1 Introduction and Problem Statement

> *One real world is enough.*
> — Santayana

When a mapping agency updates information on topographic maps or navigational charts, usually more than one map is affected by a given change. Often, the largest (most detailed) scale maps are edited first, and the changes then transferred to the appropriate map at the next smaller scale, until either the smallest scale is reached or the edited feature becomes too small to represent. As a result of changes in the world — or interpretations of it — marks on maps may change their shape or symbology, relative prominence, locations and labels, according to organizational standards and procedures as well as expert judgement, also known as "cartographic license." This tedious graphic editing process has been described as "working through the scales," as it systematically propagates edits from greater to less detailed cartographic representations. It is one strategy of attacking problems in the art of *map generalization*, one which specialists from many nations are struggling to understand, formalize and automate.

Until about a decade ago, map generalization was practiced almost exclusively by trained cartographers, who learned it by tutelage, example and intuition, deliberately, even rigorously, but not necessarily formally. Manual cartographic generalization involves an unruly set of techniques and precepts that, traditionally, were mainly practiced in public agencies and map publishing houses, and each organization tended to have its own procedures, standards, guidelines and aesthetics for compiling maps in fulfilment of its mission. Only recently has this situation begun to change, but it has done so dramatically and definitively, as maps — along with every other form of human communication — have become digital. At the same time, the media, data, tools and enterprises involved in compiling and communicating maps are rapidly becoming global. The author of this thesis, in describing, demonstrating and assessing a specific digital approach to working through scales, hopes to contribute to progress in both map generalization specifically and spatial data handling generally.

1.1 Motivation

While the foregoing describes its application focus, this thesis has a deeper, more general concern: improving the representation of geographic space — the world at large — in digital computers. The latter topic is necessary to consider because addressing the former one will achieve limited success without re-thinking how encoding conventions for geospatial data bias the development and constrain the results of methods used to generalize it. This may not appear obvious or even necessary to many who work with environmental, cartographic and other spatial datasets on a regular basis, as the software technology for handling them seems to be relatively mature and generally effective. Furthermore, a number of standards for specifying and exchanging such data are in use; these may differ in detail, but generally assume certain encoding conventions for basic map elements that are rarely questioned anymore. As a consequence, this thesis argues, progress in map generalization, spatial data interoperability and analytic applications has been hampered due to encountering a number of vexing difficulties, many of which stem from inadequate descriptions of both geographic phenomena and the space they inhabit.

1.2 Limitations to Detail and Accuracy of Spatial Data

We start by confronting a paradox: when geographic phenomena that are spatially continuous — such as terrain, temperature or rainfall — are numerically modeled, an inherently *discontinuous* ***data structure*** (regular sampling ***grids***, or ***rasters***, sometimes called ***fields***)[1] is often employed. Yet when the geometry of inherently discontinuous, linear or sharp-edged phenomena — such as administrative boundaries, land ownership, hydrography or transportation — is encoded, it tends to be done using a *continuous* data structure capable of representing the shapes and connections between entities and their neighbors in great detail (so-called ***vector-topological*** models). Some highly interpreted geospatial data, such as land use, soils, geologic structures, classified remote-sensing imagery and results of migration and accessibility models may be represented in either vector or raster format, although making either choice requires certain compromises.

If they have not worked in the raster domain, many users of geographic information systems (***GIS***) may not be aware that digital map data they use are not as definitive and concrete as they may appear to be. Those who work with raster-based GIS and image processing (IP) systems, which store and manipulate images and data grids, need little reminding of the finite resolution of their databases. Each grid ***cell*** and pixel in every dataset they handle covers a specific amount of territory on, above or under the surface of the earth, and cannot disclose details of variations within these tiny but specific domains. And raster ***data models*** — even multi-resolution schemes such as ***quadtrees*** — cannot hide the fact that only a certain amount of detail is stored, amounting to very many discrete (but not independent) observations of single-valued attributes.

---

[1] Terms defined in the glossary, appendix B, are printed in bold italics where they first appear.

In non-raster GIS environments — with which this project is almost exclusively concerned — the discrete nature of data and the limits of its ***precision*** and ***accuracy*** tend to be hidden, and may go unnoticed, but not without consequence. Systems that employ vector-topological data models actually introduce two types of discretization, illustrated in fig. 1.1:

1. Continuous contours of spatial entities are represented by finite, ordered sets of ***vertex*** locations — although sometimes, continuous functions (such as splines or elliptical arcs) may be exploited to represent certain types of curves. The *accuracy* of such data are mainly limited by sampling during their capture;

2. Based on how much computer storage is assigned to primitive data types (integers and real numbers, which in practice usually is either 32 or 64 bits), the *precision* of individual spatial ***coordinates*** (vertex locations) is always finite.

