Malta, January 2020

3rd INTERNATIONAL WORKSHOP ON SPATIAL DATA QUALITY

Count based quality control of "As Built" BIM datasets using the ISO 19157-1 framework

Francisco Javier Ariza-López, José Rodríguez-Avi, Juan Francisco Reinoso Gordo, Iñigo Ariza-López

fjariza @ujaen.es, jravi@ujaen.es, jfreinoso@ugr.es; inigoariza@us.es

Universidad de Jaén; Paraje de las Lagunillas S/N, E-23.071-Jaén

Funded by the Project: MEMCALIG (Modelización Estadística de Matrices de Confusión en Calidad de la Información Geográfica)

GOBIERNO DE ESPAÑA

MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

PROYECTO COFINANCIADO POR LA UNIÓN EUROPEA

FONDO EUROPEO DE DESARROLLO REGIONAL

Objectives

Our goals are:

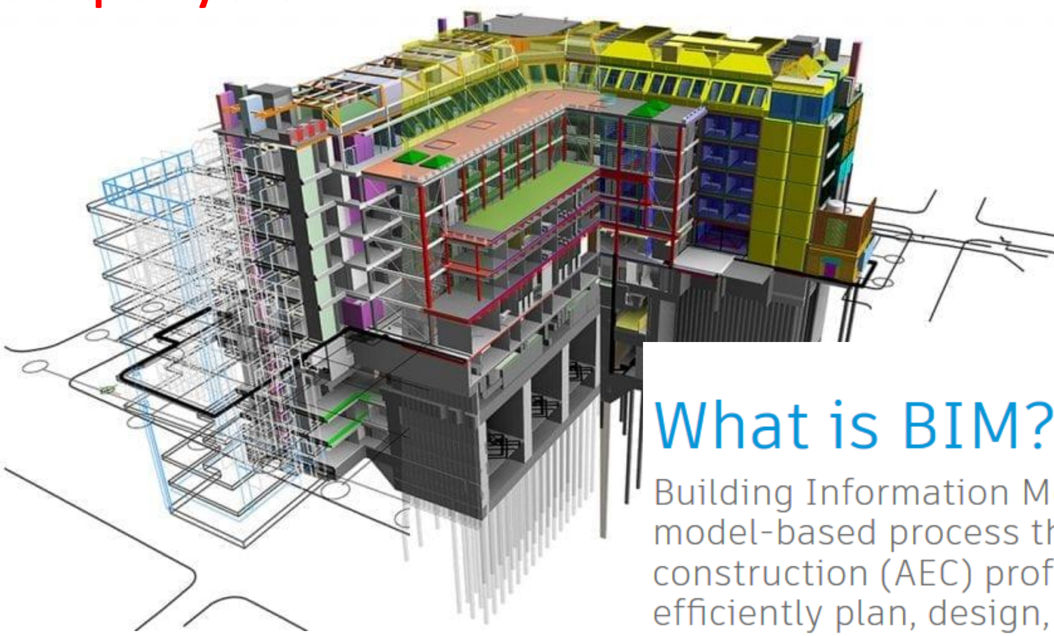
- To show that ISO 19157-1 is USEFUL for BIM data.
- Test a situation is similar to spatial data → "As built".
- That it is also possible to perform statistical controls of BIM data.

Contents

- Introduction.
- Proposal of quantitative data quality elements for BIM data.
- Proposal of a statistical method for BIM data.
- Example
- Conclusions.

Introduction

Data quality and BIM



What is BIM?

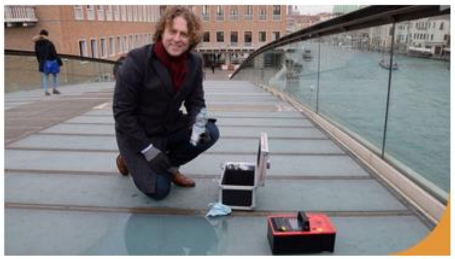
Building Information Modeling (BIM) is an intelligent 3D model-based process that gives architecture, engineering, and construction (AEC) professionals the insight and tools to more efficiently plan, design, construct, and manage buildings and infrastructure.

What is an as build?

In the architecture and construction industry, “As-Built” refers to a drawing that shows the EXISTING dimensions and conditions of a **building**, space, or area. ... Often, there are complications that arise during construction which force the contractor to make variations from the original plans.

Introduction

Data quality and BIM



EPISODE 1

Episode 1

Former aerospace engineer Justin Cunningham tries to keep the Glasgow Tower turning, and Tomo Umewaka helps engineers in Osaka keep their airport from sinking into the sea.

EPISODE 2

Episode 2

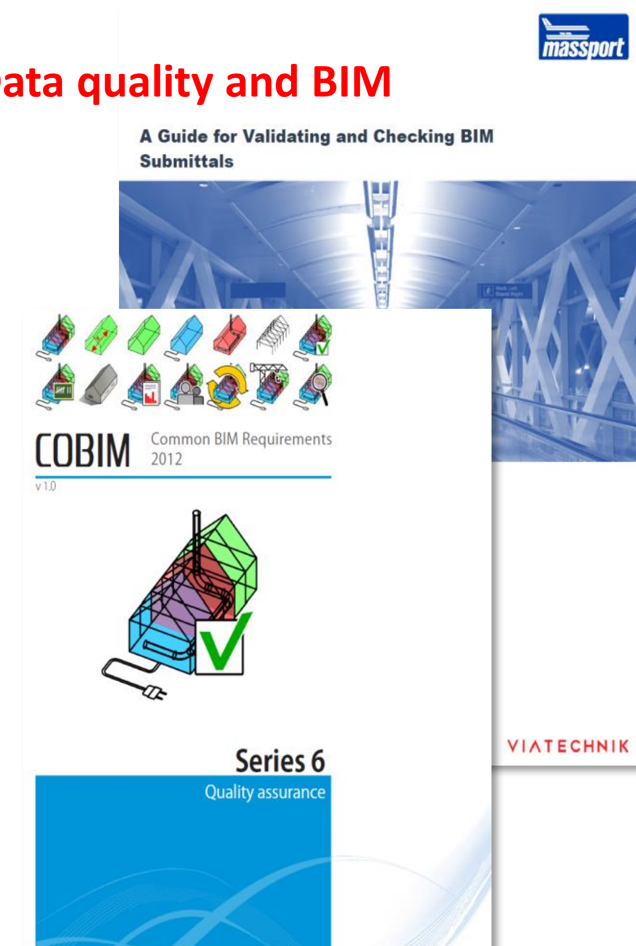
Justin climbs on top of a stadium roof that is held up by air and in danger of collapse. Meanwhile, Jimena Gascon discovers why Mexico City's Metro project has been...

