Chapter 12

Quality components and metadata

12.1 Introduction

Several years ago, databases stopped being merely simple collections of information stored in a structured format and became what they are today: indissociable from information systems that use their data and of which they are part.

Such information systems form the core of various applications both at the final level (management, systems for helping decision-making, etc.) as well at the level of end users (banks, local governments, large organizations, etc.).

In such a context, it is essential to understand what data is and to control its quality. This necessitates the active involvement of designers of information systems (IS) and the producers of the underlying data to ensure that the data fulfils the needs of the future users.

Existing geographic databases often contain errors due to the acquisition source (measuring instruments), data-input processes and information processing. In addition, the evolution in features of geographic information systems (GIS) as well as the emergence of the Internet has caused a shift in how information systems and their underlying data is used; shared information that is available online can be

Chapter written by Sylvie SERVIGNE, Nicolas LESAGE, Thérèse LIBOUREL.

'diverted' from its primary use. Mainly due to its high acquisition costs, spatial data tends to have a long life – which leads to it being used for purposes that were not originally foreseen. Originally acquired to allow cartographic plotting (which could accommodate errors that were not visible at the plotting scale), entire datasets are now being used in the field of spatial analysis which uses methods that range from interpolation to simulation for the purpose of help in decision making. Limitations, in terms of quality, of such data are more significant in this type of processing (topological consistency, precision[1] of attribute values, etc.) and it becomes imperative to define quality standards and strategies to improve this quality so that the life of currently existing batches of data and datasets can be extended. Moreover, if precision and reliability have long been the parameters of quality for qualifying geodetic networks, the quality of today's spatial or spatio-temporal databases is more difficult to define because of the complexity of spatial attributes: dimensions of definition of managed objects (1D, 2D, 3D geometric descriptions), spatial relationships between the objects (topology), potential richness of non-spatial attributes, etc.

The design of IS and databases should include, in its own right, the data quality. Thus, the quality should be specified and processes for improving and monitoring it implemented [SER 00]; some data changes rapidly (notably in the urban environment) and the data quality should also be ensured over the long term.

This integration of data and data quality is most often implemented by using metadata ('data about data' in its first meaning). Metadata allows the documentation, as precisely as possible, of data, facilitating its sharing and distribution with the overall goal of simplifying its integration and reuse. The emergence of the digital document has led to the phenomenon of annotation (well-known to librarians). The proposals of the Dublin Core and W3C (World Wide Web Consortium) attest to it. However, spatial information, due to its particularities, requires complex and voluminous metadata to be stored and organized for geographic information and this complexity and size becomes a major hindrance to its wider use. It is therefore imperative that efficient and well-conceived standards exist and take into account data quality in the appropriate measure. As an example, the information on quality should ensure the reliability of processes based on the data, as well as the system's ability to fulfil expected functions (suitability for requirements as expressed in the specifications). These two complementary notions are found in the definition of quality put forward by the International Organization for Standardization [ISO 94]:

---

[1] Editor's note: The concept described by the term 'precision' in this article should be called 'accuracy' (see section 12.2.4 for the difference between these two terms). The authors have nevertheless, chosen to use the term 'precision' because it is commonly used (erroneously) and is thus better known in the geographic information domain, both in common as well as in scientific and technical contexts.

'Set of properties and characteristics of a product or a service which confers upon it the ability to satisfy expressed or implicit requirements'. Finally, in the context of geographic information, it is necessary to keep in mind the different points of view of different users of geographic data, i.e., the data producers and data users. In fact, data producers or suppliers want to adhere to quality standards because it confers certification on batches of data that they produce or sell. Users, on their side, would like to have data whose quality is appropriate to their needs and thus to their applications.

This chapter thus takes up the concepts of quality introduced in standardization approaches. It will describe their definition and how they have been incorporated within metadata standards dedicated to spatialised information.

12.2 Concepts of quality

For long the description of quality has been reduced to a problem of the precision of stored information (see chapter 2). During the analogue age, as far as geographic data was concerned, accuracy almost exclusively concerned the position of represented objects, ignoring problems linked to their shape, representation, semantic quality or consistency. The advent of the digital age saw work on standardization, which started in the 1980s, lead up to a consensus on the definition of quality components.

The terminology surrounding spatial data quality is subject to numerous variations, and different terms are sometimes used to describe the same concept.

12.2.1 Quality reference bases

To ascertain a dataset's quality, it is necessary to have reference bases that will serve as a basis of comparison of the datasets under consideration. Two concepts, 'nominal terrain' and 'universe of discourse' can constitute possible definitions of the reference base:

– Nominal terrain: A literal, though inexact, translation of the French term, *'terrain nominal'*. It has a number of definitions, among which that of IGN, adopted by the CEN [CEN 98]: 'the real world as viewed through the specifications used when inputting a batch of geographic data'. For a shorter version, we can consider the definition of [VAU 97], 'that which should have been entered', as wholly satisfactory to describe the notion of nominal terrain and corresponding better to the English definition of 'universe of discourse' which clearly separates the producer and user aspects;

– Universe of discourse: abstractions of the real world incorporating both complementary aspects:

– of data-production specifications,

– of users' requirements, etc.

12.2.2 Quality criteria

Criteria called quantitative embody a quantitative expression of quality (for example, spatial precision = 12 m) and are also called 'quality parameters'. Criteria called qualitative (Truth in labelling) provide a qualitative expression of quality (example: lineage).

In 1987, the National Committee on Digital Cartographic Data Standards (NCDCDS, [MOE 87]) proposed the definition that describes spatial data quality by breaking it down into five criteria (one qualitative and four quantitative): lineage, geometric precision or positional precision, semantic precision or precision of attributes, completeness, and logical consistency. In 1991, the executive committee of the International Cartographic Association (ICA) established a commission on data quality [GUP 95]. This commission had as aim to develop, document and publish criteria and methods for the evaluation of digital cartographic datasets. It identified three parts in the specification and use of information on the quality of spatial data:

– the definition of elements of spatial quality;

– the establishment of metrics to measure elements of spatial quality;

– the communication of data quality.

In 1995, to the five quality criteria defined by the NCDCDS, the commission added two new parameters: 'temporal precision' and 'semantic consistency'. In 1997, the IGN [DAV 97] introduced 'specific quality' to help overcome potential lacunae not covered by the previous criteria.

12.2.2.1 Qualitative criterion

The qualitative criterion retained is designated by the term 'lineage'.

This criterion provides the material origin of the data and the methods used, as well as all subsequent transformations undergone by the data, to arrive at the final data. In other words, the lineage describes the acquisition procedures and methods of deriving and transforming data.

However, the objective of the lineage component can be interpreted differently by the data producer and the data user:

– the producer wants to ensure that standards are maintained;

– the user wants to know the origin of the acquired data so that he can be sure that it fulfils his requirements.

12.2.2.2 Quantitative criterion

The quantitative criterion or quality parameters are:

– *Geometric precision[2] (or positional precision, spatial precision, spatial accuracy).* It gives the level of conformity of data with respect to the nominal terrain from the point of view of the difference of the respective positions in these two views. It thus defines the deviation in the values of the respective positions between the database data and the nominal terrain;

– *Semantic precision[3] (or precision of non-spatial attributes):* This criterion provides information on the difference between the values of non-spatial attributes and their real value and thus gives us the deviations of measurements of qualitative attributes or quantitative attributes (classification);

– *Completeness:* It can be applied to the level of the model, the data or even objects and attributes. Data completeness helps us detect errors of omission (abnormal absence) or commission (abnormal presence) of certain objects. Model completeness, on the other hand, expresses suitability of the provided representation for users' requirements;

– *Logical consistency.* It has the goal of describing the faithfulness of relationships encoded in the database's structure with respect to all the constraints caused by data-input specifications. In other words, it describes the correspondences of the dataset with the characteristics of the structure of the model used (respecting specified integrity constraints).

