BUILDING INFORMATION MODELS ARE DIRTY

Claudio Mirarchi, and Alberto Pavan
Politecnico di Milano, Milan, Italy

Abstract

Building Information Modelling (BIM) processes imply an intensive use of information. Nevertheless, several studies revealed critical issues in the data quality of information models. While some studies presented interesting works in the evaluation of model quality with reference to IFC. The analysis of the data quality issues in native models remains a research gap as well as the understanding of where these issues are generated. This research proposes an analysis of four information models to evaluate and classify data quality issues according to three dimensions, i.e. accuracy, coherence and completeness. Results highlighted user behaviours and/or technological limitations in real-world applications.

Introduction

Building information modelling (BIM) is transforming the way construction projects are developed, delivered and managed. Information models are characterised by the aggregation of both geometrical and non-geometrical information in composing objects (Eastman *et al.*, 2011). BIM implies an intensive use of information and its uses can be specified according to the needs of the processes defined e.g. in the project context and/or in the company context. This information represents a central asset in the construction process. Nevertheless, several aspects including technical and cultural ones can limit the use of information. In this context, interoperability, i.e. the process to exchange and use information between two or more systems (IEEE, 1990), can be identified as one of the main open issues.

Since 1995, the BuildingSmart consortium (formerly known as International Alliance for Interoperability until 2008) is working on the Industry Foundation Classes (IFC), a common data model to represent and describe building processes (Laakso and Kiviniemi, 2012). Moreover, starting from early 2000s researchers start to explore the use of semantic web technologies to enhance the information exchange processes (Pauwels, Zhang and Lee, 2017). Nevertheless, a common point in the existing studies is the need to start from a building information model developed through a specific BIM authoring tool. This is the case both in the use of IFC, where the IFC model is obtained translating the native model, and in the use

of semantic web technologies or graph databases. These translating processes can introduce informative issues in the resulting models as already broadly explored in the literate for IFC (Sacks *et al.*, 2010; Törmä, 2013; Hu *et al.*, 2016). Hence, the quality of the exchanged models and the processes used to obtain these models is of central importance. However, according to the well-known concept Garbage-in-Garbage-out (GIGO), also the quality of the input models needs to be assessed to guarantee the quality of the exchanged ones.

The increasing number of experimentations based on the use of information contained in models is progressively highlighting the importance of data quality in BIM. For example, Sacks *et al.*, (2017) underlined how processes of semantic enrichment of building information models are subject to the quality of the input data. In the same way the experiment proposed by Farias, Roxin and Nicolle (2018) showed wrong results in the output of the proposed algorithms due to data quality issues derived from the input models.

The broad area of data quality is well known in the field of database management and is now gaining increasing interest in the Architecture, Engineering, Construction and Operations (AECO) industry research field. While can be found in the literature some studies focused on the analysis of IFC models and related processes quality (Solihin, Eastman and Lee, 2015), few studies have been developed in the understanding of the possible issues that can be encountered in the native models. The existing studies in the IT field showed that it is extremely difficult (if possible at all) obtaining a completely error-free database. Hence, due to the inherent data quality issues that can be found in BIM processes, their evaluation plays a critical role also in the development of automated processes to information use (e.g. rule checking processes) and to propose structured (and automated) processes to verify models' quality.

According to Donato, Lo Turco and Bocconcino, (2018), the quality evaluation of BIM processes involves two main aspects, namely the quality of the processes and the quality of the models. Moreover, the quality of the models can be related to the quality of the entire models (including e.g. clash detection and models coordination) and to the quality of the objects

composing each model including the information contained in the objects and their coherence with the project (and information) requirements. The research proposed in this paper focuses on this last quality evaluation.

The scopes of the research are twofold. On the one hand, the study highlights the types of data quality issues that can be encountered in real world models according to three data quality dimensions, namely accuracy, consistency and completeness. On the other hand, the analysis focuses on the understanding of the quantity of data that can be generated in a building information model. The identification of possible critical data quality points in a structured analysis can help developers in the study of automated means to limit the highlighted issues and pave the way for future studies in this area. Hence, the proposed research is not focused on the measurement of data quality of information models but on the understanding of the possible issues that can affect the data quality of these models.

The rest of the paper is organised as follows. The background section introduces the concept of data quality and proposes a brief literature review of the existing studies related to data quality in BIM. The methodology section explains the processes followed in the development of the study and how the analysis was organised. The case study section presents the results of the analysis. The discussion section proposes a critical explanation of the evidences emerged from the study. Finally, the conclusion section reports the summary of the research and possible future works.