EPISODE 3

Episode 3

Justin uncovers the truth behind a catastrophic landslide that wiped out a ski-jump resort in Turkey, before jetting off to Gibraltar to see an airport runway built across a road.

Data quality and BIM



https://www.heraldo.es/noticias/aragon/2018/05/05/10-grandes-fracasos-historia-ingenieria-1241909-300.html
https://es.dplay.com/dmax/grandes-fracasos-de-la-ingenieria/

Introduction

For dealing with data quality a model is needed

ISO 19100 family`s model for data quality

Quality means

- **Identify** the relevant aspects of quality
- **Assess** these aspects with adequate methods
- **Quantify** the assessments by means of comparable measures
- **Describe** adequately all aspects in metadata
- **Specify** adequately all aspects in data product specifications
- **Assuring** the quality of 3rd party supplies

ISO 19157-1

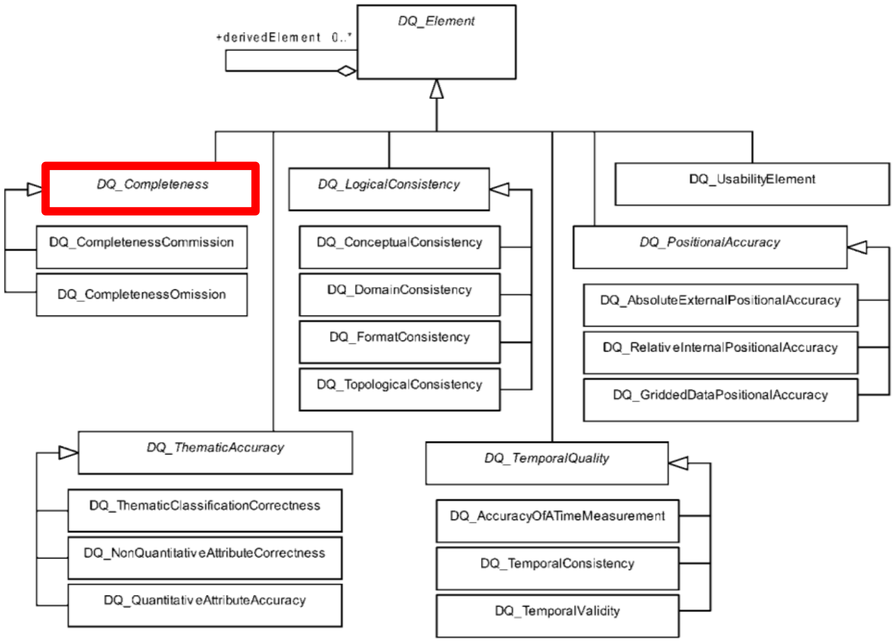
ISO 19115

ISO 19131

ISO 19158

Proposal of DQ elements

Completeness dimension



Completeness is defined as the presence and absence of features, their attributes and relationships. It consists of two data quality elements:

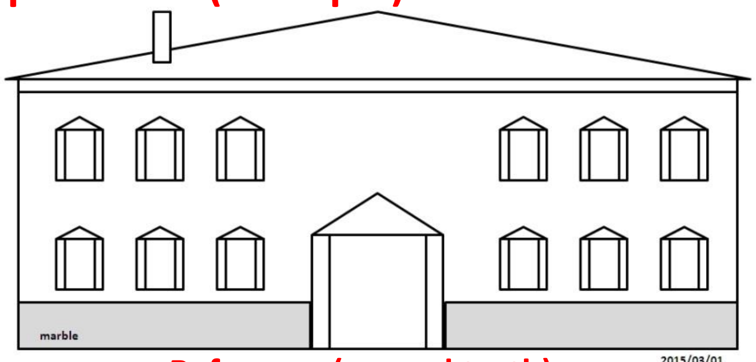
- **commission** – excess data present in a dataset;
- **omission** – data absent from a dataset.

As Built → It is necessary to check against reality

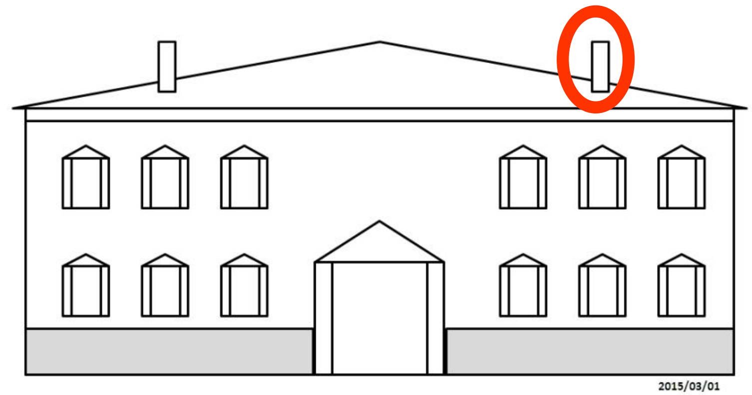
Count based quality control of “As Built” BIM datasets using the ISO 19157 framework