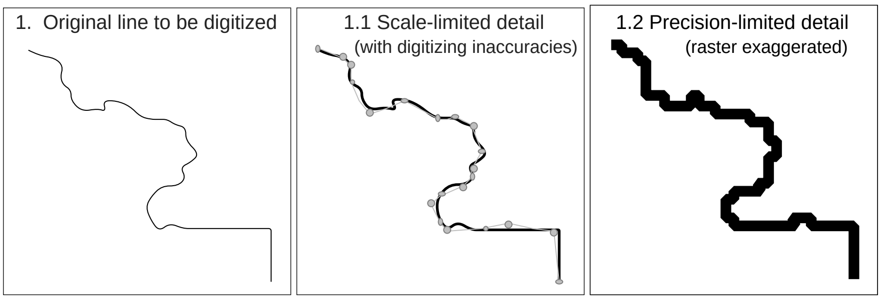

Fig. 1.1. Two types of discretization errors, along with human line tracing error

1.2.1 Scale Limitations

The first type, which we will denote as *scale-limited*[2] representation, caricatures ***lines*** and polygonizes ***regions*** by sampling networks and boundaries at small (and usually irregular) intervals. The size of intervals is mainly influenced by data capture technology and procedures and the scale and resolution of the graphic source material, which typically are aerial photographs or printed maps, captured as digital data (digitized) either automatically or manually. If assumptions are made

---

[2] In this document, *scale* is synonymous with *map scale*; a dimensionless ratio of distance on a map to the "true" distance that it represents on the earth's surface. Such ratios are always less than unity, and are notated as 1:N, where N is usually between 100 and 10,000,000. When N alone is referred to, it is termed the *scale denominator*.

regarding the smallest mark or object that a map or a display is able to represent, one may numerically relate spatial accuracy and resolution to ***map scale*** (Tobler 1988, Goodchild 1991a, Dutton 1996). The spatial data model explored in this thesis makes use of such equivalencies (see table 2.1).

GIS users and cartographers are well-conditioned by now to line segment approximations of map ***features***, and have some understanding of their limitations and of consequences to merging or overlaying digital map data. GIS system designers have come up with a variety of solutions to the "sliver problem," the generation of small, spurious shapes resulting from mismatches of line-work that represents the same features using slightly different geometric descriptions. Paradoxically, as Goodchild (1978) has shown, the more diligent and detailed is one's digitization effort, the more troublesome this problem gets. No practical, robust GIS solutions are yet available for automatically changing the scale of digital map data, or combining data from maps having major differences in the scales at which their features were captured. This is partially due to the fact that GISs — even when metadata are available to them — do not formally model the information content of the data they handle, even the critical yet restricted aspects of ***data quality*** involving spatial resolution, accuracy and positional error. In a Ph.D. thesis concerned with handling scale- and resolution-limited GIS data, Bruegger proposes an approach to data integration in which:

> ... mapping a representation to a coarser resolution becomes well-defined since the source and target knowledge content are precisely known. This compares to current GIS practice where the knowledge content of a representation is only vaguely known to the user and totally inaccessible to machine interpretation. Figure [1.2] illustrates this with an example of two vector representations of the same coast line. What knowledge about the world do the representations really contain? For example, what is found in location p? The vague definition of the knowledge content is closely related to the problem of precisely defining what it means to transform a representation from one scale to another. (Bruegger 1994: 4)

Bruegger's goal was to develop a "format-independent" representation to enable conversion between formats (specifically, raster and vector) to take place without unnecessary loss of information. Such a notation could specify the positional uncertainty of the geometry contained in a dataset as metadata equally applicable to each type of representation. This, he asserted, would avoid "problems of comparing incompatible concepts such as raster resolution and vector 'scale'"(Bruegger 1994: 4).

The latter statement is arguable, we maintain; there is nothing incompatible between these two concepts if one makes certain assumptions about what they mean in an operational sense, as they both are consequences of discretization processes, one with respect to space and the other with respect to phenomena. This aside, the goals of this thesis are in harmony with Bruegger’s, even though its approach, formalisms and application concerns are somewhat different.

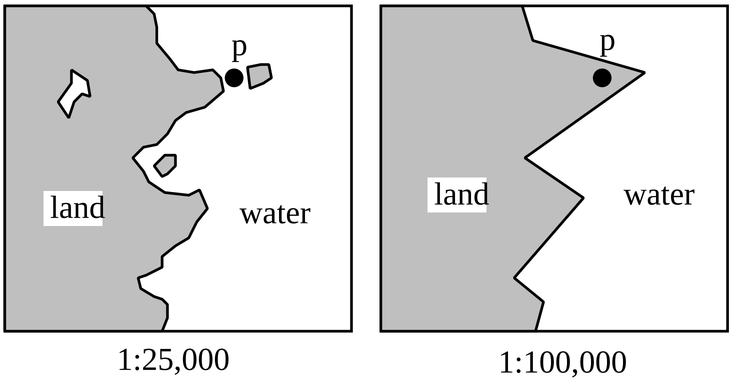

Fig. 1.2. Scale-limited representational problems (from Bruegger 1994: 4)

1.2.2 Precision Limitations

We call the second type of discretization *precision-limited*[3], as it is constrained by how many digits of precision computer hardware or software can generate or represent. This applies to data capture hardware (e.g., scanners and digitizing tables), which have finite-precision grids through which ***points*** are filtered, as well as to the word sizes supported by a computer's CPU (central processing unit) or implemented in software by programming languages. The basic issue is that geometric coordinates are often modeled as real numbers, but become represented by finite, floating point quantities that can fail to distinguish data points that are very close together and which can introduce numerical errors and instabilities, hence cause algorithms to make incorrect decisions when testing and comparing computed values. A good example of how geometric imprecision can lead to topological trouble is described and illustrated by Franklin (1984). In figure 1.3 scaling is applied to a triangle and a point, resulting in the movement of the point from inside to outside the triangle. The figure exaggerates the effect by using integer coordinates, but the principle holds true for floating point ones as well.

