ABSTRACT:

Nowadays it is easy to find many data sources for various regions around the globe. In this 'data overload' scenario there are few, if any, information available about the quality of these data sources. In order to easily provide these data quality information we presented the architecture of a web service for the automation of quality control of spatial datasets running over a Web Processing Service (WPS). For quality procedures that require an external reference dataset, like positional accuracy or completeness, the architecture permits using a reference dataset. However, this reference dataset is not ageless, since it suffers the natural time degradation inherent to geospatial features. In order to mitigate this problem we propose the Time Degradation & Updating Module which intends to apply assessed data as a tool to maintain the reference database updated. The main idea is to utilize datasets sent to the quality evaluation service as a source of 'candidate data elements' for the updating of the reference database. After the evaluation, if some elements of a candidate dataset reach a determined quality level, they can be used as input data to improve the current reference database. In this work we present the first design of the Time Degradation & Updating Module. We believe that the outcomes can be applied in the search of a full-automatic on-line quality evaluation platform.

1. INTRODUCTION

Today there are many geospatial data sources available. Space agencies and imaging companies launch new Earth observation satellites. Anyone can buy a drone with a camera in a supermarket. National mapping agencies release new products constantly. Thus, there are many data, however there are few quality information about these data. In this scenario we have presented a web service facing automatic quality assessment through web services (Xavier et al., 2015b). Though automatic, some quality evaluation procedures require an external dataset, the universe of discourse. Notwithstanding, this reference dataset experience the natural time degradation process which irredeemably decreases its quality along the time.

This paper discusses how to maintain the reference database up-to-dated. We need a solution that makes possible to useful the utopic dataset proposed by Cooper and Peled (2001), a dataset that 'gets updated automatically, continuously and transparently to the user'. On other moment the authors called 'nirvana' these automatic, continuous, and transparent updates (Peled and Cooper, 2004).

We propose to utilize the datasets sent to the quality evaluation service as a source of candidate elements that should be applied to update the reference database. In our architecture this solution is named Time Degradation & Updating Module.

This approach is built over three concepts: rules, quarantine database, and conflation. The rules are based on quality assessment and are defined to specify who can update what, when and how. The quarantine database is used to store the candidate data elements before they can be fusioned with current reference database and also to store, if exists, elements of the current database that should be rejected. Lastly, the conflation system executes the fusion between the new, updated data, and the remaining elements of the reference dataset.

This paper is structured as follows. Section 2 provides a brief explanation about temporal issues in geographic information systems (GIS), particularly models facing temporal change. Section 3 describes the first design of the Time Degradation & Updating Module. Section 4 presents some use cases in order to illustrate the theoretic solution. Lastly, section 5 brings some conclusions and future work.

2. BACKGROUND

Morris et al. (2000) argues that 'time' has been a thorn in GIS developers' flesh for a long time. Since the pioneer work of Basoglu and Morrison (1978) the integration of space and time remains an active research topic (Long and Nelson, 2013). Worboys (1994) claims that geospatial data have an 'inseparable' temporal component, i.e., systems that do not take into account the temporal component of spatial features are incomplete, limited systems.

Some authors defined methods to handle with temporal changing in geospatial features. In this direction, Langran and Chrisman (1988) presented two methods of representing geographic change: base map with overlays and space-time composite. In the first method the changes in spatial data should be registered as an overlay. So, the final map would be created by superimposing all changing overlays over the original base map. The second method, spatial-time composite, is a variation of the first one (overlays) and it is based on the study of Chrisman (1984). In this approach, each change implies in break of original object into two: the changed and the unchanged part with their own history.
Regarding land information systems, Price (1989) observed that a land parcel is an object with its own lifetime, so it is necessary to record the changes over time. The author proposed a method that keeps track of changes from creation to destruction.

