# Introduction

|  |
| --- |
| Figure 1 : Plan/Do/Check/Act pattern for data quality (source: ISO 8000) |

This document provides information to assist with completing a Mathmet Data Quality Assurance Plan (DQP). The associated interactive PDF supports the documentation of the people, procedures and activities performed to fulfil predefined quality objectives during the development and maintenance of a specific dataset or meaningful group of datasets, within the metrology community and wider scientific domains, in which quality and trustworthiness are essential.**.**. A completed data quality assurance plan facilitates the auditing of datasets, particularly with respect to the ISO 8000 [1] series of standards.

**Disclaimer**

The Quality Assurance Tools for data, software and guidelines have been provided by the Members and Partners of the European Metrology Network for Mathematics and Statistics (Mathmet). EURAMET has no influence on its correctness and completeness and does not assume any liability for it.

**Acknowledgements**

The European Metrology Network for Mathematics and Statistics is supported by the Joint Network Project ‘Support for a European Metrology Network for mathematics and statistics’ (18NET05 MATHMET). 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.

Contents

[Introduction 1](#_Toc122042635)

[Glossary 3](#_Toc122042636)

[Dataset details 5](#_Toc122042637)

[Q1 Dataset label 5](#_Toc122042638)

[Q2 Brief description 5](#_Toc122042639)

[Q3-Q6 Responsibilities 5](#_Toc122042640)

[Q7 DQM plan status / Q8a DQM plan date / Q8b DQM plan version 5](#_Toc122042641)

[Data Integrity Level 5](#_Toc122042642)

[Q9 Dataset level of complexity 5](#_Toc122042643)

[Q10 Dataset level of criticality 6](#_Toc122042644)

[Q11 Does the calculated DIL seem suitable? 7](#_Toc122042645)

[Fitness for purpose 7](#_Toc122042646)

[Q12 How will the dataset user requirements be documented? 7](#_Toc122042647)

[Q13 How will the data life cycle be documented? 7](#_Toc122042648)

[Q14 How will the project team approve the dataset user requirements internally? 8](#_Toc122042649)

[Q15 How will the customer or customer proxy approve the dataset user requirements? 8](#_Toc122042650)

[Q16 How will the objectivity of the dataset user requirements review be ensured? 8](#_Toc122042651)

[Data quality planning 8](#_Toc122042652)

[Q17 Describe how the roadmap of data quality procedures will be documented. 8](#_Toc122042653)

[Q18 Describe how the implementation progress against the roadmap for data quality procedures will be tracked. 8](#_Toc122042654)

[Q19 How will the risks associated to developing the dataset be assessed and mitigated? 8](#_Toc122042655)

[Q20 How will the potential negative impact on business/reputation of non-conformities in the dataset be assessed and mitigated? 8](#_Toc122042656)

[Quality Monitoring, Control, and Improvement 9](#_Toc122042657)

[Q21 How will non-conformities in the dataset be reported? 9](#_Toc122042658)

[Q22 How will corrective actions on non-conformities in the dataset be logged? 9](#_Toc122042659)

[Q23 How will non-conformities in the dataset be prevented? 9](#_Toc122042660)

[Q24 Who will oversee dealing with non-conformities in the dataset? 9](#_Toc122042661)

[Quality assurance 9](#_Toc122042662)

[Q25 How will data quality be documented and made available for review? 9](#_Toc122042663)

[Q26 How will the project team review data quality issues? 9](#_Toc122042664)

[Q27 How will the customer or customer proxy review data quality issues? 9](#_Toc122042665)

[Q28 How will the objectivity of the data quality issue review be ensured? 9](#_Toc122042666)

[Data understandability 10](#_Toc122042667)

[Q29 How will the meaning of the data be formalised? 10](#_Toc122042668)

[Q30 How will the structure of the dataset be documented? 10](#_Toc122042669)

[Q31 How will the metadata be documented? 10](#_Toc122042670)

[Q32 How will the customer know how to exploit the dataset? 10](#_Toc122042671)

[Metrological soundness 10](#_Toc122042672)