Background

Real world data is dirty (Hernández and Stolfo, 1998; Fan, 2012). In general dirty data means either missing data or wrong data or non-standard representations of the same data (Kim *et al.*, 2003). The management of data quality represents a fundamental area to guarantee the quality of analysis based on data. The definition of data quality proposed by Wang and Strong (1996), i.e. *"data that are fit for use by data consumers"*, underlines its complexity and dependency from the context. It is still relevant as highlighted in recent studies on data quality measurement (Mocnik *et al.*, 2018). Hence, the extent for which the quality of data needs to be evaluated depends on the context of use of the data. This evaluation can be based on more than 170 dimensions described among others by Wang and Strong (1996), Naumann and Rolker (2000), Delone and McLean (2003). Nevertheless, the most used objective dimensions are:

- Accuracy: i.e. the extent to which data are correct, reliable and certified.
- Consistency: i.e. the satisfaction of semantic rules defined over a set of data items.
- Completeness: i.e. the degree to which a

given data collection includes the data describing the corresponding set of real-world objects.
- Timeliness: i.e. the extent to which data are sufficiently up-to-date for a task.

In addition, a well-known issue in database analysis is the interpretation of *null values* (Zaniolo, 1984; Neumann, 2018). Empty cells or null values are difficult to interpret and can reduce the readability of a database. In fact, null values can be interpreted both as *unknown*, i.e. the value exist but it is not known or as *nonexistent*, i.e. the value does not exist. Moreover, there is no way to understand if the null value is related to one of those interpretations or it derives from a mistake in the development of the model.

According to the increasing need of understanding and evaluating the quality of BIM (both in terms of models and processes), in the literature can be found some recent studies in this direction. Solihin, Eastman and Lee (2015) presented a study analysing and detailing the concept of quantifiable IFC validation. The study was based on the understanding of the possible issues related to the export and import phases in the IFC process. The possibility of data quality issue in the native models is mentioned but not explored. Lee, Eastman and Solihin (2018) presented a set of product data exchange requirements, the rules of IFC data translation and exchange, the scenarios of each rule checking process, and the logic of their checking implementation creating the basis for the development of the IfcDoc (Building Smart, 2018) application that can be used to check IFC models against specific Model View Definitions (MVD). Zadeh *et al.* (2017) defined a framework for information quality assessment of building information models focusing on facility management needs. Donato, Lo Turco and Bocconcino, (2018) proposed a method for quality assessment of building information models through a structured and quantifiable BIM quality control based on a BIM quality assurance analysis. Moreover, focusing on the information flow, Tribelsky and Sacks (2010) proposed a process to measure and evaluate inefficiencies and issues. Kouhestani and Nik-Bakht (2018) proposed an automated process to collect and analyse event logs during design phases including process mining techniques to highlight procedural issues and optimise the development of the design process. Beside these studies, a systematic analysis of the user behaviour in generating models and/or of the technological limitations that can bring to data quality issue seems lacking in the literature. Hence, the proposed study focuses on these factors to improve the comprehension of the existing data quality issues that can derive from native models.
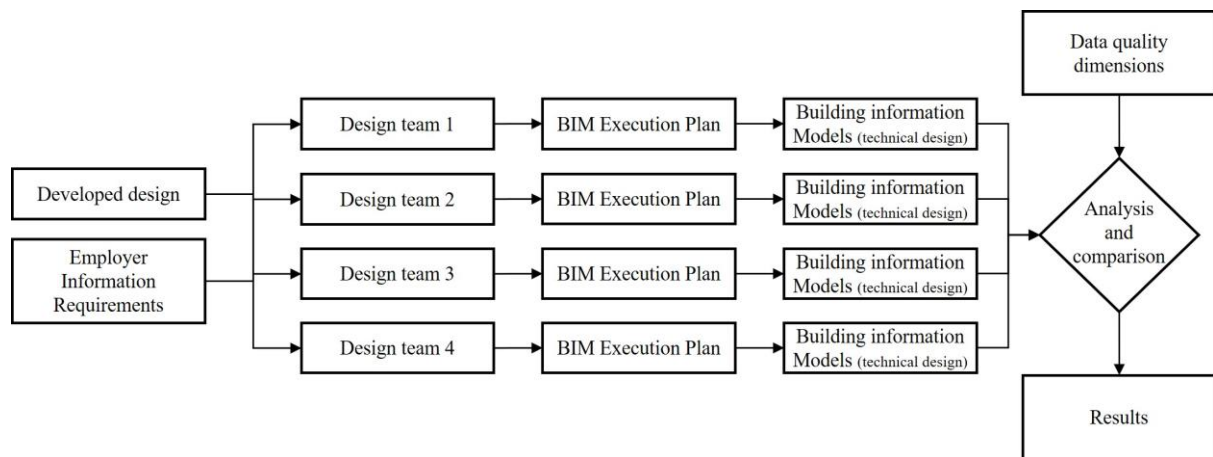
Figure 1: Methodology

