Transactions in GIS

Transactions in GIS, 2010, 14(4): 387–400

Review Article

Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities

Rodolphe Devillers
Department of Geography
Memorial University of Newfoundland

Alfred Stein
Faculty of Geo-Information Science and Earth Observation (ITC)
University of Twente

Yvan Bédard
Département des sciences géomatiques
Université Laval

Nicholas Chrisman
GEOIDE Network and Département des sciences géomatiques
Université Laval

Peter Fisher
Department of Geography
University of Leicester

Wenzhong Shi
Department of Land Surveying and Geo-Informatics Hong Kong Polytechnic University

Abstract

This article reflects on the past 30 years of academic research in the field of spatial data quality and tries to identify the main achievements, failures, and opportunities for future research. Most of this reflection results from a panel discussion that took place during the *Sixth International Symposium on Spatial Data Quality* (ISSDQ) in July 2009.

1 Introduction

Early computer mapping exercises, such as those done at the Harvard Laboratory for Computer Graphics and Spatial Analysis in the late 1970s (Chrisman 2006, 2009) and at the Canadian Geographic Information System (CGIS), pointed out that imperfections

Address for correspondence: Rodolphe Devillers, Department of Geography, Memorial University of Newfoundland, 300 Prince Philip Drive, St. John's, Newfoundland, Canada A1B 3X9. E-mail: rdeville@mun.ca

doi: 10.1111/j.1467-9671.2010.01212.x

are inherent to spatial data and have a direct influence on the reliability of spatial analysis output. For instance, a polygon overlay operation performed with objects that have slightly different boundaries would lead to a large number of small errors (slivers and gaps) that would result in errors in the results obtained from these analyses (Goodchild 1978). Software that removes slivers represented an accomplishment technologically, but one that masks the variability inherent in data capture procedures. In fact, data quality issues have been encountered more generally since early data collection efforts took place, in both social and natural sciences. Traditionally, statistics have taken care of this, as can be seen, for example, in Cox and Snell (1981) who devoted their introductory pages to data quality. Statistics have been used by geodesists, photogrammetrists, and surveyors for decades to calculate, for instance, error ellipses and error models in situations of overabundance of point measurements. The spatial component has also come into view partly from geostatistics, as is explicit in the title of the volume by Chilès and Delfiner (1999). Methods for measuring map accuracy were also studied by a number of cartographers between the 1950s and 1980s (e.g. Glusic 1961, Volkov 1950). Most of these findings have been translated, integrated, and expanded by Derek Maling (see, for instance, Maling 1989). Works on these questions increased significantly with the arrival of geographical information systems (GIS) in the early 1980s and their capability to integrate spatial and non-spatial data. A broader view of spatial data quality emerged from academia during the 1980s and the early 1990s (see, for example, Beard 1989; Bédard 1986, 1987; Burrough 1992; Buttenfield 1993; Chrisman 1983, 1990; Fisher 1994; Goodchild 1988, 1995; MacEachren 1992; Mark and Csillag 1989; Morrison 1995; Robinson and Frank 1985). This broader view also found its way into practice around the same time, when national mapping agencies made significant efforts to document spatial data quality in their standards by encompassing several quality elements such as lineage, positional accuracy, attribute accuracy, logical consistency, and completeness (see, for example, CEN/TC-287 1994, 1995; Guptill and Morisson 1995; Moellering 1987).

Since its early development, spatial data quality (SDQ) has been a core subdiscipline of geographic information sciences (GISciences)/geomatics. It has drawn considerable attention from the academic community and government agencies and, more recently, from industry. SDQ has usually been present in major national/international research initiatives (e.g. AGILE, GEOIDE, NCGIA) and conferences in the field of GISciences/geomatics. Starting with issues related to the error in spatial data, the SDQ focus has diversified over the years, addressing the technological challenges of the time and following the overall degree of maturation of the discipline.

To reflect on the achievements of the past 30 years of research, to celebrate the tenth anniversary of the *International Symposium on Spatial Data Quality* (ISSDQ), and to try and envision future research challenges, a panel discussion took place during the Sixth ISSDQ held in July 2009 in St. John's (Canada). Four panelists (Yvan Bédard, Peter Fisher, Wenzhong Shi, and Alfred Stein) were asked to present the top five achievements, top five failures, and the top five opportunities in the past 30 years of research in SDQ. The discussion was facilitated by Nicholas Chrisman, who also contributed to the reflection earlier the same day during a keynote address (Chrisman 2009). The approach used to structure the discussion was adapted from SWOT analyses (SWOT: strengths, weaknesses, opportunities, and threats), a method typically used in strategic planning sessions. After the short presentations of the panelists, the audience was asked to comment and add to these thoughts.