---

[3] Following Goodchild (1991), we define *precision* as "the degree of detail available for a measurement". Note that effective precision may be less than the theoretical maximum, due to rounding, cumulative processing errors and (in floating-point words) normalization of mantissas.

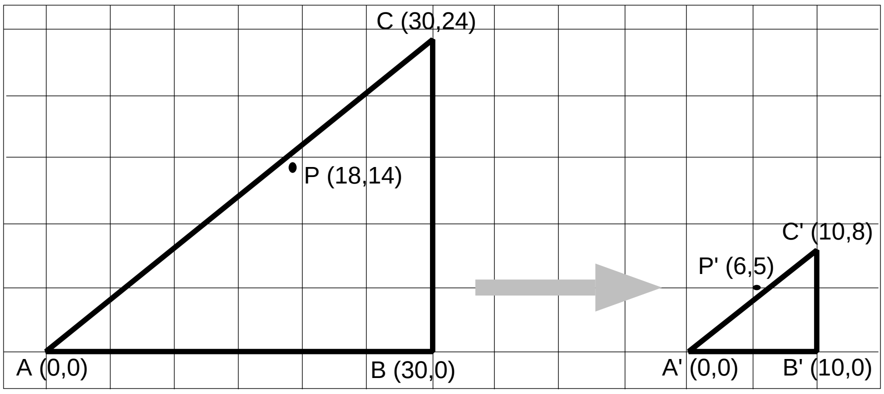

Fig. 1.3. Scaling moves point P outside triangle (from Franklin 1984)

Franklin’s paper provides other sobering lessons in the hazards of computational geometry, and provides a look at some ways to avoid them. Although numerical methods have improved considerably since he wrote, his conclusions remain valid:

> Existing computer numbers violate all the axioms of the number systems they are designed to model, with the resulting geometric inaccuracies causing topological errors wherein the database violates its design specifications. (Franklin 1984: 206)

It has found to be devilishly difficult to guarantee the robustness of algorithms that process geometric data, at least unless limiting assumptions are made about the nature and quality of input data, or should approximate answers suffice. Considerable effort in computer graphics and computational geometry has been devoted to handling the consequences of finite-precision computations in processing geometric data (Schirra 1997). While these issues are relatively well-understood, and robust techniques for dealing with some of them exist, this is not the end of the story where processing geospatial data is concerned. One reason is that many of the techniques for avoiding errors in geometric computations utilize software representation of numbers that extend their intrinsic precision generally, or as required by specific computations. While they give more accurate results than hardware floating point computations, they are hundreds or thousands of times slower to execute. This may be tolerable for some kinds of geometric applications, but GIS applications tend to be too data-intensive to make such approaches practical.

Also, even if GIS processing modules could be made provably correct in handling all the precision available for map and other locational data, its degree tends to be fixed, and is rarely sensitive to variations in the actual precision of coordinates within datasets. Finite precision is basically regarded as a nuisance, to be overcome by force, usually by allocating more bits than will ever be needed to store and process coordinate data (e.g., double-precision floating point words). This almost always fails to reflect the true amount of information that a coordinate contains about a ***geographic*** location, tending to exaggerate it.

Standardizing (hence falsifying) the precision of locations to machine word sizes may result in false confidence in the quality of computations performed on them, even if all numerical instabilities are carefully avoided. It may lead one to believe, for example, that a boundary location is known to within 2 cm, when in fact the measurement is at best accurate to about 20 meters. Some modules of some GISs do provide ways to simulate reductions to precision (by storing accuracy parameters and introducing error tolerances); these, however, are almost always based on global estimates or information about measurements (error ***attributes*** or ***metadata***), rather on the actual precision of measurements themselves. Another common strategy is to employ rational arithmetic, using integral numerators and denominators to represent real numbers. While this can ensure unambiguous results for computations such as line intersection or point-in-polygon tests, the results tend to be only as good as the least accurate coordinate involved.

1.3 Documenting Spatial Data Quality

But data — spatial and otherwise — are not only imprecise, they also (always!) contain *errors* and *inaccuracies.* The caricaturing of boundaries described above in discussing types of discretization is fundamentally a lack of accuracy[4]. Inaccuracy is inevitable, and up to a point acceptable, given that all data are abstractions of reality that by their nature must leave out details that either escape detection or would serve no useful purpose to include. If effective quality control is maintained during data capture and processing, it is normally possible to provide users with reasonable estimates of the types and magnitudes of inaccuracy within datasets. By properly interpreting such statistics, users (and software) should in principle be able to decide if a given dataset has sufficient fidelity to reality to serve a particular purpose. This is a basic tenet of the U.S. Spatial Data Transfer Standard (SDTS, USDOC 1992), and a primary impetus to the generation and cataloguing of spatial metadata, now becoming a light industry of sorts as geospatial data pops up on networked file servers around the globe.

While spatial data may be inaccurate in a variety of respects, this thesis is principally concerned with *positional accuracy,* and in particular, with how it relates to efforts to generalize topographic maps. Here is a specification for positional accuracy on map sheets published by the United States Geological Survey (USGS), which is responsible for compiling topographic maps for the U.S. at a series of scales from 1:24,000 down:

---

[4] Again, using Goodchild's definition (Goodchild 1991), *accuracy* is the degree to which measurements faithfully represent the real-world entities they purport to describe.

