

# 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

**Abstract:** There is an urgent need for a global spatial “public commons” roads data set with improved geographic and temporal coverage, consistent coding of road types, and clear documentation of sources. The private sector and military have, for different reasons, been unable or unwilling to release improved data into the public commons for use by the wide range of practitioners and researchers who need it. This document first describes the demand for such a data set (especially to make progress towards the International Strategy for Disaster Reduction and the MDGs) and then lays out a strategy for developing an initial product. Finally, it describes a method for producing regular updates that accurately reflect the world’s growing network of roads.

## Table of Contents

I. Introduction & Rationale .....	2
A. Aims and objectives .....	2
B. The need .....	4
C. Why this data set won’t otherwise be developed .....	9
D. Summary .....	12
II. Specifications .....	12
III. Methods.....	13
A. Developing roads data from scratch.....	13
B. Combining existing data from multiple sources .....	15
C. Software and information systems development .....	15
IV. Project Management .....	17
Annex 1. Participant List .....	19
Annex 2. The Data Model.....	20
Annex 3. CODATA Working Group on Roads Data Development.....	22
References.....	23

For more information on the Global Roads Data Project, including PDFs of presentations at the workshop and current updates, visit:

<http://www.ciesin.columbia.edu/confluence/display/roads/>

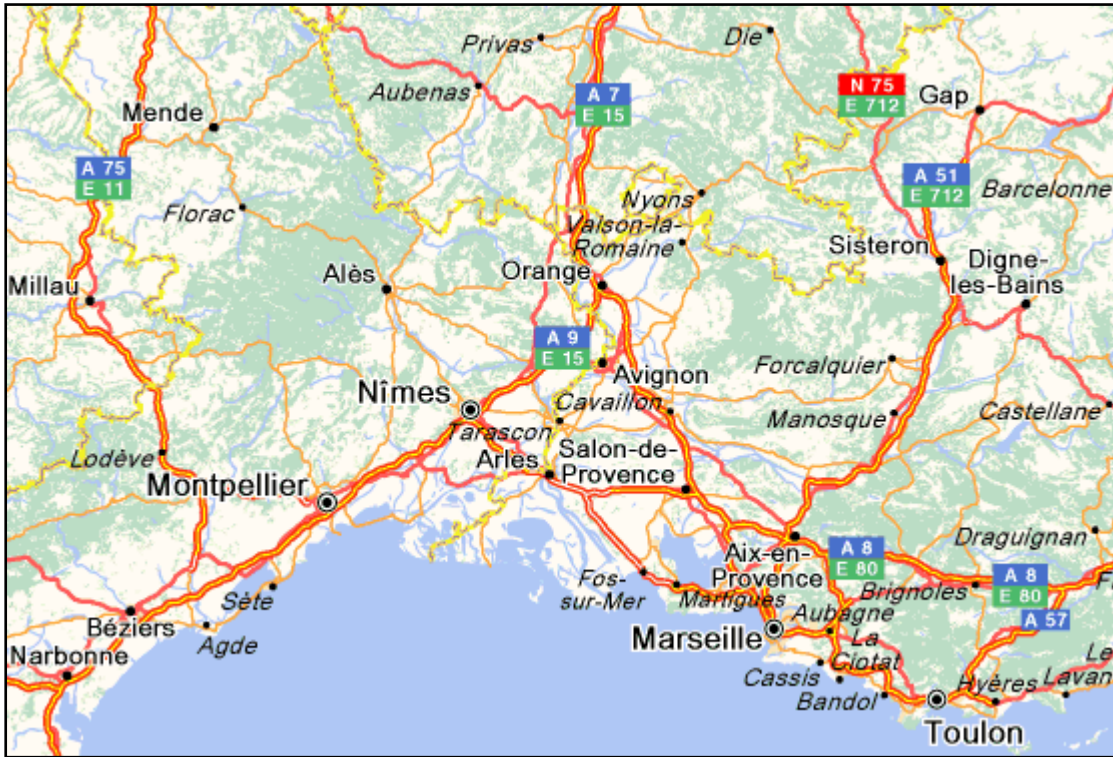
## I. Introduction & Rationale

### A. Aims and objectives

In many countries, roads and highways provide the dominant mode of land transport. They often carry more than 80 percent of passenger-km and over 50 percent of freight ton-km in a country. Consequently, roads, and highways form the back bone of the economy and provide essential links to the vast rural hinterlands (World Bank, 2005a). It is estimated that the value added by transport accounts for 3 to 5 % of GDP and 5 to 8 % of total paid employment (World Bank, 2002). Given the enormous economic and social importance of road transport, it is surprising that there is currently no good quality, freely available global spatial data set for road networks. The best available global public commons product, Vector Smart Map level 0 (known as VMAP0 or Digital Chart of the World) transport layer, covers only one-quarter to one-third of the existing road networks, and this varies considerably by region. The data from VMAP0 are of uncertain date and provenance, and there is clear inconsistency across tiles in the level of road network detail.

Among researchers and practitioners in the development, hazard response, biodiversity conservation, and public health communities there is high demand for improved global roads data. The ideal data set would have improved geographic and temporal coverage, consistent coding of road types, and good documentation of sources, and it would be available free-of-charge on an “attribution only” basis (de Sherbinin and Chen, 2005; Nelson *et al.*, 2006). Recognizing this unmet need, a range of experts and representatives of UN and government agencies participated in a three-day workshop (1-3 October 2007) on Global Roads Data at the Lamont Campus of Columbia University to develop a strategy to develop an improved global roads data set (see Annex 1 for participant’s list). The workshop was organized by the Socioeconomic Data and Applications Center (SEDAC) of Columbia’s Center for International Earth Science Information Network (CIESIN) and co-sponsored by the Consultative Group on International Agricultural Research - Consortium for Spatial Information (CGIAR-CSI), the international Council for Science’s Committee on Data for Science and Technology (ICSU-CODATA), and the World Resources Institute.

A number of the participants and their respective organizations have determined to form a consortium of interested parties to advance the development of a global roads data set. The consortium aims in a first phase to develop a global roads data set focused on inter-urban transport networks that is analogous in scale and content to a provincial-level Michelin road map (e.g. 1:200,000 scale) (Figure 1). The data set would include primary, secondary and tertiary roads, and the positional accuracy of road locations would be 100m or better. A consistent data model will be used for this global coverage, with consistent coding of road types, weight restrictions, and related information. The consortium’s initial focus will be on developing countries, starting with Africa, then moving to South Asia, South and Southeast Asia, Latin America, and Oceania. But the aim is to have a consistent global coverage within two years of project inception.

**Figure 1. Michelin 1:200,000 Scale Map for the South of France**

Source: ViaMichelin at <http://www.viamichelin.com/viamichelin/int/dyn/controller/Maps>

In a second phase, the consortium will develop an online tool to allow government, UN agencies and any organization or individual with a vested interest in developing better, free road data to edit and improve upon this map using GPS tracks, satellite imagery, and other source data in order to keep the global map up to date, incorporate more roads, and provide expanded information about each road and/or road segment. Inputs using this online tool will be evaluated by a community of experts formed as an ICSU-CODATA working group, and periodic updates will be released based on verified additions. The goal is to improve the spatial resolution over time to the equivalent of 1:50,000 scale on a paper map.<sup>1</sup> Targeted data on road types of interest to a specific community (e.g. logging roads for the conservation community, or cart tracks for rural agricultural development) would be added as time and resources allow, and on the basis of inputs from those communities.