This article integrates and develops the points raised during these discussions. Due to the nature of the exercise, this article cannot claim to be exhaustive or to represent widely accepted opinion. A number of similarities between the points identified by the panelists and some agreement with the audience, however, suggest that some issues that were raised are significant and could be of interest to the wider scientific community. This article is structured following the three aspects panelists were asked to cover. Section 2 discusses what are perceived to be the main achievements in the research on SDQ. Section 3 presents what are perceived to be failures. Section 4 suggests research opportunities that could turn into a proposed research agenda for the SDQ research community for the next 5–10 years.

2 Achievements

The past 30 years of research has resulted in a significant body of new knowledge that, in turn, has led to several significant achievements.

2.1 SDQ Community

The first significant achievement is that SDQ issues have grown outside isolated research interest and have become a well-recognized subdiscipline of GISciences/geomatics. "Accuracy of Spatial Databases" was in 1988 the first initiative of the US NCGIA research network. SDQ had a presence in their following initiatives and was present in major research funding schemes worldwide (e.g. EU Framework Program in Europe, GEOIDE network in Canada, CRCSI in Australia). SDQ has since become a significant area of focus in international spatial data standards, such as the International Standard Organisation (ISO) 19113 and 19114 standards, which focus on measurement of spatial data quality and its documentation. As such, it has helped to influence government and industry practices. A community of researchers has been built that meets in conferences focusing on spatial data quality issues, i.e. *International Symposium on Spatial Data Quality* and *Accuracy* conference series that both have been running for more than 10 years. There is also a good SDQ presence in general GISciences conferences. These meetings provide a view on an essential, and somewhat underestimated, problem field, with an open mind towards methodology and methodology development. In addition, data quality is a transversal theme in GISciences/geomatics and is hence found under a number of topics where understanding data quality is important (e.g. data quality issues in environment modeling or land-use planning). This research community has created a synergy, kept the topic alive, and produced regularly published papers, books (e.g. Ariza López 2002; Ariza López et al. 2004; Congalton and Green 1998; Devillers and Goodchild 2009; Devillers and Jeansoulin 2005, 2006; Foody and Atkinson 2003; Goodchild and Gopal 1989; Goodchild and Jeansoulin 1999; Shi 2009; Shi et al. 2002; Stein et al. 2008; Zhang and Goodchild 2002), and some special issues in journals. In this way, it has encouraged a research focus on spatial data quality issues. This also increased, in the past decade, awareness from industry that data quality issues continue to be a major challenge.

2.2 Research Discoveries

The second significant achievement encompasses the scientific contributions that proposed several methods to evaluate specific data quality elements of vector, raster, and

Digital Elevation Model (DEM) data. This includes approaches in GIS for measuring positional error for points, lines, and polygons, modeling attribute and temporal uncertainties, and understanding how uncertainty propagates, but also a number of works in remote sensing, looking at various quality aspects of images, from the collection to the processing and analysis of the images. Some of these approaches extended existing methods used in cartography, geodesy, surveying, hydrography, and photogrammetry. Some methods are now used by national mapping agencies to characterize the quality of their datasets. A related achievement lies in the international metadata standards (e.g. FGDC and ISO) that allow communicating some of the quality information to the end-users of spatial data. Although communicating this data quality information from data producers to data users is generally positive, it is also perceived as a partial failure, as will be discussed in Section 3. Several initiatives considered communicating data quality/uncertainty in other ways (e.g. by means of graphic visualization, by warnings and restrictions of responsibilities in licenses), but these approaches did not reach the community of practitioners or are of limited usefulness in their present state. The panelists identified a number of specific significant research contributions, such as the work of Shi related to error modeling of geometric features (e.g. Shi 1998). All the papers discussed will not be mentioned here as the importance of specific works can be influenced by individual research interests.

2.3 Semantics

The third significant achievement concerns recognition of the importance of semantic/ontological issues in spatial data quality. This emerged in the past decade in the context of an increased sharing of spatial data, often supported by poor, limited, or incomplete metadata. Understanding complex issues and responding to global challenges (e.g. climate change, environmental conservation, sustainable resources management) requires combining datasets collected by different organizations that used different standards and different ways to describe the world's features. But how can global changes in land cover, for instance, be understood when each country has a different classification scheme and often changes its own scheme through time (e.g. Ahlqvist 2005; Comber et al. 2005; Fritz and See 2005)? Today's trends towards spatial-centric web services, interoperability, and data mashups are facing this semantics challenge which is central to the Web 2.0. Even the semantics involved in an object's geometry (e.g. houses are represented by "polygons") may change from one data source to the other when different data acquisition specifications are used (e.g. digitizing more or fewer details for buildings, depending on map scale and map purpose). The quality of semantics has impacts on the very existence of objects, on their categorization, on their attributes, on qualitative values, and on the temporal and spatial properties of these objects. Semantics cannot be separated from spatial data quality analysis.

