Natural Resources Canada | Ressources naturelles Canada

CANADIAN GEOSPATIAL DATA INFRASTRUCTURE
INFORMATION PRODUCT 43e

Geospatial Data Quality Guide

GeoConnections
Intelli3 Inc.

2016

Canada

Geospatial Data Quality Guide

Prepared by:

Intelli[3] inc.
3700 Wilfrid-Hamel, suite 80
Quebec (Qc), Canada, G1P 2J2

Presented to:

GeoConnections
Natural Resources Canada

2015

ACKNOWLEDGEMENTS

The authors, Dr. Yvan Bédard, Sonia Rivest, M.Sc. from Intelli^3, and Dr. Marc Gervais, professor at Laval University in Quebec City, would like to thank Jean Brodeur, from GeoConnections at Natural Resources Canada, who supervised the project.

Table of Contents

Executive summary

The production, distribution and usage of geospatial data have changed dramatically in the last decade. Traditional industry and government data providers have been joined by new major IT players to facilitate the dissemination and integration of geospatial data for the mass market. This is changing the game with regards to evaluating geospatial data quality and managing the risks of inappropriate data usages. It is progressing from a traditional Business-to-Business (B2B) mindset towards a Business-to-Consumer (B2C) and Consumer-to-Consumer (C2C) mindset. In the new B2C and C2C contexts, typical geospatial data users do not understand the uncertain nature of geospatial data and take digital data for granted. This sometimes leads to erroneous decisions, possibly having significant social, political or economic consequences. In mass markets, consumers have certain rights and the providers of goods or services have obligations. Regulations and court decisions suggest that consumers must be informed about the quality of a product or service, as well as about the risks and prohibited usages. Examples of such information can be found in user manuals, guarantees and warnings.

This guide explains how to manage geospatial data quality and risks of usage at every phase of a data product life-cycle: design, implementation, production, delivery and usage. It explains the geospatial data evaluation process as presented in the ISO 19157 Geospatial information – Data quality international standard. It also presents the general risk management framework of the ISO 31000 Risk management – Principles and guidelines international standard along with numerous examples related to the management of risks of inappropriate usage of geospatial data.

Recommendations related to geospatial data quality and risk management are given for the B2B, B2C and C2C contexts.

List of Figures

List of Tables

List of Acronyms

The following table presents the meaning of acronyms used in this document.

Acronym/abbreviation	Meaning
B2B	Business-to-Business
B2C	Business-to-Consumer
BI	Business Intelligence
C2C	Consumer-to-Consumer
CGDI	Canadian Geospatial Data Infrastructure
CHS	Canadian Hydrographic Service
CSDGM	Content Standard for Digital Geospatial Metadata
DQWG	Data Quality Working Group
FAQ	Frequently Asked Questions
FGDC	Federal Geographic Data Committee
FGP	Federal Geospatial Platform
INSPIRE	Infrastructure for Spatial Information in the European Community
ISO	International Organization for Standardization
IT	Information Technology
NCCDS	National Committee on Digital Cartographic Data Standards
NSDI	National Spatial Data Infrastructure
NTS	National Topographic System
OGC	Open Geospatial Consortium
SDI	Spatial Data Infrastructure
SDTS	Spatial Data Transfer Standard
UK	United Kingdom
USA	United States of America
VGI	Volunteered Geographic Information
VQI	Volunteered Quality Information
XML	Extensible Markup Language

1. Introduction

The production, distribution and usage of geospatial data have changed dramatically in the last decade. Traditional industry and government data providers have been joined by new major IT players to facilitate the dissemination and integration of geospatial data for the mass market. Thanks to smartphones, navigation systems, numerous sources of digital maps and imagery and virtual globes, citizens are massively consuming and producing geospatial data. While the 1950s were the beginning of the hardware era and the 1980s the beginning of the software era, the present decade is the beginning of the data era.

This is changing the game with regards to evaluating geospatial data quality and managing the risks of inappropriate data usages. It is progressing from a traditional Business-to-Business (B2B) mindset towards a Business-to-Consumer (B2C) and Consumer-to-Consumer (C2C) mindset. The number of experts dealing with geospatial data has been widely surpassed by the number of non-experts and citizens. Accordingly, the roles and responsibilities of producers, distributors and users of geospatial data and services are evolving. Similarly, national geospatial data infrastructures are adapting to the new challenges of data quality evaluation, quality information communication and risk management.

1.1 Geospatial data quality: why?

Data are collected to meet well-defined goals and their quality is typically defined to meet these goals. Collecting data to fit one's need within given budgets and delays is not simple given the panoply of methods and technologies. Nevertheless, sources, users and usages of geospatial data are nowadays more diverse than ever before.

To minimize costs and delays, geospatial data reuse and interoperability have become popular. However, finding geospatial data that is the best fit for a particular purpose is not easy. Once found, these data are often shared, extracted, mashed-up, and regularly pushed wirelessly on users' devices or used on-demand for new purposes that differ from the original one.

In the new B2C and C2C contexts, typical geospatial data users do not understand the uncertain nature of geospatial data and take digital data for granted. This leads to erroneous decisions with social, political or economic consequences (Devillers, Bédard, & Gervais, 2013). Newspapers in Canada, USA, France, UK, Australia, etc. have reported accidents, injuries, material damages to bridges and houses, 911 delays, people lost in the wild and casualties related to the use of geospatial data (see project (GEOIDE (PIV-23), 2012)). Courts render decisions about faulty maps, missing information, improper warnings, liability, etc. (Gervais, Bédard, Jeansoulin, & Cervelle, 2007), (Chandler, 2010), (Chandler & Levitt, 2011).

When we read Codes of Ethics or Good Practice Guidelines of several professions, they typically state that professionals have the duty to protect their clients. In mass markets, consumers have certain rights and the providers of goods or services have obligations. Regulations and Court decisions suggest that consumers must be informed about the quality of a product or service, as well as about the risks and prohibited usages. Examples can be found in user manuals, guarantees and warnings. Similarly, governments influence what can be done with data regarding privacy (see the blog http://www.teresascassa.ca/). These typical issues of mature mass markets are entering the B2C and C2C contexts of geospatial data.

A recent survey in the Canadian geomatics industry (Gervais, Bédard, Larrivée, Rivest, & Roy, 2013) indicated that 70% of respondents believe that users are not aware of the potential risks of using geospatial data. Respondents also show strong concerns about users' ability to manage risks and 81% thought that the geospatial industry could do more to reduce users' risks. Users' primary complaints concerned poor documentation and data quality according to this survey, while a recent needs analysis by (Hickling Arthurs Low Corporation, 2011) mentioned that 15% of the issues raised by respondents regarded data quality, making it one of the most prominent issues occurring in geomatics.

Accordingly, the Canadian Geospatial Data Infrastructure (CGDI) aims to help the geospatial community meet the data quality challenge as data quality affects every link in the service chain (GeoConnections, 2013). This will facilitate comparing data sources, analyzing fitness-for-use, estimating the cost of preparing data, evaluating the risks of inappropriate data usages, and improving communication about quality and risks to users. This will also help building a community of good practices.

1.2 Geospatial data quality: who should care?

To know the quality of geospatial data, every step of the data supply chain must follow quality assurance and quality control processes. Every player contributes, from needs analysis and database design to the acquisition, integration, transformation and dissemination of data.

There is a long tradition in the geomatics industry to consider geospatial data quality because it is a key determinant of the cost of a project. In the B2B context, we often see contracts where quality is defined with clear specifications and responsibilities of the identified parties. Practices are sophisticated and tend to focus on spatial accuracy, temporal accuracy, semantic accuracy, attribute accuracy, completeness, logical consistency, lineage and conformance to standards. Such information is typically communicated using metadata. In spite of standards, this remains a B2B culture where experts talk to experts (Devillers, Stein, Bédard, Chrisman, Fisher, & Shi, 2010). In today's B2C and C2C contexts, data quality concerns and proper user information must consider potential redistribution and reuse of the data (known and unknown).

Accordingly, the following players must care about geospatial data quality:

- Every designer and developer of systems using geospatial data, for example:
  - experts in geospatial information technology (hardware, software, databases)
  - amateurs developing web-based mapping applications
  - developers of location-aware apps for smartphones
- Every provider of geospatial data, for example:
  - source data producers and distributors
  - data mashup producers and distributors
  - crowdsourcing contributors (e.g., "produsers" involved in Volunteered Geographic Information (VGI))
- Every expert involved in geoprocessing and disseminating geospatial data
- Every user of geospatial data:
  - first-hand users of the data (i.e. primary users of a system or dataset)
  - second-hand users (i.e. reusers of data, in a different system or data mashup)
  - expert users as well as non-experts and citizens

This includes all members of the geospatial community, not only the geomatics industry. For example, a computer scientist developing a web-based map, a car driver using a navigation system, a person looking for a place on a virtual globe, a hiker using his GPS to find his way, an expert designing a system with OpenStreetMap, a land owner using an aerial photograph on the web to locate his fence, etc. Geospatial data and services have become mass market products; the general rules of quality, information and responsibility-sharing apply.

1.3 Geospatial data quality and risk management

Caring about geospatial data quality requires some evaluation of this quality. The evaluation process can result in general or detailed information, it can apply to a complete dataset or only a subset (e.g., a given area, one feature type), cover one or several purposes, be well estimated or only perceived, etc. Depending upon the extent of the evaluation, it can be very complex and robust or it can remain general and show a higher level of uncertainty.

The concepts and procedures used to evaluate the quality of geospatial data are described in detail in the ISO 19157 international standard (International Organization for Standardization, 2013). It is the primary source of information to perform an evaluation of geospatial data quality in a form that is commonly understood by experts and suitable for interoperability. For this reason, the present guide builds upon concepts of this standard.

ISO 19157 was developed in a B2B context and is flexible regarding the level of detail of quality analyses. Yet, it is not geared towards the B2C and C2C contexts. In the B2C context, experts serve non-experts and have the professional duty to look at risk management strategies to reduce potential negative impacts for these non-expert users. Such a duty emanates from the Code of Ethics of licensed professionals, the concept of precaution, consumer protection and liability issues. In mature markets, this duty takes place when those who have knowledge offer products and services to those who do not, or to the masses. In the C2C context, the concepts of precaution, consumer protection and liability issues also call for risk management.

Risk management helps to select the best strategies to minimize the risk of inappropriate use of geospatial data in situations of uncertainty (uncertain data, uncertain data quality, uncertain usages, and uncertain expertise of users). Such strategies raise the awareness of users and providers, reduce the overall risk and help recognize the responsibility of every player. Without risk management actions, evaluating geospatial data quality is not socially complete. Furthermore, geospatial data quality results must be well communicated to users.

Accordingly, the present guide explains the main steps to evaluate geospatial data quality and manage the potential risks of inappropriate use of data. It is how mature markets protect both consumers and providers of goods and services.

1.4 Why this Guide?

This Guide aims at helping organizations and individuals involved in spatially-enabling the Canadian society in the contexts of B2B, B2C and C2C. In particular, this Guide will support the Canadian geospatial community into its efforts to evaluate data quality with the help of international standards such as ISO 19157 (Geospatial Data Quality) and ISO 31000 (Risk Management).

Such a global and innovative view should further encourage the widespread usage of geospatial data and the Canadian Geospatial Data Infrastructure (CGDI). It should facilitate the selection of data that best fit one's needs, facilitate interoperability and stimulate the adoption of good practices seen in mature markets. Finally, this Guide should bolster the involvement of players hesitant about open-data initiatives, crowdsourcing, data mashups and geospatial data ubiquity.

Section 2 of this Guide reviews the concepts underlying geospatial data quality. Section 3 discusses the theoretical aspects of geospatial data quality evaluation and the management of risks of inappropriate usage of geospatial data, including a description of the standards supporting these processes. Section 4 describes the detailed actions to be undertaken when evaluating the quality of a geospatial dataset or service and when managing the risks involved, along with examples. Section 5 provides recommendations for each of the contexts (B2B, B2C, and C2C) described above. Conclusions and a list of references follow.

2. Background

There is a long tradition in the geomatics industry to consider geospatial data quality. Works on the topic increased significantly with the arrival, in the early 1980s, of Geographic Information Systems (GIS) and their capability to integrate spatial with non-spatial data (Devillers, Stein, Bédard, Chrisman, Fisher, & Shi, 2010). A few years later, the use of the notion of "fitness for purpose" for defining quality, first proposed by (Juran, Gryna, & Bingham, 1974), was introduced in the geospatial community (Chrisman, 1983).

Elements of geospatial data quality were addressed by the US National Committee on Digital Cartographic Data Standards (NCDCDS), in 1987, in their report "A Draft Proposed Standard for Digital Cartographic Data" (National Committee for Digital Cartographic Data Standards, 1987). This report served as a base for the creation of the Spatial Data Transfer Standard (SDTS). In 1998, the Federal Geographic Data Committee (FGDC) created the Content Standard for Digital Geospatial Metadata (CSDGM) (Federal Geographic Data Committee, 1998) to complement the SDTS. It was in the early 2000s, along with the development of National Spatial Data Infrastructures (NSDI), that the ISO/TC 211 standards about quality principles and quality evaluation procedures were proposed (ISO 19113:2002 Geographic information – Quality principles and ISO 19114:2003 Geographic information – Quality evaluation procedures respectively, now replaced by the ISO 19157:2013 Geographic information – Data quality). In 2007, the Open Geospatial Consortium (OGC) formed the Data Quality Working Group (DQWG) with the mission to establish a forum for describing an "interoperable framework or model for OGC Quality Assurance measures and Web Services to enable access and sharing of high quality geospatial information, improve data analysis and ultimately influence policy decisions" (Trakas, 2008), (Open Geospatial Consortium, 2014). In 2012, ISO published the technical specification ISO/TS 19158:2012 Geographic Information – Quality assurance of data supply, which provides a framework for quality assurance specific to geographic information.

An overview of 30 years of research in geospatial data quality can be found in (Devillers, Stein, Bédard, Chrisman, Fisher, & Shi, 2010).

2.1 The inherent uncertainty of geospatial data

The acquisition of geospatial data involves a selection and an abstraction process geared towards a specific goal. Different goals will lead to different selections, and hence, to different datasets having different attributes and geometries, even if they represent the same entities of the real world. Models supporting geospatial datasets are only approximations of the real world. It is impossible to produce a perfect model of the reality and thus, every model is inevitably associated with a level of uncertainty (Longley, Goodchild, Maguire, & Rhind, 2001). There are 4 orders of uncertainties (Bédard, 1988):

- Conceptual uncertainty, which refers to the fuzziness in the identification of an observed reality (e.g., being or not being a "river" impacts on the existence or non-existence of an object; being a "stream" or a "creek" impacts on the category of the object, on the attributes to be measured for that object, and on the specifications used to measure its geometry).

- Descriptive uncertainty refers to the uncertainty in the attribute values of an observed reality (i.e., imprecision in quantitative values and fuzziness in qualitative values).

- Location uncertainty refers to the fuzziness in the qualitative values used for spatial or temporal referencing (e.g., city name, street name, day of week) and to the imprecision in the quantitative values used for the location in space and time of an observed reality (e.g., precision of ± 5 meters, ± 1 day).

- Meta-uncertainty refers to the degree to which the preceding uncertainties are unknown (e.g., error ellipses with a probability of 95%; being confident that a river is of class "1").

The four orders of uncertainty combine to generate the total uncertainty of the dataset. Geospatial data uncertainty can vary spatially and over time. Geospatial data uncertainty can be reduced to a certain level, and the remaining uncertainty needs to be absorbed. The higher the uncertainty reduction, the lower the uncertainty absorption needed (Bédard, 1988). Various means can be used to reduce or absorb uncertainty. Most of the ways to reduce uncertainty are technical, while most of the ways to absorb the remaining uncertainty are institutional. Example ways to reduce the uncertainty include: following guidelines and standards, improving data collection methods, receiving training, etc. Example ways to absorb the remaining uncertainty include providing a guarantee for the dataset, having insurance to cover the possible damages, relying on mutually agreed-upon contract or clauses, etc. The balance between uncertainty reduction and uncertainty absorption is an example of risk management process.

