Frank, Andrew U. "Metamodels for Data Quality Description." In *Data Quality in Geographic Information - from Error to Uncertainty*, edited by Robert Jeansoulin and Michael Goodchild, 15-29. Paris: Editions Hermès, 1998.

METAMODELS FOR DATA QUALITY DESCRIPTION

Andrew U. Frank
Dept. of Geoinformation
Technical University Vienna
frank@geoinfo.tuwien.ac.at

Abstract

Data quality descriptions are crucial, but methods to produce and use them have not significantly improved during the past 10 years. Current quality descriptions are from the perspective of the producer of the data, not the user. Actual quality descriptions are mostly verbal and not suitable for rapid comparison with a required standard to make a decision about 'fitness for use' of a certain dataset for a task. This limits business with geographic data over the net.

The paper introduces the concept of a metamodel as a framework to compare data quality from a producer and a user perspective in a single model. It is based on category theory and morphisms, which link the model of reality with the model of the GIS data, and their collection and use. The achieved quality of a decision based on using the data can be derived.

It is shown that data quality descriptions are dependent on the intended use of the data. A 'use independent', generic data quality description is not possible. Fortunately, a large set of GIS functions demand the same data quality description, therefore not every potential use requires a different data quality description of a data set.

1 Current Data Quality Descriptions Are Inadequate

Data quality descriptions are crucial for a flexible use of GIS. They are the key to the development of a sizable commerce in GIS data. Data quality descriptions are necessary for differentiated marketing strategies, and in particular product differentiation (Frank 1995; Frank 1996) . They are also the key to limit liability of data producers. Current practice of data quality description is inadequate and does not help users to decide if a potentially useful dataset should be acquired and used.

Overall, data quality descriptions have not improved much in the past 10 years. The publication of the data quality description in the Spatial Data Transfer Standard (Morrison 1988) and the report of the NCGIA Specialist Meeting in Santa Barbara 1988 (Goodchild and Gopal 1989) document the research frontier then. The list of parameters to describe geographic data has not changed in the past 15 years; there is nearly no difference between the parameters listed in (Robinson and Frank 1985) and the lists published today (Stanek and Frank 1993). There is no progress to define quantitatively the quality of GIS data - beyond positional accuracy for well-defined points. Data exchange standards and the practice of data producers rely heavily on lineage description as replacement for an effective, objective description (Chrisman 1991).

There is no sizable discussion of data quality transfer functions which link data quality of inputs to the quality of the results (for a case study see (Zeitlberger 1997). This paper first reviews the overall situation of data sharing between organisations where data quality descriptions are necessary. It then reports on a case study assessing the currently available data quality descriptions for a number of data collections and shows that these descriptions are formulated from a data producer perspective, but are not suitable to answer the potential data user's questions.

In the second part, a formal model of the relations between reality, geographic data in a GIS and its use is set up. This metamodel allows to describe the situation using morphism. This model contains

- an observation function linking reality to the data collected, and
- the decision process which links to a decision, which is termed ‘function of interest’.

A user’s query is a typical example of a function of interest which extracts some data from a database and the result is used in the decision process. For a data quality description to be usable, it must be possible for the user to link the data quality statement given for the data used to the quality of the results it deduces from the database. This presupposes a ‘data quality transfer’ function related to the function of interest. Each function of interest will in general have its own data quality transfer function.

For each function of interest, a corresponding error function can be established and the relations between observation function, function of interest and the error functions involved described. The relations can be formalized and a simple example is given in this paper.

This paper deals with the data quality description, which forms the input into the data quality transfer function of a user and demonstrates that there is a linkage between the data quality description and the ‘function of interest’ (respectively, the data quality transfer function). A single data quality description is not sufficient for all functions of interest in a GIS.

From this follows that data quality descriptions must be made to suit the intended use of the data. A fully general data quality description which is useful to decide on fitness for any potential use, is not possible. But neither is a specific data quality description for every possible use necessary: large classes of operations on spatial data can use the same data quality descriptions. The determination of these classes is an important open research question.

