



*Characterisation of metadata to enable high quality climate applications and services*

## **Deliverable 400.2**

### **Recommendations for changes to Metafor CIM to incorporate ISO 19157**

Partners providing input:

<b>version number</b>	<b>date</b>	<b>comments</b>
0.3	28/03/14	Spiros Ventouras, Philip Kershaw: initial draft for review amongst project partners
1.0	03/06/14	Spiros Ventouras, Philip Kershaw: Updated following review comments from partners



CHARMe is funded by the EC under its FP7 Research Programme

---

# 1 Contents

1	Contents.....	2
2	References.....	3
3	Introduction .....	4
4	Assessment of current CIM model and representation of quality info .....	4
5	Assessment of ISO 19157.....	6
6	Identification of ISO 19157 concepts that might improve CIM .....	9
7	The impact of CIM to CHARMe .....	9
8	Conclusions .....	9

## 2 References

1. Characterisation of metadata to enable high-quality climate applications and services, Proposal for EU call FP7-SPACE-2012-1: 9.1.2 Support to the coordinated provision of observation data SPA.2012.1.3-03 GMES Climate Change - Data archiving and exchange, 23rd November 2011
2. Geographic information - Data quality (ISO/DIS 19157:2011), Draft, July 2011
3. Metafor Common Information Model, <http://metaforclimate.eu/>
4. Geographic information – Metadata, ISO 19115:2003
5. INSPIRE, Data Specification <http://inspire.ec.europa.eu/index.cfm/pageid/2/list/7>

### 3 Introduction

Recommendations for changes to Metafor CIM [1] to incorporate ISO 19157 [2] form a deliverable for WP400 of CHARMe. These recommendations, which are primarily concerned with data models and concepts rather than specific software implementations, are represented in an implementation-neutral form.

This study of Metafor CIM in relation to ISO 19157 describes:

- An assessment of current CIM model and representation of quality info
- An assessment of ISO19157
- An identification of ISO19157 concepts that might improve CIM
- The impact of CIM to CHARMe
- The Recommendations: a summary of what action, if any, should be taken

### 4 Assessment of current CIM model and representation of quality info

The European Commission's Framework 7 project Metafor [3] defined a “quality” package as a set of specialisations around the ISO19115 –Data quality information package [4].

Data quality metadata can ensure the applicability of the data. This includes:

- the quality of the method to derive the data (design)
- the conformance to this specified method
- the user (customer) satisfaction
- the accomplishment of the needs of the application.

Figure 1 shows the main concepts defined within ISO19115 –Data quality information package that are required for describing the data quality - i.e., DQ\_DataQuality, DQ-Element, LI\_Lineage, DQ\_Scope, DQ\_Result. Further subclasses and specialisations, introduced within ISO 19115, of these main concepts are shown in Figure 2. A detailed description of this data quality modelling is provided by [3].

The “Common Information Model” (CIM) is a formal model of the climate modelling process. It includes descriptions of the experiments being undertaken, the simulations being run in support of these experiments, the software models and tools being used to implement the simulations and the data generated by the software. The “Common Information Model” (CIM) quality package, as shown in Figure 3, uses and specialises the ISO 19115 classes to compose its own data quality classes. Therefore, CIM adapts the concepts of ISO 19115 and inherits its advantages and disadvantages.