2.2 The perspectives of geospatial data quality

Many definitions of quality exist. (Juran, Gryna, & Bingham, 1974) were the first to define quality as "fitness for use". What is good quality for one may not be considered good quality for someone else. This definition is now widely recognized and is also used in the geospatial community (Chrisman, 1983), (Veregin, 1999). ISO 9000 defines quality as the totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs (International Organization for Standardization, 2005). ISO 19157 defines quality as the degree to which a set of inherent characteristics fulfils requirements (International Organization for Standardization, 2013). This standard recognizes that a data producer and a data user may view data quality from different perspectives. For a data producer, data quality is defined as how well a dataset reflects its universe of discourse, as established in the data product specifications. For a data user, data quality is defined as the ability of the dataset to satisfy the requirements of the user's application. The producer's perspective is often referred to as the *internal quality* of a dataset, or its intrinsic properties resulting from data production methods. The user's perspective follows the fitness for use definition and is often referred to as the *external quality* of a dataset, or the level of fitness between data characteristics and users needs. There are as many evaluations of external quality as there are usages of a dataset. There are as many evaluations of external quality as there are usages of a dataset.

In addition to the concepts of internal quality and external quality, the literature also distinguishes the concepts of perceived quality and metaquality. The different aspects of geospatial data quality are described in the following sections.

2.2.1 Internal quality

Internal quality is described, according to ISO 19157 (International Organization for Standardization, 2013), using five parameters: completeness, logical consistency, positional accuracy, thematic accuracy, and temporal quality.

Completeness is defined as the presence and absence of features, their attributes and relationships. Completeness is described using the following data quality elements:

- commission – excess data present in a dataset (e.g., a redundant building)
- omission – data absent from a dataset (e.g., a missing building)

Logical consistency is defined as the degree of adherence to logical rules of data structure, attribution and relationships. Logical consistency includes the following consistencies:

- conceptual – adherence to rules of the conceptual schema (e.g., all bridges cross rivers)
- domain – adherence of values to the value domains (e.g., no year older than 1900)
- format – degree to which data is stored in accordance with the physical structure of the dataset (e.g., format for months is numerical (and not alphabetical), i.e. 04 for April and not Apr)
- topological – correctness of the explicitly encoded topological characteristics of a dataset (e.g., there is a node at every street intersection)

Positional accuracy is defined as the accuracy of the position of features within a spatial reference system. It is described using the following elements:

- absolute accuracy (or external accuracy) – closeness of reported coordinate values to values accepted as, or being true (e.g., ± 5m)

- relative accuracy (or internal accuracy) – closeness of the relative, spatial positions of features in a dataset to their respective relative spatial positions accepted as, or being true (e.g., ± 1m)
- gridded data position accuracy – closeness of gridded data spatial position values to values accepted as, or being true (e.g., ± 50m)

Thematic accuracy is defined as the accuracy of quantitative attributes and the correctness of non-quantitative attributes, and of the classifications of features and their relationships. Thematic accuracy is described using the following elements:

- classification correctness – comparison of the classes assigned to features, or their attributes, to a universe of discourse (i.e. ground truth or reference dataset) (e.g., 2% of building with mixed usages can be misclassified as a commerce or as a residence)
- non-quantitative attribute correctness – measure of if a non-quantitative attribute is correct or wrong (e.g., 5% of building architecture style may be incorrect)
- quantitative attribute accuracy – closeness of the value of a quantitative attribute to a value accepted as, or known to be true (e.g., commercial value is ± 10%)

Temporal quality is defined as the quality of the temporal attributes and temporal relationships of features. It is described using the following elements:

- accuracy of a time measurement – closeness of reported time measurements to values accepted as or known to be true (e.g., ± 1 minute)
- temporal consistency – correctness of the order of events (e.g., date of cadastral subdivision before date of house construction)
- temporal validity – validity of data with respect to the format and calendar specified for the dataset (e.g., no February 30th)

2.2.2 External quality

External quality of a geospatial dataset is the degree of agreement between data characteristics (i.e., internal quality) and the explicit and/or implicit needs of a user for a given application in a given context. The external data quality values will vary from one application to the other. ISO 19157 (International Organization for Standardization, 2013) introduces the usability data quality element. *Usability* is based on user's requirements and all internal quality elements of ISO 19157 can be used to evaluate the usability of a geospatial dataset.

2.2.3 Perceived quality

Within a collaborative environment, each user may have a different view of the external quality of a dataset. This is the concept of perceived quality (Grira, Bédard, & Roche, 2009). In such context, the process of defining the fitness for use of a geospatial dataset might be complicated by the number of users involved and the variety of specific needs. Each user has his own perception of the quality of the product being used (like the quality of an album on Amazon or Apple Store). To facilitate the reach of a consensus, each user can rate the dataset based on his perception (e.g., using a 5-star rating system and comments, as Amazon and Apple, see (Jones, Devillers, Bédard, & Schroth, 2013) and (Koistinen, 2015) for examples). The global perceived quality is the result of the aggregation of all individual user perception and is hence based on a bottom-up approach. Such crowdsourced processes have been called Volunteered Quality Information (VQI) (Bédard, 2012).

2.2.4 Metaquality

Metaquality relates to the quality of information used to determine the quality of an object, a concept or a dataset (International Organization for Standardization, 2013). In the ISO 19157 context, metaquality is composed of the following elements: confidence, representativity and homogeneity. *Confidence* is defined as the trustworthiness of a data quality result (e.g., a confidence interval on a given confidence level). *Representativity* is defined as the degree to which the sample used for evaluating quality has produced a result which is representative of the dataset (e.g., all the geographic zones and concerned time periods are covered and the population is sufficiently large). *Homogeneity* is defined as the expected or tested uniformity of the results obtained for a data quality evaluation (e.g., comparison of the evaluation results of several segments of a global data set expressed using root mean square errors).

The knowledge about the quality of a given result is often of the same importance as the result itself. Metaquality can help quantifying the risk related to geospatial data uncertainties.

2.3 The subject-matters of geospatial data quality

Data quality can be evaluated at various levels of detail, or granularity levels. The hierarchy of granularity levels can be used to aggregate quality information from an attribute value of a feature up to a complete dataset. Figure 1 presents the hierarchical levels of data quality as defined in the ISO 19157 standard: dataset series, dataset, subset of a dataset, feature type, feature instance, attribute type, and attribute instance (International Organization for Standardization, 2013). Table 1 presents examples of each of the level.

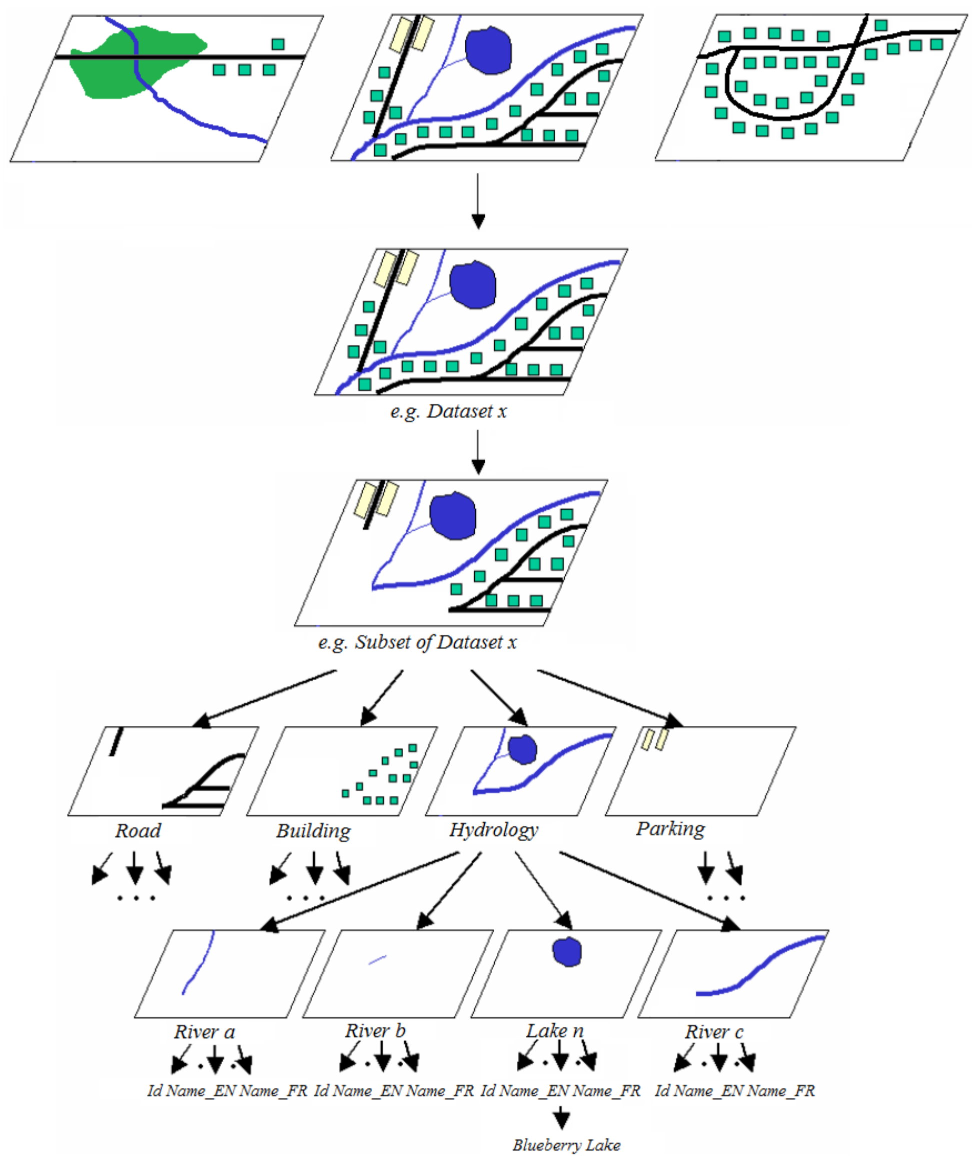

Figure 1. The hierarchical levels of geospatial data quality evaluation (adapted from (Devillers, 2004)).

Hierarchical level	Examples
Dataset series	• The National Topographic System (NTS) of Canada • The nautical charts of the Canadian Hydrographic Service (CHS)
Dataset	• A specific map of the NTS (e.g., 031G Ottawa) • A specific nautical chart (e.g., 1315 – Quebec to Donnacona)
Subset of a dataset	• Subset of data in municipality X • Subset of data for the islands north of the seaway
Feature type	• Road segment • Bridge
Feature instance	• Road 417

	• Bridge "Pierre-Laporte"
Attribute type	• Functional road class • Vertical clearance
Attribute instance	• Code 2 (Expressway/Highway) • 35

Table 1. Examples of hierarchical levels of geospatial data involved in quality evaluation.

2.4 The management of risks associated to geospatial data quality

In the changing geospatial context, the number of experts dealing with geospatial data has been widely surpassed by the number of non-experts and citizens. These new users' knowledge about the risks related to the use of geospatial data is limited. In parallel, the amount of geospatial data being made available is increasing. Geospatial data is now being shared, exchanged, integrated, mashed-up and used for purposes other than their producers' intended ones. In this context, the potential for inappropriate uses of geospatial data also increases.

In the geomatics field, a risk management process serves as a guide on how to avoid or manage the impacts of uncertainty: uncertain geospatial data, uncertain geospatial data quality, uncertain geospatial data usages, and uncertain expertise of users of geospatial data. From a legal perspective, using a risk management approach is necessary to protect both the geospatial data producer and user (Gervais, Bédard, Jeansoulin, & Cervelle, 2007).

Perfect geospatial data quality does not exist. The overall quality involves internal quality, external quality, perceived quality and metaquality. In a similar way, zero risk does not exist. Risk can be reduced, rarely eliminated. Quality and risk are interrelated. Typically, the higher the quality (and metaquality), the lower the risk to manage, as shown in figure 2.

Figure 2. The relation between quality and risk.

Geospatial data quality will be addressed in details in sections 3.1 and 4.1. Management of risks of inappropriate usage of geospatial data will be discussed in details in sections 3.2 and 4.2.

2.5 The dissemination of information about geospatial data quality and risks of usage

Information about geospatial data quality and the risks of inappropriate usages of geospatial data must be communicated to the various actors involved in the creation and use of geospatial data. The traditional way to communicate quality information is the use of metadata. However, designed by experts for experts (B2B context), metadata are less appropriate for other types of actors, particularly for the general public (B2C and C2C mass markets). The process of communicating data quality and risks must be adjusted for all audiences with new vocabularies, methods and documentation products. Since quality and risk will have different values in different usages, the information will differ for each user/usage.

The documentation proposed in this guide, according to the context, includes various methods designed to better inform users and reduce their risk of using the digital data that they used to take for granted. Based on good ethical practices and examples from more mature mass markets, these information products increase awareness of all involved players and can take various forms (e.g. text, quality-aware application, training and forum, to name a few).

Providing proper information helps producers to meet their legal duty for information, advice, and warnings. “Good data documentation and well drafted disclaimers and agreements will minimize data misuse and abuse” (National States Geographic Information Council, 2011).

Examples of the listed documentation products are presented in section 4.

2.6 The standards underpinning geospatial data quality and risk management

Many standards addressing geospatial data quality and risk management have been published. This section describes the main international standards, specific to geospatial data, or generic, as well as relevant Canadian standards.

2.6.1 Specific standards

The International Organization for Standardization (ISO) has published standards related to geospatial data quality: ISO 19115-1 and ISO 19115-2, ISO 19131, ISO 19157, and ISO 19158.

ISO 19115-1 Geographic information – Metadata – Part 1: Fundamentals

The ISO 19115-1 standard (ISO 19115-1:2014 Geographic information - Metadata - Part 1: Fundamentals), which replaces ISO 19115:2003, defines the schema required for describing geographic information and services by means of metadata (International Organization for Standardization, 2014). It provides information about the identification, extent, quality (referring to ISO 19157), spatial and temporal aspects, content, spatial reference, portrayal, distribution and other properties of digital geospatial data and services. ISO 19115-1 applies to:

- the cataloguing of all types of resources, clearinghouse activities, and the full description of datasets and services
- geographic services, geographic datasets, dataset series, and individual geographic features and feature properties

The encoding of metadata can be done in Extensible Markup Language (XML) using the ISO 19139:2007 – Geographic Information – Metadata – XML Schema Implementation (International Organization for Standardization, 2007).

ISO 19115-2 Geographic information – Metadata – Part 2: Extensions for imagery and gridded data

The ISO 19115-2 standard (ISO 19115-2:2009 Geographic Information – Metadata – Part 2: Extensions for imagery and gridded data) extends the existing geographic metadata standard by defining the schema required for describing imagery and gridded data. It provides information about the properties of the measuring equipment used to acquire the data, the geometry of the measuring process employed by the equipment, and the production process used to digitize the raw data (International Organization for Standardization , 2009).

The North American Profile of ISO19115:2003

The North American Profile of ISO19115:2003 was developed to meet the specific needs of the United Stated and Canada. This profile makes certain optional fields of ISO 19115 mandatory, supports multiple languages, and takes certain free text fields and makes them into code lists. It was published as an American National Standard by the American National Standards Institute (ANSI) and as a Canadian National Standard by the Standards Council of Canada (SCC).

Treasury Board of Canada Standard on Geospatial Data

The Standard on Geospatial data of the Treasury Board secretariat of Canada (Treasury Board of Canada, 2009) requires that the government of Canada specialists responsible for creating or using geospatial data apply the North American Profile of ISO 19115 Geographic information - Metadata (NAP - Metadata).