Methodology

According to the scope stated in the introduction, i.e. the understanding of data quality issues focused on the objects composing information models obtained from real-world applications, this study was designed to allow the comparison between similar models developed by independent design teams. This allows the identification of common issues shared by different design teams identifying these as critical ones. Following the definition of data quality that highlights its variability according to the objectives and the needs of the process where data quality is evaluated, it is required the definition of a common background and common objectives shared by all the design teams involved in the experimentation. Hence, as reported in Figure 1 the methodology starts from the development of the original Design of a medium-large building with a basic specification of the requirements in terms of performance. Moreover, to guarantee the coherence between the building information models developed by the independent teams during the process was defined an Employer Information Requirements (EIR) (BSI, 2013) were the informative needs were described. Looking at the information requirements for the objects composing the models, the EIR specified the obligatory use of the USA BIM Forum (BIMForum, 2016), requiring the LOD 300 for all the modelled objects (architectural, structural, MEP, etc.).

These two main documents were shared with four design teams in the industry. To maintain the comparability of the models, the use of a specific BIM authoring tool (i.e. Autodesk Revit) was defined. Hence, each team started with the same design guidelines, the same BIM requirements and the same objective, that was win the competition proposing the best project in terms of performance and quality of the design and of the models.

Starting from this point, each design team developed a BIM execution plan (BEP) (BSI, 2013) to support the proposed models and a project proposal defined through the development of building information models.

Collected the BEPs and the models, the analysis phase started. The analysis focused on the models objects according to the four main data quality dimensions described in the background chapter. Because of the proposed methodology, timeliness was not relevant for the study. In fact, the study is focused on a static analysis in a fixed time and with fixed objectives. Hence, not up-to-date issues can be highlighted. Comparing to other means of data quality evaluation, the proposed dimensions provide a consolidated structure to discuss data quality issues that is the focus of this work.

In the proposed case study, the organisational structure defined by each design team was considered as a black box focusing on the final outputs (i.e. the building information models and all the related documents) and not on the generation process. To maintain the privacy of all the subjects involved in the case study, every reference to the original models has been omitted and the models are numerated from 1 to 4. Each design team defined more than one model according to different internal rules. Hence, in the term "model" are comprises all the models composing the design developed by a specific team (e.g. "Model 1" includes the architectural, structural, and destination of use models).

Case study

According to the proposed methodology, the case study herein presented is based on the analysis and comparison of four projects developed using BIM processes to identify common and/or relevant issues in terms of data quality in each project and to provide insight about how real-world models are developed. In the following, for each data quality dimension, i.e. accuracy, consistency and completeness are identified issues and peculiarities revealed during the case study analysis.

Accuracy

The database structure embedded in BIM authoring tools is limited and can produce incoherent data. In the market can be found different tools some more flexible and other more rigid where the inclusion of only predefined information is allowed. Nevertheless, in contrast with traditional relational databases, the data structure of BIM authoring tools can present relational issues. Moreover, except for some basic constrains in the data structure such as the data type, there is not the possibility to define embedded constrains to limit the introduction of erroneous data. Unfortunately, this kind of issues is not quantifiable in quantitative terms because its evaluation requires a manual exploration of the model due to the difficulties in defining general rules that can be executed automatically. Nevertheless, it is possible to identify some common issues revealed in the analysed models classifying these issues in four main points:

- Inconsistency between information fields.
- Precompiled information fields:
  - The information is not known.
  - The information is known but the designer does not use it.
- Other human related accuracy issues (e.g. typos).

Starting from the first point, the most widespread issue is the inconsistency between the name of objects and their informative contents. Technically, in this case the semantic structure of the BIM authoring tool is coherent. However, the inclusion of redundant information in the name may cause this type of accuracy issue.

The issues related to precompiled values can be divided in two cases. On the one hand, some information field cannot be eliminated from the data schema and cannot be defined as empty value. Hence, the designer is obliged to leave the precompiled value if he or she does not know the information. In this case, if this situation is not explicitly defined in the BEP, it is not possible to know if the information can be used or not. A diffused example in the analysed model are the information fields "absorptance" and "roughness" associated to walls that are always compiled even if the information is not known.

On the other hand, it may happen that the designer knows the information but this last is not used in the process. Once again, the designer leaves the precompiled value in the specific object. An example diffused in the analysed models is the "function" field that is usually maintained on "exterior" even in interior elements.

Consistency

During the experimentation have been found several issues that can be related to the consistency dimension.

