

## MULTISCALE MAPPING FOR THE NSDI: DATA MODELING AND REPRESENTATION

*Lee De Cola*

*U.S. Geological Survey*

*521 National Center*

*Reston VA 22092*

*LDECOLA@USGS.GOV*

*Barbara P. Battenfield*

*Department of Geography*

*State University of New York*

*Buffalo NY 14261-0023*

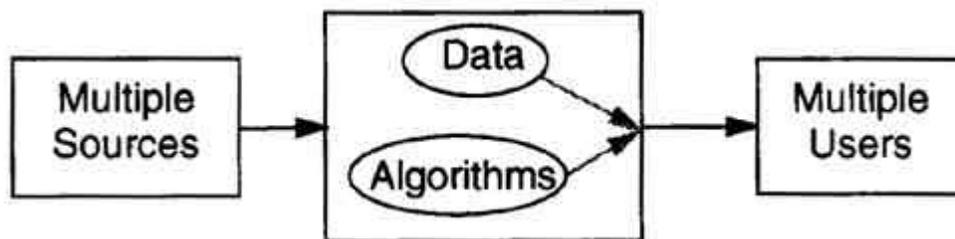
*geobabs@ubvms.cc.buffalo.edu*

### Abstract

The National Spatial Data Infrastructure system integrates spatial data from many sources into products for describing, mapping, and planning for the Nation. Two major challenges of this system are combining different kinds of spatial data from disparate agencies and providing these data at multiple scales or even ranges of scales. The transformation of a road network taken from one data source into the representation of a built-up area for a different user illustrates an approach to these challenges.

### Introduction

The National Spatial Data Infrastructure (NSDI) is intended to be a seamless multiscale geographic data framework (Mapping Science Committee 1993) (see figure\*). Expanding the NSDI requires building, upon the discrete 1:24,000- and 1:100,000-scale U.S. Geological Survey\*\* (USGS) map and digital data products (Battenfield 1994). A smoother multiscale continuum along which data can be accessed for geographic information system (GIS) representation and analysis must be developed.



Two factors need to be considered. First, the nature and appearance of geographic data change depending on the resolution at which they are collected and processed. For example, on a large-scale map a city might be represented as a pattern of buildings, at an intermediate scale the city

---

\*This paper has two kinds of figures: those in the text will be referred to by their location within the paper, and those illustrating the technical procedures are numbered on two pages below.

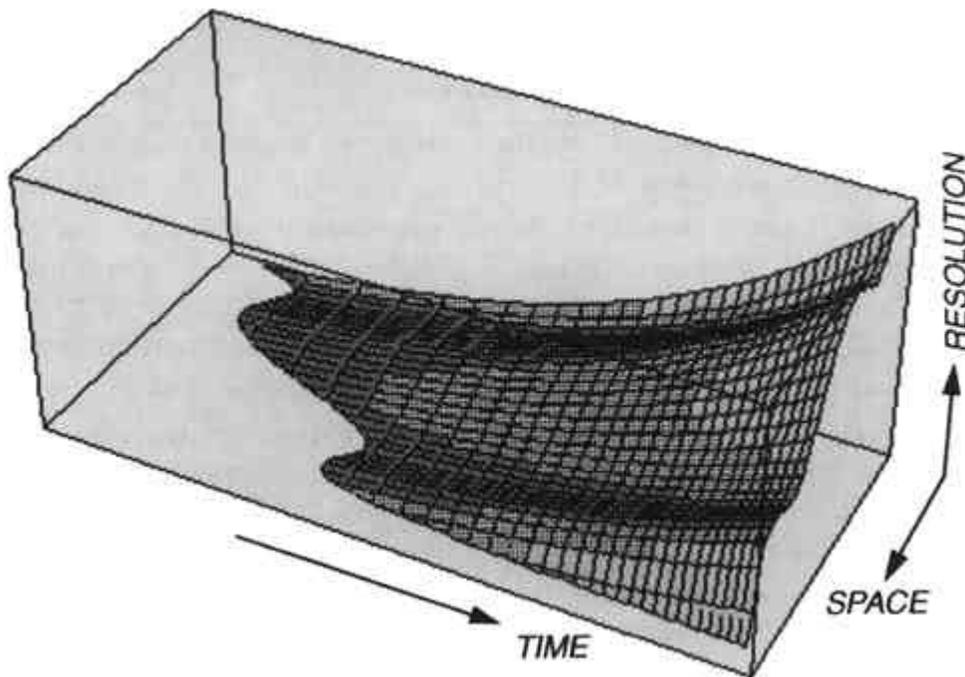
\*\*Any use of trade, product, or firm names in this publication is for descriptive purposes only and does not imply endorsement by the U.S. Government.

