Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

Interoperable GIS Operations: A Quality-Aware Perspective

Jung-Hong Hong, Associate Professor
Min-Lang Huang, Ph.D. Candidate
Department of Geomatics, National Cheng Kung University
Tainan City 701, Taiwan
junghong@mail.ncku.edu.tw yoshi.hml@gmail.com

Abstract

As the sharing of geospatial data becomes increasingly easy and convenient, the discrepancy and heterogeneity of data quality between datasets dynamically acquired from various georesources must be taken into consideration. Such awareness of data quality must extend to the design of geographic information system (GIS) operations, which GIS users often naively use to analyze and derive new information. This study proposes a quality-aware approach for incorporating standardized quality information into the design of GIS operations to ensure that the necessary quality information is always visible in a map interface for correct interpretation. Since the proposed approach can largely reduce user effort when dealing with unfamiliar datasets, it can be seamlessly integrated with current sharing frameworks based on geospatial web services to facilitate user-friendly and interoperable geospatial applications.

Keywords: Quality, Operation, Interoperability

1 Introduction

Internet-based geospatial technology has seen rapid growth, with millions of users, both novice and professional, using it on a daily basis. Although using geographic information system (GIS) operations seems to be straightforward and produces appealing visualizations in a map interface, a large portion of users either ignore the discrepancies of data quality between different datasets, or are totally unaware of the hidden risks caused by the diversity of data quality. Although data sharing has become increasingly easy and convenient, an interoperable GIS-based application should not only display available heterogeneous geospatial data, it should help users correctly interpret and use the acquired data. An approach that considers data quality in the process of integrating geospatial data is thus necessary.

Quality information is essential in decision-making. Data quality can be defined as the degree of excellence exhibited by the data in relation to the portrayal of the actual phenomena [1]. ISO/PAS 26183:2006 defines data quality as a measure of the accuracy and appropriateness of product data, combined with the timeliness with which those data are provided to the people who need them. As GIS-based applications often involve diverse types of geospatial data produced by different organizations, the quality of the selected datasets varies over a large range. Users must thus be able to (1) determine the quality of selected datasets and (2) correctly make decisions based on the quality information. This would require a comprehensive examination of the characteristics, evaluation, encoding, distribution, interpretation, and visualization of geospatial data quality from both theoretical and implementation perspectives. In the geospatial domain, the fundamental characteristics of data quality information, namely completeness, logical consistency, positional accuracy, temporal accuracy, attribute accuracy, semantic accuracy, and lineage, are covered in ISO 19113 [2] and a study by Devillers et al. [3]. Based on this framework, ISO19114 [5] and ISO19138 [6] extend the scope of standardization to include the evaluation procedure and quantitative measures of quality information. ISO19115 [7] offers a common framework for recording quality information via standardized metadata elements. These standards not only enable data suppliers to precisely convey their knowledge about data to users, but also offer the possibility for users to correctly understand the quality of acquired data and determine its fitness for use, even if data is dynamically acquired from previously unknown sources.

Although consensus agreements on the categorization and encoding frameworks for quality information have been reached, there is a lack of an effective mechanism that can seamlessly integrate quality information with its corresponding geospatial data in geospatial applications. The concept of quality-aware GIS [8][9][10][11] argues that GIS must be aware of the diversity and heterogeneity of data quality information and offer effective strategies for dealing with the selected datasets and metadata. The link of data and metadata was mentioned in [12]. Recently, the primitive modeling frameworks proposed in [13] offer a standardized structure for encoding the identification, spatial and temporal attributes, and quality of distributed geospatial datasets. Hong and Liao [14] introduced the concept of valid extent to spatially illustrate the data completeness status in a GIS-based map interface. Since the results are dynamically determined from the data completeness status of the selected datasets, the method’s success depends on the quality information provided by the data owners. The linkage between geospatial data and its metadata is the basis for developing quality-aware GIS operations in the present study.

Different from the data quality information given in ISO standards, the quality information represented by valid extent denotes a new type of quality information that can only be deduced after a particular GIS operation, e.g., map overlay, is issued. As the number of georesources increases, it becomes less likely and even impossible for users to rapidly build a comprehensive understanding of the selected datasets. Without a solid theoretical framework that considers the outcome of GIS operations, decision-making using geospatial applications may become extremely difficult because even a simple GIS operation may have to deal with a complex and unpredictable mixture of data quality information. All GIS operations must thus have a built-in capability to automatically determine the applicable conditions of the selected datasets based on the quality information. Different types of quality information may need to be taken into consideration for different operations. Extended metadata elements and implementation strategies may need to be considered in addition to the current metadata standards.

