Why Is Scale an Effective Descriptor for Data Quality? The Physical and Ontological Rationale for Imprecision and Level of Detail

Andrew U. Frank

Department of Geoinformation and Cartography
Technical University Vienna
Gusshausstrasse 27-29/E127
A-1040 Vienna, Austria
frank@geoinfo.tuwien.ac.at

Abstract

Observations and processing of data create data and their quality. Quantitative descriptors of data quality must be justified by the properties of the observation process. In this contribution two unavoidable sources of imperfections in the observation of physical properties are identified and their influences on data collections analyzed. These are, firstly, the *random noise* disturbing precise measurements; secondly, *finiteness of observations*—only a finite number of observations is possible and each of it averages properties over an extended area.

These two unavoidable imperfections of the data collection process determine data quality. Rational data quality measures must be derived from them: Precision is the effect of noise in the measurement. The finiteness of observations leads to a novel formalized and quantifiable approach to level of detail.

The customary description of a geographic data set by 'scale' seems to relate these two sources of imperfection in a single characteristic; the theory described here justifies this approach for static representation of geographic space and shows how to extend it for spatio-temporal data.

G. Navratil (eds.), *Research Trends in Geographic Information Science*,
Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-540-88244-2_4,

1 Introduction

Digital geographic data comes in different qualities and applications have different requirements for the quality of their inputs. In order to advance the use of digital geographic data, qualitative descriptions of the quality provided or required is necessary. Traditionally *map scale* is used to describe summarily the quality of static geographic data when cartographically represented. The reduction in size, expressed as proportional scale, causes a reduction in precision and detail. Users of maps have learned which map scales are suitable for which task: orienteering uses map in the scale range 1:10.000 to 1:25.000, for driving by car from city to city maps 1:250.000 to 1:500.000 are sufficient, etc. Repeated experiences taught us these practical guidelines and we follow them without asking for a theory.

In the age of digital data, the traditional definition of map scale, as the proportion between distances on the map and in reality, lost its justification: locations are expressed with coordinates and distances computed are in real world units. Only when preparing a graphical display, a proportional scale is used. The concept of scale in a digital world has been critically commented, but no solution suggested (Lam et al. 1992; Goodchild et al. 1997; Reitsma et al. 2003).

Data quality research needs a quantitative, theory based approach. The theory must relate to the physical characteristics of the observation process, where the imperfections in the data originate. Data quality measures, which are not related to universal properties of observation remain specific for some data collection technologies (Timpf et al. 1996) and impede the assessment of results from integration with other datasets obtained by other methods and with incompatible data quality descriptions.

In this paper I explore the process of geographic data collection and show how scale is introduced originally when making observations. It must be carried forward as a quality indicator with the dataset. The same "scale" quality measure is later used when considering whether a dataset can be used effectively in a decision situation (Frank 2008a). The tiered ontology, previously described in a number of articles (Frank 2001, 2003) is used as foundation for the analysis of the data collection process and reviewed here briefly in section 2. The tiered ontology commits to a physical reality in a space time continuum that can be observed. In a second tier physical objects are formed and a third tier includes the conventional, socially constructed objects (Searle 1995).

In this article the focus is on data quality describing physical objects. Section 3 describes the processes that are used to transform information between the tiers. The ontological approach distinguishes point observations from descriptions of objects and their attributes. The point observations are simpler than the prototypical measurements of measurement

theory (Krantz et al. 1971); this reduction to a more primitive type of observation allows to include imperfections in the theory, which classical measurement theory could not deliver (Orth 1974). The analysis leads to a quantitatively assessment of the imperfections introduced by each process (Frank 2007)—a goal desired since the data quality discussion in the mid 1980s (Chrisman 1987; Robinson et al. 1987), and the development of measurement theory (Krantz et al. 1971) but never comprehensively, systematically, and operationally achieved.

An analysis of the properties of real (physical) observation processes reveals that the limitations in the observation processes introduce two types of unavoidable imperfections: *random noise dilutes the precision* of the observation (section 4), and the *finiteness of the sensor limits the level of detail* (section 5). In a well-designed sensor these two effects are comprehensively characterized by scale. Map scale, in this definition, is therefore not an artifact of cartography but originates in the physical observation process itself; every observation introduces necessarily a scale to the data, independent of cartographic rendering.

The novel contributions of this paper are:

1. An ontology based analysis reveals universal limitations of all physical observation processes. These limitations are random noise and finiteness of sensors; quantitative descriptors of data quality must originate in these universal sources of imperfections.
2. A theory of data quality grounded in universal properties of the observation process and thus independent of technology, usable to integrate data from different sources and assess the quality of the result.
3. Introducing scale as a property of data resulting from the observation process (and not an artefact of cartographic rendering).
4. Scale is a justifiable summary description for data quality of static physical geographic datasets for observation processes that are with balanced precision and resolution. It can be extended from the spatial to the temporal dimension.

2 Tiered Ontology

An ontology describes the conceptualization of the world used in a particular context (Guarino 1995; Gruber 2005). The ontology clarifies the concepts and communicates the semantics intended by data collectors and data managers to persons making decisions with the data. Clarification of semantics is equally important for the semantics of data quality description.

Therefore, the description of data quality must be included in the ontology (Frank 2008b).