The data set will be available to any potential users, including UN and bilateral agencies involved in disaster response and reconstruction; the development banks and bilateral donors involved in international economic development; the biodiversity conservation and carbon-offsets investment communities; the environment and development research community; national and regional agencies and organizations in the developing world; and to the public at large. The needs of these communities are described in the next section.

<sup>1</sup> As a point of reference, the Tiger line files distributed by the US Census Bureau include every street in the United States and are at a scale of 1:100,000 with a positional accuracy of 50m.

## ***B. The need***

Applications for such a data set span multiple sectors and would be particularly valuable for a number of purposes, as described in this section. Workshop participants underscored the critical need for a good global baseline data set for disaster response, recovery, and mitigation, as well as for development planning to achieve the MDGs.

### ***Pre- and post-disaster planning***

- *Disaster response*

In the immediate aftermath of a major disaster, national disaster management agencies, UN agencies and other first responders urgently require up-to-date roads network data sets in order to assess their ability to access or evacuate victims, ship relief supplies, and send heavy equipment to the scene. Surprisingly, such data are not always readily available, and the lack of such data can mean lost time and the cost of further lives. The military may temporarily lend digital maps or high resolution imagery for humanitarian purposes, but they often require that these data be returned after the disaster, limiting their utility for relief to development work. A further problem with data from military sources is that by the time it is declassified it is usually obsolete, already commonly available, or too late to make a difference to operational planning.

- *Emergency planning*

The resilience of a transport network, or its ability to accommodate unexpected conditions, is a key concept for planning evacuation strategies when dealing with the inherent uncertainty of natural and man-made disasters (Berdica, 2002; Litman, 2004; Morlok and Chang, 2004). Such resilience can be modeled and assessed by transport planners with good spatial road network information.

- *Impact assessment*

Following a loss of transport infrastructure, how can we ensure the best appropriation of often limited reconstruction resources? Who has been affected and where? Again, good spatial transport information is vital for rapid informed assessments for damage assessment, the loss of accessibility, the demographics of the population at risk, and the pros and cons of different resource allocation strategies (Winograd, et al., 1999).

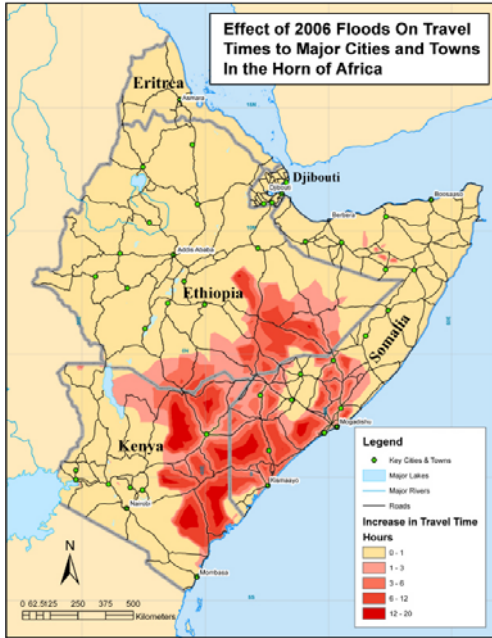
### ***Economic Development***

- *Poverty, health and inequality issues*

Lack of mobility and high transportation costs are key impediments that lead to the formation of ‘spatial poverty traps’ (Deichmann, 1999; Pender, et al., 1999; Bigman and Fofack, 2000; Chomitz et al. 2007). At the macro-level, access to safe water, electricity, and the road network have been shown to be positively correlated with national per capita income (Sarkar and Ghosh, 2000). Methods to identify areas of social and economic disadvantage are often dependent on spatial indicators of accessibility and connectivity to ensure accurate geographic targeting (Leinbach, 1995; Higgs and White, 2000). Hotspots of inequality in service provision (such as water, education, and health) can be assessed by deriving catchment areas around existing facilities in combination with population data (Williams, 1987; Airey, 1992; Hope, 2006), and the catchment areas themselves can

be defined by travel time which is mostly a function of road networks. Optimal locations for futures facilities can then be assessed.

**Box 1. The Importance of Roads Data for Emergency Response**



The map at **left** of travel time costs owing to a major flood in 2006 in the Horn of Africa region shows the value of combining road network data with digital elevation models (DEMs), flood remote sensing or meteorological data in order to plan for flood response, or to allocate additional travel time in the event of floods coupled with some other emergency.

The table and map **below** show an optimal travel time and travel distance matrix between key locations in Mozambique during the 2007 floods. Humanitarian logisticians use these products to plan their operations and decide between the use of surface or air transportation for aid distribution. The matrix is associated with maps showing optimal routes between any two locations. Travel speed and cost parameters take into account road class, surface and vehicle practicability as well as any obstacles affecting the network (see Data Model under Section II – Specifications).

Produced by Paul Bartel, HIU, US State Dept.

Distance (Km) Travel Time	Beira	Blantyre	Caia	Charre	Chenba	Chiramba	Chupanga	Guro	Lilongwe	Luabo	Manica	Maputo
Beira		12h2'	4h28'	6h13'	7h43'	9h49'	6h26'	6h42'	15h18'	16h34'	4h12'	19h49'
Blantyre	567		7h41'	5h48'	8h14'	10h20'	10h50'	6h43'	5h10'	14h1'	10h36'	29h1'
Caia	268	307		1h53'	3h22'	5h29'	3h16'	10h7'	12h37'	12h13'	7h37'	23h14'
Charre	337	230	77		2h26'	4h32'	5h2'	11h53'	10h44'	11h3'	9h23'	24h59'
Chemba	358	288	97	58		2h6'	6h31'	10h51'	13h10'	12h46'	10h52'	26h29'
Chiramba	400	330	139	100	42		8h38'	8h46'	15h16'	14h52'	12h38'	28h35'
Chupanga	265	346	47	116	136	179		12h6'	15h46'	15h22'	9h35'	25h12'
Guro	403	369	608	677	232	190	604		8h36'	20h30'	3h52'	22h17'
Lilongwe	919	306	608	531	589	631	647	516		18h57'	12h28'	30h53'
Luabo	534	628	273	226	255	297	312	993	929		19h43'	35h20'
Manica	252	601	458	527	547	422	454	232	748	723		19h47'
Maputo	1189	1707	1395	1463	1484	1526	1391	1338	1854	1660	1188	
Marromeu	310	391	92	161	182	224	45	650	692	358	499	1436
Massenguza	364	445	146	215	236	278	99	704	746	412	554	1490
Mopeia Velha	467	561	206	159	188	230	245	926	862	67	656	1593
Morrumbala	392	314	131	84	113	155	171	732	615	158	581	1518
Mutarara	324	245	63	15	44	87	102	663	546	212	513	1450
Nampula	1010	647	749	654	731	773	789	1011	948	684	1244	2349
Quelimane	587	440	326	278	307	349	365	804	741	261	776	1713
Sena	321	251	60	21	37	79	99	660	552	218	510	1447
Tambara	811	456	757	681	739	554	797	408	480	1079	640	1746
Tete	561	211	513	436	494	305	552	158	358	834	391	1496

