RESEARCH ARTICLE

A method for checking the quality of geographic metadata based on ISO 19157

Manuel A. Ureña-Cámara[a], Javier Nogueras-Iso[b], Javier Lacasta[b], F. Javier Ariza-López[a]
[a]Grupo de Investigación en Ingeniería Cartográfica (GiiC), Universidad de Jaen, Jaen, Spain;
[b]Aragón Institute of Engineering Research (I3A), Universidad de Zaragoza, Zaragoza, Spain

ARTICLE HISTORY
Compiled August 21, 2018

ABSTRACT

With recent advances in remote sensing, location-based services and other related technologies, the production of geospatial information has exponentially increased in the last decades. Furthermore, to facilitate discovery and efficient access to such information, spatial data infrastructures were promoted and standardized, with a consideration that metadata is essential to describing data and services. Standardization bodies such as the International Organization for Standardization have defined well-known metadata models such as ISO 19115. However, current metadata assets exhibit heterogeneous quality levels because they are created by different producers with different perspectives. To address quality-related concerns, several initiatives attempted to define a common framework and test the suitability of metadata through automatic controls. Nevertheless, these controls are focused on interoperability by testing the format of metadata and a set of controlled elements. In this paper, we propose a methodology of testing the quality of metadata by considering aspects other than interoperability. The proposal adapts ISO 19157 to the metadata case and has been applied to a corpus of the Spanish Spatial Data Infrastructure. The results demonstrate that our quality check helps determine different types of errors for all metadata elements and can be almost completely automated to enhance the significance of metadata.

KEYWORDS
Geographic metadata; ISO 19115; Quality; ISO 19157; Spatial Data Infrastructures

1. Introduction

To facilitate the discovery and monitoring of spatial resources, various spatial data infrastructure (SDI) initiatives have been developed over the last few years. These initiatives highlight the definition of specific standards for geographic metadata like ISO 19115 (ISO 2014) and encourage the application of such initiatives to annotate the resources produced and published through various SDIs (Nogueras-Iso *et al.* 2005). In addition, high-quality metadata is needed to assure the effective discovery and monitoring of resources. According to ISO 8402 (ISO 1994), “quality” is a measure of

---

CONTACT Manuel A. Ureña-Cámara. Email: maurena@ujaen.es
CONTACT Javier Nogueras-Iso. Email: jnog@unizar.es
CONTACT Javier Lacasta. Email: jlacasta@unizar.es
CONTACT F. Javier Ariza-López. Email: fjariza@ujaen.es

excellence or a state characterized by the nonexistence of defects, deficiencies, and significant variations. In the context of metadata, rather than defining quality in terms of the nonexistence of defects, a more appropriate definition would be to consider a record of good quality as "a record that facilitates the process of identifying, describing, managing and searching data" (Ma *et al.* 2009). However, the fact is that current metadata assets exhibit heterogeneous levels of quality because they are created by different producers with different perspectives on the purpose of such metadata (Lacasta *et al.* 2014).

To address concerns related to heterogenous quality, metadata standardization efforts usually include abstract test suites to assure the feasibility of syntactic and semantic interoperability between systems. On the one hand, XML encoding specifications are provided to avoid syntactic heterogeneity, i.e. harmonizing the exchange format and structure of metadata. On the other hand, semantic heterogeneity issues are minimized based on the definition of metadata models that establish specific data domains for each metadata element and even the use of controlled vocabularies for many metadata elements. Moreover, interoperability is usually automated using tools based on XML Schema validation, which checks the quality of metadata in terms of completeness (commission or omission of metadata elements) and consistency (metadata format and structure/domain of metadata elements). Nevertheless, there has been little focus to the accuracy of metadata, which refers to the accurate description of resources using factual and correct information.

The purpose of this study is to propose a method for analyzing the quality of metadata by considering aspects other than interoperability issues. The proposal is an adaptation of the ISO 19157 standard for geographic information quality (ISO 2013) to the metadata case. In addition to completeness and consistency, accuracy and correctness of temporal, positional, and attribute information are exhaustively reviewed in this quality standard. It is not the objective of this paper to indicate the procedures for addressing quality issues that are detected after a quality assessment, but the objective is to propose a method of only assessing the quality of published metadata such that, if producers and users consider the method appropriate, it could be integrated within the design of a general production process. In addition, it must be noted that we propose a quality check and not a quality estimation. This distinction is important because statistical tools for estimation and quality control are different. Estimation refers to the determination of the true value of a parameter (e.g. mean, deviation, proportion, etc.) of a population with limited uncertainty. The value of the parameter is determined using a sample whose size is derived based on sampling laws. The outcome is the estimated value of the parameter with limited uncertainty in estimation. Quality checks are based on controls. A control refers to the determination of whether a parameter (e.g. mean, deviation, proportion, etc.) of a product satisfies a specific requirement expressed as a quality level (e.g. no more than 5% of errors), generally whether a parameter is effective for a specified use. We adopt the perspective of a quality check because a quality control requires a smaller sample size than a quality estimation, which facilitates the execution of non-automatable controls.

The remainder of this paper is structured as follows. Section 2 introduces related state of the art. Section 3 describes our proposed method in the analysis of the quality of metadata. Section 4 demonstrates the feasibility of this method by applying it to the ISO 19115 metadata holdings of the Spanish Spatial Data Infrastructure (IDEE). Section 5 then provides a discussion on the obtained results. Finally, this paper ends with some conclusions and an outlook of future work.

2. State of the art

For decades, professionals of the Digital Libraries (DL) community have been working on various proposals for metadata standards to describe a wide range of resources, both in physical and digital form. These professionals are also aware of the significance of the quality of metadata. According to Ma *et al.* (2009), metadata quality is directly related to the quality of services provided by DL. Based on this idea, Park (2009) proposed a means of assessing metadata quality based on the "core bibliographic functions of discovery, use, provenance, currency, authenticity, and administration." According to this functional perspective, Park (2009) and other similar studies (Moen *et al.* 1998, Weagley *et al.* 2010) identified completeness, accuracy, and consistency as the most commonly used criteria for evaluating metadata quality:

- Completeness refers to the ability of describing objects as completely as economically feasible (Bruce and Hillmann 2004). However, as metadata standards exhibit considerable heterogeneity concerning their models and the number of elements, completeness is usually measured in terms of the completion of a selection of elements that seem to be the most relevant for discovery and use. For instance, (Hughes 2004) proposed an element absence penalty only focused on the analysis of the five elements deemed essential.
- Accuracy refers to the accurate description of resources using factual and correct information. Some authors like Bruce and Hillmann (2004) included the detection of typographical errors in free-text metadata elements as an accuracy measure.
- Several works (Shreeves *et al.* 2005, Park 2009) agree in considering two facets of consistency: semantics and structure. Semantic consistency refers to the use of common vocabularies or values to refer to similar concepts when describing different resources. Structural consistency involves using a common representation and format for similar metadata elements.

In addition, it must be acknowledged that there are several studies in the DL domain that provide quality metrics that can be calculated mathematically and applied to different metadata schemas. For instance, Ochoa and Duval (2006) proposed quality metrics for completeness, information content, and readability of metadata records in DL of Learning Objects. Shen *et al.* (2013) also proposed mathematical formulae for accuracy, completeness and conformance (a dimension that is partially equivalent to the idea of domain consistency of other authors) of DL metadata.

Shifting focus to the domain of geographic information domain, there has not been considerable research on metadata quality as there has been in a general DL context. However, there are several studies that highlight the importance of issues related to completeness, consistency and accuracy.

In consideration of the importance of completeness, Tolosana-Calasanz *et al.* (2006) investigated the criteria with more influence on the subjective quality of ISO 19115 metadata records that were previously annotated by a group of experts. Using Principal Component Analysis on a list of criteria recommended by experts, it was concluded that the most relevant criteria for global quality include: the percentage of completed elements related to the subset identified as core metadata for geographic datasets; the number of filled elements; and the existence of information on the distribution format, spatial reference system, dataset responsible party, lineage, and the metadata creator. Although this study provides interesting insights, it must be noted that the study was conducted in 2006. During that period, SDI recommendations concerning metadata were not widely adopted, and experts were probably more interested in the number of completed elements than the quality of the content.

In the context of consistency issues, the work of Renteria-Agualimpia *et al.* (2015) emphasizes the importance of detecting spatial inconsistencies in geographic metadata, i.e. verifying for the consistency of direct (coordinates) and indirect (place names) spatial references present in the metadata. This work proposed a semi-automatic approach based on the combination of gazetteers for geocoding (direct and reverse geocoding) and spatial ranking techniques for detecting such inconsistencies.

