New Horizons for Spatial Data Quality Research

Suzie Larrivée, Yvan Bédard, Marc Gervais & Tania Roy

Centre for Research in Geomatics, Dept of Geomatics Sciences, Laval University, Quebec City, Canada
Suzie.Larrivee@scg.ulaval.ca, Yvan.Bedard@scg.ulaval.ca, Marc.Gervais@scg.ulaval.ca, Tania.Roy.2@ulaval.ca

Abstract

Producing and using spatial data as well as web services has reached the level of mass market, leading to new research challenges. A number of concepts are appearing regularly in our community: certification, accreditation, inspection, audit, warranty, quality assessment, quality control, quality assurance which are natural ways for a society to organize itself when a mass of citizens is facing increasing risks of misusing given products or services. Such concepts help defining who is responsible or liable for these risks, who absorbs the remaining uncertainty inherent to the use of data once technical means have contributed to reduce this risk. This paper provides an overview of recent issues surrounding formal data quality endorsments in a scenario where spatial data have become a commodity.

Keywords: Spatial Data quality, Audit, Certification, Guarantee, Accreditation

1 Introduction

Research about spatial data quality has taken place for over 30 years (Devillers *et al.*, 2010). However, producing and using spatial data as well as using and offering spatial services over the web has now reached the level of mass market, leading to new research challenges.

Back in the days of paper maps, data producers had a good control over the final output and usages. The integration of data from different maps required technical skills and specialized equipment. Metadata and usage warnings were part of the map legend and easier to understand. With the arrival of digital data in the 1970s, it became easier to exchange and overlay data from different sources. The need to embed quality information within the datasets appeared as soon as the early 1980. For example, Chrisman (1983) clearly stated that “new data structures will have to evolve to encode the quality component, particularly for long-term, routinely maintained projects” and that such evolution was necessary “to assess the fitness of the spatial data to a given purpose”. However, the industry focus of the 1980s was to develop more efficient systems to produce and manage spatial data.

Then, the demand to exchange and reuse these expensive data led to international standards in the mid-1990s (cf. ISO/TC-211, OGC). The arrival of the Internet and the need for metadata played key roles into this evolution. As technology and data exchange improved, data quality became a more important issue for industry. The concept of “fitness for use” identified by Chrisman a decade earlier became a more common topic as the industry acknowledged the problem described by

Goodchild (1995): “GIS is its own worst enemy: by inviting people to find new uses for data, it also invites them to be irresponsible in their use”.

It is in the early 2000s, along the development of National Spatial Data Infrastructures (NSDI), that the ISO/TC-211 standards about *Quality principles* and about *Quality evaluation procedures* were proposed. This epoch also witnessed the arrival of GPS devices for the masses and of Google Map. Nowadays, we find web-based collaborative systems where map users also produce spatial data. We also see easy-to-deploy spatial data mashups mixing data from several sources thanks to powerful web services, free open-source software and mature interoperability standards. Spatial data and services have become mass-market !

As for every new mass market products and services, our society is adapting itself to protect the public against the increased risks of data misuses. Such misuses are increasingly taking place on a regular basis and appearing in popular literature (ex. daily newspapers), court decisions and specialized literature (scientific journals, conferences, workshops proceedings, web sites and blogs). We are witnessing a growing number of workshops about Law and Spatial Data. Meanwhile, the ISO standards about spatial data quality are evolving (cf. ISO-19157) and the scope of interests with regards to spatial data quality is enlarging. Accordingly, a number of concepts are appearing in our community, some of them more regularly than in the past: certification, accreditation, inspection, audit, warranty, quality assessment, quality control, quality assurance, etc.

These concepts are natural ways for a society to organize itself when a mass of citizens is facing increasing risks of misusing given products or services. Such concepts help to define who is responsible or liable for these risks, who absorbs the remaining uncertainty inherent to the use of data once technical means have contributed to reduce this risk. As stated by Bedard (1988), “most of the ways to reduce uncertainty are technical, while most of the ways to absorb the remaining uncertainty are institutional”. Such means characterize a market that is maturing. The goal of this paper is to propose a coherent synthesis of these institutional concepts recently surrounding spatial data quality and to indicate their impact on the data quality research agenda.