might be represented as a polygon, and at a small scale it might be just a labeled point. Second, different

kinds of data need to be integrated appropriately even if they have been derived from sources at varying scales. These problems require refining data models and developing scale-changing algorithms that are appropriate to the geographic theme and to the source and target scales of the data (Buttenfield and McMaster 1991). The mapping of urban areas illustrates these issues well.

### **Modeling and mapping urban areas**

Any geographic phenomenon may be regarded as existing at several conceptual levels, from the general to the concrete. (For particularly abstract models see e.g. Berry 1964, Harvey 1989, or Nicholis and Prigogine 1989.) Perhaps the most formal spatial model treatment of the city would be a set of points located in space and time. At the atomic level this "dust" of points (Falconer 1990) would be people, and although the typical census approach is to enumerate families and buildings, these social and material aggregations introduce significant definitional problems. In the abstract this dust could be called "the urban area," but it would be impossible to represent cartographically because points have no extent, so this collection of objects would be more a database representation than a map.



But geographic information management requires a much more concrete representation of the city. A given data object can be regarded as existing in the four dimensions of feature, resolution, location, and time (see figure above). Simply put, we decide what to look for, what details shall be recorded, and where and when we shall look. The figure illustrates how the envelope that bounds cartographic data is expanding with time as more areas are mapped to increasingly finer resolutions, creating a complex "skyline" at any given time. The figure is intended to show that for a given feature class, over time our data about the world covers a wider physical range of

locations in space as well as a wider range of resolutions. The NSDI is a system of multiple sources providing data for multiple users, with a set of algorithms used to transform data within this feature space. The challenge therefore is how to coordinate these transformations; this paper shows an example of how this might work for urban areas.

Within the Federal Government there are several different representations of urban areas, from the suggestive

to the specific (table 1). National Ocean Service (NOS) hydrological charts, for example, suggest urban areas as a pattern of streets with only significant navigational landmarks shown accurately positioned. Indeed, the need for even the street pattern in these charts is being evaluated (NOS, oral communication). The National Oceanic and Atmospheric Administration (NOAA) publishes Federal Aviation Administration (FAA) aeronautical charts that represent "populated places" as yellow polygons enclosed by thin black lines. These symbols indicate to pilots the general location of densely settled areas, lighted regions at night, and areas where residents may be particularly sensitive to noise (NOAA, oral communication). The Bureau of the Census designates "urbanized areas" as contiguous census tracts having a density  $\geq 1000/\text{mi}^2$  within 1 mi of a populated place (Torrieri 1992). This is perhaps the least ambiguous definition, but not the most phenomenologically representative.

**Table 1: Federal urban-area models**

<b>Agency</b>	<b>Geometry</b>	<b>General model</b>	<b>Reference</b>
NOS	network	generalized road pattern + navigational landmarks	NOS 1990
FAA	polygon	settled areas	NOAA 1991
Census	polygon	nested blocks...up to metropolitan statistical area	Census 1984
USGS	polygon	building exclusion area	USGS 1984

On its 1:24,000-scale maps the USGS uses the location of structures to delineate built-up areas as polygons, according to the following rules:

1. Map the location of all dwellings and industrial structures, showing smaller buildings as (closed or open) squares and landmark buildings as orthogonal polygons;
2. If the area delineated by the structures is larger than a given area, then generalize the cluster as a gray polygon representing "building exclusion area" or built-up area;
3. Then exclude all but large, significant (landmark) buildings from the built-up area.

These polygonal regions are depicted on 1:24,000-scale maps as pink polygons and on 1:100,000-scale maps as gray polygons without any buildings and outlined by a thin black line, within which major roads are represented using a reduced line width. On the 1:250,000-scale maps these polygons are further generalized. The topographic map series thus provides multiple scale representations of urban areas according to the models shown in table 2.