[Q33 Will the dataset contain measurement data or data derived from measurements? 10](#_Toc122042673)

[Q34 How will the main operations applied to the dataset be logged? 10](#_Toc122042674)

[Q35 How will the operations applied to the dataset be logged? 10](#_Toc122042675)

[Q36 Is the dataset generation intended to be repeatable? 10](#_Toc122042676)

[Q37 How will the dataset’s repeatability conditions be documented? 10](#_Toc122042677)

[Q38 Is the dataset generation intended to be reproducible? 10](#_Toc122042678)

[Q39 How will the dataset’s reproducibility conditions be documented? 10](#_Toc122042679)

[Q40 How will measurement uncertainty be expressed and evaluated? 11](#_Toc122042680)

[Q41 How will confidence in the dataset generation process be demonstrated? 11](#_Toc122042681)

[Appendix I 12](#_Toc122042682)

[References 13](#_Toc122042683)

[Document History 13](#_Toc122042684)

# Glossary

|  |  |  |
| --- | --- | --- |
| Term | Definition | Source |
| **Audit** | Systematic, independent, and documented process for obtaining objective evidence and evaluating it objectively to determine the extent to which the audit criteria are fulfilled. | [ISO 9000:2015](https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en) |
| **Business impact analysis** | Process of analysing the impact over time of a disruption on the organization. | [ISO 22300:2021](https://www.iso.org/obp/ui/#iso:std:iso:22300:ed-3:v1:en:term:3.1.24) |
| **Customer** | Person or organization that could or does receive a product or a service that is intended for or required by this person or organization. A customer can be internal or external to the organization. | [ISO 9000:2015](https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en) |
| **Data** | Reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. | [ISO 8000-2:2020](https://www.iso.org/obp/ui#iso:std:iso:8000:-2:ed-4:v1:en) |
| **Data administrator** | Person who controls and coordinates the work of data technicians by defining criteria needed to maintain data quality, by designing data schemata, and by analysing the causes of errors to prevent their recurrence. | [ISO 8000-2:2020](https://www.iso.org/obp/ui#iso:std:iso:8000:-2:ed-4:v1:en) |
| **Data dictionary** | Collection of data dictionary entries that allows lookup by entity identifier. | [ISO 8000-2:2020](https://www.iso.org/obp/ui#iso:std:iso:8000:-2:ed-4:v1:en) |
| **Data Integrity Level (DIL)** | A risk assessment metric that accounts for complexity and criticality levels. |  |
| **Data life cycle** | All stages in the process of data usage from initial establishment to its discontinuation. | [ISO 21710:2020](https://www.iso.org/obp/ui/#iso:std:iso:21710:ed-1:v1:en:term:3.12) |
| **Data manager** | Person who establishes plans for data quality improvement in an organization, grants data administrators the authority to trace and correct data over the information systems or organization, and maintains data consistency in individual information systems through the organization-wide data architecture (ISO 8000-2:2020). | [ISO 8000-2:2020](https://www.iso.org/obp/ui#iso:std:iso:8000:-2:ed-4:v1:en) |
| **Data processing** | Systematic performance of operations upon data. Examples: Arithmetic or logic operations upon data, merging or sorting of data, assembling or compiling of programs, or operations on text, such as editing, sorting, merging, storing, retrieving, displaying, or printing. | [ISO 2382:2015](https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:ed-1:v1:en) |
| **Data quality** | Degree to which a set of inherent characteristics of data fulfils requirements. | [ISO 8000-2:2020](https://www.iso.org/obp/ui#iso:std:iso:8000:-2:ed-4:v1:en) |
| **Data quality management** | Coordinated activities to direct and control an organization with regard to data quality. | [ISO 8000-2:2020](https://www.iso.org/obp/ui#iso:std:iso:8000:-2) |
| **Data set (or dataset)** | Logically meaningful grouping of data. | [ISO 8000-2:2020](https://www.iso.org/obp/ui#iso:std:iso:8000:-2:ed-4:v1:en) |
| **Data specification** | Set of requirements covering the characteristics of data being fit for one or more particular purposes. | [ISO 8000-2:2020](https://www.iso.org/obp/ui#iso:std:iso:8000:-2:ed-4:v1:en) |
| **Data steward** | Person or organization delegated the responsibility for managing a specific set of data resources. | [ISO 8000-2:2020](https://www.iso.org/obp/ui#iso:std:iso:8000:-2:ed-4:v1:en) |
