

Towards Multidimensional User Manuals for Geospatial Datasets: Legal Issues and their Considerations into the Design of a Technological Solution

Yvan Bédard¹, Rodolphe Devillers^{1,2}, Marc Gervais^{1,2} & Robert Jeansoulin³

¹Centre de Recherche en Géomatique (CRG), Pavillon Casault, Université Laval, Québec, G1K 7P4, Canada
{Yvan.Bedard, Rodolphe.Devillers, Marc.Gervais}@scg.ulaval.ca

²Université de Marne-la-Vallée, Institut Francilien des GéoSciences, France

³Laboratoire 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. Although this is generally seen as very positive, several authors have pointed to the difficulty for non-expert users to appreciate correctly the quality of the data provided by mapping organisations (e.g. Hunter and Goodchild 1996; Onsrud 1997, 1999; Phillips 1999; Gervais 2003). In particular, there are risks of mistakes in the data the users query, of inadequate use or interpretation and of incorrect results obtained from the processing of these data. Such a situation increases the potential of legal disputes between the different parties implied into a transaction or professional service that involves geospatial data. A recent study of the literature in GIScience, in Law and of 225 court decisions in Canada, US, France and other countries (Gervais 2003) has shown that several cases of data misuse have led to disputes between the users and providers of data (non-spatial and spatial data) and that several laws are involved into the protection of consumers. 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 and for better management of legal risks by providing explicit information about the use of a dataset. This new context also calls for alternatives or additions to the traditional distribution of detailed metadata which proved to be cognitively incompatible with mass users. Gervais (2003) has demonstrated that from a legal point of view, one solution is to provide extended instructions manuals that are understandable by the users. He has identified the content of such manuals in order to protect the consumer as well as the provider of geospatial data. Based on this proposed content, we have designed and developed a computer-based manual that helps the users to analyze the quality of the data they see or use. In order to have an intuitive tool supporting quality information at different levels of granularity (i.e. detailed metadata vs global quality indicators) and very fast answers, we have used a multidimensional database approach such as the one used in Business Intelligence, data warehousing and SOLAP (Spatial On-Line Analytical Processing) (see Bedard et al. 2001 for further description). The prototype is called MUM (Multidimensional User Manual).

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 an overview of existing approaches with regards to data quality as they all contribute their way to meet legal requirements. Finally, we present the technological perspective we used to build the MUM prototype that aims at reducing the risks of misuse by easily and rapidly providing

contextual and aggregated information about data quality. MUM is intended mostly to facilitate the analysis by the experts who provide advices to the users, but also for the end-users in some cases.

2 Legal issues

Geographic information is an abstraction of the reality, it is obtained from the interpretation of a model made from the data that represent the geographic phenomena of interests to given users at a given time for a given purpose. Geographic information typically suffers from limited accuracy and incompleteness, and it is frequently not up to date. Furthermore, digital data often give users a false impression of high accuracy, completeness and quality because of their technical nature and the apparently high precision of calculations performed with them (e.g. distance measurement providing six decimals with a commercial GIS).

In his PhD thesis, Gervais (2003) analyzed several legal aspects that relate to geospatial data quality in order to define an efficient way of communicating information about geospatial data quality that would better protect both data users and providers. Benefiting from 20 years of experience as a Quebec Registered Land Surveyor and as a Court expert, his research involved lawyers and a notary in addition to geomatics engineers, computer scientists and official mapping agencies. He analyzed civil rights, licensing contracts (e.g. shrink-wrap, click-wrap), professional liability, investment protection, commercial obligations and copyright issues as well as 225 court decisions (after a first overview of 1400) in several countries (e.g. Canada, France, Belgium, USA). In all jurisdictions, he identified a high level of uncertainty with regards to today's legislation and digital data, and more specifically with geographic information. Examples of areas where such uncertainty remains despite improved legislation include intellectual property rights, commercial contracts to sell data and services, civil liability of data providers, etc.