> The United States National Map Accuracy Standard (NMAS) specifies that 90% of the well-defined points that are tested must fall within a specified tolerance. For map scales larger than 1:20,000, the NMAS tolerance is 1/30 inch (0.85 mm), measured at publication scale. For map scales of 1:20,000 or smaller, the NMAS tolerance is 1/50 inch (0.51 mm), measured at publication scale.
>
> Converting to ground units, NMAS accuracy is:
>
> S / 360 feet = (1/30 inch) * (1 ft/12 in)*S,
> for map scales larger than 1:20,000 (= S / 1181 m)
> S / 600 feet = (1/50 inch) * (1 ft/12 in)*S,
> for map scales of 1:20,000 or smaller (= S / 1969 m)
>
> where S is the map scale denominator. (FGDC 1996)

Note that this statement only defines horizontal accuracy for “well-defined points,” locations which are visible and identifiable in the field. Consequently, it may be difficult to assess the accuracies of map features such as streams, political boundaries, pipelines, contour lines or even roadways, unless they include monumented points. Spatial metadata (FGDC 1994) may give spatial data users some hints about how well such features reflect reality, but the information may be qualitative or narrative, not readily usable for GIS processing purposes. In addition, as most metadata refer to an entire dataset, differentiation of data quality between or within feature classes is generally not possible unless it is specified via attribute coding.

The U.S. NMAS was defined in 1947, well before digital mapping evolved from a curiosity in the 1960’s to its status as an industry today. It is widely recognized that this standard is barely adequate for printed maps, and leaves much to be desired in the realm of digital map data. It is in the process of being updated by the U.S. Federal Geographic Data Committee (FGDC 1996); this will not change the basic approach, which may be practical for map production organizations, but is not especially helpful in certifying or assessing accuracies of distributed, heterogeneous spatial datasets that are increasingly created by and available to communities of GIS users. As a result, positional accuracy of most spatial datasets remains murky, and must often be assumed, guessed, or estimated. Sometimes this can be done using existing GIS functions, such as Goodchild and Hunter (1997) demonstrate, but many users would rather not know or will not take the trouble.

Even when the highest data capture standards are followed, and even when codebooks and metadata are scrupulously prepared to document datasets, the fact remains that geospatial data do not describe their internal quality variations very well, if at all. The reasons why this problem persists are many, and can be categorized as:

1 **Circumstantial**: Source data are not well controlled or documented, or are too diverse or convolved to compile data quality descriptions for them;

2 **Institutional**: Datasets are prepared for specific, limited or internal purposes, without a mandate to inform other potential data users;

3 **Structural**: No adequate mechanisms are in general use which are capable of documenting variations in spatial data quality at a highly detailed level.

The first two of these three categories are being dealt with in the GIS community by deliberate efforts, mostly through developing richer spatial data exchange formats and standards for geospatial metadata. As both public and private-sector data producers seek to add value to map data products and make them available over wide-area networks, particularly the internet, data documentation and quality control is receiving a great deal of recent and salutary attention. It is already possible to browse metadata repositories distributed throughout the planet on the world wide web (WWW), to identify (and in many cases also download) geospatial datasets with reasonable certainty about the nature of their contents.[5] But even with the best of intentions, efforts and tools, it remains quite problematic to assess the suitability of ***geodata*** for purposes and scales other than those for which their producers intended them to be used.

It is argued here that many limitations to reusability of geospatial data as well as many difficulties involved in their maintenance, are due to the third aspect of documenting their quality: structural deficiencies in datatypes and data models. The most glaring, yet largely unacknowledged, deficiency of GIS and cartographic vector data is its reliance on coordinate notation — (latitude, longitude, elevation) or (x, y, z) — to describe locations. This convention is so well-established and ingrained that it is hardly ever questioned, but without doubt it is responsible for millions of hours of computation time and billions of geometric errors that might have been avoided had richer notations for location been used. The author has articulated this previously (Dutton 1984; 1989; 1989a; 1992; 1996).

Regardless of how many dimensions or digits a conventional coordinate tuple contains, it is descriptive only of position, and does not convey scale, accuracy or specific role. It is possible to convey a sense of accuracy by varying the number of digits that are regarded as significant in a coordinate, but such a device rarely is used, but never to distinguish one boundary point from the next (Dutton 1992). In most digital map databases, the vast majority of coordinates are convenient fictions, as few of them represent “well-known points” on the surface of the Earth. Rather than identifying specific points on the Earth’s surface, most map coordinates should be considered as loci of events that led to their creation. These events

---

[5] In the U.S., any internet site that serves spatial data can be registered as a node in the National Spatial Data Infrastructure, by documenting its holdings in compliance with the FGDC’s Content Standards for Digital Geospatial Metadata (FGDC 1994). The FGDC, academic organizations and GIS vendors are making available tools which ease the burden of compiling, cataloging and searching such documents.

are partly natural (geological changes, for example), but may also be of human origin (such as territorial claims), and include activities involved in data capture and processing. Despite growing attention to error in spatial data (Goodchild and Gopal 1989; Veregin 1989; Guptill and Morrison 1995) spatial analysts, cartographers and their software tend to treat coordinates as if they have physical existence, like protons or pebbles. Most vector data structures tend to reify and democratize feature coordinates (although endpoints are usually given special node status). When processing boundaries, most applications treat a tuple that represents a specific location (such as monuments, corners and posts) the same way as a less well-defined one (such as inflections along soil boundaries, roadways and river courses), as just another point, or just another node. Their data structures have no way to express variations in positional data quality, and not surprisingly, their algorithms have no way to use such information. It is a vicious cycle, entrenching ignorance.