| **Data technician** | Person who creates, reads, modifies, and deletes data in accordance with the guidelines for data quality management, and measures data quality and corrects data errors found as a result of measuring data quality. | [ISO 8000-2:2020](https://www.iso.org/obp/ui#iso:std:iso:8000:-2:ed-4:v1:en) |
| **Management system** | A management system is a set of interrelated or interacting elements of an organization to establish:  • policies, intentions and direction of an organization as formally expressed by its top management,  • and objectives, results to be achieved,  • and processes, sets of interrelated or interacting activities that use inputs to deliver an intended result, to achieve those objectives. | [ISO 9000:2015](https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en) |
| **Measurement uncertainty** | parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand | [JCGM 100:2008](https://www.bipm.org/en/committees/jc/jcgm/publications) |
| **Non-conformity** | Non-fulfillment of a requirement. | [ISO 33020:2019](https://www.iso.org/obp/ui/#iso:std:iso-iec:33020:ed-2:v1:en:term:3.3) |
| **Quality** | The degree to which a set of inherent characteristics of an object fulfils requirements. | [ISO 9000:2015](https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en) |
| **Quality control** | Part of quality management focused on fulfilling quality requirements. | [ISO 9000:2015](https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en) |
| **Quality management system realization** | Process of establishing, documenting, implementing, maintaining, and continually improving a quality management system. | [ISO 9000:2015](https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en) |
| **Quality planning** | Part of quality management focused on setting quality objectives and specifying necessary operational processes, and related resources to achieve the quality objectives. | [ISO 9000:2015](https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en) |
| **Quality requirement** | Need or expectation that is stated, generally implied or obligatory, related to data. | [ISO 9000:2015](https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en) |
| **Repeatability condition of measurement** | Condition of measurement, out of a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time. | JCGM 200:2012 See reference [2] |
| **Reproducibility condition of measurement** | Condition of measurement, out of a set of conditions that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects. | JCGM 200:2012 See reference [2] |
| **Risk analysis** | Systematic use of available information to identify hazards and to estimate the risk. | [ISO 51:2014](https://www.iso.org/obp/ui/#iso:std:iso-iec:guide:51:ed-3:v1:en:term:3.10) |
| **Risk assessment** | Overall process comprising a risk analysis and a risk evaluation. | [ISO 51:2014](https://www.iso.org/obp/ui/#iso:std:iso-iec:guide:51:ed-3:v1:en:term:3.11) |
| **Risk evaluation** | Procedure based on the risk analysis to determine whether tolerable risk has been exceeded. | [ISO 51:2014](https://www.iso.org/obp/ui/#iso:std:iso-iec:guide:51:ed-3:v1:en:term:3.12) |
| **Uncertainty budget** | Statement of a measurement uncertainty, of the components of that measurement uncertainty, and of their calculation and combination. | JCGM 200:2012 See reference [2] |

# Dataset details

## Q1 Dataset label

Provide a meaningful name for the dataset(s). The name will also serve as a title for the **Data Quality Assurance Plan**.

## Q2 Brief description

Give general information about the project. Describe what the dataset is, what its purpose is, where it is stored (e.g., link to a **data dictionary**), where the project’s documentation is stored etc.

## Q3-Q6 Responsibilities

A **Data Quality Assurance Plan** is established by the dataset development team. Team members may have different roles: **data manager**(s), **data administrator**(s), **data technician**(s) and **data steward**(s). These roles are defined by the ISO 8000 series of standards. The most essential roles are **data technician** and **data steward**. Smaller development teams might be composed of a small number of **data technicians** and/or **data stewards**. For example, a team might have a single member, who creates (as **data technician**) and manages (as **data steward**) the dataset.

