Exploring the Possibility of Semi-Automated Quality Evaluation of Spatial Datasets in Spatial Data Infrastructure

Amin Mobasheri

Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, 7500 AE Enschede, The Netherlands.
Email: mobasheri23795@itc.nl

Abstract. Over the past decades, World Wide Web technology has developed greatly. One of the most important outcomes of this technology is to share data in a worldwide domain. A considerable amount of available data have spatial components and are hence called spatial data. The level of quality that spatial datasets conform plays an important role in their reliability for use in projects. This research aims to overview spatial data quality elements and select appropriate elements suitable for means of semi-automated quality evaluation. In this paper, the ISO 19100 series of standards for geographic information is used as basis for quality evaluation. The possibilities of use of different spatial data quality elements for semi-automated quality evaluation are explored and discussed. Finally, based on argumentation and reference to other research studies, a list of spatial data quality elements and sub-elements suitable for semi-automated quality evaluation of spatial datasets are presented.

1   Introduction

Quality of geo-information is often quoted as one of the critical issues in Geographic Information Science (GIS) [1,2]. Research about spatial data quality has taken place for over 30 years [2]. However, producing and using spatial data as well as using and offering spatial services over the web has now reached a mass market level, leading to new research challenges. The quality of spatial data has advanced significantly in the last years, due to quality checking by data providers, which ensures a good level of quality of datasets while capturing data [1]. Apart from the quality check in the data production phase, quality evaluation of spatial datasets during usage is also an important issue, which in practice has received less attention in the GI domain up to now [3]. Several organizations or individuals exist who discover spatial datasets on the World Wide Web for their projects, but do not know if they fit their purpose or not. In addition, they cannot find any comprehensive and complete software for spatial data quality evaluation. There are functionalities in some software applications

(e.g. ESRI or Intergraph products) that can be used for this purpose, but they have some disadvantages. For example, they can be platform-dependent and expensive. Apart from cost, most customers have a low level of knowledge about geo-information quality and its importance.

The solution is to prepare a service for customers so they can evaluate the quality of their spatial data without having knowledge about geo-data modeling or quality of geo-information. Since nowadays most datasets are available and transmitted via web servers in spatial data infrastructures (SDI), one of the best solutions is to let the evaluation process be carried out by a web processing service (WPS). Web services have several benefits, such as being standard-based, interoperable, and available at any time and any place. This research aims to overview data quality elements and select appropriate elements suitable for semi-automated quality evaluation in order to jumpstart the creation of a semi-automated quality evaluation model and web service to evaluate the quality of datasets in SDIs. The concept of this study is in line with the work being done by EuroGeographics to develop a web service for quality evaluation [4], where the aim of the research is to enable the automation of quality evaluation and conformance testing by taking accreditation principles and usability into account, as well as providing metadata guidelines for both discovery and evaluation of geo-information. However, the results of EuroGeographic's work are not published and the last news about their project goes back to the year 2010. In addition, Hunter, *et al.* [5] have argued the lack of quality evaluation services in the GI domain, although the necessary standardizations and infrastructures are mostly available to gain this goal.

2   It Is All About Quality!

Originally the term "quality" comes from the Latin *qualis*, which means "of what kind" [1]. ISO 9000 defines quality as "degree to which a set of inherent characteristics fulfils requirements" [6]. Also, the American Society for Quality states that in technical usage quality means the characteristics of a product or a service that bear on its ability to satisfy stated or implied needs [7]. Both last definitions refer to requirements as need or expectation. This is the main definition for quality in this research too. In this paper, quality is defined as a conditional and fully subjective attribute. Based on different requirements that people have, it may be understood differently.

Data quality is also a difficult term to define precisely. Different communities have different views and understandings of the subject, which causes confusion, a lack of harmonization of data across communities and omission of vital quality information [2,8]. In some of the literature data quality is defined as

features and characteristics of data that bear on its ability to meet the needs and requirements of the user [9]. By referring to this definition, in this research, the degree to which a dataset meets the requirements of a specific user implies the degree of its quality. No matter how many incorrect or missing values there may be in the dataset, if this is not against the users' needs then the dataset can still be considered to have an acceptable level of quality. This is the main definition of data quality based on "fitness for use".