2.4 Fitness for Use

The fourth significant achievement is the emergence of a cluster of research projects related to the evaluation of "fitness for use" and "external quality" (as opposed to "internal quality"). This cluster focuses on how some datasets or services fit users' needs. Although this concept of "fitness for use" is not new (Chrisman 1990), there has been a recent increase in research looking at data quality from this perspective. Timpf et al.

(1996) is an example of an early study that illustrated the lack of link that exists between the depth at which metadata are documented and the actual help metadata can provide when assessing the fitness for use. This new way of approaching the problem completely shifted the perception of some of the SDQ research from thinking "How can I measure the quality of my data and let the user know?" towards "What does the user need in terms of quality information and how can I provide what they need to avoid data misuses?" This approach has created a more direct link between the datasets and their usage, between the concerns of data producers and the expectations of mass users of spatial data. Such evolution indicates that although the most technically perfect dataset may be created, it could be completely useless, or even dangerous, if it does not fit the users' need. This evolution of the research in SDQ, to explicitly encompass the notion of fitness for use, was required to enter the mainstream of consumers' risk management within the rules of today's society. It illustrates that the subfield of SDQ is becoming more mature.

3 Failures

Although past research in spatial data quality has achieved many successes, it also encompassed a number of partial failures. These represent, in some cases, limited success of research but in other cases simply missed opportunities.

3.1 Scientific Footprint: Measuring our Research Impact

One metric that any research field and individual researchers should regularly assess is their "scientific footprint." This is the impact the researcher/field has on the non-academic community. In other words, it is asking the question "Would the world be any different without my research?" A partial failure of SDQ research is the generally poor connection that exists between academic research and the day-to-day use of spatial data by different types of users. If a number of research projects led to methods and tools used by governments and industry, a large body of scientific knowledge is still only in the hands of researchers and embedded in scientific publications (Goodchild 2008a).

Although imperfections are inherent in spatial data and have an influence on decisions based on these data, users of geospatial technologies often have the impression that data are perfect or act accordingly. Outside of a few exceptions, the main commercial GIS/remote sensing software vendors did not translate SDQ research findings into the functioning of software. Spatial data producers and distributors also have not introduced sufficient good practices to support fitness for use. A potential reason for the limited impact of SDQ research is the often fairly high level of complexity in terms of statistics and modeling, which makes it difficult for typical spatial data users to understand the basic SDQ concepts. Another possible reason is that such solutions are only a small part of the overall global picture of fitness for use, the underlying hidden part of it, and are not sufficient for preventing spatial data misuses. This points to a concern that much SDQ research is about solutions looking for problems, such as statistical approaches applied to assess SDQ components, and that research should better aim to first comprehend the nature of the problem before it provides the best possible solution. Despite an undeniable success of statistical approaches that have become an integral part of the work of certain specialists (e.g. geodesists, photogrammetrists, surveyors, hydrographers), the remaining GISciences/geomatics community has yet to reach this level of widespread use of statistics, and may never reach it. And without even considering statistical approaches, it can be argued that spatial data users seem to lack the common sense they have in other contexts. For instance, while it would be considered strange to have the weather channel reporting temperature to eight decimal places, a spatial data user will rarely question the number of significant digits reported by a GIS. Nevertheless, in all cases, the geospatial users' community would benefit from an increased awareness of issues related to SDQ. To address this issue, several works related to the visualization of spatial data uncertainty and quality have been published, in addition to the more recent developments in "error-aware" or "quality-aware" GIS (e.g. Duckham and McCreadie 1999; Zargar and Devillers 2009) and the introduction of "context-sensitive warnings" (e.g. Beard 1989; Levesque et al. 2007). Again, these works have remained academic exercises and have not reached the end-users of spatial data. Academic researchers even developed tools to help users tackling uncertainty issues (e.g. Data Uncertainty Engine – DUE – of Brown and Heuvelink 2007), but, in spite of their recent release, these tools do not seem to be used much outside of the research community, and may not be widely used until users see more clearly a rationale to do so.

3.2 Terminology

Another concern is the lack of a commonly accepted terminology, or ontology, in SDQ. Ask a national mapping agency, a researcher, and a lay user how they define "data quality" or "uncertainty," and there would likely be three different answers. Use of terms also varies largely within a community and it can be confusing for people reading academic papers to identify different concepts named the same, or similar concepts labeled using different names. Definitions of other related concepts, such as uncertainty, error, accuracy, vagueness, etc., can find different interpretations depending on the people using the terms or the scientific community they come from. Attempts have been made to clarify these definitions and their relationships as taxonomies, but these taxonomies, which are often compatible, are not yet integrated (e.g. Bédard 1986, 1987; Drecki 2007; Fisher 1999; Fisher et al. 2006; Leyk et al. 2005; Robinson and Frank 1985). Consequently, to facilitate the advance of SDQ into society, there is a need to build a generally agreed-upon ontology. Hopefully, standardization bodies such as ISO/TC-211 and Open Geospatial Consortium (OGC), which have shown increased interest in this topic lately, will help reach a level of agreement. Otherwise, it may be speculated whether we are any further with distinguishing uncertainty, accuracy, or data quality than 20 or 30 years ago.