2 Why Data Quality Descriptions?

2.1 Underlying assumption: data sharing

Data quality descriptions are only necessary if data is collected by one organisation and used by another (Frank 1992). If - as customary in the past - the same organisation collects and uses the data, a description of the data quality is not necessary. The user can directly influence the collection process and adjust it so the data are suitable for his needs, and for the process of collecting and processing minimal cost will accrue. If a direct feedback loop between data user and data collector within an agency is missing, then an explicit consideration of data quality is necessary to minimize data collection cost, while fulfilling the data user’s requirements.

When one organisation collects data and another one uses them, the producer must describe the quality of the data collected and the user prescribe the quality required for the task. From this follows that the data are fit for the intended use if the quality of the data as produced is better than the required quality. Unfortunately, today a decision about the fitness for use cannot be made without the user having a very significant understanding of the processes used for the data collection; typically a discussion between data producers and users is necessary to reach a conclusion about fitness for use.

2.2 Regular GIS assumption: sharing reduces cost

The discussion pointing to the need for data quality description is situated in the regular GIS assumption, which is that data is expensive to collect and maintain, but data once collected can be used for many purposes. Cost reduction in avoiding duplication in data collection initially is important, but cost reduction later by avoiding the duplication of the maintenance of the data is usually much larger.

Data sharing is a crucial concept for GIS (National Research Council 1980). The argument is more complex than the simple argument for reduction in cost of administration as it was initially put (Clapp, Moyer, and Niemann 1988; Gurda et al. 1987). If each agency uses its own dataset - collected and maintained individually - these data sets will necessarily differ within the error margins set for their collection process. This results in problems:

- in border line cases, two decisions based on these data collections may contradict each other; citizens affected by these decisions become aware of the errors, loose faith in the administrative decision process, etc.
- the integration of other data related to the spatial data sets cannot be integrated quickly, because the spatial reference objects are not the same.

2.3 Data quality description crucial for geographic data business

Data quality descriptions should assure the user that the data are fit for the intended use (‘fitness for use’ (Chrisman 1983)). If the quality described is better than required, the data can be acquired and used for the intended task. Data quality descriptions are necessary to facilitate the emerging commerce in geographic data:

The data quality description is part of the metadata which is part of the information a data producer makes available to prospective users of his data. Clear description of data quality brings:

- More use of data: The decision about fitness for use by a possible user (buyer) of data can be made quickly and objectively. As more data is made available using the Internet, users (or programmed agents on behalf of the user) must be able to make automatically a decision which data set can be used for a particular task (Voisard and Schweppe 1996).
- Limits to liability: If a data set is labelled by the producer with an objectively measured data quality, users who use the data for uses which cannot be achieved with this quality are clearly warned. This effectively limits liability by the producer for damage resulting from errors in the data larger than acceptable within the stated quality. Data quality descriptions are necessary to assign liability - who is liable for damage occurring as a consequence of using the data? Did the producer deliver data which was of lower quality and thus caused the damage (and incurs possibly a legal liability) or were the data according to the quality asserted, but errors in processing caused the damage (for which the user would be responsible)?

3 Critique of Current Description Methods

Data quality descriptions should be

- independent of production method,
- operational, and
- quantitative.

By the first we mean that the description of the data quality must not refer to the production method, but use a neutral formulation independent of how the data were produced. Second, a data quality description is operational if it can be used in a formalised (automated) process and does not depend on human interpretation of the terms. Third, the data quality must be measured on an (at least) ordinal scale, which allows the comparison of the quality of two data sets.

3.1 Case study - metadata according to standards

A class of surveying engineering students collected metadata on commercially available data sets in Austria, described the metadata according to the CEN (Comité Européen de la Normalisation) metadata standard, and collated the metadata in a commercial database (Timpf, Raubal, and Kuhn 1996).

The goal of the study was twofold. First, the students gave feedback to the responsible working group at CEN about the usage of the metadata standard. Second, they assessed if potential users of the metadatabase can understand the information provided and if it is sufficient for decision-making by professional users.