ISO 19131 Geographic information – Data product specification

The ISO 19131 standard (ISO 19131:2007 Geographic information - Data product specifications and ISO 19131:2007/Amd 1:2011 Requirements relating to the inclusion of an application schema and feature catalogue and the treatment of coverages in an application schema) specifies requirements for the specification of geographic data products, based upon the concepts of other ISO 19100 International Standards, such as data quality (ISO 19157) (International Organization for Standardization, 2007), (International Organization for Standardization, 2011). It also provides help in the creation of data product specifications, so that they are easily understood and fit for their intended purpose. A data product specification is 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 can also be used by users to describe their requirements.

ISO 19157 Geographic information – Data quality

The ISO 19157 standard (ISO 19157:2013 Geographic information – Data quality) (International Organization for Standardization, 2013) establishes the principles for describing the quality of geographic data by:

- defining components for describing data quality
- specifying components and content structure of a register for data quality measures
- describing general procedures for evaluating the quality of geographic data
- establishing principles for reporting data quality

ISO 19157 also contains a set of data quality measures for use in evaluating and reporting data quality. It is applicable by data producers to verify that a data product is conforming to its specifications and by data users who want to assess the fitness for use of a data product.

ISO 19157:2013 cancels and replaces ISO/TS 19138:2006, ISO 19114:2003 and ISO 19113:2002, which have been technically revised.

The details of this standard will be discussed in sections 3.2 and 4.2.

ISO/TS 19158 Geographic information – Quality assurance of data supply

The ISO/TS 19158 standard (ISO/TS 19158:2012 Geographic information – Quality assurance of data supply) (International Organization for Standardization, 2012) provides a quality assurance framework for the producer and customer in their production relationship. This technical standard identifies methods of managing the quality of production more efficiently and effectively. It enables innovation and continual improvement within the context of existing:

- geographic information quality principles and quality evaluation procedures, and
- quality management systems

2.6.2 Generic standards

ISO 9000 – Quality management

The ISO 9000 family of standards (ISO 9000: 2005 Quality management) addresses various aspects of quality management and provides guidance and tools for organizations that want to ensure that their products and services consistently meet customers' requirements and that quality is consistently improved. The ISO 9000 series comprises the following standards (International Organization for Standardization, 2005):

- ISO 9001:2008: describes the requirements of a quality management system
- ISO 9000:2005: describes the basic concepts and language
- ISO 9004:2009: describes ways to make a quality management system more efficient and effective
- ISO 19011:2011: provides guidance on internal and external audits of quality management systems

A new version of ISO 9001 will be available by the end of 2015.

ISO 31000 Risk Management – Principles and Guidelines

The ISO 31000 standard (ISO 31000: 2009 Risk management – Principles and guidelines) provides principles, framework and a process for managing risk. It can be used by any organization regardless of its size, activity or sector. Using ISO 31000 can help organizations increase the likelihood of achieving objectives, improve the identification of opportunities and threats and effectively allocate and use resources for risk treatment (International Organization for Standardization, 2009).

Related standards include:

- ISO Guide 73:2009 Risk management - Vocabulary provides terms and definitions relating to the management of risk
- ISO/IEC 31010:2009 Risk management – Risk assessment techniques focuses on risk assessment

The details of this standard will be discussed in sections 3.2 and 4.2.

3. Geospatial Data Quality and Risk Management: Getting Started

Caring about geospatial data quality involves managing quality in all the different phases of a data product's life cycle. Managing quality also requires an evaluation of this quality based on requirements or on identified needs. Depending upon the extent of the quality evaluation, results can be very complex and robust or they can remain general and show a higher level of uncertainty. To deal with uncertainty, risk management is used. Risk management helps to select the best strategies in order to minimize the risk of inappropriate use of geospatial data. Geospatial data quality results and risk management actions must be well communicated to all actors involved in the process. All these activities must be monitored and reviewed as necessary. This general process is presented in figure 3.

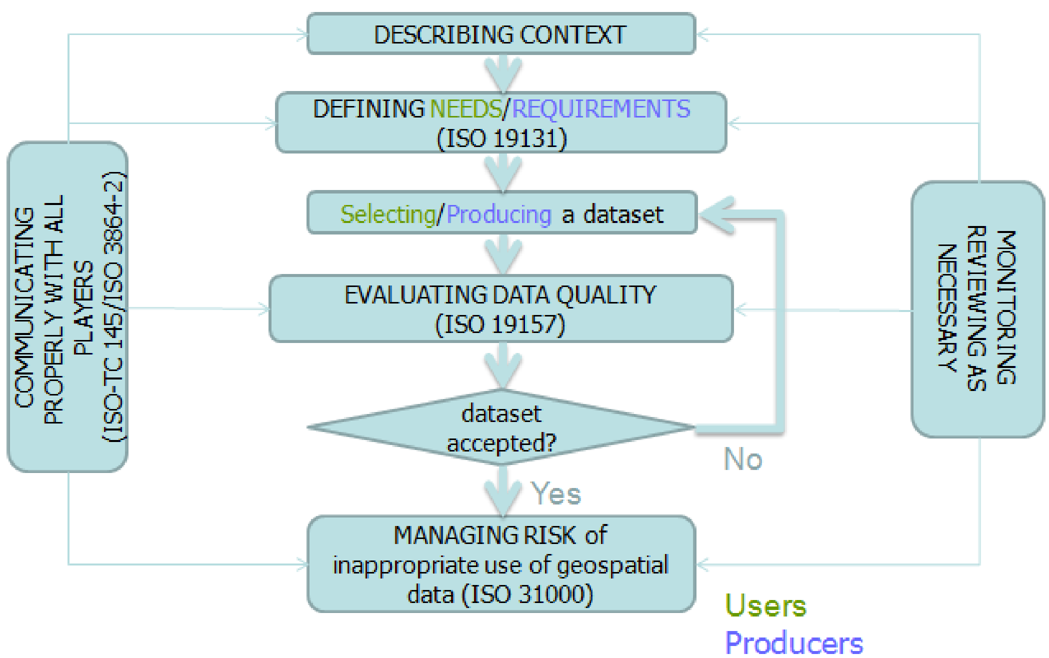

Figure 3. The complete cycle of geospatial data quality for a spatially-enabled society.

This section explains the main steps to manage geospatial data quality and manage the potential risks of inappropriate use of data. It is the way mature markets protect both consumers and providers of goods and services. Detailed examples of these activities are presented in section 4.

3.1 Geospatial data quality management

Geospatial data quality management is the activity of defining the required quality of the needed data, defining, implementing and controlling the necessary steps to ensure quality criteria are met, and evaluating, documenting and disseminating quality information (figure 4).

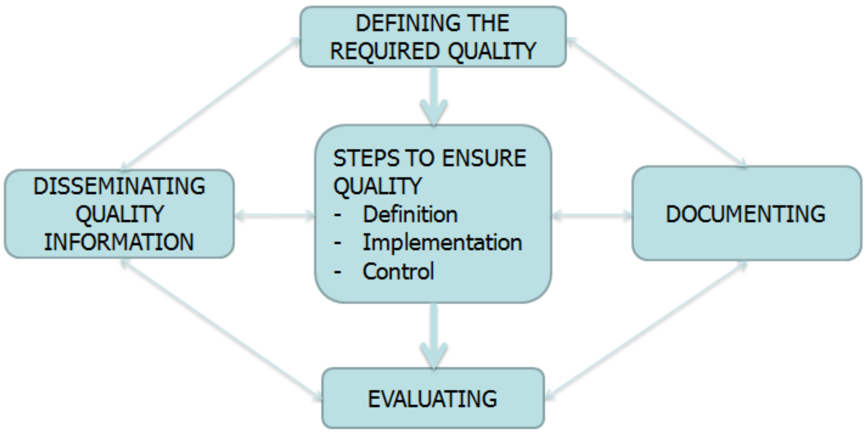

Figure 4. The geospatial data quality management process.

From a producer point of view (B2B or B2C contexts), geospatial data quality and risks of inappropriate usage must be managed at each phase of a data product life-cycle (production or update process) as illustrated in figure 5.

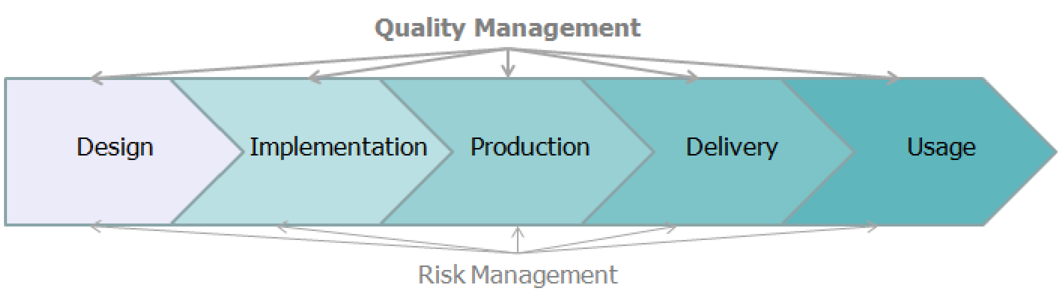

Figure 5. Management of quality at each phase of a data product life-cycle.

At the *design* phase, data quality can be managed during data modelling (application schema or schema for coverage geometry and functions), during the creation of the data dictionary (feature catalog), and when designing product specifications, as illustrated in section 4.1.1.

The following standards can be used to support the design activities:

- ISO 19109:2005 Geographic information - Rules for application schema
- ISO 19110:2005 Geographic information - Methodology for feature cataloguing
- ISO 19123:2005 Geographic information - Schema for coverage geometry and functions
- ISO 19131:2007 Geographic information - Data product specifications

At the *implementation* phase, data quality can be managed with the use of integrity constraints and spatial integrity constraints. Integrity constraints are assertions that limit the data that can be stored in a database in order to avoid inconsistencies. Examples are presented in section 4.1.2.

At the *production* phase, data quality is evaluated or controlled and metadata are managed (details in section 4.1.3). The following standards can be used to support these activities:

- ISO 19157:2013 Geographic information – Data quality
- ISO 19115-1:2014 Geographic information – Metadata – Part 1: Fundamentals

At the *delivery* phase, data quality is managed by providing users with the appropriate dataset documentation. In order for a data producer to better fulfill his/her legal duties, it is good advice to deliver datasets with, in addition to the quality related metadata, the necessary warnings about the limitations contained in the dataset and the possible risks of inappropriate use of data (Chandler & Levitt, 2011). Other possible documentation items include a traditional or interactive context-sensitive user manual, a list of recommended and non-recommended usages, training material or other deliverables identified in Section 4.1.4. In other words, risk management strategies highly influence how data quality is managed at this phase.

At the *usage* phase, data quality and risks of data misuse are also closely interlinked and they follow actions taken at the delivery phase. They can be managed by providing users with proper on-demand communication, up-to-date information, training, Volunteered Quality Information (VQI), quality-aware applications, and other strategies, as detailed in section 4.1.5.

3.2 Geospatial data quality evaluation

The evaluation of the quality of geospatial information can be extremely complex, even for an expert in geomatics. Geospatial data quality evaluation can be defined as a process used to determine whether a geospatial data product meets the objectives with regards to product specifications, from a producer point of view (i.e. internal quality), or with regards to product requirements or needs for a planned use, from a user point of view (i.e. "usability" or external quality). The evaluation process can be formal, based on the procedure described in the ISO 19157 standard, or be less formal depending upon the context defined in the first phase of risk analysis (cf. section 3.3.1). Figure 6 presents the spectrum of quality evaluation methods.

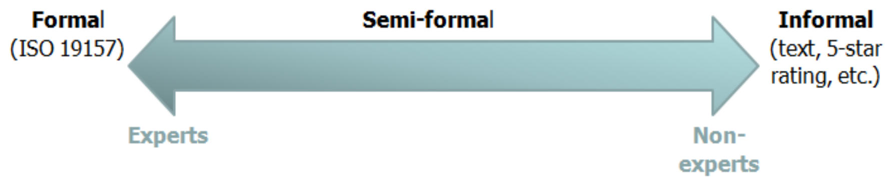

Figure 6. Spectrum of data quality evaluation methods, based on the context.

Figure 7 details the steps of the formal evaluation process (based on ISO 19157), as well as the informal evaluation process.

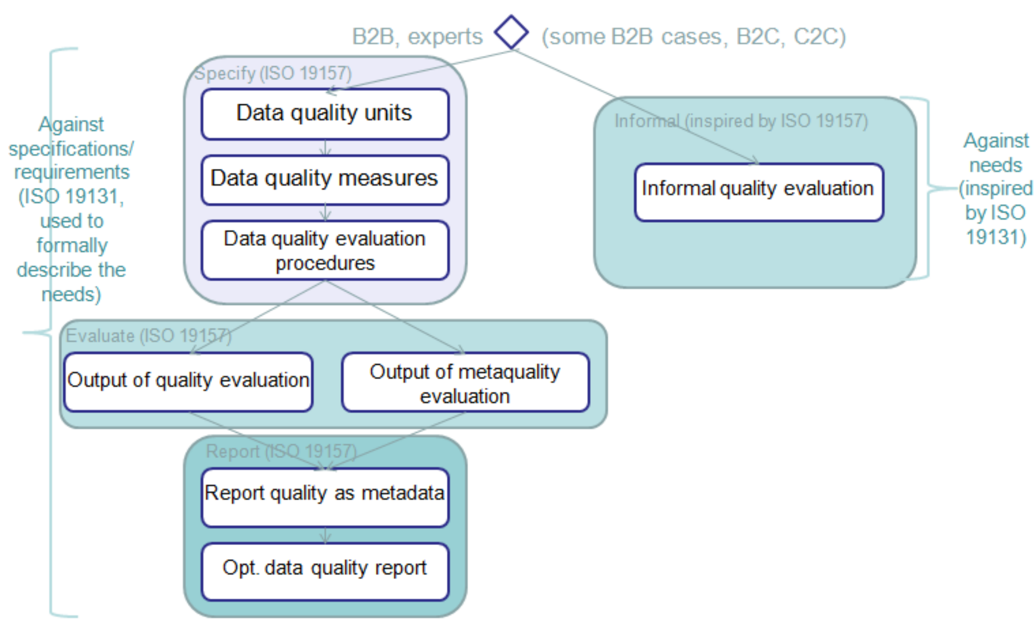

Figure 7. Formal and informal geospatial data quality evaluation processes.

The ISO 19157 standard (International Organization for Standardization, 2013) defines a general procedure for evaluating the quality of geospatial data and principles for reporting the evaluated data quality. The proposed evaluation process is comprised of the following three phases: specification, evaluation and reporting.

3.2.1 Formal geospatial data quality specification

Step 1 of the specification phase is the specification of the data quality units. Each data quality unit is composed of one scope (a scope specifies the extent, spatial and/or temporal, and/or common characteristic(s) that identify the data on which data quality is to be evaluated), and one or more data quality elements. ISO 19157 data quality elements are presented in figure 8.