## Q7 DQM plan status / Q8a DQM plan date / Q8b DQM plan version

Provide information about the **Data Quality Assurance Plan** itself: the current status of the plan, still a draft or issued, the latest date when the plan underwent significant changes, and an identifier for the current version of the plan.

# Data Integrity Level

A **risk analysis** results in the calculation of a **Data Integrity Level** (DIL). The DIL is a numeric scale, between 1 and 4, that reflects the potential impact of quality on the overall dataset development project.

## Q9 Dataset level of complexity

Evaluate the anticipated complexity of the dataset(s). Dataset complexity (CP) can be defined considering one or more of: (a) size of the dataset(s), (b) amount of effort needed for dataset consolidation, and (c) dataset understandability.

Table 1 : Data complexity levels, typical features, and examples.

|  |  |  |  |
| --- | --- | --- | --- |
| **CP** | **Complexity of data** | **Typical features** | **Example** |
| 1 | Very simple | •Commonly used datatypes (e.g. numeric). | Hourly time series of a single room temperature sensor readings over one month. |
| •Few datatypes. |
| •Small amount of data. |
| •Simple/unexpensive capture/generation/processing/storage infrastructure. |
| •Simple uncertainty budget |
| 2 | Simple | •Easy to represent visually (e.g. numeric matrix). | Hourly time series of room temperature, relative humidity and CO2 concentration sensor readings of a large building over one year. |
| •Moderate number of datatypes. |
| •Moderate amount of data. |
| •Intermediate capture/generation/processing/storage infrastructure. |
| •Intermediate uncertainty budget |
| 3 | Moderate | •Non-trivial datatypes | Combined thermographic, photogrammetric and 3D point cloud data of a large building. |
| •Fair number of datatypes. |
| •Large or very large datasets. |
| •Complex/expensive capture/generation/processing/storage infrastructure |
| •Complicated uncertainty budget |
| 4 | Complex | •Non-trivial datatypes | Multi-spectral satellite imagery of a continent. |
| •Combination of many datatypes. |
| •Very complex/expensive capture/generation/processing/storage infrastructure. |
| •Very complicated uncertainty budget |

## Q10 Dataset level of criticality

Evaluate the expected criticality of usage (CU) the product dataset(s), i.e. the consequences of customers using data of inappropriate quality, particularly on business/reputation.

Table 2 : Data criticality levels, explanations, and examples.

|  |  |  |  |
| --- | --- | --- | --- |
| **CU** | **Criticality of usage** | **Explanation** | **Example** |
| 1 | Not critical | No danger of loss of income or reputation. | Prototype dataset for internal use. |
| Short life, will not require maintenance in future. |
| 2 | Significant | Potential for loss of income or reputation. | Dataset used for scientific research. |
| 3 | Substantial | Likely to lead to loss of income or reputation if faulty. | Dataset used for commercial applications. |
| 4 | Life critical | May result in personal injury or loss of life. | Avionics reference standards. |

## Q11 Does the calculated DIL seem suitable?

After risk has been analysed, the interactive PDF tool will compute a DIL value according to the rules from Table 3. This recommended DIL can be accepted, or it could be revised by the team. For example, if inexperienced **data technicians** are involved, the team can decide to increase the DIL. If the DIL value is revised, then the project-specific factors that justify increasing or decreasing the recommended DIL value must be documented.

Once the DIL value is set, the interactive PDF tool provides an appropriate list of recommended or compulsory questions explained in the following sections. The mapping between recommended questions and DIL values is given in Appendix I.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3 : Calculating Data Integrity Level.   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | **CP1** | **CP2** | **CP3** | **CP4** | | CU1 | 1 | 1 | 1 | 1 | | CU2 | 2 | 2 | 3 | 4 | | CU3 | 3 | 3 | 3 | 4 | | CU4 | 4 | 4 | 4 | 4 | |

# Fitness for purpose

## Q12 How will the dataset user requirements be documented?

Document user requirements (**data specification**) in any case, regardless of the DIL value. However, it is expected that higher DIL levels have more detailed user requirements. For example, datasets with higher levels of complexity might have longer mathematical specifications.