Proposal of DQ elements

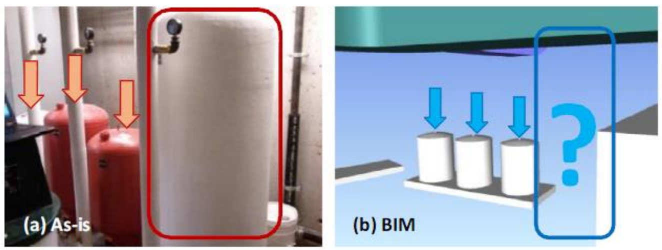
Completeness (example)



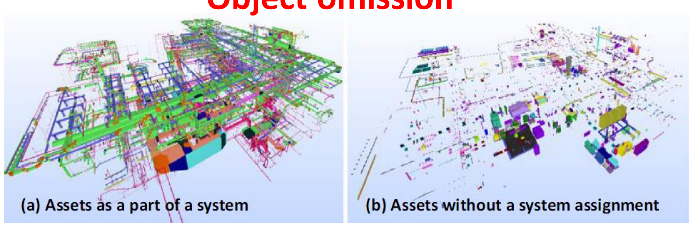
Reference (ground truth)



Dataset, case 1: COMISSION



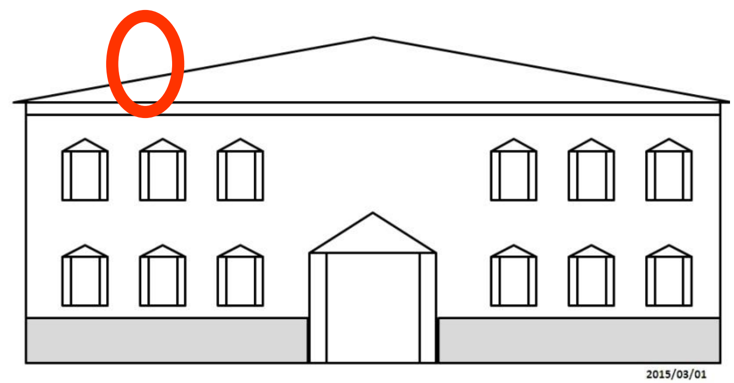
Object omission



Assignation omission

Information Quality Assessment for Facility Management

Puyan A. Zadeh [a,*], Guan Wang [a], Hasan B. Cavka [a], Sheryl Staub-French [a], Rachel Pottinger [b]



Dataset, case 2: OMISSION

Proposal of DQ elements

Logical consistency dimension

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.

Proposal of DQ elements

Logical consistency

Rule compliance and conflict detection (*clash detection*):

- Automatic
- Semi automatic
- Manual (visual)

SOLIBRI

VERITY

CLEAREDGE 3D

BIM TREE MANAGER

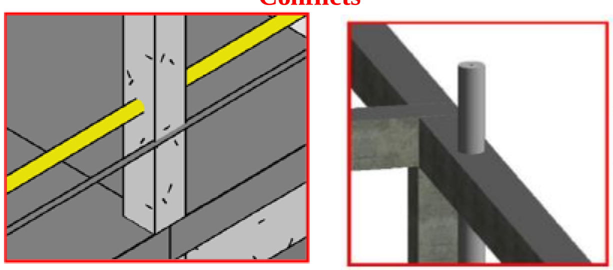
'3-CLICK & FIX'
CONSTANT QUALITY CONTROL
FOR AUTODESK® REVIT® BIM PROJECTS

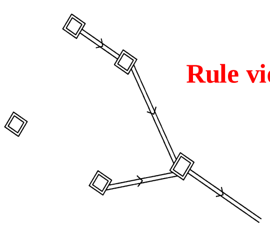
AGACAD TOOLS 4 BIM

RIB iTWO

RIB running together

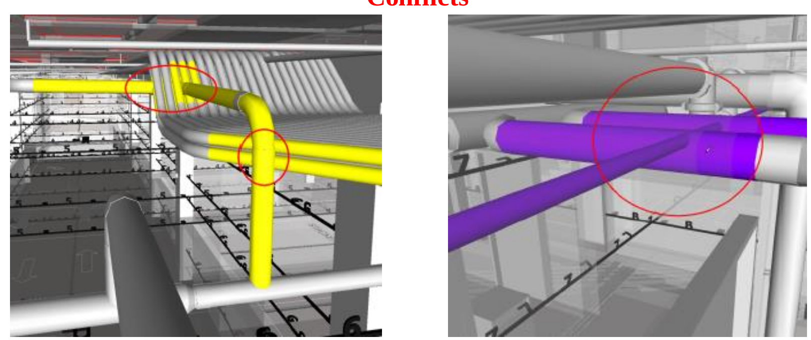
Conflicts





Rule violation

Conflicts



A Guide for Validating and Checking BIM Submittals

massport

Proposal of DQ elements

Positional accuracy dimension Proposal → Metric accuracy dimension

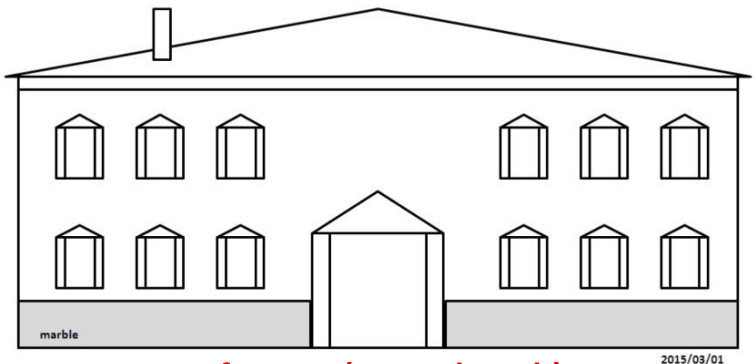
Positional accuracy is defined as the accuracy of the position of features within a spatial reference system. It consists of three data quality elements:

- 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 positional accuracy – closeness of gridded data spatial position values to values accepted as or being true.

