Chapter 7: Digital data

Joaquin Cavieres

Geoinformatics, Bayreuth University

Introduction

Many spatial data currently exist in digital forms.

Roads, political boundaries, water bodies, landcover, soils, elevation, and a host of other features have been mapped and converted to digital spatial data for much of the world.

Many spatial data currently exist in digital forms.

Roads, political boundaries, water bodies, landcover, soils, elevation, and a host of other features have been mapped and converted to digital spatial data for much of the world.

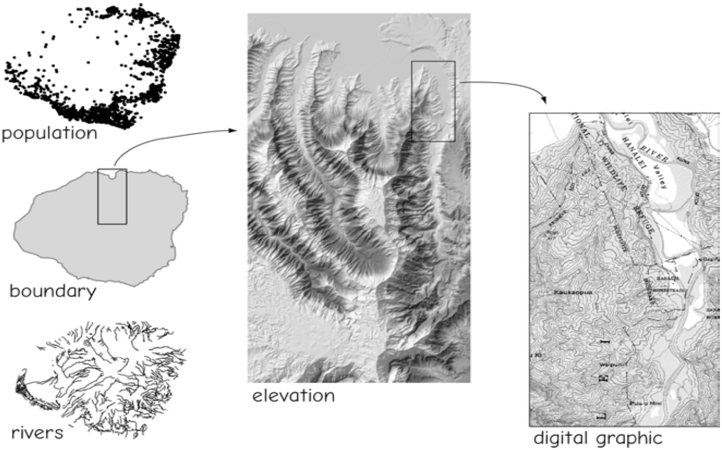

Figure 1: Examples of free digital data available at a range of themes, extents, and scales. Vector (left), raster (middle), and georeferenced digital graphic data (right) are shown for Kauai, Hawaii, USA (Bolstad (2016)).

UNIVERSITÄT BAYREUTH

Data are increasingly collected in digital formats and they are directly transferable to other digital devices and GIS systems, where they could be further processed. So, direct digital collection eliminates data transfer to viewable physical media such as maps or written lists, and then converting these media to a digital form.

Data are increasingly collected in digital formats and they are directly transferable to other digital devices and GIS systems, where they could be further processed. So, direct digital collection eliminates data transfer to viewable physical media such as maps or written lists, and then converting these media to a digital form.

Digital data are developed by governments because these could helps to:

- Basic public services
- Safety
- Health
- Transportation
- Water
- Energy

Especially, spatial data are required to:

- Planning and management
- National defense
- infrastructure development and maintenance

Thus, many national, regional, and local governments have realized that once these type of data could also be quite valuable for use outside government.

Finally,

Business, nonprofit, education, and science, as well as governmental bodies can be benefited from the digital spatial data.

Finally,

Business, nonprofit, education, and science, as well as governmental bodies can be benefited from the digital spatial data.

For the above, we must distinguish between data that are available for:

1. Transfer to, storage on, and manipulation in a local computer (locally storable)
2. Data available as a Web services, including Web Mapping Service (WMS), Web Feature Services (WFS), and Web Coverage Services (WCS).

GIS software accesses data via an internet connection, displaying these data on a local machine, although they are “served from some remote computing system.

Most data through services are currently provided as WMS, with few systems supporting and using the editing/analysis functions available through WFS and WCS.

National and global digital data

National governments commonly develop, organize, archive, and distribute national data sets and those digital data could be obtained from a central source via the internet (Fig 2).

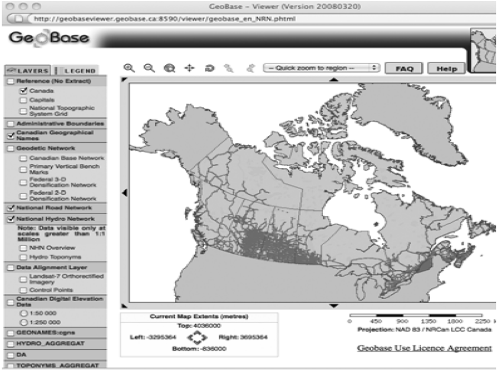

Figure 2: Australian national government, provides a suite of data layers that can be accessed via website or requested in hardcopy form (Bolstad (2016)).