The tiered ontology used here (Frank 2001, 2003) starts with tier O, which is the physical reality, that "what is", independent of human interaction with it. Tier O is the Ontology proper in the philosophical sense (Husserl 1900/01; Heidegger 1927, reprint 1993). The ordinary space-time continuum is assumed as the structure of physical reality.

2.1 Tier 1: Point Observations

Reality is observable by humans and other cognitive agents (e.g., robots, animals). Physical observation mechanisms produce data values from the properties found at a point in space and time. $v = p\ (x, y, z, t)$. A value v is the result of an observation process p of physical reality found at point (x, y, z) and time t.

Tier 1 consists of the data resulting from observations at specific locations and times (termed "point observation"); philosophers sometimes speak of "sense data" (Stanford Encyclopedia of Philosophy http://plato.stanford.edu/). In GIS such observations are often realized as raster data resulting from remote sensing, similarly to our retina that performs such point observations in parallel. Sensors and sensor networks (Stefanidis et al. 2005) in general produce point observations as well, but of a different kind, as will be seen in section 5.

2.2 Tier 2: Objects

The second tier is a description of the world in terms of physical objects. Objects are regions of space that have uniformity in some property. The object representation reduces the amount of data, if the subdivision of the world into objects is such that most properties of the objects remain invariant in time (McCarthy et al. 1969). For example, most properties of a taxi cab remain the same for hours, days or even longer, and they need not be observed and processed repeatedly. Only location and occupancy of the taxi cab change often.

The formation of objects—what Zadeh calls granulation (Zadeh 2002)—is a complex process of (1) determining the boundaries of objects (2) summarizing some properties for the delimited regions and finally (3) determine the type of the object (classification). For objects on a tabletop, object formation is dominated by spatial cohesion, which moves as a single piece is an object: a cup, a saucer, and a spoon (Fig. 1).

Fig. 1. Simple physical objects on a tabletop: cup, saucer, spoon

Geographic space does not lead itself to such a single, dominant, subdivision as objects typically do not move. Multiple aspects are used to form regions of uniform properties, leading to different objects overlapping in the same space (Fig. 2). Watersheds, areas above some height or regions of uniform soil, uniform land management, etc. can be identified and they all overlap (Couclelis 1992). Object classification is optimized to classify objects suitable for certain operations (hunting, planting corps, grazing cattle, etc.).

Fig. 2. Fields in a valley: multiple overlapping subdivisions in objects are possible.

2.3 Tier 3: Social Constructions

Tier 3 consists of constructs combining and relating physical objects to abstract constructs. This includes constructions like money (Fig. 3), legal marriage, ownership of land, etc. Constructed reality links a physical

object X to mean the constructed object Y in the context Z. “X counts as Y in context Z” (Searle 1995).

Fig. 3. Some pieces of metal and a piece of paper counting as money in the Czech Republic

Social constructions give meaning to physical objects or processes. Socially constructed objects can alternatively be constructed from other constructed objects, but all constructed objects are eventually grounded in physical objects. No “freestanding Y terms”, contrary to (Zaibert et al. 2004).

The present article focuses on physical observations and objects; the generalization of results to social construction is left to future work.

3 Information Processes

Any ontology for an information system that separates different aspects of reality must not only conceptualize the objects and processes in reality but must also describe the information processes that link the different conceptualizations and transform between them. This is of particular importance for an ontology that divides conceptualization of reality in tiers (Frank 2001; Smith and Grenon 2004). This transformation process introduces imperfections and is therefore responsible for the data quality.

Information processes transform information obtained at a lower tier to a higher tier (Fig. 4). All human knowledge is directly or indirectly the result of observations, transformed in chains of information processes. The processes that connect the tiers of ontology are described in this section before analyzing the limitations that produce the imperfections in the observation in the next two sections. All imperfections in data must be the result of some aspect of an information process. As a consequence, all theory

of data quality and error modeling has to be related to empirically justified properties of the information processes.

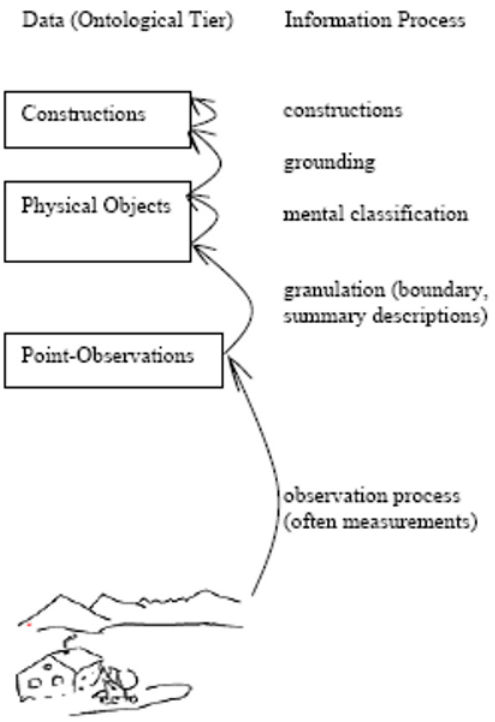

Fig. 4. Tiers of ontology and information processes transforming data between them

3.1 Observations of Physical Properties at Point

The physical process that links tier O to tier 1 is the observations of physical properties at a specific point. Observations are imperfect for two causes

- random noise disturbs the value produced and reduces precision, and
- finiteness of sensors force the observation not at a point but over an extended area and limits resolution.