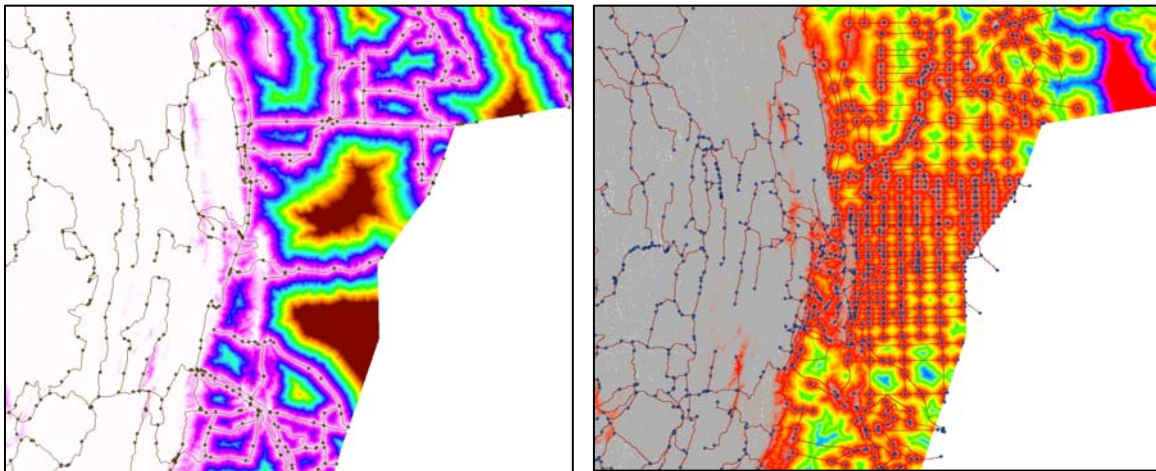
Produced by the UN Joint Logistics Centre, February 2007

- *Rural agriculture*

Better rural transportation is a principal factor for improving livelihoods in developing countries through better access to markets, increased social mobility, migration, and greater economic opportunities (Leinbach, 1995; Barwell, 1996; Dixon-Fyle, 1998). Good road information is important to allow local entrepreneurs to make better plans for the distribution of their products, and also to enable development organizations to assess the social and environmental impact of competing transportation strategies. Infrastructure, especially roads that are in good condition, are an important element of a broader strategy to reduce transactions costs and increase agricultural productivity. It has been suggested that road surface improvements can result in decreased costs of inputs and increased prices for produce at the farm gate (Hine and Riverson, 2001). Development banks need information on the state of the current road network in order to plan for future infrastructure investments. Other potential applications include assessing the availability and accessibility of agricultural inputs, optimal locations for seed distribution centers and post-harvest facilities, and road construction for “opening” agricultural lands.

### Box 2. Road Access and Development

Proximity to roads has a major impact on the price of agricultural inputs and farm gate prices for small producers. Policy makers need improved roads data to identify hotspots of poverty or unequal access to public services, as well as to plan the location of critical facilities. Inadequate roads data can lead to mis-identification of areas of greatest need. The map at **left** represents an accessibility map based on low resolution/poor quality roads data (yellow to brown colors represent the most isolated areas), and the map at **right** represents an accessibility map for the same region based on high resolution/high quality roads data (blue to red areas represent the most isolated areas). Allocation of development resources based on the roads data at left would not yield optimal results, since some of the apparently most inaccessible regions actually have dense road networks.



Maps courtesy of Glenn Hyman, CIAT

### ***Environment and Land Use***

- *Impact analysis for road development*

The ability of GIS to communicate and visualize the effects of roads on the environment is extremely valuable. Spatial models of the impact of road construction options can help to link the effect of road development on land cover change and biodiversity loss (Chomitz and Gray, 1996; Rapaport and Snickars, 1999; Laurance, et al., 2002; Nagendra, et al., 2003; Hawbaker, et al., 2004; Nelson and Leclerc, 2007).

- *Threats to protected areas and in-situ conservation issues*

Poorly planned road developments can pose a risk to protected areas and regions of high biodiversity importance (Ji and Leberg, 2002). The possible consequences of road development, such as deforestation, habitat fragmentation (Jaarsma and Willems, 2002), increased wildlife mortality, and increased population pressure, are all factors that can increase the risk of biodiversity loss (Guarino, et al., 2002). Some species are known not to cross roads of a certain width, and hence fragmented habitats may reduce population viability. Good transport planning aided by accurate and up-to-date road network maps can help to minimize the risk to such important areas and also to assess alternative transport options.

- *Carbon sequestration*

Financial incentives for carbon sequestration should pay for areas that would otherwise be deforested. Remote or inaccessible areas are not good candidates since they would be unlikely to be deforested anyway. Improved roads data would help to target financial allocations for preservation of intact forests as carbon sinks, and would help to better project likely future deforestation.

- *Planning of collections*

Plant Genetic Resource programs often face financial constraints for the collection of germplasm. Increasingly, methods for optimizing and prioritizing areas for the collection of germplasm are being based on GIS targeting strategies that include areas that are accessible by road as well as more traditional inputs such as predicted species distribution, climate, and land cover (Jarvis, et al., 2005).

### ***Research Community***

- *Agriculture and rural development*

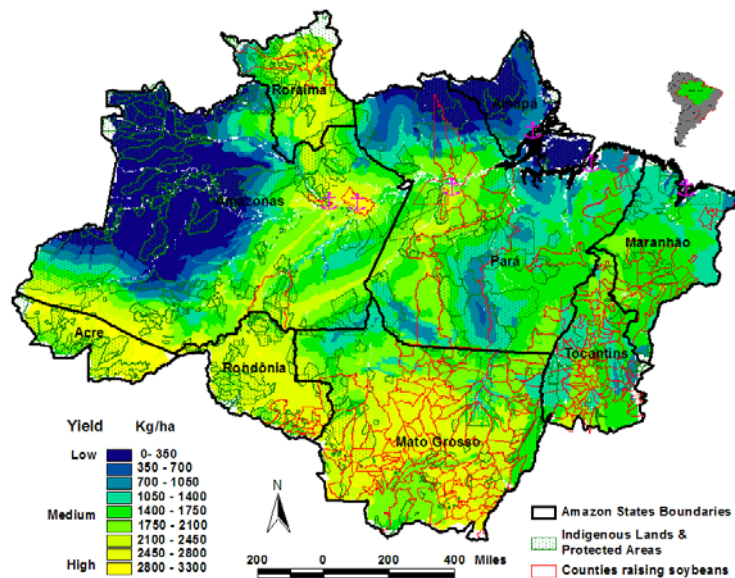
Research interest in the linkages between access (and related notions of remoteness and isolation) and rural livelihoods has spawned a growing literature on the characterization and empirical measurement of “access” and its welfare implications; including impacts on land use, enterprise mix, technology adoption, production intensification, and the degree of participation in input and output markets. Examples include Chomitz and Gray (1996) on land use, Omamo (1998) on crop choice, Staal et al. (2002) on technology adoption, Edmonds (2002) on intensification, Lanzona and Evenson (1997) on labor markets, and Fafchamps and Vargas Hill (2005) on output market participation. Road network data are critical inputs to such studies.

- *Accessibility as a driver of land use change*

Land use and land pricing models have often been based on the von Thünen model (Hite, 2000). However, some recent studies have investigated the possibility of replacing distance to market (Euclidean distance) with time to market (an economic distance) in such models to generate more realistic results for land use modeling (Chomitz and Gray, 1996; Verburg, et al., 2004; Nelson and Leclerc, 2005). Road types and conditions are the key factors in calculating economic distance.