In the context of accuracy, Olfat *et al.* (2012) stressed that existing platforms for metadata generation are unable to maintain metadata updated with dataset changes, and so the metadata is usually incomplete, out-of-date, and, in some cases, missing. This study proposes a framework for jointly managing data and metadata. Such a framework synchronizes and automates as much as possible the creation, update, and enrichment of metadata.

SDI initiatives also consider the quality of metadata. For instance, the European INSPIRE directive that promotes the creation of spatial information infrastructure in Europe formulated metadata regulation (European Commission 2008) to define a common agreement on the metadata required to describe a spatial resource. In addition to this metadata regulation there are technical guidelines (INSPIRE MIG 2017) for implementing the INSPIRE metadata regulation according to ISO/TS 19139 (ISO 2007) (the specification for the XML encoding of ISO 19115) which include a list of requirements and recommendations. Although the requirements can be considered a procedure for verifying the completeness and domain consistency of the core elements proposed by INSPIRE, the recommendations go a step further in verifying the semantic aspects of metadata because they “suggest additional, non-mandatory methods for enhancing interoperability, harmonising the provided metadata, and proposing effective defaults for the content of metadata, where several options for expression are permitted.” Similarly, the Australia and New Zealand Land Information Council (ANZLIC) and the U.S. Federal Geographic Data Committee (FGDC) have published comparable technical guidelines for metadata. In the case of ANZLIC, the “AS/NZS ISO 19115.1:2015 Guidelines”[1] provide comprehensive examples of the ways of implementing some key elements of the standard. In the case of FGDC, the “Metadata Recommendations Supporting Data Discovery and Use in Data.gov and GeoPlatform.gov”[2] describe the ways of filling some elements that are particularly relevant to the enhancement of resource sharing.

Although the literature on geographic metadata and SDI recommendations enhances awareness on the provisioning of high-quality metadata, there is no general framework for evaluating the quality of geographic metadata. Therefore, it would be sensible to adapt the quality elements already studied in a more general context of DL metadata to the peculiarities of geographic metadata. Among other features, geographic metadata focuses more on the definition of the geographic coverage of resources, identification of the timeliness of geographic information, and summarization of the content of attribute information that is hard-coded in complex geographic information formats. In addition, from a broader perspective, these features are common to geographic information. Therefore, this study proposes an analysis of geographic metadata quality in the same way as geographic information quality. In other words, we propose a redefinition of the quality dimensions of ISO 19157 standard for geo-

---

[1] http://anzlic.gov.au/sites/default/files/files/2015 Metadata Profile Guidelines Standard.pdf

[2] https://cms.geoplatform.gov/sites/default/files/document_library/GeospatialMetadataBestPractices_20141010.pdf

graphic information quality (ISO 2013). ISO 19157 proposes a general data quality description framework for spatial data that can be adapted to describe the quality of geographic metadata.

3. Method

The method proposed in this study is based on the idea of analyzing a wide range of metadata aspects according to the quality elements proposed by ISO 19157 because this standard includes some of the quality criteria for metadata of DL including completeness, accuracy, and consistency and extends such criteria to specific geographic information aspects. We first adapt the quality descriptors in ISO 19157 to metadata elements. We then describe the point of view of quality control applied to metadata. Finally, we propose a means of performing quality assessment using several methods applied to each quality element.

3.1. *Quality description (ISO 19157 data quality elements)*

Our proposal for reinterpretating the quality elements of ISO 19157 adapted to the context of quality evaluation over a collection of metadata records is as follows:[3]

- **DQ_Completeness:** presence or absence of metadata elements.
    - **DQ_CompletenessCommission:** checking for excesses of metadata elements in the metadata record.
    - **DQ_CompletenessOmission:** checking for mandatory elements that are absent in metadata records.
- **DQ_LogicalConsistency:** determining the degree of adherence to logical rules of metadata structure.
    - **DQ_ConceptualConsistency:** determining the adherence to the metadata schema or to other explicit or non-explicit requirements of a data model or a dataset.
    - **DQ_DomainConsistency:** determining the adherence of metadata element values to the value domains.
    - **DQ_FormatConsistency:** verifying the use of the appropriate format for metadata encoding.
    - **DQ_TopologicalConsistency:** determining the correctness of the way in which constituent parts are interrelated or arranged in terms of metadata.
- **DQ_TemporalAccuracy:** determining the accuracy of the temporal metadata elements and temporal relationships of metadata elements.
    - **DQ_TemporalConsistency:** determining the correctness of ordered events or sequences, if reported within a single metadata record.
    - **DQ_TemporalValidity:** determining the validity of metadata values with respect to a defined time domain.
- **DQ_ThematicAccuracy:** determining the correctness of non-quantitative metadata elements and metadata elements related to classifications.
    - **DQ_ThematicClassificationCorrectness:** determining the correctness of the thematic keywords and categories included in the metadata with respect to a universe of discourse.

---

[3] Some ISO 19157 elements are missing in this adaptation because they cannot be directly applied to metadata.

- **DQ_NonQuantitativeAttributeCorrectness:** determining the correctness of non-quantitative metadata elements in relation to a universe of discourse.

Because geographic metadata often refers to positions and are not spatial data themselves, an element focused on correction rather than accuracy is required. Therefore, we propose a new quality element for positions known as **DQ_PositionalCorrectness**. This element allows the use of a Boolean value for indicating the correctness of positions.

On the other hand, it is also important to observe that ISO 19157 has no adequate measures for free-text quality description. This standard is intended for data and not for descriptive and prosaic texts. However, metadata records include important free-text fields (e.g. abstract, purpose, lineage) that must be evaluated. For this reason and because ISO 19157 has a fixed set of elements, a new quality element has been proposed as suggested by ISO 19113. This new element is called **DQ_QualityOfFreeText** and evaluates the quality of information given within a free-text field (completeness, correctness, appropriateness, readability, etc.).

3.2. *Quality Control*

Another fundamental issue addressed by our method is the means of performing controls. As stated in the introduction, we adopt a control perspective for analyzing the quality of metadata in this study. In data quality control, we can distinguish two very different situations: first, when automation of the control process is possible; and second, when automation is not possible (ISO 19157). For the first case (automatable issues), the complete population can be checked (full inspection) for the type of risks that exist (type I and II errors). For the second case (non-automatable issues), a sample-based control is used to derive a decision involving limited risks.

Controls are performed by means of a sample (population or a portion of the population) and some rules to determine whether the quality of the product satisfies specifications, i.e. a statistical test. Regardless of sample size, the type I error (significance or producer risk) can be assured. In statistical testing, the larger the sample size, the greater is the power of the test (type II error).

Therefore, there are some metadata elements that can be automatically checked while other elements must be checked by a sample. Whatever the case, the first thing to do is to properly determine the population of interest, the size of this population, and the possibility of an automatic control. For this purpose, the first step for some elements is to determine a set of unique values within the metadata set (the population). Following this, the automatic or sample-based control will be performed. For other elements, the identification of unique cases is not possible.

In particular, in this work we initially consider 5% as the established-quality-level requirement (JGSI 2002). This 5% represents the worst or poorest level of quality that would be considered acceptable as a process average. This idea of quality corresponds to the term Acceptance Quality Limit (AQL). Therefore, AQL is equal to 5% in this case. This parameter (AQL) is the key element of the international standards ISO 2859 and ISO 3951.

However, we require certain rules in order to determine whether the quality of the product satisfies the established AQL. These rules are based on "ISO 2859 - Part 2: Sampling plans indexed by limiting quality (LQ) for isolated lot inspection." ISO 2859 series is widely used throughout the world to control attributes, e.g. for controlling

Table 1. Excerpt of “Table A - Single sampling plans indexed by limiting quality (LQ) (procedure A)” from ISO 2859-2 (ISO 1985).

Lot size		Limiting quality in percent (LQ)									
		0.5	0.8	1.25	2.0	3.15	5.0	8.0	12.5	20	32
16 to 25	n	→	→	→	→	→	25	17	13	9	6
	Ac						0	0	0	0	0
26 to 50	n	→	→	→	50	50	28	22	15	10	6
	Ac				0	0	0	0	0	0	0
51 to 90	n	→	→	90	50	44	34	24	16	10	8
	Ac			0	0	0	0	0	0	0	0
...											
3201 to 10000	n	450	315	315	200	200	200	125	80	80	80
	Ac	0	0	1	1	3	5	5	5	10	18
...											

the Land Parcel Identification System in Europe (Milenov *et al.* 2010), for geological data in China (Xie *et al.* 2008) and for diverse spatial data quality controls in New Zealand by the National Topographic Hydrographic Authority (NTHA 2004).