**Table 2: Urban areas in USGS map series**

<b>Scale 1:</b>	<b>Geometry</b>	<b>Model</b>
24,000	point	buildings
	polygon	built-up area
100,000	polygon	built-up area (See figure 10)
250,000	polygon	generalized built-up area
2,000,000	point	location

There are several drawbacks to this system. First, it requires much work on the part of photointerpreters who must locate each structure. Second, the process demands considerable judgment by editors who must later delineate the built-up polygons. Third, a map may display such paradoxes as buildings from an earlier map

edition appearing in the photorevised "building exclusion area" of a later map. Fourth-and a major problem for an NSDI-the process provides representations at fixed scales rather than within multiscale ranges. Indeed, the method is highly scale specific because it is designed to produce an optimal representation at 1:24,000 scale rather than an easily generalizable object for the urban area. For example, on 1:100,000-scale series the built-up area is mapped if its "[m]inimum size is .75" x .75" [1.9 km on the ground] or its equivalent area [3.6 km<sup>2</sup>] provided the shortest dimension exceeds .50" [1.27 km]."

### **A "heterogeometric" approach to generalization**

One approach to a seamless generalization scheme would be to incorporate each of these representations into the system and then devise some way of "morphing" each into the other in a kind of highly complex spatial interpolation (Muller and Zeshen 1992). It would be useful to explore this approach, but in the context of an NSDI it is more appropriate to illustrate how data for one kind of feature (representing an individual point in the four-dimensional feature space discussed above) might be transformed into another kind of feature (at some other point in the space). Indeed, mathematically this is just what is meant by a "mapping."

The USGS representation of roads-and of the built-up areas often associated with such networks-is a complex problem that has developed for more than a century during which mapping has evolved from a fieldwork-based activity to one using digital imagery and preexisting GIS data (Mark 1991). The approach illustrated here is based on the development of a digital representation of the urban area using the road network rather than buildings from images or other sources. (Although the discussion that follows is not extremely technical, we suggest that the reader scan the numbered figures on the next page-arranged from top to bottom-to get an overview of the proposed technique.)

Not only is the road network an indicator of high density technostructure (Taube 1985), but it is also the infrastructure that precedes all other such development. (In fact, when faced with an extensive road network in other-



Figure 1. Image of Redlands.

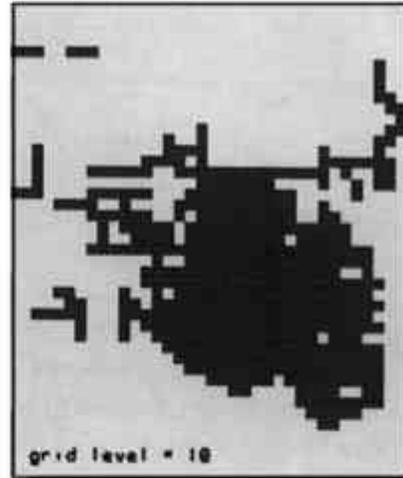


Figure 4. Streets gridded to 150 meters.

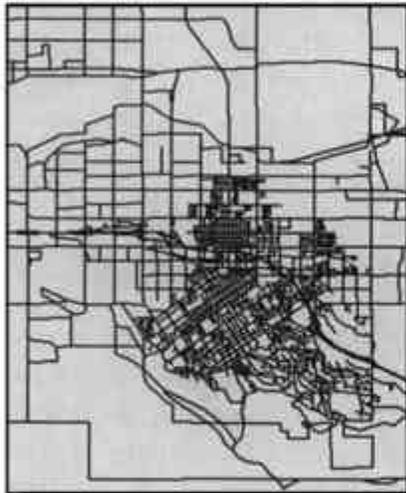


Figure 2. All Redlands roads.



Figure 5. Simplified street grid.

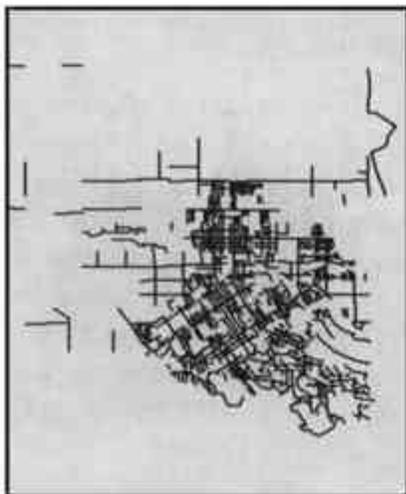


Figure 3. City streets.