### Box 3. Analyses Using Roads Data for Biodiversity Conservation

Road expansion and improvement increases the farm gate price of commodities such as beef, soybeans and palm oil, and is a powerful economic incentive for the expansion of plantations on the forest frontier (Chomitz et al., 2006). These products are also under increasing global demand as food products and biofuel feed stocks. Conservation planning with better knowledge of road networks can diminish the cost of trade-offs between biodiversity conservation and the expansion of livelihood opportunities in agriculture and forestry. By targeting economic incentives to landholders and indigenous groups for ecosystem services such as Reduced Emissions from Deforestation (RED), these services can be conserved in wilderness and working landscapes. Designing effective deterrents to uneconomic land transformation requires greater understanding of the location of areas with lower agricultural potential and lower opportunity cost, as described in Vera Diaz (2008) with respect to soybeans in the Amazon. One of the greatest influences on land opportunity cost is road proximity (Chomitz and Gray, 1996). A key input to opportunity cost models is information about planned road improvements. Planned road improvements effect land prices and agricultural transformation even before they are built (Soares-Filho, 2006).



Source: Vera-Diaz et al. (forthcoming).

- *Population modeling*

Good quality spatially referenced roads data sets are key inputs to many population-related applications. Spatial models of population distribution are frequently based on the assumption that population density is greatest in areas of good accessibility and high transport network density. Therefore, consistent and reliable road network information is



vital for improving spatial population estimates at medium and high resolution (Deichmann, 1997b; Dobson, et al., 2000; Nelson and Deichmann, 2004; Hyman, et al., 2005).

- *Road safety*

Road accidents are a growing cause of fatalities, especially in developing countries, and the public health community has an interest in tracking the phenomenon. An accurate roads data set would help researchers to study the severity of the problem in relation to indicators of road traffic so as to put in place measures that would reduce transport-related fatalities. WHO has recently allocated US\$9m for the improvement of its road safety data systems.

***Other Beneficiaries***

- *Private sector*

The development of high quality, consistent roads data in many developing countries where such data do not currently exist could have major benefits for small enterprises and geospatial consulting firms, who are likely to use these data to develop value-added products and services. Far from discouraging such use, these kinds of products and services should be encouraged in that they are likely to be a spur for economic development and job creation.

***C. Why this data set won't otherwise be developed***

This data set is unlikely to be developed by any organization or entity other than the consortium proposed in this document. This is because there is little incentive for the private sector and the military to develop a high-quality roads data set and then make it freely available, and the WIKI approach to the development of roads data by a large community of volunteers is unlikely to extend beyond the most developed countries.

***Private Sector Roads Data***

Two major companies – Navteq and Tele Atlas – currently have well developed global roads databases that are primarily used for personal navigation devices, and which also have applications in public and private sector areas. The detail and completeness of these data sets for the developed world and some rapidly industrializing countries such as China are truly impressive. In fact, for developed countries it goes beyond the detail needed in the data set proposed in this strategy paper. Yet, outside of developed regions, the data developed by these two firms is often of uncertain provenance and uneven quality. These two firms together with major partners are taking steps to improve the data in developing regions, but the fact remains that markets for the least developed countries are weak, and hence the incentive to develop such data are low.

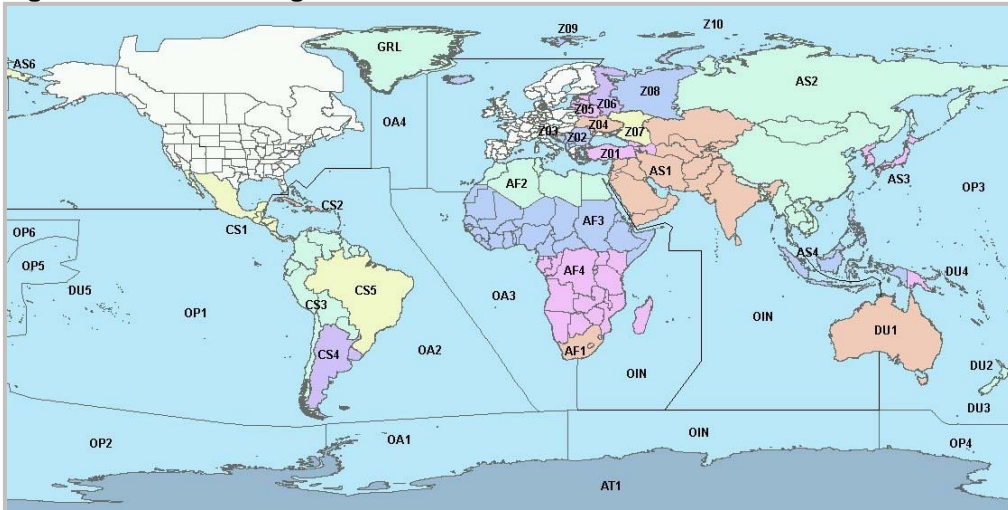
The more important issue, however, is that these data are covered by strict copyright protections and the costs are prohibitive for the user community described above. For example, to purchase Tele Atlas data for all of Africa would cost 20,000 Euros for a one-

year license and up to five users. Most of Africa’s coverage is of “low density”, described as “most important highways connecting cities.” The country group labeled AS1 (Figure 2) would cost 80,000 Euros for the same one-year license and number of users.

Although the private sector appears willing to engage in partnerships that would grant a limited user community with rights to access the data for certain limited purposes, the fundamental principle of open-access runs counter to their profit motivation and model of intellectual property rights. Thus, though some users might benefit for some period of time in engaging with the private sector and using their data, the arrangement would be precarious and would leave the wider community without access to improved data.

The *Economist* (“Location, Location, Location,” October 4, 2007) warns that with the recent acquisition of Navteq by Nokia and Tele Atlas by TomTom, the “world may end up with a digital map monopoly, as users migrate to the provider with the most comprehensive data and then further strengthen its position by adding information of their own.” The article notes that it is fears such as this that have spawned the “open-map” movement described below. But this approach also has its limitations, as will be discussed.

**Figure 2. Tele Atlas Regions**



Map courtesy of John Auble, Tele Atlas.

### ***Military-Intelligence Community Roads Data***

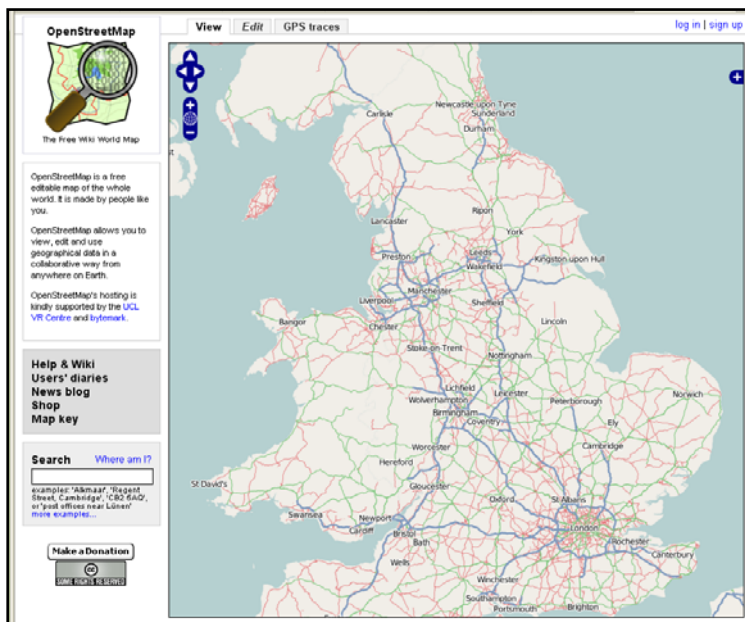
The defense and intelligence communities in the US and Europe also have potentially useful roads data sets, though their accuracy and completeness is not well known since most of the data remain classified. Approximately 65 tiles of VMAP1, developed by the National Geospatial Agency (NGA) of the US Government, are available, and they compare favorably to VMAP0. Yet the tile gaps leave large portions of the world uncovered. The military has been known to share data in the case of disaster response, but usually for limited times only.