The interoperability of GIS operations is not limited to the issues of reading GIS data formats or illustrating selected datasets in a map interface. This present study proposes a quality-aware approach for improving the interoperability of selected datasets. The rest of this paper is organized as follows. Section 2 introduces the encoding strategy of data and quality. Section 3 describes the proposed quality-aware approach for distributed application environments. Section 4 uses two examples to demonstrate the benefits of the proposed quality-aware approach. Finally, Section 5 summarizes the major findings and suggests future directions of development.

2 Encoding Strategy of Data and Quality

For a quality-aware GIS, to ensure that all the quality information can be properly interpreted by users, the distributed data must be both open and self-described. As geospatial data is dynamically selected according to application needs, the quality status after data integration must be dynamically determined. Theoretically, every quality description has its own data scope, which represents the domain of the data from which the quality information is evaluated. This scope information must be unambiguously specified for every individual quality evaluation result. Although the majority of current quality information refers to individual datasets, it may also refer to a dataset series, feature, or attribute. For example, the quantitative measures for data completeness are based on the omission error and commission error after the dataset has been compared with the universe of discourse, so the data scope by default refers to datasets. The enriched set of metadata elements given in [13] (e.g., surveyed area of the dataset) also refers to a single dataset. The positional accuracy may refer to either a dataset or a feature depending on the positioning technology or the testing specifications. Table 1 lists the data scopes and corresponding quality elements considered in this paper.

Table 1: Data quality consideration of geospatial data.

Level	Element	Component
Dataset	Completeness	Surveyed area
		Commission
		Omission
Feature	Positional Accuracy	Absolute Or External Accuracy
	Thematic Accuracy	Non-Quantitative Attribute Correctness/ Quantitative Attribute Accuracy

The strategies proposed here follow the standardized framework of self-described features proposed in [12]. In addition to the quality information, every feature is required to include a ValidTime attribute to denote its temporal status. This helps applications determine whether the GIS operations are dealing with features that coexist at the specified period of time. This type of information should be considered in many GIS operations, but are often ignored.

Feature-based modeling is commonly used for the development of object-oriented technology in GIS. The distributed data and corresponding quality information in the present study are encoded in GML and XML following ISO19136 and ISO19139, respectively. The open encoding of distributed datasets allows applications to transparently parse necessary information from acquired data. Figure 1 shows a GML encoding example of the dataset “school”.

Figure 1: Example of “school” dataset encoding.

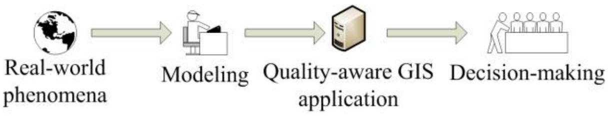
```
<igis:FeatureCollection>
 <gml:metaDataProperty>  Dataset level quality information
 <gmd:DQ_DataQuality>....
  <gmd:report>
  <gmd:DQ_CompletenessOmission>....
   <gmd:pass><gco:Boolean>true</gco:Boolean></gmd:pass>
  </gmd:DQ_CompletenessOmission></gmd:report>
 </gmd:DQ_DataQuality>
  <gmd:DQ_CompletenessCommission>....
   <gmd:pass><gco:Boolean>true</gco:Boolean></gmd:pass>
  </gmd:DQ_CompletenessCommission></gmd:report>
<igis:SurveyedArea>
  <gmd:EX_BoundingPolygon>  <gmd:polygon>
    <gml:posList>121.520 25.061... </gml:posList>
  </gmd:polygon></gmd:EX_BoundingPolygon>
</igis:SurveyedArea>
  </gmd:report>
 </gmd:DQ_DataQuality>
</gml:metaDataProperty>
 <gml:featureMember>
  <igis:School>
   <gml:validTime><gml:TimeInstant>
    <gml:beginPosition>1931-01-01T00:00</gml:timePosition>
    <gml:EndPosition>2012-01-01T00:00</gml:timePosition>
    </gml:TimeInstant></gml:validTime>
   <igis:Spatial>⋯.</igis:Spatial>
   <gmd:DQ_AbsoluteExternalPositionalAccuracy>
     <gmd:value><gco:Record>50</gco:Record></gmd:value>
   </gmd:DQ_AbsoluteExternalPositionalAccuracy>
   <igis:Area uom=" m2" >68514</igis>
   <igis:Area-Quality>
   <gmd:DQ_ QuantitativeAttributeAccuracy >
    <gmd:result> <gmd:DQ_QuantitativeResult id="ID">
    <gmd:value><gco:Record> 1</gco:Record> </gmd:value>
    </gmd:DQ_QuantitativeResult>  </gmd:result>
    </gmd:DQ_ QuantitativeAttributeAccuracy >
   </igis:Area-Quality>     Feature level quality information
  </ igis:School>
 </gml:featureMember>
 <gml:featureMember>⋯⋯.
  </ igis:FeatureCollection >
```

