

# An Introduction to Freely Available Street Network Data<sup>1</sup>

Hartwig H. Hochmair and Dennis Zielstra<sup>2</sup>

## Introduction

Projects in agricultural and natural resource management, urban planning, and community development, typically use some kind of spatial data for analysis and mapping. These tasks are often accomplished within a GIS (Geographical Information System), which is a software platform that stores, analyzes, manages, and visualizes spatial data. Part of the data will typically be collected for the project through various data collection methods. For example, point data of reptile or termite sightings from field surveys and the extent of forest damage caused by a hurricane can be assessed through remotely sensed image data from unmanned aerial vehicles. Besides these project specific data, other data will often be readily available from existing data repositories. Examples are aerial photographs, which provide realistic background mapping of a project site, or street network and elevation data which can help to model the accessibility of a given location. Various applications and websites exist, such as Google Earth or Bing Maps, which allow users to view spatial data and perform some basic spatial operations (e.g., compute the distance between two locations). This document focuses on data sources that allow users to download free street data for further processing and analysis.

In this document we use the terms *street network* and *road network* interchangeably. Such networks describe systems of connected lines and points that facilitate land-bound transportation with different modes. These modes include

motorized transportation, such as car or truck, and non-motorized transportation, such as cycling or walking. Streets can be grouped into different hierarchies according to their functions and capacities, including freeways, arterials, collectors, and local roads. GIS data are found both as freely accessible datasets and as proprietary datasets from vendors for purchase. Examples are satellite images, household survey data, or street data, which vary in data quality and price, depending on the data source. In this document we give an overview of two freely available street network datasets, describe how to retrieve them, and explain aspects of their data quality.

The rapid development of navigation systems in cars, cell phones, and other mobile devices that use GPS technology had a big impact on the demand of accurate street network data. Most manufacturers of navigation systems rely on the expertise of commercial street data providers whose data need to be purchased with the navigation system. Commercial street data can also be purchased as an independent product, where the costs vary with spatial extent and completeness of the data but can quickly exceed thousands of dollars. As an alternative resource there exist publicly accessible street data sets that come at no cost.

1. This document is FOR281, one of a series of the School of Forest Resources and Conservation Department, UF/IFAS Extension. Original publication date March 2011. Revised April 2014. Visit the EDIS website at <http://edis.ifas.ufl.edu>.

2. Hartwig H. Hochmair, assistant professor of geomatics; and Dennis Zielstra, PhD candidate and research assistant; School of Forest Resources & Conservation, Fort Lauderdale Research and Education Center, UF/IFAS Extension, Gainesville, FL 32611.

The Institute of Food and Agricultural Sciences (IFAS) is an Equal Opportunity Institution authorized to provide research, educational information and other services only to individuals and institutions that function with non-discrimination with respect to race, creed, color, religion, age, disability, sex, sexual orientation, marital status, national origin, political opinions or affiliations. For more information on obtaining other UF/IFAS Extension publications, contact your county's UF/IFAS Extension office.

U.S. Department of Agriculture, UF/IFAS Extension Service, University of Florida, IFAS, Florida A & M University Cooperative Extension Program, and Boards of County Commissioners Cooperating. Nick T. Place, dean for UF/IFAS Extension.

## Freely Available Data Sources of Street Networks

A general distinction of free geo-data can be made between authoritative data and volunteered data. The first group of data is contributed by professional organizations, whereas the second group is provided by volunteers in a collaborative effort. This document will describe datasets from both groups.

TIGER/Line data (Topologically Integrated Geographic Encoding and Referencing system), from the US Census Bureau, is freely available for download from <http://www.census.gov/geo/www/tiger/>. TIGER/Line data include a wide collection of geographic feature types, such as roads, railroads, rivers, and lakes, as well as legal and statistical geographic areas, such as counties, or census blocks. The TIGER/Line data cover the entire United States. Road features come with various attributes, including address ranges, the geographic relationship to other features, road classification, geometry length, street name, and ZIP code. The TIGER/Line data are provided in shapefile format which has been developed by ESRI, Redlands (CA). Shapefiles are used for storing the geometric location and associated attribute information of geographic features. The Census Bureau releases updates of TIGER/Line data approximately once per year. On their website the user can select a county for downloading the TIGER/Line data shapefile as part of the “All Lines” data set. That shapefile contains an attribute called MTFCC (MAF/TIGER Feature Class Code) which allows extracting the requested line types in a GIS.