In spite of these uncertainties, 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 cannot get away without properly informing the users about the datasets. Consequently, data providers must consider users' intended usage of the data and must warn them accordingly. Data providers must 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. For example, when one analyses into details these legal requirements and draws a parallel with other mass market products, he can identify the obligation for a data provider to include a user manual, or instruction manual, that is written in a language understandable by the expected users (Masse 1995; Baudouin and Deslauriers 1998; Le Tourneau 2001; Le Tourneau and Cadiet 2002; Quebec Civil Code Art. 1468 and 1718; several court decisions). In addition, this user manual must clearly prevent the user from dangerous usages and must be easily accessible.

Based on this context, a typical user manual must include at the minimum the followings:

- Licensing (typically adopted by the industry);

- Guarantees (typically used to limit the responsibility of the data providers and the quality of the data);
- Installation instructions (Le Tourneau 1995, 2001 and 2002; Balbo-Izam 2001): procedures and technical standards to follow for data feeding and updating;
- Detailed description of the product (Montero 1998): spatial and temporal coverages, thematic extent (e.g. database model, ontology) including object classes, attributes, domains, relationships and their definitions;
- Spatial, temporal and thematic resolution of the data (adapted from Montero 1998);
- General advices, verifications to do before day-to-day operation;
- Recommended usages: precautions to take when using the data (Baudouin and Deslauriers 1998; Baudouin and Jobin 1998), description of allowed operations (Montero 1998; Westell 1999), limits of the results (Marino 1997; Baudouin and Deslauriers 1998; Montero 1998), random defects and uncertainty (Trottier 1988; Marino 1997), probability of each potential risk (Agumya and Hunter 1999 and 2002; Hestler 1999; Nicolo 1999; O'Donnell and Olivier 1999), archiving practices (Le Tourneau 2002), usages to avoid (Baudouin and Deslauriers 1998), potential damages in case of erroneous usage (Baudouin and Deslauriers 1998), technical constraints to apply (Le Tourneau and Cadiet 2002);
- Warnings and security advices: care to take when using the data (Marino 1997; Montero 1998), differing points of view about the value of the data (Montero 1998; Haumont 2000), stability of the behaviour of the data when they are processed (Le Tourneau and Cadiet 2002), hidden dangers inherent to the data (Baudouin and Deslauriers 1998; Baudouin and Jobin 1998; Lefebvre 1998; Montero 1998; Le Tourneau 2001);
- After-sale service and support (Le Tourneau 2001 and 2002);
- Technical specifications (mostly the typical standard metadata);
- Index of the overall content for easy finding of the needed information.

Some of these obligations are included in today's metadata, but the obligation to explain them in a user-understandable language and to make them easily accessible are no obvious tasks. The complex nature of geographic information and today's juridical uncertainties lead us to consider the obligation to properly communicate information about the external data quality (instructions on usage) in a restricted-application context as a major challenge but also as an efficient and mandatory way to reduce juridical risk. As it is impossible to deal with geographic information quality without considering its use, data quality cannot be considered alone in a generic manner. In fact, legal obligations and court decisions convinced us that the complexity of geographic information requires (1) an expert to correctly appreciate the quality of geospatial data (internal and external), and (2) custom-made services for data users (i.e. advices), as this already exists in domains dealing with similar complexity (e.g. doctors, lawyers, brokers).

In the next section, we look at the different approaches that exist today to deal with data quality while the last section presents an overview of one solution we developed in order (1) to provide immediate access to quality information when one uses a GIS, (2) to provide detailed information as well as aggregated, synthesized quality information (e.g. using quality indicators), (3) to offer an intuitive user interface with the point and click approach typical of Spatial On-Line Analytical Processing and the speed of multidimensional databases, (4) to support spatial heterogeneity in the quality of data, and (5) to offer different ways to visualize this quality information using maps, tables, charts and to easily swap between these views.

3 Overview of existing approaches

Different approaches can contribute to help experts or end-users assessing the fitness for use of a dataset or help limiting the risks of data misuse. These approaches can be used alone or in combination with other approaches:

- *Tailor-made datasets*: in several cases, datasets can be built for uses identified *a-priori* and their usage restricted to these. 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 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. (e.g. Buttenfield, 1993; McGranaghan, 1993; Beard, 1997; Drecki, 2002).
 - o Providing warnings to users. Such an approach could both support non-expert and expert users. Hunter and Reinke (2000; 2002) suggest to provide notifications to the users when an illegal operation is performed (e.g. sound, messages, animation). This requires to identify a set of 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 (Unwin, 1995; Duckham and McCreddie, 2002). 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 (Lassoued et al., 2003).
- *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 (e.g. distance measurement, object instance count, etc.). Such a professional must be li-

able for his or her recommendation, delivered in a formal report, and have professional insurances. Such a 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, complementarily.