In order to describe the quality of a spatial dataset, data quality elements and data quality overview elements can be used. A data quality element is a "quantitative component documenting the quality of a dataset" [10]. According to [10] five quantitative data quality elements exist, namely: positional accuracy, thematic accuracy, logical consistency, completeness, and temporal accuracy. Each data quality element has two or more sub-elements that describe a certain aspect of that data quality element [10]. For instance, logical consistency has four data quality sub-elements: conceptual consistency, domain consistency, format consistency, and topological consistency [10].

In order to record information for each applicable data quality sub-element, [10] lists seven descriptors. The main items on this list are: data quality scope, data quality measure, and data quality result. Data quality scope is defined as a suitable portion of the dataset that can fulfill the user's requirements. Another important descriptor of a data quality sub-element is data quality measure, defined as "the evaluation of a data quality sub-element" [10]. For instance, the number of incorrect values of an attribute is a data quality measure used for evaluating the quality of data by means of a domain consistency check, which itself is a sub-element of topological consistency. Data quality result refers to value(s) resulting from applying a data quality measure or the outcome of comparing the obtained value against a conformance quality level [10]. The conformance quality level is a threshold value for data quality results, used to determine how well a dataset meets the user's requirements [11].

3   Data Quality Evaluation Procedure and Its Process Flow

In this research a quality evaluation procedure is defined as matching the user's requirements against the dataset itself, to see if the selected dataset is suitable for the users' needs. The process given in Figure 1 represents the sequence of steps that should be taken for obtaining a data quality evaluation result and reporting it. The process begins with two main inputs, which are the dataset and the user's requirements. The user requirements are considered to be the information given by the user based on his/her desired needs. Its main properties include data quality scope, data quality element and sub-element,

data quality measure, and quality conformance level. By considering the dataset itself and the users' needs, the process continues by defining each main property of the user's requirements, one by one. In the next step, a data quality evaluation method is chosen for handling each data quality measure. After applying these methods, each evaluation will have its own result and its related quality conformance level check will be performed to conclude information about the "fitness for use" of the dataset and report it in an appropriate manner to the user.
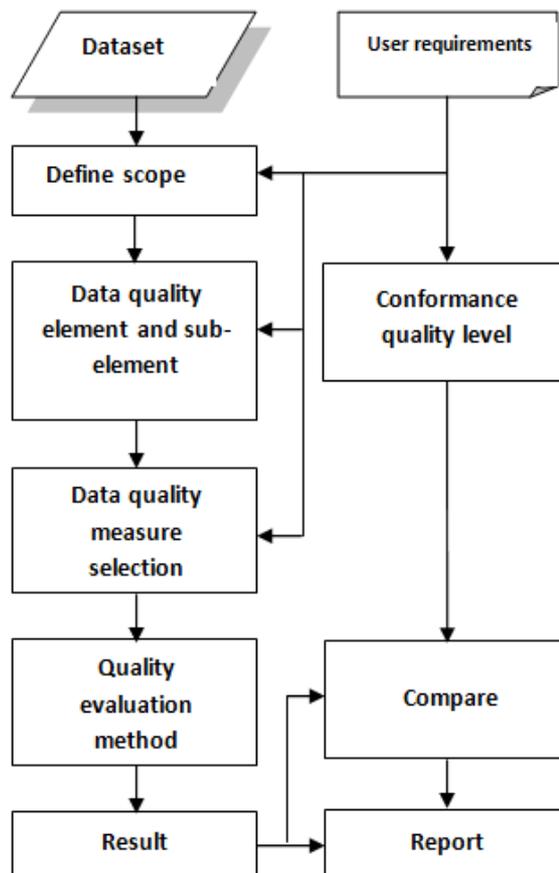

Figure 1 Data quality evaluation process flow (adopted from [10]).

