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

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.

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/

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...

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/

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

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

Dataset, case 2: OMISSION

Object omission

Assignation omission

Information Quality Assessment for Facility Management

Puyan A. Zadeh a, Guan Wang a, Hasan B. Cakva a, Sheryl Staub-French a, Rachel Pottinger b
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.
Count based quality control of “As Built” BIM datasets using the ISO 19157 framework

Proposal of DQ elements
Logical consistency

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

- Automatic
- Semi automatic
- Manual (visual)
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
Count based quality control of “As Built” BIM datasets using the ISO 19157 framework

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.

**Positional correctness:** The location is correct or incorrect.

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

Proposal of DQ elements

Thematic accuracy

Reference (ground truth)

Thematic accuracy

Information Quality Assessment for Facility Management
Puyan A. Zadeh, Guan Wang, Hasan B. Cavka, Sheryl Staub-French, Rachel Pottinger
Advanced Engineering Informatics 33 (2017) 181–205
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.
Count based quality control of “As Built” BIM datasets using the ISO 19157 framework

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.
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 ($H_0$)

$H_0$:
- Binomial Model / Hypergeometric Model
- Multinomial Model / Hypergeometric multivariate model

($\text{infinite population} / \text{finite population}$)

$H_0$: Statistical Power

Tolerance, signification, power (risks: type I & II), $\text{Tol, } \alpha, \beta$

Kind of variables

- Quantitative
- Qualitative
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:
  
  \[ \text{DQU} = \text{DQE} + \text{CoI} \]

- **DQM** → Data quality measure (Annex D of ISO 19157-1)
- **EM** → Evaluation (assessment) method.
- **QL** → Quality level (compliance level).
- **QC** → Quality control:
  
  \[ \text{Quality control QC} = \text{DQU} + \text{EM} + \text{QL} \]
Proposal of a statistical control

Process’s steps are:
• Determine the Cols. (*)
• Determine the population sizes of the Cols.
• 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
Count based quality control of “As Built” BIM datasets using the ISO 19157 framework

Example

The case

- Basement (garage)
- Ground floor (2 commercial premises)
- P1 (2 apartments)
- P2 (2 apartments)
- Roof with storage
Count based quality control of “As Built” BIM datasets using the ISO 19157 framework

**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

<table>
<thead>
<tr>
<th>Group</th>
<th>Categories of interest</th>
<th>Cases (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elements</td>
<td>C1=Doors and windows</td>
<td>119</td>
</tr>
<tr>
<td>Elements</td>
<td>C2=Bathrooms and Kitchens</td>
<td>14</td>
</tr>
<tr>
<td>Elements</td>
<td>C3=Balconies and terraces</td>
<td>29</td>
</tr>
<tr>
<td>Elements</td>
<td>C4=Other rooms</td>
<td>18</td>
</tr>
<tr>
<td>Elements</td>
<td>C5=Living rooms and bedrooms</td>
<td>16</td>
</tr>
<tr>
<td>Elements</td>
<td>C6=Common zones</td>
<td>6</td>
</tr>
<tr>
<td>Elements</td>
<td>C7=Enclosures (walls)</td>
<td>179</td>
</tr>
<tr>
<td>Elements</td>
<td>C8=Slabs and paving</td>
<td>25</td>
</tr>
<tr>
<td>Elements</td>
<td>C9=Pillars</td>
<td>105</td>
</tr>
<tr>
<td>Elements</td>
<td>C10=Sales unit</td>
<td>6</td>
</tr>
<tr>
<td>Elements</td>
<td>C11=Interior walls</td>
<td>200</td>
</tr>
<tr>
<td>Facilities</td>
<td>C12=Electricity installation</td>
<td>7</td>
</tr>
<tr>
<td>Facilities</td>
<td>C13=Heating and air conditioned installations</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>731</td>
</tr>
</tbody>
</table>
Count based quality control of “As Built” BIM datasets using the ISO 19157 framework

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)

<table>
<thead>
<tr>
<th>Data quality units</th>
<th>Cases in the population (N)</th>
<th>Sample size (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQU1=Completeness of elements</td>
<td>511</td>
<td>50</td>
</tr>
<tr>
<td>DQ = Commission + omission</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col = C1 + C2 + ... + C10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DQU2=Completeness of facilities</td>
<td>182</td>
<td>40</td>
</tr>
<tr>
<td>DQ = Commission + omission</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col = C11 + C13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DQU3=Shape Fidelity</td>
<td>1605</td>
<td>160</td>
</tr>
<tr>
<td>DQ = Fidelity in shape</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col = C1 + C2 + ... + C10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DQU4=Attributes of elements</td>
<td>462</td>
<td>50</td>
</tr>
<tr>
<td>DQ = Correction of non-quantitative attributes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col = C1 + C2 + ... + C10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DQU5=Attributes of installations</td>
<td>491</td>
<td>50</td>
</tr>
<tr>
<td>DQ = Correction of non-quantitative attributes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col = C12 + C13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DQU6=Shape Fidelity of walls</td>
<td>200</td>
<td>20</td>
</tr>
<tr>
<td>DQ = Fidelity in shape</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col = C11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Valores de α recomendados para rechazar $H_0$: $\pi = \pi_0$ cuando el verdadero $\pi_0 = \pi_0 + \delta$ con un tamaño de error de Tipo I del 5% y un tamaño de error de Tipo II del 10%.