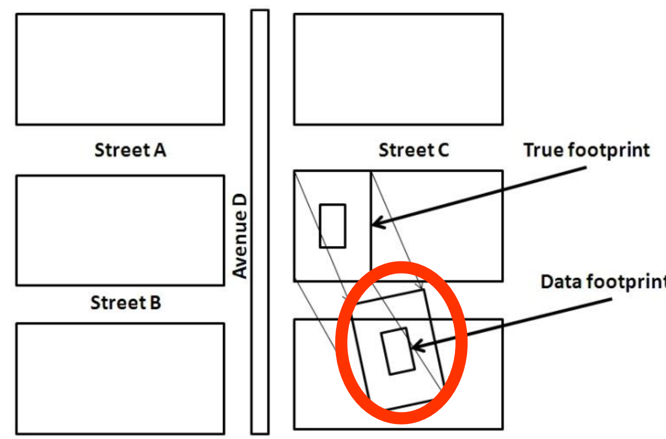
As Built → It is necessary to check against reality

Proposal of DQ elements

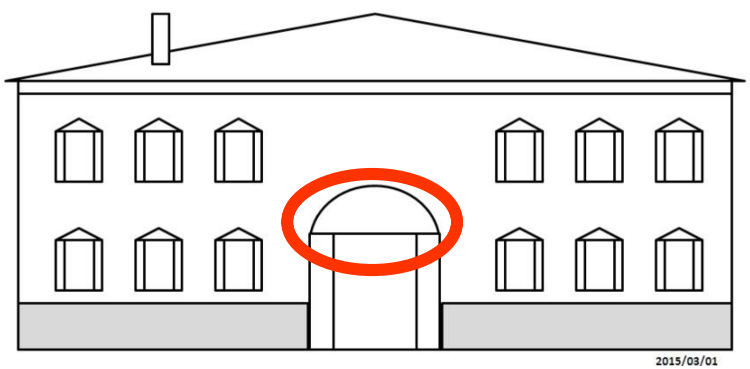
Metric accuracy



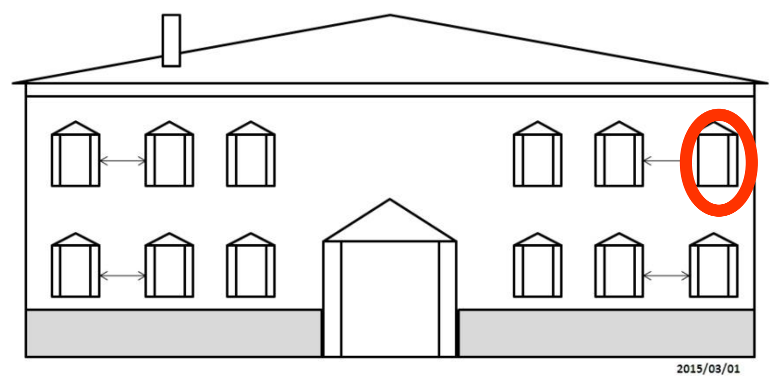
Reference (ground truth)



Absolute /external Pos. accuracy



Geometric fidelity

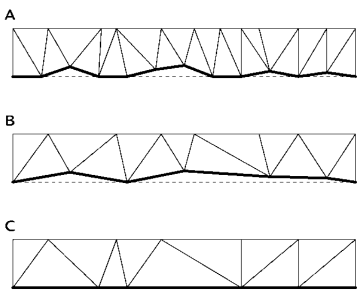


Relative / internal Pos. accuracy

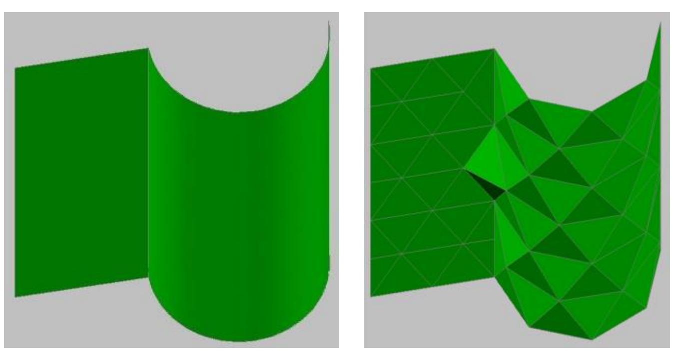
Proposal of DQ elements

Metric accuracy

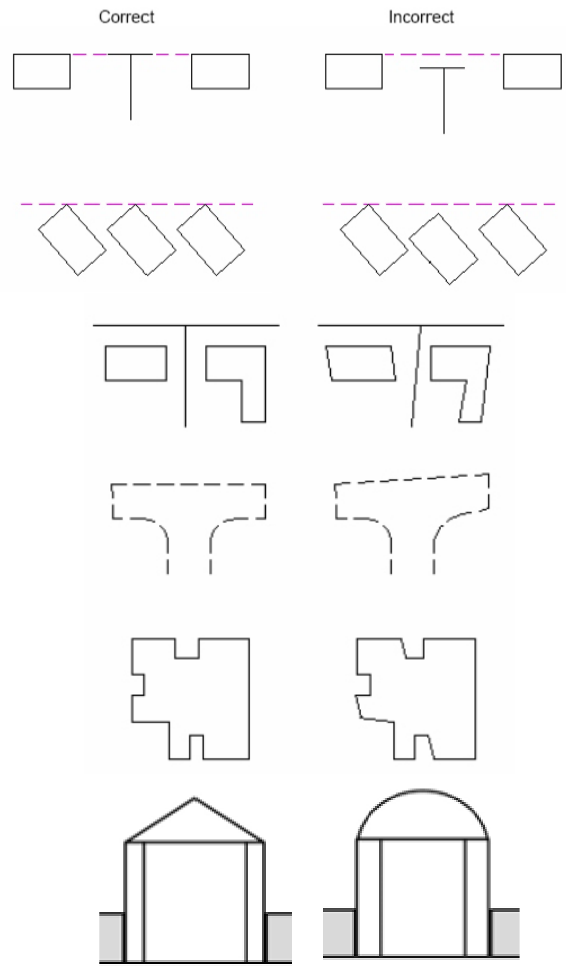
Geometric fidelity: Threshold / Tolerance compliance.