4   Semi-Automated Quality Evaluation

Considering the level of human interference in performing the evaluation procedure, three different cases can occur: non-automated, semi-automated, and automated quality evaluation. In the case of non-automated quality evaluation,

the procedure of selecting and applying the quality evaluation method is done manually by a human. After growth of computer algorithms, the idea of handling the evaluation procedure by computer services comes in view. Semi-automated evaluation – which is the aim of this study – is the case where the user still has direct interference with the evaluation procedure, and can decide to choose which method for quality evaluation should be considered, but with getting assistance from a computer service controlled by an algorithm. Finally, automated quality evaluation means that the aim is to handle the quality evaluation procedure without direct human interference, fully controlled by an algorithm that obeys a specific process. Nowadays, spatial web processing services [12] provide powerful instruments for retrieval, manipulation and dissemination of spatial data. A semi-automated quality evaluation WPS can act as an SDI node to receive requests carrying dataset and user requirements, and return the result of the quality evaluation without the need of human interaction. With this in mind, in the next section the available quantitative data quality elements are reviewed and their capability for semi-automated evaluation is discussed.

5   Which One is A Good Candidate?

As mentioned earlier, five quantitative data quality elements exist: positional accuracy, thematic accuracy, logical consistency, completeness, and temporal accuracy [10]. In this section, these five data quality elements are reviewed one by one, and the possibility of their selection for semi-automated quality evaluation is argued.

5.1   Positional Accuracy

Positional accuracy is defined as the accuracy of coordinate values [13]. For performing positional accuracy checks, obtaining true values through fieldwork is necessary. In the case where fieldwork cannot be performed, a reference dataset of the real world that has an accepted level of quality is used. Obviously, the only way to measure positional accuracy is to compare the datasets with a reference dataset [1,14]. Positional data may be specified by coordinates, addresses, or locality descriptions. Thus, positional uncertainty should be discussed according to the type of positional descriptions given. In a recent research study, both probabilistic and fuzzy methods have been used for uncertainty description in objects and fields [15]. Seo and O'Hara [16] proposed a method to measure the correspondence between line segments for assessing the geometric quality of spatial data. In their method, matching is performed on rasterized line segments, and their matching lengths and displacements are measured. Mozas and Ariza [17] described a set of new metrics for evaluating

the positional accuracy of lines in cartographic databases that is based on vertex displacements and their influence on adjacent segments.

In addition, Kronenfeld [18] proposed a polygonal modeling approach to represent boundary uncertainty on area-class maps using a simple polygon tessellation with designated transition zones, which can be conceptualized as duals of the epsilon bands. The transition-zone based representations were found to be more flexible than the epsilon bands and allow for a wide range of polygonal configurations, potentially useful for expert characterization of areal units where gradation and/or boundary uncertainty are prevalent.

Since this research relies on "fitness for use" as the definition of quality, in case of a positional accuracy test, each user might need a different accuracy and precision for positional values of objects in a spatial dataset. Therefore, the service should be able to receive user requirements as input in the form of threshold values defined for the minimum level of acceptable accuracy and precision of object features. Then the service could use proven methodologies for the evaluation of the position accuracy of objects, such as the methods provided in [15-18]. Also, several basic checks could be performed on a spatial dataset. For example, the numeric coordinate values of an object should be within the expected geographic extent (*e.g.* coordinates for a North American dataset should not have numeric values for coordinates found in South America, etc.). Thus, it is completely based on user analysis and interaction, which in turn depends on the application and domain the data is going to be used for.