A systematic bias can be included in the model of the sensor and be corrected by a function and is not further considered. Noise reducing precision is the focus of sections 4 and the finiteness of the sensor limiting resolution is discussed in section 5 (Fig. 5).

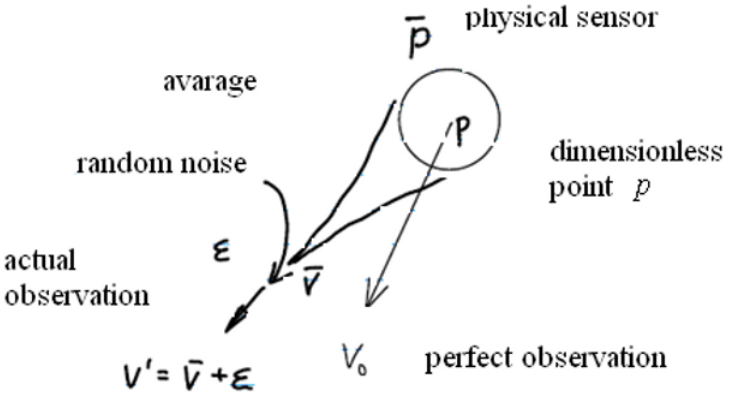

Fig. 5. Imperfections in point observations

Observations must be represented with finite length symbols. This discretization introduces yet another imperfection sometimes expressed as 'dynamic range'. It is of minor importance because observation systems can be constructed such that this influence is negligible compared to the influences of noise (see also subsection 5.3).

3.2 Object Formation (Granulation)

Human cognition focuses on objects and object properties. We are not aware that our eyes (but also other sensors in and at the surface of our body) report point observations. For example, the individual sensors in the eye's retina give a pixel-like observation, but the eye seems to report about size, color, and location of objects around us (Marr 1982). The observations are, immediately and without the person being conscious about the processes involved, converted to object data connecting tier 1 point observation to tier 2 object properties. Such processes are found not only in humans but higher animals also form mental representations of objects as well.

Object formation increases the imperfection of data—instead of having detailed knowledge about each individual pixel only a summary description of, for example, the object "middle wheat field" in Fig. 2 is retained. Reporting information with respect to objects results in a substantial reduction in size of the data. For example in Fig. 2, the area for the field includes approx. 1.5 million pixels each of which has 8 bits in three color changes. The compact representation as a region requires few points for

the boundary and a few bytes to describe the average color of the region. This is a computer model and not necessarily representative for processes in a human brain but gives nevertheless a general idea of the million fold compression achieved by object formation.

Object formation consists of three information processes

- boundary identification,
- computing summary descriptions, and
- classification,

which will be sketched in the following three subsections (more details in (Frank draft 2005)).

3.3 Boundary Identification

Objects are formed as regions in two-dimensional space (or 3D, 2D + T, 3D + T, etc.) that are uniform in some aspect. An object boundary is determined by first selecting a property and a property value that should be uniform across the object, similar to the well-known procedure for regionalization of 2D images. Tabletop objects in Fig. 1 are uniform in the material coherence and in their movement (Reitsma et al. 2003) the field in Fig. 2 is uniform in its color. Object formation exploits the strong correlation found in the real world; human life would not be possible, if not for most properties and places, nearby values are very similar. Which properties must be uniform to form an object is determined by the interactions intended and the situation. The focus of this article excludes a detailed discussion of processes and how they depend on properties of the object involved relating properties, object boundaries, and affordance of processes (Gibson 1986; Raubal 2002; Kuhn 2007).

3.4 Determination of Descriptive Summary Data

Descriptive values, summarize the properties of the object limited by a boundary. The computation is typically an integral (or similar summary function) that determines the sum, maximum, minimum, or average over the region, e.g., total weight of a movable object, amount of rainfall on a watershed, maximum height in country (Tomlin 1983; Egenhofer et al. 1986).

$$a_n = \iiint_{Fn} v(x, y, z)\, dx\, dy\, dz \tag{1}$$

where the attribute value for attribute a and object n is the integral over the 3D region Fn of the object n for the property value at v(x, y, z).

3.5 Classification

Objects once identified are mentally classified. On the tabletop (Fig. 1), we see a cup, spoon, and saucer; in a landscape (Fig. 2) forest, fields, and streams are identified. Classes—similar to types in computer languages (Cardelli 1997)—indicates which operation can be performed with an object. Gibson introduced the term affordance (Gibson 1986; Raubal 2002).

Fig. 6. Pouring requires two container objects and one liquid object

Mental classification relates the objects identified by granulation processes to operations, i.e., interactions of the cognitive agent with the world. To perform an action, e.g., to pour water from a pitcher into a glass (Fig. 6) requires a number of properties of the objects involved: the pitcher and the glass must be containers, i.e., having the affordance to contain a liquid, the object poured must be a liquid, etc. I have used the term distinction for the differentiation between objects that fulfill a condition and those that do not (Frank 2006). Distinctions are partially ordered: a distinction can be finer than another one (e.g., drinkable is a subtaxon of liquid); distinctions form a taxonomic lattice (Frank 2006).

3.6 Constructions

The tier 3 contains constructed objects and actions, which are linked through granulation and mental classification to the physical reality of

physical objects and operations. They are directly dependent on the information processes described above, but details are not consequential for present purposes.