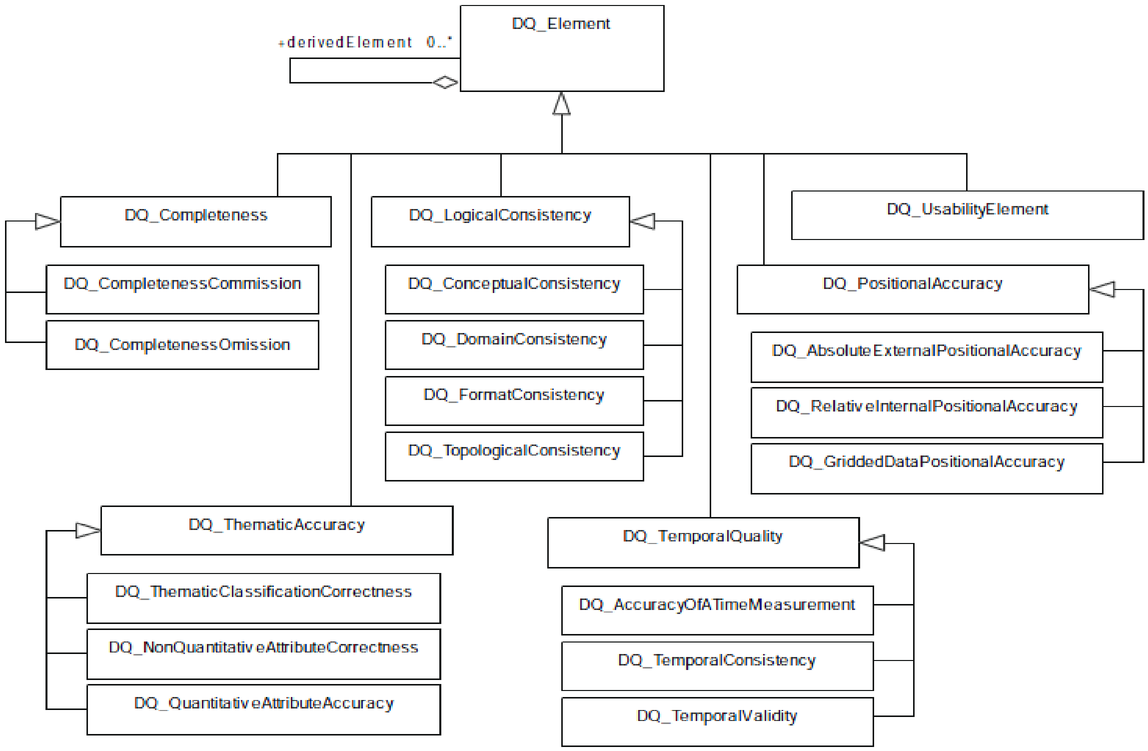

Figure 8. ISO 19157 data quality elements (from (International Organization for Standardization, 2013)).

Step 2 of the specification phase is the specification of the data quality measures used to compare the characteristics of the dataset against the formal specifications (based on ISO 19131 Geographic Information – Data product specifications). An example of data quality measure, for the "completeness" data quality element, would be the number of excess items. A list of standard measures is provided in ISO 19157. New measures can also be created.

Step 3 of the specification phase is the specification of the data quality evaluation procedures to be used to find the value of the identified data quality measures. A data quality evaluation procedure is accomplished through the application of one or more data quality evaluation methods. Data quality evaluation methods can be divided into two main classes: direct and indirect. Direct evaluation methods determine data quality through the comparison of the data with internal and/or external reference information. Indirect evaluation methods infer or estimate data quality using information on the data such as lineage. Examples are given in section 4.2.1.

3.2.2 Formal geospatial data quality evaluation

Step 1 of the evaluation phase is the determination of the output of the data quality evaluation. At least one data quality result must be provided for each data quality element. The result of the evaluation can be a quantitative result, a conformance result, a descriptive result or a coverage result (examples are found in section 4.2.2).

Step 2 of the evaluation phase is the evaluation of the metaquality. Metaquality is defined as the information describing the quality of data quality (International Organization for Standardization, 2013). Metaquality elements (confidence, representativity and homogeneity) are a set of quantitative and qualitative statements about a quality evaluation and its result. The knowledge about the quality of a given result is often as important as the result itself.

3.2.3 Formal geospatial data quality reporting

According to ISO 19157, step 1 of the reporting phase is the reporting of data quality results as appropriate metadata in compliance with ISO 19115-1:2014 Geographic information - Metadata - Part 1: Fundamentals (International Organization for Standardization, 2014), and ISO 19115-2:2009 Geographic information - Metadata - Part 2: Extensions for imagery and gridded data (International Organization for Standardization , 2009).

Step 2 of the reporting phase is the creation of a data quality report (optional). In order to provide details or nuances in an easier to understand format than the metadata, a standalone quality report may be created. However, in ISO 19157, the standalone quality report is not meant to replace the metadata. The metadata should provide a reference to the standalone quality report when it exists.

Various methods with different levels of sophistication are used to represent quality. Examples include the use of radar diagrams to compare data quality elements of a dataset to the needed level of quality, quality tables, text-only descriptions, etc. These adapted methods to communicate simply to different types of geospatial data users could be further developed and made available on SDI data access portals, for example. Such methods use a language understandable by a larger base of users. Detailed examples are presented in section 4.2.4.

3.2.4 Informal geospatial data quality evaluation

Less formal processes to evaluate and report data quality can be found in industry in B2B and B2C contexts depending upon the context defined in the first phase of risk analysis. Nevertheless, it is not surprising to see their procedures and content being similar to the ISO 19157 approach since this standard follows a highly rational logic and is flexible with regards to the level of details. It is quite possible that no metadata is provided and all information is transmitted in a data quality report.

In the B2C and C2C contexts, users of geospatial data can use various means to evaluate how a geospatial dataset fulfills their needs (i.e. external quality). Unknowingly, they typically follow the same rationale as ISO 19157 but in a less rigorous and more superficial manner. They can define their measurement methods to fit with their quality representation methods (e.g., 5 axes quality radar diagram with measurement units going from "no" to "completely" in 3 steps). In other words, they can use the more formal quality representation methods with more qualitative and less rigorous measures of quality. In the case of crowdsourced quality information, VQI systems can be implemented with the popular 5-star rating system (with average note, number of votes per star, comments, filtering per date, region or dataset version, etc.).

3.3 Risk management

Risk is "the effect of uncertainty on objectives" (Canada, 2012), (International Organization for Standardization, 2009). The effect is any deviation from the expected situation (positive and/or negative). The uncertainty is related to the deficiency of information. The objectives can be any type of goals: economic, health-related, environmental, etc. Risk is often expressed as a combination of the positive or negative consequences of an event and their likelihood of occurrence. Risk is about the effect of uncertainty, and is therefore future-oriented. In the context of geospatial data, the risk considered is the risk of inappropriate usage of geospatial data.

Risk management is "the act or practice of dealing with risk (…)" (Kerzner, 2009). Risk management is the activity of directing and controlling what an organization does to minimize unexpected impacts on its objectives. A key principle in risk management is that zero risk does not exist. Risk management implies balancing the efforts to prevent unexpected outcomes with potential negative impacts of such outcomes for stakeholders (including consumers), as shown in figure 9.

Figure 9. Risk management: a balance between efforts and unexpected negative impacts.

The goal of risk management is to respond proactively to change by mitigating the threats and capitalizing on the opportunities that uncertainty presents to an organization's or an individual's objectives (Canada, 2012). From a legal perspective, using a risk management approach is necessary to better protect both the geospatial data producer and user.

3.3.1 Risk management process

Risk management is a systematic, continuous and iterative process that successively aims to identify and assess risks in a given situation, to develop strategies to master them, and to follow, document, and communicate about them. Figure 10 presents the risk management framework described in the ISO 31000 – Risk management – Principles and guidelines standard (International Organization for Standardization, 2009).

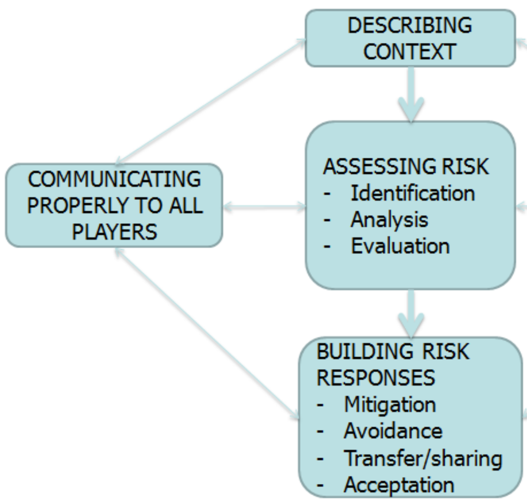

Figure 10. The steps of the ISO 31000 risk management process.

The general steps of the ISO 31000 risk management process are:

- Describing the organizational and risk management context
- Assessing risks
- Building risk responses
- Communicating risks
- Monitoring and reviewing risks

Describing the risk management context

Before starting the design and implementation of the process for managing risk, it is important to evaluate and understand the context (general, external, internal, and specific) of the organization, since this can significantly influence the design of the process (International Organization for Standardization, 2009).

The description of the general context usually includes the objectives of the organization (e.g., security-related, health-related, economic, …) and of the targeted scope of the risk management (e.g., one project, the whole business of the organization, a given community, …).

The description of the external context may include elements such as:

- The position of the organization within local, regional, national, international contexts
- Legal and regulatory requirements
- Stakeholders' perceptions
- Micro and macro economy
- Social and political environment
- Competition, trends, etc.

The description of the internal context may include:

- Organisational culture, governance, standards, structure and strategy
- Commitments, contractual relationships
- SWOT analysis (Strengths, Weaknesses, Opportunities, Threats)

The specific context of the targeted scope may include:

- Risk management objectives for targeted scope
- Resources/time required, project management constraints
- Depth, breath, inclusions, exclusions, responsibilities
- Methodologies
- Risk criteria, measures, tolerance levels, decisions to make

Assessing risks

Once the context is defined, the risk assessment process can start. Risk assessment is comprised of three main operations: risk identification, risk analysis, and risk evaluation.

Risk identification consists of building a comprehensive list of risks that might impact (in a good or bad way) the achievement of objectives. The list will include the reasons why objectives could potentially not be reached. For each risk identified, the following elements are also determined:

- The sources of risk, under control or not, known, unknown or emerging (e.g., economic, social, political, natural, markets, technological, operational, human, legal, etc.)
- The impacts and cumulative effects
- The possible scenarios

Risk analysis consists of understanding the nature of identified risks by determining their causes, sources, consequences (tangible or intangible) on objectives, likelihood to happen and interdependence. Risk analysis can:

- Be based on historical data, extrapolation, or prediction.
- Consider existing controls
- Be qualitative or quantitative
- Be undertaken at various levels of detail

Once the consequences and likelihood to happen of each risk are understood, the level of risk can be defined by combining these two parameters. The level of confidence with regards to the level of risk is also determined.

Risk evaluation is the comparison between the level of risk obtained during risk analysis and the risk criteria established when defining the context. This comparison helps decision-makers to select strategies for risk treatments and their prioritization:

- Risk cannot be tolerated, treatment is essential
- Risk can be tolerated, needs to be monitored
- Risk is negligible, to be observed

Risk evaluation must consider legal, regulatory and other requirements. People with appropriate knowledge should be involved in the risk identification operation (e.g., experts, users, support service specialists, consultants …).

The complementary IEC 31010: 2009 Risk management - Risk assessment techniques standard provides guidance on risk assessment techniques.

Building risk responses (risk treatment)

Risk treatment implies selecting and implementing one or a combination of strategies to modify a risk in order to reach accepted levels of tolerance. Four categories of strategies can be used:

- Mitigation: actions to eliminate or reduce consequences or their likelihood of occurring
- Avoidance: eliminate activity to eliminate risk
- Transfer/sharing: shift impact to another entity entirely or in part
- Acceptance: voluntarily accept and take risk (ignoring a risk is equivalent to informally accepting the risk)

Several alternatives exist to treat any given risk; they differ in cost, delays and efficiency.

This step must balance the efforts and the benefits for all stakeholders with regards to the objectives set forth in the context. Detailed examples of treatment strategies for risks related to the inappropriate use of geospatial data are presented in section 4.3.3.

Communicating risks

Risk communication must address the issues about the contexts and interest of internal and external stakeholders, the causes and origins of identified risks, their good and bad consequences, their level, the risk criteria and levels of tolerance, the treatments that already exist or that will be implemented, their monitoring and review, etc.

Examples of communication strategies are presented in section 4.3.4 and include embedding risk-related information within data quality metadata and reports, user manuals, guarantees, forums, etc. From a legal point of view, risk communication should be in a language the intended users will understand.

Monitoring and reviewing risks

Risk monitoring is the regular surveillance of the risks and the success of their treatment. This step also aims at:

- Detecting changes in contexts
- Detecting emerging risks
- Seeking continuous improvement

4. Putting Theory into Practice

This section will discuss geospatial data quality management and the management of risks of inappropriate use of geospatial data in more details and give example of each step for both processes.

4.1 Managing geospatial data quality in practice

From a producer point of view (B2B or B2C contexts), geospatial data quality and risks of inappropriate usage must be managed hand-in-hand at each phase of a data product life-cycle.

4.1.1 Design phase

When designing a vector-based data product, the ISO 19109:2005 Geographic information – Rules for application schema standard can be used to properly define the model (called *application schema*) of the data product to be built/updated (International Organization for Standardization, 2005). This standard covers:

- The conceptual modeling of features and their properties from a universe of discourse
- The definition of application schemas
- The use of the conceptual schema language for application schemas
- The transition from the concepts in the conceptual model to the data types in the application schema
- The integration of standardized schemas from other ISO geographic information standards with the application schema

From a geospatial data quality point of view, following a formal method for the design of the application schema will help improve the logical consistency of the data (DQ_LogicalConsistency data quality element of ISO 19157).

From a risk management point of view, the application schema can be enriched with risk management elements such as warnings. Figure 11 presents a road network application schema (extract from Levesque, Bédard, Gervais, & Devillers, 2007) enriched with warning symbols (based on the ISO/TC 145/ISO 3864-2 Graphical symbols - Safety colours and safety signs - Part

2: Design principles for product safety labels (International Organization for Standardization, 2004)) in order to highlight potential usage problems, likelihood to happen, and impact.

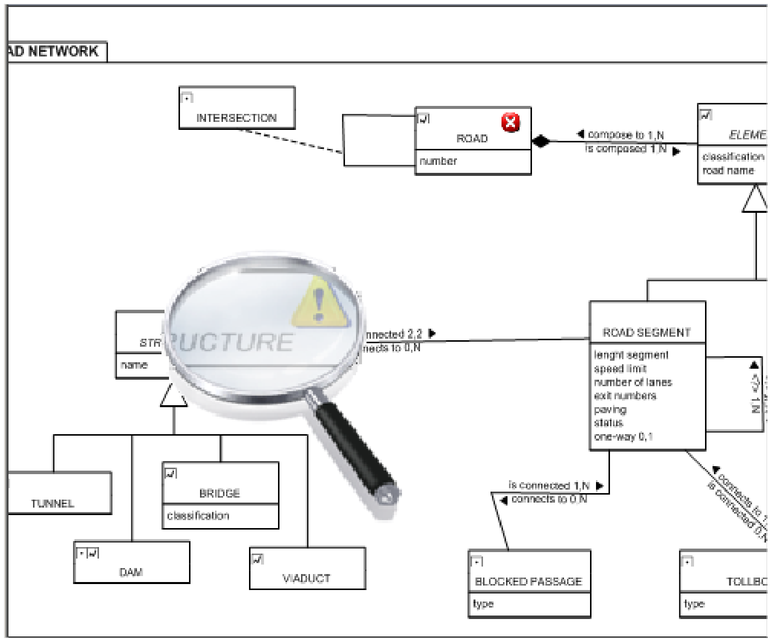

Figure 11. Road network application schema enriched with warning symbols for risk management.

When designing a vector-based data product, the ISO 19110:2005 Geographic information – Methodology for feature cataloguing standard can be used to properly define the data dictionary (called *feature catalog*) of the data product to be built/updated (International Organization for Standardization, 2005). This standard describes:

- A methodology for cataloguing feature types
- How the classification of feature types is organized into a feature catalogue and presented to the users of a set of geographic data

From a geospatial data quality point of view, following a formal method for the recording of features in a feature catalog will help improve the logical consistency of the data (DQ_LogicalConsistency data quality element of ISO 19157).

From a risk management point of view, the feature catalog can be enriched with risk management elements such as detailed descriptions of the risk, its likelihood to happen, its effects, recommendations for actions and warnings found in the corresponding application schema. Figure 12 presents an example of a feature catalog enrich with risk management information (Bédard, Chandler, Devillers, & Gervais, 2009). Figure 13 presents an extract of a CanVec+ Feature Catalog (Natural Resources Canada, 2014) enriched with a warnings section for describing in details the warning symbols used in the related application schema.