<table>
<thead>
<tr>
<th>Distancia</th>
<th>Probabilidad bajo $H_0$ $(\pi_0)$</th>
<th>$1%-\delta$</th>
<th>$3%-\delta$</th>
<th>$5%-\delta$</th>
<th>$8%-\delta$</th>
<th>$9%-\delta$</th>
<th>$15%-\delta$</th>
<th>$20%-\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1178</td>
<td>2828</td>
<td>4394</td>
<td>6610</td>
<td>8001</td>
<td>11176</td>
<td>13923</td>
<td></td>
</tr>
<tr>
<td>0.03</td>
<td>1920</td>
<td>3840</td>
<td>5560</td>
<td>7980</td>
<td>9500</td>
<td>12960</td>
<td>15930</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>880</td>
<td>1580</td>
<td>2210</td>
<td>3080</td>
<td>3620</td>
<td>4840</td>
<td>5890</td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>440</td>
<td>730</td>
<td>980</td>
<td>1320</td>
<td>1520</td>
<td>1990</td>
<td>2380</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>320</td>
<td>510</td>
<td>670</td>
<td>890</td>
<td>1020</td>
<td>1310</td>
<td>1560</td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>180</td>
<td>270</td>
<td>340</td>
<td>440</td>
<td>490</td>
<td>620</td>
<td>720</td>
<td></td>
</tr>
<tr>
<td>0.20</td>
<td>120</td>
<td>170</td>
<td>210</td>
<td>270</td>
<td>300</td>
<td>360</td>
<td>420</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3451</td>
<td>5350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Count based quality control of “As Built” BIM datasets using the ISO 19157 framework

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>
<thead>
<tr>
<th>Quality control</th>
<th>Data quality unit</th>
<th>Data Quality Measure and ID*</th>
<th>Conformity level (Maximum proportion of defects)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QC1</td>
<td>DQU1</td>
<td>Rate of excess items (ID=3) + Rate of missing items (ID=7)</td>
<td>1%</td>
</tr>
<tr>
<td>QC2</td>
<td>DQU2</td>
<td>Rate of excess items (ID=3) + Rate of missing items (ID=7)</td>
<td>3%</td>
</tr>
<tr>
<td>QC3</td>
<td>DQU3</td>
<td>Rate of unfaithful items (ID=**)</td>
<td>5%</td>
</tr>
<tr>
<td>QC4</td>
<td>DQU4</td>
<td>Rate of incorrect attribute values (ID=67)</td>
<td>10%</td>
</tr>
<tr>
<td>QC5</td>
<td>DQU5</td>
<td>Rate of incorrect attribute values (ID=67)</td>
<td>10%</td>
</tr>
<tr>
<td>QC6</td>
<td>DQU6</td>
<td>Rate of unfaithful items (ID=**)</td>
<td>80%, 15%, 5%***</td>
</tr>
</tbody>
</table>

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

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 $\rightarrow$ p.ej. Bonferroni

$\alpha/#$(simultaneous tests)

$\alpha = 5\%$

$\alpha/6 = 0.0083$

From Reality $\rightarrow$ Data $:\$ Omissions
From Data $\rightarrow$ Reality $:\$ Commissions

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

<table>
<thead>
<tr>
<th>Quality control</th>
<th>Number of nonconforming items</th>
<th>Sample size (n)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>QC1</td>
<td>0</td>
<td>50</td>
<td>1.000</td>
</tr>
<tr>
<td>QC2</td>
<td>5</td>
<td>40</td>
<td>0.179</td>
</tr>
<tr>
<td>QC3</td>
<td>11</td>
<td>160</td>
<td>0.560</td>
</tr>
<tr>
<td>QC4</td>
<td>5</td>
<td>50</td>
<td>0.966</td>
</tr>
<tr>
<td>QC5</td>
<td>2</td>
<td>50</td>
<td>0.0004</td>
</tr>
<tr>
<td>QC6</td>
<td>7.1(*)</td>
<td>20</td>
<td>0.0236</td>
</tr>
</tbody>
</table>

(*): The number of items per class is: 12 (good), 7 (acceptable), 1 (unacceptable)

Fail $\rightarrow$ 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.
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, jreinoso@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)