A REVIEW OF ROADS DATA DEVELOPMENT METHODOLOGIES

Taro Ubukawa[1], Alex de Sherbinin[2], Harlan Onsrud[3], Andy Nelson[4], Karen Payne[5], Olivier Cottray[6], and Mikel Maron[7]*

[1] *Geospatial Information Authority of Japan*
[*2] *Center for International Earth Science Information Network (CIESIN), Columbia University*
Email: adesherbinin@ciesin.columbia.edu
[3] *School of Computing and Information Science, University of Maine*
[4] *International Rice Research Institute*
[5] *Information Technology Outreach Services, University of Georgia*
[6] *Geneva International Centre for Humanitarian Demining*
[7] *Open Street Map Foundation*

ABSTRACT

There is a clear need for a public domain data set of road networks with high special accuracy and global coverage for a range of applications. The Global Roads Open Access Data Set (gROADS), version 1, is a first step in that direction. gROADS relies on data from a wide range of sources and was developed using a range of methods. Traditionally, map development was highly centralized and controlled by government agencies due to the high cost or required expertise and technology. In the past decade, however, high resolution satellite imagery and global positioning system (GPS) technologies have come into wide use, and there has been significant innovation in web services, such that a number of new methods to develop geospatial information have emerged, including automated and semi-automated road extraction from satellite/aerial imagery and crowdsourcing. In this paper we review the data sources, methods, and pros and cons of a range of road data development methods: heads-up digitizing, automated/semi-automated extraction from remote sensing imagery, GPS technology, crowdsourcing, and compiling existing data sets. We also consider the implications for each method in the production of open data.

1   INTRODUCTION

There is clear demand for an improved public domain global roads data set of inter-urban transportation linkages for a range of applications, but there remain significant shortcomings and gaps in the coverage of the most widely available public domain roads data sets (Nelson et al., 2006). We define public domain for the purposes of this article as those works whose intellectual property rights have expired, been forfeited, or are inapplicable, such as due to a waiver, license, or dedication by the owner or by legislative enactment.

A major goal of the CODATA Global Roads Data Development Task Group has been to identify appropriate options for filling the gap in the availability of public domain roads datasets and then to implement those solutions. The work of the group has resulted in a strategy paper (CIESIN, 2008), a catalog of roads data (CODATA & CIESIN, 2009), and the development of the Global Roads Open Access Data Set (gROADS), version 1 (CIESIN & ITOS, 2013).

Traditionally, "developing geospatial information" was virtually synonymous with "producing a map", a process that was highly centralized and controlled by national mapping agencies (NMAs) or private sector firms due to the high cost of required expertise and technology (GSDI, 2009). In the past decade, however, high resolution satellite imagery and global navigation satellite systems (GNSS) such as the US government maintained Global Positioning System (GPS) have come into wide use, and there has been significant innovation in web services such that a number of new methods to develop geospatial information have emerged, including automated and semi-automated road extraction from satellite/aerial imagery, crowdsourcing, and use of GPS enabled personal digital assistants (PDAs) (de Sherbinin et al., 2010).

Although researchers have sometimes evaluated their own methodologies, many algorithms are still in experimental stages, and there have been few comprehensive evaluations of roads data development methodologies to determine their suitability for filling gaps in available data at different scales and spatial extents. This paper catalogues some of the major approaches to develop roads data in order to have a systematic understanding of the existing methods. We also aim to evaluate the practical aspects of the methods and share

lessons learned from pilot efforts that would contribute to a plausible strategy for developing a global roads data set. An additional consideration, beyond feasibility, is to ensure that resulting data are available with no or few copyright restrictions and at low or no cost to end users. Thus we consider licensing issues in a separate section.

2 REVIEW OF DATA DEVELOPMENT METHODS

In this paper we review a number of automated and manual road data development methods. Each method is classified by the extraction technique applied, as shown in Table 1, and described in the following subsections. Crowdsourcing is considered as an additional approach although strictly speaking the methods used include GPS track collection and on-screen heads-up digitizing so it is not so much a method as a novel means of harnessing the labor of a large number of people using existing methods.

**Table 1.** Road data development methods

Methods	Specific technique / Existing initiative	Sources
Heads-up digitizing	- Tracing manually	- Aerial photo - Satellite imagery - Hard copy maps
Automated/semi-automated extraction	- Seeding and tracking - Snake algorithm - Segmentation and classification - Multi spectral analysis - Edge detection	- Aerial photo - Satellite imagery - Hard copy maps
GPS technology	- Passive data collection with GPS logger - Data collection with GPS/PDA	- GPS log
Compiling existing data set	- Datum/projection transferring - Generalization, Omission, Transfer	- GIS data - CAD data

2.1 Heads up digitizing

Heads-up digitizing (manual digitizing through visual observation) has been around since the 1970s, and it is still one of the major methods for data development though some elements of data processing such as georeferencing have become semi-automatic. In the past, large digitizing tables were used, and aerial photographs or paper maps were taped to their surface and digitized. Today, georeferenced and orthorectified aerial photographs or satellite imagery are shown on a computer display, and features are digitized into vector format by tracing a mouse over the imagery (Figure 1). Attributes are added at a later stage.

With the progress of GIS software, web services, and enhanced access to georeferenced imagery, this has become a widespread method for roads data development, effectively liberating this work from the hands of technocrats at NMAs. This is the basis for much of the road digitizing in OpenStreetMap (OSM), which used Yahoo! satellite imagery from 2007-2011 and now uses Bing aerial and satellite imagery. Apart from OSM, which has a special agreement with Microsoft/Bing Maps, the use of commercial satellite imagery in Bing and Google for the development of digital data such as roads or other features is restricted. This means that, from a legal standpoint, it is not possible to digitize data by using the satellite imagery as a backdrop for tracing roads or other features using GIS tools such as Quantum GIS and ArcGIS without prior written authorization from the imagery owner (see Clause 2 of Google Maps/Earth Additional Terms of Service). Yet, it is likely that many users do just that. To date, there are no well-publicized attempts to track and regulate users who do generate data using licensed imageries such as Google imagery and Bing imagery. The quality of the digitized data is generally affected by the positional accuracy and the spatial resolution of the imagery used as well as errors introduced during the digitizing process. For example, the absolute positional accuracy of ALOS/PRISM imagery, a panchromatic radiometer with 2.5m spatial resolution (1B2/Nadir), is 6.1 meter root-mean-squared error (RMSE) (JAXA website). This is suitable for updating topographic maps at a scale of 1:25,000. Positional accuracies of major imagery data sources as well as two commonly used road data sets are found in Table 2.

Many NMAs have developed their national base map at scales of 1:25,000 to 50,000 using aerial photographs and field surveys. According to the survey of GEO Task DA-06-05 (GSI & ISCGM, 2008), about 50% of developing countries surveyed (18 out of a total of 38) have developed a base map covering the whole country, and heads up digitizing methods are often used.
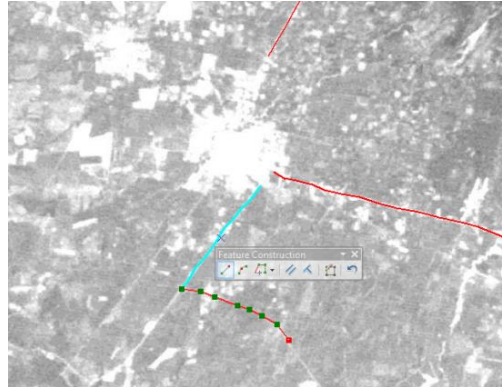
Figure 1. Example of heads-up digitizing (from satellite imagery)

Table 2. Positional accuracy by data set

Data set	# of Points (# of Scenes)	RMSE (meters)	Mean error (meters)	Std. Dev. (meters)	Range (meters)	Mean error vectors (x,y) (meters)	SD of error (x,y) (meters)
Google	140 (10)	**8.2**	7	4.2	(0.5-20.1)	(1.7,-1.9)	(4.6,6.3)
Bing	137 (10)	**7.9**	7	3.6	(1.6-22.1)	(2.2,-2.2)	(4.5,5.7)
OSM	116 (10)	**11.1**	8.8	6.9	(0.2-55.1)	(2.8,-2.2)	(8,7)
ESRI Roads	75 (10)	**121.3**	76.1	95.1	(3.1-594.7)	(14,-24.8)	(103.8,57.6)
VMAP0 Roads	54 (9)	**838.3**	652.8	530.8	(104.3-3699.6)	(-25,102.1)	(470.5,695.2)

Key: # of Points = total number of evaluated points; # of Scenes = the number of ALOS/PRISM scenes used (number of cities); Std. Dev. = standard deviation; Range = minimum to maximum spatial error.