The data completeness information (surveyed area, commission error and omission error) is modeled as a dataset-level metadata element, and positional accuracy is modeled as a feature-level metadata element. If all of the features in the dataset are evaluated following a given procedure, the quality information can be recorded at the level of the dataset and directly used as feature-level metadata in the application if necessary. This can reduce unnecessary duplicates when managing and maintaining metadata. As both the geospatial data and metadata are encoded in open data formats following well-established international standards, the reading and parsing mechanism can be readily implemented. The analysis module, however, must be able to deal with heterogeneity and diversity of quality information.

3 Quality-aware Approach

The fundamental concepts for developing a quality-aware GIS are: (1) all distributed features are self-described and published in open data formats; (2) applications have built-in knowledge to intelligently parse necessary metadata according to the GIS operations, and (3) applications present the evaluation results of quality information to users in a meaningful way to aid decision-making. A simplified workflow of the proposed quality-aware GIS approach is illustrated in Figure 2, which subdivides the whole process into three major stages, namely the modeling, quality-aware GIS application, and decision-making stages.

Figure2: The concept flow of Quality- aware approach in distributed environment.

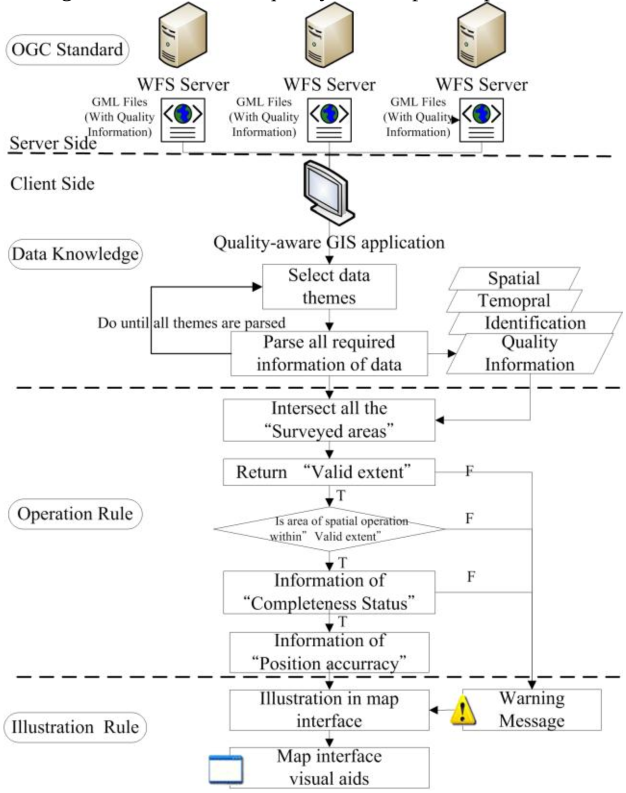


The major goal of the modeling stage is to ensure that the distributed geospatial data and quality information are self-described. Individual datasets are created according to their spatio-temporal conditions using one of the primitive frameworks proposed in [12]. Information about the identification, spatial and temporal attributes, and dataset/feature-level of quality are respectively created. Data providers are responsible for the modeling process because only organizations participating in the production and distribution of geospatial data have in-depth knowledge about the specification, restrictions, and quality of the datasets. From the perspective of SDI, this requirement must be considered while establishing the collaborative relationship between participating organizations. Otherwise users will not be able to take full advantages of the geospatial data available from SDI.