– *Temporal precision.* It provides information on the temporal aspect of data: management of data observation dates (origin), of types and frequency of updates, and the data's validity period. It could be essential to have this information, especially when evaluating the suitability for the requirements of a particular user;

---

[2] See editor's footnote on previous page.

[3] The semantic qualifier was initially associated with non-spatial attributes and, even though this nomenclature can be debated, we retain the qualifier.

– *Semantic consistency.* It indicates the relevance of the significance of objects with respect to the selected model; it describes the number of objects, of relationships and attributes correctly encoded with respect to the set of rules and specifications;

– *Specific quality.* A quality parameter (thus quantitative) expressing quality-related information that is not foreseen by the previous criteria. Thus, IGN [DAV 97] introduced the concept of 'timeliness' which helps determine the suitability for requirements by translating the offset between the produced dataset and the nominal terrain to a later instant.

12.2.3 Expression of the quality

Quality is expressed with the help of indicators, elements and measurements; their definitions follow.

– *Quality indicator:* Set of quality measurements indicating the performance of a quality parameter for an entire batch of geographic data;

– *Quality element:* Set of quality measurements indicating the performance of a quality parameter for all or part of a batch of geographic data;

– *Quality measurement:* Definition of a specific test to apply to geographic data, including algorithms and the type of value or set of values that result.

12.2.4 Precision and accuracy

There is a fundamental difference between the two concepts: precision indicates the *resolution* with which one can measure a phenomenon with a particular instrument or method (see figure 12.1) as well as the ability to obtain the same value by repeating a given measurement. In the GIS domain, precision varies most often with the cartographic scale used. A rule of thumb is that a precision is acceptable if it causes an error on the map of the order of 1/10th of a millimetre (which at 1:1000 represents an error of 10 cm and at 1:500,000, an error of 50 m). Accuracy, on the other hand, bears on the notion of truth (the centre of the target in figure 12.1), and of exact data representing faithfully the real phenomenon that it is attempting to represent. Inaccuracy arises from, among other reasons, measurement errors and can be linked to systematic methodological problems themselves caused by the imperfect nature of the method used to acquire the data and by use of unsuitable digital processing procedures (for example, a numeric range that is too narrow in a series of complex calculations with automatic truncation at each step of the process). These systematic errors should, as far as possible, be listed in the lineage elements (see

section 12.3.1), even if their effects are also felt, for example, in the domain of the geometric precision.
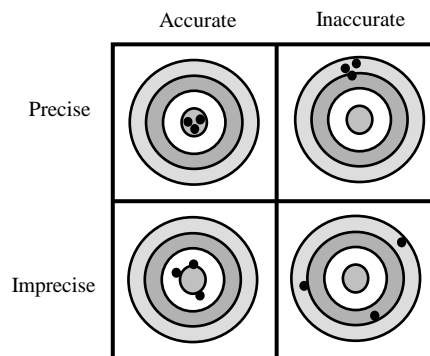

Figure 12-1. Comparison: accuracy and precision

12.2.5 Appraisal and use of quality

A very important concept for the appraisal of quality by the user is that of suitability for requirements[4] or 'fitness for use'. It represents the potential – admittedly subjective – of the data to fulfil specific requirements of the user (see chapter 15 for an example of a method for evaluating fitness for use). This is a difficult criterion to evaluate using the quality criteria defined above. Nevertheless, it is essential to do so because it allows a potential user to determine whether a particular dataset can fulfil the purpose he expects it to. Tests of deviation, appropriate to the target application, will have to be available or complementary annotation by the user will have to be authorised (based on the metaquality and his specific expertise).

12.2.6 Metaquality

The evaluation of the quality, using any one parameter, allows us to represent the corresponding performance of the dataset with respect to the considered quality element. It is essential to supply, at the same time as the result of the evaluation, a set of indications that allows one to qualify this information. We are now talking of quality of quality, and use the term 'metaquality' to describe it. The most important of these indications are the date of processing (temporal aspect), the evaluation method used (tested, calculated or estimated) and the population on which it was applied.

---

[4] Often called external quality.

– The processing date could be ad hoc (case of a quality audit conducted at regular or irregular intervals) or could be continuous, as in the case of systems for which mechanisms exist to ensure integrity of some data aspects (triggers, etc.). The processing date then corresponds to the date the quality report was created;

– The methods used could be more, or less, reliable (use of a threshold, quality of algorithms used, propagation methods – statistics);

– Finally, the population will vary depending on the method: from the entirety of the data for a general audit, to different types of sampling involving a variable number of elements. Partitions can also be used, either temporal (evaluation of the quality of entities input in the last two months, or of those that are 3 to 5 years old) or geographic (processing of a specific administrative area, for example). These two types of partitions can, of course, also co-exist within the same process.

CEN [CEN 98] has identified three main elements of metaquality. These are confidence, homogeneity and reliability.

– Confidence: 'A metaquality element that describes the accuracy of quality information.' Confidence originates primarily from the method used and of its reliability, as well, to a lesser extent, from the concerned population;

– Homogeneity: 'Textual and qualitative description of the expected or tested uniformity of quality parameters in a batch of geographic data.' In fact, a dataset can be the result of a single acquisition process or it can result from a combination of several technologically varied acquisitions (aerial photos, digitisation of paper maps, GPS, theodolites, etc.). The homogeneity depends mainly on the population that was the basis of the evaluation. In the case of a general process, it cannot be evaluated because the result is global. Homogeneity is thus only relevant when several segments were used and their evaluation results (derived using the same methods) compared. These tests are often conducted when data has been input by different operators, depending on the zone or the acquisition date.

– Reliability: 'A metaquality element describing the probability that a given sampling of a batch of geographic data, when used for quality evaluation purposes, is representative of the entire data batch.' A statistical method based on sampling could be considered as reliable as a global method when all the geographic zones and concerned time periods are covered and the population is sufficiently large.

12.3   Detailed description of quality criteria

12.3.1   Lineage

[CLA 95] identifies the information necessary for reconstructing the history of a dataset and to deduce therefrom its potential usage (processing methods and tools for a particular requirement):

– The data source (the organization's reputation, if not quantifiable, should also be taken into account), origin, reference domain (geology, etc.), characteristics of spatial data, co-ordinate and projection systems, and associated corrections and calibrations;

– Acquisition, compilation and derivation: fundamental hypotheses of observation, calibration and corrections. Then the georeferencing or application to a particular domain – taking an arbitrary 0 altitude, for example – followed by the description of methods used to interpret, interpolate or aggregate data, at the level of the structure or the format used;

– Data conversion: definition of processes, such as, for example, the stages in the vectorisation of raster data;

– Dates of different stages of processing;

– Transformations or analyses: transformation of co-ordinates, generalisation, translation, reclassification, all defined, as far as possible, in precise mathematical terms. All parameters used should be clearly defined, since these transformations can have profound effects on the produced data;

At the normalisation level, importance is often accorded to the data structure rather than to its semantics. It is possible that the real nature of the information on the lineage is not sufficiently 'closed' to be able to be represented in a standardized manner (the number of possible and successive processes perhaps ruling it out). In any case, lineage information is often provided in the form of running text describing the parameters listed above.

The collection of this information can prove to be an onerous and difficult task, especially when it concerns data originating from different acquisition processes, and having undergone numerous transformations. It is, however, in this type of case that it is most useful, indeed indispensable.

12.3.2 Positional precision or geometric precision

Positional precision is generally divided into absolute precision and relative precision. It can also be differentiated between planimetric precision and altimetric precision (for 3-dimensional data). Altimetric precision often comes down to a problem of semantic precision (see next section), since the altitude of points is often stored in the form of an alphanumeric attribute.