Data filtering A) model without reduction of points, B) and C), possible models obtained after a reduction of points.



Surface modelling



Positional correctness: The location is correct or incorrect.

Proposal of DQ elements

Thematic accuracy dimension

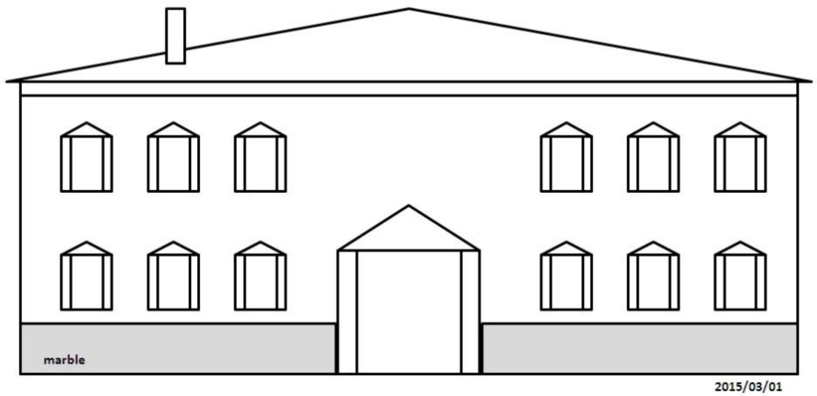
Thematic accuracy is defined as the accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships. It consists of three data quality elements:

- **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.

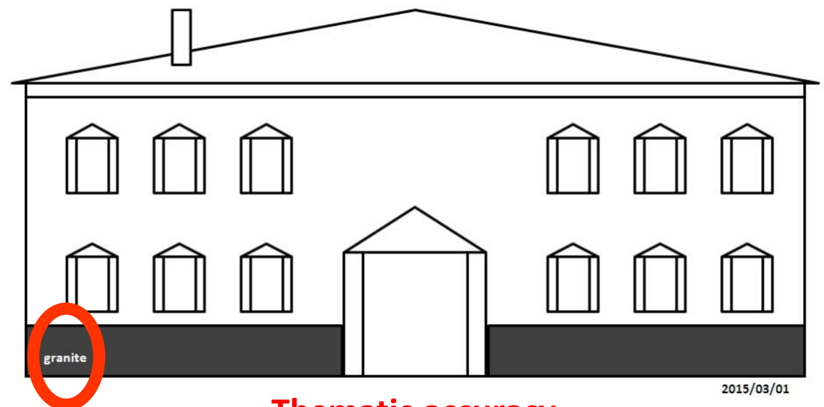
As Built → It is necessary to check against reality

Proposal of DQ elements

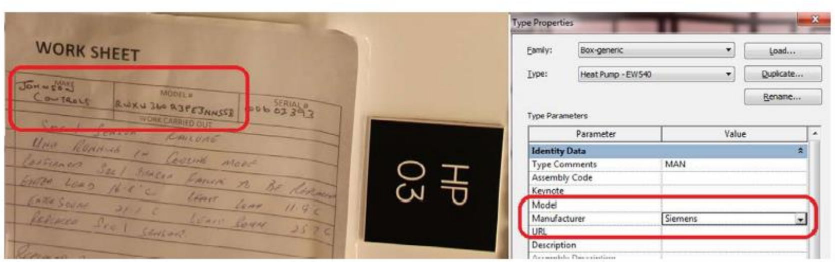
Thematic accuracy

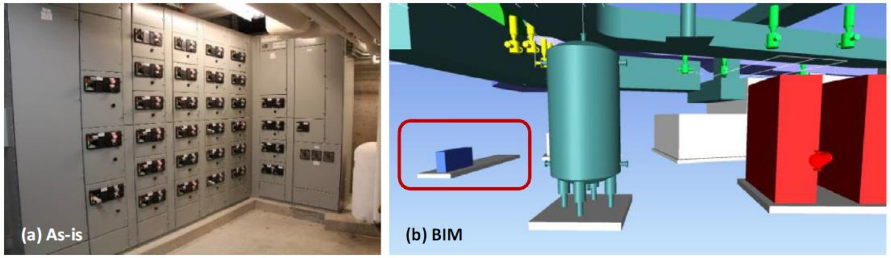


Reference (ground truth)



Thematic accuracy





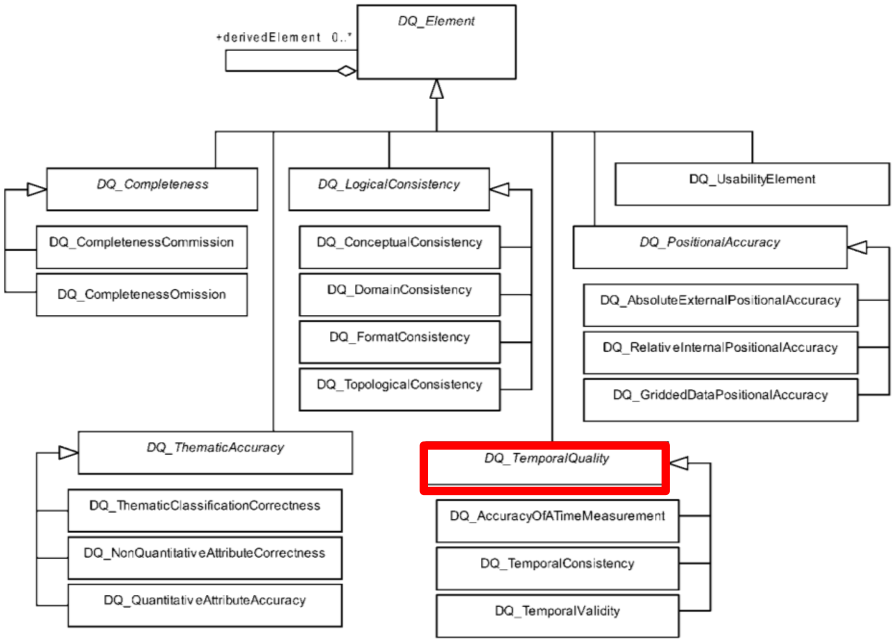
Information Quality Assessment for Facility Management