One could also call this attitude toward data quality the *fallacy of coordinates* (Dutton 1989), and is an example of the more general *fallacy of misplaced concreteness* ("if the computer said it, then it must be true"). The question for spatial data is, how can we tell if it's true, or more specifically, how true it might be?

1.3.1 Data Quality Information and Map Generalization

What connections might metadata and map generalization have? It is clear to a number of researchers (Mark 1989, Mark 1991, McMaster 1991) that the more information available to describe the nature of map features and their roles in a landscape, the more intelligently it is possible to treat them when changing the scale of a map or creating a map for a specific purpose. Some of this information might appear as tabular attributes to map features, or as global descriptors to specialized datasets. Knowing that a hydrographic feature is, for example, one bank of a braided stream can inform a generalization process applied to it, and potentially modify its behavior compared to how it would handle an ordinary stream centerline representing a main channel. Here is an example of specifications for coding and digitizing braided streams in the Digital Line Graph (***DLG***) format, taken from internal USGS documentation (USGS 1994):

> **050 0413 Braided stream**
> This code identifies braided streams that are shown by symbols 404(A), 541.6, 541.9 (C), or 2202.03(D). A braided stream is a special case where the stream subdivides into interlacing channels. In map compilation, where possible, the actual channels are shown. However, if the channels are extremely complex or obscured by vegetation, the outer limit is scribed accurately and the inner channels are represented by a conventional pattern. The use of pattern versus actual channel is not noted on the map. Therefore, the braided portion of a stream is digitized as an area that carries this code. The outer limits are digitized and carry left and right bank codes (see codes 050 0605 and 050 0606). The braided area is separated from a

> double-line stream by a closure line (code 050 0202) and from a single-line stream by nodes (see codes 050 0004 and 050 0005).

This USGS document takes nearly 250 pages (of which the section on hydrography accounts for 40) to describe just the attribute codes, associated symbol codes and instructions such as the above. While it contains a considerable amount of knowledge about map features and what they represent, it does not include specific advice on how, where or when features should be generalized. The amount of detail in the DLG coding guide is impractical to provide as file-specific metadata, but it could be turned into a knowledge base (KB), formalized as production rules or via other schemata, given sufficient effort and motivation. Once such a KB for digital database production is built, additional rules for generalization can be added incrementally, to the extent they can be derived from formal descriptions or actual cartographic practice.

As the above DLG guideline states, precisely locating a stream bank may not be feasible in places where vegetation obscures it. It is quite common for positional uncertainty of boundaries to change along a given feature for this and a variety of other reasons, such as construction or other earth moving activities, ecological succession, the presence of wetlands, and when map data from different sources is merged in developing a GIS database. When the level of uncertainty of a feature (or portion of one) changes from the norm for its data layer or feature class, most GISs — although they could — tend not record this, as it requires adding error attributes that will only be occasionally germane, and which probably would not be usable as parameters to existing commands and processes anyway. The author's research grapples with this problem, and provides a way to deal with it in as much detail as possible — at each inflection point along a curve.

1.3.2 Encoding Geospatial Data Quality Information

The notion of providing positional metadata for every coordinate location seems to imply that every point in every feature could be sufficiently unique to warrant its own metadata. Given the volume of coordinates in many GIS databases and the rather tiny amount of information that most of them provide, this may seem like an excessive amount of overhead. That, however, can be regarded as an implementation decision, which need not constrain the way in which one thinks about managing and modeling different aspects of geographic space. With this in mind, we shall describe an approach to encoding geospatial data that describes positional certainty independently of the schema used for representing spatial entities. In order to provide a context for this discussion, however, a brief description of logical elements of GIS databases may be useful. Here is a more or less conventional view of geospatial data that has a great many existing implementations in the GIS literature and industry:

Features = Identifiers + Geometry + Topology + Attributes + Metadata

In turn, these terms can be dissected too:
Identifiers = Names + ***Geocode***s + Spatial_Indices
Geometry = Identifiers + Coordinates + Other_Properties
Topology = Identifiers + ***Genus***_Properties + Invariant_Spatial_Relations
Coordinates = Spatial_Metric + Tuples_of_Scalars
Attributes = Identifiers + Data_Items + Metadata
Metadata = Data_Definition + Data_Quality + Other_Properties

Most GIS databases include most of these elements, but more specialized geospatial applications (desktop mapping packages in particular) may omit or ignore some of them, especially topology and metadata. There seems to be a trend, in addition, in GIS database design to leave out explicit topologic information, replacing it with attribute data, spatial indices or regenerating it as needed, on the fly (ESRI 1996; ORACLE 1995; Jones et al 1994).

Many different implementations of this general approach (or at least some parts of it) exist. The most common approach groups features of a given class together as a *layer* (or *coverage*) which may include topological relations between them but is independent of other feature classes. To relate different classes, layers must be geometrically and topologically combined, a process called *map overlay*. While this thesis is not directly concerned with overlay techniques, it is important to understand their implications for data quality management. Whenever spatial data from different sources are integrated, the lineage of the result becomes heterogeneous. Map overlay works at such a fine-grained scale that many individual features in the resultant layer contain data from two sources (or even more if overlays are cascaded). Attributes and metadata for the resultant layer can indicate what sources formed it, and carry over descriptions of their quality, but they cannot easily indicate which portions of particular features came from what source. Therefore, should the positional accuracies of two inputs to an overlay operation differ, the accuracy of the result will vary spatially in uncertain, uncontrolled and undocumented ways. Figure 1.4 illustrates one basis for this pervasive problem.