Other authors investigated how to model temporal changes. Ramirez (1997) presented four possible situations for a feature along the time: (1) the feature is new; (2) the feature has not changed; (3) the feature has changed; and (4) the feature no longer exists. Lemarié and Badard (2001) presented six updating type that may occur in an updating process: creation, destruction, semantic modification, geometric modification, semantic and geometric modification, and aggregation. In a more recent work, Qi et al. (2010) developed a method for detecting changes between an updated larger-scale map and a more recent work, Qi et al. (2010) developed a method for detecting changes between an updated larger-scale map and a more recent work, Qi et al. (2010) developed a method for detecting changes between an updated larger-scale map and a more recent work, Qi et al. (2010) developed a method for detecting changes between an updated larger-scale map and a more recent work.

The type of change represents the different changing possibilities that may occur with reference data. Based on the quality loss (↓Q) occurs naturally in the dataset, but in different manners for each type of object, like suggest the product specification of Ordnance Survey (2004). So this quality loss ↓Q can be quantified using the equation (2).

\[
\downarrow Q(t) = \Delta t \sum \rho_e
\]

where \(\rho_e\) is the quality loss rate for objects of data type \(e\).

On the other hand, a quality improvement only occurs when an updating procedure is executed. These processes can occur in the whole dataset, or in some specific region, or these can reach all types of elements, or a subset of them. Moreover, this improvement depends on the effectiveness of the method of updating, and it affects only the items that are 'wrong'. Therefore, the quality improvement ↑Q can be quantified by means of equation (3).

\[
\uparrow Q(t) = \sum_{z} W_z \eta_z \sum_d d_z(t)
\]

where \(z\) is an updating zone, \(W_z\) is the weight of zone \(z\) (\(\sum W_z = 1\)), \(\eta_z\) is the performance of update in zone \(z\), \(d_z(t)\) represents the elements of a data type, and \(d_z(t)\) represents the subset of objects in class \(e\) that can improve their quality (as percentage) in time \(t\).

Hence the temporal evolution model can be summarized in the equation (4). This model is generic and it can be applied to any quality element, like completeness or positional accuracy.

\[
Q(t_i) = Q(t_0) - \Delta t \sum \rho_e + \sum_{z} W_z \eta_z \sum d_z(t)
\]

### 3.1 Temporal evolution model

Ariza (2013) proposed a temporal evolution model for geospatial features based on the fact that the quality of geoinformation is valid for some period, since it is an ever-changing world. The main idea is that the quality \(Q\) in a time \(t\) is the quality in time \(t_0\) minus the quality degradation in this period (\(\Delta t\)) plus some gain in quality by updating (\(\uparrow Q\)), as show in equation (1).

\[
Q(t) = Q(t_0) - \Delta t \uparrow Q + \Delta t \downarrow Q
\]

where the values \(Q\), \(\Delta Q\), and \(\uparrow Q\) are expressed as a percentage.

3. **TIME DEGRADATION & UPDATING MODULE**

In a previous work (Ariza-López et al., 2015) we presented the architecture of a web service for the automation of positional quality control of geospatial datasets. In other study we have demonstrated the feasibility of this solution for the case of positional accuracy (Xavier et al., 2015b). Our architecture foresees that the Data Access tier is able to mitigate the natural time degradation that occurs when the reference local data is getting old. In the proposed architecture, the Time degradation & updating module plays this role.

In this section we detail this module by presenting its concepts: a temporal evolution model for quality of geospatial objects, a changing model to quantify modifications, and the rules applied to maintain the reference database up-to-date.

Time degradation & updating module, or Update DB for short, also deals with conflation techniques by calling the Conflation module that effectively executes the merging of data. This last module is not detailed in this work.

### 3.2 Changing model

The Update DB module is built over three key-concepts: rules, quarantine DB, and conflation. In order to quantify if an object should change we have defined a changing model on top of six types of elements, or a subset of them. Moreover, this improvement depends on the effectiveness of the method of updating, and it affects only the items that are 'wrong'. Therefore, the quality improvement ↑Q can be quantified by means of equation (3).