Global data sets are also available but are less common than national data sets.

Different governments specify different datums, standard map projections, data variables, and attributes, or have different requirements for survey accuracy or measurement units.

Global data sets are also available but are less common than national data sets.

Different governments specify different datums, standard map projections, data variables, and attributes, or have different requirements for survey accuracy or measurement units.

Global data sets have often been developed using global satellite data at a relatively coarse resolution (e.g., MODIS or VEGETATION canopy cover at one to eight kilometer cell sizes).

University centers or ad hoc collaborations are other rich sources of global data, for example, Natural Earth data (http://www.naturalearthdata.com/) (Fig 3)

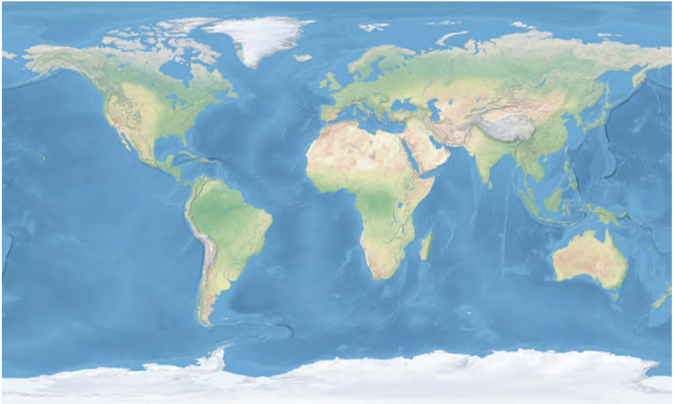

Figure 3: High-quality data suitable for small-scale mapping from the Natural Earth data project (Bolstad (2016)).

The global spatial data sets generally are organized around a theme. For example, the Max Planck Institute in Germany has led an effort to create a gridded data set of historical global precipitation by combining data from 40,000 meteorological stations in 173 countries (http://gpcc.dwd.de).

OpenStreetMap is one notable effort to develop global data through international volunteer collaboration (Fig. 4).

Figure 4: An example of OpenStreetMap data for an area in Galicia, in northwestern Spain (Bolstad (2016)).

Unfortunately, there are potential drawbacks with these data.

- Quality documentation and uniformity could be lacking.
- Data could not be complete

Given these drawbacks, the data should be verified for accuracy and completeness.

Global digital elevation data

Relatively high-resolution global digital elevation models are available from three sources:

- Shuttle Radar Topography Mission (SRTM) $\rightarrow$ (oldest)
- ASTER satellite
- DEM data

For example, next figure (Fig. 5) was obtained from the satellite pair, TerraSAR-x and TanDEM-x (DEM data)

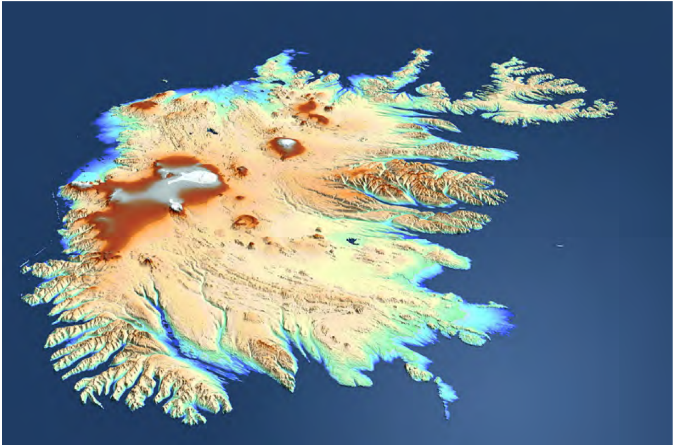

Figure 5: Example of digital elevation data obtained from the TandDEM-x satellite, here of the main island of Iceland (Bolstad (2016)).

