HERITAGE and TECHNOLOGY
MIND KNOWLEDGE EXPERIENCE

XIII International Forum Le Vie dei Mercanti

Aversa, Capri 11-12-13 June 2015

DATA QUALITY ELEMENTS FOR BIM APPLIED TO HERITAGE MONUMENTS

Íñigo ARIZA-LÓPEZ[1], Francisco Javier ARIZA-LÓPEZ[2], Juan Francisco REINOSO-GORDO[3], Antonio GÓMEZ-BLANCO[4], Concepción RODRÍGUEZ-MORENO[5], Carlos LEÓN-ROBLES[6]

(1) Architectural Constructions Dpt., School of Architecture, University of Seville, Seville, Spain.
(3) Cartographic, Geodetic and Photogrammetry Engineering Dpt. Polytechnic School, University of Jaén, Granada, Granada, Spain.
(3), (4), (5), (6) Architectural and Engineering Graphic Expression Dpt., School of Architecture, Civil Engineering School, University of Granada, Granada, Spain.
E-mail: (1)inigoariza@us.es, (2)fjariza@ujaen.es, (3)jreinoso@ugr.es, (4)agomezb@ugr.es, (5) conrodriguezmoreno@gmail.com, (6)cleon@ugr.es,

Abstract

The Building Information Models are digital data sets organized as data models that represent physical and functional characteristics of places. BIM are used for decision making and management of infrastructures and for this reason data quality is a key factor.
In this paper we analyze the possible data quality elements to be considered in BIM applications and we propose a set of such data quality elements of specific interest for BIM applied to heritage monuments.
For this analysis we review some existing international standards related with the quality of data. In this way we analyze the international standard [1] (Geographic information. Data quality) and some of the Technical Specifications of ISO/TS 8000 (Data quality) (e.g. parts 120, 130, etc.).

1. Introduction

In architecture is currently taking place the BIM revolution. As [9] indicates this is a major mental transition when one comes from the CAD world. The main difference highlighted by various authors regarding CAD is the perspective provided by the I:<<BIM is building information modeling, with emphasis on the "I", which is what differentiates the both technologies. CAD can be defined as both computer aided design and computer aided drafting, but both cases emphasize the "computer aided" aspect of the proposition. BIM can be defined as both building information modeling and building information management, and, similarly the building information is emphasized. The real value proposition of building information modeling lies in the ability to analyze data.>>
However, this transition is not so when one comes from the GIS world. GIS have always been tools of analysis. GIS are tools to analyze the real world through modeling with spatial data (geographic information). The reality that is modeled and analyzed in a GIS is usually much more complex (e.g. territorial aspects, climatic aspect hydraulic aspects, etc.) than that reality modeled and managed in a BIM.
This approach of the construction management tools (BIM) towards spatial management tools (GIS) also means an approach in many other areas, including the data handling and management. Thus, since GIS is being applied from several decades, it is logical that experiences of the GIS field can be transferred to the BIM field.
This paper proposed a perspective for evaluating the quality of BIM data. This perspective is an adaptation of the experiences that exist in the field of GI regarding the quality of spatial data.
The backbone of BIM is the understanding of materials, objects, assemblies and projects [11]. All of them are managed by BIM as objects, in a programming sense. That means that materials, objects, assemblies and projects have properties, methods, events and relations. In a BIM objects carry information about: identity, appearance, performance, operation, usage, age, location, components, restrictions or rules, etc. Some of these properties actuate as base parameters (e.g. length, high and wide) for a parametric creation of a specific instance of an object type. All

this information is managed by the BIM tool as a database; actually a BIM tool is little more than a database management system with a graphic user interface [11].
The materials (concrete, wood, steel, glass, etc.) are the basic elements of construction. Have physical, chemical, visual, tactile, etc. properties, which are fully or partially transferred to the objects that compose aggregation.

The objects and assemblies differ in how they are created, implemented and managed in a project, but really are very similar. Objects are separate components, such as a door, a window; and assemblies are groups of objects that work together to create a more complex component at a given level of detail can be managed as an object. Examples of component are: walls, floors, roofs, etc. But a window (frame + glass + blind) can also be considered as an assembly, and this depends on the level of detail with which we work.
The projects are the BIM models. It would be better to use "project" to refer to how models are managed in a software package and use "model" to refer strictly to the assembly of assemblies that make sense buildings.
Each of these elements (materials, objects, assembled and projects) requires data (attributes) to specify it. For example, for an object or assembly:

- Identification: which product is
- Process: information on the lineage
- Performance: what is good, efficient, etc.?
- Limitations and rules: indication of constraints or rules.
- Installation: how is it installed?
- Appearance: what it looks like?
- Lifecycle / sustainability: how maintenance is done?
- Etc.