Because we consider all metadata records as a whole (a unique set or lot), ISO 2859-2 must be applied. In this case, ISO 2859-2 provides the rules for quality control in a very simple way using “Table A” (see excerpt in Table 1). To use “Table A” of ISO 2859-2, we need the size of the lot under control and a limiting quality (LQ) index that is thrice the AQL ($LQ \approx 3 \times 5\% = 15\%$). “Table A” of the international standard outputs the sample size to be randomly considered and the maximum number of errors to be determined in this sample to ensure a 5% producers risk (type I error) and a 10% consumer risk (type II error). It is important to notice that Table 1 does not include LQ=15%. For this reason, we assume LQ=12.5%, which implies that AQL is equal to 4.16%. This is the final value considered in our quality control.

3.3. *Description of the proposed quality elements and measures*

The following subsections describe specific methods proposed to control the different quality elements proposed by ISO 19157 already introduced in section 3.1. In addition, it must be noted that these methods are applied to ISO 19115 / ISO 19119 metadata encoded in XML according to the ISO/TS 19139 specification. We mention the ISO 19119 standard (ISO 2016a) because prior to the approval of the revision of ISO 19115 in 2014 (first part of ISO 19115 or ISO 19115:1), the metadata extensions for describing services were defined in ISO 19119. Moreover, ISO/TS 19139 has been superseded by the third part of ISO 19115 (ISO 2016b). However, the majority of geographic metadata and cataloging systems are still working with the metadata models approved for ISO 19115 in 2003, ISO 19119 in 2005 and the XML encoding of ISO/TS 19139 in 2007. However, our proposal is also valid for ISO 19115:1 and ISO 19115:3 because the metadata elements selected for evaluating the quality elements in the following subsections are included or have a direct equivalent in the new versions.

Unless inversely expressed in each measure, the threshold value for the quality level of each measure is set to 4.16% for full inspection, as indicated in section 3.2. Moreover, if the measure is defined as automatic, the set of metadata records is the full dataset. For manual measures, a randomly selected set of records is proposed according to the number of records that must be achieved with an AQL of 4.16%.

Finally, it must be noted that the proposed measures do not impose access to the data described by geographic metadata. In most cases, metadata collections provided through cataloging systems are not directly linked to the associated datasets. Only in the case of Web Feature Services metadata (a small percentage of metadata records), the implementation of measures could access the real data and verify content on the fly.

3.3.1. DQ_Completeness

To test both completeness elements described in section 3.1, we used the XML Schemas for ISO 19115/19119 according to ISO 19139 specification.[4]

Both measures focus on ensuring that the cardinality (number of occurrences) of metadata elements follows the specification. The measures of completeness are proposed as automatically determined using the full dataset as a ratio. We consider a valid dataset if less than 4.16% of the records are wrong.

3.3.1.1. DQ_CompletenessComission. We propose an XML Schema validation that checks for metadata records with metadata elements having multiple instances, which otherwise should be unique.

The measure proposed to control the quality of the dataset is the rate of metadata records that have additional instances in one of their metadata elements. This measure corresponds to the "rate of excess items" measure described in table D.3 of ISO 19157. As explained before, the threshold value is set at 4.16%, above which the dataset does not satisfy this quality element.

3.3.1.2. DQ_CompletenessOmission. Again, an XML Schema validation checks for metadata records with missing mandatory elements.

The measure to assess quality is the rate of metadata records that do not have the mandatory elements filled correctly, which corresponds to the "rate of missing items" measure described in table D.7 of ISO 19157.

3.3.2. DQ_LogicalConsistency

Except for topological consistency, the measures to control logical consistency are based on XML Schema validation as for completeness, but in this case, attention is focused on encoding and domain values. All measures, except those related to topological consistency, are determined using the full dataset.

3.3.2.1. DQ_ConceptualConsistency. The first proposed method for conceptual consistency is the XML Schema validation that checks for metadata records that do not strictly follow the ISO 19115/ISO 19119 metadata schema and its ISO 19139 encoding on XML, e.g. adding an element not included in the metadata schema. The proposed measure is the rate of metadata records that follow ISO 19115/ISO 19119 and ISO 19139. The corresponding measure in ISO 19157 is the "compliance rate with the conceptual schema," described in table D.13.

In addition to this measure, it must be noted that quality control is usually centered on explicit quality requirements, but implicit quality requirements are also very important. An important requirement of a dataset is the principle of non-contradiction between its contents, although this is usually an implicit requirement. We consider this issue as an aspect of conceptual consistency.

---

[4] XML Schemas are available at `http://schemas.opengis.net/csw/2.0.2/profiles/apiso/1.0.0/apiso.xsd`.

To check for non-contradictory relationships between elements on the same metadata record, we propose a measure called "contradictions." We propose to apply this measure to ensure that a control on metadata elements such as scale, number of geometric objects (for vector spatial representations) or number of dimensions (for raster spatial representations) is based on the same data. In this sense, we verify the nonexistence of contradictions between information items provided within the same metadata record. We propose a determination of the number of metadata records that have inconsistent information in metadata elements. The mathematical formulation of this measure is similar to the "number of faulty point-curve connections" measure described in table D.22 of ISO 19157. It is manually controlled from a randomly selected set of metadata records.

Following the indications in the previous paragraph, we verify if the scale or resolution has the same value defined in other metadata elements, with a special focus on title/abstract. Moreover, we check the number of dimensions/number of geometric objects to determine if the number of raster dimensions is equal or greater than 2 and to determine if the number of objects is filled and defined in other metadata elements. This measure of metadata accuracy is tested internally and for this reason, it is considered to be categorized under conceptual consistency. Trueness of values can be externally checked if there is access to the original dataset/service, but this is not usually the case.

3.3.2.2. DQ_DomainConsistency. XML Schema validation checks for the presence of metadata records with elements whose values do not contain valid literals according to ISO 19139 XML encoding. The proposed measure is the rate of metadata records that follows ISO 19139 domains, which corresponds to the "value domain conformance rate" measure described in table D.17 of ISO 19157.

A specific case of domain consistency is to verify the validity of bounding boxes defining the geographic extent of the described resources. We analyze if the *EX_GeographicBoundingBox* element exists and if it contains a valid bounding box. The validity of the bounding box is marked by the fact that the provided coordinates (latitude, longitude) are in a valid range ($\pm 90, \pm 180$) and they constitute a non-inverted rectangle in the space.

3.3.2.3. DQ_FormatConsistency. We verify whether metadata records have a valid XML encoding. The measure is the rate of metadata records that do not have a valid XML encoding, which corresponds to the "physical structure conflict rate" measure described in table D.21 of ISO 19157.

3.3.2.4. DQ_TopologicalConsistency. Topological consistency is verified in two senses. The first is the relationship between our metadata record and other metadata records of the dataset. The second is the relationship between different elements in the metadata record.

The first measure, which we call "external relationships consistency," is based on the case of metadata records including information on parent metadata file identifiers: we check if the records are correctly referenced. To test this, we propose a measure of determining the rate of metadata records that reference to a non-reachable parent file. This measure is considered similar to the measure described in table D.23 of ISO 19157, which is defined as the "rate of faulty point-curve connections," in the sense that the curve is the parent file and the points are the children metadata files.

The second measure, which we call "topological contradiction," is applied in cases where a metadata record contains both a bounding box to describe the general extent of the described resource and a bounding box to define the geographic extent of a part of this resource to which data quality information applies. In this case, we determine whether both bounding boxes are compatible. Compatibility must assure at least an overlapping between both extents. This method is related to the definition of topology but in the sense of the relationship between geospatial elements in metadata elements and not their relative precision. In this sense, the method is considered internal because it does not require querying an external database: we only verify the internal information within the metadata record. In addition, it can be considered relative in the sense of comparing different elements from the same metadata record. The proposed measure for this quality element is the ratio between the number of records having non-compatible bounding boxes and the total number of records having both bounding boxes.

3.3.3. DQ_TemporalQuality

In the context of temporal quality, the quality control for accuracy in the measurement of time is not checked. The reason is that datasets and services do not require any precision in determining the time expressed in different elements of the metadata record. In other words, date-based metadata elements only require expressing a year, the month of the year or the day. More accurate values are not required and therefore, the accuracy of temporal measurements does not need to be checked.

However, we believe that it is important (and it can be checked) to maintain the logical sequence of the process until a dataset/service is published and the validity of the dataset/source and its metadata is assured.

3.3.3.1. DQ_TemporalConsistency. The objective is to verify the proper time sequence of the described steps of metadata: creation, publication, revision, harvesting, etc. The sequence may not be complete because some of these timestamps are optional. In this case, the measure assumes that the sequence is correct.

The measure is applied to the full dataset verifying the rate of correct sorted time sequences. We adapt the "chronological order" measure described in table D.62 of ISO 19157, which provides a Boolean value that verifies whether a timestamp is correctly sorted. However, because the original measure is defined only for a unique timestamp, we adapt it to a rate between all correct time sequences and the full number of checked sequences.