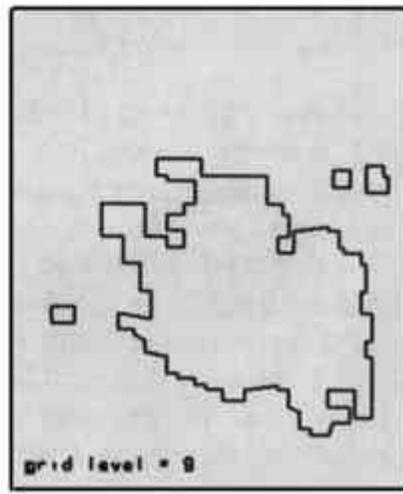


Figure 6. Orthogonal arcs.



Figure 7. Gridded boundary arcs.

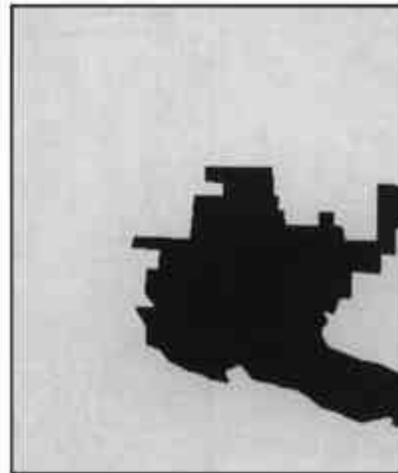


Figure 10. Redlands from 100,000-scale map.

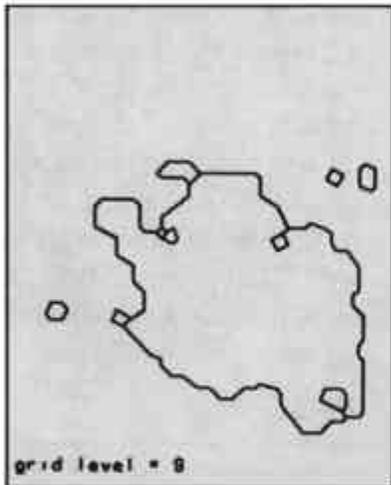


Figure 8. Octagonal boundary.

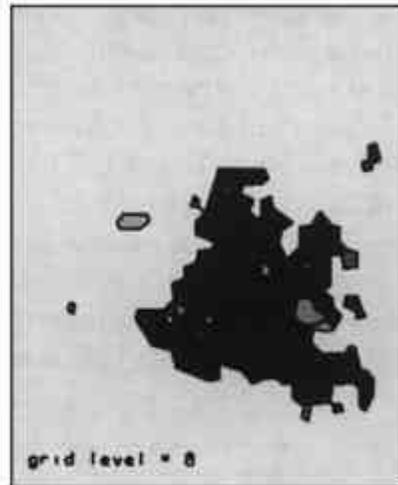


Figure 11. Level-8 gridding.

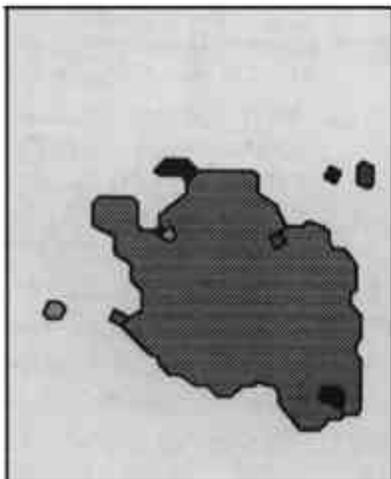


Figure 9. Octagonal polygons.