The position of objects in the database is a set of cardinal values that allow them to be positioned in three-dimensional Cartesian or polar co-ordinates. For example [AZO 00]: field mapping (*X, Y, Z*), GPS position (latitude, longitude, altitude), digitisation (*Y, X*). The only way to measure positional precision is therefore to compare the dataset, either with another dataset of better quality (and following the same specifications), also called control data [DAV 97], or with data derived from surveys and samplings (for example with a GPS sensor). Geometric precision, or precision of the co-ordinates, directly depends on the acquisition methods and processing of measurements. For example, the positional and altitudinal precision of contour lines depends on the precision of measurement of the points used to determine the contours and of the interpolation algorithms used. It specifies RMS (root mean square) errors in planimetry and altimetry in the points' co-ordinates, possibly even their mean error ellipse.
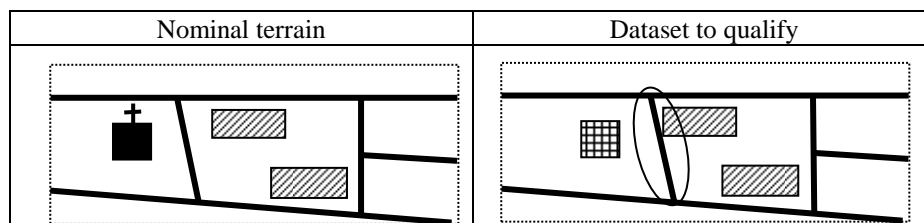
Nominal terrain	Dataset to qualify

Figure 12-2. Example of positional imprecision

12.3.3 Precision of attributes or semantic precision

An attribute can be the result of a measurement or interpretation, originate from direct human observation (such as the names of roads or lakes), or even from a historical or political census [UBE 97].

In the same manner as for geometric precision, the semantic precision is defined as the difference between a measurement and another comparable measurement known to be more accurate. This is a relative definition because it relates to the precision of the objects being compared. It also requires the knowledge of more

accurate data, namely the nominal terrain. As this does not really physically exist, reference data is used instead of the nominal terrain.

All types of attributes are subject to uncertainty because of defects in measuring instruments or data-acquisition procedures, or historical uncertainty that can afflict names. These uncertainties can be of different types depending on whether the attribute applies to a single location (attribute that is difficult to measure or valid only at a certain scale) or on a set of points (attributes are often calculated as averages or aggregations of values in the area under consideration).

To help evaluate semantic precision, a classification according to a scale of measurement was created for the specific requirements of spatialised information [GOO 95]. This classification applies to different types of simple attributes, i.e., attributes that are qualitative (names, classes used to characterise data) and quantitative (measurements, enumerations, analysis results, etc.) and introduces:

– Nominal scales (used to classify some characteristics, and though often numbered, not representing numerical values) such as residential, commercial or industrial zoning;

– Ordinal scales (to classify and sort) such as the soil richness: poor, medium or rich;

– Interval scales (when the system uses a relative zero – only measured differences make sense, as for temperature expressed in degrees Celsius – the difference in temperature is the same between 10 and 20 ℃ as it is between 20 and 30 ℃. But 40 ℃ is not the double of 20 ℃ since the zero is arbitrary);

– Ratio scales (if the ratios between measurements make sense, as is the case with temperatures in Kelvin for which 200 K = 2 × 100 K, since this scale is based on an absolute zero);

The first two scales can define both qualitative and quantitative attributes whereas the latter two only numerical values.

For attributes with cardinal values for example (interval or ratio), standard deviation can be used or, if necessary, an estimate of this standard deviation (height of trees estimated at ± 10%). For attributes with ordinal values, it becomes necessary to qualify the precision of the classification of objects when, for example, there is a possibility of confusion between object classes (for example, are the vegetation zones identified on an aerial photo not, in fact, constructed zones?). As for nominal values, a descriptive entry could be used to alert the user to the precision of the text. For Azouzi, for example [AZO 00], since the designation is one of the attributes of a

building, a qualifier of this attribute allows the user to be aware of the difficulties encountered during the assigning of the designation. By their very nature, errors linked to different types of attributes follow different statistics.

The determination of the semantic precision is sometimes similar to completeness if one considers that a difference in conceptual modelling can transform an attribute to a class or vice versa. Similarly, the geometric precision becomes a sort of semantic precision when we treat the location of objects as a specific attribute of entities [GOO 95].
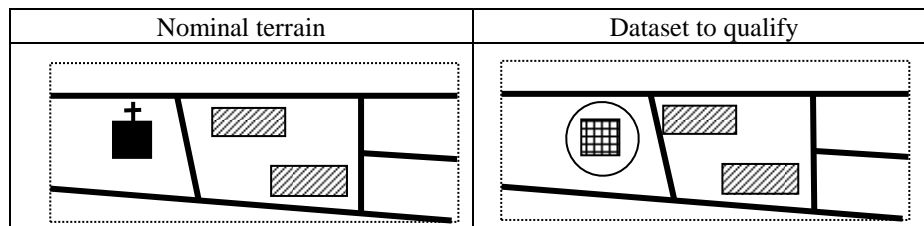
Nominal terrain	Dataset to qualify

Figure 12-3. Example of semantic imprecision

12.3.4 Completeness

'Completeness is an attribute that describes the relationships between objects represented in a dataset and is an abstraction of the same set of objects in the real world' [MOE 87]. Evaluating objects of the database with all the objects of the universe of discourse requires therefore that a formal description of both these sets be available.

Thus, depending on the domain under consideration, the completeness of a database (or a map) can be suitable for a specific task but not for another. One has therefore to relate the data quality with the fitness for use. The concept of 'fitness for requirements or use' comes into its own when data completeness has to be measured. In fact, if the information on data quality is, in principle, supplied by the producer of the dataset, the fitness for use, on the other hand, is only estimated at the time of evaluation of the *use* of the dataset (principle of 'truth in labelling'). In the useful lifetime of a dataset, the quality (considered in a general manner and not only for completeness) will be evaluated only once whereas a fitness-for-use evaluation will be conducted for each application.

Completeness is evaluated based on existing omissions and commissions between the nominal terrain and the dataset under evaluation.
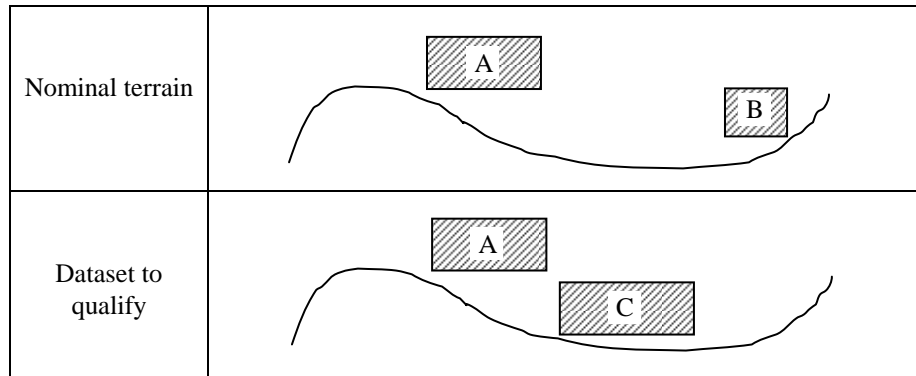
Nominal terrain	
Dataset to qualify	

Figure 12-4. *Examples of omission and commission: B = omission, C = commission*

In this context, we can distinguish between two types of completeness (see figure 12.5), *data completeness:* errors of omission or commission, which are, in principle, measurable and are independent of the application, and *model completeness:* comparison between the abstraction of the world corresponding to the dataset and the one corresponding to the application, preferably evaluated in terms of fitness for use (is the model rich enough to fulfil application requirements? [BRA 95]). Data completeness is itself broken down into 'formal' completeness (concerning the data structure – syntax, adherence to the standards and format used, presence of obligatory metadata) and object completeness, followed by that of attributes and relationships (subordinate to the that of the objects). Finally, combining the data completeness with model completeness allows one to estimate the completeness in terms of fitness for use.
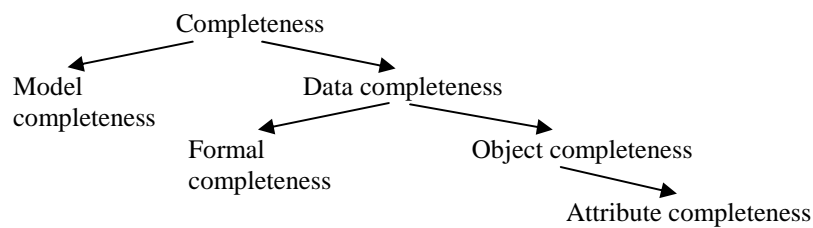

Figure 12-5. *Different types of completeness [BRA 95]*

In summary, completeness monitors the lacuna (omission) as well as the excess (commission) in information contained in the geographic database mainly by answering the following questions [AZO 96]:

– Is the coverage of the zone complete?

