# Geospatial data quality (ISO 19157-1): evolve or perish

Calidad de datos geoespaciales (ISSO 19157-1): evoluciona o muere

Francisco Javier Ariza López*
Pablo Barreira González♦
Joan Masó Pau•
Alaitz Zabala Torres▪
Antonio Federico Rodríguez Pascual▲
Gonzalo Moreno Vergara○
José Luis García Balboa□

**Abstract**

Given the circumstance that the process for the revision of the international standard ISO 19157 is currently open, this article presents a critical reflection on its content, application and some challenges posed by the new types of data (e.g. big data, BIM data, etc.), that also have a geospatial component and to which, therefore, this international standard can be applied as well. Proposals are put *forward* going along

* Universidad de Jaén, España, e-mail: fjariza@ujaen.es.
  ORCID: https://orcid.org/0000-0002-5204-3630.
♦ Instituto Geográfico Nacional, España, e-mail: pbarreira@mitma.es.
  ORCID: https://orcid.org/0000-0002-1305-4847.
• CREAF, Fac Ciencies, UAB, España, e-mail: joan.maso@uab.cat.
  ORCID: http://orcid.org/0000-0002-2983-4629.
▪ Universitat Autonoma de Barcelona, España, e-mail: Alaitz.Zabala@uab.cat.
  ORCID: https://orcid.org/0000-0002-3931-4221.
▲ Centro Nacional de Información Geográfica, España, e-mail: afrodriguez@fomento.es.
  ORCID: https://orcid.org/0000-0002-0111-1349.
○ Instituto Geográfico Nacional, España, e-mail: gmvergara@mitma.es.
  ORCID: https://orcid.org/0000-0003-4717-4561.
□ Universidad de Jaén, España, e-mail: jlbalboa@ujaen.es.
  ORCID: https://orcid.org/0000-0002-3109-5888.

three lines of improvement, on the one hand the consideration of new data quality elements and on the other, the reinforcement of the interoperability of this international standard with other standards related to data quality, and finally various improvements (e.g. standardization of evaluation methods, clearly introducing the life cycle, improvement of the definition of metaquality, etc.) of the standard, which come from experience.

Key words: *geospatial data quality, ISO 19157-1, revision.*

## Resumen

Dado que actualmente está abierto el proceso de revisión de la norma internacional ISO 19157-1, este artículo presenta una reflexión crítica sobre su contenido, aplicación y algunos desafíos planteados por los nuevos tipos de datos (p.ej. *big data*, datos BIM, etc.), que también tienen un componente geoespacial y los que, por tanto, esta norma internacional también podría aplicarse. Las propuestas se presentan siguiendo tres líneas, por un lado, la consideración de nuevos elementos de calidad de datos, por otro, el refuerzo de la interoperabilidad de este estándar internacional con otros estándares relacionados con la calidad de datos y, finalmente, varias mejoras más específicas que se basan en la experiencia (p.ej. estandarización de métodos de evaluación, introducción del ciclo de vida del producto de datos, mejora de la definición de la metacalidad, etc.).

Palabras clave: *calidad de datos geoespaciales, ISO 19157-1, revisión.*

## Introduction

The economic relevance of geographic information (GI) and geospatial services has been valued in numerous studies in the last decade (e.g. PIRA, 2000), all of which indicate large returns and added value to society (e.g. reduction of travel times, reduction of emissions of contaminating gases, etc.). For instance, one of the latest studies (αlphaβeta, 2017) estimates that the global consumer's benefits from geospatial services are more than US$550 billion annually. GI is also recognized as a tool for good governance by many organizations (e.g. United Nations, the World Bank, International Federation of Surveyors, etc.). A very popular article published in The Economist has stated that "The world's most valuable resource is no longer oil, but data"[1] and an in-depth study conducted by Hahmann and Burghardt (Dresden University of Technology) in 2012 has shown that at least 78% of the information we usually manage is geospatial information.[2]

---

[1]   https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

[2]   https://www.tandfonline.com/doi/abs/10.1080/13658816.2012.743664

Taking into account the view that defines GI as a kind of model of the real world (Burroughs, 1986), its quality aspects can be considered a key element because they describe the quality of such a model and its relationship with reality, or more precisely with the universe of discourse, which is the part of the real world that includes everything of interest for a geospatial data product (ISO 19101).

The concept of quality of geographic information, as we know it today, was introduced into the international geospatial agenda only 40 years ago. It was in 1982 when the first studies began under the auspices of the American Congress of Surveying and Mapping. Thus, a proposal for a standard was created (Moellering, 1987) which referred to suitability for use, quality reports and five categories of quality elements (lineage, positional accuracy, accuracy of attributes, logical consistency and completion). It is relevant to indicate that we are still working with these elements of quality, with almost no changes. Around this time the International Cartographic Association promoted a book edited by Stephen C. Guptill and Joel L. Morrison under the title "Elements of Spatial Data Quality" (Guptill & Morrison, 1995). This manual was the catalyst for greater international concern regarding this subject in university and research fields.

The first international standard considering geospatial data quality was probably the DIgital Geographic Exchange STandard (DIGEST),[3] whose first version was issued in 1992 by the Digital Geographic Information Working Group which was established in 1983. This proposed standard, somewhat modified, was adopted as Federal Information Processing Standard 173 by the National Institute of Standards and Technology (NIST, 1994). DIGEST has become a NATO standardization agreement (STANAG 7074). After that, the European experimental standard[4] on geospatial data quality was approved in 1998 by the CEN/TC 287, containing a complete quality model for geographic information, and four years later the first version of ISO 19113:2002 (ISO, 2002) on quality principles was published, followed by another international standard (IS) on quality evaluation procedures (ISO 19114, 2003) and a technical specifications definition on quality measures (ISO/TS 19138) (ISO 2006). Finally, in 2013 the three documents were fused in a unique IS for geospatial data quality (ISO 19157) (ISO 2013), which was amended in 2018 for describing data quality using coverages. Finally, in August 2019 ISO Technical Committee 211 resolved to revise this IS, and today it is currently under revision. Therefore, although there is a solid theoretical framework and an ample bibliography regarding this matter, we believe that geospatial data quality is not fully implemented in production processes yet, probably due to several reasons, among others:

---

[3]  http://www.dgiwg.org/dgiwg/htm/documents/historical_documents.htm
[4]  ENV 12656:1998, Geographic Information - Data description - Quality.

- Geographic information production processes are arduous, expensive and take a long time to be completed, therefore sometimes the approach of producing the best possible product with the available budget is adopted. Taking quality into account is neither cheap nor simple.
- There are very popular solutions which do not pay too much attention to data quality. Some geospatial non-official data providers such as Virtual Globes (e.g. Google Earth) and volunteered geographic information (VGI) (e.g. Open Street Map) are widely used since the cartographic democratization[5], while their quality is unknown and/or unquantified and in spite of their approach based on "take it as it is". This situation is balanced out by great usability and quality of service, and sometimes openness and global coverage.
- Not only data producers but also data users and brokers have some kind of "quality immaturity" and there is not much geospatial data quality demand. This situation is probably due, among other causes, to the fact that there are many quality indicators and measures which makes the comparison of different datasets not easy.
- Most GIS tools have not taken data quality into consideration until now.
- The transition from "the best possible quality" to a level of quality that "fits for purpose" has not been completed by geospatial data producers and users.
- Probably the international standard ISO 19157:2013 is relatively recent, and its complete application will take a while in a sector with slow production processes and big inertia.
- On the other hand, new data sources and techniques are invading us with huge amounts of geospatial data which are frequently updated (e.g. satellite imagery, UAV, LIDAR, etc.). A very high updating frequency and the fact of being sensor data seems to leave little room to consider quality assessment.