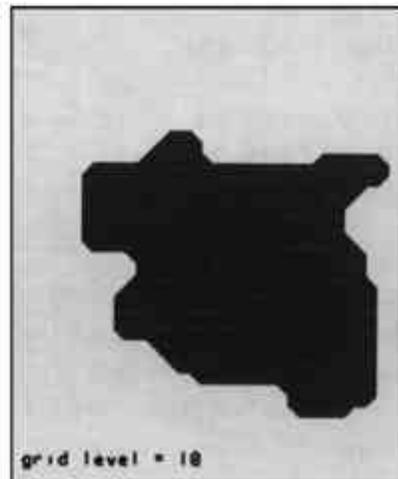
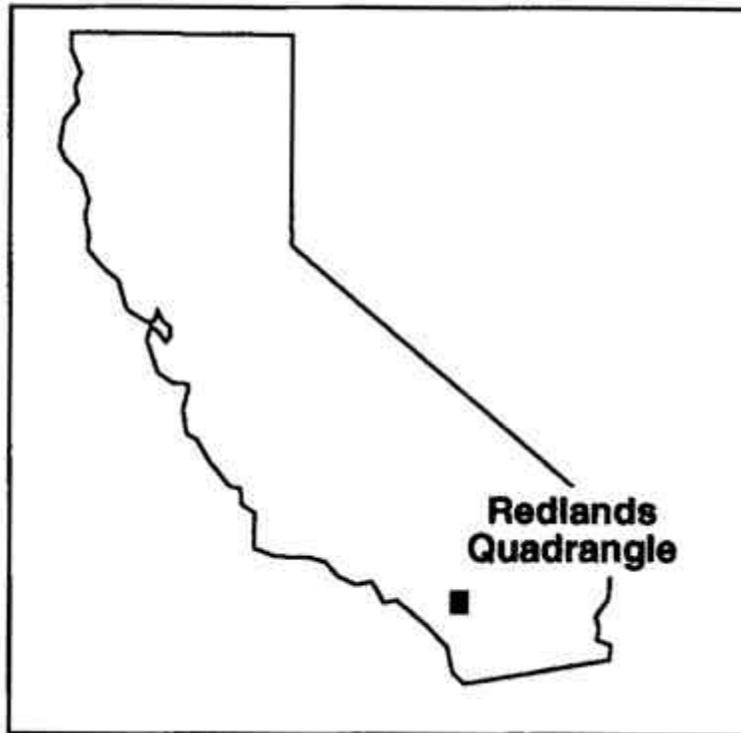


Figure 12. Level-10 gridding.



wise undeveloped space, photointerpreters will often delineate the region as built up even in the absence of buildings.) Assume that we require a representation of the city as a set of polygons to be depicted on a 1:100,000 scale map but that would also be appropriate at a range of scales "around" this scale level. In the USGS DLG data, roads are represented as six classes of lines, as shown in table 3, abstracted from USGS Technical Instructions (1985). This scheme is a hierarchy that reflects such well-ordered properties as government numbering, number of lanes, traffic type, and vehicle volume. As such, it can provide a useful alternative to buildings as a base for representing urban areas.

**Table 3: Road classes in USGS maps**

CLASS	1:24,000	1:100,000
1 Primary highways	Show	Show
2 Secondary highways	Show	Show
3 Light-duty roads	Show	Show outside of built-up areas
4 Unimproved roads	Show if it "provides access to a mapped feature..."	"Show all connecting...and those that lead to a mapped feature"
5 Trails and special service roads	Show if .24 in [146 m] long	Show if at least .63 in. [1600 m] long
6 Streets	Show	Show all Class 1 and 2 highways.

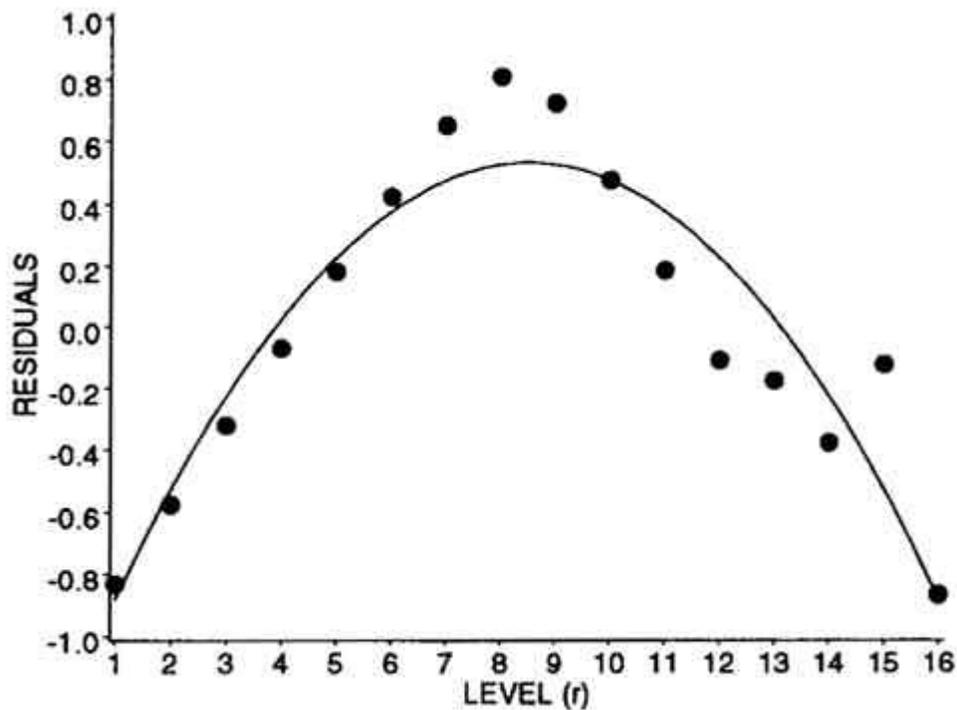
Consider the problem of representing the built-up area for Redlands, a rapidly growing edge city of the Southern California megalopolis (see figure above). Figure 1 shows a panchromatic raster image of the area corresponding to the Redlands 7 1/2" quadrangle. This area was chosen because several ARC/INFO datasets were available for the region, which contains the headquarters of the Environmental Systems Research Institute (ESRI).