It can also be argued that this confusion in terminology led to the division of the scientific community into two main groups: those doing research in "spatial data quality" and those doing research in "uncertainty." The concept of spatial data quality is often associated with the question of standards. However, the ISSDQ symposia, where the SDQ community mainly meets, has a much broader research scope. Uncertainty, on the other hand, suggests research around positional accuracy issues. However, the *Accuracy* symposia, where the uncertainty community mainly meets, also has a much broader research scope. Both communities have a lot to share but they remain too disjoint, and work done by one group remains too often unknown, and hence unused, by the other group. Fisher (2003) described these two communities as "ships passing in the night" and argues that data quality and uncertainty should be mirrors of each other.

3.3 Research Focus and Breadth

Research in the SDQ domain has been able to answer several relevant scientific and technical questions. A major problem, however, is that the solutions remain rather descriptive. A good example is the well-known list of 5–7 data quality elements (see, for instance, Chrisman 1983 for an early description of these), which includes, depending on classifications, lineage, logical consistency, completeness, positional accuracy, attribute accuracy, semantic accuracy, and temporal accuracy. Such a list was useful when first presented and may remain useful when it comes to describe data quality issues, but may now be surpassed when new questions emerge and the additional value of a spatial data quality analysis is requested. How, for example, should the quality of various layers of data combined be best summarized? How could quality in data mashups or resulting from the integration of volunteered geographic information (VGI) provided by thousands of web users be meaningfully summarized? To cope with this problem, Goodchild (2008a, b) described existing metadata that only describe one dataset as being "unary", and wondered if "binary" metadata could be documented, which would illustrate the ability of two datasets to interoperate. Instead of always considering uncertainty as a problem, uncertainty is increasingly considered as being an asset. But then again, one should seriously name, quantify, and analyze this uncertainty and use it for its benefit. Bédard (1986) brought attention to "meta-uncertainty" (uncertainty about uncertainty) and "uncertainty absorption" to describe the financial risks associated with providing/using spatial data. This concept is increasingly relevant with new questions raised by the arrival of spatial data mashups and VGI.

Finally, the link with society should be present throughout; whereas society is not really interested in understanding spatial data quality, it is interested in solving problems – in our case spatial problems that can be supported by the use of spatial data. If decision makers are not interested in data quality issues *per se*, they are, however, often interested in knowing the reliability of the data they are using to assess the reliability of the decisions to be based on these data. A number of decision makers are, however, uncomfortable when they have to explain these uncertainties, as it can open the door to people questioning their decisions. Past research in SDQ did not succeed in presenting enough convincing cases of the social, economic, and safety impacts that low-quality spatial data can generate. This could help to explain the relevance of exploring SDQ issues. While discussing with academic GIS colleagues in a recent international meeting about the impact that past decisions based on poor-quality data had (mentioning some accidents that led to people's death), a GISciences professor gave a sarcastic comment that we should not focus on "killer GIS" but should instead look at all the great aspects that the technology brings. If part of this is a fair point, it clearly shows that, even in the academic field, a number of people don't want to think about some of the consequences GIScience has on society and individuals, as long as the advantages exceed the disadvantages.

As a related issue, the general public sometimes receives exciting news from the geospatial community, such as the recent developments in virtual globes, and sensor networks that collect data about our environment to help better understand current global challenges. But could there ever be any exciting development in spatial data quality that the public, or even others in the geospatial community, would be interested in? It isn't that obvious. Research topics and solutions are often abstract, whereas concrete solutions are what matters for the end-users.

3.4 Metadata and Assessing Fitness for Use

Related to research impact is the question of metadata or, more generally, of any form of communication made of data quality information. The publication of metadata is the main approach that has been used by data producers in the past two decades to communicate spatial data quality to professional and lay users. Generating metadata is a time-consuming task and has proven to have a fairly limited impact on a user's ability to understand the possible uses of data. Successful uses of metadata are mainly related to the use of core/discovery metadata to search for spatial datasets using keywords (spatial and temporal extents, theme, etc.). Users are expected to understand the characteristics of a given dataset and the extent of its potential use from metadata. However, there is still a gap between what the quality assessment mapping experts can produce and the information users can understand and use.