Source: Ubukawa, 2013

2.1.1 Digitizing features from aerial photography

To be useful for feature extraction, first an aerial photograph needs to be orthorectified. Orthorectification is used to remove topographical effects prior to digitizing by geometrically correcting the data through projection onto an elevation model. In the past this required the identification of ground control points with well established locations and elevations. Recently, the combination of on-board GPS and an inertial measurement unit (IMU) used when capturing aerial photos enables easy georeferencing without the use of ground control points as the position and attitude of the plane can be measured directly by combining them (e.g., Nakamura et al., 2004).
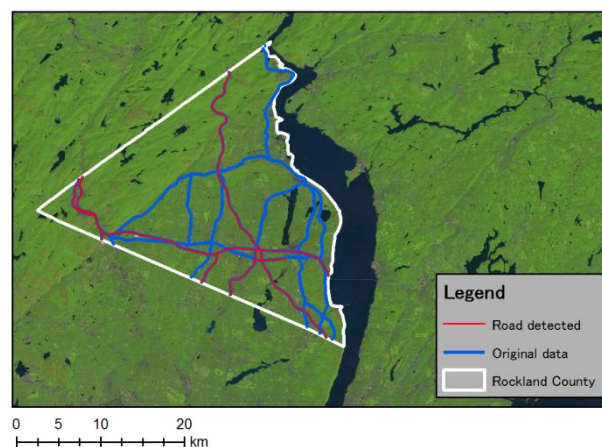
2.1.2 Digitizing features from satellite imagery

Heads up digitizing from satellite imagery is particularly common because the costs of high resolution imagery are generally lower than the costs of collecting aerial photographs. For example, a training seminar held in Kenya in 2012, supported and funded by the Japan International Cooperation Agency (JICA), aimed at developing maps using ALOS/PRISM in African countries (RCMRD, 2012). Other examples include 25,196 km of road mapping in the Amazon at a scale of 1:50,000 using Landsat TM/ETM data (Brandão & Souza, 2006) and basemap development in the Congo Basin where roads are relatively easily detected within the dense tree cover using moderate resolution Landsat 30m resolution data (de Sherbinin et al., 2010). The ability of a digitizer to recognize features during the digitizing process depends heavily on the resolution of the imagery and the size and construction material of the roads. For example, trial digitizing was executed with known roads in Rockland County, New York and Bergen County, New Jersey to evaluate efficiency and recognition of features with different resolution imagery.   The Global Map roads data set for the U.S. (USGS/ISCGM) at 1:1,000,000 scale, whose roads were generalized from much higher resolution imagery, was used as a reference (Table 2). With moderate resolution imagery, only wide roads with enough contrast with the surroundings can be detected (Figure 2). In addition, it is difficult to detect road attributes such as surface type or the number of the lanes with low resolution imagery, limiting its use to geometric feature extraction.

Table 3. Result of trial digitizing from imageries with different resolutions

Satellite	Resolution	Working scale at this practice	Efficiency	Recognition (comparing with Global Map USA)
OrbView-3/MS	4 meter	1:3,000 to 1:25,000	91 km / 40 minutes	Almost all
Landsat ETM+/ MS & Thermal	30 meter	1:20,000 to 1:100,000	187 km / 40 minutes	40 % (99 km out of 247 km in Rockland county. (Figure 2))

* Satellite imageries were downloaded from USGS.

Figure 2. Roads detected with Landsat ETM+ imagery (red) comparing with Global Map of USA (blue) (location: Rockland County, New York, USA)

2.1.3 Digitizing features from hard copy map

Digitizing road features from existing hardcopy paper maps can either be done directly from the paper map or from a scanned image of the map. The former is achieved with the use of a digitizing tablet, with the operator digitizing the features using a specialized type of mouse. This is considered an outmoded technique and rarely done in practice any more. Today it is more common to scan the paper map to create a digital image and then to rectify the image and digitize features from the scanned sheet. Scanning maps can create slightly distorted images, and for this reason, it is important to evaluate the scanned map by measuring the RMSE of the image using a set of well-known ground control points (GCPs). RMSE estimates are calculated after georeferencing the scanned image and attributing it with its projection and datum information. In the case where the map is based on a local datum or projection, the data should be transformed to a global datum/projection. Data

developers utilizing this approach need to be aware that sometimes features in paper maps are moved or generalized for cartographic purposes, resulting in the spatial displacement of features. Although there are few guidelines for this method, it is briefly mentioned in an ISCGM manual for the development of small scale GIS data sets (Manual for Development and Revision of Global Map, ISCGM, 2010).[1]

2.2   Automatic/semi-automatic extraction from imagery

There is much research on automatic/semi-automatic road extraction from aerial and satellite imagery. Mena (2003) proposed a classification of automated and semi-automated extraction methods according to three distinguishing characteristics: the objective, the extraction technique used, and the sensor or source data. Baltsavias (2004) reviewed the trends in object extraction focusing on extraction of important objects (e.g., buildings and roads) and pointed out the importance of object-oriented approaches or "knowledge-based analysis". Quackenbush (2004) also reviewed existing techniques and pointed out the lack of quantitative evaluations of results because many efforts have relied on visual assessment only. Mayer et al. (2006) gave an evaluation of several approaches through a EuroSDR (European Spatial Data Research) test comparing different approaches for automatic road extraction and indicated that they are useful for practical applications of map creation although there is some limitation on the complexity of applicable scenes. Automated extraction tends to be most useful in simple to moderately complex rural scenes. In the following subsections we provide a brief discussion of how characteristics of satellite image sensors influence the outcome of extraction techniques, followed by a synopsis of the most common automated extraction methods. As Quakenbush (2004) pointed out in his paper, it is worth noting many of the techniques require preliminary input from either a human operator or from existing data layers. The automatic/semi-automatic methods reviewed below focus on the extraction of geometric elements rather than on extraction of attributes.

2.2.1   Road extraction method and variety of satellite imagery

The appearance of roads in satellite imagery differs based on the spatial resolution, the sensor, and spectral band combination as well as the width and surface of the road and surrounding environment (Table 4). It is important to choose the right satellite imagery for each extraction method. For example, if the pixel Ground Instantaneous Field of View (GIFOV, also known as spatial resolution) in the imagery is larger than the width of the road, then the pixel values are composed of mixed land cover features, including the road surface (with its spectral properties) and other land cover classes (e.g., bare ground, buildings and their shadows, grassland, forest, etc.) with their own spectral properties. [2] Thus, methods such as spectral mixture analysis, which address spectral mixing, are required to extract the road line. On the other hand, if the pixel GIFOV is comparable to or smaller than the road width, a road may appear as bands composed of several pixel-widths, which may require other methods such as parallel edge detection.
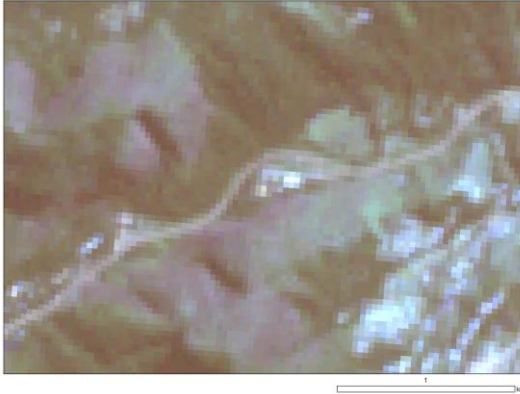
---

[1] It should be mentioned that digitizing from maps without permission raises intellectual property problems in that another map author's creative generalizations and interpretations may be copied and thus, in such an instance, copying occurs of more than just facts or standard representations that may be unprotected by copyright.

[2] A separate issue is the potential for roads to be obscured by trees or building shadows, which results in discontinuous line segments. Methods are available to connect line segments though they are not addressed here.

Table 4. Appearance of roads in imagery from satellite sensors with different resolutions

Resolution	Urban Area	Rural Area
High (1-4m)	**OrbView-3 (Panchromatic, 1m)** 	**OrbView-3/MS (3 bands composite, 4m)**
Moderate (15-20m)	**ASTER/VNIR (3 bands composite, 15m)** 	**CBERS-2/CCD (3 bands composite, 20m)**
Moderate -low (30m)	**Landsat/ETM+ (3 bands composite, 30m)** 	**Landsat/ETM+ (3 bands composite, 30m)**

* Satellite imageries were downloaded from INPE, GLCF, and USGS.
* The bars indicate 1 kilometer in the figures.

Less common imagery sources include Synthetic Aperture Radar (SAR). Edge detection in SAR has been around since the 1990s. According to Ngheim et al. (2001), "Roads typically appear as linear features in radar imagery. While automatic detection of roads from radar imagery has not been completely successful, semi-automatic, and manual road extraction is possible." The Sentinel-1 program from the European Space Agency will provide free C-band SAR with global coverage from 2014.