– Is the number of objects modelled equal to the number of objects defined in the model?

– Do the modelled objects have the correct number of attributes and are all attribute values present?

– Are all entities represented in the nominal terrain represented in the model?

– Is all that is included in the conceptual model also present in the database?

12.3.5 Logical consistency

Logical consistency relates to all logical rules that govern the structures and attributes of spatial data and describes the compatibility of a dataset item with the others.

Incidentally, this notion was used earlier in data integrity checks for non-spatial data. Its extension to geographic data was done at the time of the first analyses in the domain of topology.

Thus, a dataset is called consistent at the logical level if it respects the structural characteristics of the selected data model and if it is compatible with the attribute constraints defined for the data. There exist several different levels of logical consistency going from a simple range of attribute values to specific rules of consistency based on the geometry (example: is the contour of a polygon properly closed? [UBE 97]) (see figure 12.6) or on spatial relationships (constraints of topological integrity – example: every arc of a network should be connected by a node to another arc).

The consistency thus allows, amongst other things, to verify that:

– The objects described in geographic database respect the reality (nominal terrain) in an exact measure;

– The topology and the spatial relationships are represented and respected;

– The variables used adhere to the appropriate values (limit values, type, etc.);

– The data file is consistent (according to European standards, this aspect can even extend to the reliability of the medium on which the file is stored).
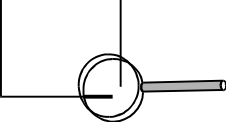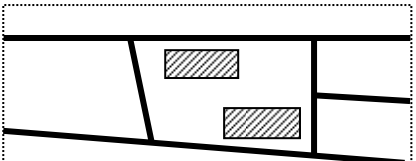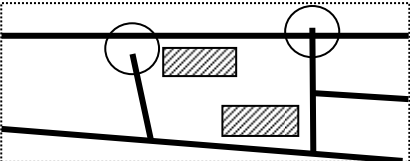
Nominal terrain	Dataset to qualify

Figure12-6. *Example of logical consistency: verification of the closure of the polygon contour, verification of the topological joining relationships*

12.3.6 Semantic consistency

The concept of semantic consistency expresses the quality with which geographical objects are described with respect to the model being used. This aspect of quality relates more to the relevance of geographical objects' significance than to their representation [SAL 95]. The semantic relevance is therefore of major importance in determining the fitness for use.

The goal of semantic consistency is to measure the 'semantic distance' between geographical objects and the nominal terrain. We can, once again, distinguish between the points of view of the producer and the user: for the former, the aim is, on the one hand, to provide documentation on the *semantic content* of his database (mainly by providing the specifications that define the nominal terrain, the model, the selection criteria, etc.) and, on the other, to provide information on the semantic performance of this database (level of conformance with the above-mentioned semantic constraints); for the latter, the goal is to define the suitability of this data for his own requirements. The knowledge of the specifications is, for the user of primary importance, especially from the semantic point of view: do the user and producer agree on a named phenomenon? (For example, does the 'hospital' class include clinics?)

As far as the specifications are concerned, two basic levels can be defined [PUR 00]: the geometric level which provides the shape and location of objects and the semantic level to describe the objects. Irrespective of whether the data's physical representation uses a vector model or a raster model, it always respects these two levels: for raster data, the geometry is made up of a collection of pixels and the semantics which are associated with these values; for vectorial data, the geometry

indicates the shape and the absolute or relative position (encoded according to the geometric primitives used) and the semantics bear upon the attributes, their values, or even the explicit relationships between the entities.

The selection criteria define, for example, the input limits (minimum size that a entity should have to be input), operated aggregations and corresponding criteria ('all crop fields will be stored as "agricultural zones" and merged as required'). The extraction is, finally, a transformation of entities of the real world into objects, attributes, fields of the selected model, and data. To indicate all the parameters used, especially in the generalisation procedures implemented, is as important in evaluating semantic consistency as it is for lineage.

In order to evaluate the semantic consistency of a database, [SAL 95] starts by introducing the concept of 'ability of abstraction' of phenomenons that have to be taken into account. Some of them are, in fact, difficult to model (edge of a forest, for example) and it is often worthwhile to evaluate whether the apprehension of the phenomenon is universal or whether it depends strongly on the observer, the context or the observation date (seasons, shadows, etc.). (See chapter 7 for a discussion of this problem.)

The methods used for evaluating the semantic consistency can be compared to those for measuring the completeness (omission/commission) of the objects, attributes and relationships. The semantic consistency also covers the field of logical consistency (data constraints), temporal precision (inconsistent dates, etc.) and semantic precision (a semantic inconsistency can also denote a classification error, for example) [PUR 00].

In conclusion, semantic consistency is composed of several parameters that cannot be easily differentiated. A flagrant error (for example, a house in a lake, see figure 12.7) is a semantic inconsistency but may be due to a temporal error (modification of the banks), a logical inconsistency (not taking into account a house on stilts), or a completeness error (forgetting an island or addition of the house or of the lake).
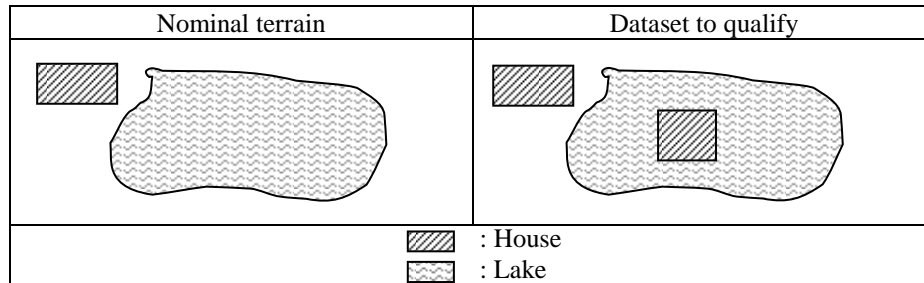
Nominal terrain	Dataset to qualify
▨ : House	
▨ : Lake	

Figure 12-7. Example of semantic inconsistency

12.3.7 Timeliness

This criterion represents the offset between a produced dataset and the nominal terrain on a reference date T. Timeliness provides information about the 'freshness' of data. It can be represented, for example, by a validity period for the data batch, a period defined by starting and ending dates.

12.3.8 Temporal consistency

The date the data is input, or the date of its revision, is an important factor for the user to judge the data quality (in the sense of fitness for use). Temporal consistency concerns the dates of data acquisition, types of updates and validity periods.

Depending on the type of phenomenon observed, the management of time-related issues will be different. Some entity classes are re-input at more or less regular intervals (aerial photography campaigns, for example), others require historical management (cadastral plots, etc.). And finally, some are placed between the two types, such as fixed phenomenons whose attributes change over time (temperature sensors) or whose location, as well as attributes, can change over time (political frontiers, coastal boundaries). In some cases, the temporal aspect has therefore to be treated as an attribute separate from the objects and sometimes modelled as a date, an interval or a temporal range (validity period) [GUP 95].

We can distinguish three types of time concepts:

– 'Logical' or factual time indicates the dates on which the phenomenon, as stored in the database, took place (in reality);

– Time (date) of observation of the phenomenon;

– Transactional time, corresponding to the date the data was entered into the database.

From the user's viewpoint, it is the concept of logical time that is the most important, but in practice, it is the transactional time that is most often stored.