The usability of the metadata was low: they found that the metadata described the data from the data producer's point of view and did not help the user to make a decision about the suitability of a data set for an intended task (Timpf and Frank 1997). Users need information on a higher level of abstraction, the information given often was too detailed and too confusingly presented. Most importantly they need to know what operations are supported by the data.

3.2 Data quality descriptions are producer oriented

The data quality descriptions are most often provided as 'lineage', which describes the process that was used for the collection and processing of the data. This implies a very detailed and accurate description of the quality of the data. The description is objective as the same process can be duplicated and should result in a data collection with similar quality characteristics.

The description of data quality as lineage is easy for the producer - it is knowledge which is available: one just describes what one has done. It shifts the burden of the interpretation of the data quality description and the decision about fitness for use to the potential user of the data, who must make the connection between the production method - unknown to him - and his intended use.

3.3 Data quality descriptions are not operational

A description can be called operational if there are standardized procedures, which can be used to determine the data quality values. These can be used without requiring interpretation. Operational methods are described in various standards to measure hard to determine values for 'noise production of a car', 'intellectual ability' etc. In every case, a well defined set of observation methods in a completely determined environment are used to assess the property, e.g., the procedures used to determine the SAT (Scholastic Aptitude Test) scores for entering students. The result of operational procedures are comparable, even if the individuals measured are incomparable.

Data quality descriptions using lineage descriptions are not operational: two data sets can have very similar characteristics, but result from different data collection efforts (e.g., photogrammetry or field survey) and their lineage descriptions are very different. The comparison of lineage descriptions is difficult and requires intimate knowledge of the different data collection methods and the applied techniques, instruments used etc.

Given a dataset of unknown origin, a lineage description cannot be produced. This implies that a lineage description given cannot be tested by the user independently of the data producer; one can only check the production records to see if the stated method was followed correctly.

3.4 Data quality descriptions are not quantitative

Data quality descriptions should be quantitative measures of the quality of the data. This is necessary so that a user can decide if the quality provided is better than the minimal quality required for a particular use. For a decision about fitness for use, qualitative quality description on an ordered scale is sufficient. To allow the propagation of data quality through the potential user's data analysis, data quality measures should be on a ratio or absolute scale (Stevens 1946).

Operational, quantitative measures for data quality are only given for the positional accuracy for sharply defined points (RMS error); statistical methods to determine a sample etc. are well known (Cressie 1991). These quality descriptions can be used to predict the quality of derived values, applying the law of error propagation.

4 Data Quality Description as ‘Product Specification’

Data quality descriptions are like other product specifications: They describe the property of the goods to be exchanged (in this case data) such that the user can decide if the goods are ‘good enough’ for the intended use and the producer can point to the asserted properties. The limits are important in case a problem occurs where the user claims that the good caused damage or was faulty and had to be replaced.

Product specifications are typically written as limits of the intended use or describe simple properties of the good which are of importance to the user: operating ranges for temperature or humidity, weight of the product, speed of processing etc.

The specifications are written

- in terms relevant for the use of the product rarely do they describe the methods of manufacturing the product,
- the specifications are quantitative values measured with an accepted method, such that a decision can be made if the product is within these limits or not.

Both these points are not fulfilled for data quality descriptions.

5 A Metamodel as a Framework

Data quality seems to be an elusive concept. The clear idea of a ‘fitness for use’ decision is hard to operationalize as a comparison of two figures on an ordinal scale.

The difficulty with the formalization of data quality is due to the definition of data quality as ‘correspondence to reality’. Data of high quality corresponds well to reality, for data of low quality, the deviation from reality is larger. The approach here is centered around functions and the composition of functions. Category theory (Asperti and Longo 1991; Walters 1991) is ‘algebra with functions’. The objects are functions and the only operation is composition. $h = f . g$ means the function h which results from applying g to a set of data and applying f to the result; it can be written as $h (x) = f (g (x))$. Category theory often uses diagrams, which show functions as arrows leading from a domain to a co-domain. Diagrams are said to commute if $f . g = h . j$ (see Figure 1).