In this context, it would be also extremely important to have an optimized quality framework based on the best possible geospatial data quality IS, and the final purpose of this article is to contribute to improving the technical content of ISO 19157 as much as possible.

ISO 19157:2013 is a good and complete standard covering all elements of geospatial data quality under a unique and consistent approach. It is better structured than the former ISO 19113, ISO 19114 and ISO/TS 19138 documents, includes sharp UML models, and introduces metaquality and some interesting considerations in the annexes about criteria on how to apply it and how to mix different quality elements.

Nevertheless, ISO 19157:2013 is a complex standard which has been thoroughly revised, and a quick analysis shows some gaps. For instance, in general there are few examples and no examples for important contents (e.g. metaquality), the user has no

---

[5]    [3] https://www.poynter.org/reporting-editing/2005/holovaty-wins-10000-batten-award-for-chica-gocrime-org/

chance to create its own quality elements, a model for quality reports is missing, raster and image quality is not taken enough into account, several measures have definition problems, etc.

The ISO/TC211 has already held some meetings in which this issue has been discussed (e.g. sessions in Toulouse (France) and Omiya (Japan)). Recently (Malta, January 2020) Eurogeographics has organized a meeting where the process as well as suggestions received by TC211 from the 28 experts who are collaborating in the review have been presented.

This paper presents some ideas coming from experts nominated by the Spanish technical committee 148 of UNE (previously known as AENOR). Our objective is multiple, on the one hand to indicate the greatest weaknesses and deficiencies in the data quality model of ISO 19157 and its application, and on the other hand to indicate a set of current opportunities and challenges in relation to data that require location and the quality of this data. Finally, we indicate a set of improvements that should be adopted in the new version of ISO 19157 if we really want it to be applied in an extensive, intense and correct way. So this article aims to examine in-depth those and other problems of this IS and try to propose solutions to improve it in the complex arena of the current geographic data ecosystem. In the next section a critical analysis of ISO 19157:2013 is presented, trying to identify weak points and areas of potential improvement and proposing solutions, new approaches and future lines of progress. Section 4 is devoted to the new challenges facing geospatial quality due to the new types of data, like the aforementioned ones, which probably requires a new approach based on some kind of "quick and big" quality information. As a consequence of sections 3 and 4, a set of proposals for the revision of ISO 19157:2013 is presented in section 5, with the aim of enhancing the applicability and usefulness of the standard. In some points a quite complete idea about how to update the standard is provided, while in other cases just some ideas and concepts are included. Finally in section 6, some conclusions summarizing the contributions are outlined.

## Critical analysis of International Standard ISO 19157 and its application

Quality should be used by producers and users. One natural way is to include data quality exigencies in the specifications. Another way is to include the data quality results in the metadata, and for this reason we will start this section talking about specifications and metadata. Ariza-López and Rodríguez-Pascual (2018) presented a small study (in April 2018), consulting the information available on 19 websites of National Mapping Agencies of the American continent and they found that only 11 times ($\approx 58\%$) metadata of the available data and services were published, that only one organization (5.3%) publishes quality information of its data beyond the lineage and that on only 6 occasions (32%), was descriptive information of the available geographic data products published, although on 4 occasions it was labeled

"Technical Documentation" and in one case presented as specifications. In addition, this is not unique to the American continent. A simple review of the sections dedicated to quality in the implementation rules of INSPIRE also give, in many cases, a sense of insufficiency. These simple studies indicate that there are problems with the inclusion of quality aspects in the specifications and metadata of geospatial data products (ISO 19131 and ISO 19115-1, respectively), or rather, that there are problems with the use of the quality framework proposed by ISO 19157.

The original ISO 19115 metadata standard is the base for the metadata records included in Spatial Data Infrastructures and clearinghouses catalogues that collect descriptions of geospatial data products. As data quality elements are an integral part of the metadata model we should expect that most producers provide a comprehensive description of the different components of the data quality in their metadata records, but this is not the common case. An analysis of the metadata harvested by an old version of the GEOSS[6] portal reveals that most datasets include no data quality indicators and if they do they rarely go beyond positional accuracy (Zabala *et al*. 2013).

ISO 19115 has been always blamed for being long and complex. Indeed, the standard is very comprehensive, with more than a hundred properties that can potentially be populated. Its complexity lies in the difficulty in separating the dataset description into so many properties, making the creation of a metadata record tedious and time-consuming. Despite these difficulties, individual properties are well defined and relatively easy to understand and populate if the information is at hand, with one exception: data quality. Actually, a quality model is included in ISO 19115 but this document alone does not provide enough details. ISO 19115:2003 mentions the 15 subclasses and some properties needed to specify quality measures but it does not include any concrete measures. We need to refer to ISO 19157:2013 to discover a list of about 80 quality measures, each one with the methodology and statistical analysis needed to extract the result. In our experience, the ISO metadata standard is very popular among practitioners but they rarely have access to ISO 19157, and so data quality remains relatively unknown. This could be remedied by metadata tools providing the necessary alternatives and information to the user, but the current metadata editors do not develop data quality in the necessary level of detail. Unwittingly, TC211 might have made the situation worse by removing the data quality element from ISO 19115-1:2014 and delegating its definition to ISO 19157. There is a need to make ISO 19157 known among the community as well as to metadata editor developers.

ISO 19157 organizes the quality measures into six classes representing mainly the components of the information (spatial, thematic, temporal, logical, etc.) that are

subdivided into 15 subclasses (data quality elements) which define what is measured (omissions, commissions, absolute accuracy, topological consistency, etc.). Conceptually, there is no reason why this could not be extended to other aspects of quality (e.g. redundancy, quality of metadata, quality of service). However, the design rules used to create the subclasses by generalization make them difficult to extend once they are encoded in a data format such as XML. This limits the extensibility to each revision of the standard. In fact, included in ISO 19157 was a new class for usability as an effort to extend the scope of the IS beyond the producer perspective into the user perspective. However, usability is described in a confused manner as an aggregation of producer's measurements and conformance to requirements, which might not be the best approach. Instead, a new model for user-created quality in the form of feedback could be better.

The measures included in ISO 19138:2006 Annex D were collected among the ones commonly used by mapping agencies. The last revision of the list, now conforming annex D of ISO 19157:2013, included a few new additions. Meanwhile geospatial information has become popular in other sectors, making the current scope quite limited. One example of a possible extension of scope is the big data world, where modeling is regularly used to investigate and predict Earth variables at regional or global level. These models are validated in different ways that are not necessary compatible with the current list of measures. The need for other quality measures is also observed in the emerging crowdsourcing, citizen science or other non-authoritative data products. The metadata records are also under scrutiny and new measures are proposed to evaluate the quality and completeness of their descriptions. Another gap is introduced by the very same ISO 19157 when providing an example for metaquality. The example used in Annex E.3 introduces a confidence quality measure called "safety factor". Unfortunately, ISO 19157 Annex D does not contain the description of this measure, obfuscating the usefulness of the example and leaving the reader with incomplete information on how to report metaquality. Ideally, ISO19157 should be encoded as an extendable list of quality measures. Annex H of ISO 19157 provides some clues on how an extensible catalogue of measures should be built as a registry. However, the annex ignores the possibility of making the measure available as a dynamic ontology which includes the current Annex D as well as other measures from other standards and best practices (e.g. ISO 8000, ISO/IEC 25012, ISO/TR 21707, GUM[7], VIM[8] guides). Ideally, this ontology would link to actual examples on how to use the measures in real live cases. Even if new quality measures seem necessary, we have to be careful in increasing the list too much: the more measures we have the more difficult it becomes to compare the quality of datasets with different origins.