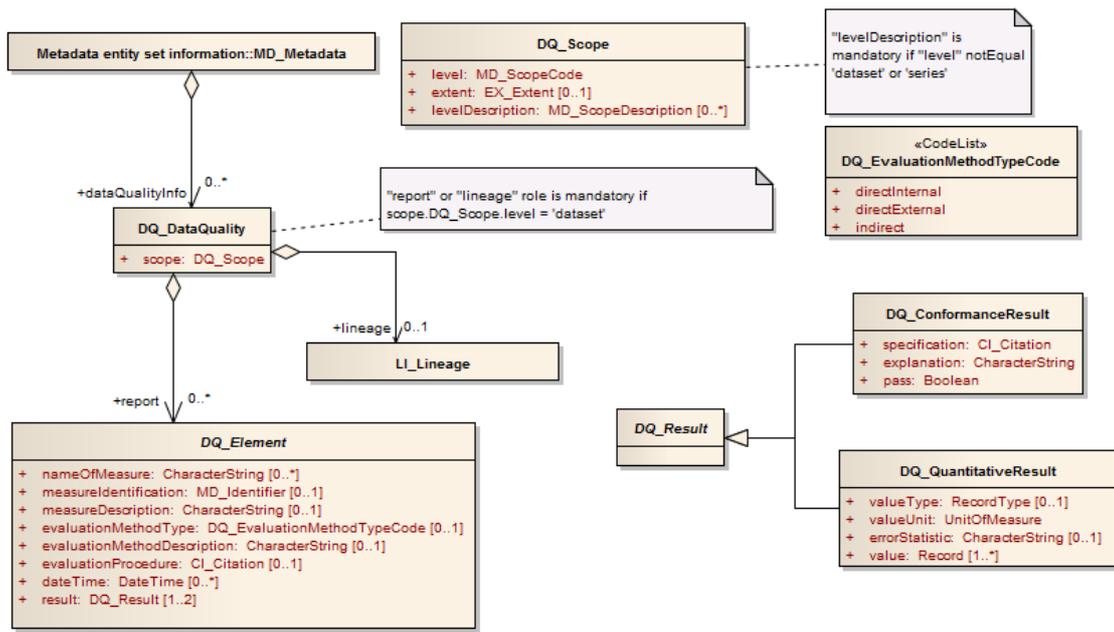


Figure 1 - ISO 19115 Data Quality Information

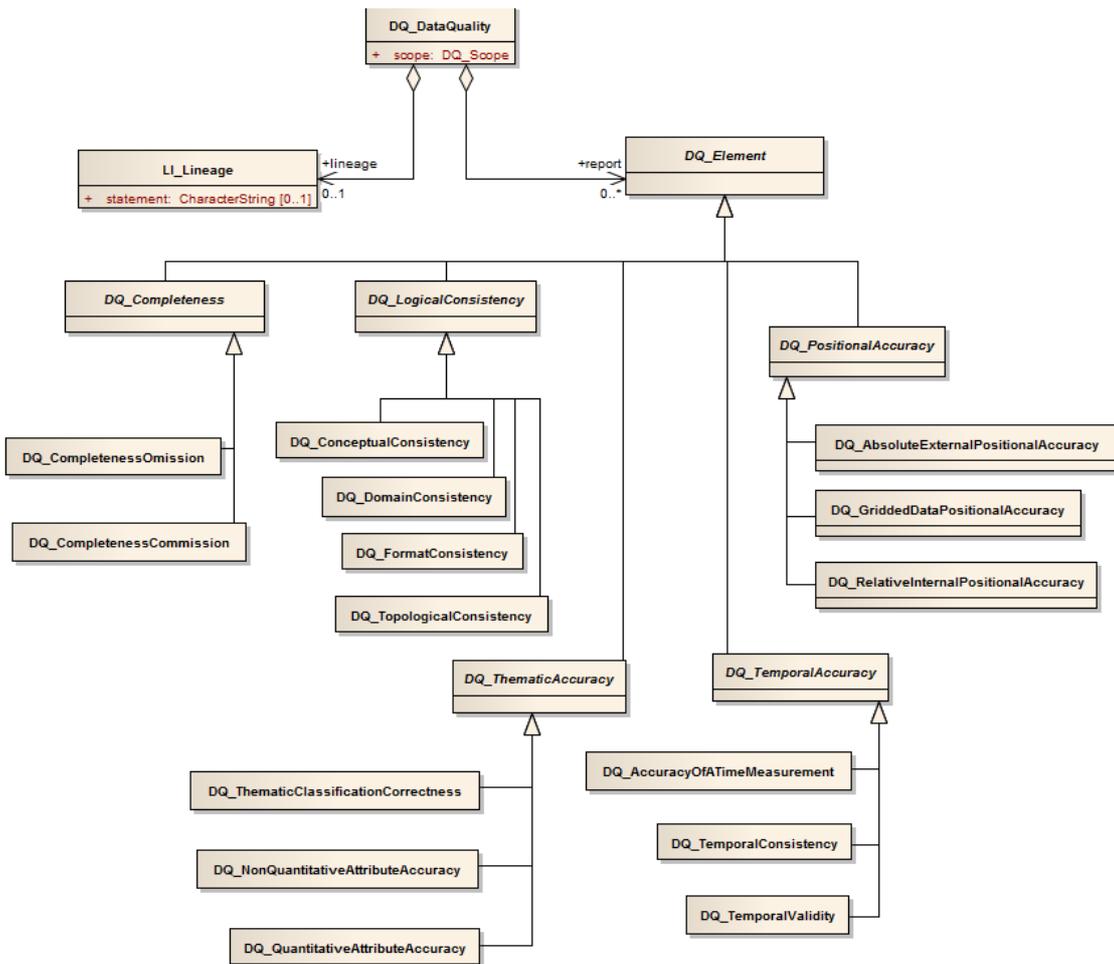


Figure 2 - ISO 19115 Data Quality Subclasses

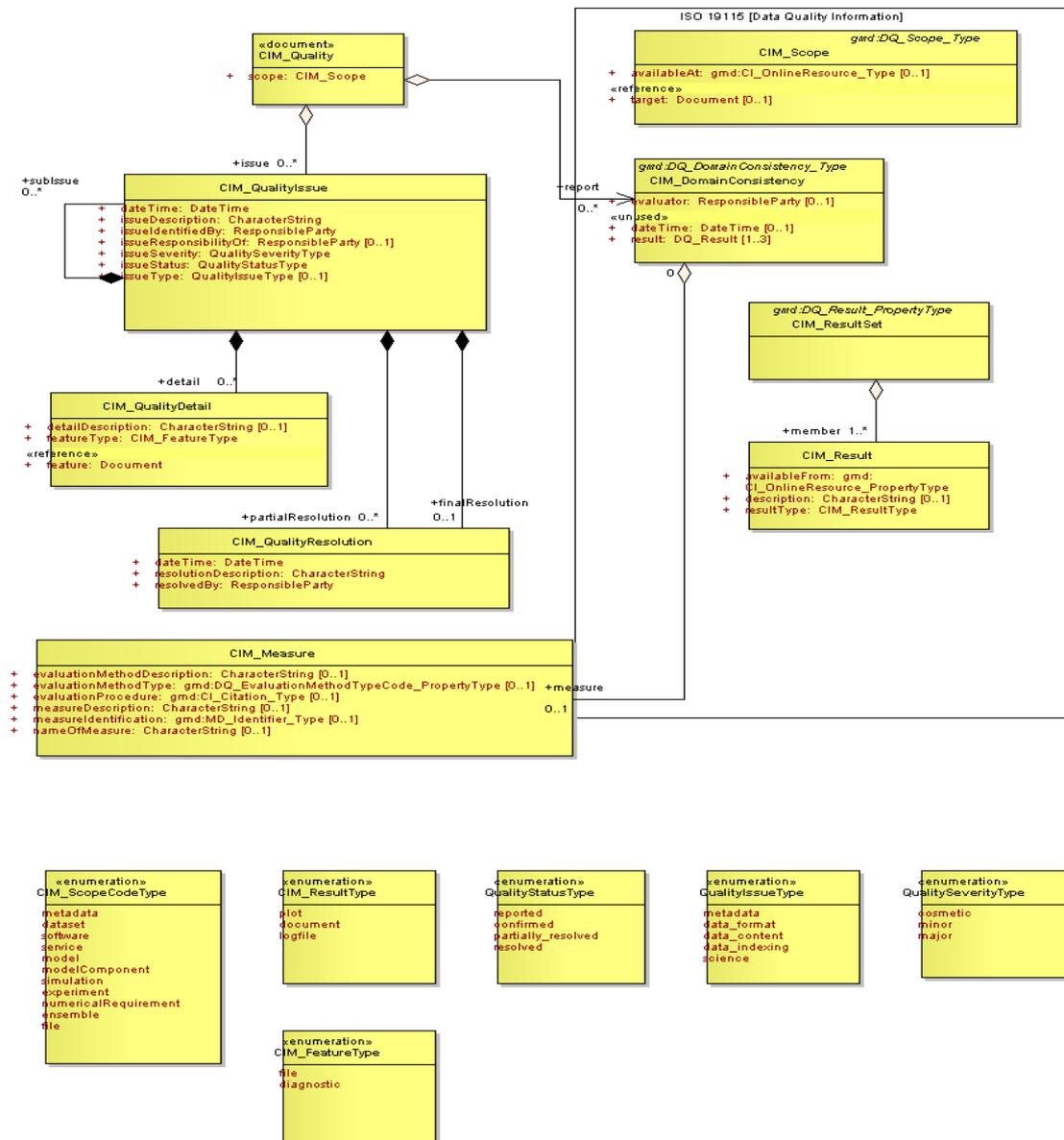


Figure 3 - CIM Quality Modelling

## 5 Assessment of ISO 19157

ISO 19157 establishes the principles for describing the quality for geographic data and concepts for handling quality information for geographic data. It provides a consistent and standards-based manner to determine and report the quality information of a dataset. It aims also to provide guidelines for evaluation procedures of quantitative quality information for geographic data (see Figure 4).