As with the private sector, the incentive structures and imperatives of this community seem fundamentally opposed to open-access distribution, although portions of their data that are available may be useful for validation of, or even for incorporation into, a new global product.

### *OpenStreetMap Roads Data*

A new and most interesting development is the emergence of a community of GPS users who digitize street and roads data in a WIKI environment and who, through the sheer force of numbers, are able to rapidly develop data (and verify existing data) for large areas. At present the development of roads and street data under OpenStreetMap (OSM) appears to be concentrated largely in urban areas of the developed world, but the potential for this approach for developing future data over wider areas should not be underestimated. OSM's founder, Steve Coast, projects that mapping of all of the European Union will be completed by 2011.

Although elements of the OSM approach (so called “crowd sourcing” – the engagement of large numbers of people on a voluntary basis) would be most useful for development of a global roads data set and ongoing updates, the on-the-ground GPS approach to crowd sourcing is unlikely to develop data for large parts of the developing world apart from urban areas and where wealthy tourists travel. Crowd sourcing the digitization of basic road networks from imagery probably has greater potential. A web service supporting semi-automated line following and object extraction techniques could greatly advance the ability of untrained volunteers to help map the world. This form of web processing service is described in greater detail in Section III.C. Unlike the private sector and military-intelligence communities, the constituency for crowd sourcing is highly committed to free and open access to data.



**Figure 3. OpenStreetMap:** Developed by a community of users in a WIKI environment, OSM represents an important approach to the current and future development of global roads data.

### ***D. Summary***

For the reasons described above, workshop participants agreed that it is time to develop a new, globally consistent data product that will meet the needs of the aforementioned user communities and fill the gaps that currently exist because of the incentive structures and mandates of the three communities currently developing roads data: the private sector, the military, and the open-source community. The next two sections describe the proposed specifications and data development methods for this new data product.

## **II. Specifications**

Well-defined data specifications are critical to the successful development of a consistent roads data set.

### ***Scale and Accuracy***

In its first phase, this project aims to develop a global roads data set focused on inter-urban transport networks that is analogous in scale and content to a provincial-level Michelin road map (e.g. 1:200,000 scale). The data set would include primary, secondary and tertiary roads, and the positional accuracy of road locations would be 100m or better.

### ***Data model***

The data model underlying the global roads data set consists of two broad elements:

#### **1. Terminology and Classification**

The data model used for this global coverage will build on UN Spatial Data Infrastructure specifications. The UNSDI is a UN-wide initiative to encourage consistent protocols in geographic data collection, processing, storing and access in order to ensure efficient data sharing and interoperability between organizations.<sup>2</sup> The data model contains consistent coding of road types, weight restrictions, and related information. Table A1 (in Annex 1) describes the fields included in this data set. Table A2 describes the metadata definition for each record in the data set. Each road segment will include information on its provider, its collection date as well as an indication of data quality and reliability. Table A3 describes the permissible values for one of these attributes (in this example, road functional class). Initially, only information on primary, secondary and tertiary roads will be collected.

#### **2. Database structure and functionality**

The database would be structured so as to allow basic network analysis and routing functions in addition to cartographic representation. These would include deriving macro and meso-level transport costs, optimal routes between population centers, contingency plans in case of shocks to the network and optimized road

<sup>2</sup> See [www.ungiwg.org](http://www.ungiwg.org) and [www.unjlc.org/mapcenter/unsdi](http://www.unjlc.org/mapcenter/unsdi)

rehabilitation investment decisions. This implies ensuring topological consistency in the data, as well as the ability to establish connectivity with external data layers such as settlements and other transportation networks.

Finally, the database would be structured in order to allow versioning and maintenance of a historical archive of the evolution of global road networks.

### ***Data format***

Users will be given a choice of formats in which to download the data. This will include the Open GIS Consortium (OGC) compliant GML format as well as some of the main proprietary formats in wide-spread use such as ESRI© Shapefiles or other common vector data exchange formats such as ESRI ARC/INFO coverages and VPF.

## **III. Methods**

This section describes the methods proposed for the development of a new, public commons data set available on an “attribution only” or “public domain” basis. In a best case scenario, with unlimited funding, a globally consistent base map would be developed from scratch using a combination of scanned base maps, orthorectified imagery, GPS tracks, and data from automated roads extraction algorithms. This approach is described in Section A below, and may well be the best means available for obtaining accurate data for countries that have not been mapped adequately. However, it must be recognized that such an approach, if carried out globally, would be prohibitively expensive, and would result in duplicate data where adequate public commons roads data already exist. Thus, Section B describes an approach for compiling the best available roads data from multiple sources, using a “patchwork” approach that would cover the globe.

Whether data are developed from scratch or using a patchwork approach, attempting to maintain a global data set over time as new roads are constructed or upgraded (e.g., from dirt to paved) would be a challenge to even the best resourced of efforts. Thus, participants in the workshop emphasized the need to develop a suite of software tools, web services, and methods for ingesting new roads data and metadata from distributed sources, updating the data, and checking the reliability of contributed data. Section C addresses this need for software and information systems development, and provides preliminary specifications for their development.

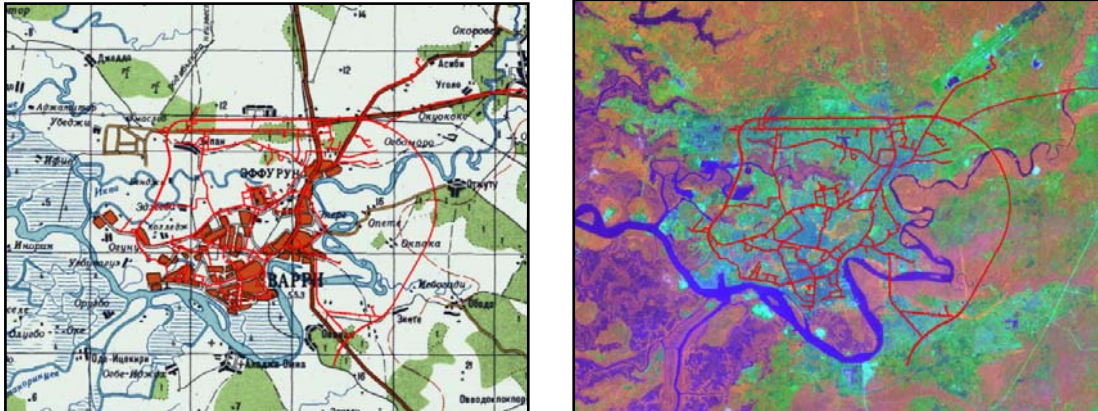
### ***A. Developing roads data from scratch***

A number of methods might be employed to develop roads data from scratch. One method, currently employed by Georigin, Ltd., in South Africa, is to manually digitize roads and attributes from the source data such as:

- Scanned 1:200,000 paper maps developed by the Russian military (ranging in dates from the late 1960s to the early 1980s) and the US Joint Operations Graphic (JOG) navigation maps.<sup>3</sup>
- Geocover Landsat pansharpened 15m imagery baselined to the year 2000, which are orthorectified and are available free of charge.
- GPS tracks wherever available to add the most recent routes.<sup>4</sup>

The relative spatial accuracy of these data are presented in Box 4. Employing this method, roads could be manually digitized and attributes assigned according to the data model described above. While digitizing, rail road networks, populated places and built up areas could be added for the sake of complete coverage of the node to node transport network in a country. Wherever possible efforts would be made to link the roads data set to data sets describing other modes of transportation, including ports and airports for shipping, air freight, and passenger transport. Altitude (or z-heights) can be added easily at a later stage in the data production by extracting elevation values for each node from a digital elevation model.