[7] https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf
[8] https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf

Data quality measures are based on applied statistics. The same statistic can be applied to different measures used in different components and with different thresholds or parameters. This is particularly true for assessing uncertainties as described in ISO 19157 Annex G. The same statistic is applied in several quality measures of annex D, making the selection of the right quality measure unnecessarily confusing. In our opinion current measures do not make enough effort in separating the domain of numbers (e.g. the individual uncertainties of each feature) from the statistical expression or mathematical metrics that will be applied to the domain in order to summarize the quality into a measurable indicator (e.g. standard deviation), adding an unnecessary level of confusion. By separating the domain from the metrics, applications and metadata editors could focus on the latter while experts and practitioners could concentrate on preparing and specifying the domain.

At present, some users as well as producers consider that ISO 19157 is too focused on quality evaluation. The relation of the data quality measurements with the life cycle management of the data product is introduced in Annex B, enumerating which stages of the life cycle quality evaluation should be performed and this could be reported as part of the quality evaluation description as metaquality. More emphasis should be placed on this topic. Only a mature and easy to use quality standard that considers all steps in the life cycle of a product will result in a certification of geospatial data products incorporated into industrial procedures.

Another problem presented by this model, as formulated and applied, is the grain. The model presents problems working at the instance level. This is problematic in the case of aiming to ensure traceability over instances and also to derive global quality values from data products generated by aggregation of instances from various sources (Ariza-López & Rodríguez-Pascual, 2018).

Finally, the revision and improvement of this international standard should be considered in a broad sense that also includes aspects of the quality of geoservices and producers. Let us think that many data are offered through geoservices and that the quality of the data producer, as an organization, affects the quality of your product (the data). Ariza-López & Rodríguez-Pascual (2018) present an analysis with this more global perspective of the challenges of quality in geospatial data.

## New types of data: the challenges

Since The Economist (2017) published a story titled, "The world's most valuable resource is no longer oil, but data", the sentence "data is the new oil" has become a usual way to indicate the value of data. In this way, the so-called data economy (Wikipedia, 2020) is a main concern of world and regional institutions (e.g. UN-EAPD 2019 and EU 2020). Geospatial data and these types of data indicated below, and many other types of data, are part of this data economy. From our point of view, ISO 19157 is focused on geospatial data but with a classical producer

perspective. For this reason, and thinking about the data economy, we consider that new perspectives must be included in the revision of this international standard in order to guarantee a higher level of application and convergence between different highly related data-types, at least considering the following types:

- Big data are "high-volume, high-velocity and/or high-variety information assets which demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation" (Gartner, 2020). There is no specific type of big data. The relationship of big data with geospatial data is clear (Eldawy and Mokbel, 2013), many of the geospatial data (e.g. remote sensing images) are true sets of big data. The relationship is so obvious that many authors talk about geospatial big data (Lee and Kang, 2015; Robinson *et al.*, 2017). Quality issues are pointed out as an important challenge in geospatial big data (Robinson *et al.*, 2017; Lee and Kang 2015; Li *et al.*, 2016).

- Building information model (BIM) data sets are model-based geometric information, enriched thematically, semantically and relationally which, managed with the right software tools, allow a more efficient management of buildings and facilities (Ariza-López *et al.*, 2019). BIM data are very similar to geospatial data and also deal with geospatial data because they must be integrated into a geographical framework (the actual location of the building) environment (the surrounding geographical-topographic reality), and also collect the presence, dimensions, positions and exact attributes of the elements of interest. BIM data are directly related to 3D geospatial city data. Puyan *et al.*2017 highlight the interest and importance of the quality of BIM data for facility management purposes and present examples of errors in BIM data that are very similar to those that occur in geospatial data. In this way, Puyan *et al.* (2017) show the close links between geospatial data and BIM data, Song *et al.* (2017) indicate the need and benefits of the integration of BIM and GIS and Ariza-López *et al.* (2019) develop BIM data quality controls based on geospatial data quality elements.

- Volunteered geographic information (VGI) (Goodchild, 2007) is a kind of participative/collaborative geospatial data where citizens, often untrained and regardless of their expertise and background, create geographic information. The quality of VGI has been a hot topic from the beginning of this trend, and there are a lot of papers dealing with this topic. The need for specific data quality elements, metrics and methods is clearly pointed out in Gusminia *et al.* (2017), Degrossi *et al.* (2018), Senaratne (2017).

- Statistical data are those produced by statistical agencies. Official statistics data are grouped into topics (e.g. economy, population, international trade, etc.) (SDMX 2009). These data have different levels of aggregation ranging from microdata (e.g. the income of a person) to an added value for a country (e.g. gross domestic product). There is a clear confluence between statistical data and geospatial data so that many statistical organizations (e.g. INEGI in Mexico, IECA in Andalucía (Spain), Eurostat in the European Union, etc.), are producing

geospatially enabled-statistical data and micro data. The current trend is that all statistical data have a location. There are already grids with geospatialized statistical data for some regions (Eurostat, 2020). There is a conceptual framework in place to include the geographical component (UN-ISGI, 2018; Moström *et al.*, 2019). Data quality and quality management are relevant issues for official statistics (UN-SD, 2019).

- Earth observation data and images. These are clearly geospatial data. They are also big data. Unfortunately, ISO 19157 is difficult to apply directly to this type of data. With the aim of process simulations in the Earth monitoring scope (e.g. climate change), there is a great concern about the quality of images and derived products (e.g. essential climate variables)[9]. The Committee on Earth Observation Satellites developed the project Quality Assurance Framework for Earth Observation (QA4EO) where a series of key guides provide a quality assurance framework for images and related processes, but ISO 19157 is not applied.

- Geolinked data. Linked Data are defined as "structured data which is interlinked with other data so it becomes more useful through semantic queries; It builds upon standard Web technologies such as HTTP, RDF and URIs, but rather than using them to serve web pages only for human readers, it extends them to share information in a way that can be read automatically by computers" (Wikipedia, 2020). Geolinked data or Linked Geodata[10] consists of enriching the web with geospatial data. Geospatial data are linked to other data and any other type of data can be linked to a position given by geospatial data. Some national mapping agencies (e.g. OS in UK, IGN in Spain) already offer linked geospatial data. The relevance of linked geospatial data is clearly indicated in López-Pellicer *et al.*, (2011). Working with the graph established by the links adds a degree of complexity to aspects of data quality. Zaveri *et al.* (2014) carry out an exhaustive work of compilation and organization of the numerous dimensions and measures that can be applied to evaluating the quality of the linked data. There have also been initiatives focused on quality dimensions and measures for geospatial linked data (GeoKnow, 2012).