Digital data for the United States

National spatial data infrastructure

The United States has defined the National Spatial Data Infrastructure (NSDI). The goal of the NSDI is:

- Reduce duplication of effort between agencies
- Improve quality and reduce the costs of geographic information
- Make geographic data more accessible to the public
- Increase the benefits of available data
- Establish key partnerships with states, counties, cities, tribal nations, academia, and the private sector to increase data availability

The framework developed for this consists in:

- Geodetic control
- Orthoimagery
- Elevation
- Transportation
- Hydrography
- Governmental unit boundaries.

A primary goal of the NSDI is to foster the efficient development of these core data.

A framework for the data are to be created and maintained by a diverse set of organizations, but will be submitted to the National Geospatial Data Clearinghouse (NGDC) for certification.

The internet page of NGDC (`http://clearinghouse1.fgdc.gov/`) is perhaps the best current resource for geospatial data for the United States.

The U.S. national map

Digital data are available for most of the United States: (http://nationalmap.gov)

Data are provided on political and civil boundaries, transportation, hydrography, geographic names, structures, elevation, aerial photographs, and land-cover (Fig. 6).

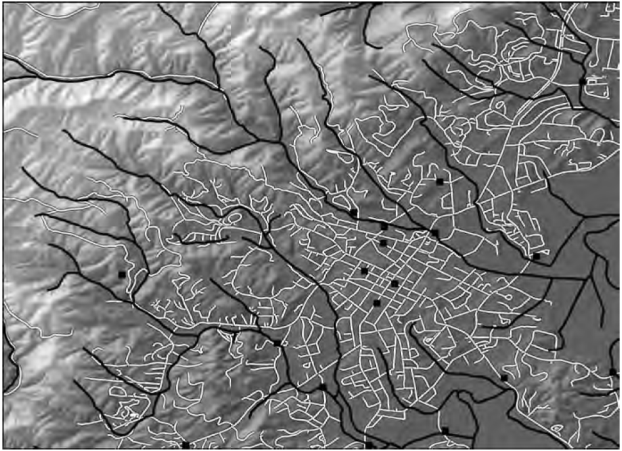

Figure 6: An example of spatial data available through the U.S. National Map, here elevation, road, river, and government building data for an area near Brevard, North Carolina (Bolstad (2016)).

Besides hydrography, the national map also distributes transportation, structures, and boundaries data in vector formats.

- Transportation data represent roads, railroads, airports, and other transportation features.
- National map boundaries data identify national, state, county, and Native American lands, as well as the boundaries for cities and towns.

Data for the national map come from a variety of sources, including new primary data collections from aerial and satellite images contributed by federal and state agencies, and older data.

Digital elevation models

Digital elevation models (DEMs) provide elevation data in a raster format and are available at 10 m or better resolution for all states (except Alaska), and are used commonly in data analysis (Fig. 7)

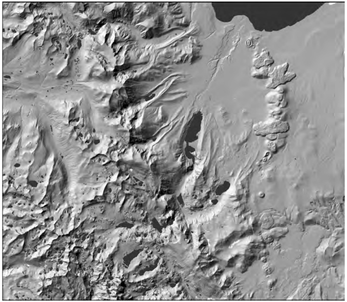

Figure 7: Digital elevation models (DEMs) are available at various resolutions and coverage areas for most of the world (Bolstad (2016)).

Commonly these methods are relatively slow and provided a sparse network of points. A dense network suitable for elevation mapping is possible only in small areas.

Although there have been some improvements, these technologies are too slow to be the sole elevation data collection method.

GIS users should be cautious because there are several versions of DEM data for many areas, and they should generally use the more current, higher accuracy, or higherresolution data (Fig. 8).