Moreover, the elements of a BIM own behavior and some associations that can be defined in the base model or applied specifically on a project in which they are indicated as specifications.
Everything mentioned so far has an overall resemblance to what happens in the field of spatial data, and therefore in this paper we analyze the applicability of the aspects of quality that are considered in the case of geographic information on the case BIM. In this sense the quality of the data from the perspective of the international standard [1] will be presented.
Moreover, much of the information (e.g. windows, steel sections, pipes, electrical elements, motors, etc.) incorporated into a BIM data can be considered as product data: a representation of information about a product in a suitable formal manner for communication, interpretation, or processing by human beings or by computers. With product data we mean information, including, among others, computer-aided design, data, models, parts information, manufacturing instructions, requirements, notes, documents, etc. In this sense the quality of the data from the perspective of the international standard [2] will be presented.
The proposal of a general model or framework for the data quality,(independent of any specific domain, means:

- Definition of data quality
- Dimensions /categories of data quality
- Attributes in each dimension
- How to measure these attributes
- How to control and improve data quality

But our proposal is related only with the definitions of the dimensions or categories of data quality.
This paper is organized as follows, first introduced in [1]. We present the definition of each of the components of the quality of the IG is performed. [2] will be presented later, its quality scheme and definitions. From the above discussion it will be decided whether to use these components in the case of BIM. The paper ends with some conclusions.

2. Data quality from the perspective of "[1]: Geographic information. Data quality"

The International Standard [1] establishes the principles for describing the quality of geographic data. It

- defines components for describing data quality;
- specifies components and content structure of a register for data quality measures;
- describes general procedures for evaluating the quality of geographic data;
- establishes principles for reporting data quality.

The conceptual model proposed by the standard is shown in Fig. 1. This embraces the four actions that have been indicated above. The key element is the data quality that must be reported and related to specific scope. For this, data quality is expressed by elements of quality described by measures, quality evaluation process, results and metaquality.

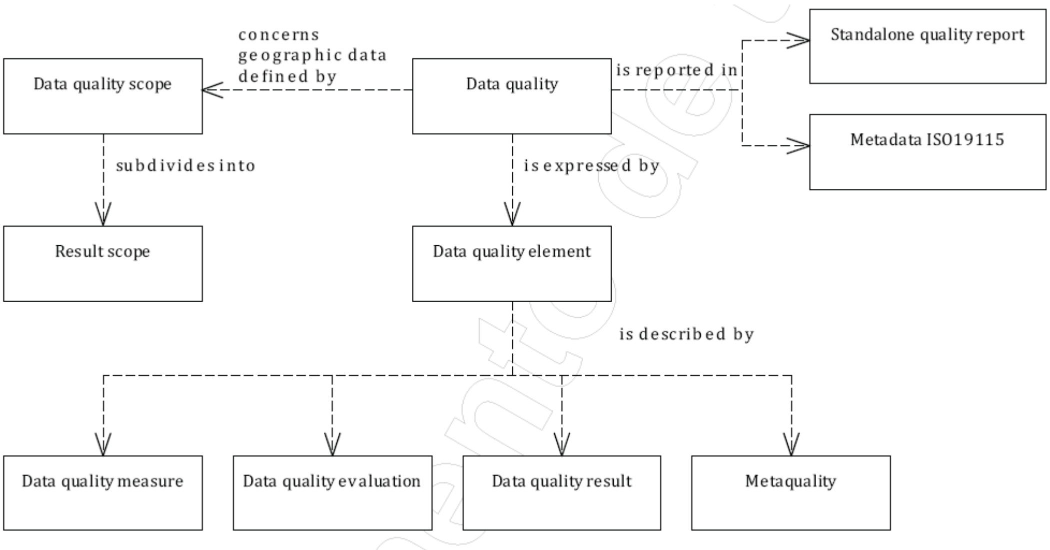

Fig. 1: Conceptual model of [1] (source: [1])

