TOWARDS MANAGING THE RISKS OF DATA MISUSE FOR SPATIAL DATACUBES

M-A. Levesque [a,b], Y. Bédard [a,b], M. Gervais [a], R. Devillers [c,a]

[a] Dép. des sciences géomatiques, Université Laval, Québec (QC), G1K 7P4 Canada – marie-andree.levesque.1@ulaval.ca, (yvan.bedard, marc.gervais)@scg.ulaval.ca
[b] Canada NSERC Industrial Research Chair in Geospatial Databases for Decision Support
[c] Dept. of Geography, Memorial University of Newfoundland, St. John's (NL), A1B 3X9 Canada – rdeville@mun.ca

KEYWORDS: Decision support, Spatial OLAP, Spatial data quality, Warnings, Legal, Misuse, Risk

ABSTRACT:

Over the years, the mass consumption of spatial data caused several concerns in the geomatics community about the risk of data to be misused, especially by people who have little expertise in spatial referencing methods and their impact on spatial analysis. These concerns increased recently with the arrival of a new category of spatial decision-support applications, called Spatial OLAP or SOLAP. These applications add a spatial component to the traditional OLAP (On-Line Analytical Processing) tools, which are one of the most widely used BI (Business Intelligence) solutions. They allow users to easily, quickly, and interactively explore spatial data of different themes and at different levels of detail. Such easiness, interactivity, speed and flexibility improve users' capability to analyze spatial data and support decisions. However, it opened the door to a larger group of users who may not completely be aware or understand the inherent strengths and weaknesses of the data. The main objective of this paper is to propose a generic approach to help producers identifying potential risks of spatial datacube misuse and different ways to manage them, such as the communication of context-sensitive warnings to end-users. The proposed approach is mainly inspired by risk management theories found in the field of project management and ISO standards. It also considers the data producers' legal duties towards the end-users which exist in Europe and America. These legal duties are mainly to inform, advise, and warn the users in their utilization of the data. Although specifically developed for spatial datacubes (SOLAP applications), several aspects of this approach are relevant for GIS and universal server transactional databases.

1. INTRODUCTION

The use of spatial data significantly changed during the last two decades. Spatial data are now used in a growing number of organizations and for many types of applications. People can easily access spatial data at low cost through various Web portals (e.g. Digital Libraries, Spatial Data Infrastructures, Web-based solutions such as Google Earth). At the same time, Geographic Information Systems (GIS) continuously improve their user-friendliness and spatial data increasingly penetrate decision support processes. This increasing mass consumption of spatial data has however caused concerns related to the risk of data misuse (Hunter and Goodchild, 1996; Hunter, 1999; Onsrud, 1999; Phillips, 1999), especially by people who have little expertise in spatial referencing methods and their impact on spatial analysis (Gervais, 2004).

These concerns increased recently with the arrival of a new category of spatial decision-support applications: Spatial OLAP or SOLAP (see Rivest *et al.*, 2005 for a detailed description). These applications add a spatial component to the traditional OLAP (On-Line Analytical Processing) tool, which is one of the most widely used BI (Business Intelligence) solutions. OLAP tools were initially introduced to solve some limitations of the traditional transactional systems (i.e. OLTP - On Line Transaction Processing – such as RDBMS and GIS) to support aggregated information, rapid comparisons in space and time, trends and knowledge discovery, quick response to unforeseen queries and other complex operations needed during tactic and strategic decision-making processes. With their architecture supported by a multidimensional database or "datacube", OLAP and SOLAP tools offer functionnalities to specifically fulfill those needs not found in transactional systems. Thus, while the GIS technology supports organizations' daily transactions, the SOLAP technology duplicates, restructures and exploits the data resulting from these transactions to better support decision processes.

More specifically, SOLAP applications provide functions to easily and interactively explore and cross-tabulate spatial data from different themes and at different levels of information granularity. For example, a health professional could, with only a few mouse clicks and within 10 seconds (i.e. Newell's cognitive band; Newell, 1990), have an answer to a query such as "What is the incidence rate of the lung cancer for the group of 40 to 55 years old women in the province of Quebec in the year 2005?". He could then apply a roll-up function to analyze this rate at the scale of the country of Canada for the last 5 years, or drill-across to make comparisons with other Canadian provinces, or drill-down to have a detailed regional view of the data. The results of these queries can be visualized using various synchronized forms: one or several maps, charts (e.g. pie charts, bar charts, diagrams) or tables.

In the recent years, tens of SOLAP applications have been introduced into public/private organizations in different countries and in various fields (health, transport, forestry, archaeology, defence, search and rescue, etc.) to better support their decisional processes. Becoming aware of improved users' capabilities to analyze spatial data, some organizations have expressed important concerns about the potential risk of data misuse by non-informed SOLAP users. Of particular concerns are:

- *Ease of use*: With their user-friendliness and intuitive interface, SOLAP applications allow users to directly interact with the data without having to learn a specific query language (e.g. SQL) or request the help of an expert (e.g. GIS specialist). This ease of use has however opened the door to a larger group of users who, when analyzing the data, may not completely be aware or understand the inherent characteristics and limitations of the data. Moreover, such ease of use can lead to a false feeling of accuracy and quality, and do not encourage users to adopt an informed behavior towards the data.