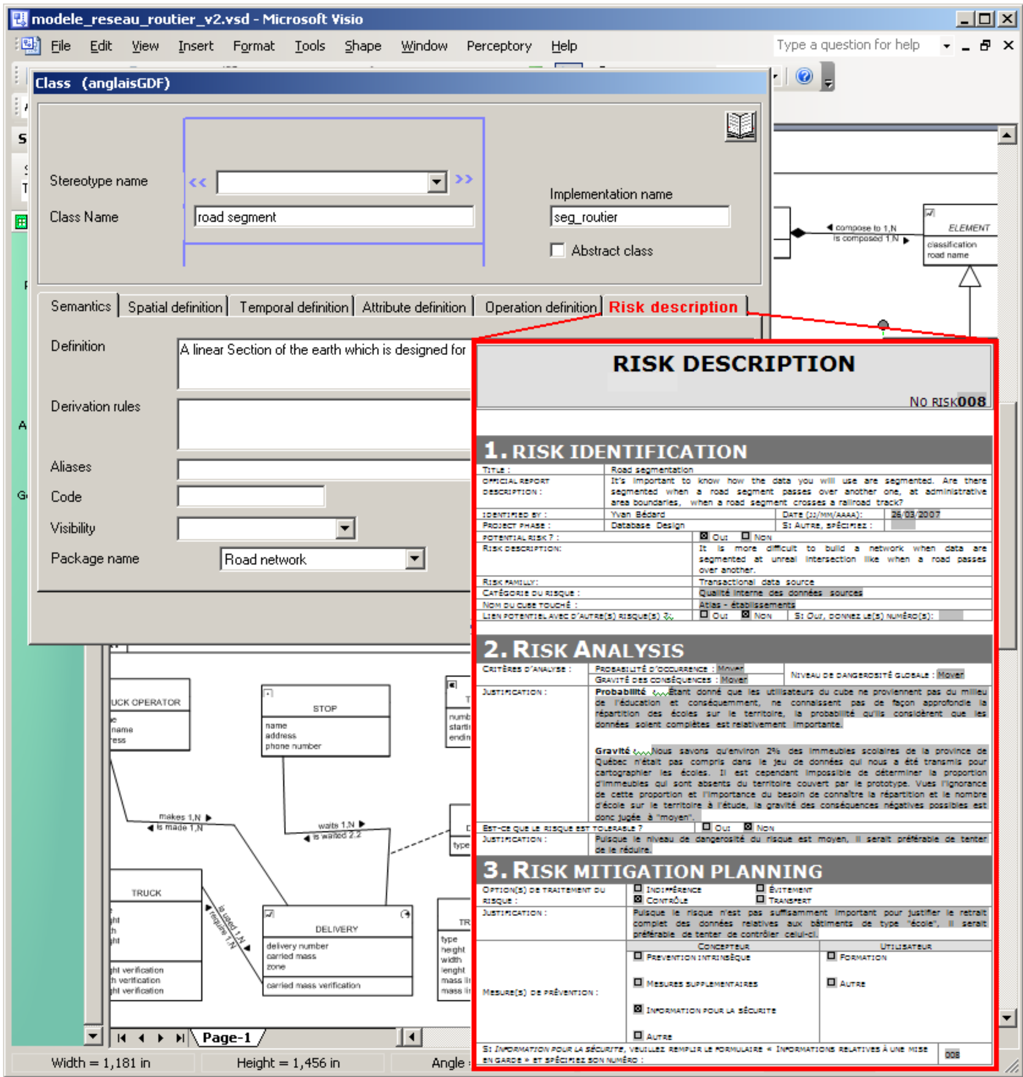

Figure 12. Example of a feature catalog enriched with risk management information (from (Bédard, Chandler, Devillers, & Gervais, 2009))

Entity

Name - *(French name)*	**Amusement park** - *(Parc d'amusement)*				
Definition	An area where entertainment is provided by game concessions, rides, etc.				
	Theme	Generic code	Is abstract	GeoBase	Geometry
	Lieux d'intérêts	2260009	No	No	Area

Attributes

Name	**Feature identifier**
Name	**Planimetric accuracy**
Name	**Specification code**
Name	**Theme**
Name	**Validity date**

Attribute values combinations (Specification codes)

Attributes combination	Point	Line	Area
Amusement park - *(Amusement park)*			2260012

Minimum sizes

Specification Code	Area (Square meter)	Lateral distance (Meter)	Length (Meter)	Longitudinal distance (Meter)	Right angle tolerance (Degree)	Spike angle tolerance (Degree)
2260012	500	1.5	---	3	---	10

Warnings

Figure 13. Extract of the CanVec+ feature catalog (from (Natural Resources Canada, 2014)).

When designing a raster-based data product, the ISO 19123:2005 Geographic information – Schema for coverage geometry and functions standard can be used to properly define the model (called *schema for coverage geometry*) of the data product to be built/updated (International Organization for Standardization, 2005). This standard describes:

- The conceptual schema for the spatial characteristics of coverages
- The relationship between the domain of a coverage and an associated attribute range

From a geospatial data quality point of view, following a formal method for the design of the schema for coverage geometry and functions will help improve the logical consistency of the data (DQ_LogicalConsistency data quality element of ISO 19157). Figure 14 presents an example schema for digital orthoimagery (from (Maitra, 2004)).

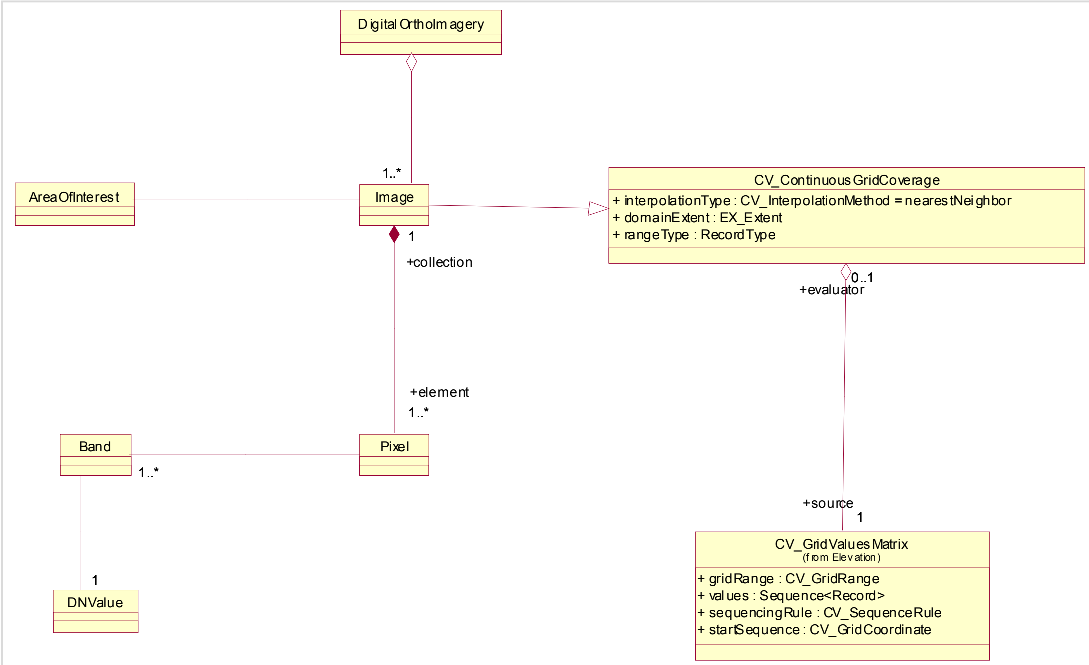

Figure 14. Example schema for digital orthoimagery (from (Maitra, 2004)).

From a risk management point of view, the application schema can be enriched with risk management elements such as risk identification, likelihood to happen, impact, remedies and suggested warnings.

The ISO 19131:2007 Geographic information – Data product specifications (with ISO 19131:2007/Amd 1:2011 Requirements relating to the inclusion of an application schema and feature catalogue and the treatment of coverages in an application schema) standard can be used to define the detailed specifications of the data product to be built/updated. This standard:

- Defines requirements for the specification of geographic data products, based upon the concepts of other ISO 19100 International Standards
- References application schema, feature catalog or schema for coverage geometry

From a geospatial data quality point of view, formally describing the specifications of a data product to be built/updated will help improve the logical consistency of the data (DQ_LogicalConsistency data quality element of ISO 19157). Data product specifications also help in the whole data quality evaluation process as they specify the expected values for the data quality elements (DQ_Element) of ISO 19157, that is the quality criteria. Figure 15 presents an extract of the CanVec+ Data Product Specifications showing the expected value for the DQ_CompletenessCommission data quality element (Natural Resources Canada, 2014).

6 DATA QUALITY

6.1 COMPLETENESS

6.1.1 Commission

The data quality assessment (including completeness and thematic accuracy) is directly performed during data production process. The validation method applied depends on the data source used. CanVec+ data come from 2 different sources: GeoBase initiative (ex.: National Road Network (NRN) data) and Natural Resources Canada digital topographic data.

For the GeoBase initiative data, data quality is ensured by the producer (and partner). The validation mechanism used may vary from one partner to another.

For Natural Resources Canada data, during data inspection, datasets produced are grouped into distinct batches. A few datasets are selected from each batch and inspected in order to check their contents and compare them to a data source used in production or to another independent source. If the percentage of error detected is less than 5%, then all datasets in the batch are normally considered acceptable.

Figure 15. Extract of the CanVec+ data product specifications (from (Natural Resources Canada, 2014)).

From a risk management point of view, product specifications standards don't explicitly include suggestions for risk management metadata (risk identification, likelihood, impacts, remedies and suggested warnings) but one can add these into the specifications.

4.1.2 Implementation phase

During the implementation phase of a data product, additional integrity constraints are implemented (i.e. in addition to those already defined in the application schema and feature catalogue) to better control the consistency of data. Integrity constraints can be spatial, temporal, or descriptive. They can reinforce application schema and feature catalogues, but at this phase they are mostly oriented towards insuring the quality of the physical structure of the data and to control potential data input errors. Different levels of integrity constraints can be defined:

- Intra-field (e.g., values of a numeric field must be between 0 and 1)
- Inter-fields (e.g., if the value of the road classification attribute is “national”, then the value of the maximum speed attribute cannot be null)
- Intra-feature (e.g., the date of an updated house assessment cannot be lower than the date of the older house assessment)
- Inter-feature (e.g., the size of a “building” must be smaller than the size of the “parcel” it is built on)
- Intra-feature class (e.g., “building” cannot intersect “building”)
- Inter-feature classes (e.g., “road” cannot cross “lake”)
- Intra-theme (e.g., “river” can connect “canal”)
- Inter-theme (e.g., “dam” can share geometry with “road”)

From a geospatial data quality point of view, the use of integrity constraints may help in controlling all data quality elements (DQ_Element) of ISO 19157. Figure 16 presents an example of a constraint repository (adapted from (Normand, 1999)). From a risk management point of view, integrity constraints may help to reduce the impacts of faulty data when applied to more sensitive features or attributes (e.g., parcel boundaries, property value, and zoning).

Constraints			
Intra-theme			
	Dam(L)	Connects	Dam (L)
	Dam(L)	Connects	Canal (L)
	Dam(L)	Connects	Stream (L)
	Dam(L)	Connects	Dyke (L)
	Dam(L)	Connects	Lake (S)
Inter-theme			
	Dam(L)	Share	Railroad (L)
	Dam(L)	Share	Road (L)
	Dam(L)	Share	Limited access road (L)
	Dam(L)	Share	Path (L)

Figure

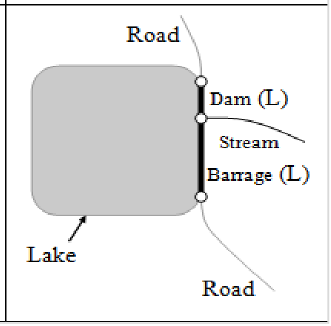

Figure 16. Example of a constraint repository (adapted from (Normand, 1999)).

4.1.3 Production phase

During the production phase, quality evaluation can be used to control the results of every data acquisition and processing step, or to control the final product. This allows one to check if the evaluated quality meets the quality criteria defined in the design phase or not.

The ISO 19157: 2013 Geographic information – Data quality standard can be used to support the detailed quality evaluation (International Organization for Standardization, 2013). This standard explains:

- The components for describing data quality
- The components and content structure of a register for data quality measures
- The general procedures for evaluating the quality of geographic data
- The principles for reporting data quality

Detailed steps of geospatial data quality evaluation are presented in section 4.2, along with examples.

The ISO 19115-1: 2014 Geographic information - Metadata - Part 1: Fundamentals and ISO 19115-2:2009 Geographic information - Metadata - Part 2: Extensions for imagery and gridded data standards can be used to report information about the data contained in the data product (i.e. the metadata) (International Organization for Standardization, 2014), (International Organization for Standardization , 2009). This standard contains:

- Mandatory and conditional metadata sections, metadata entities, and metadata elements
- The minimum set of metadata required to serve most metadata applications (data discovery, determining data fitness for use, data access, data transfer, and use of digital data and services)
- Optional metadata elements to allow for a more extensive standard description of resources, if required
- A method for extending metadata to fit specialized needs

The North American Profile of ISO19115:2003 is used in the Canadian context to describe metadata.

Metadata data can be encoded in XML using the ISO 19139:2007 Geographic Information – Metadata – XML schema implementation.

From a geospatial data quality point of view, all data quality elements (DQ_Elements) of ISO 19157 are reported using metadata. Figure 17 presents an extract of the CanVec+ 082C metadata dataset, based on the North American Profile of ISO19115, and showing the information about the DQ_CompletenessCommission data quality element (from (Natural Resources Canada, 2015). One can also produce an optional quality report. For less formal contexts, in place of, or in addition to metadata or quality reports, other means may be more appropriate to show data quality information as explained in the next section.


```
- <gmd:DQ_CompletenessCommission>
  - <gmd:result>
    - <gmd:DQ_ConformanceResult>
        <gmd:specification xlink:title="CanVec+ - Data Product Specifications, Edition 1.1"
          xlink:href="http://ftp2.cits.rncan.gc.ca/pub/canvec+/doc/CanVec+_product_specifications.pdf"/>
      - <gmd:explanation xsi:type="gmd:PT_FreeText_PropertyType">
          <gco:CharacterString>The data quality assessment (including completeness and thematic accuracy) is directly performed during the
            data production process. The validation method applied depends on the data source used. CanVec+ data come from 2 different
            sources: GeoBase initiative (ex: National Road Network (NRN) data) and Natural Resources Canada digital topographic data. For
            GeoBase initiative data, data quality is ensured by the producer (and partner). The validation mechanism used may vary from
            one partner to another. For Natural Resources Canada data, during data inspection, datasets are produced into distinct batches.
            A few datasets are selected from each batch and inspected in order to check their contents and compare them to a data source
            used in production or another independent source. If the percentage of error detected is less than 5 %, then all datasets in the
            batch are normally considered acceptable.</gco:CharacterString>
        - <gmd:PT_FreeText>
          - <gmd:textGroup>
              <gmd:LocalisedCharacterString locale="#FR">L'évaluation de la qualité des données (incluant la complétude et l'exactitude
                thématique) est faite directement lors de la production des données. La méthodologie de vérification dépend de la source
                de données. Les données du produit CanVec+ proviennent de 2 sources différentes : de l'initiative GéoBase (ex. : les
                données du Réseau routier national RRN)) et de la production de données topographiques numériques de Ressources
                naturelles Canada. Pour les données provenant de l'initiative GéoBase, la qualité des données est assurée par le
                partenaire producteur de données. Le mécanisme de vérification peut être différent pour chaque partenaire. Pour les
                données provenant de Ressources naturelles Canada, lors de l'inspection des données, les jeux de données issus de la
                production sont regroupés en différents lots. Quelques jeux de données du lot sont sélectionnés et inspectés afin de
                vérifier le contenu par rapport à une source de données ayant servi à la production ou par rapport à une source
                indépendante. Si le pourcentage d'erreurs détecté est inférieur à 5%, alors tous les jeux de données du lot sont
                normalement considérés acceptables.</gmd:LocalisedCharacterString>
            </gmd:textGroup>
          </gmd:PT_FreeText>
        </gmd:explanation>
      - <gmd:pass>
          <gco:Boolean>true</gco:Boolean>
        </gmd:pass>
```


Figure 17. Extract of metadata from the CanVec+ 082C data product (from (Natural Resources Canada, 2015)).