For [1] a data quality element is a component describing a certain aspect of the quality of geographic data. The elements have been organized into different categories. The Fig. 2 shows an UML diagram with categories and relations. Here follows a brief description of such categories and elements:

- Completeness. Completeness is defined as the presence and absence of features, their attributes and relationships. It consists of two data quality elements:
  - Commission. Excess data present in a dataset.
  - Omission. Data absent from a dataset.
- Logical consistency. Logical consistency is defined as the degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical). If these logical rules are documented elsewhere (for example in a data product specification) then the source should be referenced (for example in the data quality evaluation). It consists of four data quality elements:
  - Conceptual consistency. Adherence to rules of the conceptual schema.
  - Domain consistency. Adherence of values to the value domains.
  - Format consistency. Degree to which data is stored in accordance with the physical structure of the dataset.
  - Topological consistency. Correctness of the explicitly encoded topological characteristics of a dataset.
- Positional accuracy. Positional accuracy is defined as the accuracy of the position of features within a spatial reference system. It consists of three data quality elements:
  - Absolute or external accuracy. Closeness of reported coordinate values to values accepted as or being true.
  - Relative or internal accuracy. Closeness of the relative, positions of features in a dataset to their respective relative, positions accepted as or being true.
  - Gridded data position accuracy. Closeness of gridded data spatial position values to values accepted as or being true.
- Thematic accuracy. Thematic accuracy is defined as the accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships. It consists of three data quality elements:
  - Classification correctness. Comparison of the classes assigned to features or their attributes to a universe of discourse (e.g. ground truth or reference data).

- Non-quantitative attribute correctness. Measure of whether a non-quantitative attribute is correct or incorrect.
- Quantitative attribute accuracy. Closeness of the value of a quantitative attribute to a value accepted as or known to be true.

- Temporal quality. Temporal quality is defined as the quality of the temporal attributes and temporal relationships of features. It consists of three data quality elements:
  - Accuracy of a time measurement. Closeness of reported time measurements to values accepted as or known to be true.
  - Temporal consistency. Correctness of the order of events.
  - Temporal validity. Validity of data with respect to time.
- Usability element. Usability is based on user requirements. All quality elements may be used to evaluate usability. Usability evaluation may be based on specific user requirements that cannot be described using the quality elements described above.

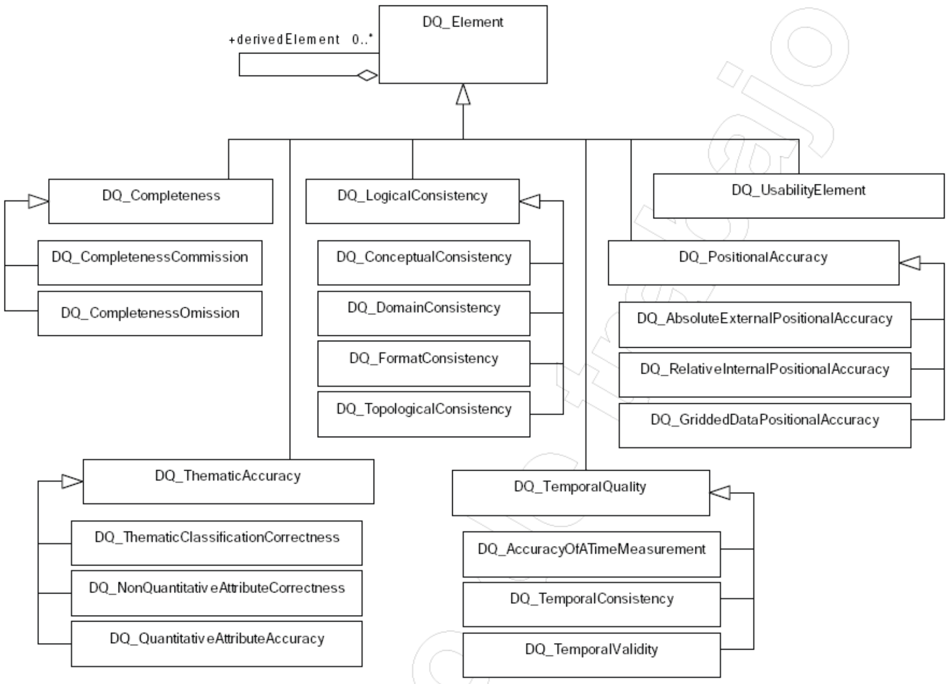

Fig. 2: Data quality elements in [1] (Source [1])