Puyan A. Zadeh [a,*], Guan Wang [a], Hasan B. Cavka [a], Sheryl Staub-French [a], Rachel Pottinger [b]

Advanced Engineering Informatics 33 (2017) 181–205

Proposal of DQ elements

Temporal quality dimension

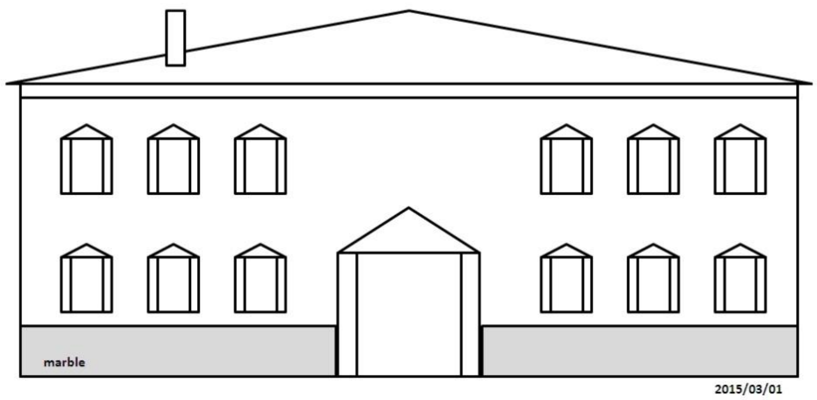


Temporal quality is defined as the quality of the temporal attributes and temporal relationships of features. It consists of three data quality elements:

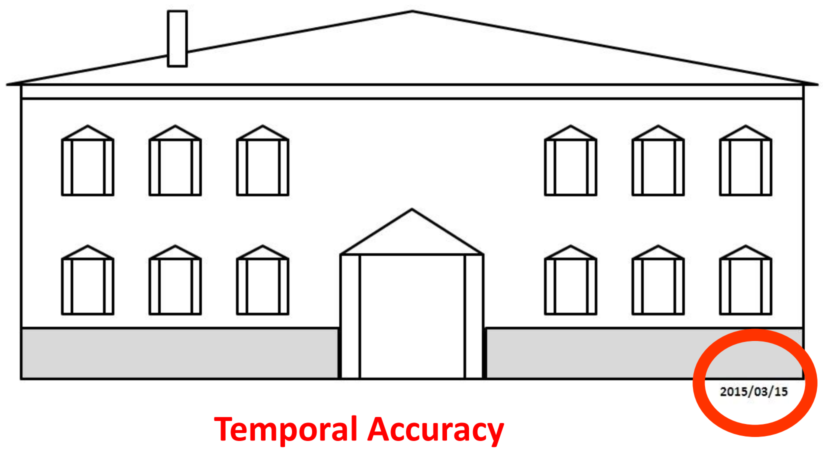
- **accuracy of a time measurement** – closeness of reported time measurements to values accepted as or known to be true;
- **temporal consistency** – correctness of the order of events;
- **temporal validity** – validity of data with respect to time.

Proposal of DQ elements

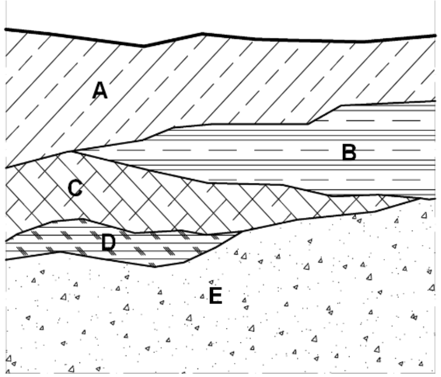
Temporal quality



Reference (ground truth)



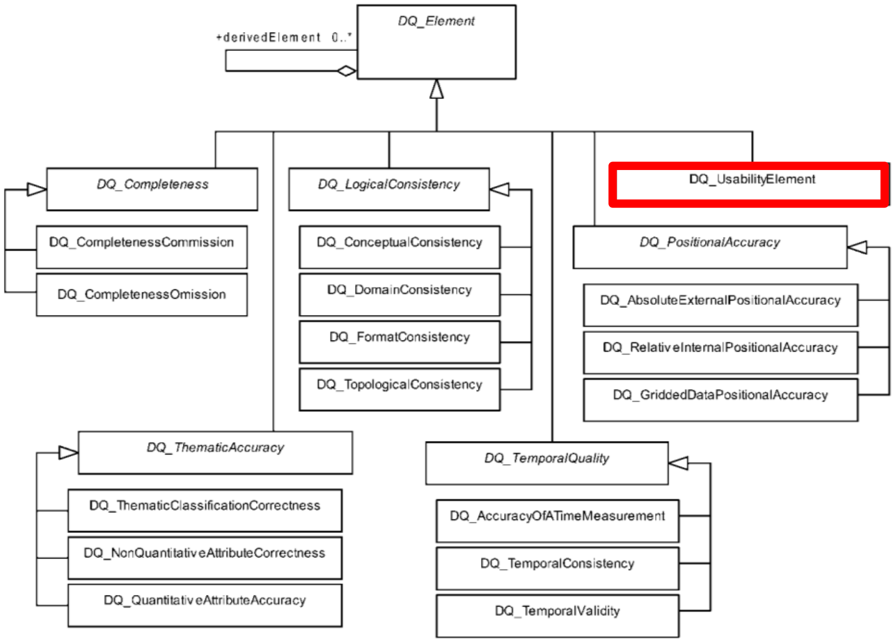
Temporal Accuracy



Stratigraphic scheme, where necessarily the age of A <B <C <D <E. If the age of C> D we would have a violation of the rules

Proposal of DQ elements

Usability dimension



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 can not be described using the quality elements described above. In this case, the usability element shall be used to describe specific quality information about a dataset's suitability for a particular application or conformance to a set of requirements.