4.1.4 Delivery phase

At the delivery phase, along with the data product, appropriate documentation product must also be delivered to the user. Proper communication typically combines several information products/services. In the B2B context, this is quality metadata with optional but recommended quality reports (potentially one per type of usage/user).

In the B2C and C2C contexts, it is strongly recommended to deliver a user manual with the following content suggested by (Gervais, 2004) based on legal considerations:

- License
- Guarantees
- Installation
- Product description
- Resolution (spatial, temporal, descriptive) of the data
- General advice
- Functional specifications
- Recommended uses
- Non-recommended uses
- Warnings and safety
- Troubleshooting
- Technical specifications

ISO 19115 metadata already contains part of this information, but in a technical jargon usually unintelligible for most users. See (Gervais, 2004) for the correspondence between elements of the user manual and quality metadata).

From a geospatial data quality point of view, quality metadata is mandatory and quality reports are optional. In every case, the value of all data quality elements of ISO 19157 (DQ_Elements) will influence the product documentation.

From a risk management point of view, the content of the delivered documentation products is a direct result of the risk management strategies adopted. Typically, it will include a combination of risk mitigation strategies and communication products (see sections 4.3.3 and 4.4.4 for details). In all cases, a good user manual is recommended as in every mature mass market. Such user guide can have different versions tailored for each type of usage (e.g., the "User Guide for the Survey of Household Spending 2012" (Statistics Canada, 2012)).

Figure 18 presents example sections of a geospatial data user manual (from (Gervais, 2004)).

Guarantee	*The manufacturer guarantees that the topology of the product dataset allows for an appropriate use of the product functions.*

Warnings and safety	Warnings related to data: *Data included in the product are updated every month. It is highly recommended that you proceed to these data updates in order to keep the product operating efficiently and safely.*

Troubleshooting	Customer service: *For any question about the use of this product, please contact the Customer Service, free, at 1-800-***-****.*

Figure 18. Sample sections of a geospatial data user manual (from (Gervais, 2004)).

Warnings can be included in the delivered documentation to facilitate reading and increase users' awareness. ISO/TC 145/ISO 3864-2 Graphical symbols (International Organization for Standardization, 2004) establishes the principles for the preparation, coordination and application of graphical symbols. Symbols and labels are powerful ways to convey the meaning of risk:

- Type of risk (danger or positive action)
- Level of risk
- Description of risk
- Actions to take in face of consequences

Figure 19 presents examples of symbols according to the level of danger (American National Standards Institute, 2006).

Figure 19. Examples symbols that can be used in warnings.

Figure 20 presents an example use of symbols to facilitate the reading of a quality report (from (Gervais, Bédard, & Larrivée, 2007)).

	Symbol	Description
	Warning	Symbol meaning that there exists a problem with the potential use of a dataset
	Prohibition	Symbol meaning that the dataset cannot be used to satisfy one or more needs identified by the user
	Duty	Symbol meaning that specific actions should be taken before using the dataset

Figure 20. Example use of symbols to facilitate the reading of a data quality report.

4.1.5 Usage phase

For the usage phase of a data product, several risk management strategies can be put into place to help users with questions about data quality and proper usage of the data. In many cases, it will be a continuation of means put into place for a quality-aware delivery of the data. Examples include a free 1-800 phone line, live chat, web-based user forum, regular webinars, email list to inform users of new recommended/forbidden usages or quality updates, etc. This can go up to the mandatory or voluntary certification of users.

More interaction can be put into place with VQI (Volunteered Quality Information) based on users voluntarily reporting quality issues (ex. TomTom MapShare) or using the 5-star/comment system regularly found on the web (cf. Amazon, Apple Store). Ideally, a quality-aware application fine-tuned for different user profiles would result from the previous actions. It is an application where, based on the user interaction with the application, warnings are displayed to inform the user about the potential risks on inappropriate usage of geospatial data.

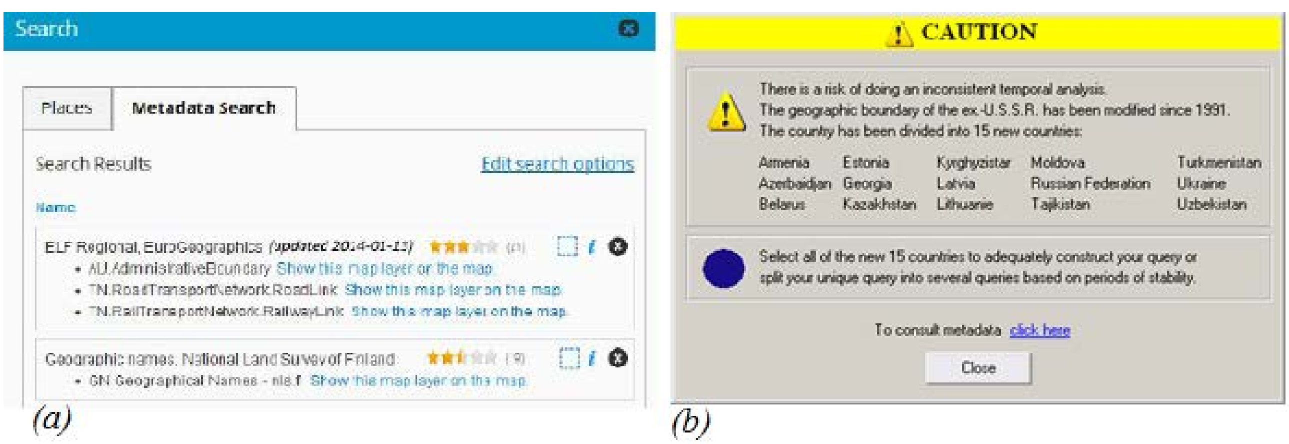

Figure 21. Example of *(a)* a VQI using a 5-star rating system (from (Koistinen, 2015)), and *(b)* context-sensitive warning for a worldwide spatio-temporal query covering a period starting in 1990 where the 3 ISO-

recommended elements are included: level of risk, nature of problem and action to solve the problem (from (Gervais, Bédard, Lévesque, Bernier, & Devillers, 2009)).

4.2 Evaluating geospatial data quality in practice

The steps of the geospatial data quality evaluation process have been presented in Figure 7. Steps for the formal process are based on ISO 19157 (International Organization for Standardization, 2013):

- Specification
  - Specifying data quality units
  - Specifying data quality measures
  - Specifying data quality evaluation procedures
- Evaluation
  - Output of data quality evaluation
  - Output of metaquality evaluation
- Reporting
  - Reporting data quality as metadata
  - Optionally providing a data quality report

4.2.1 Formal geospatial data quality specification

Specifying data quality units

A data quality unit is composed of one scope (MD_Scope) and one to many data quality elements (DQ_Element, see figure 8). A scope specifies the extent, spatial and/or temporal, and/or common characteristic(s) that identify the data on which data quality is to be evaluated.

Example quality unit 1:

- MD_Scope: dataset 082C
- DQ_Elements: DQ_LogicalConsistency, DQ_Completeness

Example quality unit 2:

- MD_Scope: feature type (hydrant)
- DQ_Element: DQ_QuantitativeAttributeAccuracy

Specifying data quality measures

A data quality element should refer to a measure, by means of a measure reference (DQ_MeasureReference). The measure reference is comprised of the following elements:

- Measure identification
- Name of measure
- Measure description

A list of standard measures is provided in ISO 19157. New measures can be created.

Examples measures for the DQ_CompletenessCommission data quality element are:

- Excess item
- Number of excess items
- Rate of excess items
- Number of duplicate feature instances

Example measures for the DQ_CompletenessOmission data quality element are:

- Missing item
- Number of missing items
- Rate of missing items

Specifying data quality evaluation procedures

Data quality measure values are evaluated using evaluation methods. A set of evaluation methods compose an evaluation procedure. Data quality evaluation methods (DQ_EvaluationMethod) can be divided into two main classes: direct and indirect. Direct evaluation methods imply a comparison of the data with internal and/or external reference information. Direct evaluation methods can be based on full inspection (i.e. every item of the population is inspected) of the related elements (DQ_FullInspection), or based on sampling (DQ_SampleBasedInspection). Indirect evaluation methods infer or estimate data quality using information on the data such as lineage (DQ_IndirectEvaluation).

Example evaluation methods based on sampling include:

- Feature-guided sampling (non-spatial sampling): based on the non-spatial attributes of features
- Area-guided regular (non-random) sampling
- Area-guided random sampling

Figure 22 presents an example of an area-guided regular sampling method.

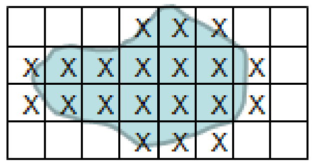

Figure 22. Example of an area-guided regular (non-random) sampling method.

4.2.2 Formal geospatial data quality evaluation

Evaluating the quality of geospatial data

At least one data quality result must be provided for each data quality element. The result of the evaluation can be a quantitative result (DQ_QuantitativeResult), a conformance result (DQ_ConformanceResult), a descriptive result (DQ_DescriptiveResult) or a coverage result.

An example of DQ_QuantitativeResult:

- (DQ_CompletenessCommission), Number of excess items: 3

An example of DQ_ConformanceResult:

- (DQ_CompletenessCommission), Number of excess items: pass

An example of DQ_DescriptiveResult:

- (DQ_LogicalConsistency), Conceptual schema compliance: “The rules of the CanVec+ conceptual schema are all recorded and validated in the source database containing the CanVec+ product. This approach ensures the conceptual consistency between the conceptual schema and the CanVec+ product.” (from CanVec+ 082C metadata dataset (Natural Resources Canada, 2015))

A coverage result is the result of a data quality evaluation, organized as a coverage. This is documented in ISO 19115-2:2009 (International Organization for Standardization , 2009).

Evaluating the metaquality of a quality evaluation

Metaquality elements are a set of quantitative and qualitative statements about a quality evaluation and its result. Metaquality can be expressed using:

- Confidence (DQ_Confidence): trustworthiness of a data quality result

- Representativity (DQ_Representativity): degree to which the sample used has produced a result which is representative of the data within the data quality scope
- Homogeneity (DQ_Homogeneity): expected or tested uniformity of the results obtained for a data quality evaluation

An example of DQ_Confidence:

- Standard deviation or a confidence interval on a given confidence level

An example of DQ_Representativity:

- All the geographic zones and concerned time periods are covered and the population is sufficiently large

An example of DQ_Homogeneity:

- Comparison of the evaluation results of several segments of a global data set expressed using root mean square errors

4.2.3 Formal geospatial data quality reporting

Reporting quality as metadata

According to ISO 19157, data quality is reported as metadata in compliance with ISO 19115-1:2014 Geographic information - Metadata - Part 1: Fundamentals (International Organization for Standardization, 2014), and ISO 19115-2:2009 Geographic information - Metadata - Part 2: Extensions for imagery and gridded data (International Organization for Standardization , 2009).

Figure 23 presents an example of the results of the evaluation of the DQ_CompletenessCommission quality element.

```xml
- <gmd:DQ_CompletenessCommission>
  - <gmd:result>
    - <gmd:DQ_ConformanceResult>
        <gmd:specification xlink:title="CanVec+ - Data Product Specifications, Edition 1.1"
          xlink:href="http://ftp2.cits.rncan.gc.ca/pub/canvec+/doc/CanVec+_product_specifications.pdf"/>
      - <gmd:explanation xsi:type="gmd:PT_FreeText_PropertyType">
          <gco:CharacterString>The data quality assessment (including completeness and thematic accuracy) is directly performed during the
            data production process. The validation method applied depends on the data source used. CanVec+ data come from 2 different
            sources: GeoBase initiative (ex: National Road Network (NRN) data) and Natural Resources Canada digital topographic data. For
            GeoBase initiative data, data quality is ensured by the producer (and partner). The validation mechanism used may vary from
            one partner to another. For Natural Resources Canada data, during data inspection, datasets are produced into distinct batches.
            A few datasets are selected from each batch and inspected in order to check their contents and compare them to a data source
            used in production or another independent source. If the percentage of error detected is less than 5 %, then all datasets in the
            batch are normally considered acceptable.</gco:CharacterString>
        - <gmd:PT_FreeText>
          - <gmd:textGroup>
              <gmd:LocalisedCharacterString locale="#FR">L'évaluation de la qualité des données (incluant la complétude et l'exactitude
                thématique) est faite directement lors de la production des données. La méthodologie de vérification dépend de la source
                de données. Les données du produit CanVec+ proviennent de 2 sources différentes : de l'initiative GéoBase (ex. : les
                données du Réseau routier national RRN)) et de la production de données topographiques numériques de Ressources
                naturelles Canada. Pour les données provenant de l'initiative GéoBase, la qualité des données est assurée par le
                partenaire producteur de données. Le mécanisme de vérification peut être différent pour chaque partenaire. Pour les
                données provenant de Ressources naturelles Canada, lors de l'inspection des données, les jeux de données issus de la
                production sont regroupés en différents lots. Quelques jeux de données du lot sont sélectionnés et inspectés afin de
                vérifier le contenu par rapport à une source de données ayant servi à la production ou par rapport à une source
                indépendante. Si le pourcentage d'erreurs détecté est inférieur à 5%, alors tous les jeux de données du lot sont
                normalement considérés acceptables.</gmd:LocalisedCharacterString>
            </gmd:textGroup>
          </gmd:PT_FreeText>
        </gmd:explanation>
      - <gmd:pass>
          <gco:Boolean>true</gco:Boolean>
        </gmd:pass>
```


Figure 23. Example of quality metadata (DQ_CompletenessCommission) from from CanVec+ 082C metadata dataset (Natural Resources Canada, 2015)).

Optionally reporting quality using a data quality report

In order to provide more details than reported as metadata, in an easier to understand format than the metadata, a standalone quality report may be created. The standalone quality report is used to complement the metadata. The metadata should provide a reference to the standalone quality report when it exists. If a dataset is intended to serve different categories of usage, different flavors of the quality report may be required since the needs vary, thus the external quality varies.

Figure 24 presents an extract of a data quality report presenting detailed DQ_CompletenessCommission results for each feature class of a dataset (from ISO 19157:2013 (International Organization for Standardization, 2013).

Feature class	Number of instances in the universe of discourse	Commission count	Commission percentage[a]	Omission count	Omission percentage[b]
Path	7	1	14	3	43
Road	5	2	40	0	0
Tree	25	3	12	2	8
Industrial building	4	0	0	2	50
House	10	1	10	1	10

[a] Commission percentage = number of included items/number of items in the universe of discourse × 100.

[b] Omission percentage = number of omitted items/number of items in the universe of discourse × 100.

Figure 24. Detailed results for DQ_CompletenessCommission reported in a data quality report (example from ISO 19157 (International Organization for Standardization, 2013)).

Figure 25 illustrates a representation framework based on radar diagrams to help compare data quality element (DQ_Elements) values of a dataset to the needed level of quality.

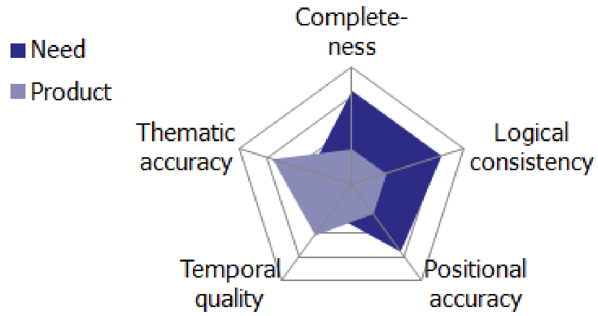

Figure 25. Use of a radar diagram to help compare data quality elements of a dataset to the required level of quality.