As a timely example, one of the first author's colleagues, working in physical geography, approached him during the writing of this article and asked what the following statement from the metadata of a topographic dataset meant, "The altimetric accuracy of the source material is provided when available. It is expressed as the Linear Map Accuracy Standard (LMAS), obtained according to the equation below: Linear Map Accuracy Standard: LMAS = 1.6449 σ_z; σ_z = standard deviation of the elevation." The colleague wanted to compare changes in elevation of a glacier in the Canadian Arctic that is melting, but could not clearly understand the vertical accuracy of the contour lines. A first surprise was to see that some users do try to access and consult metadata in attempting to avoid potential misuse of the data. If such a metadata statement can be understood by some GIS experts, this request required contacting the data producer and accessing archived technical specifications. This example shows again that data quality issues are complex, even for the data producer or for an expert in data quality. It also shows how difficult it can be to communicate the information to the end-user as, even if the metadata provide a number (e.g. "Vertical accuracy: 10 m"), it can only be assumed that it was based on some kind of sampling, involved a comparison with some ground truth data that were themselves inaccurate, and that 90% of the data fall within the 10-m threshold, but 10% do not. . . . A number of elements most data users are likely to miss. And this is only for one data source; imagine if the user wants to answer their question by combining a number of different datasets, or portions of several datasets showing some spatial heterogeneity in data quality. Such diversity of sources is often found within a dataset for different feature classes (e.g. roads *vs.* vegetation), different regions, different dates of data acquisition, but also within feature classes and individual features for each attribute and geometry. The complexity related to SDQ makes it difficult to document within a simple metadata statement.

For this reason, the balance between what should or should not be in the metadata is always in question and may vary according to users' profiles and usages. Should we document the minimum, to have the users read it and the producers being able to document it? All data quality statements are, for instance, optional in the ISO 19115 metadata standard. Or should we go into more detail and describe a number of aspects to cover more ground that may still not be enough? Or should the content and format be dictated by consumer protection regulations, professional ethics, or the type of license or contract, as is the case for a number of other commercial products? For this reason, some research teams started recently to explore the potential of having experts in data quality acting as a link between the data and some types of users. Looking at other fields, professional expertise is often used in such complex scenarios and could be relevant to at least part of the geospatial user community. Another issue that has been increasingly raised is the legal and social duty data producers/distributors have, to inform users of spatial data quality in a meaningful and understandable way (like any other product or service on the market). Several court decisions suggest such a direction. This is another interesting development in SDQ research that may end up putting some pressure on data producers and geographic information professionals to pay more attention to data quality issues. Similarly to the climate change debate with the questions of carbon credits, it is likely that spatial data quality will only become a critical issue for the geospatial community if there is a significant economic or legal incentive for industry and government to go in this direction. The cost related to data quality issues is already raised as a major concern by industry. For instance, the Gartner Group estimated, in 2008, the global market related to data quality to be around US$300 million. This is expected to reach US$677 million in 2011 and then have an estimated increase of 18% per year. Metadata are still a topic of interest in the SDQ research community and a number of new challenges may be tackled to try to improve this link between data producers and users (Goodchild 2008a).

4 Opportunities

Like any research field, our community has been making important discoveries and missing some important points. This process of looking critically backward does not aim to criticize the research that has been done, as any step forward is good, but looks at being constructive when looking forward to challenges the research community could explore. The panelists identified a number of opportunities for the SDQ community that are a consequence of past failures, but also result from a new context within which society interacts with spatial data.

4.1 Raising Users' Awareness

One aspect that should be further explored is to raise the awareness users have of data quality issues. While Box (1976) was arguing that "all models are wrong, but some are useful," it can be argued that data users should be aware that "all spatial data are wrong, but some are useful" (outside of some exceptions where the data becomes legally the reality). Raising awareness can be achieved in various ways. Some of it will result naturally from the increasing penetration of spatial data into people's day-to-day lives. A number of lay users are increasingly aware that spatial data are not perfect simply by looking at places they know well on Google Earth or Google Maps, for instance (mismatches between vector roads and aerial photos, missing roads, etc.). These people will naturally *absorb* the uncertainty (i.e. accept the risk) when using the data, mostly when the decision to be made is not critical. However, some scientists have started to be concerned about data displayed in Google Earth when being used to support more complex and important decisions (e.g. Sheppard and Cizek 2009). Some fairly easy changes to GIS/RS/GPS systems could raise awareness about spatial uncertainty issues. Most GPS users are aware that GPS do not provide a perfect position and most GPS systems will already give you an indication of the spatial accuracy of the position (e.g. a number on some systems, a blue circle of varying size on the iPhone). Similar simple approaches could be used in other mapping/RS applications to cover simple cases.

It would be interesting, as mentioned earlier, to draw the attention of users to the potential consequences of using datasets that do not fit their needs in terms of quality. This could potentially be based on series of examples of past decisions that were based on poor data. Misinterpretation of spatial data quality issues have already led to a large waste of money, a number of suboptimal policies, accidents, and deaths. Raising users' awareness could increase the interests in data quality issues. A related point that has received some attention in the past decade (e.g. de Bruin and Hunter 2003, De Bruin et al. 2001, Frank 2008) is to understand quantitatively the impact of data quality on some types of decisions to be made, to eventually understand how good the data have to be for a given task. Such approaches are often based on some type of cost–benefit analyses and assume that data do not have to be perfect to be used to answer some questions, but have to be "good enough" for the task at hand. Other works could look at ways to inform data users of imperfections with the data. A number of studies have been done in the past concerning ways to visualize, and more generally communicate, quality/uncertainty information. However, none of them clearly moved from the academic sector to the user and professional community. We should maybe explore other ways to approach the problem and look more at what users need instead of focusing on what available information could be communicated.