The reference dataset should either be passed by the user as the second input, or provided by the service itself. Since reference datasets are produced by data providers and can be too expensive for inexpert users to afford, few users are able to provide reference data to the evaluation service. Nowadays, however, many SDIs provide data free of charge and there is also open data available. Thus, the proper solution is to develop a search and discovery web processing service that searches the yellow pages of web and service catalogues for reference data in the SDIs. This geo-information search and discovery service could use information contained in the user requirements in order to select keywords and the semantics of data relating to the search task. From the list of available datasets in an SDI that could be used as reference, the most relevant should be selected and passed on to the quality evaluation service. Then by checking the coordinates of points from the input dataset and reference dataset, inaccurate objects can be identified and reported back to the user in the proper manner. In another approach, only a selection of objects would be checked, and based on the results, the service could give an estimated percentage value for the positional accuracy of the spatial dataset. Please note that volunteered geo-

graphic information (VGI), such as Open Street Map (OSM) [19], can be a good solution for reference data as long as its quality is reliable itself.

Based on this discussion, positional accuracy has the potential to be in the list of candidates for a semi-automated quality evaluation check.

5.2   Thematic Accuracy

Thematic accuracy is defined in [10] as "the accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classification of features, and their relationships". The same situation that positional accuracy has can be applied to thematic accuracy as well, where user decisions play an important role in the determination of the evaluation procedure [11]. Again, a reference dataset would be needed and through a comparison of object labels from the reference dataset and the users' spatial dataset, an estimation of thematic accuracy could be made. In this respect, OSM data could be a good source because of its richness in semantic information for object features. However, the increased usage of OSM data (as well as other VGI) makes it important to identify quality indicators for volunteered geospatial information [20-22] in order to determine fitness for the intended purpose. Based on this discussion, thematic accuracy has the potential to be in the list of candidates for a semi-automated quality evaluation check.

5.3   Logical Consistency

Logical consistency is defined as "the degree of conformance to logical rules of data structure, attributes and relationships" [10]. It consists of four sub-elements: conceptual consistency, domain consistency, format consistency, and topological consistency. Conceptual consistency involves the rules defined in a conceptual schema. In spatial datasets, the same as in non-spatial datasets, the features and their relationships are defined in the conceptual schema of the dataset. Examples of conceptual inconsistencies can be: invalid placement of features within a defined tolerance, duplication of features, and invalid overlap of features [23]. However, in practice, not all rules are explicitly defined in the conceptual schema. This is because some rules are completely application-dependent (*e.g.* not all overlapping features are necessarily erroneous). In general, the integrity constraints defined in the data model ensure that values of feature attribute geometry and topology, database schema and file formats are valid [1]. So, there will be no need to perform such kind of conceptual consistency check on a dataset. On the other hand, if the rules and relationships between data objects can be defined by a formal language (*e.g.* web ontology language), then those rules and their definition could be checked by performing a conceptual consistency test. So, to some extent logical consistency for the purpose of a conceptual consistency test could be evaluated semi-automatically.

For quality evaluation by means of a domain consistency test, the attributes of the objects within a dataset should be compared against the acceptable attribute domain, and the values that are outside the domain are determined and counted as inconsistencies. In general, a domain determines the acceptable attribute values. Whenever a domain is chosen for a field of an attribute, only the values within that domain can be entered into that field. Furthermore, two main properties of a data field should be checked, namely field type and domain type [24]. The field type is the type of field attribute that can be set to any of the standard types, such as short and long integers, double, text, date, etc. Note that a field type check has overlap with checking the type of attributes for conceptual consistency, which could be controlled and checked by defining constraints. Therefore there is no need to check the types of field. On the other hand, domain types are used for making different kinds of limitations for value choices. There exist two major kinds of domain types, *i.e.* range domains and coded domains [25]. A range domain is used for numeric attributes and specifies a valid range of values that can be entered for the domain. Coded domains can be applied to any type of text attribute, numeric, date and so on. They specify a valid set of values for an attribute. A domain consistency check is another good candidate for a semi-automated quality evaluation check, where a user can define specific domain values (*e.g.* range of numbers, enumerated list of values) because the user requirements and those values can be checked with the values of attributes in the dataset.