- *Importance for organizations*: As SOLAP tools are designed to support decision-oriented analysis processes, they are generally used at the organisational tactical and strategic levels, whereas GIS are often used at the operational level. Decisional systems are thus more likely to occupy a critical position in an organization since strategic decisions are likely to be more far-reaching in scope and consequences (Ponniah, 2001). Decisions made on the basis of incorrect deductions or interpretations of data could thus lead to stronger adverse consequences for the organization.

The need for a protection against potential data misuse is not a recent issue the spatial field (see the section 3 for an overview of proposed approaches) and more generally in any field using data to based their decisions. For example, Statistics Canada includes warnings into census statistics tables[*] made available on the Web to warn people about data uncertainty, their incompatibility regarding temporal comparisons, etc. The *Institut national de santé publique du Québec (INSPQ)*, a canadian provincial organization mainly in charge of producing health publications, also includes warnings in its reports tables and graphics[**]. With SOLAP applications, those static graphics and tables become dynamic or even maps, where people could easily and quickly interact with. We are thus facing the same issue than these other fields, but probably in a more critical way resulting from the easiness, speed of analysis, interactivity and flexibility offered by the SOLAP systems.

As government agencies took some preventing actions to reduce the risks of data misuse, we could ask ourselves what could be the role and responsibility of the spatial datacube producers in the prevention of potential misuse. As opposed to what some still believe, data producers have important legal duties towards users, which are mainly to inform and advise them in their use of the data and warn them against potential problems that could result from this use (see section 2 for details). Thus, the prevention of the potential data misuse does not only lay on the users' shoulders; producers have also an important role to play in this action and we intend to help them with our solution.

With those legal aspects in mind, this paper proposes a generic approach to help spatial data producers identifying potential risks of spatial datacube misuse and different ways to manage them, such as the communication of context-sensitive warnings to end-users. Although specifically developed for spatial datacubes (SOLAP applications), several aspects of this approach are also relevant for GIS and universal server transactional databases. In this paper, we first present in more details the liabilities of data producers with regard to the use of geographic information. We then summarize the approaches that have been proposed over the years in both the geomatics and the BI fields to prevent risks of data misuse. Finally, we describe our approach, that is a risk management approach mainly inspired from the field of project management and some ISO standards (ISO/IEC Guide 51, 1999; ISO 3864-2, 2004), and we conclude.

[*] http://www12.statcan.ca/english/census06/release/index.cfm
[**] http://www.inspq.qc.ca/pdf/publications/portrait_de_sante.asp

2. LEGAL ASPECTS

Some researchers (Gervais, 2004; Gervais, 2005; Gervais *et al.*, 2007) expressed the need to better define data producers' attitude towards end-users in the context of the mass consumption of the geographic information. Their work is based on the analysis of legal principles coming from the civil law, the common law and 225 court decisions from several North American and European countries (e.g. Canada, France, Belgium, USA).

Many duties that should be legally mandatory for data producers or geomatics professionals have been identified. Generally, data producers must show *care* and *diligence*. More specifically, one of the most important duties that a data producer should fulfill is to properly *inform* users. The importance of such a duty would vary according to the context, the nature of the product, and the knowledge of the consumer (or user). Due to the complexity, the technical nature, and the potential dangerousness of geographic information combined with the user's possible lack of expertise in spatial referencing, the importance of the duty for the producer to properly inform users in the context of mass consumption would be particularly high. The same conclusion can also be made in the context of SOLAP tools, as they share similar characteristics with GIS.

Further investigations of this duty to *inform* revealed that it involves *advising* and *warning*. Properly *informing* requires revealing a fact without necessarily orienting the decision of the user. In the context of geographic information, this duty would primarily consist in communicating information included in the metadata or any other description of the data that could be necessary to understand the characteristics of a given dataset. The duty to *advise* is characterized by the need to judge (Lefebre, 1998) the content of the information transmitted with regard to its type of information (i.e. digital geographic information), its importance, and intended use. As for the duty of *warning*, it is a constant duty, particularly when a good presents potential dangers related to its use. Warnings should be communicated to prevent misuses and allow users to adequately appreciate the value of the data. Warnings should also be clear, complete, and up to date. Finally, the nature of the duty to *inform*, *advise* and *warn*, requires that data producer understand the needs, objectives or even the intended use. In fact, this issue has been highly studied in the law and mass consumption literature, and has led to international standards on warnings (ISO/IEC Guide 51, 1999; ISO 3864-2, 2004). Although such standards cannot fit the geomatics community needs perfectly, they provide useful guidelines and cannot be ignored.

