there are issues that a QMS can help manage.

For example:

- Such software is typically developed by metrologists rather than software engineers. It is strongly arguable that software engineering good practice should be a part of every modern-day scientist's toolkit of skills. However, analogous to how the guidance of a numerical analyst should be sought for certain mathematical problems, there are situations in which it is strongly advisable to consult a software engineer (e.g., safety-critical software). The QMS provides a framework to help make, document and review such decisions.
- Non-trivial mathematics is at the heart of metrology software. Even the simplest equations can become difficult to implement and maintain if an inappropriate implementation platform is selected.
- For large scale software development projects, roles such as user and developer are distinct and held by different people. That situation is often not the case for metrology software. Also, it is not always easy to define who the customer is for this software. Is the customer somebody within a funding body? Is the customer

somebody internal to the organisation acting as a proxy for someone in an external organisation?

- Lack of clarity of roles can lead to serious issues that could be prevented easily, not least finding the software after the original developers have left the organisation. If these developers also provide the service for which the software was developed (or were authors of the paper for which the software generated results) they will know where it's located. Could others find the software, and be sure the correct version has been accessed (not an older version that contains some serious bugs)?
- A related point is that some metrology software can be in use for a long time. Can the correct versions of the source code and documentation, for example explaining how the equations were derived, be found 20 or more years after initial release?
- Perhaps the software itself will never be released outside of the organisation in which it was developed. However, results such as calibration certificates and research papers, will be released outside of the organisation. Software quality management is no less important in such situations as it is when the software itself is released.
- Even the smallest piece of software must be traceable to the results it produces. The ongoing reproducibility crisis [6] would be eased if questions such as the following could be always answered easily "Which version of which script produced those results? The exact version please, not one containing subsequent modifications", "Which versions of which libraries did the script call?" and "What were the reasons these libraries were considered appropriate for this work?". Perhaps journals should be asking for the upload of scripts as well as the data the scripts processed. What may seem like tedious, unnecessary and time-consuming bureaucracy (particularly during some interesting and exciting research) could save considerably more tedium in the longer term.
- Following on from the above point, the provenance of packages and libraries is a key point to consider. A richly featured, but new and experimental, library may be appropriate for a prototype but not for generating certificates for customers. A proprietary closed-source package from a long-established supplier with a strong reputation for well-engineered software may be the right option. Alternatively, what better guarantee of quality can there be than an open-source package that has the input of many experts in a particular field? There are often no "right" or "wrong" answers, just decisions to made, documented and reviewed. Again, the QMS provides a framework to help with these tasks.
- It should never, ever be necessary to have to look at the code of even the smallest script to work out what it does. The mathematics must be documented in a way that can be independently verified without having to examine code. In some circumstances code comments, and perhaps an accompanying README file, will be sufficient. In other circumstances more thorough documentation will be required. Again, the QMS provides a framework to help decide what documentation will be necessary.

In summary, for software the QMS aims to minimize chances of errors and manage issues such as traceability and accountability, transferability, maintenance and reproducibility. The final three points need to be managed without the original developers being available.

Some activities listed in the quality plan are considered mandatory to ensure the software is of the required quality, and the QMS provides templates to cover, for example, requirements capture. However, if an alternative template is considered more appropriate for a particular project, then this template can be used instead. In this case, it is mandatory that the reasons for such decisions are recorded.

Quality management of data is less well established than quality management of software. Accordingly, developing this component of the QMS was arguably a research activity to some extent. A key concept is that integrity levels are adapted to apply to data as well as software. Also, the ISO 8000 series of standards [7] applies the same PDCA and risk-based paradigms to data quality management

On-line risk assessment tools, called Quality Assurance Tools, will assist with assigning integrity levels and generating quality management plans. Further details are provided in section 2.4

### 2.3. Guidelines

The aim of this component of the QMS is to ensure a sufficient level of quality in the development, assessment, and recommendation of existing and future guidelines for mathematics and statistics in metrology. It is intended to include:

1. A rigorous review process for existing and new guidelines involving external experts and MATHMET stakeholders
2. Quality management activities that shall accompany and improve guidelines that will be developed in the future in projects involving one or more members of MATHMET
3. A process to provide advice and feedback that will enable third parties to adapt and improve mathematical and statistical guidelines to the needs and requirements of the European metrology community and its stakeholders.