3.3.3.2. DQ_TemporalValidity. This verifies if the temporal validity of any metadata record is constrained by the temporal domain that we define. In this sense, a metadata record has to be valid at the time of metadata harvesting or at the time of a user catalog search. Because the moment of a catalog search is unknown, this measure focuses on the validity of harvesting time according to the specified maintenance and update frequency of metadata.

Figure 1 presents the timeline of a metadata record that is published once the described dataset/service has been created and published. The temporal period since the metadata publication date until the next update (according to the metadata maintenance and update frequency element) establishes the maximum range of time during which a harvested metadata record is valid. Moreover, it must be noted that the meta-

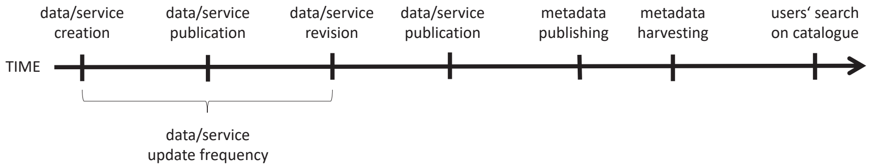

Figure 1. Timeline representing the typical development of a product/service until a user search is performed.

data maintenance and update frequency element is optional and may not be filled. In addition, the chosen value for this element may also be ambiguous (e.g., *irregular* or *as needed*). In these last two non-clear cases, we consider that the metadata record is a snapshot that is valid at the time of harvesting.

To test this quality element, we check if the metadata harvesting date falls within the temporal period defined between the latest reported timestamp and the addition of the maintenance and update frequency. The proposed measure is the rate of metadata records outside this allowed temporal range, which corresponds to the "value domain non-conformance rate" measure described in table D.18 of ISO 19157 and following the recommendations of section D.5.3 in the same standard.

3.3.4. DQ_ThematicAccuracy

Most of the measures proposed for controlling thematic accuracy are manually achieved, i.e. for each measure, an adequate set of records is selected according to ISO 2859-2.

3.3.4.1. DQ_ThematicClassification Correctness. The method proposed here consists of a manual evaluation of the correctness of a common vocabulary for the classification of resources of a sample randomly selected from the dataset (metadata corpus). For this purpose, we analyzed the use of the list of INSPIRE spatial themes included in the General Environmental Multilingual Thesaurus - GEMET.[5] Owing to the implementation of the European INSPIRE directive, this list of themes is commonly used as a controlled vocabulary for metadata descriptive keywords (Díaz-Corona *et al.* 2017). Using this list, the INSPIRE metadata regulation forces the use of one keyword, at least, to describe the spatial data theme referred by a dataset that is created and published in response to the implementation of the directive in one of the member states of the European Union.

To assess the correctness of the classifications, manual experts will review the title, abstract, and other free-text fields that provide clues on the correct assignment of the data theme. The sources can be considered both internal (i.e. comparison of keywords, title, abstract, etc.) and external as we use the list of INSPIRE themes.

The proposed measure is the "number of incorrectly classified features" that is described in table D.63 of ISO 19157, which can be interpreted as the number of metadata records that do not have a correct keyword.

3.3.4.2. DQ_Non-quantitativeAttributeCorrectness. There are many metadata elements that can be checked from the point of view of their correctness. The correctness of non-quantitative elements can be analyzed from different perspectives. Table 2 presents a selection of elements that are analyzed and the proposed approach

---

[5] `http://www.eionet.europa.eu/gemet`

for each element. In addition, it must be noted that in some cases the loss of correctness of a specific attribute value can be attributed to time decay, but this situation is not always clear. Both factors, time decay and other sources of errors, can act together.

Table 2. Summary of the measures proposed for non-quantitative attribute accuracy and a brief approach.

Element	**Approach**
E-Mail address	Automatic. Send e-mails and account the number of error responses of e-mail servers. Full inspection (100%).
Telephone number	Manual. Telephone call to the telephone number. Sampling inspection after identification of unique values. Control by ISO 2859-2.
Organization	Manual. Telephone call or e-mail and ask about the correctness of the organization name. Sampling inspection in the dataset. Control by ISO 2859-2.
Address	Manual. Telephone call or e-mail and ask about the correctness of the address. Sampling inspection in the dataset. Control by ISO 2859-2.
Responsible	Manual. Telephone call or e-mail and ask about the correctness of the responsible. Sampling inspection in the dataset. Control by ISO 2859-2.
Title	Manual. Telephone call or e-mail and ask about the correctness of the title. Sampling inspection in the dataset. Control by ISO 2859-2.

Some of the elements proposed in table 2 such as e-mail can be automatically tested and applied to the full dataset using the “rate of incorrect attribute values” measure described in table D.69 of ISO 19157. The remaining elements must be manually tested by directly asking the dataset/service responsible party via e-mail or telephone. In this last case, this control corresponds to the “number of incorrect attribute values” measure described in table D.68 of ISO 19157.

3.3.5. DQ_PositionalCorrectness

The metadata elements that describe the geographical area covered by the geospatial data are essential for the identification and use of the provided information. The most common ways of describing the area covered by a resource include a spatial bounding box or some kind of textual location reference. In ISO 19115, the spatial bounding box is contained in the *EX_GeographicBoundingBox* element of the *identificationInfo* section. The spatial textual locations can be found as keywords in the *MD_Keywords* element, but other elements such as *title*, *alternateTitle*, or *abstract* may contain textual locations in their free-text descriptions.

Our proposed method for this quality element analyzes all these metadata elements to determine if the provided spatial reference is correct. We attempt to identify if the provided bounding box covers the correct area. This is done by extracting the textual locations from the metadata and identifying if there is a match between them and the bounding box. If there is a correlation between the bounding box and the textual references, it is a guarantee that the bounding box is correctly placed. To extract the textual locations in the metadata, we use a geoparser that assumes the GeoNames[6] collection as a knowledge base of place names. Using GeoNames, the identified spatial references are converted into longitude/latitude coordinates and compared with the bounding box. Here, we considered that if some of the identified references are included inside the record bounding box, it is a good indicator of the position correctness of the bounding box (they both refer to the same place). The source of comparison is external because place names provide pointers to query external databases that allow us to compare coordinates in geographic extension metadata elements.

---

[6] `http://www.geonames.org/`

The measure will be the rate of bounding boxes that do not contain positions derived from place names within free-text elements, which is an adaptation of the "rate of positional errors above a given threshold" measure described in table D.33 of ISO 19157, with a distance threshold of 0 units because of the use of area in this context.

3.3.6. DQ_QualityOfFreeText

Free-text metadata elements represent a special case among metadata elements. These elements can be filled with any sequence of characters that accurately, or not, describe the issue indicated in the element description of ISO 19115 or the selected metadata profile. In this sense, there are two possible approaches to determining the quality of free text: (i) Determine the quality of a selection of values for each metadata element using a panel of experts and then apply the classification to the rest of the elements; (ii) Use some sort of analysis on the free-text elements that indicates the quality of the described text.

In the previous sense, this study uses both approaches to determine the quality of the selected elements. We first select a set of experts that would assign a quality value of each element according to the definition of these elements. On the other hand, we apply a readability index on the elements to test the difficulty of any reader in understanding the content.

We selected three elements to test the quality of free text with different kinds of obligations. The description of these elements (extracted from ISO 19115) is as follows:

(1) Abstract (*MD_identification/abstract*): It is a mandatory element that must be a non-empty brief narrative summary of the content of the described data set, data set series, or service to be provided. It is encoded with a Non-empty Free Text Element content in the language of metadata. Maximum occurrence number: 1.
(2) Purpose (*MD_identification/purpose*): It is an optional element. The element must describe the purpose of the creation of the resource. Maximum occurrence number: 1.
(3) Lineage Statement (*DQ_Quality/LI_Lineage/statement*): It is a conditional element that describes non-quantity information about the data. It describes the sources or events and methodologies used to obtain the dataset. Maximum occurrence number: 1 (for datasets) and 0 (for services).

The three selected elements have not been extensively described except in some technical recommendations like (INSPIRE MIG 2017), which indicates the following in its TG Recommendation C.4 for the abstract metadata element: "the resource abstract is a succinct description that can include: a brief summary with the most important details that summarize the data or service; coverage (linguistic transcriptions of the extent or location in addition to the bounding box); main attributes; data sources; legal references, and importance of the work. The most important details of the description should be summarized in the first sentence or the first 256 characters. Unexpanded acronyms should not be used."

For this new quality element, we propose the following two measures and their evaluation methods.