2.2.2   Specific methods / algorithm

There are a variety of methods for automatic/semi-automatic road extraction from remotely sensed images or scanned maps. The following are some typical methods classified according to the technique used. Some hybrid methods combine more than one technique as described below. Again, as noted above, the focus of these data extraction methods is on the geometry of the road line, and in most cases the resulting road line representations are given attribute information, such as surface composition and condition, manually after the geometry is created.

- **Seeding and tracking**

This method (e.g., Kim et al., 2004) starts by giving seed points that meet extraction criteria manually or automatically. Then, tracking is done by detecting new pixels that are similar to the pixels identified by the seed points in their spectral characteristics. The detected points are labeled as roads, and the line segments are extracted by repeating this process. This method can be done after some image-processing, such as edge detection or multispectral analysis.

- **Snakes**

The snake (active contour models) method, firstly proposed by Kass et al. (1988), is widely used in various fields for detecting boundaries of an object in an image. This method adjusts the line by minimizing a formulated energy function, which is a function that describes the shape of the spline that represents the road or road boundary, based on the digital values of the surrounding image pixels and the complexity of the line.   In other words, the curve is interpolated from known road features by adding points along a surface in such a way as to include all known and likely road features with the minimum amount of curvature change between points. Stated another way, the internal energy is imposing continuity and curvature constraints, and the image energy depends on the image intensity values along the path of the line. There are several improved snakes to extract road features such as the LSB-snake algorithm (Gruen & Li, 1996). This algorithm was employed in the CODATA global road data development project (see 4.1.).

- **Segmentation and classification**

This method is widely used in road extraction as well as land cover classification and can be found in some remote sensing software, such as ENVI and eCognition. Segmentation is the process of dividing an image into several groups of pixels based on the texture or characteristics of pixels. There are many segmentation methods, including spectral difference segmentation, multi-threshold segmentation, and contrast filter segmentation. Classification is the process of identifying the set of categories for each segment or pixels. There are also various methods for classification, including the maximum likelihood method, decision tree method, multi-level slice method, and clustering. Some methods require ground truth data for a semi- or fully supervised classification.

- **Multi spectral analysis**

The term "Multi spectral analysis" has a broad meaning; here we refer to it as an intermediate step within a road extraction methodology. This method analyzes/classifies each pixel or segments by analyzing multiple spectral bands in which each object has different spectral properties (i.e., different reflectivity across multiple bands of the electromagnetic spectrum). For example, Gomes et al. (2004) detected sub-pixel unpaved roads on Landsat images by modeling roads with a spectral response closer to that of bare soil than neighboring pixels, which required some parameters or spectral profiles to be specified manually.

This method can be used in conjunction with segmentation or it can be used on a cell by cell basis (without establishing spatial relationships among cells) to evaluate the degree of spectral mixing based on certain spectral mixture models (e.g., Small, 2003) and the likelihood of a cell belonging to a given class. Decision tree methods are often employed in this analysis. As with the segmentation methods outlined above, the result of this kind of analysis is a classified raster image that must be converted to vector data to create a road network dataset.

- **Edge detection (filtering)**

In this method, the edges of objects are detected using various filters. In principle, edges in an image can be detected by calculating the first-order derivative (gradient) or the second-order derivative (laplacian) of a pixel's brightness. Examples include the Roberts and Sobel filters (Roberts, 1963; Sobel, 1968 and 2014). As noise significantly affects edge detection with this method, a filter that employs a smoothing function was proposed (Canny, 1986) and has been widely used (e.g., Zhao et al., 2002). While the method with first-order derivative estimates image gradient and distinguishes edges, the method with second-order derivative can estimate local maxima in the gradient by extracting at zero-crossing pixels of the profile and detect parallel edges to identify the center of a bright (or dark) band. This method detects numerous false edges. For example, Hasegawa (2004) automatically extracted roads with edge detection method from ALOS/PRISM image and showed 80% of extracted lines were "false positives". For this reason, a land cover mask developed by multispectral analysis is often used to reduce the problem domain or imagery space to be classified.

2.3   Data development with GPS technologies

Many recent studies have examined the capabilities of vehicle-based mobile mapping systems with GPS technology, building on early efforts by the Center for Mapping at the Ohio State University and the Department of Geomatics Engineering at the University of Calgary (e.g., Goad, 1991; Bossler et al., 1991). Tao and Li (2007) reviewed the state of the art on mobile mapping, which means any mapping with the sensors mounted on a mobile platform including land based mobile mapping with GPS. The commercial mapping industry has provided a great deal of innovation in these systems, including, a new method to collect geospatial information with GPS using an information network system (INS) and Charge Coupled Device (CCD) camera. The system has been used for commercial services such as Google Street View. However, simpler systems devised for route tracing using both passive and active GPS equipment have also been utilized for the road data development, as described below.

- **Route tracing with passive data collection using a GPS logger**

With this method, large quantities of GPS tracks over each route are merged to produce a road segment. There is no need for the vehicle driver or user to interact with the device while in active operation. In an approach pioneered by Tracks4Africa (Tracks4Africa, *undated*), GPS tracks are collected from multiple recreational travelers and subsequently averaged to identify roads or frequently traveled tracks. The OpenStreetMap project has also employed this method though the methods for processing GPS logs are different. As GPS tracks do not record attribute information such as surface condition and route name, these attributes must be added manually although it is possible to infer road type or condition from average travel speeds, which can be derived from the log files. Current commercial GPS loggers can record road tracks at approximately a 3-10 meter horizontal accuracy. If there are high buildings, dense vegetation, or other obstacles that impede reception of the satellite signals, accuracy may be lower. In addition, acquired data need to be cleaned because of GPS "drift" when a car stops at an intersection (Figure 3). The interval of tracking (the time between recording position data) also affects the accuracy of the product (Figure 4).
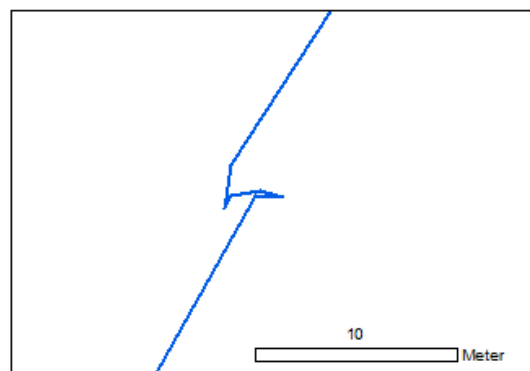
Figure 3. GPS errors while stopping at an intersection
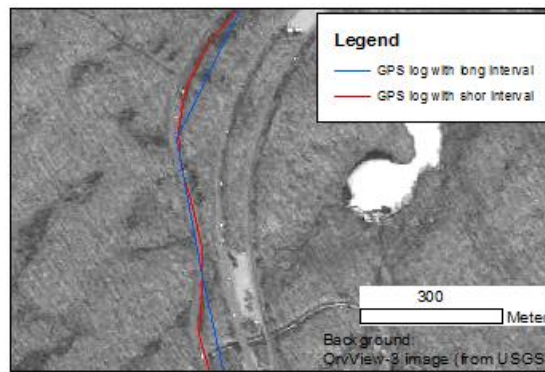
Figure 4. The same route with different interval

- **Active data collection with GPS enabled PDA**

With this method, road attribute information is recorded while GPS tracks are being acquired. IMMAP and CIESIN pilot tested this method in Ethiopia with World Food Programme (WFP) personnel on field missions (IMMAP & CIESIN, 2010). They used a customized version of the open source Cybertracker tool, which was modified to include attribute fields from the UN Spatial Data Infrastructure transport data model (UNSDI-T) and some customized fields useful to the WFP. A passenger selects standard menu options to describe road conditions or enters comments about road conditions as the vehicle travels along the route. The researchers found that while good quality road data with attributes can be acquired, the method does require trained and motivated personnel who pay attention to road conditions and features while traveling. Moreover special track cleaning algorithms needed to be developed in Python to merge data from multiple field sorties.

2.4 Compiling existing data sets

Creating regional or global road datasets requires compiling or conflating existing data. Examples of compilations of existing data include the Global Road Inventory Project (GRIP) of the Netherlands public planning bureau (PBL), the Nature Conservancy's and International Center for Tropical Agriculture's Latin America roads data set, and the Global Map assembled by the Information Technology Outreach Services (ITOS) of the University of Georgia in support of the UN Geographic Information Support Team (GIST). The Global Road Open Access Data Set (gROADS), which built upon the Global Map and included some data from GRIP, followed the following steps in its development. These steps illustrate typical procedures for the compilation of existing regional, national, or subnational data sets into a global or regional road map.