Figure 26 presents another example of content of a data quality report: a summarized feature class quality evaluation (adapted from a private report by (Gervais, Bédard, & Larrivée, 2007)).

Feature class	Result
Road	●
Building	■
Hydrology	●
Parking	▲
●	Generally good, but may present minor problems in relation to the identified needs
▲	Usable but may not completely fulfill the identified needs
■	The feature class is not fit to the identified needs

Figure 26. Summarized quality evaluation per feature class reported in a data quality report (example from (Gervais, Bédard, & Larrivée, 2007).

Figure 27 shows a dashboard view of a quality metadata along the 5 ISO quality elements (indicators) at the occurrence level and a global quality indicator for this feature type (road) (Devillers, Bédard, Jeansoulin, & Moulin, 2007).

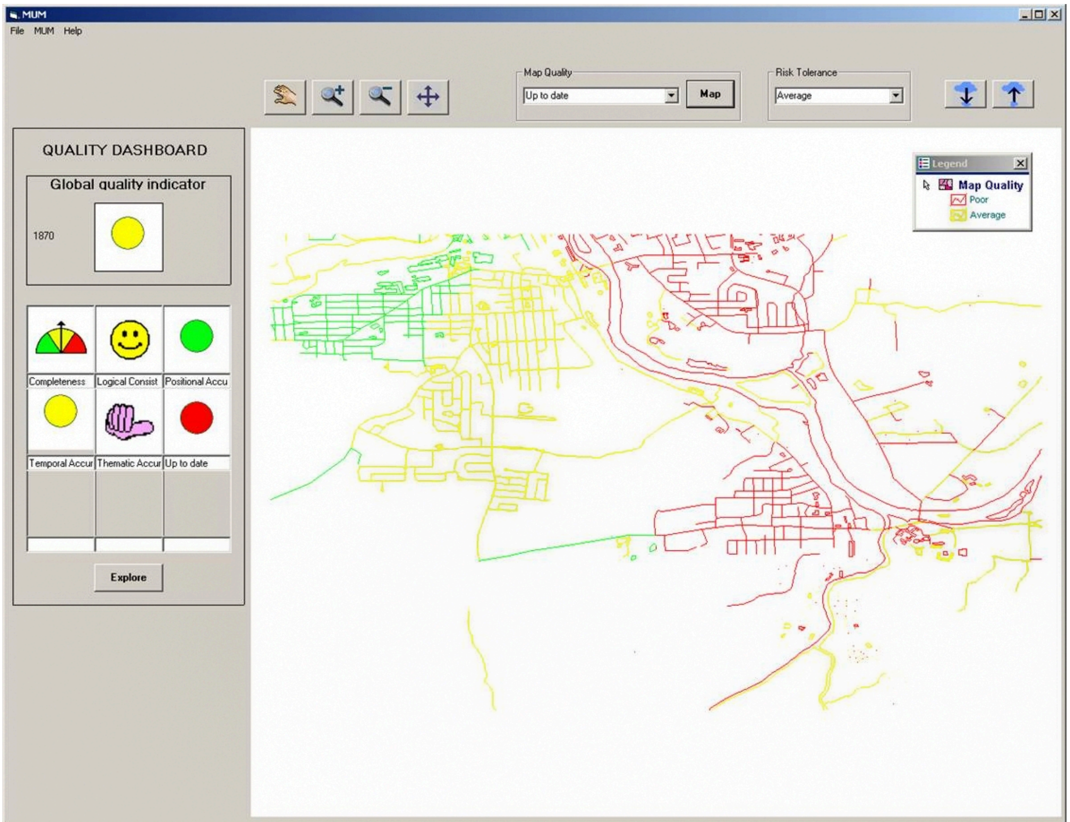

Figure 27. Dashboard view of a quality metadata (from (Devillers, Bédard, Jeansoulin, & Moulin, 2007)).

4.2.4 Informal geospatial data quality evaluation

In B2B and B2C contexts, less formal processes to evaluate and report data quality should be inspired by the ISO 19157 approach and make good use of its flexibility with regards to the level of resolution of quality analysis. In the B2B context, it is possible that no metadata is provided separately from the report, but for quality reports aimed at being reused or transmitted via a geospatial data infrastructure, the use of quality metadata is recommended. In addition to facilitating data discovery and communication/understanding between experts, metadata facilitates communication between machines. Consequently, it makes a step towards quality-aware interoperability using quality-metadata matching (i.e. context-matching as described by (Sboui & Bédard, 2012)). In the B2C context, providers of data cannot rely on metadata to inform users about data quality since it is a technical language they typically do not understand. Providers must rather use simpler methods to represent data quality such as the ones in the figures presented above or hereafter.

In the B2C and C2C contexts, users of geospatial data can use various methods to evaluate the external data quality for themselves. These methods are often centered on the data quality representation framework used and aim at gathering enough quality information to properly fill the representations. Examples of representation frameworks were given in the previous section.

Figures 28 and 29 present the use of a 5-star rating system, in a virtual globe context where the 3D representation of buildings can be rated collaboratively by users (from (Jones, 2011)), and in a SDI context, where each available dataset can be rated by users (from (Koistinen, 2015)) respectively.

Figure 28. Example of the use of a 5-star rating system to rate the representation of buildings in a virtual globe environment (from (Jones, 2011)).

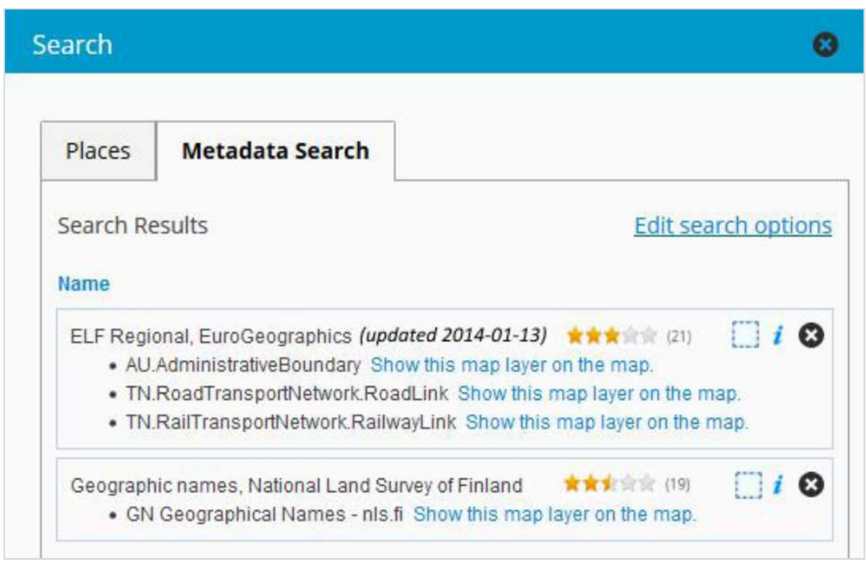

Figure 29. Example of the use of a 5-star rating system in a SDI environment (from (Koistinen, 2015)).

4.3 Managing the risks of inappropriate use of geospatial data in practice

The risk management framework described in the ISO 31000 – Risk management – Principles and guidelines standard (International Organization for Standardization, 2009) has been presented in figure 10. The general steps of the framework are:

- Describing the organizational and risk management context
- Assessing risks
- Building risk responses
- Communicating risks
- Monitoring and reviewing risks

4.3.1 Describing the risk management context

Risk management starts with a description of the context (general, external, internal, and specific) of the organization and of the targeted risk management scope. Context description helps understand how the planned activity fits into the wider organization and market/society and the organization's approach to risk management, in order to scope the risk management strategy. Figure 30 presents an example of context description.

1. Context

There are 3 other publicly available sources of data similar to ours. Their cost and overall accuracy is similar to ours and clients have difficulty to choose the one that best fit their needs. The evolution of the market is uncertain as well as the actions of the competition. By implementing a new data quality and risk management strategy, it will help to improve significantly the communication with potential clients and help them understand how our data can fulfill their geospatial needs. A potential client who knows better is reassured and has more chances of becoming client. Our objective is to increase our market share by 10% within 2 years. If we implement the strategy, the costs and risks are ... If we don't, the risks are...

Figure 30. Example of a risk management context description.

4.3.2 Assessing risks

Once the context is defined, the risk assessment process can start. Risk assessment is comprised of three main operations: risk identification, risk analysis, and risk evaluation.

Risk identification generates a comprehensive list of risks of inappropriate use of geospatial data. The risk identification step can be conducted using:

– Analysis of existing documentation (e.g., specifications, contracts, task flow charts, …)
– Interviews
– Brainstorming sessions
– Collaborative approach with users
– …

Table 2 presents an example list of potential risks of inappropriate use of geospatial data related to features/attributes in a geospatial dataset for agricultural context (extract from (Grira, 2014)).

Feature / Attribute	**Identified Risks**
Cultivated parcel	**R-1**: the user may think that the whole region delimited by the cadastral boundaries was cultivated. Some areas may not be cultivated, such as woodland, rocky button or areas near cadastral boundaries.
Floodplain	**R-2:** Floodplains are vague data. The user would think that the provided boundaries are accurate whereas they are fuzzy and large boundaries are not represented as such on the map.
Pesticide spread area	**R-3**: the areas where the pesticide is spread have large and fuzzy boundaries and uncertain location within the plot (because of the techniques and the methods of pesticide spreading). The user could think that the area is accurate whereas positional accuracy is not considered in the area calculation. Note: the uncertainty for R-3 is related to the pesticide spreading zone, i.e. its boundaries and its location. However, the uncertainty for R-1 is related to the plot (its boundaries) where the spreading zone is located.

Table 2. Example list of risks of inappropriate use of geospatial data in an agricultural context (from (Grira, 2014)).

Risk analysis consists of understanding the nature of identified risks by determining their causes, sources, consequences and likelihood to happen. The risk analysis step can be conducted using:

– Analysis of lessons learned from previous projects
– Simulation methods
– Probabilistic analysis
– …

Table 3 presents the results of the analysis of the risks presented in table 2 (extract from (Grira, 2014)).

Feature / Attribute	**Identified Risks**	**Impact of Risk**	**Probability of Occurrence**

Cultivated parcel	**R-1**	Strong overestimation of the ratio quantity of pesticide / hectare (high)	Medium
Floodplain	**R-2**	Strong underestimation of the quantity of pesticides that might be present in water (high)	Medium
Pesticide spread area	**R-3**	Strong underestimation of the quantity of pesticides that might be present in water (high)	Medium

Table 3. Example results of a risk analysis (from (Grira, 2014)).

Risk evaluation consists of prioritizing the analyzed risks according to a level of tolerance (specific to the context). Risk evaluation helps to select strategies for risk treatments and must consider legal, regulatory and other requirements. This step is usually conducted using a ranking matrix. Figure 31 presents an example of a risk ranking matrix made of two axes to show the likelihood of occurrence and the impact of consequences. The intersection of these two axes shows the overall risk value. Table 4 presents the results of the evaluation of the risks presented in table 2 (extract from (Grira, 2014)).

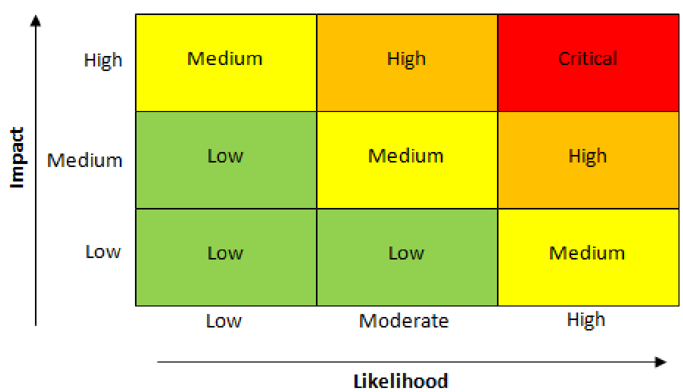

Figure 31. Example of a risk ranking matrix.

Feature / Attribute	Identified Risks	Impact of Risk	Probability of Occurrence	Overall Risk Evaluation
Cultivated parcel	**R-1**	Strong overestimation of the ratio quantity of pesticide / hectare (high)	Medium	High

Floodplain	**R-2**	Strong underestimation of the quantity of pesticides that might be present in water (high)	Medium	Medium
Pesticide spread area	**R-3**	Strong underestimation of the quantity of pesticides that might be present in water (high)	Medium	Medium

Table 4. Example results of a risk evaluation (from (Grira, 2014)).

4.3.3 Building risk responses (risk treatment)

Risk treatment implies selecting and implementing one or a combination of strategies to modify a risk in order to reach accepted levels of tolerance. Four categories of strategies can be used:

- Mitigation: modify actions to eliminate or reduce consequences or their likelihood of occurring
- Avoidance: eliminate activity to eliminate risk
- Transfer/sharing: shift impact to another entity entirely or in part
- Acceptance: voluntarily accept and take risk

When managing risks of inappropriate use of geospatial data, examples of mitigation strategies include:

- Improve database design/dataset structure
- Improve the quality control of the dataset (e.g., add integrity constraints)
- Use standards (e.g., for data quality and risk management interoperability)
- Properly inform users in a language they understand (highly recommended)
  - Provide a user manual
  - Offer a 1-800 help line or Ask@yourcompany.com
  - List target usages and non-recommended usages
  - Guide of good practices
  - Train users
  - …
- Conduct tests on the dataset (users)
- Compare with another dataset (users)

Examples of risk avoidance strategies include:

- Stop distributing or using the dataset or a part of
- Eliminate a category of users
- Eliminate a data provider
- Explicitly and clearly forbid a given usage

- …

Examples of risk transfer or sharing strategies include:

- Buy an insurance
- Obtain the dataset from a broker who can give advice related to its use
- Use a dataset with a guarantee that explains clearly risk sharing (who is responsible of what)
  - The guarantee for geospatial products is explained in (Plante & Gervais, Geospatial Data Quality Guarantee, 2015)
- Have the dataset evaluated by an expert in quality
- Replace a B2C strategy with a B2B strategy for your business by contracting a data broker who will offer the B2C strategy
- Have the data quality evaluated by an external expert for the new usages

A user accepting a risk will use the dataset no matter what the risks are and do nothing about it, i.e. the user will take the risk. For the data provider, it is important to make sure that the user does so with full knowledge of the risk.

4.3.4 Communicating risks

Risk communication involves:

- Communicating with the persons in charge of offering data/services, the persons in charge of the data production, and the users
- Developing the information products identified in the risk treatment strategies, especially:
  - B2C + C2C: paper or on-line user manuals with previously recommended content
  - B2B: embed risk-related information within data quality metadata and report
- Using a vocabulary adapted to the various audiences
- Properly informing users: distribute the information products; keep users informed of new context elements, new quality controls, new quality evaluations, new risk management strategies implemented, new usages, new restrictions, new good practices, etc.
- Promoting joint committees with various expertises
- Supporting Frequently Asked Questions (FAQ), 1-800 info lines, info@quality.yourorganisation
- Offering training, webinars
- …

Examples have been provided in previous sections.

4.3.5 Monitoring and reviewing risks

Risk monitoring is the regular surveillance of the risks and the success of their treatment. This step involves:

- Overseeing and systematically evaluating, using metrics, the effectiveness of actions taken
- Updating the initial list of identified risks and their characteristics
- Gathering information useful for the development or the update of risk response strategies
- Reviewing some aspects related to the planning process of risk management as a whole
- Collecting feedback about quality (e.g., via web-based forum, error-reporting sites, or 5-star Volunteered Quality Information (VQI))
- Implementing new restrictions, adding integrity constraints, implementing new training, building/updating a register for quality and risk management
- …

Table 5 shows an example of a risk register used to monitor risks (adapted from (Grira, 2014)).