- IoT data are data produced by a "system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction" (Wikipedia, 2020). IoT systems are related to digital twins and mirror spaces. Lot data refer to data derived from sensors (e.g., humidity, rain, heat stroke, temperature, etc.) that monitor real-world situations, and actuators (e.g. stepper motors, control valves, switches, etc.) that can modify real-world situations. Location powers the analytic capacity of IoT data-based systems. As indicated by Karkouch *et al.* (2016), data quality is crucial to gaining user engagement and acceptance of the IoT paradigm and services. There are several studies on the

---

[9]   http://www.qa4ecv.eu
[10]   http://linkedgeodata.org

quality of IoT systems (Ahmed *et al.*, 2019), but the paper by Karkouch *et al.* (2016) focuses on the quality of the Iot data. In this paper, some of the established data quality dimensions are assailable to categories of geospatial data quality elements or geospatial data quality elements. It also points out several problems (e.g. outlier management) that are very similar to those inherent to management in the geospatial data domain.

## Proposals for the revision of ISO 19157-1

As explained in previous sections, we can assert that there is a need for major changes which comes from: i) the adoption of new perspectives in terms of data, ii) the need for a greater interoperability with other ISO international standards and, iii) the experience acquired in the application of ISO 19157.

### *Candidate data quality elements*

In order to adopt new perspectives new data quality elements need to be defined, and this opportunity must be opened up as well (as in ISO 19113). Some examples of new data quality elements can be trust dimensions for open and linked data (Zaveri *et al.*, 2013), the quality of free text for descriptive texts included in metadata records (Ureña-Cámara *et al.*, 2019) and quality elements proposed for images, photogrammetric flights or other spatial gridded products (Ariza-López, 2013).

Additionally to this approach, we can consider that a DQ-Element can be split into sub-elements, however, this makes the implementation and extensibility of ISO 19157 more complex and it does not add much value. Nevertheless, creating an attribute in a DQ-element to include the quality subclass name and defining a code list makes it easier without losing any functionality. This will imply the need for new "quality categories" or dimensions, new measures, and so on.

The current version of ISO 19157 is more focused on vector data, leaving gridded data as something not frequently used. This fact makes the rule difficult to apply to this kind of geospatial data product. There is a need to place the focus on other data types (e.g. LIDAR, BIM, geo-linked data, etc.).

In order to overcome this obstacle several changes need to be undertaken so that ISO 19157 can be applied to raster data correctly, covering all its aspects.

I   Data compression: data represented in a grid is always a consequence of generalization. In this sense, raster formats tend to compress data in order to represent variables in a more discrete way, when most times they are continuous. To cover this issue a new DQ_Element called Compression will be very useful.

II  Many raster products are obtained from satellite images or photogrammetric flights. These scenes present some deficiencies most of the time: pixels failures,

rows/columns with no data, the presence of clouds or other weather conditions that hinder the extraction of information, shadows, etc. This is the case not only for specifications but also for DQ assessment of a product.

III    Representing physical variables through raster data: many remote sensing products which are geospatial data locate physical variables (such as temperature, pressure...) in space. Since these variables are most times completely different, the IS should leave open the chance of proposing specific DQ measures and elements for this kind of information.

IV    Orthophotos are another usual example of gridded data that should be able to have its own DQ elements and measures. In this sense, measures that cover the percentage of censored data in the orthophoto or the mosaic, mosaic cuts, or the degree of geometric and radiometric matching among adjacent orthophotos should be applied to this type of product.

 V    Photogrammetric flight planning assessment, as part of the lifecycle of a product (e.g. orthophoto), is of relevance for producers and needs to be considered in this sense as well (Ariza-López, 2013).

VI    Measures for thematic classification: in this sense there are several spatial metrics which come from the world of landscape metrics that can present some clues to understanding the degree of accuracy in terms of thematic classification. These metrics describe the shape, distance between patches of the same category, fragmentation of the category and so on, that could be applied to the classification obtained and the real world. The comparison of these metrics could contribute to a better DQ assessment. Also, alternatives to the confusion matrix and derived parameters (e.g. Kappa index) should be considered as they are present in remote sensing studies (e.g. ice Coefficient, Relative bias, among others (Padilla *et al.*, 2015)).

VII    Data derived from Lidar is becoming one of the main sources in geospatial products. ISO 19157 should cover this specific type of data as well as others, not only in terms of positional accuracy but also thematic accuracy and redundancy, which plays a key role in this kind of data.

Some examples of possible new elements taken from official documents are presented in Tables 1.a, b, c. Thus, the possibility of developing new elements should be left open. This opportunity existed in ISO 19113 but was eliminated in ISO 19157:2013, which goes against the backwards compatibility that should be sought with the new versions.

**Table 1a**
**Example of a new quality element (Case 1)**

| | |
|---|---|
| *Name:* | Radiometric discontinuity |
| *Definition:* | Closeness of the radiometric values of homologous pixels of two images in a common area, is defined as: "S'il y a effeccivement eu un mosaïquage, la classe de précision sera exprimée par la différence de valeur radiométrique par canal tolérée sur les raccords entre images ne correspondant pas à un linéament, divisée par la radiométrie maximale de l'image et exprimée sous forme de pourcentage" |
| *Justification:* | When creating mosaics of images, the presence of radiometric discontinuities is a common circumstance derived from many different circumstances and reasons. This situation is undesirable. The presence of a radiometric discontinuity is an aesthetic and an exploitation problem |
| *Source:* | Ministère de l'équipement, des transports, du logement, du tourisme et de la mer. Arrêté du 16 septembre 2003 portant sur les classes de précision applicables aux catégories de travaux topographiques réalisés par l'Etat, les collectivités locales et leurs établissements publics ou exécutés pour leur compte |

**Table 1b**
**Example of a new quality element (Case 2)**

| | |
|---|---|
| *Name:* | Integrity |
| *Definition:* | Is defined for aeronautical data as: a degree of assurance that an aeronautical data and its value has not been lost or altered since the data origination or authorized amendment |
| *Justification:* | This aspect is of great relevance for nautical and airspace security. The data quality model of the International Civil Aviation Organization considers this issue. This issue can be of interest for other uses and purposes (e.g. homeland security, army, medical urgencies, fiscal data, etc.) |
| *Source:* | ICAO (2010). ANNEX 15 to the Convention on International Civil Aviation. Aeronautical Information Services |

**Table 1c**
**Example of a new quality element (Case 3)**

| | |
|---|---|
| *Name:* | Geometric fidelity |
| *Definition:* | Geometric fidelity is the measure defined as: that any real world alignment or shape, when viewed at the source survey scale, must be accurately reflected in the data to the required specification |
| *Justification:* | Real world objects (e.g. buildings) can be registered in a dataset without their exact and true relationships with their surroundings. It is necessary to have an assessment of the number of such types of objects in the dataset This information is relevant for the producer (quality of his processes or supplies) and for users |
| *Source:* | OS (2007). TOPO-96 Data quality. Ordnance survey |

### Interoperability with other International Standards

In order to achieve greater interoperability with other ISO international standards or documents (e.g. ISO 8000, ISO/IEC 25012, ISO/TR 21707) and third-party documents (e.g. VIM, GUM, etc.), which are interrelated to ISO 19157, means the necessity of a great effort of coordination in concepts, terms, perspectives, etc. Along this line, a well-defined ontology is missing for the quality of geospatial data. This ontology should be compatible with other ontologies of other application domains (e.g. dqv[11] of the W3C), and allow greater interoperability between different types of data.