ISO 19157 harmonises the standards ISO 19113 – Quality principles, ISO 19114 – Quality evaluation procedures, ISO 19138- Data quality measures and ISO 19115-Metadata in order to establish a common set of metadata terminology and definitions regarding quality information (see Figure 5). In particular, it extends DQ\_DataQuality section of 19115. For

example, DQ\_Element in ISO 19157 is an aggregation of the classes DQ\_MeasureReference, DQ\_EvaluationMethod and DQ\_Result. This means, that an evaluation of a data quality element can be described by the measure, i.e. the type of evaluation, (DQ\_MeasureReference), the evaluation method (DQ\_EvaluationMethod) and the mandatory output of the evaluation (DQ\_Result). The class DQ\_StandaloneReportInformation enables connections between metadata and standalone reports, whereas the class DQ\_DescriptiveResult provides quality results described in text rather than numbers.

ISO 19157 recognizes that quantitative data quality elements may have associated quality which is termed *metaquality*. Metaquality describes the quality of the data quality results in terms of defined characteristics.

The standard is applicable to:

- Data producers providing quality information to describe and assess how well a dataset conforms to its product specification and
- To data users attempting to determine whether or not specific geographic data is of sufficient quality for their particular application.

It has to be noted that this standard does not attempt to define a minimum acceptable level of quality for geographic data.