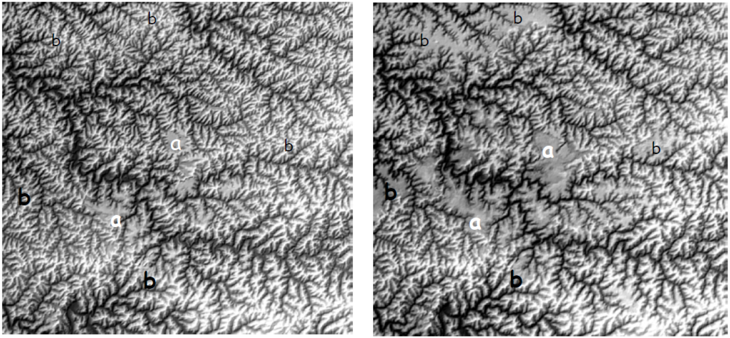

Figure 8: The extent of mountaintop removal strip mining in eastern Kentucky is evident in this comparison of older NED data (left) and year 2000 SRTM elevation data (right) (Bolstad (2016)).

Hydrologic data

The National Hydrologic Dataset (NHD, and NHD Plus) contains digital spatial data about surface waters, including rivers, streams, canals, ditches, lakes, ponds, and springs (Fig. 9).

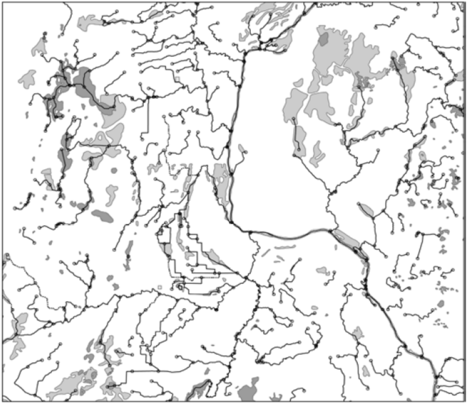

Figure 9: An example of a number of feature types are represented in an Hydrologic data, including stream segment endpoints (unfilled circles), connected stream networks (solid lines), water bodies (dark polygons), and adjacent wetlands (grey polygons) (Bolstad (2016)).

These include water bodies, canals, pipelines, dams, and other natural or control structures (Fig. 9).

NHD data also represent network topology, the connection among stream features, and include information on connections and flow directions.

NHD data are organized by areas, in a hierarchically nested set of Hydrologic Units, identified by unique codes (HUCs). These units correspond to watersheds, or basins, or logical aggregations or subareas of watersheds (Fig. 10).

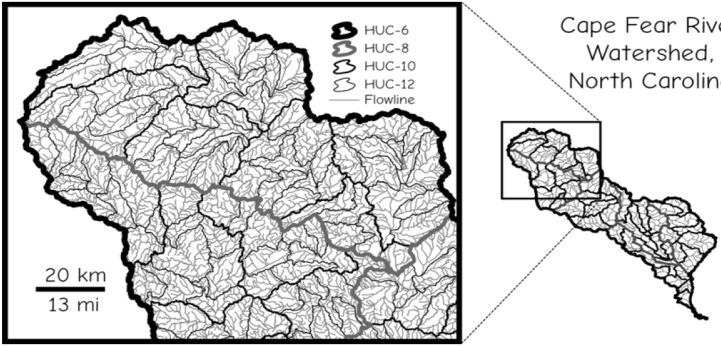

Figure 10: An example of nested HUC drainage areas for a portion of the Cape Fear River in North Carolina (Bolstad (2016)).

The United States EPA also provides data on waters and watersheds of various types and formats, organized to correspond to the HUC data at some levels.

EPA River Reach Files organize data in a series of versions, from RF1 through the most recent and detailed RF3 data.

RF3 data are designed to provide a nationally consistent hydrographic database that records the geography and assigns unique identifiers to all surface water features.

Figure 11: River reach data (RF1 through RF3) contain location and connectivity data for river and stream segments (Bolstad (2016)).

There are other improved hydrologic data, called NHDPlus. It is focused on improving the accuracy, consistency, and tools to support NHDPlus data, with work on version 2 of NHDPlus that began in January, 2011.