4 Technological perspective

With several of the above issues in mind, a project named Multidimensional User Manual (MUM) was initiated to provide a tool that will help users 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 in accordance to the legal requirements identified earlier. As such, MUM will include all functions necessary to meet these requirements once the R&D is completed. MUM also uses 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 avoid an overload of quality information and can be visualized on a dashboard and offer alerts when quality falls below some thresholds.

In order to be displayed with an information 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) (Devillers et al., 2004) 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 the ISO 19113 standard. A graphical interface allows users to select relevant indicators from an indicators database and get their descrip-

tions. This prototype communicates contextual indicators to users, providing qualitative information for different aspects of data quality (e.g. positional accuracy, completeness, logical consistency). Indicator 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 a user to get a fast insight into the spatial heterogeneity of quality information. SOLAP operators allow users to easily and rapidly navigate through 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 (e.g. 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. Figure 1 provides an example of user mindstream when navigating through quality information. Quality indicator values are recalculated each time the user modifies 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 an approach allows users to get contextual and aggregated information related to the quality of data being used and then contributes to reducing the risk of misuse of geospatial data.

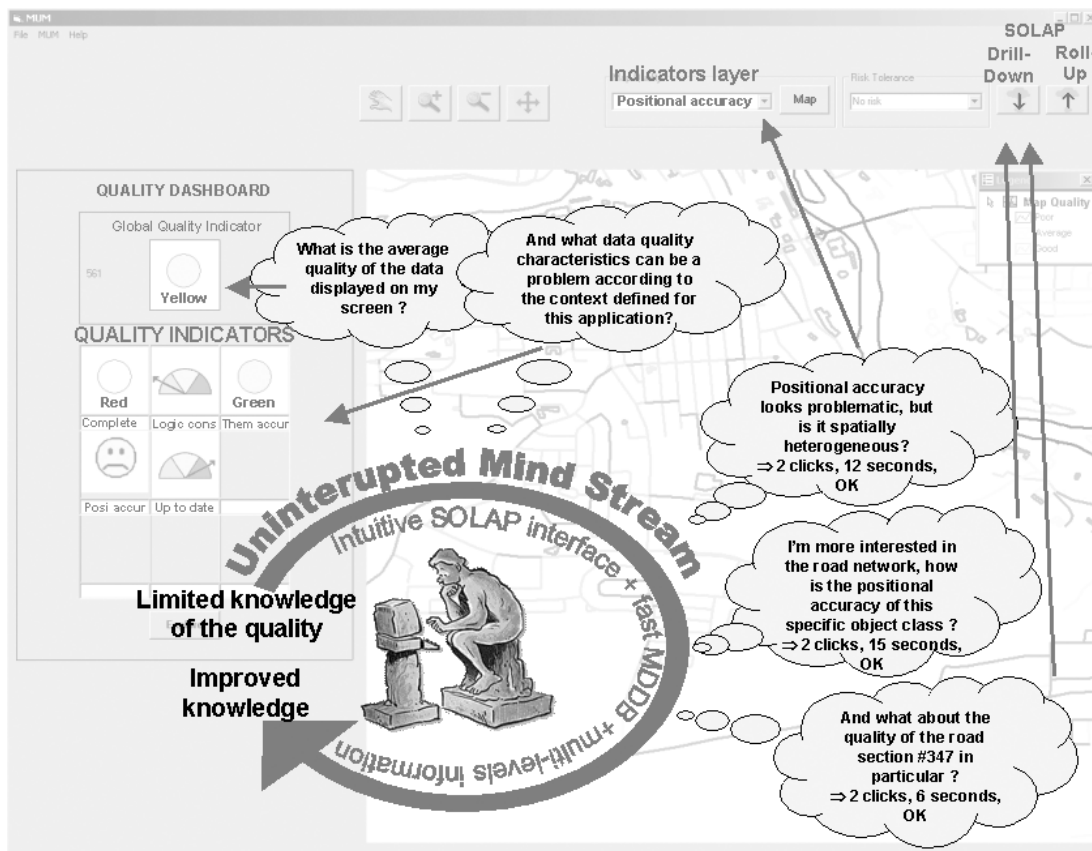


Fig. 1. MUM prototype interface with a user improving his knowledge about data quality through navigation within quality information.