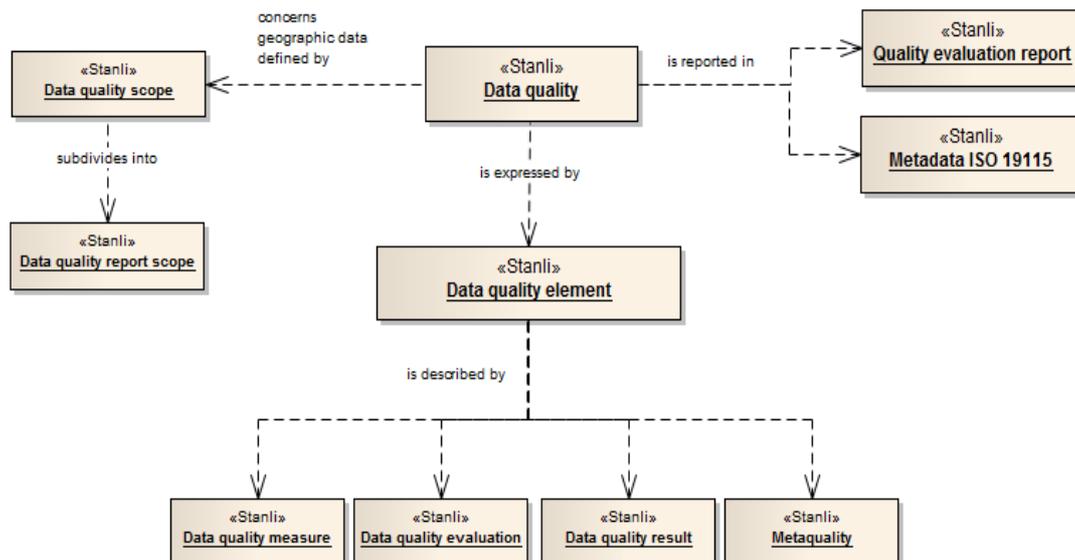


Figure 4 - ISO 19157 conceptual model on quality for geographic data

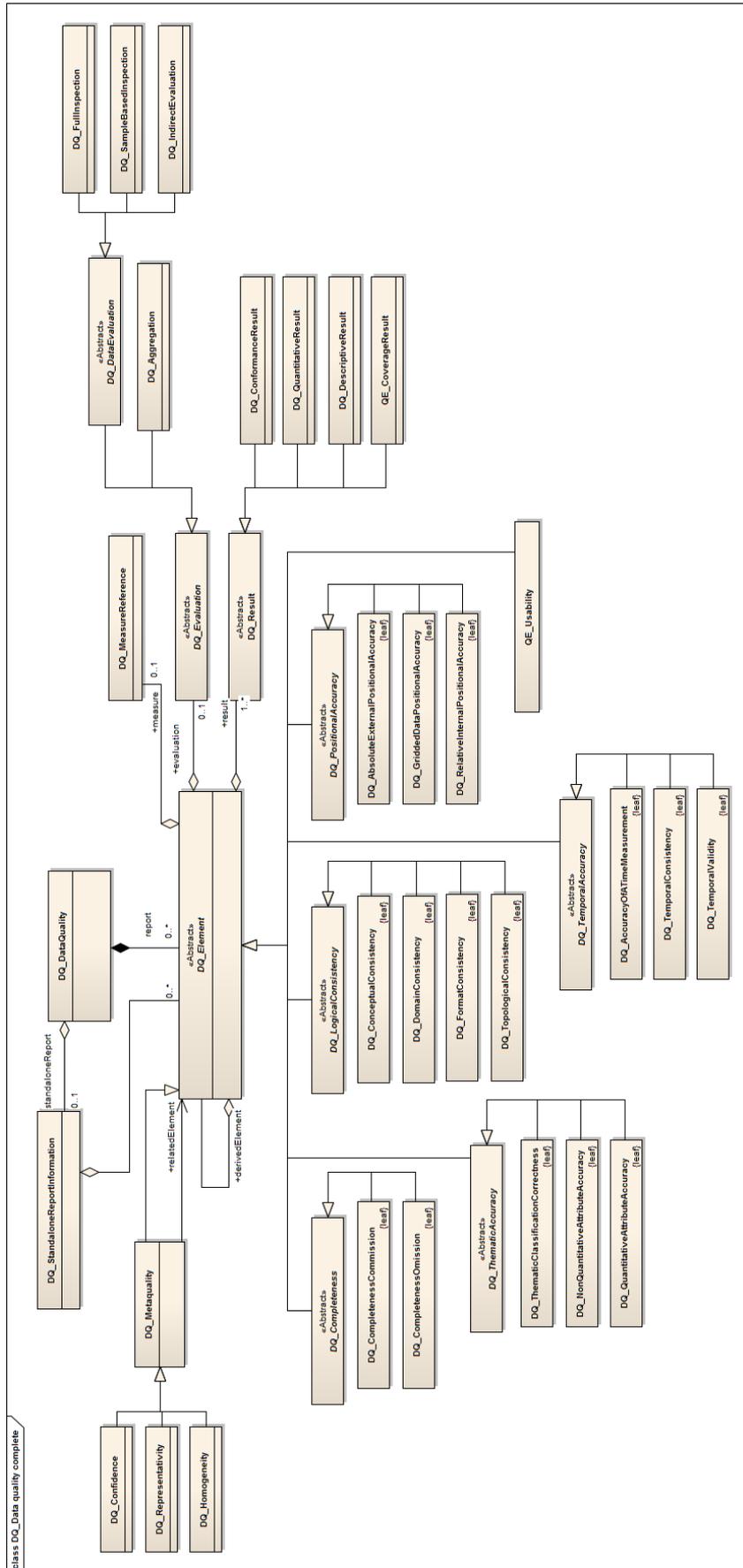


Figure 5 - ISO 19157 overview of the components of data quality

## 6 Identification of ISO 19157 concepts that might improve CIM

ISO 19157 introduces new data quality elements mainly by extending the elements already defined within ISO19115. Also, it establishes the principles for describing the quality of geographic data and concepts more vigorously compared to ISO 19115.

Therefore, CIM can be easily extended and benefit by the harmonisation and the new features provided by the ISO 19157. For example, the DQ\_Result (see Figures 1, 5) has been extended to include a) text descriptions (DQ\_DescriptiveResult) and b) data quality evaluation organized as a coverage (QE\_CoverageResult). Numerical model outputs usually go through thorough and systematic evaluations and quality assessments which can be easily described by the aggregation DQ\_Element of ISO 19157.

ISO 19157 is a very rich model which can conceptually cover a broad area of data quality information. However, there are two other main principles which also have to be considered before ISO 19157 implementation.

- a) The data quality requirements of the specific application
- b) The complexity of the designed system.

The right balance has to be found between a simple and a powerful solution. The former might be easy to be implemented but fail to meet requirements. On the other hand, the latter might provide substantial benefits available to only a few cases (users) and lead to a more complex and consequently expensive system.

## 7 The impact of CIM to CHARMe

The core requirement of the CHARMe project is to ensure that climate data is usable to a wide interdisciplinary user community, by providing users with access to the information they require to judge whether a climate dataset is fit for their purpose. Quality information associated with the climate data is part of their metadata and therefore, there is an overlap between the CHARMe and CIM projects despite their different priorities.

CIM can provide the climate data quality information but the challenge is to identify the complexity of the quality information required by a CHARMe user. The gathering of CHARMe requirements showed that there was not a strong requirement for quality information. If there was a requirement this was mainly for qualitative quality information.