Clear examples are the terms accuracy and uncertainty, which are not appropriately defined within ISO 19157, implying the need for a complete revision of many of the terms and measures proposed in it. Another example entails quantitative quality assessment, where a distinction between estimation and control is needed. Estimation means the precise determination of a parameter value and its confidence interval. Control (quality control) means taking a decision about the acceptation or rejection of a previously-stated hypothesis within a statistical framework of assumed risk.

Another issue entails the relationship between ISO 19157 and ISO 19131, for which to many of those that have applied both, the differences are not always completely clear. In the data specifications (according to ISO 19131) of many national mapping agencies there are usually no quality specifications on the data, and in some cases evaluation results appear when they are actually metadata. This is completely incorrect. More clarification is required, maybe an informative annex can help.

### Experience applying ISO 19157

Finally, the experience acquired in applying ISO 19157 shows us the bulk of the changes that we consider is relevant to introduce in its revision.

#### Distinction between the quality of a dataset and the uncertainty in individual attributes of a feature

The scope element allows narrowing the scope of data quality to a particular attribute but unfortunately it is rarely implemented. We believe this is partly because it is not clearly explained in the current version of ISO 19157. In addition, in the world of Big Data and unstructured records where the abundance/redundancy of information about a single property can be used as a quality test, the distinction mentioned above is becoming more important. This is particularly true in case as of citizen science of

---

[11]   https://www.w3.org/TR/vocab-dqv/

VGI, where redundancy in acquisitions and expert validations are used as quality indicators. A paradigmatic example is the identification of pheno-phases in a certain location and time, a situation in which redundant and coherent measures increase the veracity of each individual measurement that need to be quantified. Current measures in ISO 19157 are still applicable but they require a better definition of the individual inputs to assess.

### The quality assessment methods and their standardization

The use of the same, well specified, data quality elements, scopes and measures does not ensure the compatibility between the results of two assessments if different quality assessment methods were applied. Even if these methods have place in the

**Table 2**
**Example of a template for methods standardization**

| Line | Component | Description |
|------|-----------|-------------|
| 1 | Method identifier | Unique identifier within a namespace |
| 2 | Name | Name is the name of the method |
| 3 | Purpose | A description of the purpose of the quality assessment method |
| 4 | Method type | Indication of the method type (direct or indirect) |
| 5 | Result type | Indication of the result type (quality estimation or quality control) |
| 6 | Description | A general description of the method |
| 7 | Source | Identification of an explaining source(s), if exists |
| 8 | Detailed description | Description of the assessment |
| 8.1 | Full inspection based | Explanation of the full inspection process |
| 8.2 | Sampling based | Explanation of the sampling based process |
| 8.2.1 | Sample Scheme | Explanation of the sample scheme |
| 8.2.2 | Sample Size | Calculation of the sample size |
| 8.2.3 | Sample collection | Explanation of the sample collection process |
| 8.3 | Resources | Description of resources to be used |
| 8.3.1 | Instrumental | Specifications about instruments |
| 8.3.2 | Human | Specifications of skills |
| 8 | Measures | Identification of the standardized measures to be used |
| 10 | Procedure | A complete identification and explanation of the steps of the evaluation method |
| 10.x | Step x | Explanation of each step of the procedure |

data model we consider that quality assessment methods should be better and they require a very similar treatment to measures (standardized measures of annex D of ISO 19157). This implies the definition of a list of components and also a catalog of standardized methods. Methods must be well defined in order to be understandable, replicable and their results interoperable. A standardized quality assessment method must offer an unambiguous definition, and a structure of the methods extended in a way that is presented in Table 2. Also, the methods must be established in the data product specifications (following ISO 19131), because product specifications must include quality requirements and how (the method by which) their achievement is measured.

In the current ISO model, conformity is a binary value that should also be associated with a methodology. There is a need for presenting a relationship with the data quality assessment methods are the concepts of conformity level and the quality control decision. The first refers to the minimum good quality level (e.g. at least 90%) or the maximum bad quality level (e.g. at most 5%) that the user is willing to accept. A standardized measure, the units of the measurement (e.g. m or mm) and the value are the key elements for defining a conformity level. The second refers to how an acceptance/rejection decision is taken in a quality control. This decision must be taken by comparing a result of a standardized quality assessment method versus a conformity level by means of a given rule where producers' and users' risk are previously established. The conformity level must be established in the data product specifications (following ISO 19131) as well as the quality control decision rules.

## Metaquality and its use

There are very scarce references to the actual use of metaquality. One of the problems is that there is a lack of clear examples. An informative annex added to the new version of the IS could be adequate. In this sense, we would like to mention the use of metaquality developed in UNE 148002:2016 (UNE 2016) (see Figure 1), where metaquality elements (confidence, homogeneity and representativity) are more detailed. In order to cover several approaches to studying metaquality; for example, they divide the representativity into several topics: spatial, temporal, thematic, participative and global. They introduce the concepts of qualitative and quantitative confidence as well, which could also be interesting for the revision of ISO 19157.

## Report and its standardization

ISO 19157 indicates that in order to provide more details than those reported as metadata, a standalone quality report may additionally be created. The structure for this standalone quality report remains unspecified. This situation leads to reports being very different and resulting in a clear problem in terms of interoperability, for

"In this standard confidence on the results of a PQC process is determined by two complementary aspects:

- Qualitative. The rigorous application of the methods is the main guarantee of trust from a qualitative perspective. This aspect must be ensured by the participation of experts in the quality of geographic information in work teams and by the requirements stated in section 9 of this standard.
- Quantitative. Effective enforcement of the following aspects is the basis of trust from the quantitative perspective: sample size, randomness, independence of the control process and greater accuracy of the reference SDS. These aspects should be ensured by compliance with the requirements of sections 8 and 9 of this standard.

In this standard homogeneity of the results of a PQC is determined by:

- Production of the controlled SDS. The homogeneity of the controlled SDS is beyond the scope of this standard; however it should be noted that it can be a critical aspect in the case of SDS where numerous persons or organizations have intervened, where concur diverse backgrounds, knowledge, skills, etc., or different work methodologies are applied (e.g. OpenStreetMap).
- The extension of the control process. For PQC processes dilated in space or in time appropriate quality management measures shall be taken in order to ensure homogeneity of the PQC process at all times. Key elements to ensure homogeneity are, among others: documented procedures, the establishment of standards in education and training of personnel involved including verification mechanisms to ensure consistent processes, etc.

In this standard representativity of the result of a PQC should be evaluated from multiple perspectives. Since the assessment is based on sampling, the representativity should be:

- Considered in relation to the following aspects:
  - *Space. The spatial representativeness of the sample by its effective spatial distribution compared to the actual spatial distribution of the population.*
  - *Time. The temporal representativeness of the sample by its effective temporal distribution compared to the actual temporal distribution of the population.*
  - *Theme. The thematic representativeness of the sample by its effective thematic distribution of categories and attributes compared to the actual thematic distribution of the population.*
  - *Participation. In the case of studies with the participation of various organizations (e.g. national series) or individuals (e.g. OpenStreetMap), has the same sense as previous cases but related to the participation issue in the population.*
  - *Global. Refers to global representativeness as an interpretation of all the partial representativeness given above being considered in a specific PQC.*
- Evaluated with appropriate techniques, some techniques are applicable:
  - Visual comparison of histograms and distribution functions of the sample and the population.
  - *Adherence tests between the curves representing the distribution functions of the sample and the population (e.g. by means of the Kolmogorov-Smirnov test for continuous cases and Chi2 for discrete cases)".*

**Figure 1.**     Translation of section 10 of the Spanish UNE 148002:2016 standard.

data quality comparison and for data certification. To deal with this problem a well-defined and flexible quality report is needed.