5 Conclusion

This paper introduced an approach named Multidimensional User Manual (MUM), aiming at decreasing 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 an approach. We then presented different research challenges that can contribute to solve the problem identified by the juridical context analysis. We finally presented our approach that aims at managing quality information in a multi-dimensional data structure and communicating it through quality indicators. Such a 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

- Agumya, A. and G. J. Hunter. (1999), "Assessing fitness for use of geographic information: What risk are we prepared to accept in our decisions ?" In: K. Lowell & A. Jaton (eds), *Spatial Accuracy Assessment, Land Information Uncertainty in Natural Ressources*, Sleeping Bear Press Inc., Québec, pp. 35-43.
- Agumya, A. and G. J. Hunter. (2002), "Responding to the consequences of uncertainty in geographical data" *International Journal of Geographical Information Science*, vol. 16, n°5, pp. 405-417.
- Balbo-Izarn, N. (2001), "*Le professionnel face aux risques informatiques*" Petites Affiches, n° 34, P. 4, <http://jurisguide.univ-paris1.fr>.
- Baudouin, J.-L. and P. Deslauriers. (1998), "*La Responsabilité Civile*", Les Éditions Yvon Blais Inc., Cowansville, 1684 p.
- Baudouin, J.-L. and P.-G. Jobin. (1998), "*Les Obligations*", Les Éditions Yvon Blais Inc., Cowansville, 1217 p.
- Beard, K. (1997), "Representations of Data Quality". In: M. Craglia, and H. Couclelis (eds.) *Geographic Information Research: Bridging the Atlantic*, Taylor and Francis, pp. 280-294.
- Bédard, Y., T. Merrett and J. Han. (2001), "Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery". In: H. Miller and J. Han (eds.), *Geographic Data Mining and Knowledge Discovery*, Research Monographs in GIS series edited by P. Fisher and J. Raper, Taylor & Francis, pp. 53-73.
- Buttenfield, B.P. (1993), "Representing Data Quality". *Cartographica*, v.30, pp. 1-7.
- Devillers, R., Y. Bédard and R. Jeansoulin. (2004), "Multidimensional management of geospatial data quality information for its dynamic use within geographical information systems". *PE&RS Journal*. Submitted.
- Drecki, I. (2002), "Visualisation of Uncertainty in Geographic Data". In: W. Shi, P. F. Fisher, and M. F. Goodchild (eds.), *Spatial Data Quality*, Taylor & Francis, pp. 140-159.
- Duckham, M. and J.E. McCreddie. (2002), "Error-aware GIS Development". In: W. Shi, P. F. Fisher and M. F. Goodchild (eds.). *Spatial Data Quality*, London: Taylor & Francis, pp. 63-75.