2 Professional Services and Products to Assure Quality Processes in the Production of Spatial Data

The first quality concern typically introduced by data producers in their daily operations is a proactive process-driven step aiming at preventing defects. Such an approach is called **Quality Assurance** (QA) and is defined by ISO-9000 as “all the planned and systematic activities implemented within the quality system, and demonstrated as needed, to provide adequate confidence that an entity will fulfill requirements for quality”. Several spatial data producers are certified ISO-9001; they normally have an obligation of mean but no obligation of result. Consequently, QA is not sufficient since it cannot *guarantee* the production of quality data. However, in our emerging era of spatial data consumerism, such a *guarantee* is likely to become mandatory for spatial products and services. According to Morris (1981), to **guarantee** means “to assume responsibility for the debt, default, or miscarriage of”.

To provide adequate confidence into new products and services, it is common usage to perform an ***Audit*** which is defined as “systematic, independent and documented process for obtaining audit evidence (records, statements of fact or other information) and evaluating it objectively to determine the extent to which the audit criteria (set of policies, procedures or requirements) are fulfilled" (ISO-19011, 2003). Thus, an *audit* in context of QA, evaluates if the processes and production methods (PPM) are suitable and effective to comply with users' requirements, with standards or with product or service *specifications*. **Data product** s**pecifications** are typically defined by the producer and presented as a "detailed description of a dataset or dataset series together with additional information that will enable it to be created, supplied to and used by another party" (ISO-19131, 2007). Spatial data specifications have existed for several decades, especially in photogrammetry, geodesy and topographic mapping and have had a special emphasis on precision and completeness.

A more advanced step towards quality data is the **Certification** process. The most popular *certification* in the QA context is ISO-9001. It refers to "the issuing of written assurance (the certificate) by an independent external body that it has audited a management system and verified that it conforms to the requirements specified in the standard" (ISO Management standards). An organisation having an ISO-9001 certificate has demonstrated "its ability to consistently provide product that meets customer and applicable statutory and regulatory requirements" (ISO-9001). To deliver such a certificate, the *certification* process typically involves a quality *audit*. An *audit* made in a *certification* context can also result in an authorization of using a *certification* mark on the product (ex.: biologic certified). *Certification* is frequently used when meeting a standard is mandatory or when nonconforming products cause high risks of loss and damage.

While *certification* evaluates compliance to requirements or standards, **accreditation** usually evaluates the competency to certify. In the ISO-9001 context: "*accreditation* refers to the formal recognition (a *certification*) by a specialized body (an *accreditation* body) that a *certification* body is competent to carry out ISO *certification* in specified business sectors". However, *accreditation* is also the mechanism used by a customer to evaluate a supplier's competency, processes and production methods, thus requiring a quality *audit*. ISO-19158 defines *accreditation* in such a context, i.e. as a "procedure by which a customer assures that its suppliers are capable of consistently delivering the product to the required quality". This type of customer *accreditation* has been used by government agencies contracting private companies for the production of highly technical documents such as topographic maps and cadastral maps.

When it relates to individuals competency with spatial data, such QA methods can lead to *certifications*, *accreditations* or licenses. For example, the surveyor or the engineer ***license*** may be required to have the right to produce certain categories of spatial data and such a license typically requires a bachelor degree from an *accredited* university program. Another example to indicate a level of competency with spatial data is the *certification* delivered on a voluntary basis by professional associations like the Canadian Institute of Geomatics, the ASPRS and the URISA. Such *certification* processes can follow ISO 19122 *Qualification and certification of personnel*. Finally, short-term diplomas entitled "certificate" can be delivered by education institutions after completion of a small number of courses or by private companies after completion of equipment training (ex. ESRI, Microsoft, CISCO, Trimble *certifications*).

