Count based quality control of “As Built” BIM datasets using the ISO 19157 framework

From an informational point of view, a Building information model (BIM) is digital model based geometric information, enriched thematically, semantically and relationally that, managed by the right software tools, allows a smarter management of buildings and facilities. The corner stone of BIMs is to understand the relationships between materials, objects, assemblies and projects. All these elements are managed by a BIM tool as objects, in the sense of object-oriented programming. That means that materials, objects, assemblies and projects have properties, methods, events and relationships. In reality, a BIM tool is little more than a database management system with a graphical user interface. From this point of view, BIM models are directly linked to Geographic Information Systems (GIS) and BIM data to spatial data (geographic information).

Data quality of BIM datasets (BIMDS) is relevant and the BIM Comunity (www.bimcommunity.com) has developed a publication series which includes a guide centered on quality assurance of BIM projects (COBIM, 2012). This document proposes and develop several quality controls mainly devoted to check logical consistency issues and the use of software is proposed for examining clashes between building elements. Automatic routines for quality control of BIM has been proposed by Cheng (2018) and many others authors, also there are several software tools for this propose, e.g. iTWO by RIB (www.rib-software.co.uk), Solibri by Solibri (www.solibri.com); BIM Tree Manager by Agacad (www.aga-cad.com) or Verity by ClearEdge (www.clearedge3d.com). All these controls are based on aspects of logical consistency that, in most cases, can be automated.

Neither of the previously mentioned documents or tools develops or proposes a statistical method for a statistical quality control. Nor is there any mention of quality control standards from the industrial field (e.g. ISO 2851 or ISO 3851 series). The situation described above indicates the existence of several aspects that require research attention. One of them is that all aspects whose quality must be controlled in BIM datasets must be formalized, and another, that an appropriate method must be available so that the acceptance/rejection of BIM datasets is carried out on a statistical basis when a sampling is needed (e.g. as built perspective). In this work proposals are made in these two lines. Thus, our objective is to propose how to adequately formulate a quality control of BIM datasets and how to approach a statistical control.

BIM data quality and ISO 19157

BIM data are very similar to spatial data because they must be integrated into a geographical framework (the actual location of the building), integrated into its environment (the surrounding geographical-topographic reality), and collect the presence, dimensions, positions and exact attributes of the elements of interest. This resemblance is both conceptual (data models), and factual (e.g. capture and processing procedures), as well as exploitation (thematic, topological, temporal consultations, modeling, etc.). This proximity allows an advantageous approximation since in the field of geographic information there is a greater experience related to data quality. For instance, Sun et al. (2018) show the close links between spatial data and BIM data and review of the standards and methods currently used for ensuring quality in spatial data and BIM in Sweden (mainly), and internationally. For this reason, we adopt this international standard as the base for our proposal.

The International Standard ISO 19157 (ISO 2013) establishes the principles for describing the quality of spatial data. This is achieved by defining data quality elements, data quality measures, a general procedure for assessing and reporting data quality.

As a way of handling diverse perspectives of data quality, ISO 19157 proposes the so-called data quality elements (DQE) (e.g. absolute positional accuracy, relative positional accuracy, classification correctnes, etc.). A DQE relates to a specific aspect of data quality that can be measured and evaluated through different measures and methods. DQEs can be organized into categories which are logical groupings of DQE (e.g. DQEs related to logical consistency conform a category).