The phenomenons' temporal aspect is highly variable [PUR 00], depending on the type of phenomenon (a mountain's altitude with respect to water level in a reservoir) and the precision with which they were measured. The correct interval for confirming the validity of a database is therefore directly linked to the phenomenons which are represented therein. Similarly, the temporal consistency required between objects varies depending on the type of phenomenon: complex entities or ones with inter-relationships require good temporal consistency (topological structures, such as, for example, the road network) whereas independent elements do not require it (sign posts, etc.).

Manipulating temporal information comes down to adding the temporal dimension to the data model used and, by extension, to all the elements of the database, for example, using one or more additional 'attributes' for each entity of the database, each attribute and each relationship. In addition, to maintain a database's temporal consistency, specific mechanisms should be established to allow version-management of data. A modification such as the segmentation of a stretch of road into two parts cannot be limited to the removal of the old section and its replacement by the new ones, but should allow the modification of the characteristic of validity of the old object ('anterior', for example) and include the information that the new segments replace the old (to maintain consistency in the history). It becomes obvious that the management of time-related information requires the retention of a large amount of information and dates (modification dates, observation dates, effective dates of updates to the database) and we observe that the management of the temporal aspect can soon become complex, difficult to manage and maintain, and, above all, require large amounts of storage space. The establishment of such mechanisms should be limited as far as possible to those geomatic applications for which it is indispensable.

There exist a number of interactions between the temporal aspect and other quality elements:

– *Lineage*, which provides a lot of temporal information (sequences and processing dates);

– *Geometric precision* (for which temporal information can sometimes explain errors);

– *Semantic precision* (availability of information on the temporal validity of an attribute allows the detection of inconsistencies when suspect values change);

– *Completeness* (which should only be estimated for entities that are temporally consistent);

– *Logical consistency* (for the same reason);

– *Semantic consistency* (measuring the semantic consistency of the temporal aspects of a database allows the evaluation of the responsiveness of updates to the database with respect to changes in real phenomenons).

12.3.9 Quality criteria: difficulties and limitations

The quality parameters or criteria that have been defined partially overlap each othe which sometimes renders difficult the classification of an error (i.e., the determination of which criterion was violated). The example in figure 12.8, taken from [VAU 97], illustrates this problem:
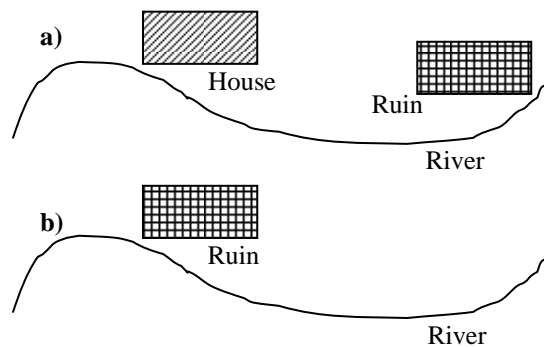

Figure 12-8. *Classification of error cases.*

In figure 12.8, the two datasets represent the same geographical area. The second (b) has one fewer item. This difference can be result from one of three different types of errors:

– An error of geometric precision (the ruin is too far to the left) added to a completeness error (the house is missing);

– A classification error, therefore of semantic precision (the house was classified as a ruin) added to a completeness error (the ruin is missing);

– A double error of temporal precision. The ruin has disappeared and the house has degraded into a ruin.

The evaluation of quality parameters is, by its very nature, useful to the user but it needs to be easily achieved. In fact, this information should be found relevant so that the users (producers and end-users) accept the limitations that they entail and understand its utility. Of course, each producer consciously wants to supply data that is as correct as possible, and each user want to acquire and use the best available information. Standardization serves as a basis for structuring and evaluating quality, but this basis is still today more oriented towards the data producer than to the data consumer.

The complexity of the standards and, above all, the difficulty in differentiating these quality elements, means that it is expensive to evaluate, store or provide the data quality in a simple and comprehensible manner. Only the evaluation of the gains arising from the use of quality information and a usage that is adapted to the users' requirements can bring home its advantages.

The use of quality criteria mentioned here is a variable depending on the organizations producing and using data. To facilitate exchange and comprehension of information on quality, standards-developing organizations have published standards which provide guidelines for using quality criteria and for the documentation of procedures for evaluating quality.

12.4 Quality and metadata as seen by standards

12.4.1 Introduction to standardization

The goal of standardization, in the meaning of decree no. 84-74 of 26 January 1984 and relating to French standardization, is to '*supply reference documents (...) solutions to problems (...) which arise repeatedly in interactions between partners (...)*.' Standardization is, above all, an activity of defining specifications in a consensual framework.

Standards emerge from a set of mandated or recognised official organizations.

The French association for standardization (French acronym: AFNOR) is the motive force behind French standardization and acts as a clearing house for official French, European and international standardization organizations, whether they are comprehensive in their scope or limited sectorally (telecommunications, electrical engineering and electricity), such as:

– International Telecommunications Union (ITU);

– European Telecommunications Standards Institute (ETSI);

– International Organization for Standardization (ISO);

– European Committee for Standardization (French acronym: CEN);

– International Electrotechnical Commission (IEC);

– *L'union technique de l'électricité et de la communication* (French National technical union for electricity and communication (UTE));

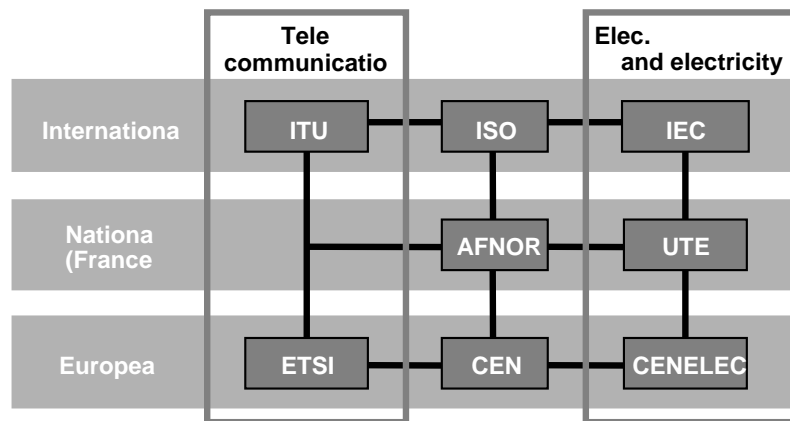– European Committee for Electrotechnical Standardization (French acronym: CENELEC).

Figure 12-9. Official standardization organizations

Around these official standardization organizations gravitate other organizations, often sectoral, self-mandated (but not necessarily less respected) that produce standards in the same consensual framework. Standardization is generally an activity that is the responsibility of organizations that have official status. The expression '*de jure* standard' is often used to designate standards.

But standards do not always result from an activity of standardization. Some specifications take a consensual character without having been designed with such a goal in mind. These specifications are called '*de facto* standards'.

These nuances around how standards are formed finally matter little. The importance of standards lies in that they provide answers to problems that arise repeatedly in inter-partner interactions. Thus, as far as quality and metadata is

concerned, one has to go beyond individual practices and rely on technical specifications having wide application in the geographical information sector.

12.4.2 Background of geographic information standards

The need to exchange geographic information was the motive force behind standardization in the domain. The first standards for exchange emanated from the defence and hydrography sectors in the 1980s:

– The military standard for exchange of geographic data 'Digital Geographic Exchange Standard (DIGEST)' [DGI 00] was established by the 'Digital Geographic Information Working Group (DGIWG)' which managed and improved it until the early 2000s;

– The exchange standard S-57 [OHI 00] was established by the International Hydrographic Organization (IHO) and is still used for exchanging nautical information destined for onboard navigation terminals;

– A little later on, national exchange standards appeared, amongst which:

  – The American 'Spatial Data Transfer Standard (SDTS)' [USG 97], which is one of the precursor standards in the domain;

  – The French exchange standard *Edigéo* [AFN 99], which was approved in 1999 after five years of testing.