## *Quality along the lifecycle.*

Geospatial data quality, as presented in ISO 19157, is not understood as something that must take place along a product lifecycle. This idea (data quality along the lifecycle, in a process) must be presented and explained in the early beginning of the IS since it is one of the keys to understanding the concept of quality related to geospatial data. In our experience, most of those who apply the 19157 standard, or who wish to apply it, only focus on the quality of the final products. They do not anticipate that ISO 19157 should be applied throughout the entire product life cycle. We believe that this problem has several causes, being one of them that the IS does not sufficiently explain it and does not present example about this. In addition, ISO 19157 scope needs to be widened to include such as photogrammetric and topographic production processes. Currently, it difficult for specialists in these fields to apply the standard.

Therefore, following this line, at least two additions are required in the revision of ISO 19157: on the one hand the idea of geospatial data products´ life cycle, and on the other hand a model of how to apply quality throughout this life cycle. For the first addition there are several proposals for geospatial data product lifecycles, for instance the United States Geological Survey[12] considers five phases (plan, acquire, process, analyze and preserve and publish/share). A proposal for quality throughout the lifecycle is that of the *Arbeitsgemeinschaft der Vermessungsverwaltungen* (AdV, 2002) which was adopted and improved by the *Instituto de Estadística y Cartografía de Andalucía* (IECA, 2011). As an example, a new version of this model is presented in Figure 2, which includes the Plan Do Check Act perspective (Deming's Cycle) using the stages of the USGS lifecycle (plan, acquire, process, analyze, preserve, publish) and the quality assessments proposed by the AdV and improved by IECA (2011). The Cross-cutting elements (describe, manage quality and Backup Secure) of the USGS data lifecycle are also labeled. This model includes the main quality management functions (quality control, quality improvement, quality assurance and quality deployment):

- Q1: Assessment of the basic model against general and strategic guidelines.
- Q2: Assessment of the application model and specifications against the basic model.
- Q3: Assessment of the application model and specifications against specific requirements.
- Q4: Assessment of the data product against its logical consistency rules.

---

[12]    https://www.usgs.gov/products/data-and-tools/data-management/data-lifecycle

- Q5: Assessment of the data product against the real world.
- Q6: Assessment of the data product performance for analysis and uses.
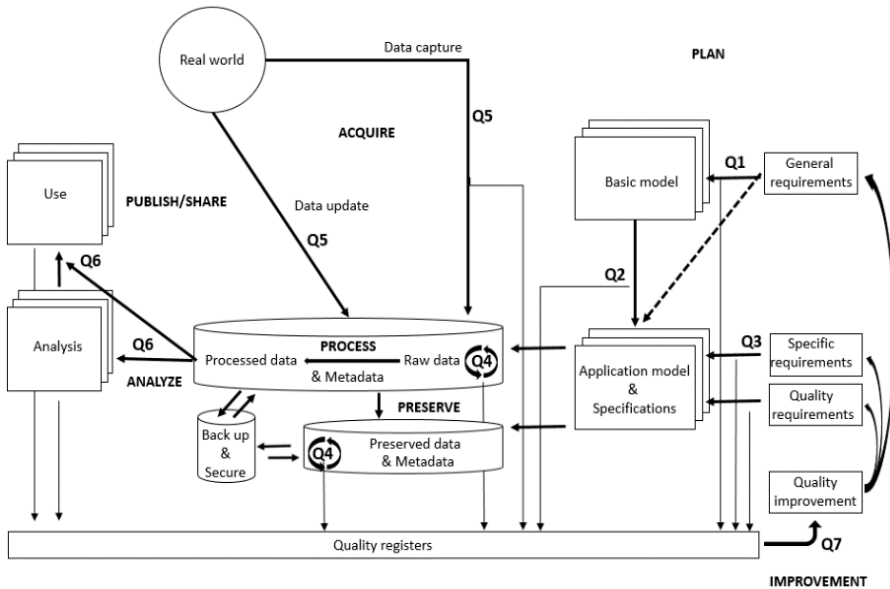- Q7: Assessment of the product continuous improvement process.



**Figure 2.**    Example model for quality throughout the cifecycle.

### *Unify UncertML and ISO 19157 to improve Annex D.*

We identified some conceptual overlapping in the list of quality measures presented in annex D. This is due to the use of similar statistical concepts applied to different quality elements but using the same measure from the statistical point of view. A simplification in the way the list of measures in Annex D is presented is necessary in order to increase comprehension. We propose a way to achieve this without losing anything, using already existing alternatives: UncertML 2.0[13] and QualityML.[14]

UncertML 2.0 provides a semantic description of statistics that can be used to compute uncertainties. A priori, it seemed simple to associate those to ISO 19157 quality measures but in practice it was not. By extending the concepts in UncertML

---

[13]  https://web.archive.org/web/20161029215725/http://www.uncertml.org/
[14]  http://www.qualityml.org/

a strategy to better normalize measures of ISO 19157 by making the list of measures more compact was achieved.

In essence, the concept of statistics is extended to include other quality metrics used to compute the result of each quality measure value when applied to a certain domain. QualityML provides a matrix of the combinations of indicators, measurements, domains and metrics most commonly used. The main idea behind this structure is to unlink measures, domains and metrics descriptions in order to maximize the generalization of descriptions and increase coherence among several measures using the same metrics (even with different domains), or several quality indicators using the same measures.

In fact, ISO 19157 already introduces the concept of data quality basic measure to avoid the repetitive definition of the same concept. There are data quality measures that have certain commonalities. Two principle categories of *data quality basic measures* are listed in the Annex: i) the uncertainty-related data quality basic measures (e.g. the LE50 basic measure used in linear error probability [id 33], time accuracy at 50% significance level [id. 55] or attribute value uncertainty at 50% significance level [id. 69]), and ii) the counting-related data quality basic measures which are based on the concept of counting errors or correct items.

QualityML goes one step beyond this generalization effort in the ISO 19157 basic measures, and groups one describing the same metric but with different parameters. For example, all the measures regarding "half length of the interval" are grouped in a single general metric called Half-lengthConfidenceInterval, which includes a parameter to describe the confidence level (or probability) of the true value being between the lower and the upper limit. Level has to be in the range [0,1]. This QualityML metric (Half-lengthConfidenceInterval) includes several ISO 19157 basic measures such as LE50 (and thus measures with id. 33, 55 and 69) but also LE68.3 (used in standard linear error, id. 34), time accuracy at 68,3% significance level (id. 54) and attribute value uncertainty at 68.3% significance level (id. 68), LE90, LE95 etc. All these ISO 19157 measures are grouped in a single QualityML with a single parameter "level" to identify the significance level. The advantage of this generalization is not only the increase of coherence in the quality measures and metrics description, but also the possibility of describing any other confidence level interval in a standardized way.