These issues, as in the case of accuracy, cannot be evaluated through a quantitative approach but it is possible to identify some common terms in the analysed models. Can be listed three main critical points.

- Misuse of the object categorisation provided by BIM authoring tools.
- Introduction of information in the wrong information field.
- Not coherent use of the same information field.

About the first point, BIM authoring tools provide a defined semantic structure that aims to improve information exchange processes and to limit the interpretability of information. However, the structure proposed in commercially available BIM authoring tools is limited and does not include all the elements that can be required in the development of a building model.

This issue together with a limited vision of the personnel involved in the construction sector on the possible future uses of data, can bring to the incorrect use of the semantic structure provided by BIM authoring tools. Some examples related to this issue are following listed.

- To model vertical wood elements related to the fixed furniture of an open space has been used "concrete pillars" characterised by wood material.
- To simulate the skirting have been used walls characterised by wood material.
- To define complex railings have been used curtain walls.

The second and third points are shared with the accuracy dimension, but they imply an incorrect use of the semantic structure embedded in BIM authoring tools. The misuse of information fields can be identified as the introduction of information in the wrong information field. Typical examples that have been found during the research are the use of text free fields such as comments or descriptions to include a performance or a dimensional parameter creating possible inconsistency e.g. with the parameters based on the geometry of the object. With reference to the third point, have been found some cases where the same information field has been used to include different data types. For example, once the specific field contains the thermal transmittance of the object and in another object the same information field contains a generic comment, or the dimension of the element.

Completeness

The completeness dimension can be related to two factors. On the one hand, one can measure if the models satisfy the information requirements defined in

the EIR. For example, in this case study all the objects contained in the building information models need to reach LOD 300 according to the USA BIM Forum (2016 version). On the other hand, the completeness dimension can evaluate the information structure that has been defined to satisfy the requirements. In this case it is of interest the evaluation of the volume of information generated in an information model and if the configuration defined in the models can cause data quality issues.

A first analysis revealed the misalignment between the USA BIM Forum requirements and the information included in the models. Table 1 reports the completeness rate analysis focused on three object types comparing the minimum information required according to the USA BIM Forum and the related information included in each object.

Table 1: Completeness rate according to USA BIM Forum 2016 version

Object type	U.B.F	M1	M2	M3	M4
Interior fixed partition	8	7	6	6	6
	100%	87%	75%	75%	75%
Exterior window wall	5	4	4	3	4
	100%	80%	80%	80%	80%
Structural frame concrete	12	4	3	4	4
	100%	33%	25%	33%	33%

Legend

U.B.F.: USA BIM Forum specification 2016. The column reports the minimum information required.

M1, M2, M3, M4: Model 1, 2, 3 and 4. The columns report how many information requirements have been satisfied with reference to the specific U.B.F. requirements. The percentage indicates the completeness rate for the specific object type.

Even if it is possible to individuate a common behaviour in the inclusion and/or exclusion of information, every design team defined different information structures to reach the defined requirements.

Figure 2 (next page) shows the relation between the number of objects and the number of fields respectively in models 1, 2, 3, and 4. Number of objects and number of fields are divided according to a logical (and technical) subdivision of the models. In the graphs are reported data for the architectural models (circle), the structural models (X), the materials (square), and the spaces (plus). Focusing for example on the architectural models, the number of objects per model ranges from around 3.400 to around 5.500, while the number of fields range from around 110.000 to around 200.000. Nevertheless, the

correlation between number of objects and number of information fields is not linear between the four models. The ratio between the number of objects and the number of information fields vary from around 29 (model 3) to around 50 (model 2). This situation can be related to the introduction of information fields defined ad hoc by each team to satisfy the requirements imposed by the EIR. To understand this configuration in the different models it is possible defining a matrix to highlight how personalised fields has been added to specific object types.

Table 2 reports the comparison between the number of fields that are inherently defined in the specific BIM authoring tool in four classes of elements, namely walls, windows, pillars and materials, and the number of fields derived from the analysed models.

It is clear how each design team defined a different structure of data to reach the required LOD. This behaviour can cause critical issues related to the accessibility and interpretability of the information because of the not standard structure defined for each model.