- 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.
- Haumont, F. (2000), "L'information environnementale : la responsabilité des pouvoirs publics" In: B. Dubuisson and P. Jadoul (eds), *La responsabilité civile liée à l'information et au conseil*, Publications des Facultés universitaires Saint-Louis, Bruxelles, pp. 103-146.
- Hesler, W. (1999), "La responsabilité du courtier en valeurs mobilières au service du particulier" In *La responsabilité civile des courtiers en valeurs mobilières et des gestionnaires de fortune : aspects nouveaux*, Les Éditions Yvon Blais Inc., Cowansville, pp. 63-88.
- Hunter, G. J. and M. F. Goodchild. (1996), "Communicating uncertainty in spatial databases". *Transactions in GIS*, Vol. 1, n°1, pp. 13-24.
- Hunter, G.J. & K.J. Reinke. (2000), "Adapting Spatial Databases to Reduce Information Misuse Through Illogical Operations". In: M. J. P. M., Lemmens and G. B. M. Heuvelink (eds.). *Proceedings of 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (Accuracy 2000)*, Amsterdam, pp. 313-319.
- Lassoued, Y., R. Jeansoulin and O. Boucelma, (2003), "Mediateur de Qualité dans les Systèmes d'information géographiques", In: *SETIT International conference (Sciences Electroniques, Technologies de l'Information et des Télécommunications)*, pp. CDindex-214, Sousse, Tunisia.
- Lefebvre, B. (1998), *La bonne foi dans la formation des contrats*, Les Éditions Yvon Blais Inc., Cowansville, Canada, 304 p.
- Le Tourneau, P. (1995), *La Responsabilité Civile Professionnelle*, Economica, Paris, 105 p.
- Le Tourneau, P. (2001), *Responsabilité des vendeurs et fabricants*, Droit de l'entreprise, Les Éditions Dalloz, Paris, 242 p.
- Le Tourneau, P. (2002), *Contrats informatiques et électroniques*, Dalloz reference, Les Éditions Dalloz, Paris, 268 p.
- Le Tourneau, P. and L. Cadiet. (2002), *Droit de la responsabilité et des contrats*, Dalloz Action, Éditions Dalloz, Paris, 1540 p.
- Masse, C. (1995), "Les premières tendances à signaler en ce qui a trait au nouveau droit de la responsabilité civile" In: *Développements récents en droit civil (1995)*, Les Éditions Yvon Blais Inc., Cowansville, pp. 47-69.
- Marino, L. (1997) *Responsabilité civile, Activité d'information et Médias*, Presses Universitaires d'Aix-Marseille et Économica, Aix-en-Provence, 380 p.
- McGranaghan, M. (1993), "A cartographic View of Spatial Data Quality", *Cartographica*, v.30, pp. 8-19.
- Montero, É. (1998), *La responsabilité civile du fait des bases de données*, Les Presses universitaires de Namur, Belgium, 564 p.
- Nicolo, M.-J. (1999), "La conformité" dans *La responsabilité civile des courtiers en valeurs mobilières et des gestionnaires de fortune : aspects nouveaux*, Les Éditions Yvon Blais Inc., Cowansville, Canada pp. 89-104.
- O'Donnell, J. V. and A. Olivier. (1999), "Les grandes tendances de la jurisprudence récente" In: *La responsabilité civile des courtiers en valeurs mobilières et des gestionnaires de fortune : aspects nouveaux*, Les Éditions Yvon Blais Inc., Cowansville, Canada, pp. 1-34.
- Onsrud, H. J. (1997), "Ethical Issues in the Use and Development of GIS". In *GIS/LIS '97 Proceedings*, ACSM/ASPRS.
- Onsrud, H. J. (1999), "Liability in the use of geographic systems and geographic datasets". In P. A. Longley, M. F. Goodchild, D. J. Maguire et D. W. Rhind (eds.). *Geographical Information Systems : Principles, Techniques, Applications, and Management*, 2nd edition, John Wiley & Sons, Vol. 2, pp. 643-652.
- Phillips, J. L. (1999), "Information Liability: The Possible Chilling Effect of Tort Claims Against Producers of Geographic Information Systems Data", *Florida State University Law Review*, vol. 26, n°3, pp. 742-781.

- Qiu, J. and G. J. Hunter. (2002), "A GIS with the Capacity for Managing Data Quality Information". In W. Shi, P. F. Fisher and M. F. Goodchild (eds.). *Spatial Data Quality*, London: Taylor & Francis, New York, pp. 230-250.
- Reinke, K.J. and G.J. Hunter. (2002), "A Theory for Communicating Uncertainty in Spatial Databases". In: W. Shi, P. F. Fisher and M. F. Goodchild (eds.). *Spatial Data Quality*, London: Taylor & Francis, pp. 77-101.
- Trottier, B. (1988), "*La distribution de masse des progiciels de micro-informatique : de l'opposabilité du contrat standard de licence et de la responsabilité contractuelle du fabricant en cas de défauts inhérents*", Master thesis, Faculté de Droit, Université Laval, Ste-Foy, 98 p.
- Unwin, D. (1995), "Geographical information systems and the problem of error and uncertainty". *Progress in Human Geography*, vol. 19, pp. 548-549.
- Westell, S. (1999), "Potential liability for defective software or data – Part 1", *GIS & The Law*, Adams Business Media, <http://www.geoplance.com>.