The processes applied by organisations such as ISO/IEC for standards development [8], Eurachem for its development of guidance documents [9], and NPL for its review and approval of documents, have been reviewed and used to steer the design of the QMS. However, the approach taken has been to present a QMS in 'skeleton' form that sets only high-level requirements on users into which the processes adopted by individual MATHMET members can comfortably fit.

For example, for a future guideline, the process comprises the stages of development, review and approval, publication, and maintenance. The stage of review and approval can be iterative and can involve separate steps that focus on different aspects, such as technical correctness or presentation and style. The stage of maintenance depends on the guideline being provided with appropriate metadata allowing versions (and changes between versions) to be tracked correctly.

For both types of guidelines (future and existing), the QMS involves completing a checklist comprising a set of questions, and making a recommendation based on the answers to those questions. For an existing guideline, the checklist considers:

- Whether the origin of the document is an established organisation
- Whether the document has been independently reviewed and approved to be issued
- Whether the document comes with appropriate metadata (such as title, author, unique identifier or version number, issuer, review date, etc.)
- Whether the document is adequately protected with respect to copyright and intellectual property rights
- What is the language of the document and whether it needs translation (for example, to English)
- Whether the document states the targeted audience or readership
- Whether the technical content of the document is relevant to the focus of MATHMET on mathematics and statistics for metrology
- Whether conclusions are clearly stated, appropriate and relevant
- Whether complete and appropriate acknowledgments are made (for example, to originating projects and funding sources)
- Whether complete, appropriate, and primary references are listed
- Whether the overall presentation of the material in the document is clear and understandable.

For a future guideline, additional questions are included covering:

- Whether the document is technically sound
- Whether the document has undergone adequate review relating to both technical and presentational aspects
- Whether notation and abbreviations have been adequately and clearly defined.

The QMS will be applied to five metrology case studies identified at the outset (see, for example, section 3 and [19]). The results of those will be assessed in terms of effectiveness, risk assessment and quality interventions by the QMS. The QMS will also be presented here at this workshop to stakeholders to gain feedback and to understand its effectiveness in meeting the needs of stakeholders.

### 2.4. On-line risk assessment tools for data and software

As noted in section 1, integrity levels for data and software are essentially a calculation involving criticality and complexity. Other factors may also need to be considered, such as the availability of suitably qualified developers. MATHMET will provide on-line risk assessment tools to guide the user through the process of calculating an integrity level and generating a quality management plan. The tool will be illustrated here at this workshop to stakeholders to gain feedback.

For the case of software, Table 1 and Table 3 list the different classifications relating to criticality of usage (CU) and complexity (CP; Table 2 concerns only data and will be discussed later). The choice of the classifications, between 'not critical' (1) and 'life critical' (4) for CU and between 'very simple' (1) and 'complex' (4) for CP, are quite subjective. However, the calculation of the software integrity level (SWIL), as detailed in Table 4, is then deterministic and undertaken automatically by the assessment tool. The user has the possibility to moderate the calculated SWIL, considering factors that influence the associated risk. For example, the SWIL might be reduced if there is an alternative

Table 1. Classifying data and software according to the criticality of usage (CU).

CU	Criticality of usage	Explanation
1	Not critical	<ul style="list-style-type: none"> <li>No danger of loss of income or reputation</li> <li>Short life, will not require maintenance in future</li> </ul>
2	Significant	<ul style="list-style-type: none"> <li><b>Potential</b> for loss of income or reputation</li> </ul>
3	Substantial	<ul style="list-style-type: none"> <li><b>Likely</b> to lead to loss of income or reputation</li> </ul>
4	Life critical	<ul style="list-style-type: none"> <li>May result in personal injury or loss of life</li> </ul>

Table 2. Classifying data according to complexity (CP).

CP	Complexity of data	Typical features
1	Very simple	<ul style="list-style-type: none"> <li>Commonly used datatypes</li> <li>Few datatypes</li> <li>Small amount of data</li> <li>Simple/unexpensive data infrastructure</li> <li>Simple uncertainty budget</li> </ul>
2	Simple	<ul style="list-style-type: none"> <li>Easy to visualise</li> <li>Moderate number of datatypes</li> <li>Moderate amount of data</li> <li>Intermediate data infrastructure</li> <li>Intermediate uncertainty budget</li> </ul>
3	Moderate	<ul style="list-style-type: none"> <li>Non-trivial datatypes</li> <li>Fair number of datatypes</li> <li>Large dataset</li> <li>Complex/expensive data infrastructure</li> <li>Complicated uncertainty budget</li> </ul>
4	Complex	<ul style="list-style-type: none"> <li>Non-trivial datatypes</li> <li>Combination of many non-trivial datatypes</li> <li>Very complex/expensive data infrastructure</li> <li>Very complicated uncertainty budget</li> </ul>

means of verification or increased if there is reliance on key staff. See Table 5 for examples of moderating factors. The reasons for such moderation of the SWIL should be recorded.