5.1 The GIS model

The model implied is characterized by the following diagram:

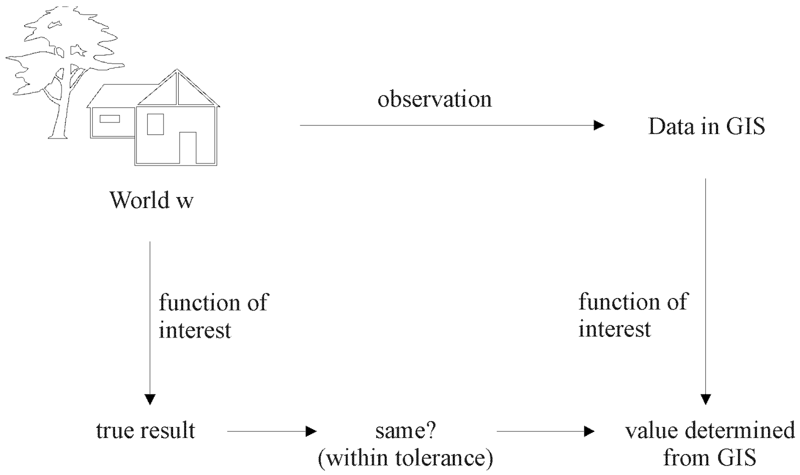

Figure 1 - The GIS model

Some value is necessary for a decision - e.g., the distance between two points. In lieu of measuring the distance in the real world, previously collected data describing the real world is used to calculate the distance. The value determined from the collected data must correspond with the result in the real world. It is easy to check the result, and often such checks occur automatically, e.g., if a plan for a building is staked out in the building process. The data in the GIS are usable, if the error between the value determined based on the data corresponds within an acceptable tolerance to the true value.

The real world is not a formal concept and thus a formalization based on the GIS model is not possible, a higher level model - termed metamodel - is required.

5.2 The Metamodel for GIS

The metamodel for GIS is a model of the GIS model: it consists of an abstract, formalized model of the real world, a model of the data, the models of the observation and correspondence processes which link the world with the data and the models of the function of interest of a potential user (see Figure 2).

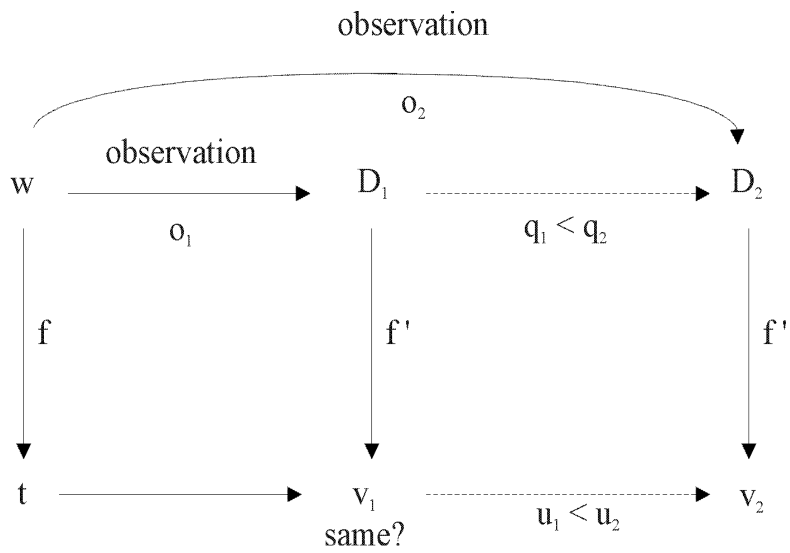

Figure 2 - The metamodel for GIS

The formalized model of the world is not known and need not be known - it is only necessary that formal models of the observation processes can be established and that the difference between data and true value can be expressed. The model is ‘categorical’ as it deals with the functions linking the models and makes statements about the composition of these functions.