These different exchange standards implemented to a lesser or greater degree the various quality components. However, their principal defect does not lie in their lack of comprehensiveness regarding these quality components but in terms of their specific implementations. Each of these standards proposes its own exchange structure within which quality information occupies a specific but also peripheral place.

The 'Content Standard for Digital Geospatial Metadata' [FED 98] of the American Federal Geographic Data Committee (FGDC) is a standard dedicated to metadata without being data-exchange centric. Its goal, defined by the presidential decree 12906 of 11 April 1994, is to capitalise and make available knowledge relating to geographic data produced by American agencies. The importance of quality information is, here too, as peripheral as in the data-exchange standards. But the standard's regulatory nature and political will in the US have led to this standard's widespread acceptance and use to this day.

A new approach to the standardization in the domain of geographic information appeared in the middle of the 1990s: one must standardize the different aspects of geographic information and then assemble these standards to respond to different needs (exchange, cataloguing, etc.). It is this new approach that the technical committee 287 of CEN (CEN/TC 287) has chosen by constructing a modular set of standards in the geographic information domain, including, most notably, an experimental standard relating to quality [CEN 99] and an experimental standard on metadata [CEN 98]. The work of CEN/TC 287 came to a premature end with the constitution of the technical committee 211 of the ISO (ISO/TC 211) in 1994.

ISO/TC 211 continued in the same vein as CEN/TC 287 but went much further. After ten years of existence, ISO/TC 211 lists more than 40 published documents of which 75% are standards or draft standards, 15% are technical specifications or draft technical specifications, and some 10% are reports. The ISO/TC 211 standards incorporate the application of new information technologies in the domain of geographic information. They create a necessary break between the relational and object-oriented eras, offering new approaches to the entire domain of geographic data. These standards are modular and, above all, extensible to respond to specific requirements of users while ensuring a sharing of standardized concepts.

In parallel to the work of ISO/TC 211, the Open Geospatial Consortium (OGC) has also established a set of standards in the geographic information domain by taking advantage of the new information technologies. The abstract standards of OGC have strongly influenced the standards of ISO/TC 211 but the originality of OGC arises from its implementation standards such as the format for vectorial-data exchange 'Geographic Markup Language (GML)' [OGC 03], the specifications for services 'Web Map Server interface (WMS)' [OGC 04a] and 'Web Feature Service (WFS)' [OGC 02] as well as in the specifications for catalogue services 'Catalogue Services (CAT)' [OGC 04b].

These implementation standards implement abstract standards of ISO/TC 211, notably those relating to the quality and metadata. In addition, these standards are generally taken up by ISO/TC 211 to be published as standards or technical specifications when they are mature enough:

　　– WMS is the subject of the draft standard ISO 19128 [ISO 04a];

　　– GML is the subject of the draft standard ISO 19136 [ISO 04b];

　　– WFS is the subject of the draft standard ISO 19142 [ISO 05a].

This trend is confirmed by a strengthened co-operation between OGC and ISO/TC 211.

Under the impetus of the project of the European directive INSPIRE, the CEN/TC 287 was reactivated in 2003 to adopt or adapt the standards of ISO/TC 211, thus affirming the importance of these international standards for the European Union. An association of European cartographic agencies and local authorities (EuroGraphics) surveyed its members on their use of ISO/TC 211 standards relating to quality. The survey's results [EGC 04] showed the clear interest that these national agencies have in these standards, but stuttering implementations demonstrated the need for a guide for implementing these standards.

The evolution of standardization of geographic information tends to encourage the joint use of ISO/TC 211 and OGC standards. This general trend does not exclude other standards from consideration, especially for those relating to the quality and metadata, even adoption of alternative technical solutions, most notably:

– The applicability of a solution that is not dedicated to geographic information should be considered before using specific solutions, even if they are of a standard character;

– The OGC and ISO/TC 211 standards should satisfactorily take into account the quality components and metadata both from the theoretical and practical viewpoints.

12.4.3 Standards relating to metadata and quality

Quality occupies a prominent and real place in the standards of ISO/TC 211, since three standards and one draft technical specification relate to it:

– The ISO 19113 standard [ISO 02] defines the principles of quality and, notably, of quality components;

– The ISO 19114 standard [ISO 03a] is dedicated to procedures for evaluating quality. It defines the ways of expressing quality measurements, either as evaluation reports or as metadata;

– The ISO 19115 standard [ISO 03b] specifies the conceptual structure of metadata. This conceptual structure takes into account the different quality components defined by the ISO 19113 standard;

– The ISO 19138 preliminary draft technical specification [ISO 04c] describes a set of quality indicators.

ISO/TC 211 is still active and other standardization documents relating to the quality could still emerge, especially for imaging requirements. The ISO 19115-2 draft standard [ISO 04d] relating to imagery metadata and the ISO 19130 draft

standard [ISO 05b] relating to sensor models extend the ISO 19115 standard. In addition, the implementation of these standards requires that other ISO/TC 211 standards, mentioned in previous section, be also considered.

OGC standards are called upon when implementing ISO/TC 211 standards and more generally, when implementing services destined for clients more, or less, specialised in geographic information.

Geographic information is, after all, primarily information. It is thus important to consider general standards relating to metadata and quality. The reference standard for generalised research applications is the Dublin Core [DCO 05] which specifies a fundamental set of 15 metadata items, such as the title, the summary, the date, etc., useful for describing different types of data.

This listing of standards relating to quality and metadata will not be complete if mention is not made of standards in the ISO 9000 series [ISO 00]. They relate to the management of quality and are fully applicable to the production of geographic data. They permit the incorporation of the evaluation of the quality of geographic data in the more general context of quality control and assurance.

From a strategic viewpoint, the four ISO/TC 211 standards relating to the quality and metadata are therefore used as a complement to the implementation standards in the domain of geographic information as well as to the general standards such as Dublin Core and the ISO 9000 series.

12.4.4   Theoretical analysis of ISO/TC 211 standards

12.4.4.1   The ISO 19113 standard

The ISO 19113 standard focuses on the description of quality parameters. It also calls upon other quality components such as:

– The use of data in terms of intention (purpose of the data) as well as feedback on the use of data;

– The lineage.

The ISO 19113 standard is mainly descriptive. It delegates the definition of the conceptual structure of quality information to the ISO 19115 standard.

The ISO 19113 standard takes into account the main quality parameters (completeness, logical consistency, semantic precision and positional precision) and offers as a supplement a parameter of temporal precision. However, the ISO 19113

standard does not address the concept of specific quality but authorises the creation of quality elements outside the standardization framework. Such elements can be considered as representing specific quality.

The ISO 19113 standard also proposes a sub-classification of the usual quality parameters:

– The completeness is broken down into omission and commission;

– The logical consistency is broken down into conceptual consistency, consistency of the domain of values, consistency of format and topological consistency;

– The positional precision is broken down into absolute (or external) precision, relative (or internal) precision and positional precision of gridded data[5];

– The temporal precision is broken down into precision of time measurement, temporal consistency and temporal validity;

– The semantic precision is broken down into classification precision, precision of non-quantitative attributes and precision of quantitative attributes.

This sub-classification is of interest because the boundaries between the different parameters are typically difficult to define:

– By how much is the measurement of temporal consistency linked to temporal consistency rather than to logical consistency?

– By how much is the consistency in the domain of values linked to logical consistency rather than to semantic precision?

These questions illustrate the risks of inconsistency between different implementations of the ISO/TC 211 standards.

Finally, the ISO 19113 standard broaches the subject of some aspects of 'metaquality' without mentioning it outright and without defining the concept.

---

[5] The relevance of this criterion is debatable since, on the one hand, the type of data representation, ideally, does not impact the classification of quality components and, on the other, the differentiation between relative and absolute precisions is as necessary for gridded data as for vectorial data.

12.4.4.2 The ISO 19114 standard

The ISO 19114 standard specifies a methodology for evaluating quality whose result can either be quantitative or be limited to an indication of data conformity vis-à-vis given specifications, which can either be product specifications or specifications of the requirements of a set of users in terms of data use.