The type of change represents the different changing possibilities that may occur with reference data. Based on the types of discrepancies presented by Qi et al. (2010), we adopt three main categories of changing:
• 1:0 changes: represents create or destroy objects in the reference DB;
• 1:1 changes: internal changes that occur with a single object. These are changing in the properties of an object, like a geometric and/or attribute change;
• m:n changes: these kind of change evolves combine or split objects, which can be common when using different scales.

We have developed a basic quality model to assess some quality elements that are relevant in updating tasks. This model is basic because it can be extended to accommodate other quality elements and measures. The model uses the concepts and measures from ISO 19157:2013 (ISO, 2013) and it is presented in Table 1. A dataset is considered valid if it is approved in two tests: format consistency and conceptual consistency. Otherwise it is considered invalid and cannot be used in the Update DB.

<table>
<thead>
<tr>
<th>Category / Element</th>
<th>Measure (identifier in ISO 19157)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness</td>
<td></td>
</tr>
<tr>
<td>Commission</td>
<td>rate of excess items (3)</td>
</tr>
<tr>
<td>Omission</td>
<td>rate of missing items (7)</td>
</tr>
<tr>
<td>Logical consistency</td>
<td></td>
</tr>
<tr>
<td>Conceptual consistency</td>
<td>conceptual schema non-compliance (8)</td>
</tr>
<tr>
<td>Domain consistency</td>
<td>value domain non-conformance rate (18)</td>
</tr>
<tr>
<td>Format consistency</td>
<td>physical structure conflicts (119)</td>
</tr>
<tr>
<td>Topological consistency</td>
<td>rate of self-intersect objects (16)</td>
</tr>
<tr>
<td></td>
<td>rate of self-overlap objects (17)</td>
</tr>
<tr>
<td>Positional accuracy</td>
<td></td>
</tr>
<tr>
<td>Absolute accuracy</td>
<td>rate of positional uncertainties above a given threshold (31)</td>
</tr>
<tr>
<td>Positional accuracy</td>
<td></td>
</tr>
<tr>
<td>Thematic accuracy</td>
<td>rate of incorrect attribute values (67)</td>
</tr>
<tr>
<td>Non-quantitative attribute correctness</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Basic quality model

The type of territory where occurs the changing/updating should also be considered. There are some regions more susceptible to change than another, as shown by some studies (Ordnance Survey, 2004). So, we can classify the regions according its susceptibility to change:

• Expansion areas: those areas under expansion, mainly the border of urban areas (high possibility of change);
• Urban: urban consolidated area (medium possibility of change);
• Rural: other areas (low possibility of change).

Some metadata elements about the assessed dataset can be used in the updating process. These are the date and scale of sent dataset. This information is optional.

After taking into account the presented six variables, we can draw our changing model. The changing of an object can be quantified by means of a 7-tuple structure: (refID, Cr, Ds, Gm, At, Cb, Sp). The first tuple (refID) represents the object's identifier in the reference database. The other tuples represent different values which indicate what should occur with the feature:

• Cr: indicates that the object should be created;
• Ds: indicates that the object should be destroyed;
• Gm: indicates that the object's geometry should be updated;
• At: indicates that the object's attributes should be updated;
• Cb: indicates that the object should be combined with other objects;
• Sp: indicates that the object should be split into other objects.

The users' reputation plays an important role in the system. The system decides in who it trust. The user's reputation is managed according some user groups:

• Anonymous users: the system permits anonymous users, so anyone can send a dataset to be evaluated. However, an anonymous dataset does not have any recognized reputation;
• Registered users: the registered users may receive more 'reputation' when their datasets get a valid evaluation;
• Trusted users: those users that have the higher level of trustworthiness.
into account the six variables: type of change, quality model, users’ reputation, territory type, date of product and scale of product, as can be seen in Figure 1. This score is used in the updating rules described in next subsection.

3.3 Rules

The rules define the possible ‘movements’ of data among three databases of the architecture: Reference (Ref), Quarantine (QA), and History (Hist). Reference database is used as the universe of discourse in quality evaluation and it should be updated by sent data. Quarantine database stores the objects before they can be inserted, removed, or updated in the reference. History database maintains the backup of changed objects.

