

# Novel Tools to Improve the Management of Spatial Data Quality in the Context of Ecosystem and Biodiversity Monitoring

Pedro CASTRO<sup>1</sup>, Joaquim ALONSO<sup>1</sup>, Carlos GUERRA<sup>1</sup>, João GONÇALVES<sup>2</sup>, Isabel PÔÇAS<sup>2</sup>, Bruno MARCOS<sup>2</sup> and João HONRADO<sup>2,3</sup>

<sup>1</sup>Instituto Politécnico de Viana do Castelo, Viana do Castelo/Portugal · pmc@estg.ipvc.pt

<sup>2</sup>CIBIO, Universidade do Porto/Portugal

<sup>3</sup>Faculdade de Ciências, Universidade do Port/Portugal

This contribution was double-blind reviewed as extended abstract.

## Abstract

The growth in data availability and in the number of users from non-spatial disciplines has increased the concern with spatial data management and spatial data quality. This reality contextualizes the need and importance of data quality management throughout the spatial data cycle related to data editing and sharing in the context of knowledge networks. The development of standard concepts, procedures and tools may foster important advances in the improvement of spatial dataset production, use and management practices. The implementation of environmental monitoring programs implies a growing number and diversity of users with specific capabilities and responsibilities. In this context, spatial data quality evaluation and management promotes communication, optimizes processes of analysis and spatial modelling, and in this sense improves research and political and technical decision-making and action. This paper focuses on the methods and tools currently available to evaluate the external quality of datasets for environmental and ecological monitoring. A novel framework is presented that allows integrating and interconnecting spatial data quality evaluation with metadata geoportals in WebGIS platforms, facilitating evaluation by users with often limited expertise in this field. The advances achieved in this research highlight the relevance of developing capacities for different users to improve data collection, data models, spatial data processing and modelling, but also the need to inform and report on spatial data quality using adequate tools.

## 1 Managing the Quality of Spatial Data

Nowadays, databases are included in information systems that use their data and of which they are part (SERVIGNE et al. 2010). This represented an increase in the length of spatial data life cycles and consequently highlighted the importance of evaluating, managing, controlling and reporting on the quality of spatial databases and datasets. Existing spatial databases often contain errors due to data model development and implementation processes, acquisition sources (measuring instruments), human-technology interaction, and data-input and modelling processes (SHI 2009).

WebGIS development and the widespread availability of geospatial technologies have promoted the processes of data sharing and integration (DEVILLERS et al. 2007). Because spatial data are transferred and shared by many users, these data must be correct and useful. To ensure that existing digital data are appropriately used, the data producer must provide, among other items, documentation about the scope of production and use contexts as well as associated data quality elements. Data developers and users have begun to document and implement data quality evaluation processes and report data quality elements using metadata (BOIN & HUNTER 2006). Spatial data are frequently relied upon as factual data, and so data quality indicators and metadata are crucial to assess their fitness for use in different application contexts. Data producers must also be aware of the implications involved in the careless development of spatial data if those datasets are intended to be used for legal, organizational, political and technical purposes. On the other hand, data users should be able to acknowledge and understand the limitations of spatial datasets (KUMI-BOATENG & YAKUBU 2010).

The increasing amount and mobility of data, associated with the heterogeneity of users and uses, influences the length and the complexity of the life cycle of spatial data. These facts place a central importance in the evaluation and management of quality in the processes of capture, management, modelling, editing, publication and sharing of spatial data. These should be reflected on the management of information systems and on technical-scientific community governance. Quality management therefore assumes a multidimensional nature in collaborative processes and in knowledge network contexts, e.g. in WEBGIS platform development and spatial data infrastructure functioning.

## **2 Spatial Data Quality for Environmental Monitoring**

Environmental monitoring programs at global, national or even local levels often engage a broad framework of entities and, besides requiring a diverse set of pre-existing and ancillary data, they also produce new datasets. These processes involve the production, mobility and publication of data in organizational contexts. Under this general framework, the partners involved in the BIO\_SOS project (FP7-SPA-2010-1-263435) developed a set of tasks aimed to: (i) assess the availability and quality of pre-existing spatial databases; (ii) define a standard metadata profile (ISO 19115 compliant); (iii) load and edit metadata entries in a metadata editor integrated within a WEBGIS geoportal; and (iv) develop online applications and routines for external quality assessment.

A new methodology was developed for evaluating external data quality supported by metadata entries, that are used to calculate the similarity between data characteristics (as documented by internal quality indicators stored in metadata fields) and user's needs or expectations (referred to as "expected quality" and highly dependent on the context of data usage) based on a process of geosemantical integration. The method allows users to specify which quality indicators have stronger importance given their own requirements or expectations. The methodology is therefore based on the comparison between internal and expected values for selected quality indicators, centered on a pre-defined rule-based system which controls how the pairwise comparisons are carried out (HONRADO et al. 2011). This methodological framework has been integrated in the BIO\_SOS Metadata Geoportal presented further below.

### 3 The BIO\_SOS Metadata Geoportal

The BIO\_SOS Metadata Geoportal has been developed as a collaborative platform with functionalities to catalogue all relevant in situ and ancillary spatial data across the consortium, based on a common metadata profile, ISO 19115 compliant, promoting the harmonization of dataset description as well as the sharing, searching and retrieval of data and metadata among project partners. It also facilitates the visualization of reference and thematic datasets, in both raster and vector formats, besides offering tools intended for external data quality assessment based on metadata entries.