3. Data quality from the perspective of "ISO/TS 8000-1:2011. Data quality"

[2] is focused on the ability to create, collect, store, maintain, transfer, process and present data to support business processes in a timely and cost effective manner. This purpose requires both an understanding of the characteristics of the data that determine its quality, and the ability to measure, manage and report on data quality. [2] covers industrial data quality characteristics throughout the product life cycle from conception to disposal. [2] addresses specific kinds of data including, but not limited to, master data, transaction data, and product data (Fig. 3). [2] is organized as a series of parts:

- parts 1 to 99: General data quality [3];
- parts 100 to 199: Master data quality [7];
- parts 200 to 299: Transaction data quality[4], [5], [6];
- parts 300 to 399: Product data quality [8].

Each of the above series addresses communication within an organization and between two or more organizations.

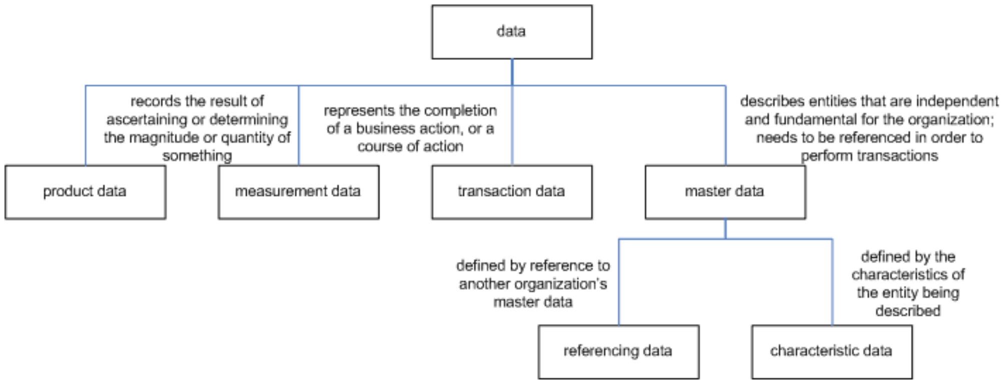

Fig. 3: General vision of the scope of [2] (Source [2])

Fig. 4 shows a general view of the data architecture proposed by [2]. This architecture is applicable to the scope of the parts of [2] that have been published until now. The relations stablished between data and other elements (data dictionary, data specifications, formal syntax and identification scheme) also can be considered valid for GI. From the point of view of our interest, the focus must be applied to the proposed accompanying quality elements for data: data provenance, data accuracy, and data completeness.

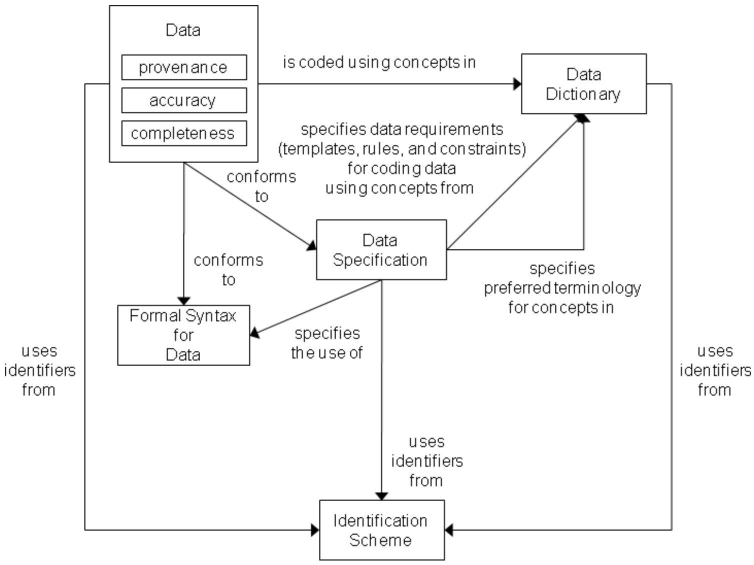

Fig. 4: general view of the data architecture proposed by [2] (Source [2])

The definitions offered by this rule are:

- Data accuracy: closeness of agreement between a property value and the true value.
- Data accuracy record: record of the information provided about the accuracy of a piece of data.
- Data completeness: quality of having all data that existed in the possession of the sender at time the data message was created.
- Data completeness record: record of the information provided about the completeness of a piece of data.