According to Gervais (2004), the duty of *advising* and *warning* also implies the duty of identifying and revealing potential *risks* related to the use of given data. This duty already exists for other types of professionals, such as physicians, brokers, real estate brokers, legal advisers, and even the database producers (Le Tourneau, 1995; Beaudouin and Deslauriers, 1998; Montero, 1998; Le Tourneau and Cadiet, 2002). From a legal standpoint, there is no reason why it would not apply to spatial data producers and geomatics professionals. Evolving in such direction indicates a higher level of maturity for a field and a better integration into the society.

3. OVERVIEW OF EXISTING APPROACHES

3.1 Approaches to prevent spatial data misuse

The transmission of metadata to end-users is currently the most adopted approach by data producers to support an adequate usage of spatial data. However, different studies raised several limitations related to the usage of metadata. Users rarely consult them (Frank, 1998; Qiu and Hunter, 2002) and if they do, only partially (Dassonville *et al.*, 2002). Metadata are considered as a technical description of spatial datasets (Walford, 2002), rather than understandable information intended for end-users (Timpf *et al.*, 1996; Harvey, 1998). Metadata would also not be sufficient to respect data producers' most important legal duties towards users (c.f. section 2 on legal aspects) (Gervais, 2004).

Considering these limitations, different approaches have been proposed, often complementary with each other, to help experts and end-users assessing the fitness for use of a dataset and reducing the risk of spatial data misuse (see Bédard *et al.*, 2004 for a description of several approaches). Some approaches recommend taking action before data come into the hands of the end-users (i.e. *a-priori*), while others can be applied after (i.e. *a posteriori*).

Some *a-priori* approaches propose to restrict the usage of a dataset, such as identifying *a-priori* the appropriate uses and restrict the usage to these. Other solutions suggest encapsulating the data within a software to ensure the adequacy between the data and the operations available in the product (e.g. GPS routing devices). Others propose to help users to better identify the data fitting their intended use by improving existing data selections tools (e.g. those used to access spatial data libraries) (Lassoued *et al.*, 2003), requiring the professional opinion of an expert in geomatics (Gervais, 2004) or even, to develop tools that could help this expert to cope with such complex task. Such tools would give the expert the capabilities to integrate, manage and visualize data quality at different levels of granularity (Devillers *et al.*, 2005; Levesque, 2007).

A posteriori approaches propose, amongst other things, to improve GIS software capabilities to better manage and use spatial data quality information. As an alternative to traditional metadata, some authors suggest to communicate spatial data quality information using visualizations techniques, such as changing the color, opacity or texture of objects, displaying a 3-D surface representing the positional variability of data quality, etc. (McGranaghan, 1993; Beard, 1997; Drecki, 2002; Devillers and Beard, 2006). Other authors suggest to provide warnings to end-users when illogical operations are performed (e.g. measuring a distance without having the map units defined) (Beard, 1989; Hunter and Reinke, 2000). Some researchers also proposed to design GIS capable of offering basic functionnalities to handle errors. Depending on their capabilities, such GIS would be named error-sensitive GIS (Unwin, 1995) or error-aware GIS (Duckham and McCreadie, 2002). Finally, some authors have talked about the design of quality-aware GIS, that would be defined as a "GIS with the added capabilities to manage, update, explore, assess and communicate quality information [...] by also addressing issues related to GIS user contexts and use patterns (e.g. user profile and needs assessment)" (Devillers *et al.*, 2007).

3.2 Approaches to prevent datacube misuse

A lot of research has been done on data quality since OLAP and data warehouses (or datacube) approaches became more popular at the beginning of the 90s. Many approaches have been proposed to improve or ensure the quality (i.e. both intrinsic quality and fitness for use) of the datacube during its design process (Wang *et al.*, 1995; English, 1999). Other works have also been done to define quality indicators (Kimball, 2000) or quality dimensions to better evaluate or characterize datacubes' quality (Ballou and Tay, 1989; Wang *et al.*, 1995). Other approaches proposed to model this quality, such as the Goal Question Metric (GQM) approach (Basili *et al.*, 1994), or a quality metamodel (Vassiliadis, 2000). Other works have also been done to identify specific situations where end-users could potentially misuse or misinterpret the data when using specific OLAP operators (Lenz and Shoshani, 1997; Lenz and Thalheim 2006). Some authors have also proposed to constrain or inform end-users when potentially inaccurate summaries could be showed by the decisional system (Horner *et al.*, 2004).

If the research community has been concerned by datacubes quality for many years, it was not the case for the BI industry. A study published in 2002 by The Data Warehousing Institute (TDWI) revealed that poor data quality costs about $600 billion per year to U.S. businesses (Eckerson, 2002). However, a recent paper published by the same organization revealed that companies only start taking quality issues seriously. Data quality has been identified as the main trend in the BI field for the year 2006 (Knightsbridge, 2006).

4. COMMUNICATING RISKS TO USERS

4.1 Risk management

As mentioned earlier, the data producers or geomatics professionals legal duty to communicate advices and warnings to end-users requires identifying and revealing potential risks related to the use of a given dataset. What does the term "risk" mean exactly? In fact, various definitions of this word can be found, depending on the context or the discipline in which it is used. In our research, we adopted the definition coming from the ISO/IEC Guide 51 (1999) international standard, which provides guidelines for the inclusion of safety aspects in standards. The notion of 'risk' thus relates to the "combination of the probability of occurence of harm and the severity of that harm" (ISO/IEC Guide 51, 1999). We consider the term 'harm' as a data misuse or misinterpretation.