A new paradigm for the collection of geodata occurred in recent years in connection with the development of Web 2.0, which allows web users to actively participate in contributing content to the Internet. This can, for example be done in form of blogs, wikis, and RSS Feeds. RSS (Rich Site Summary) is a format for delivering regularly changing web content. Two of the first widely known Web 2.0 projects, Wikipedia ([www.wikipedia.org](http://www.wikipedia.org)) and Flickr ([www.flickr.com](http://www.flickr.com)) changed the way people use the Internet. The web community changed from passive consumers of web content to active participants. The development of mobile devices with GPS functionality allows the web community to interact with each other, provide information to central sites, and thus become a significant source of geographic information. Such voluntary shared spatial data has been coined “Volunteered Geographic Information” (Goodchild 2007).

The second freely available dataset we describe is based on a project called OpenStreetMap (OSM) ([openstreetmap.org](http://openstreetmap.org)), which is one of the prominent Web 2.0 applications that allows sharing geospatial data. OSM gives all Internet users the opportunity to download data without any fees and to use it (under certain licensing conditions) for their own projects. The goal of the OpenStreetMap community is to create a detailed map of the world with data collected by volunteers. OSM covers a wide range of object types that go beyond road data. Besides many road related features, which include road segments at different hierarchies, roundabouts, street lamps, or bus stops, it also maps amenities (e.g., restaurants, libraries, bicycle rental places), historic landmarks (e.g., archeological sites, castles), physical land features (e.g., beaches, cliffs, glaciers), or railbound features (e.g., tram, subway, and monorail tracks and stations), to name a few.

For every area of the United States, the OpenStreetMap project began by importing the TIGER/Line data into the OSM data base, which enabled volunteers to update, complement, and correct these street data within the OSM platform.

When downloading data from the OpenStreetMap web page, the user can define the area of interest through a bounding box. For this bounding box the spatial data, such as road features with their attributes, will be written into an OSM file which is based on the Extensible Markup Language (XML). XML is a common format for exchanging documents and data structures over the Internet. It might limit the data usage to users who are familiar with this format and who have some computer experience necessary to convert XML to shapefile using various tools, such as “osm2shp” (<http://code.google.com/p/osm2shp/>). Alternatively, data can be downloaded from company web pages, such as Geofabrik.de and <http://www.weogeo.com/>, which provide worldwide OpenStreetMap data in shapefile format among others. These websites offer the data preformatted and divided into hierarchical regions, i.e. (country and state). Because of the fast growth of the OSM data collection the downloadable files are updated at least once a week.

## Data Quality

The quality of spatial data sets is crucial for the success of a GIS project. Data quality has several components, including attribute accuracy, positional accuracy, logical consistency, completeness, and lineage. Where TIGER/Line data are administered by a regulatory instance, i.e. (the US Census Bureau), OpenStreetMap data are primarily contributed by

non-professional individuals with generally little experience or training. Therefore, quality checks are of particular importance for data created in collective mapping and data collection efforts. For OSM there are certain guidelines on how to collect, format, and upload data, but there is no single instance for quality control. Rather, it is expected that the web community checks on the correctness of the data, as is being done in comparable projects, such as Wikipedia. The TIGER/Line data are provided through the Census Bureau and therefore have more formal quality control procedures. However, the data are not as frequently updated as OpenStreetMap data and may therefore omit some recently constructed roads or local features.

Regarding the completeness, the coverage of OSM and TIGER/Line is similar because TIGER/Line has been imported into OSM. Differences in the completeness will occur in areas where volunteers contributed to the OSM project and uploaded additional road data. The extent to which the community participates in the OSM project varies between cities and countries. In general, voluntary contribution in Europe seems to be higher than in the United States perhaps because geospatial data layers are generally not freely provided through public agencies in Europe. This means that the creation of spatial data layers, such as roads, must be started from scratch, and contributors see a significant growth of the data layer through their personal contributions and initiative. This motivation may not be as high in the United States where selected base layers are already made available.

OpenStreetMap members in the United States focus on network segments that are usually not being covered by any public datasets, such as small alleys and pedestrian paths. As an example, for user-enhanced data in OSM, Figure 1 maps street data (in black) overlaid with pedestrian-only data (in red) for the area of San Francisco. These street data in OSM originate from TIGER/Line data, whereas pedestrian data come from voluntary data collection efforts. The community-added pedestrian paths are primarily concentrated around parks and some paths also provide shortcuts within residential areas. This makes the OpenStreetMap data generally useful for pedestrian-related routing applications, such as planning a route to the nearest metro station. However, it must be noted that OSM is an ongoing project and the amount of data collected by the web community in the United States is still relatively small. Additional collective data collection effort is necessary to make OSM data a sophisticated data source for this kind of application.