- **Data assessment and selection**

The first step in data compilation is to conduct an assessment of the coverage (data extent) and positional and topological accuracy of available data to determine if the data sets meet the minimum requirements or criteria for the final data product.   The threshold or required accuracies should be determined prior to the assessment and will depend on the scale and purpose of the compiled data set. The legal status of the source data sets and/or difficulty in acquiring permission to include the existing data in compilation may also preclude use of the data.

> **Coverage:** Although there is a method to evaluate completeness of the roads data set comparing the coverage of the data set with whole roads that exist, it is difficult to apply this method to the data sets at a smaller scale (or coarse resolution). This is because some features are omitted on purpose at smaller scales, and it is difficult to find a data set that contains all existing roads in an area of interest at a small scale. There are two primary methods to evaluate the completeness of a candidate road data set. The first is to examine the spatial extent of the candidate data to see if it contains data outside of the existing spatial domain. The second is to determine the amount of road information in the candidate data, as measured by the total kilometers of digitized roads. It is worth mentioning, however, that the total length of roads per area can vary depending on the degree of urbanization or the scale of the data set. Generally, for comparison purposes across candidate data sets, it is possible to ascertain the coverage of the road network for highways and primary and secondary roads by calculating the kilometers of roads by class by country.   Candidates with greater total length of the roads may be considered to have greater coverage, but this needs to be balanced against positional accuracy.

> **Positional Accuracy:** The root mean square error (RMSE) of a data set can be calculated against satellite imagery such as imagery available in Google Earth or Bing Maps to assess its positional

accuracy. To do this, gROADS assembled a global grid (grid cell size is either 0.5 or 1 degree depending on country) and picked ten random locations within countries near the grid vertices, usually road intersections. Distances were measured between the intersection found in the satellite imagery and the overlaid candidate data. If errors are normally distributed, the RMSE can be regarded as the standard deviation, and about 68% of candidate vector sets errors are within the distance of the RMSE.

➢
➢ **Topology:** Topology is an important metric of data quality. The following elements should be taken into consideration.

   o Overlapping roads (duplicate segments)
   o Gaps, undershoots, and unconnected short lines
   o Overshoots and dangles
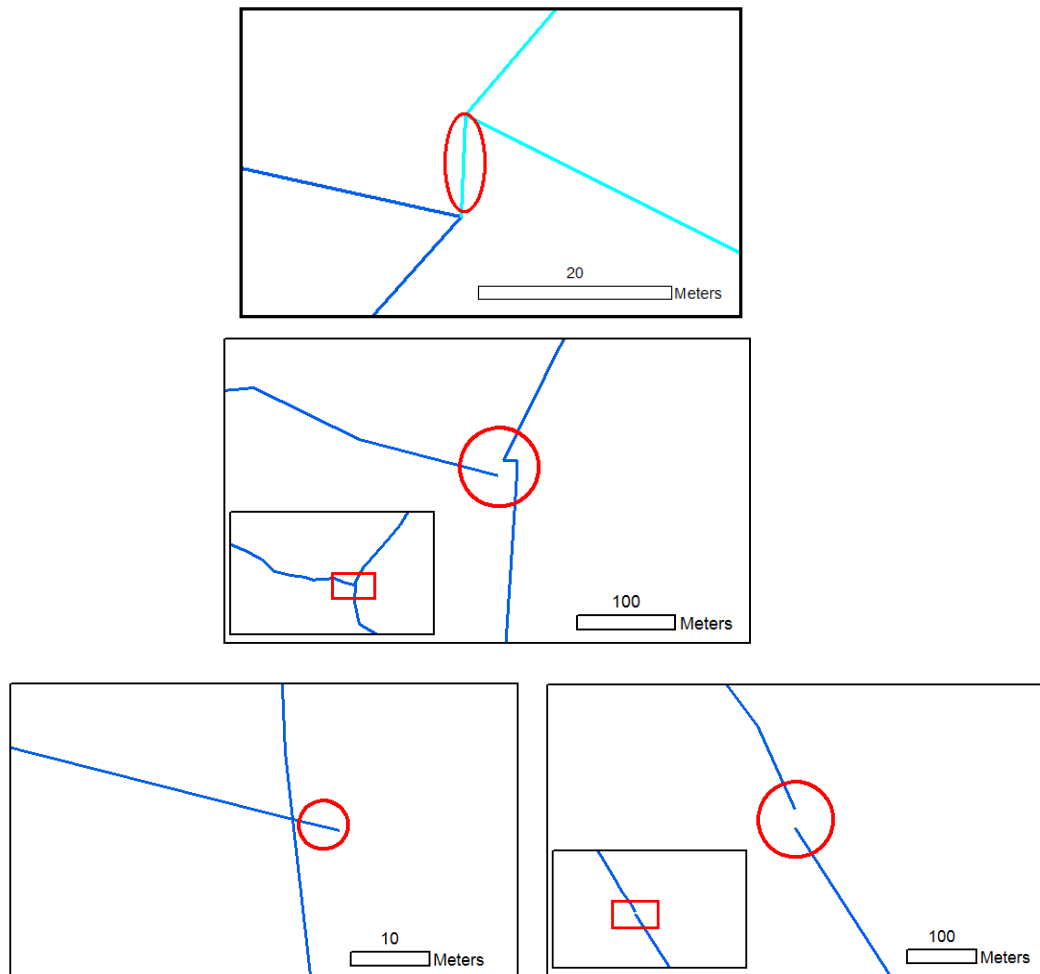
Figure 5. Examples of errors (Overlap, unconnected intersection, overshoot resulting in a short dangle and a gap)

• **Data editing**
   Merging data sets requires a series of data preparation and editing steps, including:

➢ **Projection/Datum transfer:** If the source has a coordinate system that is different from the target data set, it will need to be reprojected. Similarly, some countries use local datums so it is necessary to change the datum to the target datum (e.g., generally WGS 1984 for gROADS).

➢ **Editing topology:** In order to merge data sets, it is necessary to clean the topology by removing duplicate lines, connecting nodes, and removing dangles.

➢ **Attribute editing:** The candidate data must be transformed so that its attribute information conforms to a defined target schema of the newly developed data set (e.g., UNSDI-T for gROADS). This is often done using a custom ETL (Extract, Translate, Load) model. Once this is done, the new data set may then be merged into the target data set.

> ➢ **Compiled data topology edits:** After the candidate data has been appended, a second round of topological edits may be necessary to connect roads at borders to create a seamless coverage. This is particularly true if the compiled data are in two different scales.

2.5   Crowdsourcing through the Internet

Although crowdsourcing could be considered a workflow more than a method (because it actually employs multiple methods), we include it here for the sake of completeness and because it is an innovative use of emerging Internet-based tools. The generation of geospatial information through volunteer effort via the Internet has become increasing popular (Goodchild, 2007; Sui et al., 2012). Prominent examples of platforms for gathering and displaying crowdsourced road data include OSM and Google Map Maker. Of all of the road extraction methods described above, heads up digitizing and active and passive GPS traces are the most commonly utilized in crowdsourced road development projects. To our knowledge none of these projects pay crowd contributors financially for data collection or updates; most contributors simply gain satisfaction from producing data useful for themselves and others in their own community, for a larger cause, or for a sense of community with other amateur mappers. Although there is no "industry standard" or comprehensive review of workflows available, the OSM "new user's guide" describes the general approach and methods commonly used in these projects (OpenStreetMap, 2011). One of the advantages is the ability for real time data updating that enables communities to add detailed/updated information at any time through field observations, image interpretation, GPS tracks, etc. This has enabled the community to respond rapidly to natural disasters, such as the Haiti Earthquake in January 2010, when the Humanitarian OSM Team (HOT) was able to compile a complete street map of Port-au-Prince within days.

- **OpenStreetMap**

OSM is the most widely known crowdsourcing geospatial data development project. OSM is an online live service of a global data set not only of streets but also of a large array of user-defined features, including building footprints, parks, populated places, water bodies, pipelines, and hundreds of other relevant geographic entities built entirely by volunteers. Moreover, tags or attributes (such as a building designated as "hospital" or road names) are often translated into multiple languages. Anyone who has registered an account with OSM may contribute or edit data, and as of December, 2012, OSM reported over 900,000 registered users from all over the world although only a small percentage of these users actively contribute data. Several methods to edit the data set are provided through the website including GPS log submission, online heads-up editing, and a paper map (so called "walking paper") printing function for the field survey.