Digital images

Digital images are available from a range of sources, including national, state, and county governments, or from private contractors andsatellite imaging companies.

High-resolution digital image data are typically collected every five to ten years by the USGS, in partnerships with states or other government agencies.

The High Resolution Orthophotographs (HROs) are among the highest resolution, widely available image data sets.

These data are often collected at 0.3 m (1 ft) resolution, and at times up to 10 cm (4 in) resolution. They are orthophotographs, thus corrected for tilt and terrain distortion at ground height (Fig. 12).

Figure 12: An example of a High Resolution Orthophotograph (HRO), distributed by the U.S. Geological Survey (Bolstad (2016)).

Since those photos record the surface at a fixed point in time, they can be used to create new maps or to monitor change (Fig. 13).

Figure 13: An example of a historical aerial photograph from the 1940s (left) and 2008 (right), for an area in east-central Minnesota (Bolstad (2016)).

NAIP digital images

The images of the National Aerial Imagery Program (NAIP) are mainly to monitor agricultural landscapes. The images are typically collected from June through August and are useful as a base for digitizing, particularly when information on vegetation type or condition is important (Fig. 14).

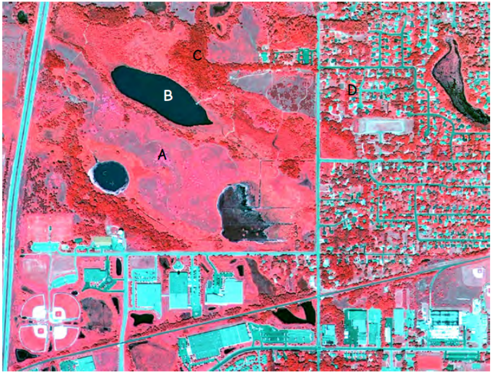

Figure 14: An example NAIP image showing wetlands (A), lakes (B), forest (C), and residential areas (D) (Bolstad (2016)).

National landcover data

While landcover is important when managing many spatially distributed resources, data on landcover are quite expensive to obtain over large areas.

The National Land Cover Database (NCLD) is the most recent and detailed source of national landcover information (Fig. 15).

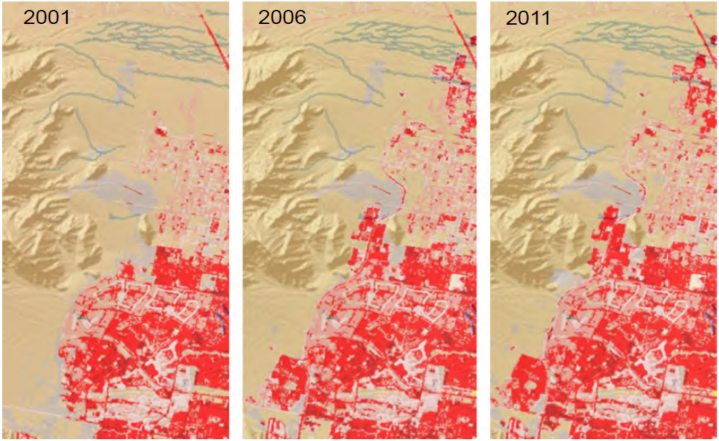

Figure 15: A spot scanning system. The scanner sweeps an instantaneous field of view (IFOV) in an across-track direction to record a multispectral response (Bolstad (2016)).

NASS CDL

The National Agricultural Statistical Service (NASS) produces yearly Crop Data Layer (CDL) data, landcover maps that focus on distinguishing major crop types and rotations (Fig. 16).

CDL data produced annually for most regions, allowing analysis of trends in planting, crop rotations, and harvest.