This is done in QualityML not only for uncertainty-related data quality basic measures describing one dimensional random variables (Z, using "Half-length Confidence Interval", examples above) but also for uncertainty-related data quality basic measures describing two dimensional variables and for counting-related data quality basic measures. More details about this approach can be found in Zabala & Maso (2016).

*User feedback in the broader sense.*

Gray knowledge about the data can be as useful as quality indicators to assess "fit for purpose". A structured way to include experiences of users of products that report applications not initially foreseen as well as problems and workarounds should be included in this IS (or in another document related to it). Users can produce valuable and complementary metadata about resources structured in feedback about each resource they are interested in, have used, etc. There are plenty of elements that can be included in a feedback item about a resource (or a group of them) such as ratings, comments, usage, related publications, additional lineage steps, quality elements or significant events description. All these user feedback metadata elements complement producer metadata and add value to the dataset descriptions. It also helps increasing users' engagement as they can see a real opportunity to create a community and establish social links on a geospatial portal around the datasets they are interested in. The data producers may also take advantage of this situation, being able to respond to users' demands in creating new versions of the resources or answering their concerns as new feedback items (related to the previous ones).

The work developed on previous projects such as GeoViQua,[15] CHARMe[16] and Melodies[17] and on the OGC Geospatial User Feedback (GUF) working group, led to the approval of the OGC Geospatial User Feedback Conceptual Model Standard (Masó & Bastin, 2016) using the ISO schemas as a baseline. Geospatial User Feedback is metadata that is predominantly produced by the consumers of geospatial data products as they use and gain experience with those products. This standard complements existing metadata conventions whereby documents recording dataset characteristics and production workflows are generated by the creator, publisher or curator of a data product. As a part of metadata, the GUF data model reuses some elements of ISO 19115-1:2014 but not the general structure. This selective use of ISO metadata elements was intended to prioritize future interoperability with developing ISO metadata models, and would allow an easy integration into the new version of ISO 19157.

**Conclusions**

Geospatial data are relevant for the decision making of many daily activities and large investments, and therefore the quality of this data is important. Thus, ISO 19157 is a significat IS in the domain of geospatial data. ISO 19157 is currently in a revision process and therefore it is desirable that all interested parties are aware of this and

[15]  http://www.geoviqua.org/
[16]  https://www.copernicus.eu/en/characterisation-metadata-enable-high-quality-climate-applications-and-services
[17]  https://www.the-iea.org/projects/melodies-research-project/

their ability to propose improvements. This document presents the contributions from a Spanish experts group.

In our view, ISO 19157 is a standard that has adequately fulfilled its function as a model to define, quantify and report quality, although its application has not been as widespread as desirable and when applied, has not been without problems. Many of the application problems are not particular of ISO 19157, they are typical of all standards in general, and this is a situation that requires a specific analysis that is outside the scope of this study.

Since technology and data availability have changed tremendously in recent years, we consider this revision of ISO 19157 appropriate. The challenges presented by the new types of available data as well as certain parts of traditional production processes that were not adequately covered with the ISO 19157:2013 version makes this revision essential if we want the proposed quality model that is continuously being applied. We need a model that ensures the highest degree of interoperability in the definition (conceptualization), quantification and reporting of data quality. For this reason the convergence with other standards and the inclusion of new dimensions of quality and new quality elements are aspects that we consider critical.

Additionally, the experience in the application of the model offers us a set of very specific guidelines focused on the field of spatial data production, such that they can make the application of the IS better understood, more efficient, powerful and versatile.

Therefore, we consider that the review is a good opportunity to improve this IS, otherwise the new types of data will bring other proposals for data quality that will surely require more additional efforts. There is no other option than to evolve or perish.

We are convinced that quality aspects are going to be more and more important in the near future for the proper use of geographic data and correct decision-making based on them in order to face the global challenges threatening the survival and well-being of humankind in the long term, as expressed in the 17 Sustainable Development Goals defined by the United Nations.

## References

αlphaβeta (2017). "The economic impact of geospatial services: how consumers, businesses and society benefit from location-based information". https://www.alphabeta.com/wp-content/uploads/2017/09/GeoSpatial-Report_Sept-2017.pdf.

AdV (2002). Erarbeitungeines Qualitätssicherungssystems für die Geodaten des amtlichen Vermessungswesens. Grundsätze für Qualitätskriterien und standard-isierte Prüfverfahren für die Anwendung des AFIS-ALKIS-ATKIS-

Basisschemas bei der Entwicklung der Anwendungsschemata. Arbeitsgemein-schaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutsch-land.

Ahmed, B.S.; Bures, M., Frajtak, K.; Cerny, T. (2019). "Aspects of Quality in Inter-net of Things (IoT) Solutions: A Systematic Mapping Study", IEEE Access, vol. 7, pp. 13758-13780. https://doi: 10.1109/ACCESS.2019.2893493.

Ariza-López, F.J.; Rodríguez-Pascual, A.F. (2018). "Calidad en datos geográficos, geoservicios y productores de datos: análisis crítico", *Revista Cartográfica*, núm. 97, pp: 105-143. https://doi.org/10.35424/rcar.v0i97.178.

Ariza-López F.J. (ed.) (2013). *Fundamentos de evaluación de la calidad de la infor-mación geográfica*, Servicio de Publicaciones de la Universidad de Jaén.

Ariza-López, F.J.; Rodríguez-Avi, J.; Reinoso-Gordo, J.F.; Ariza-López, Í.A. (2019). "Quality Control of "As Built" BIM Datasets Using the ISO 19157 Framework and a Multiple Hypothesis Testing Method Based on Proportions", *ISPRS Int. J. Geo-Inf.*, 8, 569.

Burrough, P.A. (1986). "Principles of Geographic Information Systems for Land Re-source Assessment", *Monographs on Soil and Resources Survey* No. 12, Oxford Science Publications, New York.

Castro, L., Porto, J.; dos Santos, R.; Zipf, A. (2017). A taxonomy of quality assess-ment methods for volunteered and crowdsourced geographic information. *Tran-sactions in GIS*, 22(2): 542-560. https://doi.org/10.1111/tgis.12329.

Degrossi, L.C., Porto de Albuquerque, J., Dos Santos Rocha, R., & Zipf, A. (2018). "A taxonomy of quality assessment methods for volunteered and crowdsourced geographic information", *Transactions in GIS: TG*, 22(2): 542-560. https://doi.org/10.1111/tgis.12329.

Eldawy, A.; Mokbel, M.F. (2013). *The Era of Big Spatial Data: A Survey*, Founda-tions and Trends in Databases, 6(3-4): 163-273, Hanover, USA, Now Publishers Inc.

EU (2020). "Building a European data economy". https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy

Eurostat (2020). "Population grids". https://bit.ly/2TwlPc7.

Gartner (2020). https://www.gartner.com/en/information-technology/glossary/big-data

GeoKnow (2012). "Metrics for Linked Geospatial Information. Deliverable 3.4.1.GeoKnow - Making the Web an Exploratory for Geospatial Knowledge".

Goodchild, M. (2007). "Citizens as Sensors: The World of Volunteered Geography", *GeoJournal*, 69: 211-221. https://doi.org/10.1007/s10708-007-9111-y.

Guptill, S.C.; Morrison, J.L. (1995). *Elements of Spatial Data Quality*, Pergamon Press.