4.2 Spatial Data: A Changing Landscape

A first aspect of the changing landscape under which spatial data operate is the overall democratization of spatial data that occurred in the past 10 years, resulting, among other things, from the adoption and evolution of the Internet. Ten years ago was the early age of web mapping, when being able to use a web browser to retrieve dynamically a raster map from another computer linked to the Internet was very impressive. Since then, improvements in the software, in computing power, and the web bandwidth have resulted in completely new ways to think about spatial data. The standard data user is not a GIS expert working on a powerful and expensive workstation and software for professional usage anymore, but the lay person with no expertise in GISciences using web mapping applications, such as Google Earth or Google Maps, for all kinds of unpredictable tasks. This has opened the door to a number of new research challenges in SDQ. For instance, in the new Web 2.0 era, where anyone can contribute to the content of a web site, how do you assess the quality of data produced by a web user? Should mapping be restricted to professional cartographers or is there something good in opening the door to the public to contribute to this? How do you assess the quality of a single dataset potentially produced and updated by thousands of users? Some organizations like Open Street Map appear to have developed reliable procedures to address these issues outside of the traditional production chain. At the same time, governments are pushing national and international spatial data infrastructures that aim, among other things, to give easier access to spatial data. This changing landscape is also increasingly raising a number of legal questions. Who is liable if someone uses spatial data of poor quality and causes some harm? Is it the producer of the data, who did not do a perfect job or did not document the quality of the data sufficiently? Is it the distributor of the data? Is it the user, who took his own risk by using the available data? Or is the responsibility shared? The case becomes even more complex when dealing with data mashups and VGI.

4.3 What is so Special about Spatial?

Another interesting question can be asked: "After having a number of people tell us that spatial data were special (which is to some extent true), did we end up thinking that it wasn't worth looking at what other non-spatial people are doing?" Spatial data quality is, first of all, about data quality in general and there is a very important research community looking at (non-spatial) data quality issues. They have their conferences (e.g. *Information and Data Quality Conference*, *Data Management and Information Quality Conference*, *International Conference on Information Quality*), journals (e.g. *Journal of Data and Information Quality*, *International Journal of Information Quality*), books (a search on Amazon provided 74 results when using "spatial data quality" and 11,718 when using "data quality"), etc. Obviously, spatial data have a number of specific aspects, but there are probably a lot of good things to get from research done in data quality generally, and few SDQ researchers have looked into this. Research methods developed in non-spatial fields for handling data quality could have a lot to contribute to the research in SDQ. Questions about providing the best possible information is also an issue discussed in a number of other disciplines, such as journalism, marketing, and human communication. Our community should be careful not to end up being a microcosm that operates on its own, with a limited and consistent list of researchers and for the satisfaction of its own members. We should aim at integrating new groups of people into our community, who could help us look at the issues differently. This could include, for instance, lawyers, psychologists, human–machine interface specialists, spatial data infrastructure actors, insurance companies, risk managers, and cognition scientists.

5 Conclusions

This article presented the outcome of a fairly uncommon exercise in academia, which was to ask a research community to reflect on its achievements and failures, in order to shape a research agenda through the use of a strategic planning approach. Academics are perhaps not the best placed to adopt the formula of strengths, weaknesses, opportunities, and threats. It is too much in our culture to emphasize what is going wrong, not necessarily to build a future. It is also much easier to describe and criticize the past, than it is to predict the future. Yet, the research sector plays a key role in exploring weaknesses, turning them into strengths; and also in turning threats into opportunities. The past decades have seen dramatic improvements, particularly in mobilizing a community engaged on issues of spatial data quality. This strength creates opportunities. The combination of the growing penetration of geospatial technologies in day-to-day life, of the omnipresence of uncertainty in spatial data, and of an increasingly risk averse society is likely to put spatial data quality in an increasingly important place on the scientific agenda.

Acknowledgments

Thanks are due to Mahmoud Delavar and two anonymous reviewers for their suggestions and critical comments that helped to improve this article. The manuscript also benefited from the comments and careful review of Randal Greene, Aaron Hase, and René Enguehard from the Memorial University of Newfoundland Geography department. We are also grateful to the Natural Sciences and Engineering Research Council (NSERC) of Canada and the GEOIDE Network of Centres of Excellence for their financial support to the research programs on spatial data quality of R. Devillers and Y. Bédard and to the Social Sciences and Humanities Research Council (SSHRC) of Canada for their financial support to the research of N. Chrisman.