The identification of potential risks related to the use of a datacube and how to react to these risks can quickly become complex. Datacube production involves many decisions and data manipulations that can impact the way data should be considered and interpreted. We thus suggest to use a risk management approach to reduce this complexity and consequently, to deal with the different types of risks that could be discovered during the design and feeding of the datacube.

A similar approach has already been proposed in the past by Agumya and Hunter (1999) for transactional spatial data. However, our approach proposes to identify potential risks (1) *during* the design process, (2) for spatial datacubes (or databases), and (3) in a more preventive mode, thanks to the *raison d'être* and capabilities of datacubes. Enriching system development cycles such as the OMG Model-Driven Architecture* or IBM Rational Unified Process** with a risk management process regarding potential spatial data misuse is a novel approach and has never been attempted for spatial datacubes. Since in practice the totality of risks cannot be identified *a priori*, the two above approaches are complementary and could be used jointly.

4.2 Spatial datacube risk-based approach

The risk-based approach we are presenting in this section is inspired from the one given in the ISO/IEC Guide 51 (1999) standard that aims at reducing the risk arising from the use of products, processes or services and also by risk management approaches found in the field of project management (Courtot, 1998; Kerzner, 2006).

More specifically, the approach (cf. Figure 1) is a continuous and iterative process that goes on during the phases of the datacube development cycle, from needs analysis to design, implementation and feeding, with stronger emphasis during the last two stages. This approach successively consists into identifying and evaluating potential risks of misuse or misinterpretation, prepare responses towards these risks, and finally document the risk management process as required for quality audits.

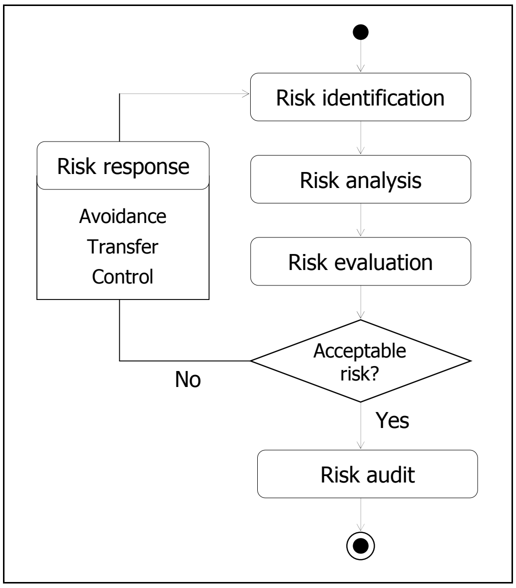

Figure 1 Risk-based approach (a*dapted from* ISO/IEC Guide 51, 1999 and Courtot, 1998)

Potential risks related to the use of a spatial datacube can be illustrated using the following simple example. Consider a university president who wants to analyze the origin of students (e.g. country and degree) who attended his university over the last 20 years to make better decisions for future recruitment. To do this, he contracted a geomatic firm which has made the proposition of building a SOLAP application that will meet his requirements. In the following paragraphs, we will illustrate how the datacube producer might use the risk management process in order to detect and manage potential risks of misuse.

1. Risk identification

Risk identification is the most important step in the risk management process. As mentioned by Agumya and Hunter (1999): "a risk identified is a risk controlled". The main goal here is to find, in the most exhaustive manner possible, what could go wrong during the analysis and the interpretation of the data. Several techniques can be used to make an inventory of the potential risks: the analysis of existing documentation about the data to be integrated in the decision-support system (i.e. metadata, data dictionary), the quality and characteristics of the conceptual and implementation models of the datacube to be produced, the extraction, transformation and loading (ETL) processes used to populate this datacube, the functions used to aggregate the data, the analysis of existing practices of warning (e.g. notes below tables, report forewords, restricted access), etc.

One of our research goals is to provide a classification of potential datacube's risks and to develop a questionnaire to support datacube designers in the complex task of identifying and documenting potential risks. For example, the designer could systematically detect a risk related to the *potential of data to lead to inconsistent temporal analysis* by asking the following question suggested by the guidelines of our solution: *did any change occur in the geometry, the semantic or the attributes of the datacube dimension members*? *Did any change occur in the structure or semantics of the data source? Etc*. These are the types of questions an experienced datacube designer already goes through in an informal manner but which typically get undocumented and not systematically communicated to the users, leading to higher risks of misuse and increased efforts for *a posteriori* prevention.

In our previous example, the designer could have detected a change in the administrative boundaries of the ex-U.S.S.R. Indeed, it was replaced by 15 new countries in 1991. This evolution can potentially make comparisons over time more difficult and will necessarily demand that end-users be particularly careful when doing this. For certain calculations (e.g. total students from Georgia in 20 years), this historical fact may remain hidden and the user would not see that students prior to 1991 are missing, leading potentially faulty analysis. This is a risk that the data producer must then be taking care of.