From Q12, all answer fields support rich and long text, however good practice is to provide a short explanation and a link to a working document (stored in a persistent accessible location).

## Q13 How will the data life cycle be documented?

Document the **data life cycle**: from creation to long-term preservation. For example, provide a short explanation and link to a flow diagram, or for a more complex project, link a Gantt chart, a task board etc. **Data life cycle** processes can include how data is captured (how it is discovered, acquired, recombined from static or dynamic sources), how it is created (data pipeline, if it is synthetic data), how it is formatted (and what format conversions have been applied), how it is annotated (are there metadata?), how it is validated (detection/mitigation of abnormalities is also covered by the data quality monitoring section of the plan), how it is anonymised (for privacy purposes), how the data will be preserved (e.g., is there a retention period?) etc.

## Q14 How will the project team approve the dataset user requirements internally?

Describe the procedure for approval of the user requirements within the development team. Who oversees collating user requirements? Who approves them within the team? How is the approval traced (email thread, merge request if using a version control system, digital signature, meeting minutes…)?

## Q15 How will the customer or customer proxy approve the dataset user requirements?

Who is the **customer**? Who is the **customer** proxy, if any? Describe the procedure for approval by the **customer**. How is user requirement documentation communicated to **customer**? How is the approval traced (email thread, merge request, digital signature, meeting minutes…)? If the **customer** is not involved in this procedure, which team members act as a proxy?

## Q16 How will the objectivity of the dataset user requirements review be ensured?

Describe how an independent person/organisation will review the user requirements. Depending on the criticality of usage of the dataset, it could be someone in the same organisation with expertise but not directly involved in the development, or an external assessor with an accepted level of authoritativeness/accreditation.

# Data quality planning

## Q17 Describe how the roadmap of data quality procedures will be documented.

Detail the roadmap for the implementation of the data quality interventions, for example by providing a link to a document: a flow diagram (for example, a sub-diagram of the data life cycle diagram) or for more complex projects: Gantt chart, task board… It could also be a link to an external platform, e.g., a project tracker.

## Q18 Describe how the implementation progress against the roadmap for data quality procedures will be tracked.

Describe the procedure that will be implemented to track progress against the roadmap. For example, progress review meetings with minutes, or approval procedure in a project tracker.

## Q19 How will the risks associated to developing the dataset be assessed and mitigated?

Document the risk assessment associated to the development of the dataset. What could go wrong and how to mitigate? For examples, having redundant instruments, setting up data backup procedures.

## Q20 How will the potential negative impact on business/reputation of non-conformities in the dataset be assessed and mitigated?

For higher DIL values, provide a more detailed criticality assessment and mitigation plan, e.g., a **business impact analysis**.

# Quality Monitoring, Control, and Improvement

## Q21 How will non-conformities in the dataset be reported?

Describe how non-conformities are identified and documented. Internal reporting and customer reporting could have different channels. For example, the development team can set up a procedure to assign priority levels to non-conformities. ISO 8000 standards characterise data quality according to its consistency with its format (syntactic quality), its meaning (semantic quality), and its usefulness (pragmatic quality).

## Q22 How will corrective actions on non-conformities in the dataset be logged?

Provide a short description and/or a link to an external document, or a link to an interactive platform such as an issue tracker.

## Q23 How will non-conformities in the dataset be prevented?

Describe the procedures that will be implemented to prevent non-conformities. For example, provide a link to a testing/verification plan or to a comparison dataset.

## Q24 Who will oversee dealing with non-conformities in the dataset?

Assign team members or organisational units to the management of non-conformities in the dataset(s).

# Quality assurance

## Q25 How will data quality be documented and made available for review?

Provide a link to a document (for example, a shared Word document in review mode), or a link to a dedicated platform (for example, an issue tracker).

## Q26 How will the project team review data quality issues?

Provide, for example, a link to an agenda for review meetings with minutes, or a link to a chat history that is private to the team.