To check the general quality of free-text elements we propose a measure called "overall quality of free text." The measure uses a manual method based on a pool of experts. We established five levels of quality, and following the definition of the technical recommendations, we developed a set of suggestions for the experts to define what level of quality should be assigned. A set of examples has been prepared as a guide for a common evaluation criterion. In summary, a guide of examples for each quality level and their definition were agreed and provided to each expert. We have chosen a random sample of metadata records, and a set of seven experts to define a level of quality. The quality levels, sorted from the worst to the best, include the following:

- Not defined: The element is not filled or has been filled blank.
- Bad defined: The element does not provide information about the element.
- Regular defined: The element provides a general description.
- Good defined: The element describes an adequate description.
- Very good defined: The element provides a detailed description.

To develop a measure according to a quality control, we considered that very good, good, and regular levels pass the control. In contrast, a bad level does not pass the quality control. The non-defined case is a bit more complex: when the metadata element is optional, it is considered to pass the test; otherwise, when the metadata element is mandatory, the test is not passed.

In addition, we propose a second measure called “readability of free text”, which is an automated readability index based on Flesch definition (Flesch 1948). The computation has been modified to adapt it to the native language of metadata elements, Spanish, by Fernández-Huerta (1959). This index is defined by the following equation:

$$Index_{FH} = 206.835 - 1.015\frac{N(words)}{N(sentences)} - 60.0\frac{N(syllables)}{N(words)} \tag{1}$$

where $N(words)$ represents the number of words in the free-text element, $N(sentences)$ is the number of sentences in the element, and $N(syllables)$ is the number of syllables. $N(words)$ and $N(sentences)$ are computed directly. However, $N(syllables)$ is determined using the algorithm developed by Hernández-Figueroa *et al.* (2013). It is interesting to observe that the index cannot be determined using acronyms or other abbreviations as suggested in the aforementioned INSPIRE technical recommendation C.4.

The quality measure is determined as the ratio between the sum of all elements with an $Index_{FH}$ greater or equal to 50, plus the unknown elements if the field is optional or conditional (e.g. *purpose* and *lineage*), and the total number of metadata elements. The threshold of 50 in $Index_{FH}$ is selected because it represents the limit between difficult and very difficult to read text (Flesch 1948). The threshold value to pass the quality check is 4.16%, i.e. not more than 4.16% of metadata elements should be wrong.

4. Results

In this section we describe the results of applying the measures defined in the previous section to a corpus. First, we describe the corpus of metadata records that has been used, which contains the description of both datasets and services and has been developed by different types of producers.

Table 3. Completeness results.

Quality element	**#**	**fail**	**error rate**	**result**
DQ_CompletenessCommission	4824	9	0.19%	Pass
DQ_CompletenessOmission	4824	147	3.05%	Pass

4.1. *Corpus*

For the purpose of our experiments, we used the metadata records created under the umbrella of IDEE. In particular, we used the collection of metadata records contained at the central metadata catalog on 2nd September 2016 with 4,824 records: 3,640 records describing spatial datasets or series; and 1,184 records describing geographic services.[7] The contents of this central catalog are regularly updated with a harvesting process that retrieves the contents of the catalogs running at the different SDI initiatives that belong to either national governmental offices or regional governments. This dataset of metadata is the population we want to characterize.

4.2. *Results of experiments*

For automatic methods where full inspection is possible, we developed a Java program (any alternative programming language is also applicable) that takes advantage of XML parser libraries to validate and upload the XML files of the metadata corpus, check the problems raised by each quality element and related measure as indicated in section 3.3, and finally compute the rates proposed for these measures. For the manual methods, which require the interaction with experts and cannot be automatically implemented in the previous program, a random sample of the corpus has been selected by means of a random number generator and the sample size is determined according to the population and the AQL defined in section 3.2.

With respect to the layout of the tables that present the results in this subsection, it must be noted that in the case of the application of manual methods the tables contain the following columns: the population size (#), the size of the sample (column *sample*), the AQL, the number of items that fail the proposed test (*fail*), and the final result of the measure (*result*). In contrast, the tables that present the results of automatic methods contain the following columns: population size (#), the number of items that pass or fail the proposed test (*pass* or *fail*), the rate of items passing or not the quality control (*correct rate* or *error rate*), and the final result of the measure (*result*).

4.2.1. DQ_Completeness

Table 3 presents the results of the omission and commission errors of the automatic evaluation of metadata records for datasets/series and services by means of an XML Schema validator. It must be noted that although one metadata record may have several problems involving completeness commission or omission, it is only counted once as a failing record.

As can be observed in Table 3, the error rate is very low and does not surpass the threshold value of 4.16% in both cases. However, commission has a slightly lower rate perhaps because the wide use of metadata editors forces the uniqueness of some elements. In addition, it must be noted that metadata records describing services have no errors. This is maybe because of the fact that metadata for services are used in

---

[7] http://www.idee.es/csw-inspire-idee/srv/spa/catalog.search#/home

Table 4. Logical consistency results. Notes: (1) 60 with non-blank values; (2) 17 records cannot be evaluated by comparison with other metadata elements; (3) 11 with non-blank values; (4) records with non-blank bounding box extent; (5) records with non-blank parent identifier; (6) records with both bounding boxes.

Quality element	**#**	**sample**	**AQL**	**fail**	**result**
DQ_ConceptualConsistency (scale)	4824	80 (1) (2)	5	6	Fail
DQ_ConceptualConsistency (grid dimensions/geometric object count)	4824	80 (3)	5	2	Pass

Quality element	**#**	**pass**	**correct rate**	**result**
DQ_ConceptualConsistency (general)	4824	3792	78.61%	Fail
DQ_DomainConsistency (general)	4824	4578	94.90%	Fail
DQ_DomainConsistency (bounding box)	4801 (4)	4651	96.88%	Pass

Quality element	**#**	**fail**	**error rate**	**result**
DQ_FormatConsistency	4824	0	0.00%	Pass
DQ_TopologicalConsistency (parent identifier)	299 (5)	276	92.31%	Fail
DQ_TopologicalConsistency (topological contradiction)	74 (6)	31	41.89%	Fail

digital environments for service chaining that do not permit any completeness errors. In contrast, metadata describing datasets are prone to manual annotation (despite using metadata editors) and the introduction of some errors.

4.2.2. DQ_LogicalConsistency

Table 4 presents the results for evaluating conceptual consistency, domain consistency (general and for bounding boxes), format consistency and topological consistency.

With respect to the conceptual consistency referring to schema validation, it must be noted that it may be problematic in some cases. Some errors may arise because an XML element node contains unexpected subnodes. For instance, this happens in the case of dates: instead of having an XML element that contains the date, there is a sub-node with the date. The problem in these cases involves the way the error is classified because the error could also be considered for domain consistency.

On the other hand, to test contradictions beyond the schema validation for elements on the same metadata record, we have selected two of the more important elements for geospatial information: the scale (resolution); and the number of grid dimensions in the case of raster data or the count of geometric objects in the case of vector data. The most typical error in the first element is the inconsistency between scale denominator and the indication of scale contained in the title or abstract text. Moreover, there are some gross errors such as assigning a zero as a scale denominator when the scale is unknown or forgetting to add zeros in the denominator to indicate thousands. Considering the second metadata element, the most typical error involves filling the number of grid dimensions when the dataset actually uses a vector spatial representation.

In the context of domain consistency, it can be observed that we distinguish between the general domain consistency and bounding box consistency. General domain consistency is automatically evaluated by an XML Schema validator and it is mainly focused on checking the domain of literals. However, there are some numeric literals constrained by the standard that cannot be forced through the metadata XML Schema. For instance, this is the case of bounding box values. For this specific element, we created a specific program to assess this consistency: bounding boxes should use geographic coordinates and represent a valid rectangle. This is why there is a separate row in table 4.

From a general point of view, format consistency and bounding box domain consistency (including domain consistency) exhibit a low rate of errors and high rate of correct items, respectively, passing the quality check in both cases. Even general

Table 5. Temporal accuracy results. Note: (1) 88 with blank values.

Quality element	#	fail	error rate	result
DQ_TemporalConsistency	4824 (1)	12	0.25%	Pass
DQ_TemporalValidity	4824	1323	27.43%	Fail

domain consistency is close to passing the quality check (94.90%). However, the conceptual consistency quality control does not pass the quality check. This is because of the fact mentioned in previous paragraph: the identification of subnodes is not permitted by ISO 19115.

Table 4 also presents the results of topological consistency among different metadata records. As can be observed, the error rate is very high. Probably this is because metadata cataloguers have misunderstood the semantics of the *parentIdentifier* metadata element. Instead of referencing to a metadata record that describes the series of higher level dataset, this metadata element is used to indicate a common prefix of the *file identifiers* of a set of metadata records describing related datasets.

Finally, Table 4 also presents the intersection results between the bounding box that describes the general extent of a resource and the bounding box, when available, used to describe the data quality of a subregion of the resource. We restricted the analysis to the metadata records that contain both a bounding box for defining the geographic extent and a bounding box for describing the geographic scope in a data quality report. For the few cases of metadata records that defined the scope of data quality, the number of errors is numerous. Most errors are a consequence of confusing east and west coordinates or the exclusion of the negative sign, which can be attributed to human typing errors.