Before executing quality control, the population of elements of interest must be defined, and this is carried out by means of a scope. The scope is a filter based on time, location, classification, attributes… or, in general, in any other criteria that establish an element selection rule. The scope is usually defined by a category of elements of interest (e.g. windows, walls, pipes, etc.), but it can also be defined by a set of categories of elements of interest that share some aspect of common interest (e.g. windows and doors and walls, when our interest is the correction of the finish color). We call this set of categories of elements of interest the category of interest (CoI). The joint of a CoI with a DQE is known as data quality unit (DQU) in ISO 19157 terminology. So the same CoI can be linked to different DQE in order to control several perspectives of the data quality (e.g. those of all the DQE). Also, the same DQU can be assessed by means of different DQM (data quality measures) and by different evaluation methods (EM). ISO 19157 defines more than 70 standardized data quality measures (see Annex C of ISO 19157) but only a general EM. The last is not problematic because ISO 19157 allows the use of whatever evaluation method considered adequate for the assessment purpose, e.g. ISO 28590 (ISO 2017), ISO 3951 (ISO 2007), etc. Finally, quality control of a product is a statistical decision on the acceptance or rejection of a product with respect to its specifications, for this purpose a quality level (QL), or conformity level, must be established. This QL must be expressed in the same way and units as the DQM used for the DQE being considered. By this way, a quality control is well defined if a DQU (=DQE + Scope) and its corresponding QL (=DQM) and EM are properly stablished. These are the elements that must be managed to unequivocally establish quality control when using the ISO 19157 framework.

Count-based quality control

Products are defined by specifications, so that a nonconformity is the non-fulfilment of a specified requirement: e.g. a specification can be that 95% of the instances of a BIMDB must carry correct attributes in relation to reality. The presence of nonconforming/defective items is then quantified and a decision is made about the compatibility of this amount with respect to the conformity level. If sampling is required, e.g. in an "as built" BIM quality control, this decision must be taken in a statistical context in which the risks of the parties are controlled. The appropriate statistical tool for this is the hypothesis testing framework. Thus, adopting a hypothesis (distribution and value) on the behavior of the nonconforming items, by taking a sample (of a given sample size *n*), this statistical technique allows a decision to be taken where the producer's risk (Type I error), and the user's risk (Type II error) are bounded. The appropriate statistical models for working with proportions are: the binomial and hypergeometric models for working with one single class, in an infinite or finite population, respectively, and the multinomial and multivariate hypergeometric models for working with multiple classes, in an infinite or finite population, respectively.

Thus, the procedure is:

- Take an independent sample for each DQU.
- Count the number of nonconforming items found in the sample of each DQU.
- Calculate the corresponding p-values for each DQU.
- Check whether or not the global H0 hypothesis is accepted or rejected according to a MHTM correction.

EXAMPLE OF APPLICATION

As an example of the application of the proposed method, the case of a BIMDB control corresponding to the delivery of an ended project ("as built") will be considered. It is a building with 4 floors (basement, F0, F1 and F2); with garages in the basement, two commercial premises in F0 and 4 apartments distributed between F1 and F2, that is, two per each floor.

Table 1 Categories of interest in the BIMDB

Group	Categories of interest	Cases (N)		Group	Categories of interest	Cases (N)

Elements	C1=Doors and windows	119			C8=Slabs and paving	25
	C2=Bathrooms and Kitchens	14			C9=Pillars	105
	C3=Balconies and terraces	29			C10=Sales unit	6
	C4=Other rooms	18			C11= Interior walls	200
	C5=Living rooms and bedrooms	16		Facilities	C12=Electricity installation	7
	C6=Common zones	6			C13=Heating and air conditioned installations	7
	C7=Enclosures (walls)	179			Total	731

In relation to the DQU for the control, Table 2 summarizes their configuration, population and sample sizes. Sample sizes have been set arbitrarily with the criteria set forth above (≈10%), except for case C2, for which a size that assumes a proportion is adopted of the population.

Table 2 Definition of data quality units to be considered for the control (cases in the population and sample size) and the quality controls by means of the data quality units and the conformity levels

Data quality units	Cases in the population (N)	Sample size (n)	Quality control	Data Quality Measure and ID*	Conformity level (Maximum proportion of defects)
DQU1=Completeness of elements DQE = Commission + omission CoI = C1+C2+ ··· + C10	511	50	QC1	Rate of excess items (ID=3) + Rate of missing items (ID=7)	1%
DQU2=Completeness of facilities DQE = Commission + omission CoI = C11+ C13	182	40	QC2	Rate of excess items (ID=3) + Rate of missing items (ID=7)	3%
DQU3= Shape Fidelity DQE = Fidelity in shape CoI = C1+C2+ ··· + C10	1605	160	QC3	Rate of unfaithful items (ID=**)	5%
DQU4=Attributes of elements DQE = Correction of non-quantitative attributes CoI = C1+C2+ ··· + C10	462	50	QC4	Rate of incorrect attribute values (ID=67)	10%
DQU5=Attributes of installations DQE = Correction of non-quantitative attributes CoI = C12+ C13	491	50	QC5	Rate of incorrect attribute values (ID=67)	10%
DQU6= Shape Fidelity of walls DQE = Fidelity in shape CoI = C11	200	20	QC6	Rate of unfaithful items (ID=**)	80%, 15%,5%***
Total	**3451**	**350**			