Once the SWIL is fixed, the assessment tool automatically generates a Software Quality Management Plan identifying those activities and interventions that are mandatory, recommended and not required. Table 6 lists activities related to the capture of user requirements for software. For example, for a SWIL of 2, documented user requirements are mandatory, and their review by the project team and (a proxy for) the customer is also mandatory but review by an independent person is not required. In contrast, an independent review is recommended for a SWIL of 3 and mandatory for a SWIL of 4.

Software quality requirements are set related to the stages in the software development cycle of:

- Capturing software functional requirements
- Software design
- Software coding
- Verification and validation
- Delivery, use and maintenance.

For the case of data, the assessment tool operates in a similar manner to that for software. It takes the user through a series of questions collected into the following sections:

- Dataset details
- Responsibilities (in terms of data managers, data administrators, data stewards and data technicians)
- Document control
- Complexity and criticality leading to the assignment of a data integrity level that can be moderated by the user
- Fitness for purpose
- Quality planning
- Quality monitoring, control and improvement
- Quality assurance
- Data understandability
- Metrological soundness.

Table 3. Classifying software according to complexity (CP).

CP	Complexity of program	Typical features
1	Very simple	<ul style="list-style-type: none"> <li>Elementary functionality, easy to understand</li> <li>Little or no control of an external system</li> <li>Simple mathematics</li> </ul>
2	Simple	<ul style="list-style-type: none"> <li>Simple functionality</li> <li>Straightforward control of a system</li> <li>Intermediate mathematics</li> </ul>
3	Moderate	<ul style="list-style-type: none"> <li>Large or very large programs</li> <li>Difficult to modify</li> <li>Complicated mathematics</li> </ul>
4	Complex	<ul style="list-style-type: none"> <li>Extremely complex functionality</li> <li>Complex feedback systems</li> <li>Very complicated mathematics</li> </ul>

Table 4. Calculating the integrity level for data (DIL) and software (SWIL).

	CP1	CP2	CP3	CP4
CU1	1	1	1	1
CU2	2	2	3	4
CU3	3	3	3	4
CU4	4	4	4	4

Table 5. Moderating factors for a calculated software integrity level (SWIL).

Moderating factors	Possible effect on SWIL
Alternative means of verification	Decrease
Modular approach	Decrease
Suitably trained staff available	Decrease
Difficult to test	Increase
Reliant on key staff	Increase
Inexperienced staff	Increase
Ambitious timescales	Increase
Ambitious requirements	Increase
New technology	Increase
Novel design	Increase

Table 6. Quality interventions for capturing software user requirements and their dependence on the calculated integrity level (X, R and M denote not required, recommended and mandatory, respectively)

Quality Requirement	SWIL1	SWIL2	SWIL3	SWIL4
Documented user requirements	M	M	M	M
Review by team	R	M	M	M
Review by suitably qualified independent person	X	X	R	M
Review by customer or proxy	M	M	M	M

In the above, ‘metrological soundness’ considers whether the dataset contains measured data or results derived from measurements or simulated (measured) data or combinations of those, whether the generation of the dataset is intended to be repeatable and reproducible (and how the repeatability and reproducibility conditions for the measurements are documented), how measurement uncertainty is evaluated and expressed, and how confidence in the generation of the dataset is demonstrated. These issues are important for datasets generated for, and to be used in, a metrological context.

The calculated data integrity level (DIL) determines the comprehensiveness of the data plan. The DIL, in turn, depends on the risks associated with the criticality of usage (see Table 1) and complexity (see Table 2) of the dataset. As with software, the DIL is calculated as detailed in Table 4. For example, for a DIL of 1, it is not necessary to provide, under the section on fitness for purpose, information about how the data life cycle will be documented whereas such information is mandatory for a dataset having a DIL of 4. Considering the section on ‘metrological soundness’, there is no difference between datasets having different data integrity levels. For a dataset having the highest DIL, the associated plan comprises information relating to 41 questions in total.