4.2.3. DQ_TemporalAccuracy

Table 5 presents the results for temporal consistency between the recorded dates for creation, revision, publication, and harvesting. We considered, because some date types are optional, that each missing date represents a valid date. Following this, the temporal consistency has a low rate of errors (0.25%). This implies that the temporal sequences of provided dates are valid in general. It is interesting to note that 7 records do not have metadata dates and for this reason they are considered unknown.

On the other hand, Table 5 also presents the results of temporal validity. Temporal validity is evaluated by comparing the harvesting date with respect to the temporal range defined by the metadata publication date plus the metadata maintenance and update frequency. The temporal validity exhibits a higher error rate with respect to temporal consistency. An error rate of 27.43% indicates that many metadata records are old or the metadata publication date has not been appropriately updated.

4.2.4. DQ_ThematicAccuracy

Table 6 presents the results of the control of the thematic classification correctness and the non-quantitative attribute correctness for the metadata elements indicated in Table 2 of section 3.3.4. This section explains the way in which the sample used for the quality control of each metadata element is defined.

First, in the context of classification correctness, the selected sample is a set of 80 random metadata records. From this set, only 76 have a non-blank set of keywords. Because this element is optional in ISO 19115, we have decided not to consider those metadata records with no keywords. A set of experts then manually controlled the

Table 6. Thematic accuracy results. Notes: (1) 76 with non-blank values; (2) unique values; (*) DQ_NQAC = DQ_NonQuantitativeAttributeCorrectness.

Quality element	**#**	**sample**	**AQL**	**fail**	**result**
DQ_ThematicClassificationCorrectness	4824	80 (1)	5	3	Pass
DQ_NQAC (*) (telephone number)	247 (2)	20	0	3	Fail
DQ_NQAC (*) (organization)	4824	80	5	17	Fail
DQ_NQAC (*) (address)	4824	80	5	1	Pass
DQ_NQAC (*) (responsible)	4824	80	5	5	Pass
DQ_NQAC (*) (title - a name that describes someones position or job)	4824	80	5	10	Fail

Quality element	**#**	**fail**	**error rate**	**result**
DQ_NQAC (*) (e-mail address)	390 (2)	57	14.61%	Fail

sample to check if the provided keywords are coherent with respect to the rest of descriptive information in the metadata record. The number of metadata records that do not pass the test is low, only 3 in this case. This implies that the selected sample, and therefore the estimated value of the whole corpus include a set of keywords well defined according to the list of data themes proposed by the INSPIRE directive (and included as part of the GEMET thesaurus). The typical errors detected by the experts include an incorrect understanding of the the meaning of the INSPIRE spatial data theme (e.g. a kilometric point is not a “Utility and governmental service”), or the considerable number of spatial data themes that have been assigned by default (e.g. an administrative units dataset has nothing to do with land use or land cover).

The case of qualitative elements is slightly different. In the full set of metadata records, there are 14426 elements that contain *e-mail addresses* (i.e. some metadata records may have more than one e-mail). These e-mails present some different formats and, in some cases, other contents are included (e.g. an URL) instead of a valid e-mail. After a simple process of normalizing values, we obtained 390 unique e-mails. Because e-mails can be checked automatically, in this case we prove all e-mail addresses by sending a mail and then account for the number of error responses from e-mail servers. The number of error responses of e-mail servers was 57. Therefore, 14.61% (57 out of 390) of e-mails in metadata records are wrong. This is not an estimation or control but an actual situation of e-mail address correctness.

With respect to *telephone numbers*, the corpus contains 13324 elements because a metadata record can include several instances of telephone number elements. These telephone number values have many different formats and, in some cases, other contents (e.g. an URL) instead of a phone number are included. After a simple process of normalizing values, we obtained 247 unique telephone number values. For this case the population size is N=247, and the control can be defined based on “Table A” of ISO 2859-2 (see Table 1) . Because there was one invalid telephone number in the sample, it can be stated that the corpus fails this quality control.

For the elements *organization*, *address*, *responsible*, and *title* it is not possible to normalize the corpus and determine unique values. For this reason, a sample is extracted from the complete metadata corpus (N=4824) using “Table A” of ISO 2859-2 (see Table 1) to define the parameters of the control in such cases.

The content of *organization* refers to the name of the responsible party or entity editing a metadata record. In this case, there were 17 errors in organization names. Some of these errors are because of official changes in organization names. This situation is common in Spanish public administration where ministries, departments, etc. change their names after election processes (e.g. MAGRAMA → MAPAMA). This situation involving name changes is a clear case that is related to the temporal validity

Table 7. Positional correctness results. Notes: (1) records with non-blank and valid geographic bounding box extent (2) 676 records cannot be evaluated because they lack textual reference.

Quality element	#	fail	error rate	result
DQ_PositionalCorrectness	4651 (1) (2)	337	7.25%	Fail

of data. It can be stated that the corpus fails this quality control.

The *address* refers to the physical address of the organization. In this case only 1 case was incorrect. We cannot distinguish between input errors, errors because of changes in names, or errors because of changes in the address of the organization. It can be stated that the corpus passes this quality control.

The *responsible* refers to the individual name of a person acting on behalf of a responsible party and is not a mandatory element. In this case, 5 metadata records including a responsible were incorrect. It can be stated that the corpus passes this quality control. However, the value is within the limit of the AQL.

The *title* element is a name that describes someone's position or job, which in this case is the title of the previously analyzed *responsible*. Contact with the organizations was performed by telephone calls. In this case, 10 of the metadata records were incorrect. It can be stated that the corpus fails this quality control.

4.2.5. DQ_PositionalCorrectness

Table 7 presents the results of the evaluation of correctness of the geographic bounding boxes found in metadata records defining the geographic extent of described resources. It only presents results for those metadata records with a non-empty and valid geographic bounding box extent.[8] In addition, it must be noted that because the proposed automatic method uses textual location references in other free-text metadata elements, there are 676 of those records whose positional accuracy could not be evaluated (14% of records lack textual location references). As the textual location is not mandatory, we assumed that these metadata records are correct.

The rate of the obtained measure is 7.25%, which is above the threshold of 4.16%. The selection of GeoNames as a knowledge base probably influences the obtained results. Although the coverage of place names in GeoNames is quite exhaustive, some fine-grain place names may not be found. In addition, GeoNames only provides point geometries for comparison with the bounding boxes defining the geographic extent.

4.2.6. DQ_QualityOfFreeText

We first present the results obtained for overall quality of the *lineage*, *purpose*, and *abstract* metadata element. The readability analysis of these free texts is then described. It must be noted that: mandatory elements like *abstract* have been completed in 100% of the records; the *purpose* element, which is optional, has been filled in 70% of the records; and the *lineage* element, which is conditional, is filled in a similar manner as *purpose* (70%).

The set of experts who evaluated the overall quality have a background in the areas of SDI, natural language processing and geographical information. The results obtained from the pool of experts (mode values) are shown in Figure 2 and the proposed measure is described in Table 8. The first conclusion is that the number of free texts identified as "very good" is very low. It can be even observed that the *abstract* element has not received any "very good" qualification. In the context of the "good defined" level,

---

[8] The analysis of valid bounding boxes is analyzed as part of the domain consistency.

Table 8. Results of the proposed measure for quality of free-text elements. Notes: (1) 3342 records with non-blank values; (2) 3363 records with non-blank values.

Quality element	#	sample	AQL	fail	result
DQ_QualityOfFreeText (abstract - overall quality of free-text)	4824	80	5	29	Fail
DQ_QualityOfFreeText (purpose - overall quality of free-text)	4824	80	5	40	Fail
DQ_QualityOfFreeText (lineage - overall quality of free-text)	4824	80	5	21	Fail

Quality element	#	pass	correct rate	result
DQ_QualityOfFreeText (abstract - readability of free-text)	4824	2946	61.07%	Fail
DQ_QualityOfFreeText (purpose - readability of free-text)	4824 (1)	2862	59.33%	Fail
DQ_QualityOfFreeText (lineage - readability of free-text)	4824 (2)	2554	50.10%	Fail

the results are very similar for all elements. Following this, the "regular defined" level represents a low percentage of the sample. Finally, it is very surprising that the percentage of "bad defined" free texts is very high: a 40% in the case of *abstract* elements, and 60% in the case of *purpose* elements. These high levels of bad descriptions probably indicate that there is little interest in metadata creation and its ulterior use. According to the measure defined in section 3.3.6, the quality check fails for all free-text elements.

