Towards Multidimensional User Manuals for Geospatial Datasets: Legal Issues, Research Challenges and Technological Perspective

Yvan Bédard[1], Rodolphe Devillers[1,2], Marc Gervais[1,2] & Robert Jeansoulin[3]

[1] Centre de Recherche en Géomatique (CRG), Pavillon Casault, Université Laval, Québec, G1K 7P4, Canada
{Yvan.Bedard, Rodolphe.Devillers, Marc.Gervais}@scg.ulaval.ca
[2] Université de Marne-la-Vallée, Institut Francilien des GéoSciences, France
[3] Laboratoire des Sciences de l'Information et des Systèmes (LSIS), Centre de Mathématiques et d'Informatique (CMI), Université de Provence (Aix-Marseille I), 39 Rue Joliot Curie, 13453 Marseille Cedex 13, France.
Robert.jeansoulin@cmi.univ-mrs.fr

1 Introduction

Over the last two decades, geographic information has been moving through two important revolutions, namely (1) the move to a digital mode and (2) mass consumption of low-cost data. The difficulty for non-expert users to appreciate correctly the quality of the data they use brings several concerns, in particular regarding the risks of mistakes in the data they query, of inadequate use or interpretation, and of incorrect results obtained from processing. Such situation increases the potential of legal disputes between the different parties implied into a transaction or professional service that involves geospatial data. Several cases of data misuse are mentioned in the literature, media and jurisprudence [3]. This new context of mass consumption along with legal trends over the last decade call for improved methods to bring geographic information to the market, for alternatives to the traditional metadata that are cognitively not compatible with mass users, and for better management of legal risks by providing explicit information about the use of a dataset. This latter issue leads to providing professional advices and user manuals like the ones found with mass-market products; this allows to better convey the information necessary for a user to assess the external quality of geospatial datasets (fitness for use).

In this paper, we start with an overview of legal issues which fix some parameters to the development of practical solutions aiming at better informing the users about external data quality. Then, we present the resulting research challenges with regards to data quality. Finally, we present a technological perspective using Spatial On-Line Analytical Processing technology (SOLAP) to build a prototype called MUM (Multidimensional User Manual) to reduce the risks of misuse by providing contextual and aggregated information about data quality, either for the experts who provide the advices or, in some cases, for the end-users.

2 Legal issues

Because geographic information represents an abstraction of the reality, a model of the reality of interest for the user at a given time, it typically suffers from limited accuracy and incompleteness, as well as not being up to date. Furthermore, digital data often give users a false impression of high accuracy, completeness and quality because of its technical nature and the high precision of calculations performed with them (e.g. distance measurement providing six decimals). Based on a survey of different juridical aspects in several countries (e.g. Canada, France, Belgium, USA), we identified a high level of uncertainty in several aspects of today's legislation regarding digital data, and specifically geographic information. Examples of areas where such high uncertainty remains in spite of improved legislation include intellectual property rights, commercial contract to sell data and services, civil liability of data providers, etc. Gervais [3] has presented a detailed analysis of these uncertainties and of the way they create legal uncertainty over geographic data quality and the liability of data providers and users.

In spite of these uncertainties about legal issues, it is possible to identify duties which are mandatory for data providers regarding the quality of data, duties that every professional is expected to do. Among these duties, the data providers can't get away without properly informing the users about the datasets. Consequently, data providers must consider the users' objectives before usage of the data to warn them accordingly, to explain potential defects, risks, controversial interpretation, and so on. Such obligation already exists in several countries with mass market products and is typically regulated by several laws (e.g. professional liability, consumer protection). These legal issues aim at insuring the quality of the products for the mass as well as professional clients and they influence the design of solutions that assess the quality of geographic datasets.

Our research results, which also involved mapping agencies, lawyers and a notary claim that the complex nature of geographic information and today's juridical uncertainties lead to consider the obligation to properly communicate information about the external data quality (instructions on usage) in a restricted-application context as an efficient way to reduce juridical risk. We conclude from that study that, based on juridical considerations, it is impossible to deal with geographic information quality without considering its use. We also conclude that the complexity of geographic information requires (1) an expert to correctly appreciate the data and its quality (internal and external), and (2) custom-made services for data users (i.e. advices), as this already exists in domains dealing with similar complexity (doctor, lawyer, broker, etc.).