All edits to OSM are unmoderated though other users may flag issues concerning specific contributions. This is approached by monitoring tools, discussion, and quick action. The Data Working Group (DWG) is the backstop if issues are not solved in the community; only authorized DWG members can issue blocks, and this happens very infrequently. Technical safeguards, such as the ability to roll back data and not allowing users to commit large edits at once, also exist as deterrents to vandalism. A basic training module and a huge website (OSM wiki) provide information to ensure consistency of approach. There are also some tools for quality assurance, such as bug reporting tools, error detection tools, and so forth. However, some researchers have reported significant differences in the quality and the rate of mapping for different regions (e.g., Siebritz et al., 2012), a problem shared by most global mapping efforts. The data are distributed under the Open Database License, as discussed below in the section addressing licensing.

- **Google Map Maker**

Google Maps is one of the most popular commercial map services on the Internet and includes not only a basemap but also routing and geolocation services. Using Google Map Maker, users can add or update geospatial information to Google Map. Users can also edit third party user submissions by opting to moderate them. The updates by users are reviewed and appear on line if approved. This system is available for more than 190 countries and regions. Some Google Map data is owned exclusively by Google while other data is provided to Google through licenses by third party commercial geospatial data providers. All contributions provided by the crowd are legally owned by the contributors, but Google receives a perpetual, irrevocable, worldwide, royalty-free, and non-exclusive license to the contributions and restricts use of the resulting data to research use only.

Potere (2008) found that Google Map imagery has an absolute positional accuracy of approximately 39.7m RMSE, which is in the range of Landsat GeoCover (less than 50 meters RMSE) while Ubukawa (2013) found that it has an RMSE of 8.2m. Google makes no representation or warranties regarding the accuracy or completeness of any content or product (Google Maps/Earth Additional Terms of Service).

- **GISCorps**

GISCorps is a program of the Urban and Regional Information Systems Association (URISA), which coordinates short term and volunteer-based GIS services to support humanitarian relief, community development, local capacity building, health, and education. GISCorps has over 2,800 registered volunteers from 94 countries and to date has deployed 366 of these volunteers on 105 missions in 46 countries (http://www.giscorps.org/). When a disaster occurs, GISCorp calls for volunteers by sending them emails about the mission and puts them in touch with the agency that is hosting the project. GISCorps volunteers are considered an "Expert Crowd" since they are typically seasoned GIS professionals and as such they are able to follow any standards and protocols provided. If a requesting agency (RA) asks the volunteers to develop criteria for data collection and/or quality control (QC), they are able to implement those standards. Volunteers use a variety of applications programming interfaces (APIs) and software products. Use of the products is always dictated by the RA. GISCorps volunteers have been involved in several data collection and data quality control exercises in the past several years. Features are often digitized from the Google Maps API, and layers are created in KML format and then converted to other file formats. GISCorps has been involved with several other crowdsourcing missions since 2010, including HOT, USAID, and CrisisCommons in Japan and Alabama. For example, in the Shkodra region of Albania, GISCorps volunteers created a detailed road network along buildings and points of interest using the OSM database. Currently, GISCorps volunteers are creating primary data for the People's Democratic Republic of Korea, including data depicting roads, railroads, bridges, airports, rivers, and villages.

3   LESSONS LEARNED FROM PILOT EFFORTS

In this section we describe the accomplishments of two pilot projects in exploring potential alternatives to heads up digitizing for tracking roads and designating road features. They were accomplished in the country of Ethiopia, which has a need for improved road data.

3.1   Road development in Ethiopia with semi-automatic extraction algorithms from high resolution remote sensing imagery (NASA-SERVIR project)

This work was conducted by the Center for International Earth Science Information Network (CIESIN) with the cooperation of the Center for Spatial Information Science (CSIS) of the University of Tokyo and the Japanese National Institute of Advanced Industrial Science and Technology (AIST) as part of the NASA-funded Expansion of Regional Visualization and Monitoring System (SERVIR) project (de Sherbinin, 2010). The road data was developed from ASTER imagery with approximately 15m spatial resolution using a semi-automatic road extraction tool named "the Global Road Mapping Tool" developed by CSIS.

Targeted area: The central plateau area of Ethiopia with a bounding box of 12° 44' 11" N and 36° 31' 00" E (upper left) and 6° 19'21"N and 41° 34' 00"E.
Tool: The Global Road Mapping Tool (GRMT), developed by the Center for Spatial Information Science (CSIS) of the University of Tokyo, was used to extract roads from ASTER imagery. The GMRT is a stand-alone software package using a "snake" algorithm (see section 2.2.2) to connect points that are seeded along a road. Given sufficient spectral contrast between the road and surrounding land covers, it can connect a sparse array of points by following the road edges. Initial road seeds are generated by manually clicking over the imagery.
Imagery used: The roads were extracted from orthorectified ASTER imagery contributed by AIST. CIESIN worked with 27 ASTER scenes, each scene representing an area of 60 km by 60 km on the ground.
Process: The roads were extracted from ASTER imagery with the tool and functionally classified into several classes such as highway, primary, trail, and so forth (FCLASS attribute of UNSDI-T model). Not only the ASTER imagery but also other information was referred to when a road was classified. Then, acquired data was merged into the existing WFP data set for Ethiopia.
Result: From 27 scenes of Aster imagery, 5,148 unique road segments were digitized and 111 of these segments were selected and added to the existing data set.

Lessons learned:

- This method of using moderate resolution ASTER imagery held some promise, insofar as the footprint is quite a bit larger than that of most high resolution imagery and is at a resolution that makes it much easier to pick out roads than using Landsat imagery.
- It was possible to extract road segments from ASTER imagery through visual inspection.
- As initial road seeds need to be specified manually and need to be very closely spaced for the algorithm to track the road, the method is similar to heads-up digitizing and does not dramatically reduce the workload.
- This result may have been in part due to a lack of spectral contrast between the roads and the surrounding areas. Even paved roads in Ethiopia can be dust-covered and therefore spectrally indistinguishable from surrounding areas even though the human eye can detect the difference. Thus the method may achieve greater efficiency when used with other landscapes.
- It was difficult to distinguish the existence or function class of a road by the ASTER imagery only.

3.2   Road data development with GPS enabled PDAs (AGCommons Project: Ethiopia)

This project was funded by the Gates Foundation's AGCommons (Agricultural Geospatial Commons) initiative and implemented by iMMAP with the collaboration of CIESIN and the Regional Center for Mapping of Resources for Development (RCMRD). Through the project, roads data for a portion of Ethiopia were developed using a customized Global Roads Open Access Data Set (gROADS) PDA-based GPS data collecting tool (iMMAP, 2010).

Target area: Ethiopian provinces of Afar, Gambella, Oromiya, SNNP, Somali, and Tigray.
Tool: The gROADS PDA-based data collecting tool was developed based on CyberTracker, a free software developed for ecological field surveys but modified for the collection of roads data using the UNSDI-T data model. Figure 6 shows the interface, with icons representing different road types and features of interest.
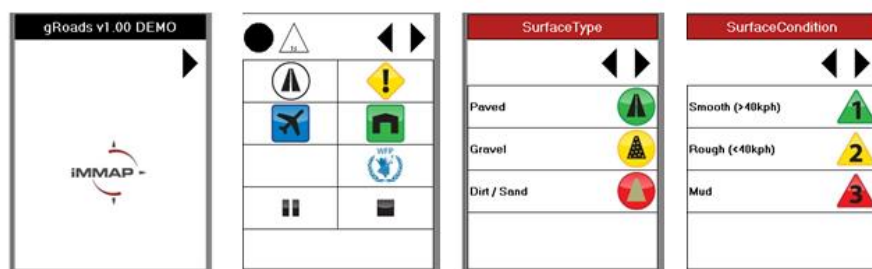
Figure 6. Screenshots from the customized Cybertracker tool

Process: One day workshops were conducted to train the World Food Program (WFP) field teams in the use of the PDAs. These teams then used the PDAs to collect roads data while on field sorties for various missions to remote parts of southern Ethiopia. The teams collected data in the field from June to November 2009. Finally, collected data were compiled and cleaned.
Result: The project produced about 5,200 km of improved roads data for Ethiopia.
Lessons learned: In principle the method held promise, in the sense that the added labor required to do the roads data collection was minimal. Yet a number of obstacles were encountered:

- Significant training time was required for data collectors to learn how to use the PDAs, and the time required increased with group size. Some collectors never developed full proficiency.
- Collectors recorded attributes inconsistently and, despite the importance of recording surface type, often forgot to record when the type changed.
- Collectors needed an incentive or a very clear mandate from superiors to pay close attention to the data collection.
- The area covered by field sorties was limited to areas of interest to WFP, largely between Addis Ababa and the Somali border. An original idea of training truck drivers to do the collection proved impractical, partly for the reason of incentives listed above.
- Practically speaking, units with integrated car chargers and window mounts were found to be easier to use.