Gusminia, M.; Jabeurb, N.; Karama, R.; Melchioria, M.; Renso, C. (2017). "Reputation evaluation of georeferenced data for crowd-sensed applications", *The 8th International Conference on Ambient Systems, Networks and Technologies* (ANT 2017), May 16-19, Madeira, Portugal.

IECA (2011). NTCA 01-002. "Modelo Aseguramiento Calidad productos de IG en Andalucía". Recuperado de https://www.juntadeandalucia.es/institutodeestadisticaycartografia/ieagen/sea/ntca/01_modelos/01002_Modelos_Aseguramiento_Calidad.pdf

ISO (2002). "Geographic information - Quality principles"

ISO (2003). ISO 19114:2003 Geographic information - Quality evaluation procedures

ISO (2006). ISO/TS 19138:2006 Geographic information - Data quality measures

ISO (2013). ISO 19157:2013 Geographic information - Data quality

ISO (2014). ISO 19115-1:2014 Geographic information - Metadata - Part 1: Fundamentals

Karkouch, A.; Mousannif, H.; Al Moatassimeand, H.; Noel, T. (2016). "Data Quality in Internet of Things: A state-of-the-art survey", *Journal of Network and Computer Applications*, 73: 57-81. http://dx.doi.org/10.1016/j.jnca.2016.08.002.

Lee, J.-G., Kang, M. (2015). Geospatial Big Data: Challenges and Opportunities. *Big Data Research*, 2(2): 74-81. https://doi.org/10.1016/j.bdr.2015.01.003.

Li, S., Dragicevic, S., Anton, F., Sester, M., Winter, S., Coltekin, A.; Cheng, T. (2016). "Geospatial Big Data Handling Theory and Methods: A Review and Research Challenges", *ISPRS Journal of Photogrammetry and Remote Sensing*, 115: 119-133. https://doi.org/10.1016/j.isprsjprs.2015.10.012.

López-Pellicer, F.; Vilches-Blazquez, L.; Zarazaga, F.J.; Muro-Medrano, P.; Corcho, O. (2011). "The Delft Report: Linked Data and the challenges for geographic information standardization", in *Jornadas Ibéricas de Infraestructuras de Datos Espaciales 2011* (JIIDE 2011), November 9-11, 2011, Barcelona, Spain.

Masó, J.; Bastin, L. (2016). *OGC Geospatial User Feedback Standard. Conceptual model*. OGC 15-097r1. In Internet: http://docs.opengeospatial.org/is/15-097r1/15-097r1.html.

Moellering, H. (1987). A draft proposed standard for digital cartographic data. National Committee for Digital Cartographic Data Standards. https://doi.org/10.3133/ofr87308.

Moström, J.; Hedeklint, K.; Dysterud, M.; Engelien, E.; Bloch, V.; Tammisto, R., Lillethun, A. (2019). "GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe". Proposal from the GEOSTAT 3 project. https://bit.ly/2Ts8PEG.

NIST (1994). *Spatial Data Transfer Standard (SDTS). Federal Information Processing Standards Publication 173-1*, National Institute of Standards and Technology.

Padilla, M., Stehman, S.V., Ramo, R., Corti, D., Hantson, S., Oliva, P., Chuvieco, E. (2015). "Comparing the Accuracies of Remote Sensing Global Burned Area Products using Stratified Random Sampling and Estimation", *Remote Sensing of Environment*, 160: 114-121. https://doi.org/10.1016/j.rse.2015.01.005

PIRA (2000). Commercial exploitation of Europe's public sector information, Final Report, Pira International, European Commission, Directorate General for the Information Society.

Puyan, A.; Zadeh, G.W.; Cavka, H.B.; Staub-French, S.; Pottinger, R. (2017). "Information Quality Assessment for Facility Management", *Advanced Engineering Informatics*, 33: 181-205. https://doi.org/10.1016/j.aei.2017.06.003.

Robinson, A.C.; Demšar, U.; Moore, A.B.; Buckley, A. Jiang, B.; Field, K., Sluter, C.R. (2017). "Geospatial big data and cartography: research challenges and opportunities for making maps that matter", *International Journal of Cartography*, 3 (sup1): 32-60. https://doi.org/10.1080/23729333.2016.1278151.

SDMX (2009). List of subject-matter domains.

Senaratne, H.; Mobasheri, A.; Loai Ali, A.; Capineri, C.; Haklay, M. (2017). "A review of volunteered geographic information quality assessment methods". *International Journal of Geographical Information Science*, 31(1). 139-167 https://dx.doi.org/10.1080/13658816.2016.1189556

Song, Y.; Wang, X.; Tan, Y.; Wu, P.; Sutrisna, M.; Cheng, J.C.P.; Hampson, K. (2017). "Trends and Opportunities of BIM-GIS Integration in the Architecture, Engineering and Construction Industry: A Review from a Spatio-Temporal Statistical Perspective", *ISPRS International Journal of Geo-Information*, 6(12): 397. https://doi.org/10.3390/ijgi6120397.

Sun, J.; Harrie, L.; Jensen, A.; Eriksson, H.; Tarandi, V.; Uggla, G. (2018). *Description of geodata quality with focus on integration of BIM-data and geodata*. Smart Built Environment. Available online in: https://bit.ly/3clKusH.

UNE (2016). UNE 148002:2016 *Metodología de evaluación de la exactitud posicional de la información geográfica*.

UN-EAPD (2019). "Data Economy: Radical transformation or dystopia?", *Frontier Technology Quarterly*, january 2019. https://bit.ly/2TcnmVQ.

UN-ISGI (2018). *Global Statistical Geospatial Framework: Linking Statistics and Place*. United Nations Expert Group on the Integration of Statistical and Geospatial Information.

UN-SD (2019). *United Nations National Quality Assurance Frameworks Manual for Official Statistics* (UN NQAF Manual). UN Statistics Division

Ureña-Cámara, M.A.; Nogueras-Iso, J.; Lacasta, J.; Ariza-López, F.J. (2019). "A method for checking the quality of geographic metadata based on ISO 19157", *International Journal of Geographical Information Science*, 33(1): 1-27. https://doi.org/10.1080/13658816.2018.1515437

Wikipedia (2020a). "Data economy". https://en.wikipedia.org/wiki/Data_economy

Wikipedia (2020b). "Internet of things". https://en.wikipedia.org/wiki/Internet_of_things

Wikipedia (2020c). "Linked data". https://en.wikipedia.org/wiki/Linked_data

Zabala, A.; Masó, J. (2016). *Testbed-12 Imagery Quality and Accuracy Engineering Report*. OGC 16-050, in Internet: http://docs.opengeospatial.org/per/16-050.html.

Zabala, A.; Riverola, A.; Serral, I.; Díaz, P.; Lush, V.; Masó, J.; Pons, X.; Habermann T. (2013). Rubric-Q: Adding Quality-related Elements to the GEOSS Clearinghouse Datasets. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (IEEE J-STARS), 6(3): 1676-1687. https://doi.org/10.1109/jstars.2013.2259580.

Zaveri, A.; Rula, A.; Maurino, A.; Pietrobon, R.; Lehmann, J.; Auer, S. (2014). "Quality Assessment Methodologies for Linked Data: A Survey", *Semantic Web Journal*. http://www.semantic-web-journal.net/content/quality-assessment-linked-data-survey.