- Data provenance: This definition is not included in the rule. Provenance refers to the ability to trace and verify the creation of data, how it has been used or moved among different databases, as well as altered throughout its lifecycle.
- Data provenance record: record of the ultimate derivation and passage of a piece of data through its various owners or custodians. A data provenance record can include information about creation, update, transcription, abstraction, validation, and transferring ownership of data.

4. Discussion and proposal of Heritage BIM data quality elements

All the standards developed by ISO, and under development, focus on some specific domains, e.g. manufacturing, intelligent transportation, geography, and so on. There is still not a general model/framework of data quality independent of any particular domain or application and the existing domain models of data quality differ on dimensions, attributes, measurements, evaluations and so on. A justification is the great diversity of themes that work the very many technical committees of ISO, and the miscommunication and absence of partnerships between the technical committees that develop standards.

The two standards ([1] and [2]) that have been discussed have two different perspectives on data quality, and both are of interest in the BIM field. Besides presenting different perspectives, the definitions that they provide and how they manage these elements in their models is different. Also, there are other standards related to data quality (e.g. ISO / TR 21707-2008, ISO / IEC 25012: 2008), but the situation is similar. This circumstance is usual in ISO standards.

As stated previously, we are not going to propose a general framework for data quality, even a complete framework for the quality of BIM data in the heritage field. Our aim in this paper is more modest: we only propose a set of categories or dimensions of quality applicable to heritage BIM. This proposal focuses only on the intrinsic quality (accuracy, objectivity, believability, reputation). In later works these categories will be used to develop a UML model and to define measures, control methods and improvement, etc., but now it is out the scope of this paper.

The quality aspects are organized into elements and sub-elements of quality, here we use these terms with the same sense that were used in ISO 19113. The elements are logical groupings of quality sub-elements. The sub-elements relate to specific aspects of quality that can be measured and evaluated by means of different measures and methods.

The elements and sub-elements are:

- Provenance or lineage. This element appears in [2], as provenance, and in ISO 19113 (the predecessor of the [1]), as lineage. We believe that the inclusion is important because it provides information relevant to the credibility and reputation. For example, in the case of heritage BIM will allow authorship or featured processes. In the area of GI this element is treated by ISO 19115, and includes references to dated processes and sources. Both elements seem appropriate for consideration for heritage BIMs.
- Completeness. Completeness. Completeness appears in both [2] and [1]. The completeness's defects or errors are important when working with data, both when they occur in running BIM projects and when managing BIM related to existing buildings. In this case can be considered two sub-elements:
  - Omission. It is the absence of elements. This means that something is not on the BIM model but that really does exist in real world. In heritage BIM this may mean no consideration (omission) of any relevant element of the real world.
  - Commission. It is the presence of excess items. This means that something appears in the BIM model, but that does not exist in real world. An example in a heritage BIM project may be the inclusion of an item that really should not appear, the inclusion may involve the purchase of the good with the consequent extra cost.
- Metric accuracy. This element does not appear in [2] or [1]. However, the definition given to accuracy in [2] is fully applicable to this proposal. In this case we consider the following sub-elements:
  - Absolute positional accuracy. This sub-element does not appear in [2], since its scope is not geographical. But the position is the most significant aspect of the GI and therefore relevant in [1]. Under our consideration BIMs should be understood fully integrated with the GI and Geoservices (e.g. SDI services, virtual globes, etc.). This means that the positional accuracy, relative to a reference coordinate system and a projection system, is a critical requirement. For example, the absolute positional accuracy will be a requirement to adequately integrate a BIM model with its cadastral parcel, and place it properly on geographic viewers and virtual globes.
  - Relative positional accuracy. This sub-element does not appear in [2] since its scope is not geographical. But the position is the most significant aspect of the IG and therefore relevant in [1]. This sub-element means that heritage BIMs should pick accurately the relative positions of objects of the real world. For example, the distance between a door D and a window W must be