4 COPYRIGHT AND OTHER INTELLECTUAL PROPERTY IN SPATIAL DATA

In attempting to construct a regional or global roads data set legally suitable for general purpose use upon which many individuals, agencies, and private companies might extend from or build upon to meet their own needs, it must be kept in mind that the source data often have legal restrictions regarding their use. These existing legal rights must be ascertained and accommodated.

In most jurisdictions across the globe, copyright subsists in original works of authorship upon creation in a tangible form whether or not the author desires it (Berne Convention). The practical result of this rule of law is that users should assume that for all datasets and other data sources they gain access to on the Internet or elsewhere, one or more other parties probably have an ownership right in the work. Just because one finds a video or music file openly available on the Internet does not mean it is legally permissible to take all or a portion of that file and incorporate it into a new file that you would like to use or distribute. The same holds true for most geographic datasets. Not getting caught, challenged, or sued is not the same as having a clear legal right to create derivative products from the works of others.

Whether a person or team is able to copy a geographic data set or digitize information from an existing map and incorporate the data into another product without asking permission depends on answers to a string of legal questions. As a general rule, however, compilers should be cautious and assume that some originality aspects of those other works will be copied. If so and even though counter to much common practice, the law assumes that one must have permission to include all or some of the source work in the derivative product (Onsrud, 2010).

One might argue that roadway tracks or feature descriptions drawn from another's database or dataset qualify as *facts* and are therefore not protected by copyright under most national and international copyright laws and conventions. However, even if individual data elements are not protected in a specific instance, the original or creative selection, arrangement, and coordination of facts is still protected in many jurisdictions (Uruguay Round Agreement). Further, some jurisdictions protect such concepts as "sweat of the brow" and "industriousness" under their copyright laws even without a showing of originality or creativity in the material copied. Finally, many jurisdictions protect datasets and databases far beyond the protections of copyright with additional intellectual property protections through database protection legislation and such concepts as moral rights and catalogue rules.

The violation of copyright, database legislation, and other intellectual property protections, whether knowingly or not, can result in severe penalties. Potential violation of these rights imposes by far the greatest liability exposure for those parties attempting to compile geographic data from a variety of sources (Onsrud, 2009). Intellectual property laws exist and are being strengthened continually in most nations across the globe against those who would compile enhanced digital data resources without permission, even if the goal is for promoting the general public welfare. This is the current reality.

As a result, the compiler of a global road dataset with the intention of making it widely accessible and usable by others for expansive purposes typically needs to pursue one or a combination of the following approaches:

1. Create a web-based global platform that facilitates contributions by the crowd with contributors, whether novice or expert, agreeing to adhere to a specified open access license or dedication to the public domain for all of their contributions;
2. Convince previous road data set creators to make their data available to the world on the web using metadata to declare that the work is available under a standardized open access license or public domain dedication so that others are free to use the data; or
3. Seek agreements with previous road data set creators to allow use of their work in the new compilation under the open access license or public domain dedication conditions chosen for the new road data set compilation. While the agreement clearly needs to cover copying of already digital data, it may also need to cover the digitizing off of maps of generalized geographic parameters and feature classifications.

Under the first approach, Open Street Map has been highly successful in defining the open access rights to which its contributors and users are expected to adhere. Under its published legal constraints, it is conceivable that a subset of the current roads in OSM could be chosen as a base upon which to expand to create a global roads network if the current data met the technical requirements sought and the added contributions of the

gROADS data set effort were fed back into Open Street Map. However, OSM data is not in the public domain if one defines public domain as including those works whose intellectual property rights have expired, been forfeited, or are inapplicable. OSM uses a restrictive open access license. That project now uses the Open Database License (ODbL) to restrict the public use of any adapted version of the database and works produced from the database. It further uses the Creative Commons Share-Alike 2.0 (CC BY-SA) license to restrict the use and creation of derivative works from significant cartographic content of OSM. Both of these licenses contain share-alike provisions meaning that any derivative works must carry and announce that those works are also available only under the terms of the ODbL. Thus imposing fewer or greater restrictions when attempting to combine data from multiple datasets for a specific geographic area is not allowed unless all other data owners, whether government owners or otherwise, agree to allow their contributions to be controlled by the ODbL. This has been problematic for many public agencies. Further, crowdsource contributors transfer to OSM all rights, but such content rights are not necessarily transferred to those creating derivative works.

The Google Maps platform and its Map Maker tool is an additional example of a global platform for crowdsourcing of location and feature data. Both Open Street Map and Google Maps point out the utility of using a global platform to call on thousands of individuals to contribute to a software and data repository platform that is centrally controlled. One great advantage of the approach is that legal clarity is provided because conditions of contribution and use may be clearly established and made consistent for all of the data. This approach requires building a software and data management infrastructure in order to gain the legal rights clarity desired.

Under the second approach, some of the most prevalently used or recommended standard open access licenses and similar instruments used by geographic data and location based information contributors in the legal status fields of their metadata include: (a) the Creative Commons instrument of CC0 (http://creativecommons.org/publicdomain/zero/1.0/), which allows creators to opt out of copyright and database protection to place their works in the public domain by waiving any rights to the extent possible and providing a public license fallback for those rights that might not be able to be waived in some jurisdictions, (b) the Open Data Commons Public Domain Dedication and License (PDDL) (http://www.opendatacommons.org/licenses/pddl/1-0), which accomplishes similar objectives to that of CC0, (c) the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/legalcode), requiring users to provide author attribution, (d) the Open Data Commons Attribution License (http://opendatacommons.org/licenses/by/), accomplishing similar objectives to that of the Creative Commons Attribution License and (e) the previously mentioned Open Data Commons Open Database License (ODbL) (http://opendatacommons.org/licenses/odbl/), which requires attribution and share-alike for databases. In addition, government agencies might use an acknowledgement of Public Domain status in their posting of metadata (e.g., Public Domain Mark of Creative Commons at http://creativecommons.org/choose/mark/).

The primary problem with the second approach is that many organizations and individuals fail to post metadata about their road and other geographic datasets and most that do tend to not complete the appropriate metadata fields for declaring the legal status of the dataset. Finally, many that describe the legal status in the metadata tend to not use a standard widely recognized open access license or instrument. A further challenge is that information in metadata fields is not readily found through global web searches. Finding such information typically requires a seeker to search each and every distributed repository node to find out if it is even possible to discover online the legal status of datasets of interest. The GEOSS Common Infrastructure (http://www.earthobservations.org/gci_gci.shtml) is an example of an effort attempting to lessen the need to search thousands of individual repository nodes. However, the lack of reported metadata on the legal status of geographic data sets remains a substantial barrier to knowing whether found data that meets technical and content standards can be used without seeking further permissions.

The third approach of seeking permission to use the works of potentially thousands of others in creating a compilation such as a global roads data set is daunting. However, it remains the default approach unless the compiling organization is willing to expose the organization and its employees to very substantial liability exposure. This is the primary approach that was used to begin creating the Global Road Open Access Data Set (gROADS), which with a few clearly documented exceptions (mostly requiring attribution to the original data provider for individual country data sets) is fully open access. Such explicit permissions need to be acquired not only from private sector firms and non-profit organizations but also from most national and local governments as well as international agencies that have created useful road data holdings.

For a few data sets that were considered of possible interest to the project in meeting content and technical requirements, Table 5 summarizes the published restrictions imposed. The generally published terms of use may of course be overcome if owners are willing to enter into an agreement to alter their generally applicable terms. However, most large entities are unwilling to do so for limited projects such as creating a global roads data set for general-purpose use by all. It should be noted that gROADS is defined as a Collective Database in order to allow differing license approaches to be applied for the data from some nations. No consistent set of open access legal language applies to the entire collection.