3 Research challenges

In this context, different approaches can contribute to help experts or end-users in assessing the fitness for use of their data:

- *Improve datasets*: Production of datasets for certain uses identified *a-priori.* These datasets include objects, attributes, spatial accuracy, completeness, etc. required by the users for the application defined.
- *Encapsulate data within software*: we have also seen such an approach for a few years with mobile navigation systems encapsulating road data with real-time GPS positioning and routing functions. Risks are then be minimized because of the adequacy of the data for the operations offered in these products, and the presence of user-understandable warnings on the use of the product. Such approach benefits non-expert users and follow legal requirements expressed in the section 2;
- *Improve existing GIS software*:
    - o Visualization of quality information. Visualizing quality information can provide a fast insight on potential quality problems. This research area was for instance explored within the NCGIA research initiative 7 "Visualizing the Quality of Spatial Information" led by K. Beard & B. Buttenfield between 1991 and 1993. Numerous ways to visualize data quality were proposed, such as changes in object colors, textures or opacity, use of fuzzy representation for objects, display of a 3-D surface representing the positional variability of data quality, quality sliders applying threshold allowing the display of high quality data only, etc;
    - o Providing warning to users. Such approach could both support non-expert and expert users. Hunter and Reinke [4,5] suggest to provide notifications to the users when an illegal operation is performed (e.g. sound, messages, animation). This requires to identify a set rules linking data quality information to the system operators;
    - o Some authors propose the development of broader approaches by adding the possibility for instance to integrate techniques for handling error within GIS. These researches aim at designing what is named error-sensitive, error-aware or sometimes quality-aware GIS [2,6]. Such systems can for instance integrate AI or advances database techniques to provide additional functionalities to the traditional GIS functions. These tools also require to better formalize the relations between data quality and operators to identify potential risks.
- *Improve existing data selection tools*: instead of protecting the users once they got their geospatial data, it is possible to act earlier, i.e. when users are getting these data. One way to achieve this requires to enhance existing data selection tools (e.g. geospatial digital libraries) to better help users getting data fitting with their intended use. This reaches some orientations of the REVIGIS project, for example to use ontologies to formalize user needs and data characteristics, or using mediators to make the link between data and users;
- *Require the professional opinion of an expert in geomatics*: for complex situations technological improvements may not be sufficient to avoid data misuse.

Users can then consult a geomatics professional who will help them describing their requirements, identifying relevant datasets for the planned application, and describe the possible operations a user could perform using his GIS (ex: distance measurement, object instance count, etc.). Such professional must be liable for his recommendation, delivered in a formal report, and have professional insurances. Such situation already exists in many professions where experts are needed for complex or risky situations;

- *Develop tools to help experts to provide advices*: the above-mentioned geomatics experts need data quality tools helping them to analyze facts, make their opinion and provide advices to non-expert users. These tools need capabilities to integrate, manage, visualize quality information and identify potential risks related to the combinations of data with certain operators.
- *Educate the users*: Instead of bringing systems and data closer to the users, another solution could be to bring users closer to existing data and systems. This roughly means teaching users geomatics concepts such as spatial reference systems, data acquisition and production techniques, precision assessment, problems of external and internal quality, etc.

Each of these approaches contributes to solve the problem of assessing geographic data quality and can be used, in several occasions, in complementarity.

4 Technological perspective

With several of the above issues in mind, a project named Multidimensional User Manual (MUM) was initiated to provide tools that will help users to assess the external quality of their datasets (i.e. compare data specifications to their needs). This work aims at improving existing GIS functionalities by improving management and communication of quality information. This research introduces the concepts of quality indicators that aim at communicating contextual and summarized/aggregated quality information to users. The use of such indicators, structured at different levels of detail, helps avoiding an overload of quality information and can be visualized on a dashboard and offer alerts when quality goes under some thresholds.

In order to be displayed into the system, existing quality information issued from metadata is integrated, summarized/aggregated and structured with different levels of detail within a multidimensional database (such as the ones used in Data Mining). This management of quality information uses a model named QIMM (Quality Information Management Model) [1] that is based on a multidimensional database design. Multidimensional databases, as defined in the database field, are not restricted to spatial and temporal dimensions. These models allow the management of information at different levels of granularity along certain themes (or axes) named “dimensions”, from which a user can analyze the data. This model suggests the use of two dimensions, namely “data dimension” and “indicator dimension” and supports the description of data quality from the global dataset level to the detailed primitives level (geometric and semantic).

This model was implemented into a Spatial OLAP (On-Line Analytical Processing) prototype relying on a multidimensional data structure, associated with SOLAP tools and a cartographic interface. Data from the Canadian National Topographic Database (NTDB) were used for the prototype and quality information was described according to ISO 19113 standard. A graphical interface allows users to select relevant indicators from an indicators database and get their description. This prototype communicates contextual indicators to users, providing qualitative information for different aspects of data quality (e.g. positional accuracy, completeness, logical consistency). Indicators values are based on the difference between user expectations and datasets internal quality. These values are displayed on a dashboard using user-understandable representations (e.g. street lights, speed meter, smileys) and have a cartographic display (thematic mapping of the quality indicators selected by the user) (cf. Figure 1). The cartographic display allows user to get a fast insight on the spatial heterogeneity of quality information. SOLAP operators allow users to easily and rapidly navigate into quality information at different levels of details along data and indicators hierarchies. For instance, a user can look at the average positional accuracy of a dataset (ex: topographic map including several object classes), then drill down to get the average positional accuracy of a single object class (e.g. road network) and drill down again to get it for an instance. Quality indicator values are recalculated each time the user modify the area visualized (e.g. zoom in, out, pan) in order to provide information about the objects visible in the cartographic extent (or within a fence around an ad hoc polygon). Such approach allows users to get contextual and aggregated information related to the quality of data being used and then contributes to reduce the risk of misuse of geospatial data.