References

Ahlqvist O 2005 Using uncertain conceptual spaces to translate between land cover categories. *International Journal of Geographical Information Sciences* 19: 831–57

Ariza López F J 2002 *Calidad en la producción cartográfica*. Madrid, Ra-Ma

Ariza López F J, García Balboa J L, and Pulido R A 2004 *Casos prácticos de calidad en la producción cartográfica*. Jaén, Universidad de Jaén

Beard K 1989 Use error: The neglected error component. In *Proceedings of AUTO-CARTO 9*, Baltimore, USA: 808–17

Bédard Y 1986 A study of data using a communication based conceptual framework of land information systems. *The Canadian Surveyor* 40: 449–60

Bédard Y 1987 Uncertainties in land information systems databases. In Chrisman N R (ed) *Proceedings of AUTO-CARTO 8*. Baltimore, MD, American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping: 175–84

Box G E P 1976 Science and statistics. *Journal of the American Statistical Association* 71: 791–99

Brown J D and Heuvelink G B M 2007 The Data Uncertainty Engine (DUE): A software tool for assessing and simulating uncertain environmental variables. *Computers and Geosciences* 33: 172–90

Burrough P A 1992 Development of intelligent geographical information systems. *International Journal of Geographical Information Systems* 6: 1–11

Buttenfield B P 1993 Representing data quality. *Cartographica* 30: 1–7

CEN/TC-287 1994 WG 2. *Data description: Quality*. Brussels, European Committee for Standardization Working Paper No. 15

CEN/TC-287 1995 PT05. *Draft Quality Model for Geographic Information*. Brussels, European Committee for Standardization Working Paper D3

Chilès J P and Delfiner P 1999 *Geostatistics: Modeling Spatial Uncertainty*. New York, John Wiley & Sons

Chrisman N R 1983 The role of quality information in the long-term functioning of a geographic information system. In Douglas D H (ed) *Proceedings of AUTO-CARTO 6*. Ottawa, ON, Canadian Institute of Surveying and Canadian Cartographic Association: 303–21

Chrisman N R 1990 The error component in spatial data. In Maguire D J, Goodchild M F, and Rhind D W (eds) *Geographic Information Systems: Principles and Applications*. London, Wiley: 165–74

Chrisman N R 2006 *Charting the Unknown – How Computer Mapping at Harvard Became GIS*. Redlands, CA, ESRI Press

Chrisman N R 2009 A difference that makes a difference – Reflections on 30+ years in the field of Spatial Data Quality. Unpublished keynote presentation at the Sixth International Symposium on Spatial Data Quality (ISSDQ'09), St. John's, Canada (available at http://www.mun.ca/issdq2009/ISSDQ2009_Chrisman_Keynote.pdf)

Comber A, Fisher P F, and Wadsworth R 2005 What is land cover? *Environment and Planning B* 32: 199–209

Congalton R G and Green K 1998 *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, FL, Lewis Publishers

Cox D R and Snell E J 1981 *Applied Statistics: Principles and Examples*. London, Chapman and Hall

De Bruin S and Hunter G 2003 Making the trade-off between decision quality and the information cost. *Photogrammetric Engineering and Remote Sensing* 69: 91–8

De Bruin S, Bregt A, and Van de Ven M 2001 Assessing fitness for use: The expected value of spatial data sets. *International Journal of Geographical Information Sciences* 15: 457–71

Devillers R and Goodchild H (eds) 2009 *Spatial Data Quality: From Process to Decisions*. Boca Raton, FL, CRC Press

Devillers R and Jeansoulin R (eds) 2005 *Qualité de l'Information Géographique*. Paris, Hermès

Devillers R and Jeansoulin R (eds) 2006 *Fundamentals of Spatial Data Quality*. London, ISTE

Drecki I 2007 Geographical information uncertainty: The concept and representational challenges. In *Proceedings from the Twenty-third International Cartographic Conference*, Moscow, Russia: CD-ROM

Duckham M and McCreadie J E 1999 Error-aware GIS development. In Shi W, Goodchild M F, and Fisher P F (eds) *Spatial Data Quality*. New York, Taylor and Francis: 62–75

Fisher P F 1994 Animation and sound for the visualization of uncertain spatial information. In Hearnshaw H M and Unwin D J (eds) *Visualization in Geographic Information Systems*. London, John Wiley and Sons: 181–85

Fisher P F 1999 Models of uncertainty in spatial data. In Longley P, Goodchild M F, Maguire D, and Rhind D (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications*, Vol. 1. New York, John Wiley and Sons: 191–205

Fisher P F 2003 Data quality and uncertainty: Ships passing in the night! In Shi W, Goodchild M F, and Fisher P F (eds) *Proceedings of the Second International Symposium on Spatial Data Quality*. Hong Kong, Hong Kong Polytechnic University: 17–22