#### Box 4. Topographic Map, Landsat, and GPS Integrated



This example from Warri, Nigeria, illustrates how a Russian 1:200 000 topographic map at **left** (georeferenced, cropped, datum shifted to WGS84) can be integrated with data from Landsat 7 (geometrically enhanced with GPS ground control points) and GPS tracks at **right** to produce a road map. There is a fair coincidence level (50%) between the GPS tracks and the main roads on the topographic maps, but there is poor geometric accuracy between the topo maps and GPS Tracks (200m – 400m). On the other hand, there is good coincidence (90%) between GPS Tracks and main roads on the Landsat image, and better geometric accuracy between enhanced Landsat and GPS Tracks (50m to 100m). The GPS tracks represent the true position of the roads.

Source: John Dann, Georigin, Ltd.

<sup>3</sup> The Russian military maps and UN JOG maps covering many parts of the world are available through UN agencies.

<sup>4</sup> GPS tracks could be obtained from UN agencies or companies such as Tracks4Africa.

The consortium will assess the costs of deploying this method, whether through volunteers, employing so-called “click workers” in digitizing shops in India or elsewhere, or through a sub-contract with a private company.<sup>5</sup> It may be possible to develop an online software application similar to the one employed by OSM to facilitate this crowd sourcing, which is addressed further in Section C.

Another approach that the consortium will pilot test is automated road extraction from satellite imagery. Although algorithms for road extraction are mostly at an experimental or research phase, operational deployment may be possible with further testing and development. In order to provide the best results, the satellite imagery would need to be orthorectified. Once road segments are extracted, they will need to be verified and possibly corrected using a manual process that would compare the segments to “ground truth”, namely GPS tracks and moderate to high resolution orthorectified satellite imagery. Attribute information could also be added at the same time.

### ***B. Combining existing data from multiple sources***

A range of organizations support the development of roads data (e.g., the World Bank, World Resources Institute, International Steering Committee for Global Map (ISCGM), etc.) or disseminate data created by others (e.g., FAO’s Geonetwork, CGIAR-CSI, the Southern Africa Humanitarian and Disaster Management GIS Library, etc.). The data are of varying degrees of completeness and accuracy, but nevertheless represent a resource that should not be ignored.

As a first step, the CODATA working group (see Section IV) will oversee the development of a catalog of existing data, many of them compiled by Andrew Nelson at the European Commission Joint Research Centre (JRC). The catalog will be maintained by CIESIN, and depending on copyright restrictions, the best available data will be made available for download. The catalog will include information on, among other things, the source, the data and methods used to develop the data set, the total kilometers of roads included (by country), and use and redistribution restrictions.

Out of this collection it is hoped that the highest quality data can be stitched together and edited into a global mosaic covering a significant number of countries, with road segments connecting across borders (to form a topologically integrated network), and consistent attributes.

### ***C. Software and information systems development***

Although compiling a global roads data set using the methods described above is certainly feasible, improving its spatial resolution and coverage, and keeping it up to date with new road information, would tax the resources of even a multi-million dollar effort.

---

<sup>5</sup> A preliminary estimate by Georigin in South Africa suggests that wall-to-wall coverage of the globe using this methodology could be achieved within 18 months at a cost of approximately US\$12 million.

The navigation mapping industry has the resources needed, and even they face immense challenges in improving road data for the least developed countries (where markets for navigational products are thinnest). It is unlikely that any donor would underwrite the development of a public commons data set absent a strategy to maintain and improve it.

Thus, a key component of the overall strategy is to foster the development of a software tool that would allow “crowd sourcing” of roads data, using an OSM-type approach in which moderate to high resolution orthorectified satellite data are provided as a back drop, and users are able to upload GPS tracks or other ancillary data.<sup>6</sup> Users would then be able to manually digitize road segments and add attribute information according to the data model described above. The term “crowd sourcing” is used here to cover everything from voluntary individual contributors to potentially large governmental agencies that seek to build data for their countries based on this freely available tool.<sup>7</sup> The idea is to develop an OSM-type tool that is faster and more robust, given that it would need to service far more users.

The OSM-type tool would need to be developed in close collaboration with programmers. One approach would be to have users download an application to their desktop computer for digitization, and then allow them to upload results to the web mapping tool. This would increase processing speed and potentially allow the users to add more attribute information following the standard data model.<sup>8</sup> The same tool might allow selected “power users” to take new high resolution roads data, edit them to the data model specifications, and load them to the web mapping tool in place of the existing networks. Another approach would be to develop a robust ArcIMS custom application that would allow on-screen digitizing using the above combination of data sets with different opacity levels. If this tool could be developed to robust enough standards, it would be possible to have large numbers of people working on different tiles of the globe simultaneously in different locations in the world under the coordination of the consortium. In both cases, algorithms could be developed to statistically test the work of multiple click workers working on the same road network so as to identify significant differences and flag them for visual validation by supervisors.

Once data are entered, regional custodians would be alerted to additions and would validate newly entered data on a monthly or quarterly basis and, based on this validation, add selected road segments into the overall global quality controlled product. Since new segments would be date-stamped on the date of creation (see Table A2.2), custodians will be able to readily identify new segments that were added since the previous round of

---

<sup>6</sup> The consortium will look into different approaches to foster the development of this tool, either through public-private partnerships (e.g., with ESRI, Microsoft, or Google), in collaboration with groups that are already developing tools in this area (e.g., the OSM Foundation, the CERN/BOINC/UNOSAT Africa@home project, Bright Earth Project, the New Mexico Consortium, or Google Earth), or by administering a competition to develop such a tool in return for prize money.

<sup>7</sup> It would be possible to provide an online training program to be “certified” as an official contributor. Contributors would need to affirm that their contributions are free and clear of any intellectual property restrictions.

<sup>8</sup> The OSM WIKI allows users to enter basic attribute information, but does not currently support relational tables.



validation. A specific tool box in the web mapping tool would permit the custodian to do statistical testing and use other validation techniques. Since validation would generally only check to see if a road is correctly placed, the process of adding attributes would take place similarly to that described in the section above.

On an annual or bi-annual basis a new version of the data set would be rolled out that would reflect the most up-to-date snapshot of the product, including the validated crowd sourced data and any additional coding or corrections of road segments that was possible through validation by field teams. As mentioned above, additional attribute information might be added, for example, by adding attributes such as Z-heights by overlay of the data product on a global DEM (e.g., to identify areas at risk of flooding or landslides). This system will not permit globally consistent updates, since presumably some countries will be checked and updated more often than others. But it will represent a great improvement on the current decades-old data provided by VMAP0.

Additional tools to provide incentives for contributions would include user accounts, so that individuals can see what they've added to the global data set, and selected individuals might even be recognized for significant contributions, perhaps using the metric of numbers of kilometers digitized (e.g., 1,000km, 10,000km, and 100,000km clubs).

It must be acknowledged in advance that not all fields in the data model can be filled out based on remote on-screen digitization. For instance, road name, road and lane width, weight restrictions, and road surface material or condition may not be easily obtained from either paper, image, or GPS sources.<sup>9</sup> Field validation for further data development will be a critical component of building the depth of this data set. This can be accomplished in collaboration with UN operational agencies with an on-the-ground presence that have an interest in seeing this data set developed for their own operational purposes.