4 Random Effects on the Observations

Physical sensors are influenced by random processes that produce perturbations of the observations. The unpredictable disturbance is typically modeled by a probability distribution. For most sensor a normal (Gaussian) probability distribution function (pdf) is an appropriate choice described by expected value (mean μ) and variance (standard deviation σ).

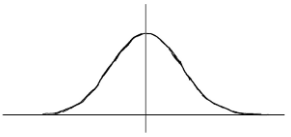

Fig. 7. Normal distribution

$$f(x,\,\mu,\,\mu) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2e^2}} \tag{2}$$

If the same observation could be repeated multiple times (which is strictly speaking not possible, because these observations would be at different times), we could compute an average and a standard deviation from the values observed.

4.1 Influence on Object Formation and Summary Values

Errors in the observation influence the determination of the object boundary. The statistical error of the boundary follows for simple cases from Gauss' law of error propagation; the standard deviation σ_f for function $f(u,v,w)$ in terms of σ_u, σ_v, and σ_w is:

$$\sigma_r^2 = \sigma_u^2\left(\frac{df}{du}\right)^2 + \sigma_v^2\left(\frac{df}{dv}\right)^2 + \sigma_w^2\left(\frac{df}{dw}\right)^2 \tag{3}$$

If the observation information processes allow a probabilistic description of the imperfections of the values, then the imperfections in the object boundary and summary value are equally describable by a probability

distribution. Assuming a pdf for the determination of the boundary, one can describe the pdf for the boundary line (Fig. 8). It is an open question whether the transformations of probability density functions associated with boundary derivation and derivation of summary values preserve a normal distribution, i.e., if observation processes described by imperfections with a normal distribution produce imperfections in boundary location and summary values that are describable again by a normal distribution.

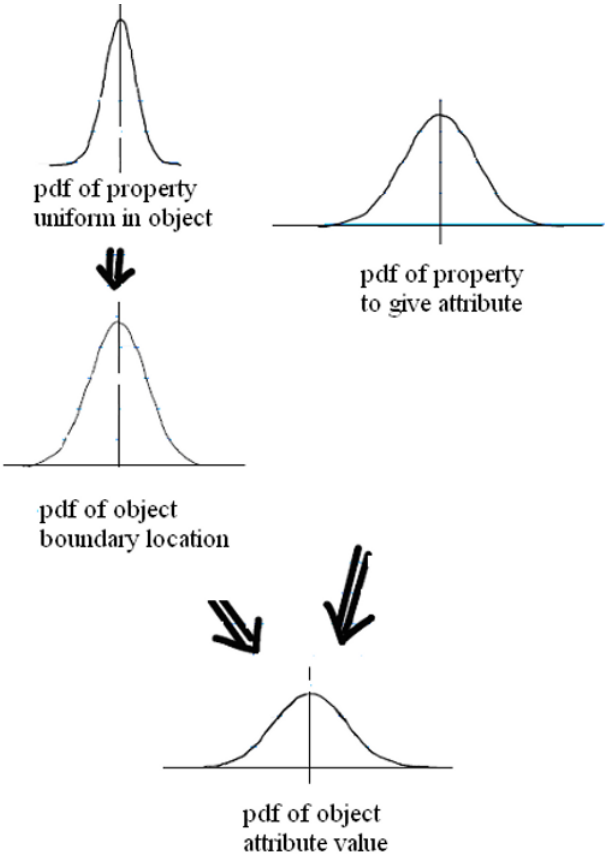

Fig. 8. The transformation of pdf from observations to object attribute values

4.2 Classifications

Distinctions describe limits in the attribute values of an object, whether the object can or cannot be used for a specific interaction and thus is mentally

classified as a particular category. The decision whether the values for an object are inside the limits or not is more or less sharp and the cutoff usually gradual (Fig. 9). The distinctions and classifications are therefore fuzzy values, i.e., membership functions as originally defined by Zadeh (1974). This is a fuzzy membership value for the category because neither the relevant attribute values of the object nor the limits for classification are known accurately. Here error propagation usually comes to an end, because the situation and the mental assessment include various correlations between the multiple relevant aspects; it is more complex than the prototypical engineering decision. Classification reduces precision to what is relevant in the context of an operation; it also reduces uncertainty and increases reliability.

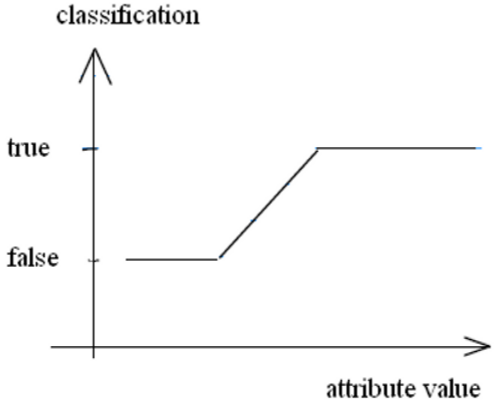

Fig. 9. Classification of objects results in fuzzy membership values

4.3 Qualitative Description

The assumption that the influence of the random effects follows a normal distribution is usually justified for physical observations. It is then reasonable to use the standard deviation σ to describe quantitatively the *precision* of observation data, even approximatively for data derived from point observations.

5 Finite Observation Devices

In this section the effects of the finiteness of physical observation instruments are discussed. The limits are threefold:

- the sensors are not infinitely small but measure over an extended area and time,
- only a finite number of observations are possible, and
- only a finite number of different readings to represent the observation value is possible.