## Q27 How will the customer or customer proxy review data quality issues?

Describe how data quality issues will be communicated to the customer (through email, in meetings with minutes, via an access to an issue tracker...).

## Q28 How will the objectivity of the data quality issue review be ensured?

Describe how the processes for review and mitigation of data quality issues will be reviewed independently. Depending on the criticality level, the independent reviewer can be someone in the same organisation who is not involved in the development of the dataset, or an external expert.

# Data understandability

## Q29 How will the meaning of the data be formalised?

Provide a link to a **data specification** document. A machine-interpretable format specification must be provided if compliance with the [FAIR principles](https://www.go-fair.org/fair-principles/) is desired, e.g., the dataset is intended to be shared in an open science data repository.

## Q30 How will the structure of the dataset be documented?

Provide a link to a data model document/diagram (e.g., a UML diagram), or a database schema with metadata (e.g., the dataset is kept in a relational database). A machine-interpretable format specification needs to be provided to comply with the [FAIR principles](https://www.go-fair.org/fair-principles/).

## Q31 How will the metadata be documented?

Provide a link to metadata database schema, a controlled vocabulary, a taxonomy… In simpler cases, a README file that explains the headers, if the data is shared in a tabular format. A machine-interpretable format specification needs to be provided to comply with the [FAIR principles](https://www.go-fair.org/fair-principles/).

## Q32 How will the customer know how to exploit the dataset?

Detail how customers will learn about the dataset usage. For example, provide a link to a user manual in PDF format, or a link to a documentation wiki/web site. For more complex datasets/intended usage, describe how training will be provided, e.g., online or via in-person workshops/classes…

# Metrological soundness

## Q33 Will the dataset contain measurement data or data derived from measurements?

## Q34 How will the main operations applied to the dataset be logged?

If the dataset contains measurement data, describe the logging procedure for the main operation, e.g., by providing a link to a document/diagram.

## Q35 How will the operations applied to the dataset be logged?

For higher levels of criticality/complexity, describe all operations performed on or in relation to the dataset, e.g., provide a processing pipeline specification.

## Q36 Is the dataset generation intended to be repeatable?

## Q37 How will the dataset’s repeatability conditions be documented?

*Condition of measurement, out of a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time.* ([VIM: International Vocabulary of Metrology](https://www.bipm.org/en/committees/jc/jcgm/publications) [2])

## Q38 Is the dataset generation intended to be reproducible?

## Q39 How will the dataset’s reproducibility conditions be documented?

*Condition of measurement, out of a set of conditions that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects.* [2]

## Q40 How will measurement uncertainty be expressed and evaluated?

Document how **measurement uncertainty** is expressed/evaluated. Refer to the [Guide to the expression of uncertainty in measurement (GUM) [3]](https://www.bipm.org/en/committees/jc/jcgm/publications) when possible, or provide links to explanatory documents (reports, papers) when the uncertainty evaluation is not traceable to the GUM. If there is an uncertainty budget, provide a link and explain how it is traceable to the GUM, if not, justify why there isn't one.

If the evaluation is performed using commercial software: specify the tools and version used. If it is performed using in-house software, specify tools and version used, and link to the source code and/or a software quality plan.

## Q41 How will confidence in the dataset generation process be demonstrated?

For higher levels of criticality of usage, this could be a link to a recent ISO 17025 accreditation certificate.