Feature / Attribute	Identified Risks	Impact of Risk	Owner	Probability of Occurrence	Overall Risk Evaluation	Response	Status	Action items
Cultivated parcel	R-1	Strong overestimation of the ratio quantity of pesticide / hectare (high)	Project manager	Medium	High	Mitigate (…)	(date) open	Follow-up with users
Floodplain	R-2	Strong underestimation of the quantity of pesticides that might be present in water (high)	Project manager	Medium	Medium	Mitigate (…)	(date) open	Follow-up with users

Table 5. Example of a risk register (from (Grira, 2014)).

4.4 Communicating about geospatial data quality and risks of usage in practice

Examples of quality and risk documentation have been presented in previous sections:

- In the B2B (experts) context:
  - Communication products and services that were identified in the B2B contract
  - Data quality metadata (ISO 19115, ISO 19157), see figure 23
  - Initial specifications as additional information for new external expert users

    - Depending upon the detailed context: data quality report (ISO 19157) potentially for each type of usage (to complement the metadata), see figure 24
    - Optional: user manual and other communication products/services offered in B2C and C2C contexts, see figure 18

- In the B2C and C2C (mass market) contexts:
    - User manual highly recommended (Gervais, 2004), it may include:
        - Warnings (including symbols (ISO/TC 145/ISO 3864-2)), see figure 21
        - Disclaimer
        - Recommended and non-recommended usages
        - and much more
    - License
    - Guarantee
    - …

5. Recommendations

In this section, we make some recommendations for the Geospatial Community with regards to geospatial data quality. Recommendations have been grouped according to the context (B2B, B2C, and C2C).

5.1 Recommendations in a business-to-business context

In the B2B context, the challenge related to geospatial data quality is to use more formal methods for data quality evaluation and reporting.

The recommendations are:

- Facilitate communication between experts about data quality by adopting a common language, based on ISO 19157 (International Organization for Standardization, 2013), ISO 19115-1 (including quality metadata) (International Organization for Standardization, 2014), and on Synthesized Quality Reports, Aggregated Quality Information, and Automated Q&A Advisory System.
- Foster more efficient geospatial data reuse and interoperability by adopting a common frame of reference regarding quality (i.e. set of concepts), based on ISO 19157 (International Organization for Standardization, 2013) and ISO/TS 19158 (International Organization for Standardization, 2012).
- Facilitate contractual agreements by adopting the common frame of reference and common language proposed above.
- Decrease the risks of inappropriate uses of geospatial data using risk management concepts and strategies, based on ISO 31000 (International Organization for Standardization, 2009). Balance these efforts with data quality increase.
- Gradually become familiar with the concepts and standards mentioned above by participating in specialized training.
- Encourage high data quality with strong quality assurance procedures and quality controls (i.e. improve meta-quality), based on ISO 9000 (International Organization for Standardization, 2005), ISO 19157 (International Organization for Standardization, 2013), ISO 19158 (International Organization for Standardization, 2012), quality auditing, certification and use accreditation.
- Develop a Guide of Good Practices to visually represent geospatial data quality (quality maps, quality radars, quality tables, quality warning symbols, etc.)
- Develop a Guide of Good Practices to mitigate the risks of geospatial data misuse
- Further clarify the roles and responsibilities between parties by including quality guarantees.
- Gradually implement good practices with the help of geospatial data quality experts.

- Develop a series of quality-related products and services such as quality guarantee, quality certificate, quality audit, quality control and quality assurance mechanisms, accreditation of quality experts.

5.2 Recommendations in a business-to-consumer context

In the B2C context, the challenge related to geospatial data quality is to manage risks for a better protection of consumers (and providers).

The recommendations are:

- Facilitate geospatial data selection based on users' needs (cf. external quality) by providing:
  - Lists of recommended and non-recommended uses
  - 1-800 free line or Ask@yourcompany.com
- Contribute to the advancement of the geospatial community and spatially-enabled society by offering:
  - User manuals written in a language understandable by the target users
  - Emphasize on clear advices and warnings (use symbols, cf. ISO/TC 145/ISO 3864-2 (International Organization for Standardization, 2004))
  - Real guarantees (as for any other product/service in a mature market)
  - Guides of Good Practices
  - Synthesized quality reports and aggregated quality information
- Stimulate quality analysis and users' awareness with:
  - Web-based participatory VQI (Volunteered Quality Information) (e.g., 5-stars + comments)
  - Web-based users forums
  - Web-based or in-person training
- Reduce the uncertainty related to certain law-related topics by investing into studies to:
  - Understand the new trends and rights regarding privacy, data ownership, copyright, data vs. service
  - Further develop the concept of guarantee (see (Plante & Gervais, Geospatial Data Quality Guarantee, 2015))
  - Further clarify the responsibility of non-experts VGI data contributors
  - Clarify responsibilities when geospatial services and data cross borders
  - Stimulate legal interoperability of geospatial data (see (Uhlir, 2013))
- Rapidly make the move towards becoming a mature mass-market:
  - Specialized training, innovation, collaboration
- Improve metadata (i.e., easier to use, new types of quality-centered and risk-centered metadata written for the end-users, explicitly illustrate warnings with ISO 3864-2:2004 (Graphical symbols - Safety colours and safety signs - Part 2: Design principles for product safety label).

- Increase geospatial data providers' and users' awareness of potential risks of data misuses by gathering, on a crowdsourced web site, examples of damages (N.B. such a site was developed under GEOIDE funding at http://dataquality.scg.ulaval.ca).

5.3 Recommendations in a consumer-to-consumer context

In the C2C context, the challenge related to geospatial data is about increasing awareness.

The recommendations are:

- Rapidly inform developers of public-oriented web-based systems and smartphone apps about their duty and potential liability with regards to geospatial data quality
- Facilitate geospatial data selection based on users' needs (cf. external quality) by providing:
    - Lists of recommended and non-recommended uses
    - 1-800 free line or Ask@yourcompany.com
    - User manuals written in a language understandable by the target users
        - Emphasize on clear advices and warnings (use symbols, cf. ISO/TC 145/ISO 3864-2)
    - Guides of Good Practices
    - Training and collaboration for the providing C (e.g., when publishing data mashups)
- Reduce the uncertainty related to certain law-related topics by investing into studies to:
    - Understand the new trends and rights regarding privacy, data ownership, copyright
    - Further clarify the responsibility of data contributors, integrators and distributors
    - Clarify responsibilities when geospatial services and data cross borders

6. Conclusion

As geospatial data is increasingly being produced and (re)used by new types of actors, the question of geospatial data quality is becoming a major concern. The objective of this guide was to support the Canadian geospatial community into its efforts to make the spatially-enabled society more aware of geospatial data quality and of the risks related to the inappropriate use of geospatial data. Based on international standards such as ISO 19157 (Geospatial data quality) and ISO 31000 (Risk management), this guide presented the concepts underlying geospatial data quality, the management of geospatial data quality, the geospatial data quality evaluation process in details (based on ISO 19157), and the management of risks of inappropriate use of geospatial data (based on ISO 31000). Detailed examples of quality management, evaluation and risk management tasks to be undertaken in the B2B, B2C and C2C contexts were presented.

7. Bibliography

American National Standards Institute. (2006). *ANSI Z535.6 American National Standard for Product Safety Information in Product Manuals, Instructions, and Other Collateral Materials.* ANSI.

Bédard, Y. (2012, May 18). Geospatial Data Quality Awareness, The Next Challenge: Are We Ready? Invited Keynote Speaker, GEOIDE IV-23 International Workshop on Geospatial Data Quality, Legal, Ethical and Technical Aspects. Quebec City, Canada.

Bédard, Y. (1988). Uncertainties in Land Information System Dayabases, Proceedings of the Eighth International Symposium on Computer Cartography. Baltimore, Maryland.

Bédard, Y., Chandler, J., Devillers, R., & Gervais, M. (2009, March 22-27). System Design Methods and Geospatial Data Quality, Association of American Geographers Professional Ethics - Session on Geographic Information Ethics and GIScience. Las Vegas, USA.

Canada, T. B. (2012). *Guide to Integrated Risk Management.* Treasury Board Secretariat of Canada. Treasury Board Secretariat of Canada.

Chandler, J. A. (2010). User generated legal issues: Potential legal issues raised by user-generated content for geographical information services. *Proceedings GSDI 12 World Conference.* Singapore.

Chandler, J. A., & Levitt, K. (2011). Spatial Data Quality: The Duty to Warn Users of Risks Associated with Using Spatial Data. *Alberta Law Review , 49* (1), pp. 79-106.

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

Devillers, R. (2004). Conception d'un système multidimensionnel d'information sur la qualité des données géopatiales. *Thèse de doctorat présentée en cotutelle, Faculté de foresterie et géomatique, Université Laval* .

Devillers, R., Bédard, Y., & Gervais, M. (2013, April 9-13). We need to adopt more ethical practices for geographic information data and products dissemination, 2013 AAG Annual Meeting. Los Angeles, California, USA.

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) , 21* (3), pp. 261-282.

Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P., & Shi, W. (2010). 30 years of research on Spatial Data Quality : Achievements, failures and opportunities. *Transactions in GIS , 14* (4), pp. 387-400.

Federal Geographic Data Committee. (1998). *FGDC-STD-001-1998 Content Standard for Digital Geospatial Metadata.* FGDC.

GeoConnections. (2013). CGDI Data Quality Poster.

GEOIDE (PIV-23). (2012). Retrieved from Public protection and ethical dissemination of geospatial data - Social and legal aspects: http://dataquality.scg.ulaval.ca/

Gervais, M. (2004). La pertinence d'un manuel d'instructions au sein d'une stratégie de gestion du risque découlant de la fourniture de données géographiques numériques, Ph.D. Thesis, Department of Geomatics Sciences, Laval University.

Gervais, M., Bédard, Y., & Larrivée, S. (2007). Spatial Data Quality Audit, June 19th, Quebec, Canada. *Third Annual National Surveyors Conference.*

Gervais, M., Bédard, Y., Jeansoulin, R., & Cervelle, B. (2007). Qualité des données géographiques. Obligations juridiques potentielles et modèle du producteur raisonnable. *Revue internationale de géomatique , 17* (1), pp. 33-62.

Gervais, M., Bédard, Y., Larrivée, S., Rivest, S., & Roy, T. (2013, October 3-4). Enquête canadienne sur la qualité des données géospatiales et la gestion du risque, Géomatique 2013, 3-4 octobre, Montréal, Canada. Montréal, Canada.

Gervais, M., Bédard, Y., Lévesque, M.-A., Bernier, E., & Devillers, R. (2009). Data Quality Issues and Geographic Knowledge Discovery. In *Geographic Data Mining and Knowledge Discovery, Second Edition.* CRC Press.

Grira, J. (2014). Improving knowledge about the risks of inappropriate uses of geospatial data by introducing a collaborative approach in the design of geospatial databases. *Ph.D. Thesis, Geomatics Sciences Department, Laval University* .

Grira, J., Bédard, Y., & Roche, S. (2009). Spatial Data Uncertainty in the VGI World: Going from Consumer to Producer. *Geomatica , 64* (1), pp. 61-71.

Hickling Arthurs Low Corporation. (2011). *Final Report: CGDI Operational Policies Needs.* GeoConnections.

International Organization for Standardization . (2009). *ISO 19115-2:2009 Geographic Information - Metadata - Part 2: Extensions for imagery and gridded data.* ISO.

International Organization for Standardization. (2005). *ISO 19109:2005 Geographic Information - Rules for application schema.* ISO.

International Organization for Standardization. (2005). *ISO 19110:2005 - Geographic information - Methodology for feature cataloguing.* ISO.

International Organization for Standardization. (2014). *ISO 19115-1:2014 Geographic information - Metadata - Part 1: Fundamentals.* International Organization for Standardization.

International Organization for Standardization. (2005). *ISO 19123:2005 Geographic Information - Schema for coverage geometry and functions.* ISO.

International Organization for Standardization. (2007). *ISO 19131:2007 Geographic Information - Data product specifications.* International Organization for Standardization.

International Organization for Standardization. (2011). *ISO 19131:2007/Amd.1:2011 Geographic Information - Data product specifications Amendment 1 Requirements relating to the inclusion of an application schema and feature catalogue and the treatment of coverages in an application schema.* International Organization for Standardization.

International Organization for Standardization. (2007). *ISO 19139:2007 Geographic information - Metadata - XML schema implementation.* ISO.

International Organization for Standardization. (2013). *ISO 19157:2013 Geographic information – Data quality.* International Organization for Standardization.

International Organization for Standardization. (2009). *ISO 31000:2009 Risk Management - Principles and Guidelines.* International Organization for Standardization.

International Organization for Standardization. (2004). *ISO 3864-2:2004 Graphical symbols -- Safety colours and safety signs -- Part 2: Design principles for product safety labels.* International Organization for Standardization.

International Organization for Standardization. (2005). *ISO 9000:2005 Quality management systems - Fundamentals and vocabulary.* International Organization for Standardization.

International Organization for Standardization. (2012). *ISO/TS 19158:2012 Geographic information - quality assurance of data supply.* International Osganization for Standardization.

Jones, K. (2011). Communicating perceived geospatial data quality of 3D objects in virtual globes, M.Sc. Thesis, Department of Geography, Memorial University of Newfoundland.

Jones, K., Devillers, R., Bédard, Y., & Schroth, O. (2013). Visualizing perceived spatial data quality of 3D objects within virtual globes. *International Journal of Digital Earth* .

Juran, J. M., Gryna, F. M., & Bingham, R. S. (1974). *Quality Control Handbook.* New York: McGraw-Hill.

Kerzner, H. (2009). *Project Management, A Systems Approach to Planning, Scheduling and Controlling, 10th Edition.* New Jersey: John Wiley & Sons.

Koistinen, K. (2015). Quality information representation in ELF Geo Product Finder. *International Workshop on Spatial Data and Map Quality.* Malta.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2001). *Geographic Information Systems and Science.* England: John Wiley & Sons.

Maitra, J. B. (2004). Applying ISO/TC 211 Standards in the Development of Data Content Standards. *GSDI-7.*

National Committee for Digital Cartographic Data Standards. (1987). *A Draft Proposed Standard for Digital Cartographic Data.*

Natural Resources Canada. (2015). canvec_082C_pna.

Natural Resources Canada. (2014). *CanVec+ Data Product Specifications Edition 1.0.*

Natural Resources Canada. (2014). *CanVec+ Feature Catalog Edition 1.0.2.*

Normand, P. (1999). Modélisation des contraintes d'intégrité spatiale, théorie et exemples d'applications. *Essai de M.Sc., Faculté des Sciences Géomatiques, Université Laval* .

Open Geospatial Consortium. (2014). *Data Quality DWG.* Retrieved from Open Geospatial Consortium: http://www.opengeospatial.org/projects/groups/dqdwg

Plante, K., & Gervais, M. (2015, March). Geospatial Data Quality Guarantee. *Geomatica* .

Sboui, T., & Bédard, Y. (2012). Universal Geospatial Ontology for the Semantic Interoperability of Data: What are the Risks and How to Approach them? In T. Podobnikar, & M. Ceh (Eds.), *Universal Ontology of Geographic Space: Semantic Enrichment for Spatial Data* (pp. 1-27).

Statistics Canada. (2012). *User Guide for the Survey of Household Spending 2012.*

Trakas, A. (2008). Benefits of Open Standards - how to engage with OGC.

Treasury Board of Canada. (2009). *Standard on geospatial data.*

Uhlir, P. F. (2013). The Legal Interoperability of Data. *NSGIC Midyear Meeting Archive.*

Veregin, H. (1999). Data Quality Parameters. In *Geographic Information Systems : Principles and Technical Issues* (pp. 177-189). John Wiley & Sons, Inc.