The ISO 19114 standard also defines two methods for evaluating quality:

– A direct method of comparing data with other data, either within the dataset (in this case the method is direct and internal) or external data;

– An indirect method of deducing or estimating a measure of data quality from metadata and, more specifically, from lineage information or data usage.

Whichever be the method used, the evaluation can bear on all or part of the dataset, and can be conducted in a systematic manner on the entirety of the selection or by sampling on a representative subset of the selection.

Finally, ISO 19114 specifies that the evaluation result can be expressed in the form of metadata and/or quality evaluation reports. The standard authorises an aggregated expression of evaluation results within the metadata; summary results are used rather than detailed results. In such a case, an evaluation report is asked for. Quality evaluation reports are covered briefly within annexure I of the ISO 19114 standard but no conceptual structure is offered. The result is that this aspect of the standard is often overlooked.

12.4.4.3 The ISO 19115 standard

It is paradoxically the ISO 19115 standard which formalises, in UML, the quality concepts defined in the ISO 19113 standard and the expression of results of quality evaluations conforming to the methodology defined in the ISO 19114 standard. The experts in the field of quality must feel that their ideas have been appropriated by the metadata experts!

The ISO 19115 standard is organized in metadata sections. The quality information is mainly found in one dedicated section. Some information on the use and timeliness of the data appears in the identification section. Information relating to updating of data is to be found in the section devoted to maintenance. This structure upsets the quality experts, but it ensures a certain consistency of metadata and its use by the users.

A collection of metadata can include several sets of quality data with each applicable to a selection of the dataset or, more generally, of the metadata *resource* object. Each set of quality data can consist of a set of lineage information and a set of quality-evaluation reports.

The lineage of the resource can consist of information on the sources used and, if applicable, supplementary information on the procedures applied to these sources, but can also very well be limited to a simple textual description. Information relating to the source can be relatively detailed without being limiting. The resource zone covered by a source can be indicated clearly. On the other hand, it is inconvenient that one cannot specify the resolution of a source image but can express the scale of a cartographic source.

Each quality evaluation report, called 'quality element', is the expression of the results of evaluating a quality indicator. Some aspects of the conceptual definition of these reports are to be noted:

– The quality elements are subject to a classification that follows the classification of quality parameters proposed by the ISO 19113 standard, thus forcing the indicator to relate to a quality parameter;

– The requirement of taking into account specific quality parameters forces the extension of the proposed classification, which, in practice, is somewhat impractical;

– The evaluation result can be expressed in a quantitative and/or a qualitative manner by a simple indication of conformity with a product specification or user specification;

– It is not possible to describe the sampling used for evaluation without extending the ISO 19115 standard;

– The designers of the standard did not want to exclude any type of quantitative result (covariance matrix, for example), thus rendering the expression of a quantitative result somewhat difficult and its use practically impossible in most general cases;

– It is not possible to express an evaluation result in the form of homogenous quality zones. This limitation forces the users to make a dangerous mixture of the ability to select evaluated geographic data and the need to express these zones over which the evaluation result is constant.

However, these problems do not diminish our interest in ISO 19115. Moreover, the establishment of an amendment process for ISO 19113 and ISO 19115 is being discussed within ISO/TC 211 to resolve them.

12.4.4.4 ISO 19138 preliminary draft technical specification

On the one hand, the future technical specification ISO 19138 defines information necessary to describe a quality indicator and, on the other, provides a description of a list of quality indicators. These standard quality indicators are, beyond question, a factor for interoperability, but the users' requirements are such that each community should be allowed to describe its own indicators. By standardising the manner of describing quality indicators, the ISO 19138 standard will allow the emergence of community indicator registries, thus simplifying the approach to quality by organizations for whom the production of geographic data is a secondary activity and who do not necessarily have the means to manage these still-esoteric matters. However, the implementation of such registries will have to wait until this standard, currently under development, has attained a sufficient maturity.

12.4.5 Standardized implementation of metadata and quality

12.4.5.1 Preamble

Issues of standardization are pertinent only during interactions between different actors. By relying on the ISO quality and metadata standards, actors in the domain of geographic information can share credible concepts and principles. To share knowledge of data quality, one has to go further and actually implement these standards.

ISO/TC 211 includes infrastructure standards such as the ISO 19019 standard which bears directly on the implementation of ISO/TC 211 standards in two axes:

– The model for geographic data exchange impacts the relationship between metadata and the resources concerned by this metadata;

– Only one semantic model (General Feature Model – GFM) governs the manner in which metadata and quality are applied at the level of defining geographic objects.

12.4.5.2 The model for exchange by transmission

In the *model for exchange by transmission*, the user invokes services that respond on a case-by-case basis to his queries formulated through a client application. The CAT standard of OGC is the standard reference for services for querying metadata warehouses. It defines the interface between the client applications and the

cataloguing service that delivers the metadata. A cataloguing service can be the client of another cataloguing service via the CAT standard.
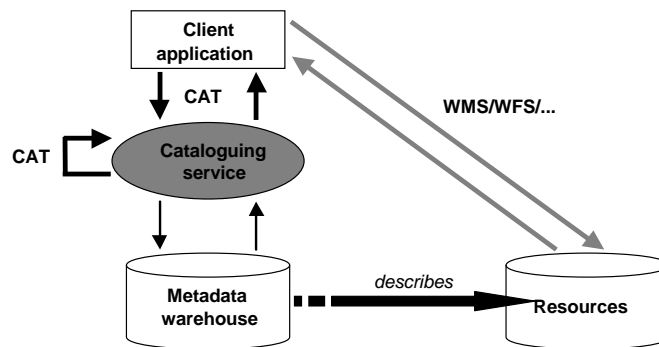

Figure 12-10. *The context for metadata exchange by transmission*

Client applications access resources that are described by the metadata using the OGC access services (WFS, WMS, etc.) or any other non-standard solution.

By default, the CAT standard is based on a subset profile of the Dublin Core which is designed for wide-ranging general applications. But CAT can meet the expectations of specialist applications by offering the possibility of specifying the level of metadata detail expected by the client in the formulation of his query. Within such a framework, it is possible to expect ISO 19115 metadata coded in XML and conforming to the ISO 19139 standard.

12.4.5.3 Data transfer

In the traditional *model of exchange by transfer*, the data supplier creates a batch of data which is transferred to the user with the information necessary for its use, most notably its metadata. Several batches of data can be assembled into a set of batches of data having their own metadata. A batch of data contains geographic objects with their own metadata (see section 12.3.5.4).

The concepts involved in the model of exchange by transfer are introduced in the ISO 19115 standard and are detailed in the ISO 19139 standard which proposes XML encoding of metadata and associated resources.

12.4.5.4 Metadata of geographic objects

In the sense of the ISO 19109 standard, the relationship between geographic objects and their metadata is ensured through particular attributes of geographic object classes:

– The attributes of type metadata hold complete sets of metadata relating to the concerned geographic objects. Such attributes are accessed by ISO 19115 metadata-consulting applications, the metadata resource being the geographic object with the attribute;

– The attributes of type quality hold the quality elements in conformance with the ISO 19115 standard.

The GML standard, subject of the ISO 19136 draft standard, is recommended for encoding geographic objects and their metadata, for which it is based on the ISO 19139 standard. Moreover, GML is the format used by the WFS and WMS services for accessing data. It can also be used as a format for geographic data in a context of exchange by transfer, conforming to the recommendations of the ISO 19139 draft standard.

12.5 Conclusion

This chapter has presented the growing importance of the concept of quality for the exchange and distribution of geographic data. It has presented the quality components followed by the standards relating to this topic. The most widely held and used point of view is that of the data producer. Quality and the growing importance of metadata are normally acknowledged and adopted by the producers, whether they be institutional or casual. For the users, access to this knowledge is essential. However, we must admit that the 'fitness for use' concept, even though not recent, is not really widely implemented. The community must, in the end, develop major types of targeted applications to be able to have the necessary context for evaluating this concept. Tools, indeed even evaluation standards, are yet to be defined, but this new challenge is an unavoidable stage in the evolution of geographic information systems.