There are three kinds of movement for the considered data (Figure 2): insert to Quarantine (Insert to QA), Quarantine to Reference (QA to Ref), and Quarantine to History (QA to Hist). Each movement is regulated by its own rules.

![System architecture and data movements between databases.](image)

When a dataset to assess arrives at the WPS interface, it is evaluated in the Evaluation module. After the evaluation the dataset becomes available to the Update DB module, where the first rules are applied. The rules for Insert to QA are the following:

1. Invalid datasets are discarded (non-conformance with format consistency or conceptual consistency);
2. If the date of product is previous to reference DB (inside its bounding box), this product is discarded;
3. Objects with an invalid evaluation are discarded (invalid domain consistency or topological consistency);
4. The score of remained objects is calculated according the changing possibilities listed in Table 2;
5. When a create occurs, the system should look at the QA in order to find its homologous.

### Table 2. Insert to QA changing tuples.

<table>
<thead>
<tr>
<th>Change ID Value assigned to each tuple</th>
</tr>
</thead>
<tbody>
<tr>
<td>No change₁ refID -score for Cr – Sp</td>
</tr>
<tr>
<td>Commission newID score, for Cr, 0 for others</td>
</tr>
<tr>
<td>Omission refID score, for Ds, 0 for others</td>
</tr>
<tr>
<td>Positional accuracy² refID score, for Gm, 0 for others</td>
</tr>
<tr>
<td>Thematic accuracy² refID score, for At, 0 for others</td>
</tr>
<tr>
<td>Multi-case mₕ&lt;n refID score, for Cb, 0 for others</td>
</tr>
<tr>
<td>Multi-case mₕ&gt;n refID score, for Sp, 0 for others</td>
</tr>
</tbody>
</table>

Notes: (1) score uses all quality elements. The other changing possibilities, no. (2) Positional and thematic accuracy can occur in the same object.

Using some regular period (e.g. daily), the system should check if some candidate changing should occur. The rules for the QA to Ref are the following:

6. Considering all 7-tuples grouped by its refID, the system calculates the final score for each changing possibility. If some score is greater than a given tolerance, the Condilat module is called in order to execute the change in the reference using the data in QA. If a destroy occurs, the object deleted from Ref is inserted in QA;
7. Any change applied to the reference DB generates a new ‘date’ for this data in the affected region;
8. Any change applied should reward the source registered user with more reputation.

Finally, the last set of rules is applied for the QA to Hist movement:

9. After some period (e.g. one year), any 1:0 change (create/destroy) shall be moved from QA to Hist;
10. After some period (e.g. 200 days), any 1:1 change (alter geometry/attribute) shall be moved from QA to Hist;
11. After some period (e.g. 200 days), any m:n change (combine/split) shall be moved from QA to Hist.

The examples above (rules 9-11) for valid periods were extracted from the temporal validity of some feature types according to emergency services in New Zealand (LINZ, 2004).

### 4. USE CASES

In order to develop our proposal for the Update DB module, in this section we present two feasible use cases that may occur in the quality evaluation service. The first one deals with datasets with similar level of detail in relation to the reference database, while the second case deals with datasets with distinct levels of detail.

#### 4.1 Use case 1: same scale

In this scenario, our quality evaluation service receives three datasets to be evaluated which have a level of detail similar to the reference database. The first one (E1) was sent by a registered user and it reaches the changing score of 0.81. The second (E2) was sent by an anonymous user and it has a 0.65 score. The last evaluation (E3) was sent by a trusted user and its calculated changing score is 1.32. The reference database is
represented in Figure 3a, and the assessed datasets are represented in Figure 3b for E1, 3c for E2 and 3d for E3.