Format consistency deals with the format and the type of fields that data is stored in. When conceptual consistency was discussed, it was mentioned that data models have constraints defined for the format of the fields inside the dataset. Software has the capability to ensure these integrity constraints. In special cases, based on user requirements, the user might want to define a specific structure and check the values inside fields to see whether they obey this structure. For example, postal codes are defined as string fields in the data model, but apart from that, a user might want to check and see if the postal code values obey a specific structure like [1234 AB] (six characters in total: first four characters should be digits and the last two should be letters). In this example, all items that have a postal code field and do not obey this user-defined field structure are counted as inconsistencies. This type of check could be performed semi-automatically with a format consistency test in a logical consistency check.

In addition, due to data measurement methods and map generalization operators such as aggregation, displacement, and simplification, topological inconsistencies occur in spatial datasets. This is because these operators often

reduce the shape and structure of spatial objects. There exist several methods that focus on topological consistency [26-28]. The main considerations in a topological consistency check are checking polygon boundary closures, checking true connections in linear features (every arc of a network should be connected by a node to another arc), checking the topology and the spatial relationships, and checking for polygon overlaps. The first two cases can be checked by means of automated evaluation, since every arc is stored in the database as a straight line connecting a start and end node, and each node has its own identifier. By checking positional values of nodes, a boundary closure check or network connectivity check can be performed. The same procedure can be applied for polygon overlaps. For checking the topology and spatial relationships of features, topological rules should be defined and used. One research study [29] defines four approaches for the establishment of topological relationships between regions with each other, and of line/region relations as well. Apart from that, several other articles define topological relationships of features [27,28]. In an interesting research study, different categories of polygons have been defined based on their topological consistency, namely invalid, valid and clean polygons [30]. The authors of this study argue that during their tests and benchmarks, they noticed subtle but fundamental differences in the way polygons are treated (even in a 2D situation and using only straight lines). The consequences can be quite unpleasant. For example, a different number of objects are selected when the same query is executed on the same data set in different environments. Another consequence is that data may be lost when transferring it from one system to another, as polygons valid in one environment may not be accepted in the other environment.

It is believed that an automated service can check the topological consistency of spatial datasets by using the definitions and rules provided in [30], and thus report the number of invalid, valid and clean polygons in a specific dataset. Therefore, it is concluded that topological consistency is also a good candidate for a semi-automated quality evaluation check.

5.4   Completeness

Completeness is defined as errors of omission (measure of the absence of data), and errors of commission (measure of the presence of extra data) [13]. The completeness of a dataset can be suitable for a specific task but not for another. So, when completeness has to be measured, the concept of fitness for use comes in mind. Generally speaking, two types of completeness exist, namely: data completeness and model completeness [1]. Data completeness refers to the abovementioned errors of omission and commission. It is measurable and independent of application. Model completeness is defined as the "comparison between the abstraction of the world corresponding to the dataset and the one

corresponding to the application, preferably evaluated in terms of fitness for use" [1]. Furthermore, data completeness contains both formal completeness and object completeness. Formal completeness concerns the data structure, adherence to the standards used, and presence of metadata [1]. Object completeness concerns attributes and relationships of objects. Completeness monitors both omission and commission in information contained in geographic databases by answering the following questions: [1]

1.   Is the number of objects modelled equal to the number of objects defined in the model?
2.   Do the modelled objects have the correct number of attributes and are all attribute values present?
3.   Are all entities represented in the reference data represented in the model?

For the purpose of a semi-automated quality evaluation check, the completeness of a dataset can be performed only if a reference data is at hand (for the purpose of comparison). However, this reference dataset can be a simple text file containing specific values for attributes, or could be a complete dataset itself.

5.5   Temporal Accuracy

The concept of quality in this research is defined based on "fitness for use". Due to this, the date of data input and the date of update become important factors [1]. Some users may want to use this type of date and time information. Based on the type of feature, the management of time-related issues is different [1]. Some entities are updated at regular time intervals, such as aerial photographs, while others require historical management, such as cadastral maps. This is the reason why the temporal aspects of features are treated in different manners, sometimes as a date, an interval, and sometimes as a temporal range [1]. Therefore, temporal consistency can be a candidate for semi-automated quality evaluation check if the metadata of a dataset is available.