In the following formulae, letters will be used with the meaning: *W* for world, *D* for data, *o* for observation operation, *f* for the function of interest of the user, *q* for the function to determine the quality of the result of *f*, *v* for values, *e* for errors. The data are the result of the observation of the world, the result of interest to the user is computed with the function *f* from the data.

$$f\,(o_1\,(W)) = f'(W)$$
$$dist\,(p_1, p_2) =\sim dist\,(P_1, P_2)$$

The function of interest computed in the model is (within a tolerance) the same as the measured value. For a concrete example assume that the distance between two points in reality (P_1, P_2) is needed; it can be measured as $dist\,(P_1, P_2)$ or it can be computed using the coordinate data for the two data sets (p_1, p_2), using Euclid’s formula. p_1, p_2 represent the result of observing P_1 *and* P_2 $(p_1 = o_1\,(P_1), p_2 = o_1\,(P_2))$.

The actual functions and the actual values are not important in this model. We assume only that functions can be composed i.e. that the result of an observation can be input into the function of interest etc. and we are interested that the diagram ‘commutes’, i.e. that the value derived from the observed data is, within tolerance, the same as the value directly measured.

This is, incidentally very similar to the approach taken in statistics: the world is represented by a *'true' value* - which cannot be observed in most cases - and *processes* to observe the true value with certain behavior, typically normal distribution of error.

5.3 Observation functions with error

The observation of the world is with error and the result of the observation is a reduced, generalized and abstract model of reality. In this very abstract model, we call error any difference between the world (the true value) and the observed value. The observation function o maps the world to data $W > D$ and it introduces an error e.

With each data value v for any property of an object in the world or in the data an error e is associated. The world has obviously the error 0

$$q\ (W) = 0$$

The observation function consists of two functions, one mapping values (o), one mapping the errors (o')

$$D_v = o\ (W)$$
$$D_e = o'\ (W)$$

There are no assumptions made about the type of data, the type of observation or the kind of data quality considered. Data quality can be described with a set of values (for example, to characterize attribute quality, update level etc.) and the error function in the observation function then consists correspondingly of a vector of functions, each computing the corresponding term.

5.4 The user's function of interest with error

The function of interest, for example, the distance calculation formula, must be extended with an error propagation function, which takes the error term associated with the data and calculates the error term of the result.

$$v = f\ (D) = f\ (o(W))$$
$$e = q\ (D) = q\ (o'\ (W))$$

5.5 Decision about fitness for use

The computed value of interest v can be compared with the true value t. The true value can be observed with an observation process o_2 with the corresponding error function o_2', resulting in an observed value t with the error u.

$$t = o_2\ (W)$$
$$u = o_2'\ (W)$$

5.5.1 When is the GIS an acceptable model?

The GIS is a 'good' model of reality within the stated data quality if the value computed from the model does agree with the true value, within a tolerance which is a function of the data quality. In general, the value with which one compares is measured with a process which introduces error, and a more complex comparison is necessary. This function compares the error terms and values: the set of possible values for v and t are compared and a level of accordance is determined and compared with the level desired.

5.5.2 Fitness for use as a comparison of data quality

The GIS is usable if the results produced are good enough for the decision process. Assume that the decision process requires a result v' of quality e'. A dataset D is 'fit for use' if the error e resulting from using D is $e < e'$. In order to make this decision quickly, the potential user must convert his desired quality of the result back to the quality of the input data D_e: the desired quality of the input data D_e results from the inversion of $e = q\ (D)$ giving $D_e' = q\ (e)$. A dataset is then 'fit for use' if the data quality of the data set D_e is better than the worst acceptable quality D_e' (i.e. $D_e < D_e'$). A potential user can then quickly compare his desired data quality with the available quality and make his decision.

The data quality description must be selected such that the data quality transfer function preserves order, which means that for any two data sets with errors D_e' and D_e" the quality of the results for the same 'function of interest' *f* is

$$\text{if } e' = f(D_e') \text{ and } e'' = f(D_e'') \text{ then } D_e' \leq De'' \text{ implies } e' \leq e''$$

6 Data Quality for Spatial Data

The model developed is fully general and does not imply that the data describe spatial objects or situations. For spatial data, the model can be made more specific:

- space consists of spatial elements,
- observation functions group (typically nearby) elements (Tobler 1979b) and determine a single value for the group of elements.