The BIO\_SOS Metadata Geoportal has a simple and easy-to-use graphical user interface, accessible to authorized users, consisting of four distinct components which are accessible through a web page address: (1) a geoviewer interface providing an interactive map that allows the visualization of cartographic data; (2) a metadata management component providing tools to query and retrieve information from the metadata catalogue and to edit and publish metadata information using a predefined metadata profile; (3) an external quality evaluation interface which facilitates the analysis and visualization of data quality assessments based on metadata entries; and finally (4) a users management component allowing the definition of groups, users and their privileges, thus providing differential access and management capabilities to metadata entries. Regarding the spatial data quality evaluation based on metadata entries, the BIO\_SOS Geoportal provides a user-friendly interface for users less familiarized with data quality concepts (Fig. 1), to perform an evaluation of fitness for use of spatial data. This component allows the user to select the context/objective (dataset quality assessment, or dataset search and rank), to define the expected data quality values for a set of predefined quality indicators based on metadata fields that correspond to the values that the user would like the dataset to have, and to visualize a final summary with the results of the evaluation indicating the conformity or non-conformity of each quality indicator and an overall fitness value for each selected dataset.

The screenshot displays the BIO\_SOS Metadata Geoportal interface. The top navigation bar includes 'BIO\_SOS Biodiversity Multisource Monitoring System: from Space TO Species', 'Welcome', and 'Logout'. The main interface is divided into three panels:

- Left Panel (Context/Objective Selection):** Shows a 'New dataset quality evaluation' window with a 'New dataset search' section. It lists several datasets: 'EDGE\_FT1\_sampling design for the collection of dat...', 'EDGE\_FT2\_Production of habitat maps', 'EDGE\_FT2\_Production of habitat maps', and 'FT1 - evaluation effects'.
- Center Panel (Definition of Expected Values):** Shows a 'Define expected quality values' window. It includes sections for 'Topic category (\*)', 'Spatial extent (\*)', and 'Temporal extent (\*)'. The 'Spatial extent' section is active, showing a table with columns for 'West Bound Longitude (\*)', 'East Bound Longitude (\*)', and 'South Bound Longitude (\*)'. The values are: West Bound Longitude (-8.32039), East Bound Longitude (-8.26727), and South Bound Longitude (48.8197).
- Right Panel (Visualization of Results):** Shows a 'Summary' window with a table of quality indicators. The table is divided into two sections: '1. Fit datasets' and '2. Partially fit datasets'. Each section has columns for 'Datasets', 'Fitness value', 'Typology', 'Spatial extent', 'Temporal extent', 'Language description', 'Reference system', 'Spatial scale', 'Producer description', 'Access and use restrictions', and 'File formats'.

**Fig. 1:** Quality evaluation in the BIO\_SOS metadata geoportal: context/objective selection (left); definition of expected values for quality indicators (center); and visualization of results of the dataset quality evaluation (right)

## 5 Conclusion and Perspectives

Spatial data quality evaluation plays a central role in knowledge development and management. The BIO\_SOS WEBGIS platform assumes a collaborative dimension in its development and functioning, and a modular and evolving dimension as a guarantee of its utility and continuity. This platform is intended to improve the internal and external communication in the context of the project, and to promote the quality evaluation of relevant datasets for specific use (e.g. modelling and indicator extraction tasks), as well as of the datasets resulting from image processing and classification.

The overarching goals of the project emphasize the importance of data quality evaluation procedures, of the availability of data quality management tools associated to minimum and adequate user's quality needs, of appropriately defined metadata profiles, as well as of accurate filling of metadata. Therefore, the continuous and iterative character of quality evaluation and communication is once again stressed, particularly considering the use of spatial datasets for specific purposes.

## References

- BOIN, A. T. & HUNTER, G. J. (2006), Do spatial data consumers really understand data quality information? In: 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Lisbon.
- DEVILLERS, R., BEDARD, Y., JEANSOULIN, R. & MOULIN, B. (2007), Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *International Journal of Geographical Information Science*, 21 (3), 261-282.
- HONRADO, J., ALONSO, J., CASTRO, P., MARTINS, L., PÔÇAS, I., GONÇALVES, J., GUERRA, C. & MARCOS, B. (2011), The BIO\_SOS metadata geoportal and the external quality of pre-existing datasets. Accompanying report to Deliverable No: D4.5. Project "BIO\_SOS Biodiversity Multisource Monitoring System: from Space TO Species" (FP7-SPA-2010-1-263435).
- KUMI-BOATENG, B. & YAKUBU, I. (2010), Assessing the Quality of Spatial Data. *European Journal of Scientific Research*, 43 (4), 507-515.
- SERVIGNE, S., LESAGE, N. & LIBOUREL, T. (2010), Quality Components, Standards and Metadata. In: DEVILLERS, R. & JEANSOULIN, R. (Eds.), *Fundamentals of Spatial Data Quality*, ISTE – GIS Series, 179-210.
- SHI, W. (2009), *Principles of Modeling Uncertainties in Spatial Data and Spatial Analyses*. CRC Press, 1st Edition, 135-156.
- VASSEUR, B., JEANSOULIN, R., DEVILLERS, R. & FRANK, A. (2006), External data quality of geographical applications: an ontological approach. In: DEVILLERS, R. & JEANSOULIN, R. (Eds.), *Fundamentals of Spatial Data Quality*, ISTE – GIS Series, 255-270.

*Research developed in the context of the FP7 project "BIO\_SOS – Biodiversity Multisource Monitoring System: from Space TO Species" (FP7-SPA-2010-1-263435).*