![Figure 3a](image1.png)  ![Figure 3b](image2.png)  ![Figure 3c](image3.png)  ![Figure 3d](image4.png)

**Figure 3. Use case 1 datasets: (a) reference, (b) evaluation E1, (c) evaluation E2, (d) evaluation E3.**

The dataset in evaluation E1 differs from the reference database in three aspects: new building B3, alter geometry in road R2, and alter the name (attribute) of lake L1. The differences in the dataset E2 are: alter the geometries of roads R1 and R2, destroy the building B1, and alter the name of lake L1. Lastly, the third dataset (E3) introduces other three changes: new building B3, destroy B1, and alter the geometry of lake L1.

After receive these three datasets to be evaluated, Insert to QA rules (1-5) generated the list of changing tuples listed in Table 3 (without combine and split for simplification). The small fluctuations in scores are due to the selection of different quality results according to the type of change. Each tuple was calculated using the values described in Table 2.

The next step is to verify if the Quarantine data should migrate to Reference, so the QA to Ref rules (6-8) are applied. The system computes what should occur with changing features grouping by each identifier. Considering a changing threshold equal to one, the results are shown in Table 4.

**Table 3. Changing tuples after assessment.**

<table>
<thead>
<tr>
<th>Evaluation</th>
<th>ID</th>
<th>Cr</th>
<th>Ds</th>
<th>Gm</th>
<th>At</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1 – registered</td>
<td>B1</td>
<td>-0.81</td>
<td>-0.81</td>
<td>-0.81</td>
<td>-0.81</td>
</tr>
<tr>
<td></td>
<td>B2</td>
<td>-0.81</td>
<td>-0.81</td>
<td>-0.81</td>
<td>-0.81</td>
</tr>
<tr>
<td></td>
<td>B3</td>
<td>0.87</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>R1</td>
<td>-0.81</td>
<td>-0.81</td>
<td>-0.81</td>
<td>-0.81</td>
</tr>
<tr>
<td></td>
<td>R2</td>
<td>0</td>
<td>0</td>
<td>0.84</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>L1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.83</td>
</tr>
<tr>
<td>E2 – anonymous</td>
<td>B1</td>
<td>0</td>
<td>0.68</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>B2</td>
<td>-0.65</td>
<td>-0.65</td>
<td>-0.65</td>
<td>-0.65</td>
</tr>
<tr>
<td></td>
<td>B3</td>
<td>-0.65</td>
<td>-0.65</td>
<td>-0.65</td>
<td>-0.65</td>
</tr>
<tr>
<td></td>
<td>R1</td>
<td>0</td>
<td>0</td>
<td>0.69</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>R2</td>
<td>0</td>
<td>0</td>
<td>0.69</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>L1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.71</td>
</tr>
<tr>
<td>E3 – trusted user</td>
<td>B1</td>
<td>0</td>
<td>1.35</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>B2</td>
<td>-1.32</td>
<td>-1.32</td>
<td>-1.32</td>
<td>-1.32</td>
</tr>
<tr>
<td></td>
<td>B3</td>
<td>1.36</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>R1</td>
<td>-1.32</td>
<td>-1.32</td>
<td>-1.32</td>
<td>-1.32</td>
</tr>
<tr>
<td></td>
<td>R2</td>
<td>-1.32</td>
<td>-1.32</td>
<td>-1.32</td>
<td>-1.32</td>
</tr>
<tr>
<td></td>
<td>L1</td>
<td>0</td>
<td>0</td>
<td>1.37</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 4. Final score for each changing possibility.**

<table>
<thead>
<tr>
<th>ID</th>
<th>Cr</th>
<th>Ds</th>
<th>Gm</th>
<th>At</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>-0.81</td>
<td>1.22</td>
<td>-0.81</td>
<td>-0.81</td>
<td>1.22 &gt; 1, destroy</td>
</tr>
<tr>
<td>B2</td>
<td>-2.78</td>
<td>-2.78</td>
<td>-2.78</td>
<td>-2.78</td>
<td>No change</td>
</tr>
<tr>
<td>B3</td>
<td>1.58</td>
<td>-0.65</td>
<td>-0.65</td>
<td>-0.65</td>
<td>1.58 &gt; 1, create</td>
</tr>
<tr>
<td>R1</td>
<td>-2.13</td>
<td>-2.13</td>
<td>-1.44</td>
<td>-2.13</td>
<td>No change</td>
</tr>
<tr>
<td>R2</td>
<td>-1.32</td>
<td>-1.32</td>
<td>0.21</td>
<td>-1.32</td>
<td>0.21 &lt; 1, no change</td>
</tr>
<tr>
<td>L1</td>
<td>0</td>
<td>0</td>
<td>1.37</td>
<td>1.54</td>
<td>1.37 &gt; 1 and 1.54 &gt; 1, Change geometry and attribute</td>
</tr>
</tbody>
</table>