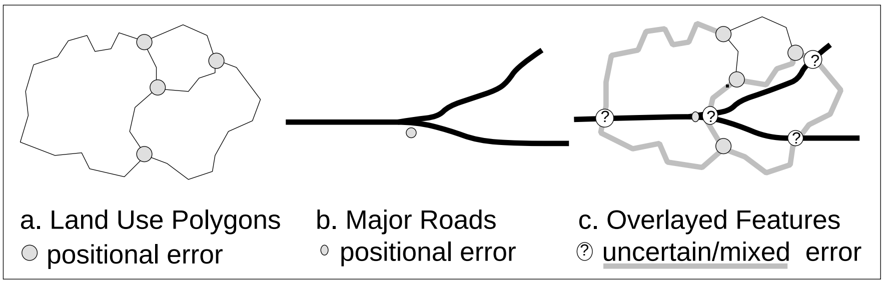

Figure 1.4: Overlay of data with differing positional accuracy

Some GISs model features as sets of primitive geometric elements ("Primitives") which are stored together or separately, and are linked together to form semantic elements ("features") either via identifiers or topology. For example, the Earth's land masses may be represented as a set of polygons that describe continents, islands and lakes. A motor way can be a set of linked arcs, and a village can be modeled as a set of point features. Even a relatively small feature, such as a family farm, may consist of multiple primitives of several types and include many topological relationships. In general, such "complex features" may consist of groups of points, lines and areas, with or without explicit topology. Obviously, how features are encoded and modeled has strong implications for what operations can be applied to them, and how easy these are to perform. Certainly it is not simple to describe the details of data quality variation where complex features are modeled, just as it is difficult for overlaid data. Alas, there is no optimal data model for geospatial data in general, although there may be nearly optimal ones for restricted sets of data in the context of specific applications. This is one reason why GISs all have different (and usually proprietary) data models, which may be difficult to translate from one system to another.

The problem of documenting heterogeneous data quality is largely unsolved. The U.S. standard for geospatial metadata (FGDC 1994) was a major stride forward in documenting the quality and other aspects of GIS data, as was that country's Spatial Data Transfer Standard (USDOC 1992) before it (the latter was ten years in the making, and is already somewhat obsolete). But both standards deal with data at the dataset level (leaving the scope of datasets up to producers), and make no real provisions for providing finer-grained qualifying data, aside from feature attribute coding conventions. GIS users, of course, are free to embed lower-level metadata in their systems wherever and however they may, normally by creating tables and text associated with specific layers, feature classes and features. This information might be keyed to feature identifiers to link it to geometric objects. The left side of figure 1.5 shows this type of schema, where lower levels of metadata record changes in data quality from the norms for higher levels of abstraction, hence need not be specified for any "normal" object.

There are several problems with this approach, discussed below and in (Dutton 1992). However defined, such metadata records are "auxiliary" data structures, stored separately from descriptions of the geometry and topology of spatial features. Should a set of features be transferred from its host system to another GIS, while metadata records could be provided as well, the receiving system might not be able to use some or all of them, depending on how standardized the metadata records and similar and robust its data modeling capabilities were. Object-oriented systems make this easier to do, but object models must be mapped from one schema to another, not a trivial problem in the general case. Finally, users and producers are burdened with providing (or estimating) data quality information for a complete hierarchy of data elements, not just an entire dataset, unless tools are crafted to determine the metadata and automate their insertion, including formatting them and linking them to associated data entities.