Figure 1 summarises the QMS processes for data and software.

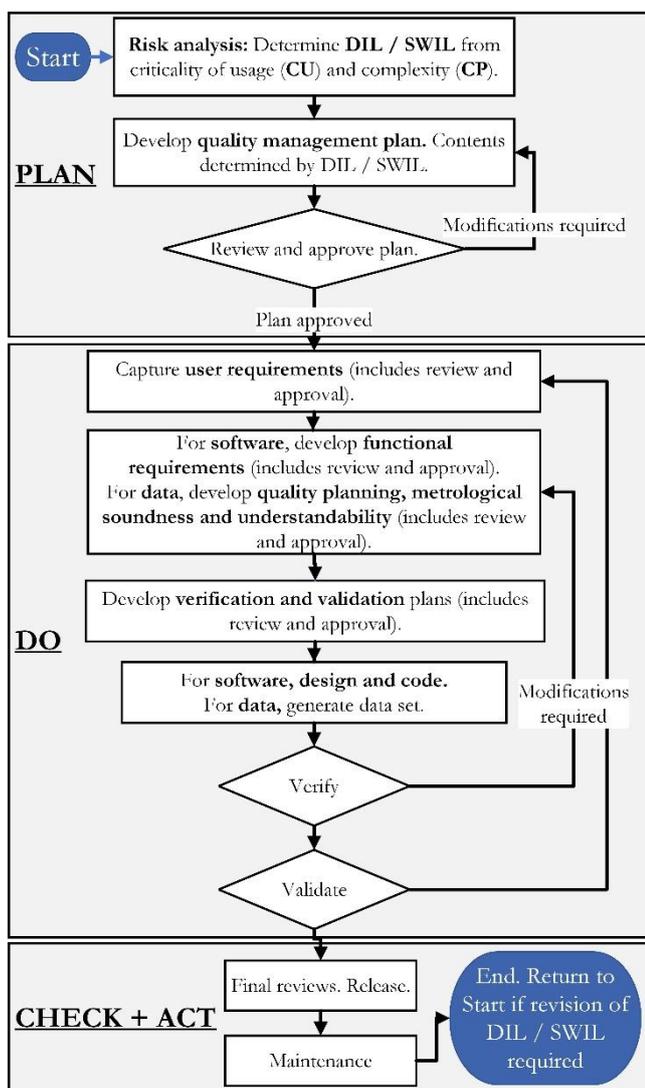


Figure 1. MATHMET QMS process flowchart.

### 3. CASE STUDIES

The acceptance and success of the MATHMET QMS will depend on its ability to address a variety of needs set both by the ‘owner’ of the research output as well as the user or customer at which the research output is aimed. To this end, case studies are being undertaken to support the development of the QMS and help to promote it to MATHMET members and stakeholders. They are chosen to illustrate the wide range of possible research outputs and are undertaken by MATHMET members having different experience of using a QMS. Case studies focussed on the application of the QMS to software, e.g., [10], [11], [12], and those focussed on data, e.g., [13], [14], [15], were described at MSMM 2021 [2]. Additional case studies focussed on guidelines have since been chosen to demonstrate the application of the QMS to that form of research output, and include:

- A Eurachem document on the use of uncertainty in compliance assessment [16]
- A good practice guide describing methods of metrological data processing for industrial process optimization, focusing on aspects of redundancy, synchronization and feature selection applied to data affected by measurement [17]
- Best practice guides on Bayesian inference for regression problems, uncertainty evaluation for computationally expensive models, and decision-making and conformity assessment [18]
- An internal MATHMET document containing a glossary and ontology of terms to support the QMS described in this abstract.

In [19] the application of the QMS to several of these case studies by different MATHMET members is presented, and the ease of use and possible pitfalls of the QMS are discussed. The lessons learned from the different case studies will be reported elsewhere.

### 4. CONCLUSIONS

The components of a Quality Management System (QMS), created by the European Metrology Network for Mathematics and Statistics (MATHMET), have been described. The aim of the QMS is to ensure that research outputs in the forms of data, software and guidelines are fit-for-purpose, achieve a sufficient level of quality, and are consistent with the aims of National Measurement Institutes to provide quality-assured and trusted outputs.