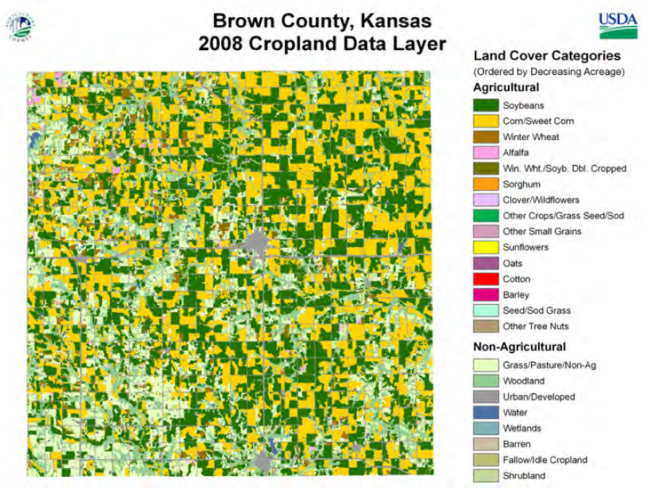

Figure 16: An example of NASS Crop Data Layer information on agricultural landcover, here for a region in eastern Kansas (Bolstad (2016)).

National wetlands inventory

Data on the location and condition of wetlands are available for much of the United States through the National Wetlands Inventory (NWI) program.

NWI data describe the extent and characteristics of wetlands, including open water (Fig. 17).

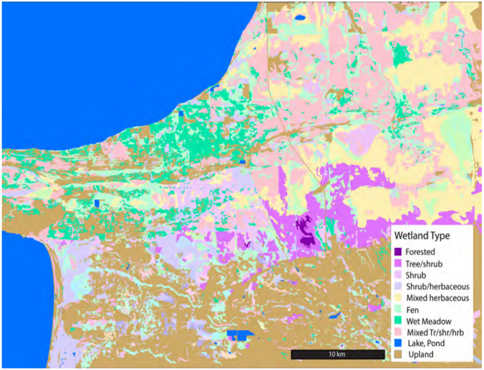

Figure 17: An example of national wetlands inventory (NWI) data (Bolstad (2016)).

Digital soils data

There are three digital soils data sets developed by the Natural Resource Conservation Service (NRCS). They differ in the scale of the source maps or data, and thus in the spatial detail and extent of coverage. These are:

- The National Soil Geography (NATSGO)
- State Soil Geographic (STATSGO)
- Soil Survey Geographic (SSURGO)

NATSGO: The data set is a highly generalized soils map for the continental United States, developed from small-scale maps.

STATSGO: The data are intermediate in scale and resolution

SSURGO: The data provide the most spatial and categorical detail.

SSURGO maps indicate the geographic location and extent of the soil map units within the soil survey area (Fig. 18).

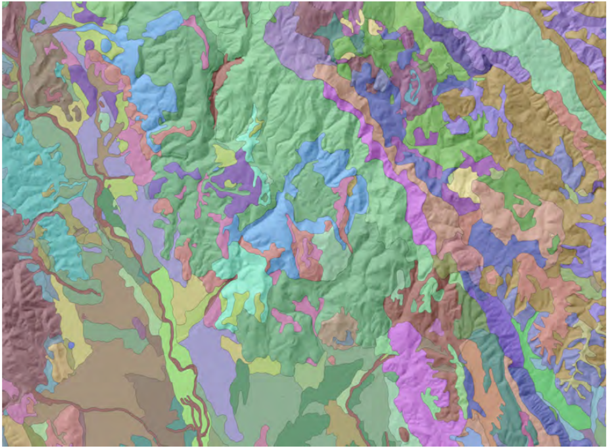

Figure 18: Landcover and land use classification is a common application of satellite images. The spectral reflectance patterns of each cover type are used to assign a unique landcover class to

SSURGO data are more appropriate for broader-scale application, such as identifying areas most sensitive to erosion, or planning land use and development.

STATSGO digital soil maps are smaller scale and cover broader areas than SSURGO soil data (Fig. 19).