Table 5. Restrictions on data use

Data set	Type	Licensing document	Permitted Use			
			Copy	Modify/edit	Make derivative works	Redistribute
VMAP0	GIS data (Public domain)	None	Allowed	Allowed	Allowed	Allowed
Google Map (Google)	Online map (free for usage below 25,000 page views per day)	Terms of Use	Not allowed*	Not allowed*	Not allowed*	Not allowed*
			(*) Without a prior written authorization			
Bing Maps (MSN)	Online map Free of charge	Terms of Use	Not allowed*	Not allowed*	Not allowed**	Not allowed*
			(*) Without a prior written authorization (**) OpenStreetMap is allowed to use Bing aerial imagery for tracing			
OpenStreetMap	Online map Free of charge	Open Database License	Allowed	Allowed	Allowed***	Allowed
			(***) users must provide attribution and "share alike"			
OS OpenData (Ordnance Survey)	GIS data Free of charge	OS OpenData License	Allowed	Allowed	Allowed	Allowed
			With the condition "show the attribute (data source)"			
Landsat	Satellite image (By NASA)	Landsat Data Distribution Policy	Allowed	Allowed	Allowed	Allowed
			With the condition "show the attribute (data source)"			
IKONOS	High resolution satellite image (Commercial)	License agreement	Allowed (Internal use in single organization)	Allowed (Internal use in single organization)	Allowed (Unless derivative product preserves original imagery)	Allowed (non-commercial; e.g. scientific paper)

Table 5 does not list all of the restrictions imposed on most of the products. Further, the terms of use of any of the products may change at any time so it is important that users incorporating the works into derivative products document the existence of the terms at the time of creation of the derivative product or use.

5   EVALUATION AND DISCUSSION

## 5.1   Comparison table and keys for consideration

The best method to develop a good quality data set depends on the purpose for which the data will be used and the desired scale, which itself relates to purpose. Some aspects of data quality include completeness, positional accuracy, thematic accuracy, and temporal accuracy. These quality characteristics depend heavily on the scale or resolution of sources from which the data is derived. Each data development approach may have different impacts on data quality.

5.1.1   Positional Accuracy

Positional accuracy is perhaps the most important element of data quality. There are many ways to regulate horizontal and vertical spatial accuracy of geospatial information. For example, for Japanese topographic maps at a scale of 1:25,000, features are required to have a RMSE within 0.7mm on the map sheet. In the United States, for topographic maps of the scale less than 1:20,000, more than 90 percent of well-defined points should be within 1/50th of an inch on the sheet. USGS defines "well-defined points" as those that are easily visible on the ground. Errors include the errors caused during the digitizing process, whether onscreen or using a digitizing table, as well as the errors in source materials. Given that all errors caused during the digitalizing process are small compared with the errors that their sources originally have, the accuracy of original data would give certain limitation on the scale of the data developed. For example, given that some GPS logged data have positional errors in 10^1 order (unit: meter), the scale of developed data should be at a scale of 1:10,000 (10^{-4} order) or smaller if the accuracy of 1 mm (10^{-3}) RMSE on the map sheet would be achieved. It is worth noting that GPS or satellite based methods work best at the scale of 1:1,000 to 1:10,000. There are a lot of methods proposed for data development at large scale while a few methods are available at a small scale.

5.1.2   Coverage

The spatial coverage of the data (area mapped) has a close relationship with the scale of the resulting maps and the source data used. Although it is easy to recognize features and maintain good positional accuracy with higher resolution imagery, the footprint of this imagery is much smaller than that of the imagery with lower resolution, which means that more work is required to tile imagery together and digitize visible road segments. Therefore, it seems that there is a tradeoff between coverage and scale in existing data sets (Figure 7). It is important to choose proper resolution of the source and proper scale in light of the desired spatial coverage. Traditionally, data with global coverage, such as VMAP0, have been at a small scale (low spatial resolution). In the survey of VMAP0 data conducted by gROADS in assembling data for the final global map, the RMSEs ranged from 530m in Burundi to 1,265m in Sudan, meaning on average road locations could be anywhere from ½ km to 1.2km from where the road is actually positioned on the map. While these spatial errors are unacceptable for navigation purposes, they may be fit for purpose for rather coarse scale global or regional modeling.
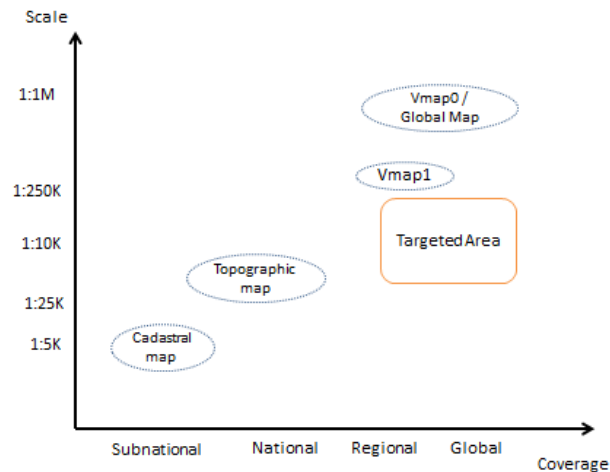
Figure 7. Conceptual figure on the scale and coverage

5.1.3 Completeness

Completeness is also important element of the data set, which represents the absence and presence of features. Many research projects on auto/semi-auto road extraction from satellite imagery evaluate their methods by estimating their completeness (both errors of commission and omission). It is regarded as a key of data quality especially on a large scale map. On the other hand, for a map of inter-settlement transportation such as gROADS, inclusion of streets or dirt tracks in rural areas may provide more detail than is needed for most requirements (see Nelson et al., 2006). Satellite imagery shows states of the ground as they are, and selection of acquired features are up to an algorithm or the data developer while only roads that are driven are recorded with a GPS logger.

5.1.4 Thematic Accuracy

This is also an important evaluation from the aspects of the accuracy of thematic classification, such as condition of the road surface, intended route use, and route name. UNSDI-T has developed a data model for transportation, which includes rather detailed thematic information, and particularly information of relevance to humanitarian or relief operations. Data development methods from satellite imagery or scanned map sheets are widely used to extract linear features, but there are few studies attempting to automatically capture thematic information along with the extracted line work. To improve thematic accuracy, additional information from ground surveys or ancillary maps is required.

5.1.5 Temporal quality and freshness

This represents accuracy of a time measurement, temporal validity, and freshness of the data. Satellite imagery, GPS logs, and scanned maps clearly show the existence of a road at the time when data were acquired. Commercial firms have developed comprehensive workflows that involve amassing information on new developments through media reports and automated comparisons between imagery of different dates to identify locations of new road development.

5.2 Final considerations

Reviewing several methods of road data development, many studies have worked with high to medium resolution satellite imagery, which can increase efficiency in data development over relatively small areas.

- **Directly developing data at a large scale in a globally consistent manner, making best use of automated methods**
  Crowdsourcing platforms are seeking to develop a large scale (highly detailed) spatial database with global coverage while there are still density gaps in the current maps. The Google Maps Live Traffic

Works is a service that shows the state of current road congestion by tracing and analyzing movement of undefined individuals who use the Google Map service in a GPS-enabled mobile phone. This technology could be applied to road data development.

- **Directly developing/updating data at a medium to small scale making best use of semi-auto/auto methods**
  In the field of the land cover classification, many researchers focus on data development with medium to low resolution satellite imagery to achieve global coverage (e.g., Globecover; GLCNMO), where recognition of linear segments is not required. In order to maintain a small scale global roads data set, it would be worth trying to make use of medium to low resolution satellite image as well as existing geospatial information.

- **Introducing a global standard for geospatial information**
  Many national mapping agencies have their own map elements based on their own drawing rules following their laws. If there was a global standard or data profile for certain features, it would be easy to merge several data sets developed by different agencies. Several stakeholders adopted this strategy. ISO/TC 211 already defined the format for geospatial information (ISO19136). UN-JLC defined the UNSDI-T model for transportation features. ISCGM defines Global Map Specifications at 1:1,000,000 scale map elements sharing this with national mapping agencies.

- **Compiling existing data sets transferred into consistent specifications**
  This is one of the most practical ways to develop global data sets. By assembling several data sets, there are inevitably differences in road density and some differences in the feature-definition. There are limits to the improvements possible for small scale data such as VMAP0, and the strategy taken by gROADS is generally to wait for improved data from national mapping agencies or other third-party sources with improved spatial accuracy as this quality is deemed more important than others for most use cases.

In conclusion, we can say that there is no single best way to develop global scale roads data for inter-urban (or inter-settlement) transportation links, but it is important to be aware of the strengths and limitations of each approach with regards to different data quality criteria.

6   ACKNOWLEDGMENTS

7   DISCLAIMER

The view expressed in this report are those of the authors and do not necessarily represent the views of their respective organizations.

8   REFERENCES

Baltsavias, E. P. (2004) Object extraction and revision by image analysis using existing geospatial data and knowledge: current status and steps towards operational systems. *ISPRS Journal of Photogrammetry & Remote Sensing 58*, 129-151.

Berne Convention for the Protection of Literary and Artistic Works (September 9, 1886 and subsequent revisions as amended). Retrieved from the World Wide Web, March 23, 2014:
http://www.wipo.int/treaties/en/ip/berne/trtdocs_wo001.html
Bossler, J., Goad, C., Johnson, P., & Novak, K. (1991) GPS and GIS map the nation's highways. *GeoInfo System Magazine, March*, 26-37.