6.1 Model of space

For modeling purpose the model of space for the world and for the observed representations of the world have the same structure:

Space is modeled as set of spatial objects - which will be called resels (Tobler 1979a), which are jointly exhaustive and mutually disjoint (JEPD). For each of these spatial objects a value is known (or observed). For each resel the area is known.

The resels are arranged in a data structure *S*, which allows the traversal of all resels to apply a certain function (map) and to merge resels.

$$S = a_1, a_2, a_3, \ldots, a_n$$
$$\Sigma a' = a_1 + a_2 \ldots + a_n$$

This model of space is not the most general one - it does not include values given as functions over the space, (this is left to future work) - but covers a very large set of situations and includes at least raster and vector models as customary used in GIS.

6.2 Model of Observation

A spatial observation function determines the value of the observed quantity for an area which is larger than a single resel. It assumes that there exists some grouping function, which groups sets of resels (JEPD). The grouping function returns a set of resels (Bruegger 1995). The function 'instant-field-of-vision' determines then a generalized value for this group of resels.

The concept of an 'instant field of vision' has been applied to the discussion of error treatment by Bruegger (Bruegger 1994). It mimicks the human visual system, which perceives a locally aggregated form of the real world. This is usually an averaged value of the values encountered in this field of vision. This is directly related to the concept of an averaging filter for convolution (Horn 1986).

This concept is essentially spatial because it relies on spatial correlation. Only if the values within an instant field of vision are highly correlated, the resulting picture is informative. A situation with low spatial autocorrelation, seen through an instant field of vision aggregation (which is blurring), results in an uniform 'gray' image and does not inform.

7 Data Quality Description for Area Calculation

As a simple example, a user of data is interested in the calculation of the area. We will prove that the area of the resel is an appropriate data quality measure which allows the prediction of the error of the area calculated.

7.1 Observation

Without restriction of generality, the world is represented by a very fine raster and geographic regions are represented by values for each resel to be in or out. The observation function, as described above, considers all resel values within the field of vision and results in In or Out if all resel are inside (outside) and Maybe if both In and Out resels are found.

For a description of the data quality, the area of the resel is given. The observation process sums the area of the cells in the group.

7.2 Calculation of area

The calculation of the area is the sum of all the resels and 0 otherwise. The error measure is then the sum of the area of the resel with value Maybe.

The most likely value for the area is then the mean between In and (In + Out) resel and the true value must lie between the values (sum In) and (sum In + sum Maybe).

7.3 Proof for properties of a data quality value

It is necessary to prove that this is

- an operational method to determine data quality,
- a data quality value and allows to predict the quality of the result,
- the data quality measured on an ordinal scale, for which the data quality transfer function preserves order.

7.3.1 Data quality is predictable

To show that data quality is predictable, given the value of the data quality of the data, means to show that the diagram in Figure 2 above commutes (within the tolerance). In particular, from a given observation function follows a determined data quality. The instant field of vision function sketched above is a coarsening function, which has this property:

v = area (D) = area (coarsened (W))
e = area (sum of mixed (coarsened (W))

and using maximal error, the interval $[t, t + u]$ must be included in $[v, v + e]$, which simplifies with $u=0$ to the statement $v <= t <= v + e$, which is here:

f.w = area (D) = area (coarsened (W)) < area (W)

because only if all four pixels in a group have the value In, the coarsened representation has the value In; in all other cases the value is Mixed or Out.

v + e = area In (coarsened (W) + area Mixed (coarsened (W)) = area (in or mixed) coarsened (W)
= total_area (W) - area_out (coarsened(W)) > total_area (W) - area_out (W)
area_out (coarsened (W)) < area_out (W)