Proposal of a statistical control

Remember that:

- Automatable processes → Total inspection (100%)
- Non-automatable processes → Sampling → Statistical tests

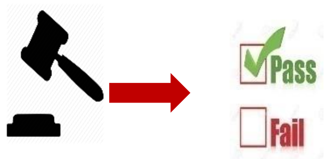
As Built → It is necessary to check against reality

Operational needs of a statistical control for BIM data quality:

- Jointly control of variables and attributes.
- Consider different seriousnesses.
- Allow joint control of various data quality elements, even of different dimensions.
- Establish a clear risk framework (user´s and producer´s risks).

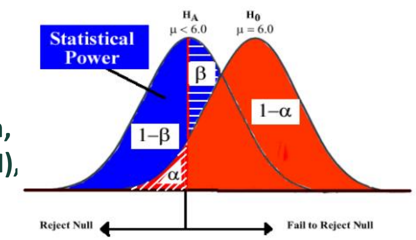
Proposal of a statistical control

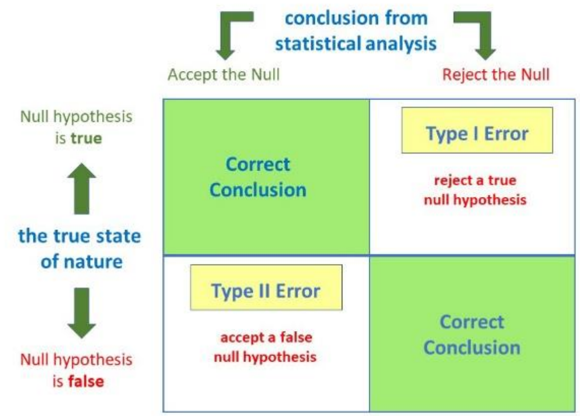
Statistical test. They are based on a distributional hypothesis. **So called Null Hypothesis (H_0)**



Tol, α, β

Tolerance, signification, power (risks: type I & II),





Kind of variables

- Quantitative
- Qualitative

H_0 : Binomial Model / Hypergeometric Model

(infinite population / finite population)

H_0 : Multinomial Model / Hypergeometric multivariate model

Proposal of a statistical control

Applying the jargon of ISO 19157-1:

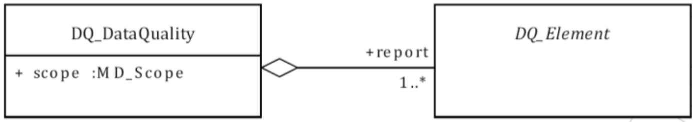
DQE → Data quality element.

Scope → Scope of interest. Composition of spatial, thematic, temporal… filters.

CoI → Category of Interest for a control. Set of classes filtered by a scope.

DQU → Data quality unit:

DQU = DQE + CoI



DQM → Data quality measure (Annex D of ISO 19157-1)

EM → Evaluation (assessment) method.

QL → Quality level (compliance level).

QC → Quality control:

Quality control QC = DQU + EM + QL

Proposal of a statistical control

Process’s steps are:

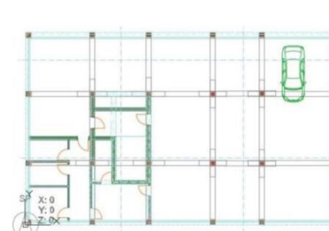
- Determine the CoIs. (*)
- Determine the population sizes of the CoIs.
- Determine the DQUs. (*)
- Determine population sizes for each DQUs.
- Determine the statistical model to apply for each DQU. (*)
- Determine the sample size. (*)
- Define the QC of each DQU (*) : QC = DQU + EM + QL
- Take a simple random sample (SRS) for each QC.
- Count the number of nonconformities for each QC.
- Calculate the p-values for each QC.
- Check global acceptance/rejection. Apply an MHTC, if applicable.

(*) They must be specified in the product specifications

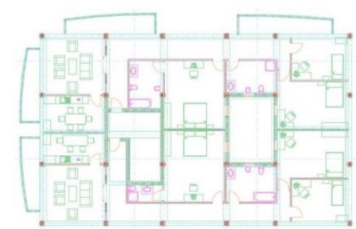
Example

The case

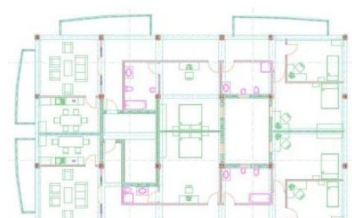
- Basement (garage)
- Ground floor (2 commercial premises)
- P1 (2 apartments)
- P2 (2 apartments)
- Roof with storage



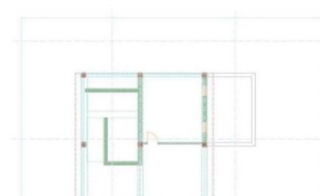
Basement



Apartments



Apartments



Roof with storage

Example

The case

- Determine the CoIs.
- Determine the population sizes of the CoIs.
- Determine the DQUs.
- Determine population sizes in DQUs.
- Determine the statistical model to apply.
- Determine the sample size.
- Define the QC of each DQU.
- Take a simple random sample for each QC.
- Count the number of nonconformities for each QC.
- Calculate the p-values for each QC.
- Check global acceptance/rejection. Apply an Multiple Hypothesis Testing Correction (MHTC) if applicable.