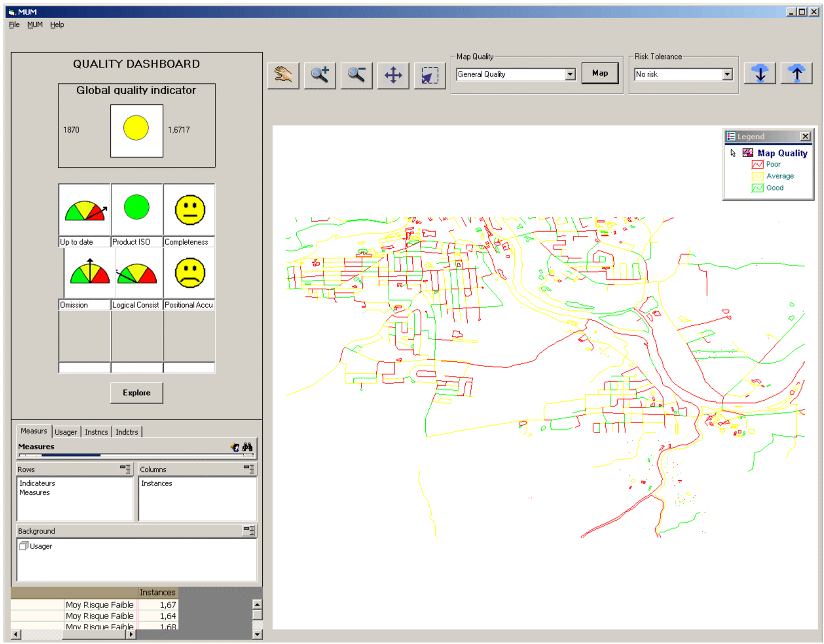

Fig. 1. Example of a MUM prototype interface with quality indicators selected by the user (left) and the mapping of a specific indicator (right)

5 Conclusion

This paper introduced an approach named Multidimensional User Manual (MUM), aiming at decrease the risks of misuse of geospatial data by providing quality information, helping the users to reduce their uncertainty within spatial decision processes. We first discussed different legal issues, related to geospatial data, supporting such approach. We then presented different research challenges that can contribute to solve the problem identified by the juridical context analysis. We finally present our approach that aims at managing quality information into a multidimensional data structure and communicating it through quality indicators. Such tool can support experts in geomatics to assess the external data quality in complex situations. A simplified version of this tool could also support end-users in the context of restricted applications.

Acknowledgements

The authors are grateful to the Ministère de la Recherche, de la Science et de la Technologie du Québec as well as the European Commission project REVIGIS, the

Ministère des Relations Internationales du Québec and the Commission Permanente Franco-Québécoise for their financial contribution.

References

1. Devillers, R., Bédard, Y. and Jeansoulin, R. 2004, Multidimensional management of geospatial data quality information for its dynamic use within Geographical Information Systems. PE&RS Journal. Submitted.
2. Duckham, M. and J.E. McCreadie, 2002, Error-aware GIS Development. In *Spatial Data Quality*, edited by W. Shi, P. F. Fisher and M. F. Goodchild (London: Taylor & Francis), pp. 63-75.
3. Gervais, M. 2003. La pertinence d'un manuel d'instructions au sein d'une stratégie de gestion du risque juridique découlant de la fourniture de données géographiques numériques. PhD thesis, Université Laval, Québec, Canada & Université Marne-la-Vallée, France, 337 p.
4. Hunter, G.J., and K.J. Reinke, 2000, Adapting Spatial Databases to Reduce Information Misuse Through Illogical Operations. In *Proceedings of 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (Accuracy 2000)* (Amsterdam), pp. 313-319.
5. Reinke, K.J. and G.J. Hunter, 2002, A Theory for Communicating Uncertainty in Spatial Databases. In *Spatial Data Quality*, edited by W. Shi, P. F. Fisher and M. F. Goodchild (London: Taylor & Francis), pp. 77-101.
6. Unwin, D., 1995, Geographical information systems and the problem of error and uncertainty. *Progress in Human Geography*, 19, pp. 548-549.