3. Quality Control Professional Services and Products to Ensure Spatial Data Quality

While QA is a proactive process-driven approach aiming at preventing defects and focusing on the obligation of means, Quality Control (QC) is a reactive product-based approach aiming at finding defects and focusing on the obligation of results. QC directly addresses the quality of spatial datasets, i.e. of the final products. **Quality Control** (QC) is defined by ISO-9000 as “the operational techniques and activities that are used to fulfill requirements for quality”. To verify each spatial dataset, QC activities such as tests and inspections are performed by skilled people accordingly to quality requirements and specifications. When defects are found, they are reported to managers who can adjust the process chain and QA. Normally, QC answers an obligation of results and can be used as support to produce a **commercial guarantee.** Such a *guarantee* is any promise given voluntarily by the producer in writing to the customer, “to reimburse the price paid or to replace, repair or handle goods in any way if they do not meet the specifications set out in the guarantee statement or in the relevant advertising” (Consumer Affairs Act, art.72). A *guarantee* assumes responsibility for a default but does not *guarantee* the absence of default. To ensure safety and no risk, there exist many mandatory standards for which products must conform. In this circumstance, *certification* is practically a necessity. With voluntary standard, *certification* also certifies conformance to standard and allow products to use *certification* mark, such as Certified OGC-compliant logo.

4. Levels of Trustability into Professional Activities and Documents with regards to QA and QC

The value of previous QA and QC activities depends on who examines the quality. According to Parker (2005) and Raynolds *et al* (2006), there is 3rd, 2nd and 1st party assessments. Each of them has a different level of trust.

A ***3rd-party QC or QA assessment*** is an evaluation made by an independent and neutral outside body. Third-party examination typically entail: (1) an outside *audit* of an organization’s documentation of requirements compliance if we are in a context of QA or (2) a product quality control inspection in a context of QC. Third-party bodies are typically *accredited* to be able to assess, inspect or certify. For example in the context of QA, ISO-9001 *certification* is a 3rd-party *certification* done by an ISO-9001 accredited auditor. In a QC context, many products are 3rd party certified by laboratories or accredited bodies. Third party assessment is generally viewed as the most comprehensive and accurate method to ensure quality.

A ***2nd-party QA or QC assessment*** is performed by a user or customer to evaluate for themselves the fitness of products or processes. In a QA context, 2nd-party examinations entail *audit* by them and are often utilized to ensure quality of an organization’s supply chain (ex. ISO-19158 *accreditation)*. In a QC context, one can also assess if a product meets requirements different than those originally intended; it is the case when GIS professionals have contracts from their customers to validate if certain commercial spatial datasets meet his requirements. The latter requirements are usually expressed using spatial data quality elements and techniques described in standards such as ISO-19113, ISO-19114, National Standard for Spatial Data Accuracy (NSSDA) and ASPRS.

A ***1st-party QA or QC assessment*** is a producer's self-examination of compliance of processes or products with self-defined criteria. Self-inspection and self-*audit* are regularly made by producers to ensure that their products and processes reach expecting quality. Such *self-certifications* can be used by the producer to deliver *guarantee certificates* that spatial data meet quality measures written in the specifications. To increase the trustability of self-*certification*, organizations can obtain an ISO-9001 *certification*, meaning that the organization gives oneself the means to produce quality products.

5 Ethics and Legal aspects of Professional Activities

The current state of the law raises many legal aspects related to data quality delivered by spatial data producers, including liability, *legal guarantee*, citizens' privacy and copyright to name a few. Insofar, many spatial data producers use liability exclusions and *guarantee* exclusions about their data such as: (1) no warranty of any kind to the use or appropriateness of the use, (2) no *guarantee* of completeness or currentness, (3) no obligation to correct defects, errors and to update. When spatial data are distributed in a controlled environment where a contract exists between the producer and the customer (typically two organizations), the liability and *guarantee* clauses are clearly defined. However, when spatial data are distributed to the general public, there may be a voluntary *Commercial guarantee* offered by the producer, but there is always a **legal guarantee** which is the obligation to deliver goods in conformity with the descriptions and specifications in the contract of sale to consumers. According to art.73 of Consumer Affair Act and if we assume that spatial data are goods like other goods, then, they are in conformity with the contract if : "(a) they comply with the description given by the trader and possess the qualities of the goods which the trader has presented to the consumer as a sample or model; (b) they are fit for any particular purpose for which the consumer requires them and which he made known to the trader at the time of the conclusion of the contract and which the trader has accepted; (c) they are fit for the purposes for which goods of the same type are normally used or (d) they show the quality and performance which are normal in goods of the same type and which the consumer can reasonably expect". Such obligation may extend to a given period.