2. Risk analysis

Once identified, the risk is then analyzed according to its *probability of occurence* and the *severity of the consequences* if it occurs. Different techniques can be used to do such analysis; some use simulation techniques or probabilistic analysis, others look at relevant lessons learned during past projects, consult experts and specialists, etc. Independently to the technique used, those two criteria are more often than ever presented according to an ordinal scale of three or five levels. We thus suggest to classify the probability of occurence and the severity of the consequences as *low, moderate* or *high.* Like other risk-based approach, the risk evaluation step is not a simple task to do; as it demands to look in a certain way in the future, it often requires experience, judgment, and sometimes intuition. In addition to these, we also consider that an excellent knowledge and understanding of the datacube users' needs and skills, which also represent a legal duty of the datacube producer, are necessary to have the best risk analysis possible.

* http://www.omg.org/mda/

** http://www-304.ibm.com/jct03004c/businesscenter/smb/us/en/solutionsummary/xmlid/29811/nav_id/product

In our example, if there are several similar cases in the datacube, the data producer could conclude that the *probability* of making inconsistent temporal analysis is *moderate* because making comparisons in time is an important part of the client's needs, but only a smaller proportion of their analysis goes back more than 10 years. He could also predict that the *severity of the consequences* will be *low* because not taking into account the evolution of the countries could lead to false conclusions for older data but not for the most recent ones which are critical to decide about future actions.

3. Risk evaluation

The results coming from the risk analysis step are then combined together in order to determine the overall level of danger related to the risk. In the project management field, this task is often made by using an hierarchisation matrix, which is composed by two axes representing respectively the probability of occurence of the risk and the severity of the consequences if it occurs (cf. Table 1). The intersection of the two axes thus gives the overall level of danger related to the risk.

		Probability of occurence		
		Low	Moderate	High
Severity level	High	M	H	H
	Moderate	L	M	H
	Low	L	L	M

L = Low M = Moderate H = High

Table 1 General hierarchisation matrix (*adapted from* Kerzner, 2006)

Table 1 shows the matrix commonly used in the project management field. It could however be modified according to the risk tolerance level of the data producer. He could then adopt a more careful behavior by asking that the overall level of danger be *high* when the probability of occurence of the risk is *high*, no matter what its severity level is. According to the Table 1, we could say that the overall level of danger related to the risk in our example is *low*. However, even if we consider this risk as *low,* this does not mean that a civil court would conclude the same. For that reason, we must take advantage as often as possible of the lessons learned by others or into the organization itself.

4. Acceptable risk?

Once the overall level of danger related to the risk is known, a decision must me made to decide if the risk is acceptable or not. If *acceptable*, the producer of the datacube would decide to accept it without engaging special efforts to control it. This decision could possibly be taken when the risk is considered as *low*. However, this decision could involve some legal consequences. If the end-user is victim of damages because the risk has not been communicated to him, the data producer can then be legally declared liable to have chosen to ignore it. If the risk is considered as *unacceptable*, the producer must then choose a response mechanism in order to manage it (cf. 5. Risk response). In our example, the data producer could think that this specific risk is *unacceptable* even if the overall danger level is *low*.

Until now, the datacube producer has been the main player involved into the risk management process. He could however ask to the datacube client to participate in the process. This participation could be particularly helpful to determine if the risk is acceptable or not and eventually, to select the appropriate response mechanism.

5. Risk response

At this stage of the process, a decision must be made to determine how the data producer can cope with the unacceptable risk. Several mechanisms can be used, such as:

- *Avoidance*: the producer of the datacube can decide to reduce or eliminate the risk by refusing to provide the data to the identified category of end-users. This could be done for example when the data are considered as being too dangerous or not reliable for the users or their intended usages. This type of action is frequent in the field of statistics when data have a coefficient of variation above a certain threshold.

- *Transfer*: the producer of the datacube could also transfer the risk to another party (e.g. private firm of geomatics professional or an insurance company). If this other party is of equal competence compared to the data producers, the duties to *inform*, *warn*, and *advise* will then be transferred to this new party (Le Tourneau, 1995; Baudouin and Deslauriers, 1998; Le Tourneau and Cadiet, 2002).

- *Control*: the last possible mechanism for the producer could be to take preventive actions in order to reduce the risk. According to the ISO/IEC Guide 51 standard (cf. Figure 2), the risk reduction must first be accomplished at the conception level of the product. The designer could for example decide to modify the conceptual model of the datacube (or database) or implement some integrity constraints. In addition to the intrinsic prevention, the Guide 51 also suggests to produce information for security purposes, which is commonly named a "warning". Warnings could be communicated to datacube end-users by using different media. For example, warnings of general nature could be integrated in a user-manual while specific ones (i.e. context-sensitive warnings) could be prompted in the SOLAP application's interface at the moment end-users are facing a risky query or situation.