(*) The ID is the identifier for this measure given in Annex D of ISO 19157.
(**) This measure is not included in Annex D of ISO 19157.
(***) This proportions are linked to good, acceptable and unacceptable cases.

Prior to the control and by agreement between the parties, QL must have been established. For this example, the specifications are those presented in Table 2. When indicating completeness, we refer to both omissions and commissions, considering both types of error as equivalent for error counting proposes. Finally, it should be noted that the QLs are themselves an order of the importance of the different aspects considered in the control. Naturally, these values must be determined based on the experience and the requirement of greater or lesser rigor for the BIM application. By this way, as indicated by Eq (4), the global control on the BIMDB means that: QC1 is passed AND QC2 is passed AND QC3 is passed AND QC4 is passed AND QC5 is passed AND QC6 is passed.

Defect case counts are computed (Table 3). From them, applying the functions (*pbinom* and *phyper*) (R Core Team, 2019), the p-values that appear in Table 3 are obtained. As can be seen, the hypergeometric model has been considered for the case QC2, in the rest of the cases the binomial model is applied. Here a MHTM is needed, and we apply Bonferroni by its simplicity. Since α = 5% was adopted, the global null hypothesis should be rejected if any p-value were less than 0.05 / 6 =

0.0083. Given that the lowest obtained p-value is 0.0004 <0.083, it is possible to reject the hypothesis that the BIMDB complies with the specifications imposed by Table 4 since the observed data give evidence of this.

Table 3 Results of the defective count and p-values by quality control

Quality control	Number of nonconforming items	Sample size (n)	p-value		
			Binomial	Hypergeometric	Multivariate Hypergeometric
QC1	0	50	1.000		
QC2	5	40		0.0004	
QC3	11	160	0.179		
QC4	5	50	0.569		
QC5	2	50	0.966		
QC6	7,1(*)	20			0.0236
(*) The number of items per class is: 12 (good), 7 (acceptable), 1 (unacceptable)					

CONCLUSIONS

The quality of BIMDB is an issue of great importance but, so far, it is not acquiring the appropriate relevance compared to the current boom of its applications. The quality of BIMDB is not fully formalized, but directly applicable knowledge can be transferred from the field of geospatial data. The framework established by ISO 19157 (ISO 2013) has already been proposed for its application to BIM data due to its great similarity with geographical information. This paper has presented the statistical basis of a method of global quality control of BIMDB with multiple DQUs, which means different scopes and diverse DQEs. The method has a valid, affordable and known statistical formulation as it is based on known distribution functions that are applied in the field of quality control. The main contributions of this work are two, first the proposal and example of use of ISO 19157 data quality framework to BIM data, and second the statistical approach formulation including an example of use on how to handle the joint control of several types of errors with different quality specifications for each of them.

Funding: This work has been supported by grant CMT2015-68276-R of the Spanish Ministry of Economy and Competitiveness.

References

Cheng, Y.M. (2018). Building Information Modeling for Quality Management. In Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS 2018) - Volume 2, pages 351-358

COBIM (2012). Series 6 Quality assurance. Common BIM Requirements.

ISO (2013). Geographic Information – Data Quality. International Standardization Organization.

ISO (2007). *ISO 3951-3:2007 Sampling procedures for inspection by variables — Part 3: Double sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection*. International Organization for Standardization.

ISO (2017). *ISO 28590:2017 Sampling procedures for inspection by attributes — Introduction to the ISO 2859 series of standards for sampling for inspection by attributes*. International Organization for Standardization.

R Development Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, R Foundation for Statistical Computing.

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.