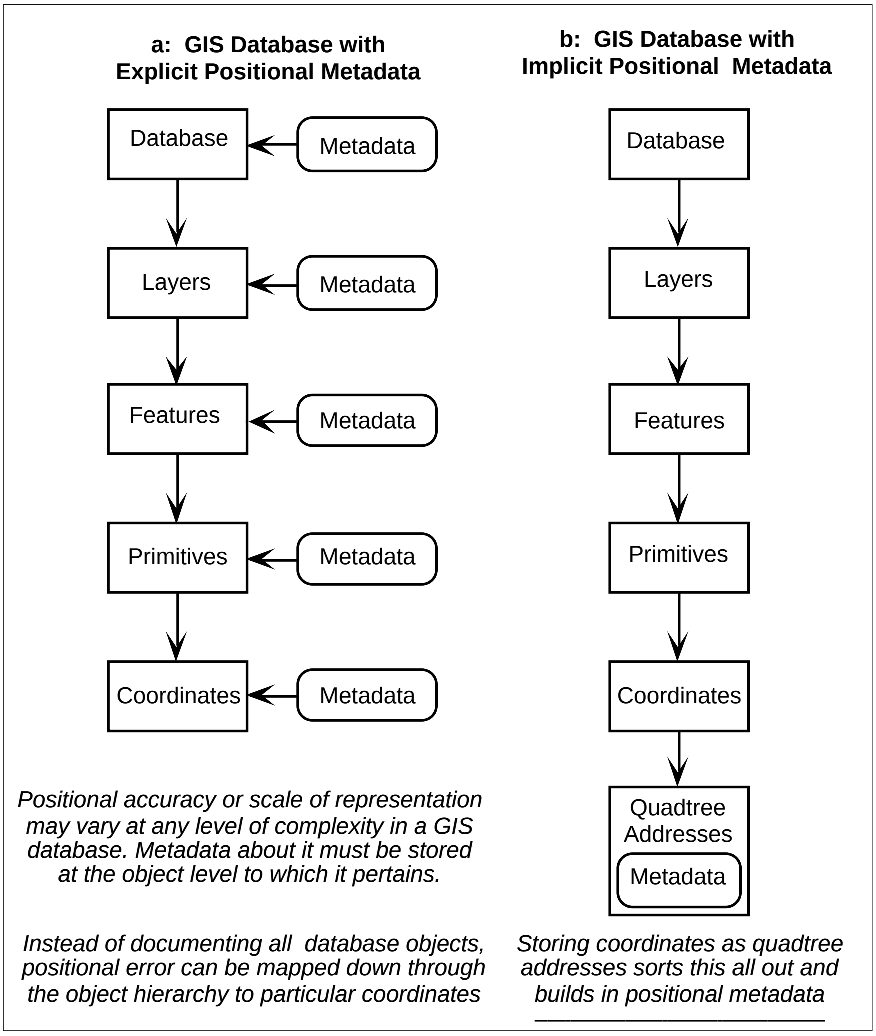

Fig. 1.5. Alternative ways to encode positional metadata (from Dutton 1992)

An alternative way to handle storage of *positional data quality information* (not necessarily all metadata) is illustrated on the right side of figure 1.5, and is described in subsequent chapters. The author believes strongly that this approach provides a much more effective way to embed positional metadata into GIS databases, putting it where it is most useful; at the spatial coordinate level. Properties and consequences of this approach will be explored throughout this document.

1.4 Challenges of Digital Map Generalization

Generalization of digital map data has received an increasing amount of attention in the past several years, not only from the GIS community but from computer scientists as well. This is a multi-faceted problem which will require a multiplicity of approaches, and will keep researchers busy for a number of years to come. Most of the research and solutions in this area are concerned with a single sub-problem, that of *line simplification*, for which a large number of algorithms have been developed. Other work has focused on characterizing point features, aggregating areal features (including administrative units, land use polygons and groups of buildings), text placement, reclassifying thematic values and collapsing feature hierarchies. Besides simplification, other manual techniques (generally called generalization *operators*) that have been identified and automated include *selection*, *smoothing*, *displacement*, *aggregation*, *amalgamation*, *collapse*, *exaggeration*, *enhancement* and *typification* (McMaster and Shea 1992; Brassel and Weibel 1988). These operators are sometimes given different names and definitions in the literature (some may be ignored entirely); professional cartographers are not at all agreed on how to characterize the fundamental activities these terms denote, as a survey by Rieger and Coulson (1993) documented. The most general agreement seems to be that generalization begins with *selection* of data to be displayed, after which other operators come into play.

A special journal issue (Weibel 1995) offers papers from well-known researchers examining a number of aspects of digital map generalization, including line simplification techniques and much more. A recent book from a GISDATA workshop also provides an overview of generalization problems and solutions in the GIS context (Muller et al 1995). Most past work has dealt mainly with the isolated generalization of single features or even portions of features, one or several at a time. But most maps depict more than one feature class or layer, and at a small enough scale a given feature can come into conflict with ones from any other class anywhere in its locality, in arbitrary and complex ways. This has led to a more recent emphasis on *holistic feature generalization* (Ruas and Plazanet 1996), which attempts to assess interactions among features. This is a much different (and more difficult) problem in a vector environment than in a raster-based GIS or image processing system. This is because most map generalization problems result from *competition for map space*, and in vector-based system space is not modeled, phenomena are. Unless features are contiguous, and their topology encoded, extensive analysis may be needed to determine which features are in conflict, especially if they inhabit different classes or layers. In order to handle these kinds of scale-sensitive, combinatorial problems, the trend is for researchers to exploit greater numbers of more complicated data structures, which must all work in concert as they solve specific generalization tasks.

Among the data structuring techniques being investigated to support holistic map generalization are Delaunay triangulations — often constrained by the inclusion of several feature classes at once (DeFloriani and Puppo 1995; Jones et al 1992 and 1994; Ruas 1995). Also being exploited are hierarchical structures for vector data, such as KDB-trees (Robinson 1981), R-trees (Guttman 1984), and many related and derived approaches (Robinson 1981; Cromley 1991; Devogele et al 1996; van Oosterom and Schenkelaars 1995). These techniques are often combined with one another, as well as with other ways to logically integrate and structure multiple versions of map features, such as directed acyclic graphs (Timpf and Frank 1995). The general goal of most of this activity is the creation and exploitation of multi-scale geographic databases, both for cartography and spatial analysis. Jones and Abraham (1986) and van Oosterom (1993) offer overviews of the problems, opportunities and some possible solutions in this increasingly complex area of GIS design.

1.5 Scope of Work

Based on the intuition that generalization of spatial data can be improved by development of multi-scale databases, a specific approach to multi-scale data storage, retrieval and manipulation has been designed, implemented and tested in the course of this research project. A software prototype has been built which, while not a full GIS by any means, provides many of the basic data modeling functions found in one. This testbed has enabled us to assess the essential utility, effectiveness and efficiency of a geoprocessing system based on hierarchical positioning, in the context of performing basic map generalization tasks.