#### **IV. Project Management**

Coordination of the consortium will be provided by CIESIN with guidance from the CODATA Working Group on global roads data development including participants at the Global Roads Workshop and other experts in the field (see Annex 3 for a list of members). CIESIN and the Working Group will begin by developing a catalog of existing roads data, will help pilot test methods for automated road extraction from satellite imagery, and will investigate other means of developing the data. The Working Group will also strategize on the specifications for the software and information systems necessary for the development and maintenance of the data, and explore partnerships to develop the tool.

---

<sup>9</sup> Road condition can be derived from GPS tracks if one assumes that condition is directly related to travel speed.

*20 February 2008*

Over time, this initiative will ideally evolve into a community approach with data stewards for quality control (using automated statistical processes and visual validation) for a validated product.

Data will be distributed on an “attribution only” basis through CIESIN servers, and on the servers of selected partners.

## Annex 1. Participant List

The following individuals participated in the Global Roads Data Workshop hosted by the Socioeconomic Data and Applications Center of CIESIN, Columbia University, 1-3 October 2007, at the Lamont Campus of Columbia University.

<b>Name</b>	<b>Organization</b>
Keith Alger	Conservation International
Dalia Bach	Columbia University - LDEO
Imed Ben Hamadi	International Roads Federation
Robert Chen	Columbia University - CIESIN
Steve Coast	OpenStreetmap
Olivier Cottray	UN Joint Logistics Centre
Lorant Czarán	UN Geographic Information Working Group Secretariat
John Dann	Georigin, Ltd.
Alex de Sherbinin	Columbia University - CIESIN
Chris Elvidge	NOAA
Meredith Golden	Columbia University - CIESIN
Larry Gorenflo	Dept of Landscape Architecture, Penn State University
Johann Groenewald	Tracks4Africa
Timothy Haithcoat	Geographic Resources Center, University of Missouri
Glenn Hyman	CGIAR-CIAT
Koki Iwao	AIST/GEOGRID
Christopher Lenhardt	Columbia University - CIESIN
Marc Levy	Columbia University - CIESIN
Susan Minnemeyer	World Resources Institute
Jordan Muller	Humanitarian Information Unit, U.S. State Department
Maria Muniz	Columbia University - CIESIN
Siobhan Murray	World Bank
Andrew Nelson	EC Joint Research Centre
Harlan Onsrud	Dept of Spatial Info. Science and Engineering, University of Maine
Deborah Salon	Columbia University - Earth Institute
Christopher Small	Columbia University - LDEO
Carmelle (C.J.) Terborgh	ESRI
Suha Ulgen	UN Office of the Coordinator of Humanitarian Affairs
Stanley Wood	IFPRI
Greg Yetman	Columbia University - CIESIN

## Annex 2. The Data Model

**Table A2.1. Data Set Fields:** The following fields will be included in the roads data set.

Priority	FieldName	Type	Length	Description	AliasName	DomainName	DefaultValue	IsNullable	Precision
1	SourceID	Integer	4	SourceID	Source ID	null	null	false	0
1	ONme	String	100	ONme	Official Road Name	null	null	true	0
1	NtlClass	String	200	NtlClass	National Inventory Road Class	null	null	true	0
1	FClass	Integer	4	FClass	Functional Class	RdClass	0	true	0
2	Crgway	Integer	4	Crgway	Carriageways	Carriageways	0	true	0
2	NumLanes	Integer	4	NumLanes	Number of lanes	null	null	true	0
2	LneWidth	Integer	4	LneWidth	Lane Width (m)	null	null	true	0
2	RdWidth	Integer	4	RdWidth	Road Width (m)	null	null	true	0
2	Srf	Integer	4	Srf	Surface Type	SurfaceType	0	true	0
2	SrfCond	Integer	4	SrfCond	Surface Condition	SurfaceCondition	0	true	0
2	GradDeg	Integer	4	GradDeg	Gradient (specify +/-degrees)	null	null	true	0
2	Paved	Boolean	0	IsPaved	Is Paved	Boolean	false	true	0
2	Clearance	Float	5	Clearance	Clearance (m)	null	null	true	4
3	RteNme	String	50	RteNme	Alias Name	null	null	true	0
3	MaxAxleLoadMT	Integer	4	MaxAxleLoadMT	Maximum Axle Loading (MT)	null	null	true	0
3	MaxTotLoadMT	Integer	4	MaxTotLoadMT	Maximum Total Loading (MT)	null	null	true	0
3	IsSeasonal	Boolean	4	IsSeasonal	Affected by season	Boolean	0	true	0
3	DefaultPrac	Integer	4	DefaultPrac	Non-difficult Season Road Practicability	RdPracticability	0	true	0
3	SeasonalPrac	Integer	4	SeasonalPrac	Difficult Season Road Practicability	RdPracticability	0	true	0
3	SpeedLimit	Integer	4	SpeedLimit	Speed Limit (Km/hr)	null	null	true	0
3	DefAvgeSpeed	Integer	4	DefAvgeSpeed	Default Average Speed	null	null	true	0
3	SeaAvgeSpeed	Integer	4	WetAvgeSpeed	Seasonal Average Speed	null	null	true	0
3	Access	Integer	4	Access	Access	Access	0	true	0
3	WrksEDC	Date	8	WrksEDC	Road Works Est Date of Completion	null	null	true	0
3	DrivSide	Integer	4	DrivSide	Driving Side	DrivingSide	0	true	0
3	BiDirectional	Boolean	0	IsTwoWay	Is Two Way	Boolean	null	true	0
4	Notes	Blob	0	Notes	null	null	null	true	0

**Table A2.2. Source Identification:** This table links to the SourceID field described in Table A1. Each feature will have source information associated with it.

Priority	FieldName	Type	Length	Description	AliasName	DomainName	DefaultValue	IsNullable	Precision
1	UID	Integer	4	UID	User ID	null	null	true	0
1	ADt	Date	8	ADt	Acquisition Date	null	null	true	0
1	EDt	Date	8	EDt	Edit Date	null	null	true	0
1	FPNme	String	50	FPNme	Focal Point Name	null	na	true	0
1	FPPhn	String	50	FPPhn	Focal Point Phone	null	na	true	0
1	FPEml	String	50	FPEml	Focal Point Email	null	na	true	0
1	SrcType	Integer	4	SrcType	SrcType	SourceType	0	true	0
1	GeoSrce	String	50	GeoSrce	Geometry Source	null	na	true	0
1	AttSrce	String	50	AttSrce	Attribute Source	null	na	true	0
1	GeoQual	Integer	4	GeoQual	Geometry Quality	DataQuality	0	true	0
1	AttQual	Integer	4	AttQual	Attribute Quality	DataQuality	0	true	0
1	Editor	String	50	Editor	Editor	null	na	true	0
4	Notes	Blob	0	Notes	null	null	null	true	0

**Table A2.3. Road Class:** This table links to the Domain Name RdClass described in Table A1.

<b>DomainName</b>	RdClass	
<b>DomainType</b>	CodedValue	
<b>FieldType</b>	Integer	
<b>MergePolicy</b>	DefaultValue	
<b>SplitPolicy</b>	DefaultValue	
<b>Description</b>	null	
<b>Owner</b>	null	
<b>Coded Values</b>		
<b>Name</b>	<b>Code</b>	
Primary	2	connects important cities
Secondary	3	important city, city, or to town
Tertiary	4	village to anywhere
Local/ Urban	5	within or surrounding settlements (does not include primary, secondary, or tertiary roads within settlements)
Trail	6	
Unspecified	0	

### Annex 3. CODATA Working Group on Roads Data Development

The following individuals are members of the proposed CODATA Working Group on Global Roads Data Development. Additional members may be added at a future date. More information on the objectives of this working group can be found at <http://www.codata.org/taskgroups/WGglobalroads/index.html>.