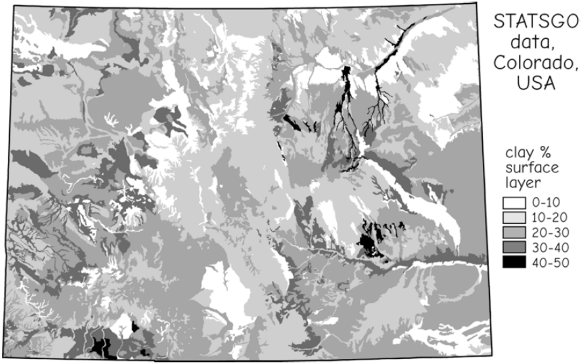

Figure 19: An example of STATSGO data for Colorado (Bolstad (2016)).

If SSURGO data are not available, STASGO data could be generated from a combination of topographic, geologic, vegetation, land use, and climate data.

STASTGO map units are larger, more generalized, and don't necessarily follow the same boundaries as SSURGO map units.

If SSURGO data are not available, STASGO data could be generated from a combination of topographic, geologic, vegetation, land use, and climate data.

STASTGO map units are larger, more generalized, and don't necessarily follow the same boundaries as SSURGO map units.

Climate, geology, and environmental data

The National Climatic Data Center (NCDC) maintains historical climate records and provides their data through a Web portal (http://gis.ncdc.noaa.gov/maps).

Mineral resources data are available from the United States Geological Survey, at http://mrdata.usgs.gov/ (Fig. 20)).

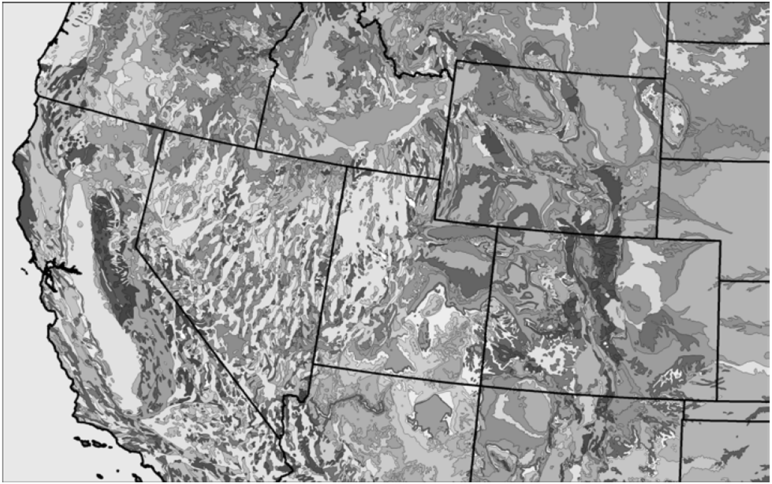

Figure 20: A general geologic map of the United States, based on spatial data from the USGS (Bolstad (2016)).

Or, for example, climate data have been converted to spatial fields, and are distributed through the PRISM initiative (`prism.oregonstate.edu`, see Fig. 21).

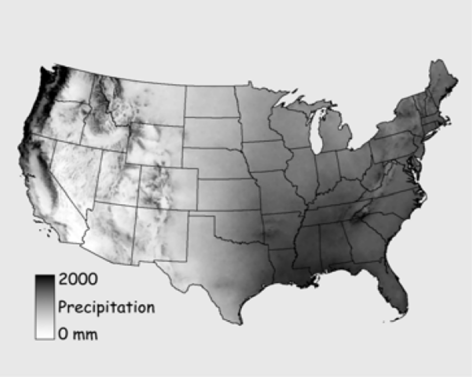

Figure 21: U.S. average precipitation, measured from 1971 - 2000, interpolated from weather stations across the United States to create a raster grid (Bolstad (2016)).

Digital census data

The United States Census Bureau developed and maintains a database system to support the national census. This system is known as the Census TIGER system.

The census TIGER system links geographic entities to census statistical data on population size, age, income, health, and other factors (Fig 22).