Table 2: Identification of the number of information fields per object in the different models

	Walls			Windows		
	S	M	Δ	S	M	Δ
Model 1	62	89	27	58	75	17
Model 2	62	87	25	58	76	18
Model 3	62	62	0	58	70	12
Model 4	62	67	5	58	65	7
	Pillars			Materials		
	S	M	Δ	S	M	Δ
Model 1	59	75	16	42	47*	5
Model 2	59	89	30	42	50*	8
Model 3	59	61	2	42	42*	0
Model 4	59	64	5	42	42*	0

Legend

S: number of parameter in a standard model (i.e. number of information fields that the specific BIM authoring tool contains as starting point)

M: number of parameter in the analysed model (i.e. number of information fields calculated for each object classes in the analysed models)

Δ: $\Delta = M - S$

*: the number shows the material completed with physical and thermal properties (these are not included in all the materials included in the model)
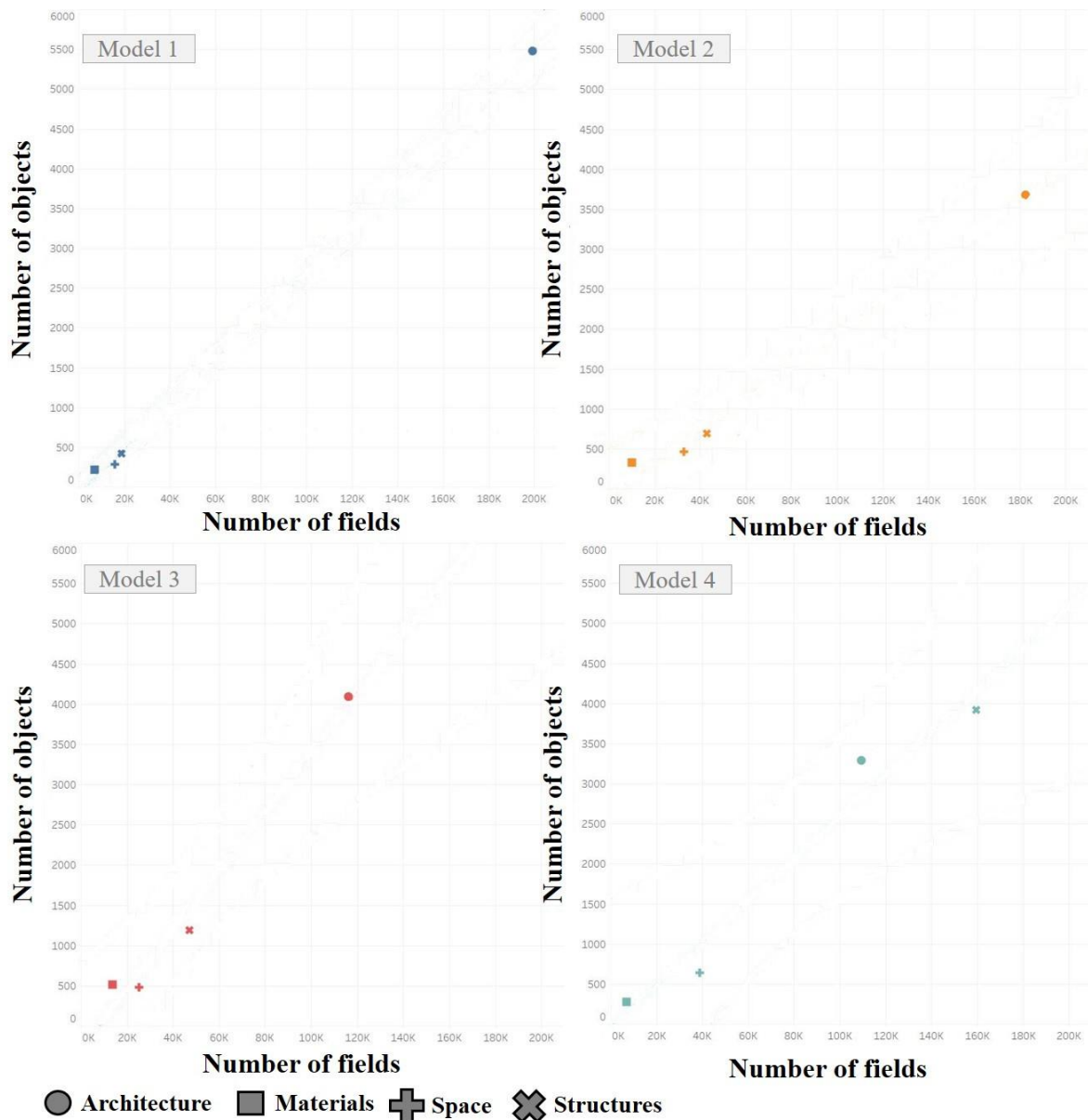
Figure 2: Analysis of the relation between number of objects and number of informative fields included in each model. The analysis is divided in four main classes of objects, i.e. architectural objects, materials, rooms (spaces), structural objects

Nevertheless, the proposed table highlights that it is possible to reach similar levels of information on the object with a very different number of information fields, in some cases even without including additional information fields in the embedded structure of the BIM authoring tool.