12.6 Bibliography

[AAL 99] AALDERS H. J. G. L. Geo-quality and its Impact on GIS Applications. *Proceedings of 21st Urban Data Management Symposium.* Venice, Italy, April 21-23 1999, pp. VI.2.1-VI.2.11

[AFN 99] Association Française de Normalisation. Traitement de l'information – Echanges de données informatisées dans le domaine de l'Information Géographique. AFNOR, 1999. 234 p. NF-Z52000

[AZO 96] AZOUZI M., MERMINOD B. Qualité des données spatiales. *Revue Mensuration, Photogrammétrie, Génie Rural*, Switzerland, 1996.

[AZO 00] AZOUZI M. *Suivi de la qualité des données spatiales au cours de leur acquisition et de leurs traitement*s. Thesis presentation, EPFL, 2000. 107p.

[BRA 95] BRASSEL K., BUCHER F., STEPHAN E., VCKOVSKI A. Completeness. *Elements of spatial data quality.* Edited by S.C. Guptill and J.L. Morrison. Oxford: Elsevier, 1995. pp. 81-108

[CLA 95] CLARKE D. G., CLARK D. M. Lineage. *Elements of spatial data quality.* Edited by S.C. Guptill and J.L. Morrison. Oxford: Elsevier, 1995. pp. 13-30

[CEN 98] Comité Européen de Normalisation – Comité Technique 287. Information Géographique – Description des données – Métadonnées, March 1998, 62 p. XP ENV 12657:1998, NF Z52-007

[CEN 99] Comité Européen de Normalisation – Comité Technique 287. Information Géographique – Description des données – Qualité, August 1999, 60 p. PrENV 12656:1999, NF Z52-006

[DAV 97] DAVID B., FASQUEL P. *Qualité d'une base de données géographique: concepts et terminologie*. Saint-Mandé (FR.): IGN, 1997. 40 p. Bulletin d'information de l'IGN no. 67

[DCO 05] Dublin Core Metadata Initiative (DCMI). DCMI Metadata Terms http://dublincore.org/ (last consulted: 31 January 2005)

[DGI 00] Digital Geographic Information Working Group, Digital Geographic Information Exchange Standard (DIGEST) – September 2000, Part 1 – 72 p, Part 2 – 643 p, Part 3 – 70 p, Part 4 – 348 p.

[EGC 04] Eurogeographics – Expert group on Quality. Use of the ISO 19100 Quality standards at the NMCAs – Results from a questionnaire – Version 1.0, 17 December 2004, 15 p.

[FED 98] Federal Geographic Data Committee. Content Standard for Digital Geospatial Metadata, June 1998, 90 p. No. FGDC-STD-001-1998

[GOO 95] GOODCHILD M. F. Attribute Accuracy. *Elements of spatial data quality*. Edited by S.C. Guptill and J.L. Morrison. Oxford: Elsevier, 1995. pp. 59-80

[GUP 95] GUPTILL S. C. and MORRISON, J. L. (ED.). *Elements of Spatial Data Quality*. Oxford: Elsevier, 1995. 202 p.

[ISO 94] International Organization for Standardization. *Management de la qualité et assurance de la qualité - Vocabulaire*. ISO, 1994. 50 p. Standard ISO 8402

[ISO 00] International Organization for Standardization. Systems for Quality Management – Set of standards ISO 9000: ISO 9000, ISO 9001 and ISO 9004. 140 p. ISO 9000:2000

[ISO 02] International Organization for Standardization – Technical Committee 211. Geographic Information - Quality principles. ISO, 2002, 36 p. ISO 19113:2002

[ISO 03a] International Organization for Standardization – Technical Committee 211. Geographic Information - Quality Evaluation Procedures. ISO, 2003, 71 p. ISO 19114:2003

[ISO 03b] International Organization for Standardization – Technical Committee 211. Geographic Information - Metadata. ISO, 2003, 148 p. ISO 19115:2003

[ISO 04a] International Organization for Standardization – Technical Committee 211 – Working Group 4. Geographic Information – Web Map Server interface – Draft International Standard, February 2004, 83 p. ISO/DIS 19128

[ISO 04b] International Organization for Standardization – Technical Committee 211 – Working Group 4. Geographic Information – Geography Markup Language – Committee Draft, February 2004, 580 p. ISO/CD 19136

[ISO 04c] International Organization for Standardization – Technical Committee 211 – Working Group 7. Geographic Information – Data Quality Measures, Preliminary Draft Technical Specification, November 2004, 88 p. ISO/PDTS 19138

[ISO 04d] International Organization for Standardization – Technical Committee 211 – Working Group 6. Geographic Information – Metadata – Part 2: Metadata for imagery and gridded data, Working Draft, September 2004, 47 p. ISO/WD 19115-2

[ISO 05a] International Organization for Standardization – Technical Committee 211 – OGC Joint Advisory Group. Geographic Information – Web Feature Service – New Work Item Proposal, February 2005, 135 p. ISO/TC 211 N1765

[ISO 05b] International Organization for Standardization – Technical Committee 211 Working Group 6. Geographic Information – Sensor data model for imagery and gridded data, 2nd Committee Draft, February 2005, 160 p. ISO/CD2 19130

[MOE 87] MOELEERING, H. *A draft Proposed Standard for Digital Cartographic Data.* National Committee for Digital Cartographic Standards, American Congress on Surveying and Mapping, 1987. Report no. 8, 176 p

[OGC 02] Open Geospatial Consortium. OpenGIS® Implementation Specification – Web Feature Service – Version 1.0.0, May 2002, 105 p. OGC 02-058

[OGC 03] Open Geospatial Consortium. OpenGIS® Implementation Specification – Geography Markup Language (GML) – Version 3.0, January 2003, 529 p. OGC 02-023r4

[OGC 04a] Open Geospatial Consortium. OpenGIS® Implementation Specification – Web Map Service – Version 1.3, August 2004, 85 p. OGC 04-024

[OGC 04b] Open Geospatial Consortium. OpenGIS® Implementation Specification – Catalogue Services Specification– Version 2.0, May 2004, 192 p. OGC 04-021r2

[OHI 00] International Hydrographic Organization. IHO Transfer Standard for Digital Geographic Data – Edition 3.1 – November 2000, 114 p. Special Publication No. 57 (S-57)

[PUR 00] PURICELLI A. *Réingénierie et contrôle qualité des données en vue d'une migration technologique*, Thesis presentation, Institut National des Sciences Appliquées de Lyon, December 2000.

[SAL 95] SALGE F. Semantic Accuracy. *Elements of Spatial Data Quality.* Edited by S.C. Guptill and J.L. Morrison. Oxford: Elsevier, 1995. pp. 139-152

[SER 00] SERVIGNE S., UBEDA T., PURICELLI A. and LAURINI R. A Methodology for Spatial Consistency Improvement of Geographic Databases. *GeoInformatica,* 2000, Vol 4 No 1, pp. 7-34

[UBE 97] UBEDA T. *Contrôle de la qualité spatiale des bases de données géographiques: cohérence topologique et corrections d'erreur,* Thesis presentation, Institut National des Sciences Appliquées de Lyon, December 1997.

[VAU 97] VAUGLIN F. Statistical Representation of Relative Positional Uncertainty for Geographical Linear Features. *Data Quality in Geographic Information: From Error to Uncertainty*, Edited by M. Goodchild and R. Jeansoulin. Paris: Hermès, 1998, pp. 87-96

[USG 97] United States Geological Survey – National Mapping Division. Spatial Data Transfer Standard, American National Standards Institute Ed., 1997, Draft for review. Part 1, 193 p., Part 2, 53 p., Part 3, 29 p., Part 4, 38 p.
http://mcmcweb.er.usgs.gov/sdts/whatsdts.html (last consulted: 31 January 2005)