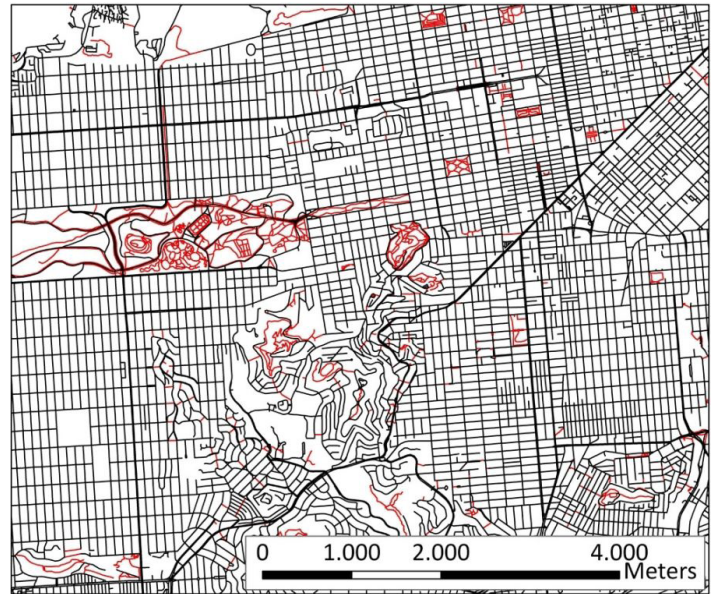


Figure 1. Visualization of pedestrian-only segments on top of streets for San Francisco  
Credits: OpenStreetMap

## Further Differences

Completeness is just one of many factors that need to be considered when choosing between OpenStreetMap and TIGER/Line datasets. Both sources offer a significant amount of information in addition to road geometry.

As opposed to OSM, the TIGER/Line dataset provides addresses and zip code information for most of its segments, which simplifies geocoding. Geocoding is the process of finding associated coordinates, such as latitude and longitude, from other geographic data, such as street addresses. TIGER/Line data provide a more detailed classification of recreational entities (e.g., National Park, State Park, Regional Park), and legal and statistical geographical areas (e.g., census blocks, block groups, tracts, counties, states, urban areas), compared to OSM.

OSM provides a variety of railbound features, such as tram, subway, monorail, and public transit stations, while TIGER/Line only includes railroads. Further, OSM includes surface and smoothness attributes with their road geometries. This facilitates the development of routing applications that consider the surface type, such as bicycle route planners for users of road bikes or mountain bikes. Additional information, such as turn restrictions and landmarks, can be useful for routing applications as well.



## Sample Applications

This section briefly describes two possible ways to use OSM and TIGER/Line data in GIS projects.



Figure 2a. OpenStreetMap coverage before the Haiti Earthquake 2010 (source: Where 2.0, 2010 Conference Presentation)

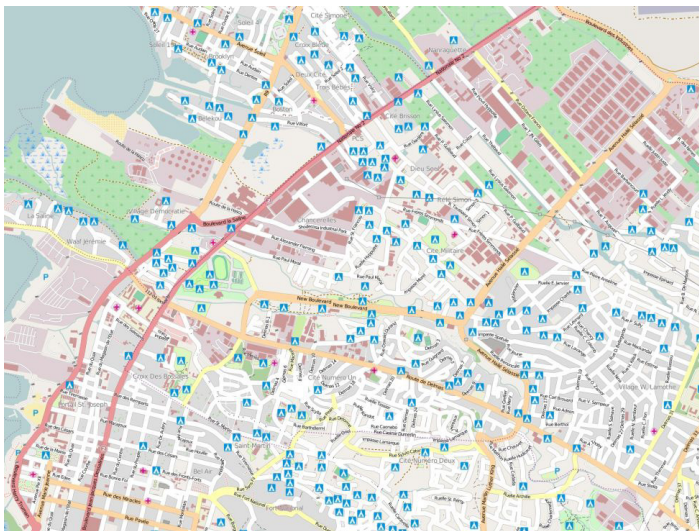


Figure 2b. OpenStreetMap coverage after the Haiti Earthquake 2010 (source: Where 2.0, 2010 Conference Presentation)