Fisher P F, Comber A, and Wadsworth R 2006 Approaches to uncertainty in spatial data. In Devillers R and Jeansoulin R (eds) *Fundamentals of Spatial Data Quality*. London, ISTE: 43–59

Foody G M and Atkinson P M (eds) 2003 *Uncertainty in Remote Sensing and GIS*. New York, John Wiley and Sons

Frank A U 2008 Analysis of dependence of decision quality on data quality. *Journal of Geographical Systems* 10: 71–88

Fritz S and See L 2005 Comparison of land cover maps using fuzzy agreement. *International Journal of Geographical Information Sciences* 19: 787–807

Glusic A M 1961 *The Positional Accuracy of Maps*. Washington, DC, US Army Corps of Engineers Army Map Service Technical Report No. 35

Goodchild M F 1978 Statistical aspects of the polygon overlay problem. In Dutton G (ed) *Harvard Papers on Geographic Information Systems*, Vol. 6. Reading, Addison Wesley

Goodchild M F 1988 Stepping over the line: Technological constraints and the new cartography. *The American Cartographer* 15: 311–20

Goodchild M F 1995 Sharing imperfect data. In Onsrud H J and Rushton G (eds) *Sharing Geographic Information*. New Brunswick, NJ, Rutgers University Press: 413–25

Goodchild M F 2008a Epilogue: Putting research into practice. In Stein A, Shi W, and Bijker W (eds) *Quality Aspects in Spatial Data Mining*. Boca Raton, FL, CRC Press: 345–56

Goodchild M F 2008b Spatial accuracy 2.0. In Zhang J and Goodchild M (eds) *Spatial Uncertainty: Proceedings of the Eighth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Volume 1. Liverpool, World Academic Union: 1–7

Goodchild M F and Gopal S (eds) 1989 *Accuracy of Spatial Databases*. London, Taylor and Francis

Goodchild M F and Jeansoulin R (eds) 1999 *Data Quality in Geographic Information: From Error to Uncertainty*. Hermès, Paris

Guptill S C and Morisson J M (eds) 1995 *Elements of Spatial Data Quality*. Amsterdam, Elsevier

Levesque M A, Bédard Y, Gervais M, and Devillers R 2007 Towards managing the risks of data misuse for spatial datacubes. In Stein A, Shi W, and Bijker W (eds) *Fifth International Symposium on Spatial Data Quality (ISSDQ)*. Enschede, The Netherlands, International Institute for Geo-Information Science and Earth Observation: CD-ROM

Leyk S, Boesch R, and Weibel R 2005 A conceptual framework for uncertainty investigation in map-based land cover change modelling. *Transactions in GIS* 9: 291–322

MacEachren A M 1992 Visualizing uncertain information. *Cartographic Perspectives* 13: 10–9

Maling D 1989 *Measurements from Maps: Principles and Methods of Cartometry*. Oxford, Pergamon Press

Mark D M and Csillag F 1989 The nature of boundaries in area-class maps. *Cartographica* 26: 65–78

Moellering H (ed) 1987 *A Draft Proposed Standard for Digital Cartographic Data*. Report 8. Columbus, OH, National Committee for Digital Cartographic Data Standards (NCDCDS)

Morrison J L 1995 Spatial data quality. In Guptill S C and Morrison J L (eds) *Elements of Spatial Data Quality*. New York, Elsevier: 1–12

Robinson V B and Frank A 1985 About different kinds of uncertainty in collections of spatial data. In *Proceedings of AUTO-CART0 7*, Washington, D.C.: 440–49

Sheppard S R J and Cizek P 2009 The ethics of Google Earth: Crossing thresholds from spatial data to landscape visualization. *Journal of Environmental Management* 90: 2102–17

Shi W 1998 A generic statistical approach for modeling error of geometric features in GIS. *International Journal of Geographical Information Sciences* 12: 131–43

Shi W 2009 *Principles of Modeling Uncertainties in Spatial Data and Spatial Analyses*. Boca Raton, FL, CRC Press

Shi W, Fisher P F, and Goodchild M F (eds) 2002 *Spatial Data Quality*. New York, Taylor and Francis

Stein A, Shi W, and Bijker W (eds) 2008 *Quality Aspects in Spatial Data Mining*. Boca Raton, FL, CRC Press

Timpf S, Raubal M, and Kuhn W 1996 Experiences with metadata. In *Proceedings of the Seventh International Symposium on Spatial Data Handling*, Delft, The Netherlands: 12B31–B43

Volkov N M 1950 *O Tochnosti Kart*. Trudy TsNIIGAiK, Vup 52: 67–91

Zargar A and Devillers R 2009 An operation-based communication of spatial data quality. In *Proceedings of GEOWS'09 Conference*, Cancun, Mexico: 140–45

Zhang J X and Goodchild M F 2002 *Uncertainty in Geographical Information*. London, Taylor and Francis