Similarly, during the INSPIRE data specification campaign, the majority of Thematic Groups – bearing in mind the balance between a simple and a powerful solution - required qualitative instead of quantitative quality information [5].

## 8 Conclusions

The quality information associated with the climate data is the overlapping area between the CHARMe and CIM projects despite their different priorities.

CIM quality package is rooted on the ISO 19115 data quality information package. The new data quality information standard ISO 19157 enhances the functionality of ISO 19115 by mainly extending elements defined within ISO 19115. Therefore, from a pure technical perspective, CIM can be easily extended and benefit by the harmonisation and the new features provided by the ISO 19157.

However, from a practical point of view regarding a system development, an optimum balance has to be found between a simple and a powerful solution. The two principles which have to be considered are:

- a) The data quality requirements of CHARMe
- b) The complexity of the designed system.

To meet the project goal of enabling users to judge fitness for purpose for a given dataset, a number of factors need to be considered including but not only quality-related aspects; for example, accessibility and provenance. The first priority for CHARMe is to provide the ability to bring together and link these pieces of often-disparate information together and enable a user-driven process to support this. CHARMe has been building an extensible data model based on Open Annotation. This has a baseline with the ability to extend or profile to suit specific use cases. This would allow the addition of classes from ISO19157 expressed as annotation targets or bodies. Application of ISO 19157 to CHARMe would provide potential benefit in terms of increased interoperability with other systems and/or the application of classes to express particular quality concepts. The former is dependent on the more widespread adoption of what is still a relatively new standard. The latter is more likely to originate from domain-specific use cases rather application to the model globally. Consequently the best approach for CHARMe is to ensure that the model is open to inclusion of ISO 19157 classes where use cases dictate but not explicitly incorporating it in the baseline model.