This situation is further emphasised in Figure 3 (next page) where are identified the number of objects (vertical axes) and the number of empty information fields (horizontal axes) in the architectural models. As mentioned in the background chapter, empty fields can cause several issues in the use of the models and can be listed as data quality issues with reference to completeness. Hence, a high number of empty fields brings to a low data quality evaluation due to the

difficulties in the use of model information. Moreover, calculating the percentage of completeness for each model (intended as how many fields have been compiled), many information fields included in the models were useless with reference to the objectives and scopes of the models. For example, Model 2 shows a percentage of completeness equal to 52,4 % that means that around 85.680 fields were left empty. At the same time, the model does not satisfy the information requirements related to LOD 300 (Table 1).
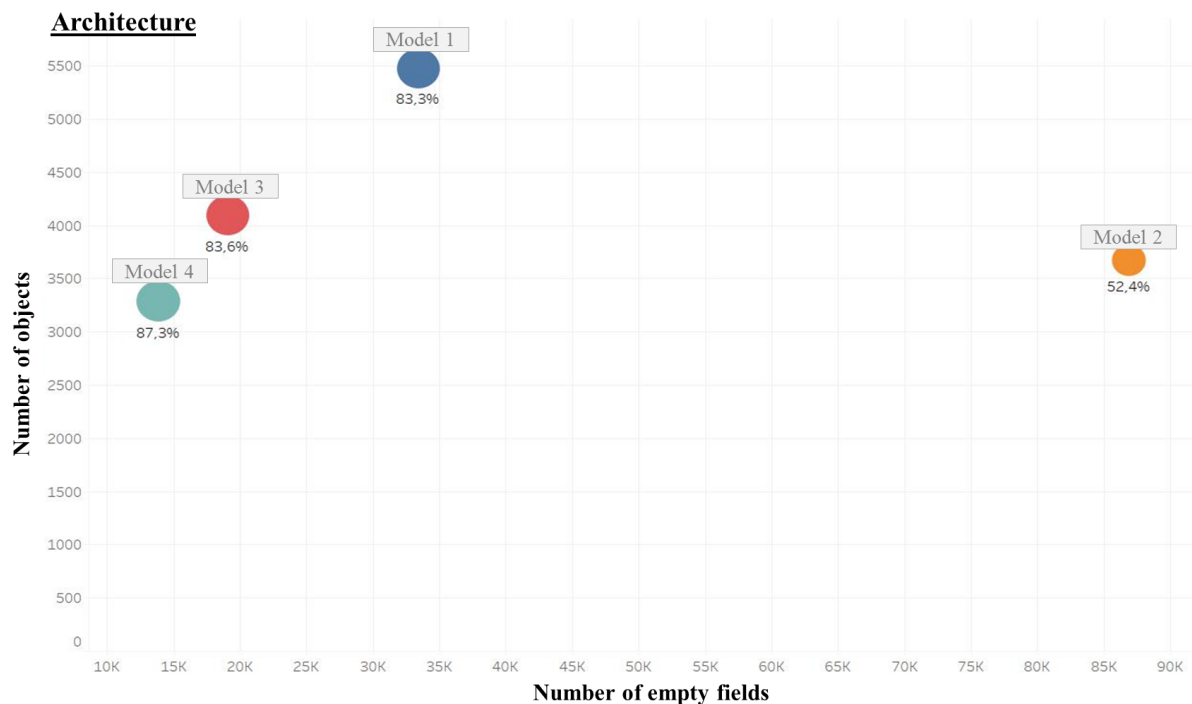
Figure 3: Analysis of the relation between number of objects and number of empty fields. The percentage under each circle indicates the percentage of completion of the informative fields in each model. The graph is limited to the architectural objects

Discussion

Considering the existing studies in the field of quality assessment of information models, we proposed a specific analysis devoted to understanding the possible data quality issues in the source information models, i.e. information models in their native format focusing on the component's objects. The results of the experimentation highlighted several critical points in the management of data quality in BIM revealing common user behaviours and/or technological limitations in real-world applications. In comparison to existing studies, this paper proposes a detailed identification of data quality issues generated in real world models highlighting not only procedural and technical problem e.g. related to IFC export processes but also the behaviour of the industry in the development of information models. The authors believe that the evaluation of the societal context is of critical importance in BIM data quality area.

Some of the issues highlighted in the paper can be identified as software dependent issues that means that are issues related to the specific software and not to BIM in general. Nevertheless, because building information models are developed using specific software, it is useful to know possible related issues to improve the quality of the software and to improve the quality of the research linking the results to the peculiarities of each source instrument.