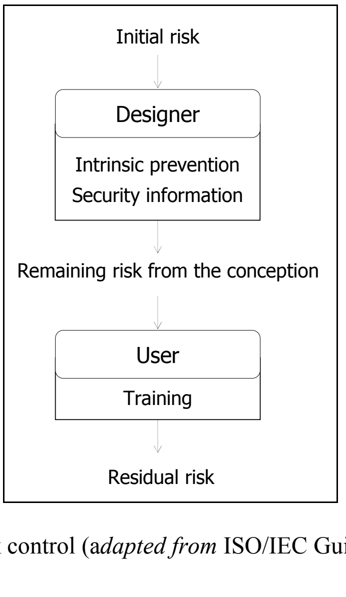

Figure 2 Risk control (*adapted from* ISO/IEC Guide 51, 1999)

We suggest to design the warnings according to the ISO 3864-2 (2004) standard, which regulates the design of product safety labels. This standard states that safety signs and safety labels should communicate the following four elements: 1) the overall *danger level* of the risk (cf. Section 3 on risk evaluation) which is caracterized by an alert word (i.e. DANGER, WARNING or CAUTION) 2) the *nature* of the risk which is often illustrated with a symbol, 3) the *consequence* of interacting with the risk and 4) how to *avoid* the risk. Figure 3 shows, for our previous example, the warning message that could be prompted in the user's interface when analyzing the data. Together with the content of the message, the warning includes symbols indicating the type of information displayed in addition to an hyperlink directing to others sources of information (e.g. detailed metatada, data dictionary). This allows end-users to improve their understanding of the message.

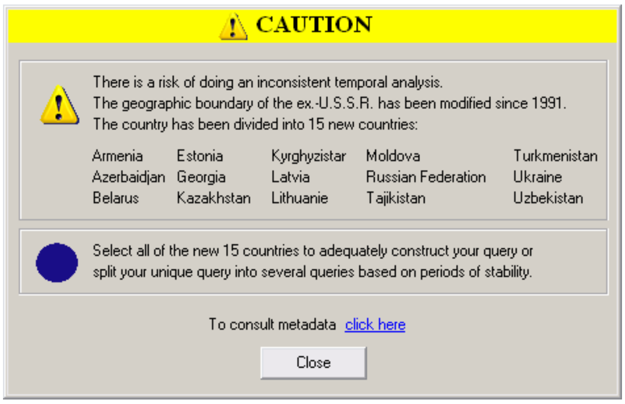

Figure 3 Example of warning message

Finally, the risks remaining after the conception must in a second time be treated at the end-user level, by providing him a training for example. This activity could even be mandatory when there is a high number of warnings, which could potentially indicate a lack of end-users' skills or a SOLAP application having a high degree of complexity and dangerousness.

6. Risk audit

The last step of the risk management procedure is to rigorously documenting all the steps discussed above as if preparing for an audit: the potential risks identified, their analysis, their evaluation, the response to these risks if needed, the justification, etc. The documentation of the risks should ideally be done regularly during the design and feeding of the datacube (i.e. at the end of some important steps of this process or at the end of each risk management step). When communicating a warning message to the end-users, the message itself, the involved elements and the time the message should be directed to the end-users (e.g. before the SOLAP query or once the results are displayed) should also be documented. To do so, we have designed two paper forms to be implemented into a UML-based CASE tool (Computer-Assited Software Engineering), in addition to a dictionary of terms and definitions that can help data producers to documente these processes. Such documentation is a very important step that may look expensive at first, but should not be neglected by the data producer from a legal standpoint because 1) this documentation can help them to prove they complied to their legal duties, and 2) it constitutes an important source of information that will help them to manage risks in futures datacube developments. More generally, it would also help the system designers to build more robust systems.

We presented in this section a preventive process aiming at reducing the risks for both users and producers of spatial datacubes. This reduction is made to obtain an acceptable risk using informed risk management decisions. However, there will always remain some residual risks at the end as it is impossible to eliminate them all (Bédard, 1988).

5. CONCLUSION

We presented an approach which aims at preventing and decreasing the risks of spatial datacubes misuse supported by SOLAP technology. This process mainly relies on a risk management approach adapted from the ISO/IEC Guide 51 standard and the field of project management. In this paper, we first showed the increased need to prevent the risks of spatial datacube misuse for SOLAP applications (i.e. in a context of easy and interactive access to multi-resolution cross-tabulated data) and the important role of the data producer in this process. We then did an overview of existing approaches in the fields of geomatics and BI to prevent data misuse. We finally presented our approach that specifically aims at helping spatial data producers to fulfill their legal duties towards end-users by identifying the potential risks of data misuse and by managing them in the best way possible.This approach takes place during the datacube design and feeding phases of the system lifecycle and completes other mechanisms to handle possible risks during the use of the datacube. This solution is being tested in a public health datacube project and will be the main subject of a future publication.

