European Partnership on Metrology
Call 2022 – Digital Transformation, Health, Integrated European Metrology, Normative and Research Potential

Selected Research Topic number: **SRT-d04**
Version: 1.0

Title: Quality of data in the European Open Science Cloud

Abstract

The European Open Science Cloud (EOSC) is part of European Union's policy agenda 2022-2024 with the objective to deepen open science practices in Europe. The aim of the EOSC is to provide European researchers with an open, multi-disciplinary platform (data space) where they can publish, find and reuse data, tools and services for research, innovation and educational purposes. However, the sharing and reuse of measurement data from different domains poses new challenges for ensuring data quality and interoperability. Metrological traceability, uncertainty, FAIR (Findable, Accessible, Interoperable, and Reusable) principles and semantics (i.e. the meaning of data) are all key drivers for assessing data quality. Thus, data spaces such as the EOSC, urgently need agreed processes for determining data quality, based on these drivers. Only by providing a metrological basis for data quality principles and interoperability can the EOSC, be successfully used for the exchange of data and fully achieve the EU's objective.

Keywords

European Open Science Cloud, Data Quality, Metadata, FAIR principles, Data interoperability, Semantics.

Background to the Metrological Challenges

Modern data spaces like the EOSC have been established to provide European researchers with an open platform where they can publish, find and reuse data. The EOSC is based on open science and FAIR principles, however, the successful reuse of data relies on 'good' data quality and interoperability. In turn data quality, and its availability, relies on how well data is processed, curated, its metrological traceability, uncertainty, semantics and integration in a digital workflow. But there are currently no principles for ensuring the high quality or consistency of data and metadata across different domains, which limits the reuse of data.

A vision for a digital framework has been published by the International Conference of Weights and Measures (CIPM). The SI Digital Framework uses fundamental metrological entities (i.e. SI units) to provide a basic layer for digital applications with future machine-actionable data. The vision also emphasised a need for semantics, FAIR principles compliance, and suitable data stewardship as key tools for enabling data and metadata from metrology to be ready for automatic analysis by machines. However, the SI Digital Framework is currently still a vision and further work is needed to implement it.

Currently, metrology communities, standardisation bodies and accreditation bodies have well established standards for conveying to humans, the necessary metrological information for different applications. So far there have been some adoptions of these standards to digital formats (e.g., by the US's National Information Standards Organization Standards Tag Suite (NISO STS) but progress has been hampered due to (i) a lack of common principles on data quality, (ii) agreed trustworthiness of data and (iii) agreed reproducibility of data for autonomous processing. A collaboration between the IEC and ISO recently developed a utility model for smart standards outlining five levels of maturity for digital content: (i) digital document; (ii) machine-readable document; (iii) machine-readable content; (iv) machine-interpretable content; and (v) machine-controllable content. So far digital formats for smart standards (e.g. NISO STS) are only at the second level of this utility model. Further initiatives within ISO/IEC, on data quality, have also emphasised the following needs:

- accurate, reliable, and verifiable data, structured according to a common understanding of the meaning and representation of metrological information in a harmonised and translatable way.
- data quality for Machine Learning (ML) and Artificial Intelligence (AI) methods.
- cooperation with existing initiatives to harmonise vocabularies and guidelines.

Objectives

Proposers should address the objectives stated below, which are based on the PRT submissions. Proposers may identify amendments to the objectives or choose to address a subset of them in order to maximise the overall impact, or address budgetary or scientific / technical constraints, but the reasons for this should be clearly stated in the protocol.

The JRP shall focus on metrology research necessary to support digital transformation and the quality and interoperability of data in the EOSC.

The specific objectives are

1. To develop robust principles and guidelines for 'good' data quality and interoperability, that can be used to ensure the consistency of metrologically relevant data and metadata for at least 4 different use cases, each in a different domain of application. The principles should include pragmatic definitions for at least 6 different situations (e.g. sensing, data fusion, interoperation, analysis, information retrieval, and data exchange through the EOSC). The use of (i) semantics, (ii) FAIR data principles, (iii) Human-to-Machine and Machine-to-Machine communication, and (iv) AI and ML should also be evaluated.
2. Using the principles from Objective 1, to determine reliable metrics for assessing the quality of data, across at least 4 different domains of application (e.g. manufacturing, healthcare, environmental monitoring, calibration). The metrics should include (i) indicators for levels of agreement and (ii) different maturity levels of implementation. In addition, to develop methods for the automated application of these metrics.
3. To compare the methods from Objectives 1 and 2, against existing resources for digital representations of metrological data. Evaluation should focus on (i) traceability to the International System of Units (SI), (ii) assessment of the metrological soundness of data and metadata and (iii) alignment of vocabularies and guidance.
4. To validate the principles from Objective 1 and the metrics and methods from Objective 2, by demonstrating their use with the EOSC in at least 4 different test cases e.g. manufacturing, healthcare, environmental monitoring, calibration. Based on the results to produce recommendations and guide practice guidelines for end users for ensuring 'good' data quality & interoperability in the EOSC.
5. To facilitate the take up of the technology and measurement infrastructure developed in the project by the EOSC, the measurement supply chain, standards developing organisations (ISO TC 215, ISO/IEC JTC 1/SC42, EURAMET Technical Committee on Interdisciplinary Metrology Working Group on Digital Transformation in Metrology (TC-IM WG M4D) and end users (industry, manufacturing, healthcare, environmental monitoring)

These objectives will require large-scale approaches that are beyond the capabilities of single National Metrology Institutes and Designated Institutes. To enhance the impact of the research, the involvement of the appropriate user community such as industry, standardisation and regulatory bodies, and other European Partnerships is strongly recommended, both prior to and during methodology development.

Proposers should establish the current state of the art and explain how their proposed project goes beyond this. In particular, proposers should outline the achievements of the EMRP project NEW06 TraCIM and the EMPIR project 17IND02 SmartCom and how their proposal will build on those.

EURAMET expects the average EU Contribution for the selected JRPs in this TP to be 1.7 M€ and has defined an upper limit of 2.4 M€ for this project.

EURAMET also expects the EU Contribution to the external funded beneficiaries to not exceed 35 % of the total EU Contribution across all selected projects in this TP.

Any industrial beneficiaries that will receive significant benefit from the results of the proposed project are expected to be beneficiaries without receiving funding or associated partners.

Potential Impact

Proposals must demonstrate adequate and appropriate participation/links to the 'end user' community, describing how the project partners will engage with relevant communities during the project to facilitate knowledge transfer and accelerate the uptake of project outputs. Evidence of support from the 'end user' community (e.g. letters of support) is also encouraged.

You should detail how your JRP results are going to:

- Address the SRT objectives and deliver solutions to the documented needs,
- Feed into the development of urgent documentary standards through appropriate standards bodies,
- Transfer knowledge to industry, manufacturing, healthcare and environmental monitoring sectors.

You should detail other impacts of your proposed JRP as specified in the document “Guide 4: Writing Joint Research Projects (JRPs)”

You should also detail how your approach to realising the objectives will further the aim of the Partnership to develop a coherent approach at the European level in the field of metrology and include the best available contributions from across the metrology community. Specifically, the opportunities for:

- improvement of the efficiency of use of available resources to better meet metrological needs and to assure the traceability of national standards
- the metrology capacity of EURAMET Member States whose metrology programmes are at an early stage of development to be increased
- organisations other than NMIs and DIs to be involved in the work.

Time-scale

The project should be of up to 3 years duration.