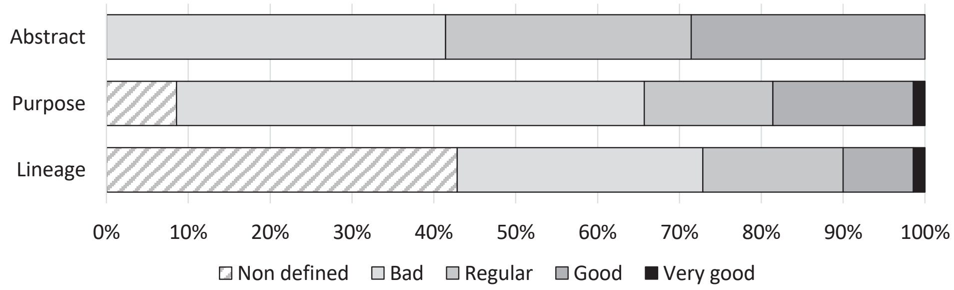

Figure 2. Analysis of experts on the three free-text elements.

In the context of readability, the readability difficulty has been defined by a number in the range of 100 to 0, with 100 being very easy-to-read texts, and values below 50 indicate complex texts at university degree level. In this sense, the three elements provide different results. Texts contained in *abstracts* have a medium-level readability difficulty (values from 80 to 60 are the most common in Figure 3). However, *purpose* texts have very high difficulty with almost 60% of values below 50. Finally, *lineage* texts, because they describe processes, also have a very high difficulty based on their readability index of almost 70% of the filled elements. It is important to notice that *purpose* and *lineage* have a great number of repeated elements, 75% of the elements are completely equal.

In the context of the measure proposed in section 3.3.6, the obtained values are 61.07%, 59.33% and 50.10% for the *abstract*, *purpose*, and *lineage* elements, respectively (see Table 8). Although all values are below the threshold value of 96.84%, the *abstract* appears to have been written more carefully.

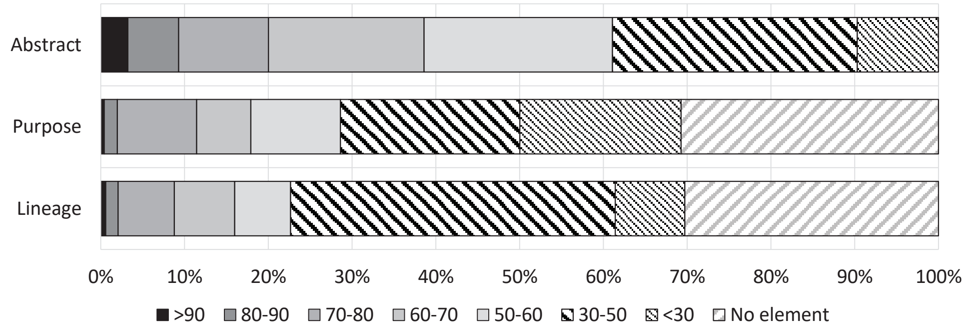

Figure 3. Readability index (Fernandez-Huertas modification of Flesch index for Spanish Language) of three free-text elements: *abstract* (M), *purpose* (O) and *lineage* (C).

5. Discussion

In section 4.2, we presented the results of applying the methodologies proposed in section 3.3. The results obtained for each quality element and measure are summarized in Table 9. The measures taken from ISO 19157 are linked to specific elements of quality. This implies that a defined measure for one quality element cannot be applied to another element. With "similar," we want to indicate that the mathematical construction of the proposed measure is identical to that of the identified measure.

From a general point of view, Table 9 presents three separate groups (first column). The first group represents the group of measures that can be checked using current services and methodologies like completeness, general domain consistency (that has almost passed the quality control), and format consistency. The second group represents metadata elements that are relevant and are filled with more interest by the producers, e.g. *bounding box* or *keyword.* The last group represents some semantic, internal quality checks and quality checks such as free-text measures, positional correctness or all thematic accuracy measures, which are not reviewed using the previously proposed measures and methodologies that are only oriented for interoperability. From this last group of measures the majority do not pass the quality check. In addition, table 9 indicates which measures are computed automatically (over the full corpus) or manually (over a randomly selected number of samples). It must be noted that most of the manual measures (8 out of 11) belong to the third group.

Moreover, by focusing on these three groups, we can highlight that the proposed quality control framework and measures include both current checks for interoperability (group 1) and new measures (groups 2 and 3) that would make metadata more expressive to DL search and to users in general. In addition, although we focused our approach of quality evaluation on ISO 19115 metadata, it could be easily adapted to other metadata standards by considering the most appropriate metadata elements of a new standard for each quality element.

On the other hand, it is fair to say that the proposed method covers the main quality elements of ISO 19157. However, it must be noted that given the complexity of ISO 19115 metadata standard, with more than 400 metadata elements, we had to select a set of representative elements, namely those used typically for discovery purposes (e.g., title, abstract, keywords, geographic extent, or responsible party) and those additional elements that illustrate quality elements (e.g., conceptual consistency of spatial representation properties) that are not easily applied to other elements. This selection of metadata elements probably has some influence on the obtained quality

Table 9. Summary of quality elements and the results obtained after applying the proposed measures to the corpus. Notes: (*) DQ_NQAC = DQ_NonQuantitativeAttributeCorrectness

Gr	quality element	measure	automatic	result
1	DQ_CompletenessCommission	D.3 (ISO 19157)	Y	Pass
	DQ_CompletenessOmission	D.7 (ISO 19157)	Y	Pass
	DQ_ConceptualConsistency (general)	D.13 (ISO 19157)	Y	Fail
	DQ_ConceptualConsistency (scale)	Similar to D.22 (ISO 19157)	N	Fail
	DQ_ConceptualConsistency (grid dimensions/geometric objects count)	Similar to D.22 (ISO 19157)	N	Pass
	DQ_DomainConsistency (general)	D.17 (ISO 19157)	Y	Fail
	DQ_FormatConsistency	D.21 (ISO 19157)	Y	Pass
2	DQ_DomainConsistency (bounding box)	D.17 (ISO 19157)	Y	Pass
	DQ_TopologicalConsistency (parent identifier)	D.23 (ISO 19157)	Y	Fail
	DQ_TopologicalConsistency (topological contradiction)	Topological contradiction	Y	Fail
	DQ_ThematicClassificationCorrectness	D.63 (ISO 19157)	N	Pass
3	DQ_PositionalCorrectness	Similar to D.33 (ISO 19157)	Y	Fail
	DQ_TemporalConsistency	Similar to D.62 using a rate	Y	Pass
	DQ_TemporalValidity	D.18 (ISO 19157)	Y	Fail
	DQ_NQAC (*) (e-mail address)	D.69 (ISO 19157)	Y	Fail
	DQ_NQAC (*) (telephone number)	D.68 (ISO 19157)	N	Fail
	DQ_NQAC (*) (organization)	D.68 (ISO 19157)	N	Fail
	DQ_NQAC (*) (address)	D.68 (ISO 19157)	N	Pass
	DQ_NQAC (*) (responsible)	D.68 (ISO 19157)	N	Pass
	DQ_NQAC (*) (title)	D.68 (ISO 19157)	N	Fail
	DQ_QualityOfFreeText (abstract)	Overall quality of free text	N	Fail
	DQ_QualityOfFreeText (purpose)	Overall quality of free text	N	Fail
	DQ_QualityOfFreeText (lineage)	Overall quality of free text	N	Fail
	DQ_QualityOfFreeText (abstract)	Readability of free text	Y	Fail
	DQ_QualityOfFreeText (purpose)	Readability of free text	Y	Fail
	DQ_QualityOfFreeText (lineage)	Readability of free text	Y	Fail

measures, but as most of the selected metadata elements are mandatory, we expect that the results using other elements would only improve (optional missing elements are evaluated as correct).

Finally, we want to remark that the evaluation methods, either manual or automatic, are subject to considering other alternatives. For instance, in the case of evaluating the quality of INSPIRE metadata, we can replace the ISO 19139 XML Schema validator with the INSPIRE metadata validation service to measure completeness commission, completeness omission, conceptual consistency (general) and domain consistency.

6. Conclusions

In this paper, we adapted ISO 19157 to assess the quality of metadata records of geospatial information. The analysis of quality for this type of metadata is a special case with respect to digital libraries because they follow a specific standard in this domain (i.e. ISO 19115) and describe complex geospatial information datasets and services. The proposal is complex, as are the standards that are referred to, but fulfills all aspects of the quality check from completeness to thematic accuracy. In addition, we proposed two new quality elements, named *DQ_Positional Correctness* and *DQ_QualityOfFreeText*, which are not considered in ISO 19157 but produce interesting insights in different contexts. Moreover, the new element of *DQ_QualityOfFreeText* and its measures include some semantic analysis of the metadata, which can also be applied to analyze the quality of thematic attributes included in datasets and containing free-text.