6. ACKNOWLEDGEMENTS

The authors wish to thank for its support the NSERC Industrial Research Chair on Geospatial Databases for Decision Support funded by the Natural Sciences and Engineering Research Council of Canada, Laval University, Hydro-Québec, Research and Development Defence Canada, Natural Resources Canada, Ministère des Transports du Québec, KHEOPS Technologies, Intélec Géomatique, Syntell, Holonics, and Groupe ALTA. They also acknowledge Canada's GEOIDE Network of Centers of Excellence, Health Canada, the Institut national de santé publique du Québec and Ouranos for their support in the testing project named "An Innovative Interactive Web Tool to Better Understand Climate-Related Health Vulnerabilities".

7. REFERENCES

Agumya, A., Hunter, G.J., 1999. A Risk-Based Approach to Assessing the 'Fitness for Use' of Spatial Data, *URISA Journal*, Vol. 11, No. 1, pp. 33-44.

Basili, V.R., Caldiera, G., Rombach, H.D., 1994. The Goal Question Metric Approach, *Encyclopedia of Software Engineering*, John Wiley & Sons, New York, pp. 528-532.

Ballou, D.P., Tayi, K.G., 1989. Methodology for allocating resources for data quality enhancement, *Communications of the ACM*, Vol. 32, No. 3, pp. 320-329.

Beard, K., 1989. Use error: the neglected error component. *Proceedings of AUTOCARTO 9*, Baltimore, MD, pp. 808–817.

Beard, K., 1997. Representations of Data Quality, *Geographic Information Research: Bridging the Atlantic* (M. Craglia and H. Couclelis, Eds), Taylor and Francis, pp. 280-294.

Beaudouin, J-L., Deslauriers, P., 1998. *La responsabilité civile*, Les Éditions Yvon Blais, Cowansville, 1684 p.

Bédard, Y., 1988. Uncertainties in Land Information Systems Databases, *Proceedings of the 8th International Symposium on Computer Cartography (AUTO-CARTO 8)*, Baltimore, USA, March 29 – April 3, pp. 175-184.

Bédard, Y., Devillers, R., Gervais, M., Jeansoulin, R., 2004. Towards Multidimensional User Manuals for Geospatial Datasets : Legal Issues and their Considerations into the Design of a Technological Solution, *Proceedings of the 3rd International Symposium on Spatial Data Quality (ISSDQ)*, Bruck an der Leitha, Austria, April 15-17th, Vol. 28b GeoInfo Series, pp. 183-195.

Courtot, H., 1998. *La gestion des risques dans les projets*, Éditions Économica, Paris, pp. 17–74.

Dassonville, L., Vauglin, F., Jakobsson, A., Luzet, C., 2002. Quality Management, Data Quality and Users, Metadata for Geographical Information, *Spatial Data Quality* (W. Shi, P.F. Fisher and M.F. Goodchild, Eds), Taylor & Francis, London, pp. 202-215.

Devillers, R., Bédard, Y. Jeansoulin, R., 2005. Multidimensional management of geospatial data quality information for its dynamic use within geographical information systems, *Photogrammetric Engineering & Remote Sensing (PE&RS)*, Vol. 71, No.2, pp. 205–215.

Devillers, R., Beard, K., 2006. Communication and use of spatial data quality information in GIS. *Fundamentals of Spatial Data Quality*. ISTE Publishing, London, pp. 237-253.

Devillers, R., Bédard, 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 Sciences (IJGIS)*, Vol. 21, No. 3, pp. 261-282.

Drecki, I., 2002. Visualisation of Uncertainty in Geographic Data, *Spatial Data Quality* (W. Shi, P.F. Fisher and M.F. Goodchild, Eds), Taylor & Francis, London, pp. 140-159.

Duckham, M., McCreadie, J., 2002. Error-aware GIS Development, *Spatial Data Quality* (W. Shi, P.F. Fisher and M.F. Goodchild, Eds), Taylor & Francis, London, pp. 63-75.

Eckerson, W., 2002. *Data Warehousing Special Report: Data quality and the bottom line*, Data Warehousing Institute, http://www.adtmag.com/article.aspx?id=6321&page= (accessed 06 Feb. 2006)

English, L.P., 1999. *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*, John Wiley & Sons, New York, 544 p.

Frank, A.U., 1998. Building a geospatial data framework – finding the best available data, *Data Quality in Geographic Information: from error to uncertainty,* Hermès, Paris, 192 p.

Gervais, M., 2004. *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, Université Laval, Québec et Université de Marne-La-Vallée, France, 344 p.

Gervais M., Y. Bédard, R. Jeansoulin et B. Cervelle, 2007. Obligations juridiques potentielles et modèle du producteur raisonnable, *Revue Internationale de Géomatique*, Vol. 17, No 1, pp. 33 à 62.

Harvey, F., 1998. Quality Needs More Than Standards, *Data Quality in Geographic Information – From Error to Uncertainty* (M.F. Goodchild and R. Jeansoulin, Eds), Editions Hermès, 192 p.

Horner, J., Song, I-Y., Chen, P.P., 2004. An Analysis of Additivity in OLAP Systems, *Proceedings of the 7th ACM International Workshop on Data Warehousing and OLAP,* Washington, USA, pp. 83-91.