7.3.2 Data quality transfer function preserves order

To show that the data quality transfer function preserves order, means to show that if the data quality of a data set 1 is less than the data quality of a data set 2, then the result achieved from data set 1 will be less accurate then the result from data set 2

$D_e' < D_e''$ *follows* $q(D_e') < q(D_e'')$

A dataset d_2 derived from a dataset 1 by the given coarsening operation, has a data quality corresponding to the area of the groups in the 'instant field of vision'. For simplification, we assume here that the space is represented by a regular raster and the grouping takes four neighboring cells - similar to the quadtree construction rule. This simplifies the argument of the proof, but even if irregular groupings are selected, the result is the same (but the proof requires induction over the size of the groups). Under this assumption, the error measure for the dataset D_2 is 4 (assume the area of a single cell is 1), thus clearly larger than $D_1 = 1$.

It remains to show that $e' = f(D') < e'' = f(D'')$. The data values must be the same (except for the coarsening), thus we must show that $f(D') < f(D'')$, or here $f(1) < f(4)$.

Consider that the error D_e for the area calculation is the sum of the area of the mixed pixels. Coarsening translates each group in a single pixel, with value Mixed if not all values in the group are the same (and not mixed). The area of the mixed pixels can only increase, thus *f (D') < f (D")*.

7.4 Formalisation for a raster model

The above model is intentionally very abstract. To translate it quickly into a working program, a raster representation of the world is assumed and the grouping collects four cells to a generalized one.

The world is represented by a very fine raster (the raster cell can be thought of as arbitrarily but finitely small). A region is represented as a raster with cells which are marked in or out, and it is assumed that the regions represent natural phenomena like forest, urban area etc. with high spatial autocorrelation. The function of interest is an area calculation, which reduces to a count of raster cells multiplied by the area of a cell. The observation is producing a raster of coarser resolution, taking into account the many cells from the world which fall into a data cell; cell values for data are either In, Out or Mixed (in case not all world cells are in or out). The area calculation in the data values counts the In pixels for the value and the Mixed pixels for the error. A computed value is acceptable if the true value is in the interval between the In count and the sum of the In and the Mixed cell count.

8 Conclusions

Data quality descriptions are related to the intended use of the data, no generic solution.

Large classes of 'functions of interest' exist to work with the same data quality description.

Show that I propose is 'operational, quantitative and ..'

A few hypotheses relate to this framework:

- which operations can be integrated in this framework for data quality description by area of cell (i.e. raster cell size). It appears as if most (if not all) operations in Tomlin's Map Algebra (Tomlin 1983; Tomlin 1994) propagate error in the described form.
- one might be tempted to call operations which behave reasonable under this error measure as 'essentially spatial' and operations which depend on the
- the method assumes a strong spatial autocorrelation - which is also dependent on the size of the neighborhood. Phenomena which are not strongly autocorrelated at the scales (i.e. raster sizes) considered, cannot be covered by this approach.

Acknowledgments

This article is based on a presentation given at the First Cassini International Workshop "Data Quality in Geographic Information: from Error to Uncertainty" in Paris in April 1997, organized by Robert Jeansoulin. Discussions with Mike Goodchild and Barbara Buttenfield helped me to clarify some of the issues and I appreciate their contribution. Martin Staudinger made the figures.

References

Asperti, A., and G. Longo. 1991. *Categories, Types and Structures - An Introduction to Category Theory for the Working Computer Scientist*. Edited by G. M. a. M. Albert, *Foundations of Computing*. Cambridge, Mass.: The MIT Press.

Bruegger, B. P. 1995. Theory for the Integration of Scale and Representation Formats: Major Concepts and Practical Implications. In *Spatial Information Theory-A Theoretical Basis for GIS*, edited by A. U. Frank and W. Kuhn. Berlin-Heidelberg: Springer.

Bruegger, B. P. 1994. Spatial theory for the integration of resolution-limited data. Ph.D. thesis, University of Maine, Orono, Maine.

Chrisman, N. R. 1983. The Role of Quality Information in the Long-Term Functioning of a Geographic Information System. Proceedings of Sixth International Symposium on Automated Cartography (Auto Carto 6), at Ottawa, Ontario, Canada.