In the following, we discuss the results according to technical and procedural issues exploring possible solutions e.g. using standard and shared structures. In fact, even if the processes used to generate the models were considered as black boxes, the results highlighted a correlation with the processes that play a fundamental role in improving data quality performance. It is worth mention that standard and shared information structures can be seen as means to limit data quality issue. This does not mean that there is always a direct link between specific standards and the data quality evaluation. Data quality can be evaluated even without considering specific standards, while these lasts can be identified as valuable means to improve the control of information models data quality.

Technical issues

The misuse of information field highlighted in the consistency dimension brings to the impossibility to develop automated algorithms and/or applications based on the data contained in information models because it is not possible to know where to find the right information and if a specific information field reports always the same information type. Moreover, the inconsistency between information fields, underlined in both accuracy and consistency, can limit the reliability of information models with the direct consequence that information is not used, loosing value.

The analysis focused on completeness highlighted the misalignment between the EIR requirements and the information included in the analysed models. This

issue may be analysed by different perspectives. On the one hand, the direct use of information requirements defined for foreign markets (USA BIM Forum) can generate difficulties in the interpretation and in the alignment with the national regulation. On the other hand, the knowledge background (also related to the societal aspects) of the personnel involved in the development of the model can create the highlighted limitations.

Procedural issues

Some issues that have been underlined in the different data quality dimensions can be related to procedural components. Most of the data accuracy issues revealed during the analysis can be limited (and probably eliminated) identifying precise processes to develop new objects. In the same way, this can help in limiting consistency issues because the responsible for the development of the objects should guarantee for their consistency according to the objectives of the model. Moreover, the variety of data structures used to reach the required LOD can be limited through a structured process that, starting from the client, defines a rigid matrix of information with specific name, data type, semantic, etc.

Nevertheless, in many cases the application of these rigid and structured processes in the construction industry demonstrated to be all but easy. In fact, due to the complexity of the construction products (buildings, infrastructures, etc.) it is not possible to think through the entire process from the beginning with the consequent need to introduce changes and adjustments on the way in a reduced time lap. Moreover, some data quality issues related to human factors cannot be eliminated and they must be considered in the analysis and use of information models.

Conclusions

The implementation of BIM can provide effective means in the management of data quality in the construction processes. However, the inclusion of geometrical and non-geometrical information in an information model imposes new issues related to the verification and validation of the model quality. Evaluating and controlling the quality of information models is a significant challenge and requires further research. In order to address this gap and with the scope of understanding data quality dimensions in building information models developed in native formats, a systematic analysis of real-world models is presented in this work. Data quality issues were classified according to three dimensions, namely accuracy, coherence and completeness. This structure provides the background to explore and discuss technical and behavioural issues in the processes of information models development.

The authors believe that the proposed results can represent a good starting point to understand where

data quality issues are generated and how (and if) these issues can be limited or solved. They can be used in different applications, for example to improve existing and future studies that are based on intensive data use of BIM objects such as export procedures (from native to IFC, from native to graph databases, from native to OWL, etc.), expert systems, machine learning and artificial intelligence applications, design support systems, etc. evaluating the sources of errors and including possible processes to limit their influence on the results. Moreover, understanding the typical sources of errors in native models can help in the evaluation of exporting processes highlighting issues related to the input models and those related to the export and/or import processes. Moreover, relating this study with the concept of Model View Definition (MVD) it is possible to optimise the research devoted to automate and optimise the translation process used to share information. Existing studies highlighted the dependency between data quality and performance of the translation process. Understanding the nature of the input errors can help in defining better algorithms and information flows.

On the other hand, the proposed analysis can be used by the industry to improve the existing modelling processes, i.e. the rules that can be defined in a project team and/or in a collaborative environment to guarantee the quality of the overall process that cannot forget the quality of every subject (people) and of every object (models) that are included in the process.

As a future work, we envision to apply the same experimentation structure to other BIM authoring tools to highlight software dependent issues and/or issues that are shared among different tools. Moreover, it would be interesting to develop the same experimentation in a different societal context to highlight how different knowledge and contextual backgrounds can affect the process development with reference to data quality.

References

BSI (2013) PAS 1192-2:2013, Specification for information management for the capital/delivery phase of construction projects using building information modelling. *UK: The British Standards Institution.*

Building Smart (2018) IfcDoc Tool Summary. Available at: http://www.buildingsmart-tech.org/specifications/specification-tools/ifcdoc-tool (Accessed: 14 January 2019).

Delone, W. H. and McLean, E. R. (2003) The DeLone and McLean Model of Information Systems Success : A Ten-Year Update, *Journal of Management Information Systems*, 19(4), pp. 9–

30.

Donato, V., Lo Turco, M. and Bocconcino, M. M. (2018) BIM-QA/QC in the architectural design process, *Architectural Engineering and Design Management*, 14(3), pp. 239–254. doi: 10.1080/17452007.2017.1370995.