# Appendix I

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Data Quality Planning** | | | | **DIL 1** | **DIL 2** | **DIL 3** | **DIL 4** |
| **Fitness for purpose** | | **Documented user requirements** | | M | M | M | M |
| **Documented data life cycle** | | X | R | M | M |
| **Review by team** | | R | M | M | M |
| **Review by customer or proxy** | | X | R | M | M |
| **Review by independent expert** | | X | X | M | M |
| **Data quality planning** | | **Roadmap of data quality processes** | | X | R | M | M |
| **Document progress and updates against initial implementation roadmap** | | X | R | M | M |
| **Metrological soundness and traceability** | | **Log of main operations applied to the data set** | | R | M | X | X |
| **Log of all operations applied to the data set** | | X | R | M | M |
| **Document repeatability conditions** | | R | M | M | M |
| **Statement of reproducibility** | | R | R | M | M |
| **Measurement uncertainty expression and evaluation traceable to the GUM** | | R | R | R | R |
| **Dataset produced in an ISO 17025 compliant environment** | | R | R | R | R |
| **Data Quality Monitoring, Control and Improvement** | | **Risk assessment report** | | X |  | R | M |
| **Business impact report** | | X | R | R | M |
| **Data non-conformity report** | | R | M | M | M |
| **Log of data non-conformity corrections** | | X | R | M | M |
| **Document procedures to prevent data non-conformities** | | X | R | M | M |
| **Data quality committee/team meeting reports** | | X | R | M | M |
| **Data Quality Assurance** | | **Data Quality Report** | | X | R | M | M |
| **Review of data quality issues by team** | | R | M | M | M |
| **Review of data quality issues by customer or proxy** | | X | M |  |  |
| **Review of data quality issues by non-team internal expert** | | X |  |  |  |
| **Review of data quality issues by non-team external expert** | |  |  |  |  |
| **Data understandability** | | **Existence of data model** | | R | R | M | M |
| **Existence of data dictionary** | | R | R | R | R |
| **Provision of guidance for data usage** | | R | R | R | M |
| X | Not required | |
| R | Recommended | |
| M | Mandatory | |

# References

[1] ISO 8000-1:2022 Data quality — Part 1: Overview. Retrieved 22/11/2022 from ISO:

<https://www.iso.org/standard/81745.html>

[2] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. International vocabulary of metrology | Basic and general concepts and associated terms (VIM). Joint Committee for Guides in Metrology, JCGM 200:2012. (3rd edition). URL (Retrieved 14/12//2022 from BIPM): [https://www.bipm.org/documents/20126/2071204/JCGM 200\_2012.pdf/f0e1ad45-d337-bbeb-53a6-15fe649d0ff1](https://www.bipm.org/documents/20126/2071204/JCGM%20200_2012.pdf/f0e1ad45-d337-bbeb-53a6-15fe649d0ff1)

[3] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Guide to the expression of uncertainty in measurement | Part 6: Developing and using measurement models. Joint Committee for Guides in Metrology, JCGM GUM-6:2020. URL (Retrieved 14/12/2022 from BIPM): <https://www.bipm.org/documents/20126/2071204/JCGM_GUM_6_2020.pdf/d4e77d99-3870-0908-ff37-c1b6a230a337>

# 

# Document History

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version** | **Date** | **Revised by** | **Description** | **Approved by** |
| Draft 0.1 | 26/05/2021 | JLH | Created draft outline | N/A |
| Draft 0.1 | 07/06/2021 | JLH | Started glossary | N/A |
| Draft 0.1 | 21/06/2021 | JLH | Started introduction | N/A |
| Draft 0.1 | 22/06/2021 | JLH | Started responsibilities and DIL sections | N/A |
| Draft 0.1 | 06/07/2021 | JLH | Added revision history table | N/A |
| Draft 0.2 | 27/04/2022 | JLH | Amended glossary | N/A |
| Draft 0.3 | 26/05/2022 | JLH | Revised introduction | N/A |
| Draft 0.4 | 01/06/2022 | JLH | Added to outline | N/A |
| Draft 1.0 | 15/09/2022 | JLH | Wrote up explanations for questions. Simplified structure of the document. | N/A |
| Draft 1.0.1 | 06/10/2022 | JLH | Addressed Peter Harris’ comments | N/A |
| Draft 1.0.2 | 17/11/2022 | JLH | Added disclaimer and made changes to comply with Euramet’s requests | N/A |
| Draft 1.0.3 | 15/12/2022 | KL | Added feedback from Mathmet chair and KL, e.g. added references. | N/A |
| Draft 1.0.4 | 20/12/2022 | KL | Added DRAFT watermark | N/A |