Chrisman, N. R. 1991. The error component in spatial data. In *Geographical Information Systems: principles and applications*, edited by D. J. Maguire, M. F. Goodchild and D. W. Rhind. Essex: Longman.

Clapp, J. L., D. .D. Moyer, and B. J. Niemann. 1988. The Wisconsin Land Records Committee: Its Background, Status, Impact, and Future. Proceedings of GIS/LIS'88, Third Annual International Conference, at San Antonio, Texas.

Cressie, N. A. 1991. *Statistics for Spatial Data*, *Wiley Series in Probability and Mathematical Statistics*. New York: John Wiley.

Frank, A. U. 1992. Acquiring a digital base map - A theoretical investigation into a form of sharing data. *URISA Journal* 4 (1):10 - 23.

Frank, A. U. 1995. Strategies. In *Geographic Information Systems - Materials for a Post-Graduate Course: Vol.6 - GIS Organization*, edited by A. U. Frank: Dept. of Geoinformation, Technical University Vienna.

Frank, A. U. 1996. Der Nutzen und der Preis von Geographischer Information. Proceedings of AGIT'96, at Salzburg.

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

Gurda, R. F., D. D. Moyer, B. J. Niemann, and S. J. Ventura. 1987. Costs and Benefits of GIS: Problems of Comparison. Proceedings of Int. Geographic Information Systems (IGIS) Symposium (IGIS'87), at Arlington, VA.

Horn, B. K. P. 1986. *Robot Vision*, *MIT Electrical Engineering and Computer Science Series*. Cambridge, MASS: MIT Press.

Morrison, J. 1988. The proposed standard for digital cartographic data. *Amer. Cartographer* 15 (1):9-140.

National Research Council. 1980. *Need for a Multipurpose Cadastre*. Washington DC.: National Academy Press.

Robinson, V., and A. U. Frank. 1985. About Different Kinds of Uncertainty in Collections of Spatial Data. Proceedings of 7th Int. Symposium on Computer-Assisted Cartography (Auto-Carto 7), at Washington, DC.

Stanek, H., and A. U. Frank. 1993. Data quality - Necessary Complement for GIS Based Decision Making. Proceedings of 25th International Symposium: Remote Sensing and Global Environmental Change, at Graz.

Stevens, S. S. 1946. On the theory of scales of measurement. *Science* 103 (2684):677 - 680.

Timpf, S., and A. U. Frank. 1997. Metadaten-multimedial. Proceedings of AGIT'97, at Salzburg.

Timpf, S., M. Raubal, and W. Kuhn. 1996. Experiences with Metadata. Proceedings of 7th Int. Symposium on Spatial Data Handling, SDH'96, at Delft, The Netherlands (August 12-16, 1996).

Tobler, W. R. 1979a. Cellular Geography. In *Philosophy in Geography*, edited by S. Gale and G. Olsson. Dordrecht Holland: D. Reidel Publishing Company.

Tobler, W. R. 1979b. A transformational view on cartography. *The American Cartographer* 6 (2):101-106.

Tomlin, C. D. 1983. A Map Algebra. Proceedings of Harvard Computer Graphics Conference, at Cambridge, Mass.

Tomlin, C. D. 1994. Map algebra: one perspective. *Landscape and Urban Planning* 30:3-12.

Voisard, A., and H. Schweppe. 1996. A Multilayer Approach to the Open GIS Design Problem. Proceedings of 2nd ACM GIS Workshop, at Gaithersburg, MD.

Walters, R. F. C. 1991. *Categories and computer science*. Vol. 1, *Cambridge Computer Science Texts*. Cambridge, UK: Carslaw Publications.

Zeitlberger, K. 1997. Der Einfluss von Datenqualität auf GIS-basierte Entscheidungen: Eine Untersuchung am Beispiel eines Ertragsmodells in der Landwirtschaft. Master Thesis, Dept. of Geoinformation, Technical University, Vienna.