Brandão Jr., A. O. & Souza Jr., C. M. (2006) Mapping unofficial roads with Landsat images: a new tool to improve the monitoring of the Brazilian Amazon rainforest. *International Journal of Remote Sensing 27*(1), 177-189.

Canny, J. (1986) A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8* (6).

CIESIN (Center for International Earth Science Information Network, Columbia University) (2008) A Strategy for Developing an Improved Global Roads Data Set. Developed by Participants at The Global Roads Workshop 1-3 October 2007, Lamont Campus, Columbia University Palisades, New York, USA.

CIESIN (Center for International Earth Science Information Network, Columbia University) & ITOS (Information Technology Outreach Services, University of Georgia) (2013) Global Roads Open Access Data Set, Version 1 (gROADSv1), Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Retrieved from the World Wide Web, March 23, 2014: http://sedac.ciesin.columbia.edu/data/set/groads-global-roads-open-access-v1.

CODATA (Committee on Data for Science and Technology) Global Roads Data Development Working Group &CIESIN (Center for International Earth Science Information Network, Columbia University) (2009) CODATA Catalog of Roads Data Sets, Version 1, Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Retrieved from the World Wide Web, March 23, 2014: http://sedac.ciesin.columbia.edu/data/set/groads-codata-roads-catalog-v1.

de Sherbinin, A., Yetman, G., & Steil, M., (2010) The Global Roads Open Access Data Set (gROADS): Pilot Efforts to Develop Improved Roads Data. *Presentation to the 12th Annual Meeting of the Global Spatial Data Infrastructure Association (GSDI-12).*

de Sherbinin, A. (2010) Roads Data Development in East Africa: Final Report, US National Aeronautics and Space Administration (NASA) SERVIR project.

Goad, C. (1991) The Ohio State University Highway Mapping System: The Positioning component. *Proceedings of the Institute of Navigation Conference, Williamsburg, VA*, 117-120.

Gomes, O. F. M., Feitosa, R. Q., & Coutinho, H. L. C. (2004) Sub-pixel unpaved roads detection in Landsat images. *Proceedings of the IRPRS 35*, 1196-1200.

Goodchild, M. (2007) Citizens as sensors: the world of volunteered geography. *GeoJournal* 69(4), 211-221.

Gruen, A. & Li, H. (1996) Linear feature extraction with LBS-Snakes from multiple images. *International Archives of Photogrammetry and Remote Sensing 31*, 266-272.

GSI & ISCGM (2008) Guideline for Basic Geographic Data. Retrieved from the World Wide Web, March 23, 2014: http://www.iscgm.org/cgi-bin/fswiki/wiki.cgi?action=ATTACH&page=Application%2FGEO%2FDA-06-05&file=Guideline+for+Basic+Geographic+Data.pdf

GSDI (2009) Spatial Data Infrastructure Cookbook 2009. Retrieved from the World Wide Web, July 5, 2012: http://www.gsdi.org/gsdicookbookindex

Hasegawa, H. (2004) A semi automatic road extraction method for ALOS satellite imagery. *Proceedings of the ISPRS Congress 35*, 402-407.

iMMAP & CIESIN (Center for International Earth Science Information Network) (2010). AGCommons Roads Data Development in Ethiopia Final Report. Retrieved from the World Wide Web, March 23, 2014: http://www.groads.org.

ISCGM (International Steering Committee for Global Map) (2010) Manual for Development and Revision of Global Map.

Kass, M., Witkin, A., & Terzopoulos, D. (1988) Snakes: *Active contour models, International Journal of Computer Vision 1*, 321-331.

Kim, T., Park, S., Kim, M., Jeong, S., & Kim, K. (2004) Tracking road centerlines from high resolution remote sensing images by least squares correlation matching, *Photogrammetric Engineering & Remote Sensing 70* (12), 1417-1422.

Mayer, H., Hinz, S., Bacher, U., & Baltsavias, E. (2006) A test of automatic road extraction approaches. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science 36*, 209-214.

Mayer, H. (2008) Object extraction in photogrammetric computer vision. *ISPRS Journal of Photogrammetry & Remote Sensing 63*, 213-222.

Mena, J. B. (2003) State of the art on automatic road extraction for GIS update: a novel classification. *Pattern Recognition Letters 24*, 3037-3058.

Nakamura, T., Shimono, T., Oki, S., & Suzaki, T. (2004) Research on Application of GPS/IMU to Topographical Mapping. *Report of GSI (in Japanese) 105*, 17-22.

Nelson, A., de Sherbinin, A., and Pozzi, F. (2006) Towards development of a high quality public domain global roads database. *Data Science Journal 5*, 223-265.

Ngheim, S., Balk, D., Small, C., Deichman, U., Wannebo, A., et al. (2001) Global Infrastructure: The Potential of SRTM Data to Break New Ground. CIESIN White Paper.

Onsrud, H. J. (2009) Liability for Spatial Data Quality. In Devillers, Rodolphe & Goodchild (Eds.) *Spatial Data Quality: From Process to Decisions,* New York: CRC Press, 187-196.

Onsrud, H.J. (2010) Legal Interoperability in Support of Spatially Enabling Society. In Abbas, Crompvoets, Kalantari, & Kok (Eds.) *Spatially Enabling Society: Research, Emerging Trends and Critical Assessment,* GSDI Association and Leuven University Press, 163-172.

OpenStreetMap (2011) A new user's guide. Retrieved from the World Wide Web, July 6, 2012: http://en.flossmanuals.net/openstreetmap/index/

Potere (2008) Horizontal Positional Accuracy of Google Earth's High-Resolution Imagery Archive. *Sensors 8*, 7973-7981; DOI: 10.3390/s8127973.

Quackenbush, L. J. (2004) A Review of Techniques for Extracting Linear Features from Imagery. *Photogrammetric Engineering & Remote Sensing 70* (12), 1383-1392.

RCMRD (Regional Center for Mapping of Resources for Development) (2012) News: 3rd Country Group Training on Application of ALOS/Daichi for Mapping Opens – January 31, 2012. Retrieved from the World Wide Web, July 6, 2012: http://www.rcmrd.org/index.php?option=com_content&view=article&id=146:3rd-country-group-training-on-aplication-of-alosdaichi-for-mapping-opens-january-31-2012&catid=1:latest-news&Itemid=55

Roberts, L.G. (1963) Machine perception of three-dimensional solids. Massachusetts Institute of Technology. Retrieved from the World Wide Web, April 15, 2014: http://dspace.mit.edu/handle/1721.1/11589#files-area

Siebritz, L., Sithole, G., & Ziatanova, S. (2012) Assessment of the homogeneity of volunteered geographic information in South Africa. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 39*, 553-558.

Small, C. (2003) High spatial resolution spectral mixture analysis of urban reflectance. *Remote Sensing of Environment 88*, 170-186.

Sobel, I. (1968) "A 3x3 Isotropic Gradient Operator for Image Processing", presented at the Stanford Artificial Intelligence Project (SAIL).

Sobel, I. (2014) History and Definition of the Sobel Operator. Retrieved from the World Wide Web, April 15, 2014: http://www.researchgate.net/publication/239398674_An_Isotropic_3_3_Image_Gradient_Operator

Sui, D.Z., Elwood, S., & Goodchild, M.F. (Eds.) (2012) *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*. New York: Springer.

Tao, C. V. & Li, J. (2007) Advances in Mobile Mapping Technology. *ISPRS Book Series, Taylor & Francis Group, London, ISBN 978-0-415-42723-4.*

Tracks4Africa (undated) *Standards for Field Data Collection,* Retrieved from the World Wide Web, April 15, 2014: http://tracks4africa.co.za/media/flatstuff/3_Standards_for_Field_Data_Collection_1.pdf

Ubukawa, T. (2013). An Evaluation of the Horizontal Positional Accuracy of Google and Bing Satellite Imagery and Three Roads Data Sets Based on High Resolution Satellite Imagery. Unpublished manuscript for the CODATA Roads Task Group, available at http://www.groads.org.

Uruguay Round Agreement (Part II, Section 1, Article 10) of the World Trade Organization (effective in 1995). See also Feist Publications, Inc., v. Rural Telephone Service Co., 499 U.S. 340 (1991).

Yuan, J., Wang, D., Wu, B., Yan, L., & Li, R. (2011) LEGION-based automatic road extraction from satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *vol. 49*, No. 11.

Zhao, H., Kumagai, J., Nakagawa, M., & Shibasaki, R. (2002) Semi-automatic road extraction from high-resolution satellite image. *Proceedings of the ISPRS Congress*, Photogrammetric Computer Vision, 406-411.