Table 4 indicates what should occur with the features regarding the QA to Ref movement. The building B1 has to be destroyed because the final changing score have reached the changing threshold (=1). Despite the registered user has not informed any change in this object, the anonymous and trusted user have, so the system decision is to destroy the object based on the final score. Similar procedure is executed for the other objects. It can be observed that despite the road R2 has been modified (geometry) by two users (registered and anonymous), the unchanged status informed by the trusted user has reduced its changing score to a value below the threshold. Thereafter the object R2 remains unaltered.

4.2 Use case 2: different scales

The reference database in this scenario for the considered regions is compound by buildings and roads (Figure 4a). The quality control service receives two datasets to be evaluated. The first one has a more detailed scale (Figure 4b), while the second presents less details (Figure 4c).

![Figure 4a](image5.png)  ![Figure 4b](image6.png)  ![Figure 4c](image7.png)

**Figure 4. Use case 2 datasets: (a) reference, (b) detailed, (c) less detailed.**

The more detailed dataset suggests that the three roads should receive a rotatory in their connection. This implies in three new objects (R4-R6), plus some geometric changes in existing roads (R1-R3). Other identified changing is in the geometry of building B1, by adding more vertices. The last alteration is the split of building B2 into two new objects B21 and B22.

On the other hand, the less detailed dataset did not be able to identify the new rotatory (R4-R6) due to its acquisition limitations. The dataset also suggests to combine the buildings B2 and B3 into only one (B2), what should be reject due the information provided by the more detailed dataset. Despite of the less detailed dataset has not given any accepted geometric changing, it has supplied a name for the unnamed road R3, an alter attribute change.
rules using a spatiotemporal algebra.

The proposed solution is built over three concepts: rules, quarantine database, and conflation in order to keep the reference database up-to-date. This approach presents interesting features. It uses a quality model based on ISO standards. After some time, the system will be able to determine the changing rates of affected classes. It could be applied to the temporal evolution model in order to identify when an update should be executed. In the course of time, we will be able to update the territory types by identifying new expansion areas.

This approach presents interesting features. It uses a quality model based on ISO standards. After some time, the system will be able to determine the changing rates of affected classes. It could be applied to the temporal evolution model in order to identify when an update should be executed. In the course of time, we will be able to update the territory types by identifying new expansion areas.

Since it is a first design, the solution requires refinements like: how should be calculated the score using the proposed six variables? Other limitation is that the approach does not take into account the relations among objects of distinct classes.

Future work includes: (1) define how to calculate the changing score; (2) detail and develop the Conflation module; (3) upgrade the quality model by testing new elements; and (4) describe the rules using a spatiotemporal algebra.

ACKNOWLEDGEMENTS

This work has been funded by the Ministry of Science and Technology (Spain) and the European Regional Development Fund under grant no. BIA2011-23217. The authors also wish to acknowledge the support of the Brazilian Army's Department of Science and Technology, and the support of the Regional Government of Andalucía (Spain) for the Research Group “Ingeniería Cartográfica” (TEP-164).

REFERENCES


5. CONCLUSIONS

In our ever-changing world the natural time degradation that suffer the geospatial data can easily leads us to a geographic misunderstanding. We should maintain our geospatial databases as up-to-date as possible. In this paper we have presented the first design of a solution to use third-party data as a source for updating a reference database used in quality evaluation procedures.

Figure 4. Use case 2 datasets: (a) reference, (b) more detailed, (c) less detailed.


