DATA QUALITY EVALUATION in NSDI

Cemre YILMAZ, Çetin CÖMERT
Karadeniz Technical University, Geomatics Engineering
Trabzon, TURKEY

cemre.yilmaz@ktu.edu.tr

www.harita.ktu.edu.tr
Presentation Content

- Aim
- Related Studies
- Case Study
- Conclusion
Aim of Study

• How to evaluate spatial data quality?

• FOSS?

• A preliminary application
• The quality of the data provided is important in terms of the decision-making process and the accuracy of transactions.

• What is the quality of data acquired by the user?
Spatial Data Quality

Data Quality elements ISO/DIS 19113 (2005)

**Completeness** - presence and absence of features, their attributes and relationships

**Logical consistency** - degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical)

**Positional accuracy** – accuracy of the position of features

**Temporal accuracy** – accuracy of the temporal attributes and temporal relationships of features

**Thematic accuracy** – accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships
Logical Consistency (ISO/DIS 19113, 2005)

- **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
Logical Consistency

Logical consistency, may be defined as the degree of consistency of the data with respect to its specification. (Mostafavi, 2004)

One of the most important data quality elements.
Related Studies

Specifications

• INSPIRE Data Specifications

• Standards and Specifications of the National Topographic Database (Canada)

• Turkish Large Scale Map and Map Information Production Regulation (LSMMPR)
1Spatial Group has developed a “standard-based rules language".

They used Radius Studio software to implement this language. In INSPIRE Annex 1 testing process, consistency of the datasets to its specifications have been tested.

The Online Data Validation Service validates spatial data against the different kinds of standard rules including, geometric, polygon, network and Irish MapRoad check rules.
Overview of the main Road Transport Networks objects (INSPIRE Data Spec., 2010)

- In a Transport Networks data set which contains nodes, these nodes shall only be present where Transport Links connect or end.

- Transport link ends shall be connected wherever an intersection exists between the real world phenomena they represent. No connections shall be created at crossing network elements when it is not possible to pass from one element to another.
Related Studies

Illustration – Example of use of elements forming the Road Transport Network

(INSPiRe Data Spec., 2010)
Related Studies

Standards and Specifications of the National Topographic Database (Canada Geomatics)

Connection relation between different themes
For the connection relation, if the objects are from two different themes, an intersection point must be added to the geometric representation of the objects involved.

In the case where the objects are from the same theme, the objects should be segmented at the connection point.
• A rule base is created to test the inconsistencies according to the rules in specifications.

• Prolog, major logic programming language, was used to create the rule base. (Mostafavi, 2004)
Case Study

Turkish Large Scale Map and Map Information Production Regulation (LSMMMPR)

Currently, there is not enough rule about the relations of entities in the regulation.

Only the entity types, their attributes and geometry types are defined in the regulation.

Although there is no rule about the relations in the specification, dataset needs to be tested, and if needed new rules should be created according to the errors in the dataset.
**Rule 1** - Contours do not intersect with the details which are double lined such as roads, river, canal, arc etc. and closed type details. (LSMMPR, 2008)

**Rule 2** - One of the most commonly faced error in the base map is the non continuous road entities when they have common borders with the building details.
A street view in base map of Trabzon with building and road details
Case Study

A street view in base map of Trabzon with only road details
Case Study

• A rule base has been created with Spatial YAP Prolog.

• To test the results of algorithms created, a new sample database created with Quantum GIS software.

• Dataset imported to MySQL with ogr2ogr library.
Spatial-YAP Prolog

- Logic based Geographic Information System.

- Spatial predicates + MySQL database with Yap Prolog.

- M:-DDAS (MySQL Deductive Database System)
Case Study

MySQL Deductive Database System (MYDDAS), (LIACC, University of Porto)
A sample test database with errors in Quantum GIS software
Case Study

Data imported to MySQL query browser
Rule base created to test inconsistencies

Case Study
Case Study

Building, park and road layers

Park and road layers
% Find the boundary of the building
dis(ID,C):-bina(ID,X),ogc_boundary(X,C).

% Finding the vertices of road
koseler(ID,P):-yol(ID,X),ogc_boundary(X,P).

% Creating buffer to building
bufo(ID,B):-bina(ID,X),D is 0.1,ogc_buffer(X,D,B).

% If exist finding the vertices of road close to
yakinkose(YolID,BinaID,Q):-koseler(YolID,P),bufo
(BinaID,B),ogc_intersects(P,B),ogc_intersection(P,B,Q).

% Creating line with close points
doldurma(YolID,BinaID,L):-yakinkose
(YolID,BinaID,Q),ogc_convex_hull(Q,L).

% Intersection of this line with the boundary of the building
sonkesisisim(YolID,BinaID,M):-doldurma(YolID,BinaID,L),dis
(BinaID,Y),ogc_intersection(L,Y,M).

% If there is an intersection between line and boundary of the
building, there is an error!
sonkesisiyo(YolID,BinaID):-sonkesisisim
(YolID,BinaID,M),ogc_length(M,L),L>0.
Ubuntu Xterm interface to execute Spatial yap and query for inconsistencies
### Case Study

#### Attribute Table - yol :: 1 / 8 feature(s) selected

<table>
<thead>
<tr>
<th>uzunluk</th>
<th>genislik</th>
<th>yol_tipi</th>
<th>aciklama</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NULL</td>
<td>5</td>
<td>NULL</td>
</tr>
<tr>
<td>1</td>
<td>NULL</td>
<td>3</td>
<td>NULL</td>
</tr>
<tr>
<td>2</td>
<td>NULL</td>
<td>5</td>
<td>NULL</td>
</tr>
<tr>
<td>3</td>
<td>NULL</td>
<td>4</td>
<td>NULL</td>
</tr>
<tr>
<td>4</td>
<td>NULL</td>
<td>1</td>
<td>NULL</td>
</tr>
<tr>
<td>5</td>
<td>NULL</td>
<td>2</td>
<td>NULL</td>
</tr>
<tr>
<td>6</td>
<td>NULL</td>
<td>6</td>
<td>NULL</td>
</tr>
<tr>
<td>7</td>
<td>NULL</td>
<td>5</td>
<td>NULL</td>
</tr>
</tbody>
</table>
## Case Study

<table>
<thead>
<tr>
<th>Katadedi</th>
<th>Malik</th>
<th>Ä°l</th>
<th>Ä°lce</th>
<th>Mahalle</th>
<th>Cadde</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
</tr>
<tr>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
</tr>
<tr>
<td>2</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
</tr>
<tr>
<td>3</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
</tr>
<tr>
<td>4</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
</tr>
<tr>
<td>5</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
</tr>
<tr>
<td>6</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
</tr>
<tr>
<td>7</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
</tr>
</tbody>
</table>
We create rules in a rule base according to the LSMMPR. Creating new rules will be the subject of future studies.

MyDDAS gives good results with small amount of data.

1/1000 scaled base map of Trabzon also tested, but it takes very long time to answer this query.

This is because of the lack of Spatial Index in MyDDAS.

User Defined Indexing (UDI) is in developmental stage. This indexing can be used to query such large databases.
Thanks......