A pragmatic approach has been taken to the development of the QMS, which sets only high-level requirements on users to ensure there are no conflicts with the processes in-place and adopted by individual MATHMET members. A key element of the QMS is a risk assessment, and on-line tools that guide a user through the process of assigning an integrity level for the research outputs of data and software has also been presented.

### ACKNOWLEDGEMENT

The project 18NET05 MATHMET has received funding from the EMPIR programme co-financed by the Participating States and from the European Union’s Horizon 2020 research and innovation programme.

We thank the other members of the EMN MATHMET for their support in the development of the QMS described here.

## REFERENCES

- [1] MATHMET: European Metrology Network for Mathematics and Statistics home page. Online [Accessed 01 December 2022] <https://www.euramet.org/european-metrology-networks/mathmet/>
- [2] MSMM 2021 Mathematical and Statistical Methods for Metrology. Online [Accessed 01 December 2022] <http://www.msmm2021.polito.it/programme>
- [3] ISO 9001: Quality management systems – Requirements, 2015. Online [Accessed 01 December 2022] <https://www.iso.org/standard/62085.html>
- [4] TickIT<sup>plus</sup> home page. Online [Accessed 01 December 2022] <https://www.tickitplus.org/en/>
- [5] IEC 61508: Functional safety of electrical/electronic/programmable electronic safety-related systems, Part 0: Functional safety and IEC 61508, 2005
- [6] M. Baker, 1,500 scientists lift the lid on reproducibility, Nature volume 533, 2016, pp. 452–454. DOI: [10.1038/533452a](https://doi.org/10.1038/533452a)
- [7] ISO 8000-63:2019 Data quality — Part 63: Data quality management: Process measurement. Online [Accessed 01 December 2022] <https://www.iso.org/standard/65344.html>
- [8] ISO/IEC Directives and Policies. Online [Accessed 01/12/2022] <https://www.iso.org/directives-and-policies.html>
- [9] Procedure for the development of Eurachem guidance. Online [Accessed 01 December 2022] [https://www.eurachem.org/images/stories/Policies/Development\\_of\\_Eurachem\\_Guidance\\_2020.pdf](https://www.eurachem.org/images/stories/Policies/Development_of_Eurachem_Guidance_2020.pdf)
- [10] CASoft: Software for Conformity Assessment taking into account measurement uncertainty. Online [Accessed 01 December 2022] <https://www.lne.fr/en/software/CASoft>
- [11] MET4FOF: Metrology for the Factory of the Future. Online [Accessed 01 December 2022] <https://www.ptb.de/empir2018/met4fof/software/>
- [12] ISO 6142-1:2015, Gas analysis - Preparation of calibration gas mixtures - Part 1: Gravimetric method for Class I mixtures, ISO, Geneva, 2015.
- [13] MedalCare: Metrology of automated data analysis for cardiac arrhythmia management. Online [Accessed 01 December 2022] <https://www.ptb.de/empir2019/medalcare/home/>
- [14] P. Wagner, N. Strodthoff, R. Bousseljot, W. Samek, T. Schaeffter, PTB-XL, a large publicly available electrocardiography dataset. Online [Accessed 01 December 2022] <https://physionet.org/content/ptb-xl/1.0.1/>
- [15] TraCIM: Traceability for Computationally-Intensive. Metrology. Online [Accessed 01 December 2022] <https://www.tracim.eu/>
- [16] Eurachem: Use of uncertainty information in compliance assessment. (2nd ed. 2021). Online [Accessed 01 December 2022] <https://www.eurachem.org/index.php/publications/guides/uncertcompliance>
- [17] Y. Lo, P. Harris, L. Wright, K. Jagan, G. Kok, L. Coquelin, J. Zaouali, S. Eichstädt, T. Dorst, C. Tachtatzis, I. Andonovic, G. Gourlay, B. Xiang Yong, Good Practice Guide on Industrial Sensor Network Methods for Metrological Infrastructure Improvement, 63 pp. DOI: [10.5281/zenodo.6342744](https://doi.org/10.5281/zenodo.6342744)
- [18] Novel mathematical and statistical approaches to uncertainty evaluation. Online [Accessed 01 December 2022] <https://www.ptb.de/emrp/2976.html>
- [19] G. J. P. Kok, Use case examples for the MATHMET Quality Management System at VSL, IMEKO-MATHMET Symposium, Porto, Portugal, 31 August – 2 September 2022.