5.1 Effects of Size of Sensor

A sensor cannot realize a perfect observation at a perfect point in space or time. Any physical observation must integrate the effects of a physical process over a region during a time. The time and region over which the integration is performed can be made very small. For example, a pixel sensor in a camera has a size of 5/1000 mm and integrates (counts) the photons arriving in this region for as little as 1/5000 sec but it is always of finite size and duration. The size of the area and the duration influences the result and give a "scale" to an observation.

The sensor can be modeled as a convolution of the perfect observation with a Gaussian. The effects of the size of the observation device is as unavoidable as the random perturbations of the observations, which is more widely recognized, discussed and given a formal model (summarized in the previous section). The intended point-observation $v = f(x, y, z)$ cannot be realized but effectively the device reports the average value for a small area and a small time interval $\underline{\varepsilon}$.

$$v(x, y, z, t) = \int_{-\varepsilon}^{+\varepsilon} f((x, y, z, t) - \underline{\varepsilon}) d\underline{\varepsilon} \tag{4}$$

where $\underline{\varepsilon}$ ranges over the size (x, y, z) and the time (t) interval used by the sensor. The region ε (in space and time) is called the support of the sensor. The imperfection in the observation process can be formalized as a special case of convolution. Convolutions are shift and linear invariant transformations, meaning shifting a signal in space gives the shifted result (5.1) and the addition of two inputs gives the addition of the two outputs (5.2) (for two signals g and h and a transformation f). This is:

$$
\begin{aligned}
& g'(x) = f(g(x)) \\
& h'(x) = f(h(x)) \\
& g'(x-a) = f(g(x-a)) \qquad (5.1) \\
\textit{mit} \quad & \alpha g'(x) + \beta h'(x) = f(\alpha g(x) + \beta h(x)) \qquad (5.2)
\end{aligned}
\tag{5}
$$

These seem to be fundamental rules for observation systems of a spatio-temporal reality.

The formal model is a convolution of $f(\underline{x}, t)$ with a kernel $k(\varepsilon)$

$$k_{(\varepsilon)} \text{ is } v(\underline{x}, t) = \int f((\underline{x}, t) - \underline{\varepsilon}) k(\underline{\varepsilon}) d\underline{\varepsilon}. \tag{6}$$

The observation $f(\underline{x}, t)$ is multiplied by the kernel value $k(\underline{\varepsilon})$, which is non-zero only for a small region around zero (the support) and for which

$$\int k(\varepsilon) d\varepsilon = 1. \tag{7}$$

For sensors in cameras, the effect can be modeled with a convolution with the size of the sensor elements (Fig. 10a). Together with the inevitable blurring of the optical system, the imperfection of the observation can be approximated by a convolution with a Gaussian kernel (Fig. 10b).

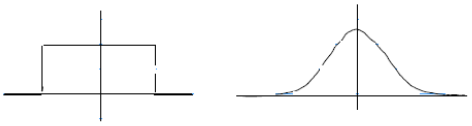

Fig. 10. a) a pillbox function describing a camera sensor b) the Gaussian with a variance σ

5.2 Effects of Finite Number of Observations

The sampling theorem addresses another related limitation of real observations: It is impossible to observe infinitely many points; real observations are limited to sampling the phenomenon of interest at finite number of points.

Sampling introduces the danger that the observations may suggest a signal that is not present and an artifact of the sampling (Horn 1986). The sampling theorem (a.k.a. Nyquist law) states that sampling must be twice as frequent as the highest frequency in the signal to avoid artifacts (so-called aliasing). The signal must be filtered and all frequencies higher than half the sampling frequency cut off (low-pass filter). In the audio world the sampling theorem is well known, but it applies to multi-dimensional

signals as well: including sampling by remote sensing or sensor network in geographical space. It maybe appear strange to speak of spatial frequency, but it is effective to make the theory available to GIScience, where it must be applied to all dimensions observed (2 or 3 spatial dimensions and the temporal dimension).

Filtering out high frequencies, i.e., small objects in the image, must be performed before sampling. Cameras, our eyes and remote sensing devices by their construction from small sensor elements tightly packed attenuate high frequencies sufficiently to avoid aliasing modeled in a convolution with a pillbox function (Fig. 10a). The limitations of the optical system producing blur is similarly a low-pass filter, which can be modeled as a convolution with a Gaussian kernel (Fig. 10b). The two effects (finite highly packed sensors and optical blur) create observations, which are suitably filtered by a low-pass filter to avoid aliasing. Point sensors spread at a distance however do not filter high frequencies and artifacts due to aliasing may appear in the data, if frequencies higher than half the sampling rate are present in the terrain. The regular patterns of vineyards, a high frequency signal in space has been observed to produce artifacts in laser scanning data (personal communication Wolfgang Wagner January 2009).

The finiteness of observations introduces effectively a scale into the data. It limits the resolution to objects twice the size of the sampling rate. From observations at one scale more generalized data on a larger scale (i.e., cartographically speaking, a smaller scale) can be produced, approximatively by convolution with a Gaussian, but data of higher resolution cannot be derived. The sampling rate effectively limits the zooming-in to obtain more detail. Proper observations avoiding aliasing can be formally modeled as a convolution with a Gaussian kernel acting as a low-pass filter followed by sampling. For this situation it appears reasonable to say that the observation has the "scale" corresponding to half the sampling frequency v, which is the cutoff frequency of a proper low-pass filter. A numerical description of "scale" could then be $1/v$ with dimension time (second) and length (meter) respectively; giving the size of the smallest detail included.