While it is based on a general-purpose, global data model, our project has not attempted to assess its suitability for all GIS and cartographic purposes, only for cartographic line generalization. Nor, as much as we would have liked to, can we claim to have developed more than a partial solution for problems of holistic map generalization. Our fundamental aims are quite specific and relatively modest: *to enrich the content of geospatial data at the coordinate geometry level, so that common operations on data are supported more by the data themselves rather than requiring external metadata, or elaborate post-processing and complex algorithms to cope with missing information*.

Still, there are some senses in which our approach is more universal than existing vector and raster data architectures: it is global (planetary) in scope, and represents features at a hierarchy of resolution (scales), properties which make it applicable to a broad range of geospatial applications, not just world-wide ones, and not simply for cartographic purposes. And while the new generalization methods developed by this effort seem to work quite well, perhaps the basic contribution to map generalization is more diagnostic than prescriptive. Using the a hierarchical framework may make it easier to detect and negotiate conflicts for map space, but additional hard work will be needed to translate these insights into working systems that yield a long sought-after solution: generating acceptable map products across a range of scales from a single detailed geospatial database.

The potential utility of the approach to location modeling described in this thesis extends beyond map generalization, to indexing of point, vector and raster terrestrial and astronomical data, modeling global environmental phenomena and processes, and multi-scale terrain modeling. All of these areas have been addressed by researchers exploring hierarchical polyhedral planetary representations, some derived from the author's work, others closely related to it. The scope and context of these investigations are summarized in the next chapter, following a descriptive overview of our specific approach to modelling geographic location. Differences in purposes and backgrounds of cited researchers cause some inconsistencies in the organization and nomenclature of their models, which are all basically quite similar, varying only in a few basic respects.

Chapter three descends into greater detail to describe salient geometric, topological and numerical properties of the model. Details concerning notation and coordinate conversion are presented (pseudocode for the latter is provided in appendix A). Also dealt with is *spatial indexing*; a derived hierarchical construct for this purpose is introduced and discussed in this context, with some implementation details provided in chapter 4. Use of our model to index global point data is summarized in appendix E. Chapter 3 concludes by illustrating how hierarchical methods can both ascertain and verify important aspects of positional data quality for cartographic data.

Chapter 4 narrows the focus to map generalization, describing a feature-oriented data model and some hierarchical algorithms designed for characterizing and filtering cartographic detail. It starts with an overview of digital map generalization, providing a context for our own approach to line simplification. Also discussed are assessing and conditioning line data and approaches to identifying conflicts among map features due to scale change. Details of an operational data model are presented, followed by those of a line simplification algorithm for hierarchical coordinates and the control parameters it uses. The chapter closes with an explanation of the generation and handling of point-specific attributes (metadata), showing how they are used to guide point selection for line generalization.

Chapter 5 describes how generalization methods presented in chapter 4 were implemented, and along with appendices C and D, presents results of generalization experiments using two shoreline datasets. The software and hardware environment used for development and testing is described, as well as the methods used for data handling and presentation. Results of processing each of the test files are discussed and illustrated, followed by an evaluation of strengths and weakness of our methods in relation to the dominant line simplification paradigm.

In chapter 6, we summarize what our empirical studies seem to say about a major tenet of cartographic generalization, the notion of ***characteristic points***, looking at it from several perspectives, including scale imperatives, point weeding, local and global generalization strategies and evaluation of results from the algorithms compared in chapter 5. We then propose a counter-intuitive reconstruction of this notion, along with a way to implement it. The chapter concludes with a research agenda for refining our approach to make it a more robust basis for multi-scale database construction and automated map generalization.

1.6 Chapter 1 Summary

This chapter has attempted to describe types difficulties that can result from poor or undocumented positional data quality. These need to be overcome in order to properly integrate, update, analyze and display geospatial data in a GIS environment. While many aspects of data quality obviously affect these activities, the research reported here is focused on improving the descriptive power of locational notation. In particular, the assertion is made that reliance on spatial coordinate tuples, such as latitude and longitude, constrain the ability of software to manage geodata, and that it is possible to minimize this shortcoming by using a hierarchical approach. The particular application focus of the research, cartographic generalization, has been chosen to demonstrate how map scale and spatial resolution can be linked in a data model that documents the precision of feature locations and permits them to be accessed across a range of spatial resolutions and map scales.

The remainder of this first chapter focused on how digitization of andcoordinate notation for spatial geodata hampers its use. First, several types constraints arising from the process of discretization were described; limitations of scale of capture and the inherent numerical precision of computer hardware and software both are shown as contributing to lack of certainty about location. The next section discussed measuring and documenting positional data quality. After presenting an overview of typical elements of GIS datasets, a set of reasons were described that contribute to poor documentation of spatial accuracy in geodata, and why this is inevitable given how such data are conceived of and structured. These problems were put into focus by considering the difficulties in generalizing digital map data, in which locational uncertainty may vary within features in ways that data structures are unable to document. This was further addressed in a brief review of current map generalization research directions, in which one common task is to add contextual information to map data, by enriching temporary or archival data structures; this focus on data enrichment is maintained throughout the thesis. The chapter ended with a plan of the thesis, indicating what the project has sought to accomplish, describing topics covered, problems addressed and solutions presented in each subsequent chapter.