6   Discussion and Conclusion

The main goal of this research study was to explore the possibility of semi-automated quality evaluation of spatial datasets. This is important because of the increasing spatial data volume, as well as the call of volunteered geographic information for (semi-)automated approaches of spatial data quality evaluation. Based on the literature review discussed in section 5, several data quality elements and their sub-elements were identified as having the potential for a semi-automated quality evaluation check of spatial data sets. Table 1 shows the final list of nominated elements and their sub-elements that have the possibility of a semi-automated quality evaluation check, with a brief comment on the required conditions, if any.

This research study is carried out as part of a research project supported by Khavaran Institute of Higher Education in Iran, and the results will be used to develop a semi-automated quality evaluation model. Therefore, for future work, a model for evaluating the quality of each data quality element will be designed and implemented in a web processing service to act as a SDI node. This service should provide users with a tool for performing a semi-automated quality evaluation check of spatial datasets. Different users in various organizations can pass their spatial datasets to the web service and have quite definite information regarding the quality of the dataset for their intended use. The users can be human users and/or other web services.

Table 1   Possibility of semi-automated quality evaluation check based on data quality elements and their sub-elements.

Data quality element And sub-element	Possibility status	Condition
Positional accuracy – Check of coordinate values	Possible	If reference dataset is available (needs users interaction and opinion)
Thematic accuracy – Check of object labels and semantic information	Possible	If reference dataset is available
Logical consistency – Conceptual consistency	Possible	If conceptual relationships are defined formally
Logical consistency – Domain consistency	Possible	--
Logical consistency – Format consistency	Possible	In case of structure definition
Logical consistency – Topological consistency	Possible	If object relationships are defined formally
Completeness	Possible	If reference dataset is available
Temporal consistency	Possible	If metadata is available

The design of the quality evaluation model and web service is not discussed in this paper. However, the initial steps and investigations towards the development of such a model and service have been discussed and presented, which gives valuable insight to the geo-spatial community in the issues involved with preparing semi-automated services for quality evaluation of spatial data.

7   Acknowledgement

References

[1]   Devillers, R., & Jeansoulin, R., *Fundamentals of Spatial Data Quality, Second Chapter, Spatial Data Quality: Concepts*, ISTE, pp. 31-42, 2010.

[2]   Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P., & Shi, W., *Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities*, Transactions in GIS, **14**(4), pp. 387-400, 2010.

[3]   Devillers, R., & Goodchild, H., *Spatial Data Quality: from Process to Decisions*, CRC Press, 2010.

[4]   EuroGeographics, *ESDIN Project Delivers A World First in Quality Evaluation Web Service*, http://www.eurogeographics.org/sites/default/ files/ESDIN_first%20quality%20ev_services%20published_1.pdf   , (1 April 2013).

[5]   Hunter, G.J., Bregt, A.K., Heuvelink, G.B., De Bruin, S., & Virrantaus, K., *Spatial Data Quality: Problems and Prospects*, In Research Trends in Geographic Information Science, Springer Berlin Heidelberg, pp. 101-121, 2009.

[6]   ISO, *ISO 9000: 2005, International Standard: Quality Management Systems–Fundamentals and Vocabulary*, Geneva, Switzerland: International Organization for Standardization, 2005.

[7]   American Society for Quality, *Glossary-entry: Quality*, http://www.asq. org/glossary/q.html, (1 April 2013).

[8]   Yang, X., Blower, J.D., Bastin, L., Lush, V., Zabala, A., Masó, J., Conford, D., Diaz, P., & Lumsden, J., *An Integrated View of Data Quality in Earth Observation,* Phil. Trans. R. Soc. A., **371**, 20120072, 2013.

[9]   Life Cycle Assessment, *Glossary-Entry: Data Quality*, available at http://www.lcacenter.org/LCA/LCA-definitions.html, 2013.

[10]   ISO TC211, *ISO/DIS 19113, Geographic Information-Quality Principles*, Text for DIS 19113, The International Organization for Standardisation, 2001.

[11]   Mobasheri, A. *Automated Quality Evaluation in The Context of Spatial Data Infrastructure*, Master of Science Thesis, Faculty of Geo-

Information Science and Earth Observation (ITC), University of Twente, Enschede, 2010.

[12] Schut, P., & Whiteside, A., *OpenGIS Web Processing Service*, Open Geospatial Consortium, Wayland, MA, USA, 2007.

[13] van Oort, P.V., *Spatial Data Quality: from Description to Application*, PhD dissertation, Wageningen Universiteit, Wageningen, 2006.

[14] Congalton, R.G., & Green, K, *Assessing The Accuracy of Remotely Sensed Data: Principles and Practices*, CRC press, 2008.

[15] Li, D., Zhang, J., & Wu, H., *Spatial Data Quality and Beyond*, International Journal of Geographical Information Science, **26**(12), pp. 2277-2290, 2012.

[16] Seo, S., & O'Hara, C.G., *Quality Assessment of Linear Data,* International Journal of Geographical Information Science, **23**, pp. 1503-1525, 2009.

[17] Mozas, A.T., & Ariza, F.J., *New Method for Positional Quality Control in Cartography Based on Lines, A Comparative Study of Methodologies*, International Journal of Geographical Information Science, **25**, pp. 1681-1695, 2011.

[18] Kronenfeld, B.J., *Beyond The Epsilon Band: Polygonal Modeling of Gradation/Uncertainty in Area-Class Maps*, International Journal of Geographical Information Science, **25**, pp. 1749-1771, 2011.

[19] OpenStreetMap, *The Free Wiki World Map*, http://www.openstreetmap. org/, (1 April 2013).

[20] Haklay, M., & Weber, P., *OpenStreetMap: User-Generated Street Map*, IEEE Pervasive Computing, **7**(4), pp. 12-18, 2008.

[21] Goodchild, M.F., *Citizens As Sensors: The World of Volunteered Geography*, GeoJournal, **69**(4), pp. 211-221, 2007.

[22] Flanagin, A.J., & Metzger, M.J., *The Credibility of Volunteered Geographic Information*, GeoJournal, **72**, pp. 137-148, 2008.

[23] ISO TC211, *ISO/TS 19138. Geographic Information - Data Quality Measures*, Text for TS 19138, The International Organization for Standardisation, 2006.

[24] ArcGIS 9.2 desktop help, *An Overview of Attribute Domains*, http://web help.esri.com/arcgisdesktop/9.2/index.cfm?topicName=An_overview_of _attribute_domains, (1 April 2013).

[25] ISO TC211, ISO/TS 19103, *Geographic Information-Conceptual schema Language*, Text for TS 19103, The International Organization for Standardisation, 2005.

[26] Clementini E., Di Felice P., & Egenhofer, M.J., *Topological Relations between Regions with Holes*, International Journal of Geographical Information Systems, **8**(2), pp. 129-144, 1994.

[27] Sharma J., & Egenhofer, M.J., *Topological Consistency*, in Fifth International Symposium on Spatial Data Handling, pp. 335-343, 1992.

[28]   Sharma J., & Egenhofer, M.J., *Assessing the Consistency of Complete and Incomplete Topological Information*, Geographical Systems, **1**(1), pp. 47-68, 1993.

[29]   Kim T.W., Li K.J., & Kang, H.K., *Topological Consistency for Collapse Operation in Multi-Scale Databases*, in S. Wang, *et al.* (eds.), 23rd International Conference on Conceptual Modeling, Lecture Notes in Computer Science, **3289**, pp. 91-102, 2004.

[30]   Van Oosterom, P., Quak, W., & Tijssen, T., *About Invalid, Valid and Clean Polygons*, In Developments In Spatial Data Handling, Springer Berlin Heidelberg, pp. 1-16, 2005.