accurate, or the distance between wall M1 and another wall M2 must have an appropriate degree of accuracy.
  - Shape accuracy. This sub-element is directly related to the proposal of accuracy that is done in [2]. Moreover, in some applications of ISO 19113 (the [1] predecessor) it was defined the geometric fidelity sub-element, and it approximates well what we propose here. This sub-element is proposed to consider all geometrical aspects relating to the element itself, as opposed to the positional relationships between the element and its environment (e.g. absolute or relative positional accuracy). Shape accuracy includes, among others, manufacturing tolerances. Thus, depending on the aspect (roughness, roundness, height, width, etc.) we can define different measures
- Thematic accuracy. This element does not appear, as such, in [2], although interpretations of accuracy can cover this field properly. The thematic accuracy does appear in [1]. We propose to include this element in order to incorporate all aspects of the accuracy that are thematic type (not metric aspects). The following sub-elements are proposed:
  - Classification correctness. It refers to the correct classification of elements by its assigned categories.
  - Non-quantitative attributes correctness. It refers to the correctness of the registered values of object attributes.
  - Accuracy of quantitative attributes. Objects may have quantitative attributes, for example, a heat or light transmission values. This sub-element means that values are accurately registered.
- Temporal accuracy. This sub-element does not appear, as such, in [2], although interpretations of accuracy can cover this field properly. The temporal accuracy does appear in [1]. We consider the temporal aspect in BIMs is very important, although not much is highlighted in the BIMs manual. In the case of BIM applied to heritage this aspect is particularly relevant.
  - Date accuracy. It means that the dates are accurately recorded.
- Logical consistency. This sub-element does not appear, as such, in [2], although interpretations of accuracy can cover this field properly. Also, the proposed architecture by [2] for the quality does consider. Logical consistency does appear in [1]. The following sub-elements are proposed:
  - Format consistency. It means that the records are coded following the rules of the formats set in the specifications.
  - Conceptual Consistency. It refers to that elements recorded in a BIM must follow some conceptual rules that are proper of a construction (e.g. constructions rules).
  - Domain Consistency. It refers to that recorded values for any given characteristic (e.g. an attribute, category, etc.) must be within the logical domain (e.g. interval for numerical values, or an enumerated list for classes) of values for that feature. For example, an enumerated list may be relative to the material forming a gate: {wood, PVC, iron, glass}. In this case you can only use one of the items on the list; the use of any other material is a consistency error.

5. Conclusions

This work has focused on the quality of BIM data. The data that make up a BIM have a great similarity with spatial data managed by GIS, and also with product data managed by product data management systems. For this reason we have presented both, the [1] and [2] standards, covering each of the two perspectives mentioned above. It is observed that in the ISO organization there is no a general or common framework for defining the quality of data, in each application domain standards are proposed.

This paper has made a proposal of quality elements and sub-elements for heritage BIMs. This is a very basic proposal; it is not exhaustive, so that in the future it is possible to add more elements and sub-elements. This design is the first part of a more comprehensive design that should include measures, assessment methods and reporting, but these issues are beyond the scope of this work. The proposal combines aspects from the [1] and [2] standards, with the perspective of intrinsic quality. For each of the proposed elements and sub-elements it was performed a justification and included an example.

Suggestions welcome, please send an e-mail to the corresponding author.

6. Acknowledgement

The authors would like to thank the ERDF of European Union for financial support via project “Calidad en modelos BIM (Building Information Models) aplicados al Patrimonio” of the “Programa Operativo FEDER de Andalucía 2007-2013”. We also thank to Public Works Agency and Regional Ministry of Public Works and Housing of the Regional Government of Andalusia. And we also thank to the TEP 164 group (Ingeniería Cartográfica) belonging to the PAIDI (Plan Andaluz de Investigación, Desarrollo e Investigación).

7. Referencias

[1] ISO 19157:2013. Geographic information. Data quality.

[2] ISO 8000-2:2012. Data quality. Part 2: Vocabulary.

[3] ISO/TS 8000-1:2011. Data quality. Part 1: Overview.

[4] ISO/TS 8000-120:2009. Data quality. Part 120: Master data: Exchange of characteristic data: Provenance.

[5] ISO/TS 8000-130:2009. Data quality. Part 130: Master data: Exchange of characteristic data: Accuracy.

[6] ISO/TS 8000-140:2009. Data quality. Part 140: Master data: Exchange of characteristic data: Completeness.

[7] ISO/TS 8000-150:2011. Data quality. Part 150: Master data: Quality management framework.

[8] ISO/TS 8000-311:2012. Data quality. Part 311: Guidance for the application of product data quality for shape.

[9] Weygant, R. BIM Content Development: Standards, Strategies and Best Practices. NJ: John Wiley and Sons, 2011.