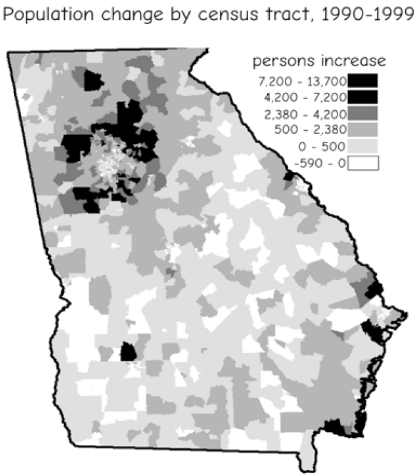

Figure 22: Laser mapping systems operate by generating and then sensing light pulses (Bolstad (2016)).

Polygon features include census tabulation areas such as census block groups and tracts, and area landmarks such as parks and cemeteries. Point landmarks such as schools and churches may also be represented. Points, lines, and polygons are used to define these features (Fig. 23).

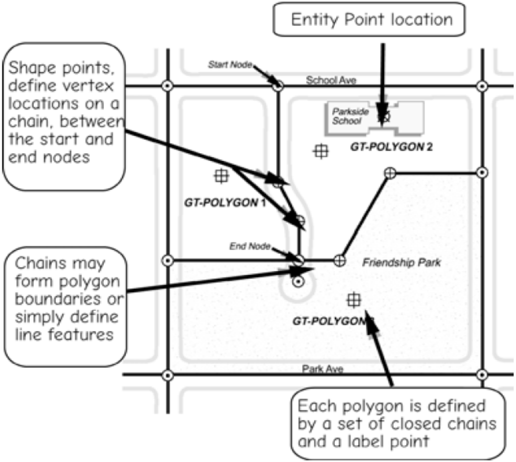

Figure 23: TIGER data provide topological encoding of points, lines (chains), and polygons (U.S. Dept. of Commerce) (Bolstad (2016)).

TIGER/Line files contain information to identify street address labels. Starting and ending address numbers are recorded corresponding to starting and ending nodes (Fig. 24).

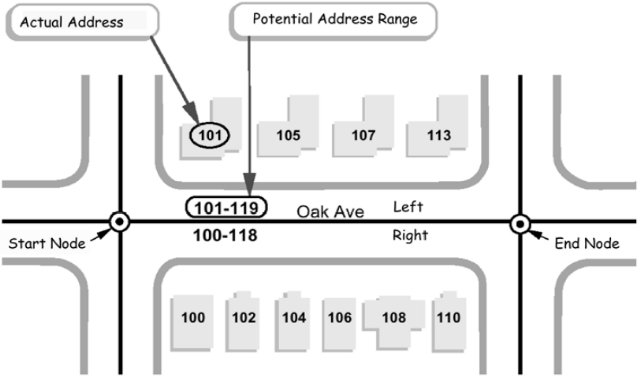

Figure 24: TIGER data provide address ranges for line (chain) segments. These ranges may be distributed across the line, giving approximate building locations on a street (Bolstad (2016)).

Many United States government data sets are provided with codes compatible to United States Census Bureau data. For example, data are delivered in census-compatible units and codes by the United States Department of Education (Fig. 25).

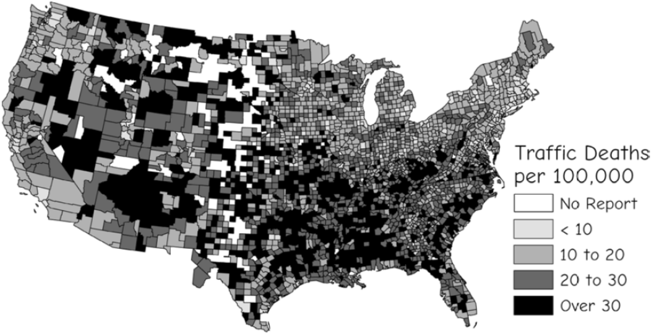

Figure 25: Traffic fatality data, showing average number of deaths per 100,000 persons over 1997-2006, derived from data reported by the U.S. Centers for Disease Control (Bolstad (2016)).

Thank You

References I

Bolstad, P. (2016). *GIS fundamentals: A first text on geographic information systems*. Eider (PressMinnesota).