To illustrate one way of defining the urban area in a multiuser/multiscale context, consider the following approach. Let there be a network of topologically 1-dimensional lines each of class  $c$ , where lower values of

class c represent higher capacity links. Figure 2 shows all of the classes of roads and streets for Redlands, and figure 3 shows just the city streets (level 5 in the training dataset). These data are from the ETAK Corporation, which based them on updated USGS transportation data (ESRI oral communication).

Because a dense road network is a usually reliable indicator of urban development, the network can be used to make a cartographic representation of the city. The problem is that we want to create a topologically 2-dimensional representation of the urban area (one or a few polygons) from a set of topologically 1-dimensional streets. A common geometric ground for these sets is that of fractals, and one way to transform the latter into the former is to determine the fractal dimension of the street lines (Lam and De Cola 1993).

Perhaps the simplest way to compute the fractals dimension of a set is to cover it with a (usually regularly spaced) set of boxes at varying resolution sizes and then count the number of boxes at each resolution level.\* At each aggregation level  $r$  we cover the road network with a set of boxes of side length  $\delta = 2^{-r}$ . Then for each  $r$  the number of boxes  $N(\delta)$  that cover the set can be related to  $\delta$  by  $N(\delta) = A(\delta)^{-D[b]}$  where  $D[b]$  is the boxcount dimension. A linear regression of the model  $\log N(\delta) = a - D[b]r$  fits our data quite well, with  $R^2 = .992$  and a boxcount fractal dimension  $D[b] = 1.26$ , suggesting that the road network is more nearly linear than space-filling.



Real phenomena are multifractal: at some scales sets have different  $D$  than at others. Standardized residuals from the above regression (see figure above) suggest that at around  $r = 9$  (corresponding to about 150 meters) the aggregated road network shifts to a different spatial regime.

---

\*Base 2 is used here mainly for convenience (especially for machine-level operations), but any base greater than 1 will do.

Indeed, this is roughly the scale of the city blocks of the region, so that this multifractal behavior reflects the coalescing of the grid sets into more coherent polygonal structures.

From the fractal analysis it appears that level  $r = 9$  is the minimum box-covering level at which the road

network can be covered by a compact and connected region. One obvious way to create polygons that represent this road network is to use box-covering (the ARC/INFO LINEGRID operation) to cover the network with a grid. Figure 4 illustrates what happens when the street network is covered at level  $r = 9$ , corresponding to 150 meters (the effects of  $r = 8$  and 10 are shown in figures 11 and 12).

Table 4 outlines the operations that are used to create a polygonal coverage of a road network. This scheme is based on the liberal use of heterogeometric tools that perform transformations between arcs, polygons, and grids. Figure 5 shows the area after low-dimensional fingers and regions have been removed from the data, and figure 6 represents the arcs that outline this "cleaned region. This representation may be quite adequate for many purposes: it retains the orthogonal nature of many urban road networks. But it is possible to recursively grid and skeletonize (figures 7 and 8) these arcs, resulting in the somewhat greater directional freedom shown in the polygons of figure 9.