In the context of reporting the results of metadata quality, although beyond the scope of our research, the decision to adapt ISO 19157 has the advantage, if necessary, of using ISO 19157-2 (ISO 2016c) for encoding metadata quality reports in XML. According to certain studies (Boin and Hunter 2007), there is no empirical evidence on the usefulness of such types of technical reports for communicating quality to final users, but, at least, these reports can be generated in a machine-readable format and can be further processed by other systems.

In light of the results obtained through experiments for testing the feasibility of our method, the analyzed metadata do not surpass a majority of the quality controls. However, we consider that this does not imply that they are particularly ineffective in relation to the metadata that exists in the metadata repositories of other SDIs. Current metadata cataloging policies attempt to comply with interoperability between catalog services to assure proper functionality, e.g. the tested corpus passes the measures of completeness and format consistency. Nevertheless, this interoperability has not benefited the semantics and internal relations of metadata elements because they do not pass the proposed measures. In any case, it is a reason to reflect on the current processes of creation and maintenance of metadata and, above all, on the use and implementation of very static and complex metadata models.

In general, we envision that the proposed quality elements computed automatically (e.g., checking accuracy and consistency issues related to elements like e-mail addresses, bounding boxes or date stamps) could be integrated within current software of metadata editors in addition to the completeness and format verifications that are already typically applied (through XML Schema validation). Most thematic accuracy controls failed in our test, but small improvements in metadata editors could enhance the quality of published metadata. However, the existence of quality elements that must be evaluated manually and the study of existing metadata holdings justify the need of a separate framework for evaluating metadata quality.

Finally, we believe that our proposal can be improved by considering the addition of new measures on other metadata elements: the temporal consistency of process steps in the description of a dataset lineage, the topological consistency of the *MD_AggregateInformation* section (i.e., references to aggregate datasets and their metadata), or the relative internal positional accuracy between the geographic extent described by a bounding box and a bounding polygon (an overlap test should be considered). In addition, an increase in semantic measures of free-text metadata elements or the proposal of a global usability value could be valuable. As a future extension of this work, a full report of our ISO 19157 adaptation (with full details of UML diagrams describing quality element classes and tables explaining the related measures) can be submitted to the ISO technical committee for geographic information as a new part of ISO 19157 for metadata quality.

Acknowledgement(s)

This work has been partially supported by the Spanish Government (project TIN2017-88002-R), the Regional Government of Aragon (Spain) and the European Social Fund (code T59_17R), and the Regional Government of Andalusia (Spain) for the financial support since 1997 to the research group (Ingeniería Cartográfica) with code PAIDI-TEP-164. The authors also wish to express their gratitude to National Center of Geographic Information (CNIG) of Spain for providing the corpus (dated on September 2nd 2016) used to test the methodology, and to the experts that contributed to the

manual evaluation of some quality elements.

References

Boin, A.T. and Hunter, G.J., 2007. What communicates quality to the spatial data consumer. *In*: *Proc. 2007 Int. Symposium on Spatial Data Quality (ISSDQ 2007)*. 320–329.

Bruce, T.R. and Hillmann, D.I., 2004. *Metadata in practice.* Chicago: American Library Association. ALA editions, Ch. The continuum of metadata quality: defining, expressing, exploiting.

Díaz-Corona, D., Lacasta, J., and Nogueras-Iso, J., 2017. The quality of thesauri used in geographic metadata. *In*: *Proc. AGILE 2017 conference.*

European Commission, 2008. *COMMISSION REGULATION (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata.*

Fernández-Huerta, J., 1959. Medidas sencillas de lecturabilidad. *Consigna*, 214, 29–32.

Flesch, R., 1948. A new readability yardstick. *Journal of Applied Psycology*, 32, 221–233.

Hernández-Figueroa, Z., Carreras-Riudavets, F.J., and Rodríguez-Rodríguez, G., 2013. Automatic syllabification for spanish using lemmatization and derivation to solve the prefixs prominence issue. *Expert Systems with Applications*, 40, 7122–7131.

Hughes, B., 2004. Metadata quality evaluation: Experience from the open language archives community. *In*: *Proc. Int. Conf. on Asian Digital Libraries.* 320–329.

INSPIRE MIG, 2017. *Technical Guidelines for implementing dataset and service metadata based on ISO/TS 19139:2007.* INSPIRE Maintenance and Implementation Group (MIG). Version 2.0.1. `http://inspire.ec.europa.eu/id/document/tg/metadata-iso19139`.

ISO, 1985. *ISO 2859-2:1985 Sampling procedures for inspection by attributes – Part 2: Sampling plans indexed by limiting quality (LQ) for isolated lot inspection.*

ISO, 1994. *ISO 8402:1994 - Quality management and quality assurance.*

ISO, 2007. *ISO/TS 19139:2007. Geographic information - Metadata - XML schema implementation.*

ISO, 2013. *ISO 19157:2013 Geographic information - Data quality.*

ISO, 2014. *ISO 19115-1:2014. Geographic information - Metadata - Part 1: Fundamentals.*

ISO, 2016a. *ISO 19119:2016. Geographic information - Services .*

ISO, 2016b. *ISO/TS 19115-3:2016. Geographic information - Metadata - Part 3: XML schema implementation for fundamental concepts.*

ISO, 2016c. *ISO/TS 19157-2:2016 Geographic information - Data quality - Part 2: XML schema implementation.*

JGSI, 2002. *User's Manual for Spatial Data Product Specification Description, GSI TR A1-No.264. Technical Report of the Japan Geographical Survey Institute.* `http://www.gsi.go.jp/ENGLISH/RESEARCH/GIS/jsgi-manual.pdf`.

Lacasta, J., *et al.*, 2014. Improving the visibility of geospatial data on the web. *In*: *Proc. IEEE/ACM Joint Conference on Digital Libraries (JCDL 2014) and Int. Conf. on Theory and Practice of Digital Libraries (TPDL 2014)*. 155–164.

Ma, S., *et al.*, 2009. Evaluating the metadata quality of the IPL. *In*: *Proceedings of the American Society for Information Science and Technology.* vol. 46, 117.

Milenov, P., Sagris, V., and Wojda, P., 2010. *Report on LPIS Quality Assessment feasibility trial.* JRC technical notes.

Moen, W.E., Stewart, E.L., and McClure, C.R., 1998. Assessing metadata quality: Findings and methodological considerations from an evaluation of the US government information locator service (GILS). *In*: *Proc. IEEE Int. Forum on Research and Technology Advances in Digital Libraries, ADL 98*. IEEE, 246–255.

Nogueras-Iso, J., Zarazaga-Soria, F.J., and Muro-Medrano, P.R., 2005. *Geographic information metadata for spatial data infrastructures – resources, interoperability and information retrieval.* Springer Verlag.

NTHA, 2004. *Standards for geospatial locators.* Officials Committee for Geospatial Information. New Zealand by the National Topographic Hydrographic Authority, ESA Core Data Specification.

Ochoa, X. and Duval, E., 2006. *Advances in conceptual modeling - theory and practice.* Ch. Towards Automatic Evaluation of Learning Object Metadata Quality, 372–381.

Olfat, H., *et al.*, 2012. Spatial metadata automation: A key to spatially enabling platform. *International Journal of Spatial Data Infrastructures Research*, 7, 173–195.

Park, J.R., 2009. Metadata quality in digital repositories: A survey of the current state of the art. *Cataloging & Classification Quarterly*, 47 (3-4), 213228.

Renteria-Agualimpia, W., *et al.*, 2015. Identifying geospatial inconsistency of web services metadata using spatial ranking. *Earth Science Informatics*, 8, 427–437.

Shen, R., Gonalves, M.A., and Fox, E.A., 2013. *Key issues regarding digital libraries: Evaluation and integration.* Synthesis Lectures on Information Concepts, Retrieval, and Services, vol. 5. Morgan & Claypool Publishers.

Shreeves, S.L., *et al.*, 2005. Is 'quality' metadata shareable metadata? the implications of local metadata practices for federated collections. *In*: H. Thompson, ed. *Proceedings of the Twelfth National Conference of the Association of College and Research Libraries, April 7-10 2005*, Minneapolis, MN . Chicago, IL. Association of College and Research Libraries.

Tolosana-Calasanz, R., *et al.*, 2006. *Research and Advanced Technology for Digital Libraries, ECDL 2006*, LNCS, vol. 4172. Ch. On The Problem Of Identifying The Quality Of Geographic Metadata, 232–243.

Weagley, J., Gelches, E., and Park, J.R., 2010. Interoperability and Metadata Quality in Digital Video Repositories: A Study of Dublin Core. *Journal of Library Metadata*, 10 (1), 37–57. DOI: 10.1080/19386380903546984.

Xie, H., Tong, X., and Jiang, Z., 2008. The quality assessment and sampling model for the geological spatial data in china. *In*: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing. vol. XXXVII.