<b>Name</b>	<b>Organization</b>
Olivier Cottray (co-chair)	UN Joint Logistics Centre, Italy
Alex de Sherbinin (co-chair)	CIESIN, Columbia University, USA
Steve Coast	OpenStreetmap, UK
John Dann	Georigin, Ltd., South Africa
Johann Groenewald	Tracks4Africa, South Africa
Timothy Haithcoat	Geographic Resources Center, University of Missouri, USA
Glenn Hyman	CGIAR-CIAT, Colombia
Koki Iwao	AIST/GEOGRID, Japan
Andrew Nelson	EC Joint Research Centre, Italy
Harlan Onsrud	Dept of Spatial Info. Science, University of Maine, USA
Jinnian Wang	Institute of Remote Sensing Applications (IRSA), Chinese Academy of Sciences (CAS), China

## References

- Airey, T. (1992) The impact of road construction on the spatial characteristics of hospital utilization in the Meru district of Kenya, *Social Science & Medicine*, 34, 1135-46.
- Berdica, K. (2002) An Introduction to Road Vulnerability, *Transport Policy*, 9, 117-27.
- Bigman, D., & Fofack, H. (2000) Geographical Targeting for Poverty Alleviation, Methodology and Applications, Working Paper, 1-307 pp, The World Bank, Washington, D.C.
- Chomitz, K. M., & Gray, D. A. (1996) Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize, *World Bank Economic Review*, 10, 487-512.
- Chomitz, K.M., Buys, P., De Luca, G., Thomas, T.S. and Wertz-Kanounnikoff, S. (2007) *At Loggerheads? Agricultural Expansion, Poverty Reduction, and Environment in Tropical Forests*. The World Bank, Washington, DC.
- de Sherbinin, A., and Chen, R. S. (2005) *Global Spatial Data and Information User Workshop: Report of a Workshop*, 21-23 September 2004, Socioeconomic Data and Applications Center, Center for International Earth Science Information Network, Columbia University, Palisades, NY. Accessed on 14 January 2006, <http://sedac.ciesin.columbia.edu/GSDworkshop/>
- Deichmann, U. (1997b) Accessibility indicators in GIS, 24 pp, Information and Policy Analysis. United Nations Statistics Division, NY, USA.
- Deichmann, U. (1999) Geographic aspects of inequality and poverty, Accessed on 30 December 2005, <http://www.worldbank.org/poverty/inequal/povmap/povmap.pdf>
- Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., & Worley, B. A. (2000) LandScan: A Global Population Database for Estimating Populations at Risk, *Photogrammetric Engineering & Remote Sensing*, 66, 849-57.
- Higgs, G., and White, S. (2000) Alternatives to census-based indicators of social disadvantage in rural communities, *Progress in Planning*, 53, 1-81.
- Hite, J. P. (2000) The Thunen model as a paradigm for rural development, Accessed on, [www.strom.clemson.edu/opinion/thunen.html](http://www.strom.clemson.edu/opinion/thunen.html)
- Hope, R. A. (2006) Evaluating water policy scenarios against the priorities of the rural poor, *World Development*, 34, 1, 167-179.

Hyman, G., Nelson, A., Lema, G., & Deichmann, U. (2005) Latin America and Caribbean Population Database Version 3, International Center for Tropical Agriculture (CIAT) and United Nations Environment Programme (UNEP), Cali, Colombia.

Leinbach, T. (1995) Transportation and Third World Development: Review, Issues, and Prescription, *Transportation Research*, 29A, 337-44.

Litman, T. (2004) Lessons From Katrina and Rita: What Major Disasters Can Teach Transportation Planners, Accessed on 05/12/05, <http://www.vtpi.org/katrina.pdf>

Morlok, E. K., and Chang, D. C. (2004) Measuring Capacity Flexibility of a Transportation System, *Transportation Research A*, 38, 405-20.

Nelson, A., & Deichmann, U. (2004) African Population Database Version 4, United Nations Environment Programme, Sioux Falls, SD, USA.

Nelson, A., A. de Sherbinin, and F. Pozzi. (2006). Towards Development of a High Quality Public commons Global Roads Database. *Data Science Journal*, Vol. 5, pp. 223-265. Available from [http://www.jstage.jst.go.jp/article/dsj/5/0/223/\\_pdf](http://www.jstage.jst.go.jp/article/dsj/5/0/223/_pdf).

Nelson, A., & Leclerc, G. (2005) A Spatial Model Of Accessibility: Linking Population And Infrastructure To Land Use Patterns In The Honduran Hillsides, in *Making Development Work: A New Role for Science*, edited by Hall, C. and Leclerc, G., New Mexico University Press, Albuquerque, NM, USA.

Pender, J. L., Scherr, S. J., & Durón, G. (1999) *Pathways of development in the hillsides of Honduras : causes and implications for agricultural production, poverty, and sustainable resource use*, International Food Policy Research Institute (IFPRI), Washington, D.C.

Sarkar, A. K., & Ghosh, D. (2000) Meeting the Accessibility Needs of the Rural Poor, *IASSI Quarterly* (Indian Association of Social Science Institutions), 18, 1-5.

Soares-Filho, B.S., D.C. Nepstad, L.M. Curran, G.C. Cerqueira, R.A.Garcia, C.A. Ramos, E. Voll, A. McDonald, P. Lefebvre, and P. Schlesinger. (2006). Modeling conservation in the Amazon basin. *Nature*, 440, 23 March 2006, doi:10.1038/nature043893

Souza Jr., C., A. Brandão Jr., A. Anderson, and A. Veríssimo. (2005). The Expansion of Unofficial Roads in the Brazilian Amazon. Available from [http://www.imazon.org.br/upload/ea\\_1e.pdf](http://www.imazon.org.br/upload/ea_1e.pdf)

Vera-Diaz, M.C., R.K. Kaufmann, D.C. Nepstad, and P. Schlesinger. (Forthcoming 2008). "An interdisciplinary model of soybean yield in the Amazon Basin: the climatic, edaphic, and economic determinants." *Ecological Economics* Available from [http://www.bu.edu/cees/people/faculty/kaufmann/documents/soybean\\_model.pdf](http://www.bu.edu/cees/people/faculty/kaufmann/documents/soybean_model.pdf)



Verburg, P. H., Overmars, K. P., & Witte, N. (2004) Accessibility and land-use patterns at the forest fringe in the northeastern part of the Philippines, *The Geographical Journal*, 170, 238-55.

Williams, R. (1987) Meningitis and unpaved roads, *Social Science & Medicine*, 24, 108-15.

Winograd, M., Farrow, A., & Nelson, A. (1999) Hurricane Mitch: What Happens When Natural Disaster Takes Out Your Transport System?, Accessed on 05/12/05, <http://worldbank.org/transport/expopres/winograd.ppt>

World Bank. (2002). Transport Sector Overview, Accessed on 30/11/2005, <http://www.worldbank.org/transport/whytsimp.htm>

World Bank. (2005). Roads & Highways, Accessed on 30/11/2005, <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTTRANSPORT/EXTROADSHIGHWAYS/0,,menuPK:338667~pagePK:149018~piPK:149093~theSitePK:338661,00.html>