Hunter, G.J., Goodchild, M.F., 1996. Communicating uncertainty in spatial databases, *Transactions in GIS*, Vol.1, No.1, pp. 13-24.

Hunter, G.J., 1999. Managing uncertainty in GIS, *Geographical Information Systems: Principles, Techniques, Management and Applications* (P.A. Longley, M.F. Goodchild, D.J. Maguire and D.W. Rhind, Eds), John Wiley & Sons, London, pp. 633-641.

Hunter G.J., and Reinke K.J. 2000. Adapting Spatial Databases to Reduce Information Misuse Through Illogical Operations, *Proceedings Spatial Accuracy Assessment, Land Information Uncertainty in Natural Resources Conference,* Amsterdam, July, pp. 313-319.

ISO/IEC Guide 51, 1999. Safety aspects – Guidelines for their inclusion in standards.

ISO 3864–2, 2004. Safety colours and safety signs – Part 2: Design principles for product safety labels.

Kerzner, H., 2006. *Project Management: a systems approach to planning, scheduling, and controlling (9th Edition)*, John Wiley & Sons, New Jersey, 1004 p.

Kimball, R., 2000. Indicators of Quality, http://www.intelligententerprise.com/000410/webhouse.jhtml (accessed 23 Nov. 2006)

Knightsbridge, 2006. *Top 10 trends in business intelligence for 2006*, Data Warehousing Institute, http://www.tdwi.org/Marketplace/WPContent.aspx?PID=174 (accessed 16 Jan. 2007)

Lassoued, Y., Jeansoulin, R., Boucelma, O., 2003. Médiateur de qualité dans les systèmes d'information géographique, *SETIT International conference (Sciences Electroniques, Technologies de l'Information et des Télécommunications),* Sousse, Tunisia, pp. CDindex-214.

Lefebvre, B., 1998. *La bonne foi dans la formation des contrats*, Les Éditions Yvon Blais, Cowansville, 304 p.

Lenz, H-J., Shoshani, A., 1997. Summarizability in OLAP and Statistical Databases, *Proceedings of the 9th International Conference on Scientific and Statistical Database Management,* Washington, USA, August 11-13, pp. 142 - 153.

Lenz, H-J., Thalheim, B., 2006. Warning: Cube May Mislead!, *4th International Multiconference on Computer Science and Information Technology (CSIT 06)*, Vol.2, Amman, Jordan, April 2006, pp. 7-16.

Le Tourneau, P., 1995. *La responsabilité civile professionnelle*, Éditions Economica, Paris, 105 p.

Le Tourneau, P., Cadiet, L., 2002. *Droit de la responsabilité et des contrats*, Éditions Dalloz, Paris, 1540 p.

Levesque, J., 2007. *Évaluation de la qualité des données géospatiales, Approche top-down et gestion de la métaqualité*, Mémoire (M.Sc.), Université Laval, Québec, 127 p.

McGranaghan, M., 1993. A cartographic View of Spatial Data Quality, *Cartographica*, Vol. 30, pp. 8-19.

Newell, A., 1990. *Unified theories of cognition*, Harvard University Press, Cambridge, MA, 549 p.

Onsrud, H.J., 1999. Liability in the use of GIS and geographical datasets, *Geographical Information Systems: Principles, Techniques, Management and Applications* (P.A. Longley, M.F. Goodchild, D.J. Maguire and D.W. Rhind, Eds), John Wiley & Sons, London, 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, No.3, pp. 742-781.

Ponniah, P., 2001. *Data Warehouses Fundamentals : A Comprehensive Guide for IT Professionnals*, John & Wiley Sons, New York, pp. 291-314.

Qiu, J., Hunter, G.J., 2002. A GIS with the Capacity for Managing Data Quality Information, *Spatial Data Quality* (W. Shi, P.F. Fisher and M.F. Goodchild, Eds), Taylor & Francis, London, pp. 230-250.

Rivest, S., Bédard, Y., Proulx, M-J., Nadeau, M., Hubert, F., Pastor, J., 2005. SOLAP technology : Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data, *ISPRS Journal of Photogrammetry & Remote Sensing*, Vol.60, No.1, pp. 17-33.

Timpf, S., Raubal, M., Kuhn, W., 1996. Experiences with Metadata, *Proceedings of Symposium on Spatial Data Handling, SDH'96, Advances in GIS Research II,* Delft, The Netherlands, p. 12B31-12B43.

Unwin, D., 1995. Geographical information systems and the problem of error and uncertainty, *Progress in Human Geography*, Vol. 19, No. 4, pp. 549-558.

Vassiliadis, P., 2000. *Data Warehouse Modeling and Quality Issues*, National Technical University of Athens, Athens, Greece, pp. 1.1-2.19.

Walford, N., 2002. *Geographical data, Characteristics and Sources*, John Wiley & Sons, England, 274 p.

Wang, R.Y., Storey, V.C., Firth, C.P., 1995. A Framework for Analysis of Data Quality Research, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 7, No.4, pp. 623–640.