Eastman, C. et al. (2011) BIM Handbook: A guide to Building Information Modeling for owners, managers, architects, engineers, contractors, and fabricators. Hoboken, NJ, USA.

Fan, W. (2012) Data Quality: Theory and Practice, In: *Proceeding of the 13th International Conference, WAIM*, Harbin, China, pp. 1–16.

Farias, T. M., Roxin, A. and Nicolle, C. (2018) A rule-based methodology to extract building model views, *Automation in Construction*, 92, pp. 214–229. doi: 10.1016/j.autcon.2018.03.035.

Hernández, M. A. and Stolfo, S. J. (1998) Real-world Data is Dirty: Data Cleansing and The Merge/Purge Problem, *Data Mining and Knowledge Discovery*, 2(1), pp. 9–37. doi: 10.1023/A:1009761603038.

Hu, Z. Z. et al. (2016) Improving interoperability between architectural and structural design models: An industry foundation classes-based approach with web-based tools, *Automation in Construction*, 66, pp. 29–42. doi: 10.1016/j.autcon.2016.02.001.

IEEE (1990) Standard Computer Dictionary. A Compilation of IEEE Standard Computer Glossaries. New York: *Institute of Electrical and Electronics Engineers*.

Kim, W. et al. (2003) A Taxonomy of Dirty Data, *Data Mining and Knowledge Discovery*, 7(1), pp. 81–99. doi: 10.1023/A:1021564703268.

Kouhestani, S. and Nik-Bakht, M. (2018) Towards level 3 BIM process maps with IFC & XES process mining, In: *Proceedings of the 12th european conference on product and process modelling (ECPPM)*, Copenhagen, Denmark, pp. 103–111.

Laakso, M. and Kiviniemi, A. (2012) The IFC standard - a review of history, development, and standardization, *Journal of Information Technology in Construction*, 17, pp. 134–161.

Lee, Y. C., Eastman, C. M. and Solihin, W. (2018) Logic for ensuring the data exchange integrity of building information models, *Automation in Construction*, 85, pp. 249–262. doi: 10.1016/j.autcon.2017.08.010.

Mocnik, F. B. et al. (2018) A grounding-based ontology of data quality measures, *Journal of Spatial Information Science*, 16(16), pp. 1–25. doi: 10.5311/josis.2018.16.360.

Naumann, F. and Rolker, C. (2000) Assessment Methods for Information Quality Criteria, In: *Proceedings of the 2000 International Conference on Information Quality*, pp. 148–162.

Neumann, T. (2018) Reasoning in the Presence of NULLs, In: *Proceedings of the 34th International Conference on Data Engineering (ICDE)*. Paris, France.

Pauwels, P., Zhang, S. and Lee, Y. C. (2017) Semantic web technologies in AEC industry: A literature overview, *Automation in Construction*, 73, pp. 145–165. doi: 10.1016/j.autcon.2016.10.003.

Sacks, R. et al. (2010) The Rosewood experiment — Building information modeling and interoperability for architectural precast facades, *Automation in Construction*, 19(4), pp. 419–432.

Sacks, R. et al. (2017) Semantic Enrichment for Building Information Modeling: Procedure for Compiling Inference Rules and Operators for Complex Geometry, *Journal of Computing in Civil Engineering*, 31(6), doi: 10.1061/(ASCE)CP.1943-5487.0000705.

Solihin, W., Eastman, C. and Lee, Y.-C. (2015) Toward robust and quantifiable automated IFC quality validation, *Advanced Engineering Informatics*, 29(3), pp. 739–756. doi: 10.1016/j.aei.2015.07.006.

Törmä, S. (2013) Semantic linking of building information models, In: *Proceedings - 2013 IEEE 7th International Conference on Semantic Computing, ICSC*, pp. 412–419. doi: 10.1109/ICSC.2013.80.

Tribelsky, E. and Sacks, R. (2010) Measuring information flow in the detailed design of construction projects, *Research in Engineering Design*, 21(3), pp. 189–206.

Wang, R. Y. and Strong, D. M. (1996) Beyond Accuracy: What Data Quality Means to Data Consumers, *Journal of Management Information Systems*, 12(4), pp. 5–33. doi: 10.1080/07421222.1996.11518099.

Zadeh, P. A. et al. (2017) Information Quality Assessment for Facility Management, *Advanced Engineering Informatics*, 33, pp. 181–205. doi: 10.1016/j.aei.2017.06.003.

Zaniolo, C. (1984) Database relations with null values, *Journal of Computer and System Sciences*, 28(1), pp. 142–166. doi: 10.1016/0022-0000(84)90080-1.