It is debatable whether to call this scale, adding one more sense to the dozen already existing, or to use a term like resolution or granularity. I prefer resolution because speaking of a dataset and stating its resolution, for example as "30 m in space and 1 year in time", extends the current usage reasonably.

5.3 Effects of Finite Representation of Observations

The representation of the observation is again finite. Only values from a range can be used. For example in photography and remote sensing, the intensity (amplitude) of the signal is given by an 8 bit value allowing values between 0 and 255 (2^8 -1)—the so-called dynamic range. In a properly designed observation system the dynamic range is smaller than the precision of the sensor and has therefore a negligible effect.

5.4 Effects of Scales on Object Formation

5.4.1 Size of smallest objects detected

The scale of the observation limits the smallest object that can be detected; objects with extension in one dimension less than the scale are not observed and their extent is merged with the neighbors. This applies to small separating objects as well; roads or streams separating two fields are not picked up at large scale (low resolution) sampling and two separate fields appear as one.

5.4.2 Effects on attribute values

Attribute values are derived from two different observations: one signal is used to determine the object boundary the other is integrated over the region of the object to give the attribute value (subsection 3.4). If the two scales are comparable, the result will be meaningful at this scale. If, however, the scales are different, the interpretation of the result is difficult. The result has the larger scales of the two; it seems possible to treat this question formally in the framework designed here and I leave this as open question.

5.4.3 Effects on object classification

The scale of observation, influencing directly the object formation influences indirectly the classification in geographic data. This is relevant, where the size of an object influences the classification especially if the class is distinct by a size, e.g., small buildings vs. larger buildings. A recent study on reserves of land zoned 'residential' assumed that a building to qualify as a residential building had a minimal footprint of 60 sq.m. (Riedl 2009). In data of a scale m one does not expect objects smaller than $f \bullet m^2$, where f is the maximally expected indicating how different such objects are from a square (respective cube).

5.5 Qualitative Descriptors

The finiteness of the sensor is affecting the data in 3 ways:

- the sampling rate
- the support of the sensor
- the dynamic range of the sensor.

The support of the sensor produces in a well-designed observation system the necessary low-pass filter to satisfy the sampling theorem for the sampling rate used. In this case a description with the sampling rate in space and time alone is sufficient. Dynamic range in a well-designed sensor is such that the effects are less than the noise in the observations.

6 Scale as a Summary Description

In a well-designed observation system, the inevitable imperfections introduced by the observation system are by design balanced. There is no point in taking more samples than necessary from a band—with limited signal or observing with more precision a low resolution (band limited) signal. The appropriate relation between the low-pass filter and the appropriate precision of the observation depends on the amplitude (energy) of the signal for different frequencies. In general more precision in the observation than what is filtered away will be unnecessary.

Noting that quantitative descriptors for data quality are neither obtained nor required to be very precise, it is sufficient if the four characteristics (precision, sampling rate, highest frequency in signal, and dynamic range) of the observation system are approximately corresponding. Then they can be described with a single quantity, for which I suggest to use the term scale.

If the highest frequency in the signal is v, corresponding to a wavelength λ, then the resolution is 2λ and the smallest objects discernable are at least of size 2λ in any dimension. The spatial-temporal precision should then be of the same order ($\sigma \approx \lambda$) and the attribute precision comparable to the amplitude in signal frequencies higher than v.

The traditional map scale, describing the result of a cartographic rendering process, is organized around the accuracy of the human eye and limitation of the reproduction process. Assuming that one tenth of a millimeter is the graphical resolution, the "scale" describes the size of this minimal graphical element ($^1/_{10}$ to 1 millimeter) in reality. A map scale of 1:50.000 indicates that the smallest object expected is 5 m to 50 m, and precision of location is approximately the same. Spatial resolution expressed in milli-

meters (50m = 50.000 mm) gives the customary scale denominator. Different national mapping standards vary somewhat, but this describes the expectations of a map reader and his assessment what use the map is fit for realistically.

As a guideline, traditional map scale is therefore a reasonable comprehensive descriptor for the quality of a balanced data product. Combining datasets of similar "scale" produces most likely reasonable results. Combining datasets with very different "scales" requires care and the four different characteristics of observation quality must be considered separately. It is probably acceptable that for datasets for which only a summary description with scale is given, the individual characteristics are "reconstructed" assuming a balanced observation method.

7 Conclusions

Physical observation systems deviate in two inevitable and non-avoidable respects from the perfect point observation of the properties of reality:

- Random perturbation of results
- Finite spatial and temporal extent over which the observation system integrates.

Random effects are described by probability distribution function pdf and the propagation of these follow in simple cases Gauss' Law of error propagation, in general transformations for the probability distribution must be computed.

The effects of finite support for the observation can be modeled as a convolution with a Gaussian Kernel and the non-zero extend of the kernel determines the "scale" of the observation. The signal must be filtered before the sampling with a low-pass filter to cut off all frequencies above half the sampling frequency. The typical sensors for remote sensing or photogrammetry achieve this and can be modeled as convolution with a Gaussian kernel.

In a balanced, well-designed observation system, the attenuation of frequencies above the half the sampling frequency is sufficient to avoid artifacts in the result (aliasing). Precision for the signal corresponds to the resolution.