Table 4: Operations used to create various representations

OPERATION	OUTPUT	PARAMETER	RESULTING DATA	FIG
image	raster	10-m.resol.	image of quadrangle	1
digitize	arc	weed tolerance	arc coverage of roads	2
SELECT	arc	street-code 5	level-5 roads	3
LINEGRID	grid	size = $2^{\text{level}}$	grid representing roads	4
BOUNDARYCLEAN	grid	none	grid with small regions deleted	5
GRIDPOLY	arc	none	orthogonal arcs outlining grid	6
LINEGRID	grid	size = $2^{\text{level}}$	orthogonal grid boundary lines	7
GRIDLINE	arc	none required	octagonal arcs outlining region	8
BUILD	polygon	none	octagonal polygons	9

## Conclusion

This exercise has demonstrated the use of various GIS tools to create a multiscale representation of urban areas from an existing street database. For example, the source data could just as well have been the Bureau of the Census Topologically Integrated Geographic Encoding and Referencing (TIGER) data. Although the process involves a number of steps, the scheme is completely automated, requiring only the specification of three parameters: the road classes to be gridded, the resolution of the initial grid

cells, and the resolution of the grid cells that cover the resulting polygon. Moreover, the technique is completely consistent, avoiding the significant variations in results that are inevitable with human judgment.

The cartographic quality of the resulting product can be evaluated in several ways. We can examine its similarity to the existing 1:100,000-scale representation of the built-up area for this quadrangle, shown for comparison in figure 10. We can also compare the level-9 results to those of levels 8 and 10 (figures 11 and 12), suitable for larger or smaller scale representations. Finally, we can compare these various representations to other ways of visualizing the city, such as images, Census maps, idealized surfaces, and so forth.

But the essential point of this exercise is not so much to produce a cartographic representation of a given quality, for indeed no such measures of quality have been established in the present instance. Rather, the goal has been to demonstrate the feasibility of using one data representation from one source to produce a quite different feature for another possible user. Such an approach-in which customers can seek from a distributed database not only data (streets) but also new features (urban areas from streets) as well as new techniques

(streets plus algorithms to make new features)--lies at the heart of a useful National Spatial Data Infrastructure.

## References

Berry, Brian J.L. 1964 Cities as systems within systems of cities, in Friedmann and Alonso 1984, pp. 116-137.

Buttenfield, Barbara P. 1994 Object-oriented map generalization: modeling and cartographic considerations, European Science Foundation Meeting on Generalization, Compiègne France.

\_\_\_\_\_ and Robert B. McMaster 1991 Map generalization: making rules for knowledge representation, NY: Longman.

Falconer, K.J. 1990 Fractal geometry: mathematical foundations and applications, NY: Wiley.

Friedmann, John and William Alonso 1964 Regional development and planning, Cambridge MA: MIT Press. 722 pp.

Harvey, David 1989 The urban experience, Baltimore: Johns Hopkins.

Lam, Nina S. and Lee De Cola 1993 Fractals in geography, Englewood Cliffs NJ: Prentice-Hall, 308 pp.

Mapping Science Committee 1993 Toward a coordinated spatial data infrastructure for the nation, Washington: National Academy Press.

Mark, David M. 1991 Object modelling and phenomenon-based generalization, in Buttenfield and McMaster 1991 pp. 103-118.

Muller, J.C. and Wang Zeshen 1992 Area-patch generalization: a competitive approach, The Cartographic Journal 29(2):137-144.

NOAA 1991 NOAA aeronautical chart user's guide, 3rd ed., October Rockville MD: NOAA.

NOS 1990 Nautical chart symbols abbreviations and terms, Washington DC: National Ocean Service.

Nicholis, G. and Ilya Prigogine 1989 Exploring complexity: an introduction, New York: Freeman.

Taube, M. 1985 Evolution of matter and energy on a cosmic and planetary scale, NY: Springer-Verlag.

Torrieri, Nancy and Joel Sobel 1992 The urban/rural dichotomy: an overview of current criteria and future research, The Operational Geographer 10(2):46-48.

U.S. Bureau of the Census 1984 1980 Census of Population: population and land area of urbanized areas for the US and Puerto Rico: 1980 and 1970, Washington DC: Census.

U.S. Geological Survey 1984 Standards for 1:50,000 and 1:100,000-scale county maps, Reston VA.

U.S. Geological Survey 1985 Standards for 1:100,000-scale quadrangle maps, Reston VA.

---