On January 12, 2010, a 7.0 earthquake struck Haiti. The OpenStreetMap community was able to help the response teams by building a reliable and accurate database of the functional road network and utilities in the affected area. Figure 2 demonstrates how quickly a base map can be built and improved with community-based data collection efforts. It shows two images in the area of Port-au-Prince, Haiti, before and after the earthquake. It is an impressive amount of data that volunteers contributed within days either from their computers at home or by collecting data in the field using GPS-enabled devices. The data include details about the current street network situation (e.g.,

impassable or blocked streets caused by debris or damage), water and sanitation infrastructure, health and medical facilities, and ad hoc settlements and refugee camps.

Freely available data generated by volunteers helped during this crisis. With these data, first aid forces were able to pinpoint the locations where help was urgently needed. The same type of emergency management could be used in hurricanes, floods, or forest fires where evacuation routes could be developed based on the latest data provided by volunteers.

Hurricanes present substantial challenges to city and county officials in terms of post-hurricane response to damage and removal of forest debris. In a research study, US Census Bureau TIGER/Line data were combined with FEMA Project Worksheet reports that itemize vegetation and construction debris amounts and costs of cleanup as well as hurricane damage related to hazard tree pruning and removal (Staudhammer et al 2009). With this method, researchers revealed that the amount of debris generated depended upon urban forest characteristics such as landscape-level tree cover, tree density, amount of tree cover in urbanized areas, and the amount of urbanized land. Spatial analysis indicated that debris results were clustered into northwest and southeast areas of Florida.

This case study is an example of a GIS project where freely available street data are of sufficient quality for the conducted spatial analysis task, and the purchase of commercial data is not necessary.

## Conclusions

In this document we gave an introduction to two freely available street datasets, including information on how to retrieve the data. We pointed out some of their differences. Two sample applications demonstrated the usefulness of free street data in various projects.

Different aspects need to be weighted against each other when evaluating the suitability of a free dataset for a given project. If the GIS project includes street data, a possible approach is to download both TIGER/Line and OpenStreetMap datasets for comparison and evaluation. A sample of a commercial data set or some aerial background images could be an additional valuable source for comparison and for spotting potential errors in the road datasets and assessing their quality.

Thus, the final recommendation with free data is not to rely on a single data source but to compare and deliberate on which source might be more useful for the specific project.

## Additional Resources

Goodchild, M. F. 2007. Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0 (Editorial). *International Journal of Spatial Data Infrastructures Research (IJSDIR)*, Vol. 2.: 24–32.

Flanagin, A. J., and M. J. Metzger. 2008. The credibility of volunteered geographic information. *GeoJournal*, 72:, 137–148.

Haklay, M. 2010. How good is Volunteered Geographical Information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B, Planning and Design*, Vol. 37, 4: 682–703.

O'Reilly, T. 2005. What is web 2.0: Design patterns and business models for the next generation of software. O'Reilly Media.

Staudhammer, C., F. Escobedo, C. Luley, and J. Bond. 2009. Patterns of urban tree debris from the 2004 and 2005 Florida hurricane season: A technical note. *Southern Journal of Applied Forestry* 33(4): 193–196.

Zielstra, D. & Hochmair, H. H. (2011). A Comparative Study of Pedestrian Accessibility to Transit Stations Using Free and Proprietary Network Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2217, 145-152.

## Internet

USCensus Bureau TIGER/Line Data  
<http://www.census.gov/geo/www/tiger/>

Weogeo—Provider of preformatted OpenStreetMap Data and OSM related Tools  
<http://www.weogeo.com/>

Geofabrik—Provider of preformatted OpenStreetMap Data and OSM related Tools  
<http://www.geofabrik.de/>

Official Website of the OpenStreetMap Project.  
<http://www.openstreetmap.org/>

Where 2.0, 2010 Conference Presentation: “Haiti: Crisis Mapping the Earthquake”  
<http://whereconf.com/where2010/public/schedule/detail/13201>

Functional classification of roads from the Federal Highway Administration (FHWA)  
<http://www.fhwa.dot.gov/environment/flex/ch03.htm>

MAF/TIGER Feature Class Code (MTFCC) definitions for TIGER/Line data from the US Census Bureau  
<https://www.census.gov/geo/maps-data/data/pdfs/tiger/tgrshp2009/TGRSHP09AF.pdf>