A quality-aware GIS application must parse the required metadata created in the modeling stage and intelligently determine how to use the acquired data. Although the algorithms for each operation are designed independently, common principles may be shared by a number of operations. Quality-aware GISs are mainly designed and operated by application users with professional expertise because the functions required are application-dependent. Finally, every GIS operation must be examined to ensure that the design of its processing strategies and outcomes takes the necessary data quality information of geospatial data into consideration. This is the responsibility of GIS software developers or vendors. In the decision-making stage, the results are transformed into visual aids and presented to users.

Using data completeness as an example, Figure 3 shows the modified workflow after the consideration of quality information has been added to the design of a GIS operation.

Figure 3: Workflow of quality-aware spatial operation.



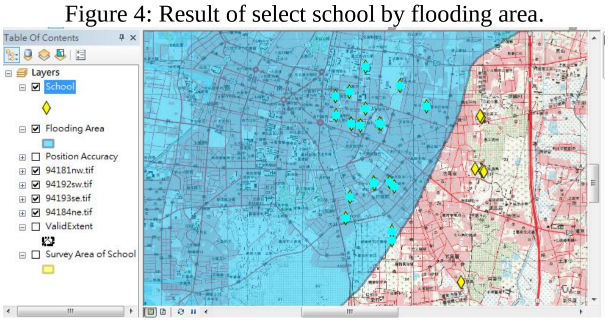
The operation begins with the user selecting a number of datasets. The application automatically parses the information of the surveyed area, completeness status, and positional accuracy of the datasets from their metadata. Whenever a new dataset is selected, the parsing procedure is triggered to update the quality information for the analysis module. To provide meaningful visual aids about the data completeness status, three conditions must be tested. Firstly, the valid extent is determined by calculating the geometric intersection of the surveyed area of selected datasets. If the valid extent is empty, a warning message must be issued to users to indicate that some information is missing. Secondly, the analysis module needs to test whether the spatial extent of the queried region is entirely within the valid extent. If not, a message should be issued to users. Finally, the completeness status of all the selected datasets must be "YES" to make the valid extent meaningful. If it is not, a warning message is shown to users. Since the judgments are based on the features being within the valid extent, positional accuracy is also included for user reference. Note that this analysis depends on the selected datasets and metadata, so the valid extent must be recalculated after any modification to the selected datasets. In the decision-making stage, the information of the valid extent is transformed into visual aids. If different measures for completeness (e.g., count or percentage) are used, the concept of valid extent is still valid, but the completeness status for each dataset must be presented to users. Visual inspection is still possible, but decision making becomes much more complicated when different types of measures are considered simultaneously.

4 Use Cases

The following discussion uses two widely used GIS operations, select by region and distance measurement, to demonstrate how the quality information can be assimilated into the design of GIS operations. Decision-making for a school evacuation caused by flood is chosen as the test case. A system prototype for these two operations was developed in Visual Basic using ESRI ArcGIS 9.3.

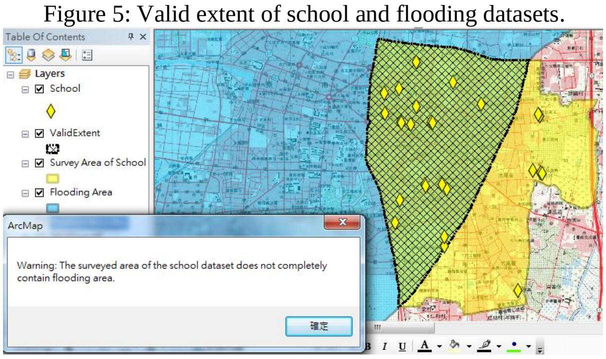
The commonly-used algorithm for the "selection by region" operation is very simple; users specify a region as the spatial constraint and the GIS responds with the information of the features within the specified region. It is typically regarded as a geometric operation, where only location is taken into consideration (e.g., Figure 4).

Figure 4: Result of select school by flooding area.



Users can easily determine the schools that need to be evacuated by specifying the predicted flooding region as the constraint and using the operation "selection by region" to determine whether the location of schools meet the "within" condition. This, however, only works when the surveyed area of the school dataset contains the spatial extent of the flooding region. Under such circumstances, it can be ensured that all of the schools within the flooding region are found. Figure 5 shows the school evacuation scenario. The results would be rather different if such data quality information is not taken into consideration.