For a dataset obtained with a well-balanced observation system, a characterization of the quality by a spatial and temporal scale is reasonable. It allows decision by users about the uses the data is fit for. A detailed analysis is necessary if multiple signals are observed by different observation systems, which is the regular case for GIS. The improvements of interoperability

allow the use of datasets from different sources. If they are combined, the analysis must detail the four characteristics for each signal and consider its combination. The limitations resulting from analog overlay of detailed and less detailed maps are known—they are not less severe in a digital system but less visible. The description given here shows how they can be dealt with analytically.

Using a tiered ontology where point observations are separated from object descriptions allows to follow how the imperfection introduced by random error and scale propagate to objects and their attributes. The analysis of the physical observation process and its formalization as a combination of random noise and a convolution with a Gaussian Kernel opens the door for a formal treatment of the effects in particular situation. Recommended is research to understand how data of different scale interacts and how from a dataset with small scale a dataset with a larger scale can be derived. Previous research by Openshaw et al. (1987) on the modifiable areal unit problem (MAUP) documents the importance of the question. The framework allows a formal treatment, but does not answer the question what the right scale to describe a phenomenon is. A recent paper by Gabora, Rosch and Aerts (2008) discusses the transformation of concepts (classes) between contexts. Changes in scale can be modeled as scale change and it appears promising to see if the approach of Gabora et al. applies.

Information is used to make a decision; this may be a simple, everyday decision in street navigation—"do I turn right or left here?"—decisions leading to important and complex actions—"is a new hospital building at location X necessary?"—or even indirectly connected to action as in the testing of scientific hypothesis that leads to scientific rules. A decision can always be reduced to a binary question and thus brought to a comparison of two values, from which a yes or no answer follows. Formally a decision is described as a test: $R - S > 0$. If imperfections affect R and S and formal models exist for these imperfections, the imperfection of the decision—i.e., the risk that the decision is wrong—can be tested. In particular, the scale for the observation of R and S must be comparable; this means for ordinary supplications that the scale of the observation should be comparable to the scale of the action we decide on.

Scale effects in geographic data are not yet well understood, despite many years of being listed as one of the most important research problems (Abler 1987; NCGIA 1989b; NCGIA 1989a; Goodchild et al. 1999). It is hoped that the conceptual clarification achieved here and the formalization using convolutions contributes to advancing research in scale effects in information processes.

Acknowledgements

These ideas were developed systematically for a talk I presented at the University of Münster. I am grateful to Werner Kuhn for this opportunity.

References

Abler R (1987) Review of the Federal Research Agenda. In: International Geographic Information Systems (IGIS) Symposium (IGIS'87), The Research Agenda, Arlington, VA

Cardelli L (1997) Type Systems. In: Tucker AB (ed) Handbook of Computer Science and Engineering, CRC Press, pp 2208–2236

Chrisman N (1987) Fundamental Principles of Geographic Information Systems. In: Auto-Carto 8, Baltimore, MA, ASPRS & ACSM

Couclelis H (1992) People Manipulate Objects (but Cultivate Fields): Beyond the Raster-Vector Debate in GIS. In: Frank AU, Campari I, Formentini U (eds) Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, Springer, Berlin Heidelberg New York, LNCS 639, pp 65–77

Egenhofer MJ, Frank AU (1986) Connection between Local and Regional: Additional "Intelligence" Needed. In: FIG XVIII International Congress of Surveyors, Toronto, Canada (June 1-11, 1986)

Frank AU (2006) Distinctions Produce a Taxonomic Lattice: Are These the Units of Mentalese? In: International Conference on Formal Ontology in Information Systems (FOIS), Baltimore, Maryland, IOS Press

Frank AU (2001) Tiers of Ontology and Consistency Constraints in Geographic Information Systems. International Journal of Geographical Information Science (IJGIS) 75(5 (Special Issue on Ontology of Geographic Information)): 667–678

Frank AU (2003) Ontology for Spatio-Temporal Databases. In: Koubarakis M, Sellis T, Frank AU, Grumbach S, Güting RH, Jensen CS, Lorentzos N, Manolopoulos Y, Nardelli E, Pernici B, Schek H-J, Scholl M, Theodoulidis B, Tryfona N (eds) Spatiotemporal Databases: The Chorochronos Approach, Springer, Berlin Heidelberg New York, pp 9–78

Frank AU (2007) Data Quality Ontology: An Ontology for Imperfect Knowledge. In: Winter S, Duckham D, Kulik L, Kuipers B (eds) Spatial Information Theory, 8th International Conference, COSIT 2007, Melbourne, Australia, September 19-23, 2007, Proceedings, Lecture Notes in Computer Science 4736, Springer, Berlin Heidelberg New York, pp 406–420

Frank AU (2008a) Analysis of Dependence of Decision Quality on Data Quality. Journal of Geographical Systems 10(1): 71–88

Frank AU (2008b) Data Quality - What Can an Ontological Analysis Contribute? In: Spatial Accuracy Assessment in Natural Resources and Environmental Sciences 2008, Shanghai, China, WorldAcademicPress

Frank AU (draft 2005) Ontology for GIS. Vienna, Technical University Vienna, Institute for Geoinformation and Cartography

Gabora L, Rosch E, Aerts E (2008) Toward an Ecological Theory of Concepts. Ecological Psychology 20(1): 84–116

Gibson JJ (1986) The Ecological Approach to Visual Perception, Hillsdale, NJ, Lawrence Erlbaum

Goodchild MF, Egenhofer MJ, Kemp KK, Mark DM, Sheppard E (1999) Introduction to the Varenius Project. International Journal of Geographical Information Science (IJGIS) 13(8): 731–745

Goodchild MF, Proctor J (1997) Scale in a Digital Geographic World. Geographical & Environmental Modelling 1(1): 5–23

Grenon P, Smith B, Goldberg L (2004) Biodynamic Ontology: Applying BFO in the Biomedical Domain. In: Pisanelli DM (ed) Ontologies in Medicine, IOS Press, Amsterdam, pp 20–38.