Table 1 Categories of interest in the BIMDB

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

Example

The case

- Determine the CoIs.
- Determine the population sizes of the CoIs.
- Determine the DQUs.
- Determine population sizes in DQUs.
- Determine the statistical model to apply.
- Determine the sample size.
- Define the QC of each DQU.
- Take a simple random sample for each QC.
- Count the number of nonconformities for each QC.
- Calculate the p-values for each QC.
- Check global acceptance/rejection. Apply an MHTC, if applicable.

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

Data quality units	Cases in the population (N)	Sample size (n)	
DQU1=Completeness of elements DQE = Commission + omission CoI = C1+C2+ ··· + C10	511	50	B
DQU2=Completeness of facilities DQE = Commission + omission CoI = C11+ C13	182	40	H
DQU3= Shape Fidelity DQE = Fidelity in shape CoI = C1+C2+ ··· + C10	1605	160	B
DQU4=Attributes of elements DQE = Correction of non-quantitative attributes CoI = C1+C2+ ··· + C10	462	50	B
DQU5=Attributes of installations DQE = Correction of non-quantitative attributes CoI = C12+ C13	491	50	B
DQU6= Shape Fidelity of walls DQE = Fidelity in shape CoI = C11	200	20	H
Total	**3451**	**350**	

Valores de n recomendados para rechazar $\mathbb{H}_0: \pi = \pi_0$ cuando el verdadero valor es $\pi_1 = \pi_0 + \delta$ con un tamaño de error de Tipo I del 5 % y un tamaño de error de Tipo II del 10 %

Distancia δ	Probabilidad bajo $\mathbb{H}_0$ (π_0)						
	1 %	**3 %**	**5 %**	**8 %**	**10 %**	**15 %**	**20 %**
0.01	1178	2828	4394	6610	8001	11176	13923
0.03	192	381	555	798	950	1296	1593
0.05	88	158	221	308	362	484	589
0.08	44	73	98	132	152	199	238
0.10	32	51	67	89	102	131	156
0.15	18	27	34	44	49	62	72
0.20	12	17	21	27	30	36	42

Example

The case

- Determine the CoIs.
- Determine the population sizes of the CoIs.
- Determine the DQUs.
- Determine population sizes in DQUs.
- Determine the statistical model to apply.
- Determine the sample size.
- Define the QC of each DQU.
- Take a simple random sample for each QC.
- Count the number of nonconformities for each QC.
- Calculate the p-values for each QC.
- Check global acceptance/rejection. Apply an MHTC, if applicable.

QC = DQU + EM + QL

Table 3 Definition of the quality controls by means of the data quality units and the conformity levels

Quality control	Data quality unit	Data Quality Measure and ID*	Conformity level (Maximum proportion of defects)
QC1	DQU1	Rate of excess items (ID=3) + Rate of missing items (ID=7)	1%
QC2	DQU2	Rate of excess items (ID=3) + Rate of missing items (ID=7)	3%
QC3	DQU3	Rate of unfaithful items (ID=**)	5%
QC4	DQU4	Rate of incorrect attribute values (ID=67)	10%
QC5	DQ5	Rate of incorrect attribute values (ID=67)	10%
QC6	DQ6	Rate of unfaithful items (ID=**)	80%, 15%,5%***

(*) 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.

Usability

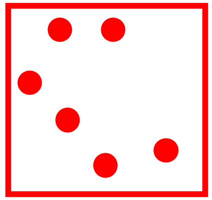
Example

The result

- Determine the CoIs.
- Determine the population sizes of the CoIs.
- Determine the DQUs.
- Determine population sizes in DQUs.
- Determine the statistical model to apply.
- Determine the sample size.
- Define the QC of each DQU.
- Take a simple random sample for each QC.
- Count the number of nonconformities in each QC.
- Calculate the p-values for each QC.
- Check global acceptance/rejection. Apply an MHTC, if applicable.

MHTC → p.ej. Bonferroni

simple random sample



From Reality → Data : Omissions
From Data → Reality : Commissions

Table 4 Results of the defective count and p-values by quality control

Quality control	Number of nonconforming items	Sample size (n)	p-value	
			Binomial	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)

α/#(simultaneous tests)

α = 5%

α/6 = 0,0083

Fail → **rejection**

Conclusions

- The quality of the BIM data is important. ISO 19157-1 must be adapted for dealing with BIM data and other data types

- There is a great previous experience in spatial data quality.
- The quality elements proposed for spatial data are directly applicable to BIM data
- New data quality elements are needed.
- In the case of requiring quality controls by sampling (“as built” case), adequate statistical models are available.

- Statistical models allow flexibility: control numerous data quality elements, give importance to some aspects or others, consider different levels of conformity at the same time, etc.
- The presented model can be applied in different phases of a BIM project.

Thank you

Count based quality control of "As Built" BIM datasets using the ISO [illegible] framework

Francisco Javier Ariza-López, José Rodríguez-Avi, Juan Francisco Reinoso Gordo, Iñigo Ariza-López

fjariza @u[illegible]es, [illegible]jaen.es, j[illegible]einoso@ugr.es; inigoariza@us.es

Universidad de Jaén; Paraje de las Lagunillas S/N, E-23.071-Jaén

Funded by the Project: MEMCALIG (Modelización Estadística de Matrices de Confusión en Calidad de la Información Geográfica)

GOBIERNO DE ESPAÑA

MINISTERIO DE ECONOMIA Y COMPETITIVIDAD

PROYECTO COFINANCIADO POR LA UNIÓN EUROPEA

FONDO EUROPEO DE DESARROLLO REGIONAL