Figure 5: Valid extent of school and flooding datasets.



Because the surveyed area of the school dataset only covers a portion of the map interface, users have no information about the number and spatial distribution of schools outside the surveyed area. The green polygon in Figure 4 represents the overlapped region between the predicted flooding region and the surveyed area of the school dataset. With the addition of quality information, it is validated that all schools within this region have been found. The yellow polygon indicates the surveyed area of the school dataset outside the flooding region. Schools within this region do not need to be evacuated. The blue polygon indicates the subpart of the flooding region where no information about schools is available. Without the consideration of data completeness, a user may naively assume that the schools being found within the green region (e.g., the ValidExtent in Figure 5) are all the schools that need to be evacuated, which may lead to erroneous decisions. Therefore, the operation of "selected by region" is not as simple as it appears to be. The testing of the surveying area for each selected dataset needs to be continuously updated with changes of the selected datasets, and a warning message should be alert when the surveying area doesn't completely contain the flooding region. The message box in Figure 4 is automatically prompted to users to indicate that users should be cautious about the data completeness status of the searched results. Furthermore, visual aids (e.g., the green and blue polygons in Figure 5) must be promptly presented to remind users of the data quality status of the illustrated content in the map interface.

The parsed quality information can be presented to users in an integrated interface for reference. Figure 6 and Figure 7 show an interface that simultaneously presents information about the surveyed area, completeness, and position accuracy of each dataset. If the measure of the data completeness is Boolean, users can make reasonable decision based on the features within the valid extent (e.g., Figure 6) and must be aware of the possible missing information outside the valid extent. Users will notice that only 23% of the flooding region is within the valid extent from the presented information, so the queried result may not represent all the schools in the flooding region. If the measure is based on the percentage (e.g., 98% in Figure 7), the selected features may not represent all the schools within the valid extent, so the number "98%" must be presented to users to avoid wrong decision making.

None of these mechanisms work if no metadata about the selected datasets is available. The proposed approach thus allows the contributions from data providers, application developers, and software designers to be effectively integrated.

Figure 6: Completeness status (Boolean) for selection by region operation.

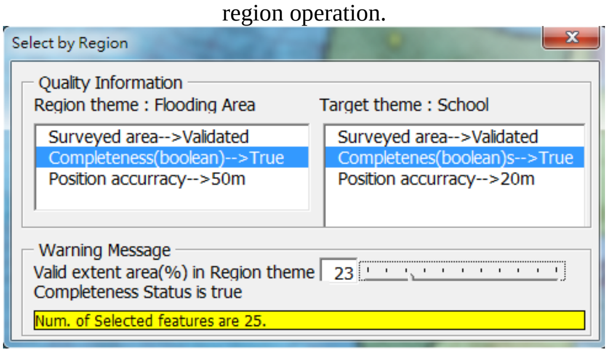
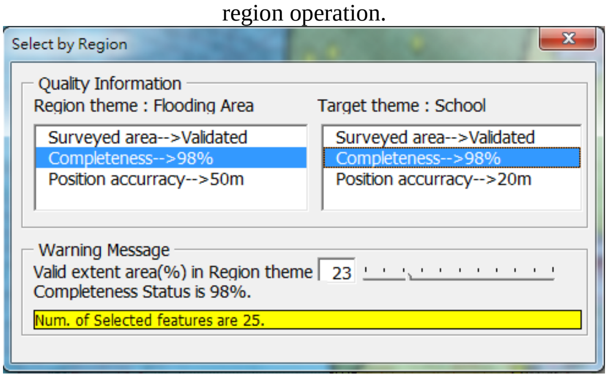


Figure 7: Completeness status (percentage) for selection by region operation.



Next, the distance measurement between features of rainfall stations and stream gauging stations is used as an example to demonstrate the proposed quality-aware approach (Figure 8). The commonly-used distance measurement operation determines the distance values by calculating the coordinates of the selected points. The positional accuracy and the level of abstraction for the selected features, however, influence the final results. The integrated interface not only presents the calculation results, but also the parsed information to remind users that the two datasets were created with different levels of positional accuracy, we use sample data to demonstrate positional accuracy in Figure 9. An analysis of the Validtime attributes of the two measured features indicates that the calculated distance value is only meaningful in the period from 1995/3/8 to 2011/9/21 (Figure 9). This example clearly demonstrates that the addition of quality information can effectively improve interoperability and decision-making reliability.