Besides laws to protect consumers, there also exist obligations for professionals which are dictated by contracts, codes of ethics and laws. According to Gervais (2003), contractual obligations of GIS Professionals include: (1) to consider user requirements and wishes, (2) effectiveness, (3) to verify and control spatial and descriptive dimensions, (4) to ensure technology compatibility, (5) to ensure evolution capabilities, and (6) to ensure database monitoring. Injuries sustained due to breach of contract, negligence or misfeasance are awarded by monetary compensation. Most codes of ethics have similar guidelines based on concept of morality. In the GISCI code of ethics, we can find guidelines such as: "encouragement to make data and findings widely available, to document data and products, to be actively involved in data retention and security, to show respect for copyright and other intellectual property rights, and to display concern for the sensitive data about individuals discovered through spatial or database manipulations" (Craig *et al*, 2003). If a professional violates a code of ethics, his licence or certificate can be revoked.

6 Conclusion

As spatial data are entering the consumer world, a new era of rights and obligations has begun. Nowadays, many users *de facto* perceive spatial data as reliable for their usages. Their perception of quality is different than that of experts. The increasing number of incidents and accidents involving spatial data is driving Society to protect these users against the risks of data misuses. Accordingly, we see a growing number of workshops, conferences, blogs and publications involving Law and Spatial Data. Such trend suggests Society is maturing regarding digital spatial products. Nevertheless, if someone complains about damages and wants to know who is liable for the quality of the data involved, do we know immediately what to look for? If the answer is "no", then research is still needed to offer Society the services required. We must develop the knowledge and services to stand in court as experts in front of judges and provide professional answers when needed. We must expand our R&D horizons towards those concepts involved in QA and QC. Accordingly, we started by investigating *which type of services* (audit, inspection, certification, etc.) can be performed *by who* (1st, 2nd and 3rd party) *regarding what* (system, product, individual). Analysing the pertinence of consumer affair acts for spatial data and doing a cross national comparaisons should be next.

Acknowledgments

GEOIDE Project #PIV-23 and Canada NSERC.

References

Bédard, Y., (1988), "Uncertainties in Land Information Systems Databases". *In: Proceedings of AutoCarto 8*, Baltimore, USA, pp. 175-184.

Chrisman, N. (1983), "The role of quality in the long-term functioning of a Geographic Information System". *In: Proceedings of AutoCarto 6*, Hull, Canada. pp.303-321

Consumer affairs Act (1996), chapter 378, Laws of Malta.

Devillers, R., Stein, A., Bédard, Y. *et al* (2010), "30 years of research on Spatial Data Quality-Achievements, failures and opportunities", *Transactions in GIS*, Vol. 14(4):387-400.

Gervais,M. (2003), *Élaboration d'une stratégie de gestion du risque juridique découlant de la fourniture de données géographiques numériques*. PhD thesis, Laval Univ., Canada

Craig, W. J., Fetzer, J. H., Onsrud, H., Somers, R. and Olson J. M. (2003), *A GIS Code of Ethics*. Approved by URISA Board of Directors.

Goodchild, M.F.(1995), Sharing Imperfect Data. In Sharing Geographic Information, H.J.

ISO-19011 (2003), Guidelines for quality and environmental management systems auditing.

ISO-19131 (2007), Geographic information – Data product specifications.

ISO/DTS-19158 (2010), Geographic information – Quality assurance of data supply.

ISO-9001 (2008), Quality management systems –Requirements.

Morris, W. (1981), *The American Heritage Dictionary of the English Language*. Publisher Houghton Mifflin Company, Boston, USA, 1550p.

Parker, B. (2005), *Introduction to globalization & business*. London: Sage Publications.

Raynolds, L.T., Murray, D. L., Wilkinson, J. (2007), *Fair trade: the challenges of transforming globalization,* Routledge, Taylor & Francis Group: London, UK. 240p.