Gruber, T. (2005). "TagOntology - a way to agree on the semantics of tagging data." Retrieved October 29, 2005., from http://tomgruber.org/writing/tagontology-tagcapm-talk.pdf.

Guarino, N. (1995). "Formal Ontology, Conceptual Analysis and Knowledge Representation." International Journal of Human and Computer Studies. Special Issue on Formal Ontology, Conceptual Analysis and Knowledge Representation, edited by N. Guarino and R. Poli 43(5/6).

Heidegger, M. (1927; reprint 1993). Sein und Zeit. Tübingen, Niemeyer.

Horn, B. K. P. (1986). Robot Vision. Cambridge, Mass, MIT Press.

Husserl (1900/01). Logische Untersuchungen. Halle, M. Niemeyer.

Krantz DH, Luce RD, Suppes P, Tversky A (1971) Foundations of Measurement. New York, Academic Press

Kuhn W (2007) An Image-Schematic Account of Spatial Categories. In: Winter S, Duckham D, Kulik L, Kuipers B (eds) Spatial Information Theory, 8th International Conference, COSIT 2007, Melbourne, Australia, September 19-23, 2007, Proceedings, Lecture Notes in Computer Science 4736, Springer, Berlin Heidelberg New York

Lam N, Quattrochi DA (1992) On the issues of scale, resolution, and fractal analysis in the mapping sciences. The Professional Geographer (44): 88–98

Marr D (1982) Vision. New York, N.Y., W.H. Freeman

McCarthy J, Hayes PJ (1969) Some Philosophical Problems from the Standpoint of Artificial Intelligence. In: Meltzer B, Michie D (eds) Machine Intelligence 4. Edinburgh, Edinburgh University Press, pp 463–502

NCGIA (1989a) The U.S. National Center for Geographic Information and Analysis: An Overview of the Agenda for Research and Education. International Journal of Geographical Information Science (IJGIS) 2(3): 117–136

NCGIA (1989b) Use and Value of Geographic Information Initiative Four Specialist Meeting, Report and Proceedings, National Center for Geographic Information and Analysis; Department of Surveying Engineering, University of Maine; Department of Geography, SUNY at Buffalo

Openshaw S, Charlton M, Wymer C, Craft A (1987) A Mark 1 Geographical Analysis Machine for the automated analysis of point data sets. International Journal of Geographical Information Systems 1(4): 335–358

Orth B (1974) Einführung in die Theorie des Messens. Verlag W. Kohlhammer, Stuttgart, Berlin, Köln, Mainz

Raubal M (2002). Wayfinding in Built Environments: The Case of Airports. Münster, Solingen, Institut für Geoinformatik, Institut für Geoinformation.
Reitsma F, Bittner T (2003) Process, Hierarchy, and Scale. In: Spatial Information Theory, Cognitive and Computational Foundations of Geographic Information Science, International Conference COSIT'03
Riedl M (2009) Erstellung von Baulandbilanzen in Tirol. In: 15. Internationale Geodätische Woche Obergurgl, Ötztal Tirol, Wichmann
Robinson V, Frank AU (1987) Expert Systems Applied to Problems in Geographic Information Systems: Introduction, Review and Prospects. In: Auto-Carto 8, Baltimore, MA, ASPRS & ACSM
Schneider M (1995) Spatial Data Types for Database Systems. Hagen, FernUniversität
Searle JR (1995) The Construction of Social Reality. New York, The Free Press
Stefanidis A, Nittel S (2005) Geosensor Networks. Boca Raton, Florida: CRC Press
Timpf S, Raubal M, Kuhn W (1996) Experiences with Metadata. In: 7th Int. Symposium on Spatial Data Handling, SDH'96, Delft, The Netherlands (August 12-16, 1996), Faculty of Geodectic Engineering, Delft University of Technology
Tomlin CD (1983) A Map Algebra. Harvard Computer Graphics Conference, Cambridge, Mass.
Zadeh LA (1974) Fuzzy Logic and Its Application to Approximate Reasoning. In: Information Processing, North-Holland Publishing Company
Zadeh LA (2002) Some Reflections on Information Granulation and Its Centrality in Granular Computing, Computing with Words, the Computational Theory of Perceptions and Precisiated Natural Language. In: Data Mining, Rough Sets and Granular Computing, Heidelberg, Germany, Physica-Verlag GmbH
Zaibert L, Smith B (2004) Real Estate - Foundations of the Ontology of Property. In: Stuckenschmidt H, Stubkjaer E, Schlieder C (eds) The Ontology and Modelling of Real Estate Transactions: European Jurisdictions, Ashgate Pub Ltd, pp 35–51