Figure 8: Measure distance operation between rainfall and stream gauging station.

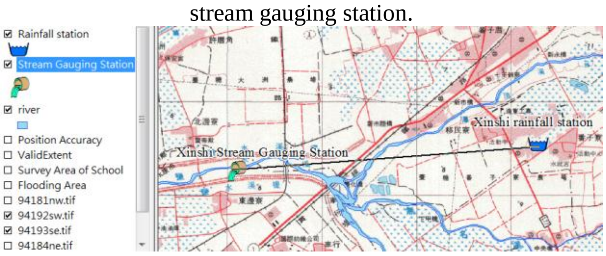
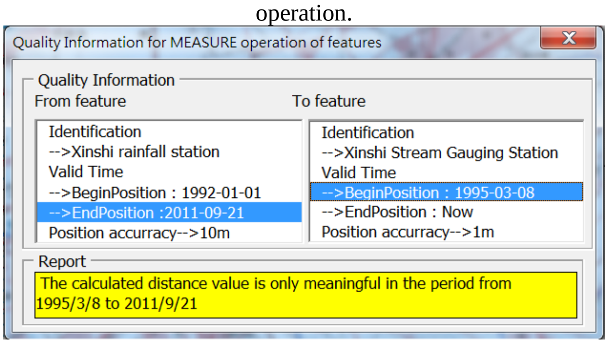


Figure 9: Report and warning message of measure distance operation.



5 Conclusion

This study integrated the standardized encoding of quality information and the fundamental characteristics of GIS operations to improve the correctness and interoperability of results. By incorporating the proposed quality-aware approach into GIS operation design, the operations detect possible risks during data processing and then provide meaningful visual aids in a map interface. With increased sharing of data from various organizations, the proposed approach offers a comprehensive framework for considering quality information and increasing the interoperability of GIS operations. With its open nature, the proposed framework can be easily developed in a service-oriented architecture environment to improve internet-based applications.

References

[1] http://www.archives.nysed.gov/a/records/mr_pubGIS03_glossary.shtml

[2] ISO/TC 211, ISO 19113: Geographic information -- Quality principles, 2002.

[3] Devillers, R., Bédard, Y., & Jeansoulin, R., Multidimensional management of geospatial data quality information for its dynamic use within GIS, *Photogrammetric Engineering and Remote Sensing*, 71(2), pages 205-215, 2005.

[4] ISO/TC 211, ISO 19113: Geographic information -- Quality principles, 2002.

[5] ISO/TC 211,ISO 19114 : Geographic information -- Quality evaluation procedures, 2003a.

[6] ISO/TC 211,"SO 19138 : Geographic information -- Data quality measures, 2006.

[7] ISO/TC 211,ISO 19115 : Geographic Information - Metadata, 2003b.

[8] Yang, T., Visualisation of spatial data quality for disturbuted GIS, *The University of New South Wales, Australia*, 2007.

[9] Devillers, R., Bédard, Y., & Jeansoulin, R., Multidimensional management of geospatial data quality information for its dynamic use within GIS. *Photogrammetric Engineering and Remote Sensing*, 71(2), pp. 205-215, 2005.

[10] Devillers, R. and Zargar, A. Towards quality-aware GIS: Operation-based retrieval of spatial data quality information.*Spatial Knowledge and Information (SKI) Conference*, Fernie (BC), Canada, February 21st. 2009.

[11] Hong, J.-H. and Huang, M.-L., A Quality-aware Approach towards the Integration of Feature-based Geospatial Data, *Multimedia Technology and its Applications (IDCTA) 7th International Conference*, 2011.

[12] Giger, C., and C. Najar., Ontology-based integration of data and metadata, *Proceedings of the 6th AGILE Conference on Geographic Information Science, Lyon, France*, 2003.

[13] Hong, J.-H. and Huang, M.-L, Kuo, C.-L. and Tseng, Y.-W. A Spatio-Temporal Perspective towards the Intelligent and Interoperable Data Use in SDI Environment, *AGIS 2010 International Conference*,2010.

[14] Hong, J. H., Liao, H. P., Maps or Not A New Insight of the Map Interface to the Open and Distributed Geospatial Web Service Environment, *ASPRS 2009*, 2009.

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke