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 domain 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 domain for use by the wide range of practitioners and researchers who need it. This document describes the demand for such a data set – especially to make progress towards the International Strategy for Disaster Reduction and the MDGs – and lays out a strategy for developing an initial product and then producing regular updates that accurately reflect the world's growing network of roads.

Table of Contents

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 domain 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.

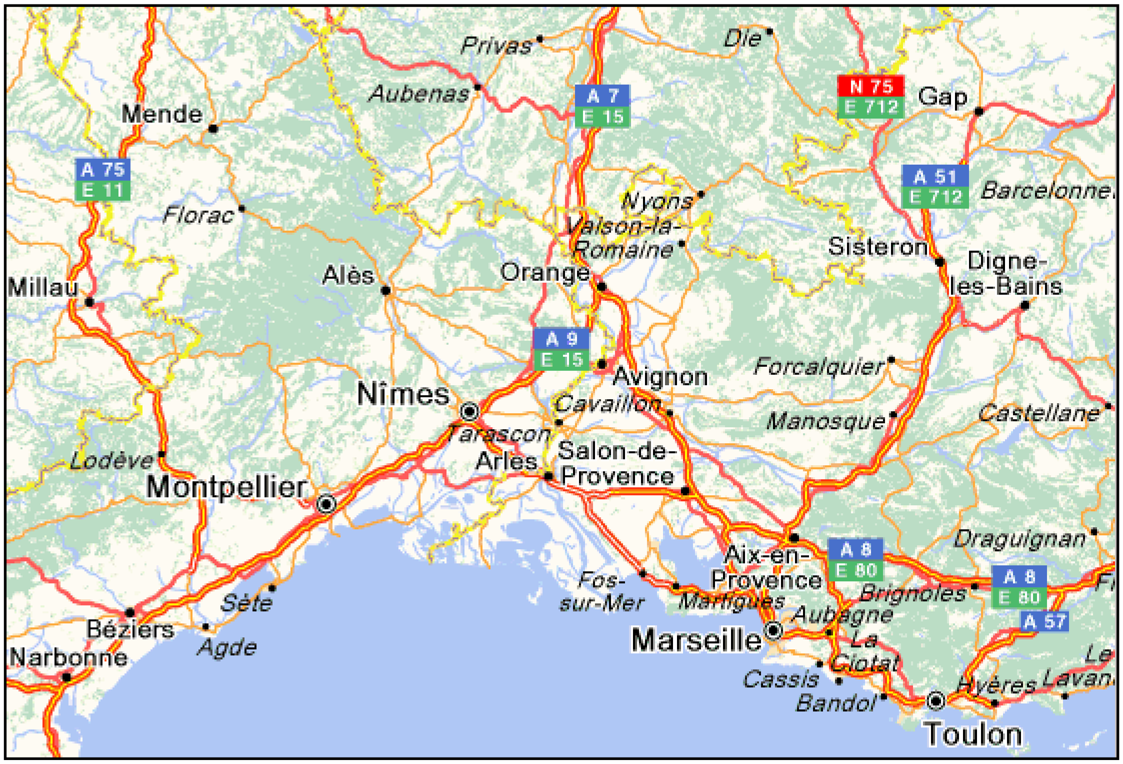
Among researchers and practitioners in the development, hazard response, biodiversity conservation, and public health communities there is high demand for a public domain[1] global spatial roads data set with improved geographic and temporal coverage, consistent coding of road types, and good documentation of sources (de Sherbinin and Chen, 2005; Nelson *et al.*, 2006). In order to address this issue, 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 (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 (see Section IV).

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. Our initial focus will be on developing countries, starting with Africa, then moving to South Asia, South and Southeast Asia, Oceania, and Latin America. But the aim is to have a consistent global coverage within two years of project inception.

---

[1] The term "public domain" is here construed to mean a data set that can be used free-of-charge by a range of users on an attribution-only basis, covered by a Creative Commons license.

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, we 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. Inputs using this online tool will be evaluated by a community of experts formed as a working group of the International Council for Science (ICSU) Committee on Data for Science and Technology (CODATA), 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.[2] 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 free of charge, on an attribution only basis, to the communities that most need it: 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; and national and regional agencies and organizations in the developing world. The needs of these communities are described in the next section.

[2] 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. We underscore 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*

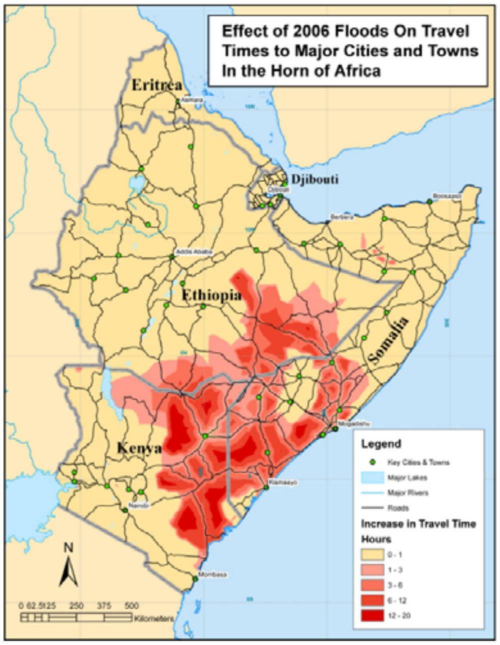
Following a loss of transport infrastructure, how can we ensure the best appropriation of often limited reconstruction resources? Who has been affected and where? Again, good spatial transport information is vital for rapid informed assessments for damage assessment, the loss of accessibility, the demographics 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



Produced by Paul Bartel, HIU, US State Dept.

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).

Distance (Km)/ TravelTime	Beira	Blantyre	Caia	Charre	Chemba	Chiramba	Chupanga	Guro	Lilongwe	Luabo	Manica	Maputo	
Beira		12h2'	4h28'	6h13'	7h43'	9h49'	6h26'	6h42'	15h18'	16h34'	4h12'	19h49'	8h4
Blantyre	567		7h41'	5h48'	8h14'	10h20'	10h50'	6h43'	5h10'	14h1'	10h36'	29h1'	13h
Caia	268	307		1h53'	3h22'	5h29'	3h16'	10h7'	12h37'	12h13'	7h37'	23h14'	5h3
Charre	337	230	77		2h26'	4h32'	5h2'	11h53'	10h44'	11h3'	9h23'	24h59'	7h1
Chemba	358	288	97	58		2h6'	6h31'	10h51'	13h10'	12h46'	10h52'	26h29'	8h4
Chiramba	400	330	139	100	42		8h38'	8h46'	15h16'	14h52'	12h38'	28h35'	10h
Chupanga	265	346	47	116	136	179		12h6'	15h46'	15h22'	9h35'	25h12'	2h1
Guro	403	369	608	677	232	190	604		8h36'	20h30'	3h52'	22h17'	14h
Lilongwe	919	306	608	531	589	631	647	516		18h57'	12h28'	30h53'	18h
Luabo	534	628	273	226	255	297	312	993	929		19h43'	35h20'	17h
Manica	252	601	458	527	547	422	454	232	748	723		19h47'	11h
Maputo	1189	1707	1395	1463	1484	1526	1391	1338	1854	1660	1188		27h
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	

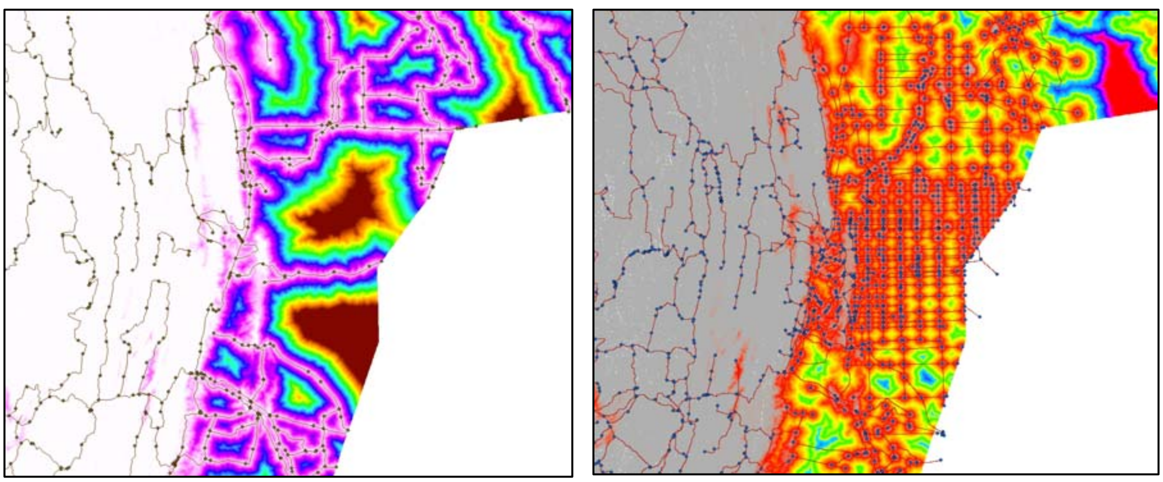
Produced by the UN Joint Logistics Centre, February 2007

• *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 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, and the map at **right** represents an accessibility map for the same region based on high resolution/high quality roads data. 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 germaplasm. 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*

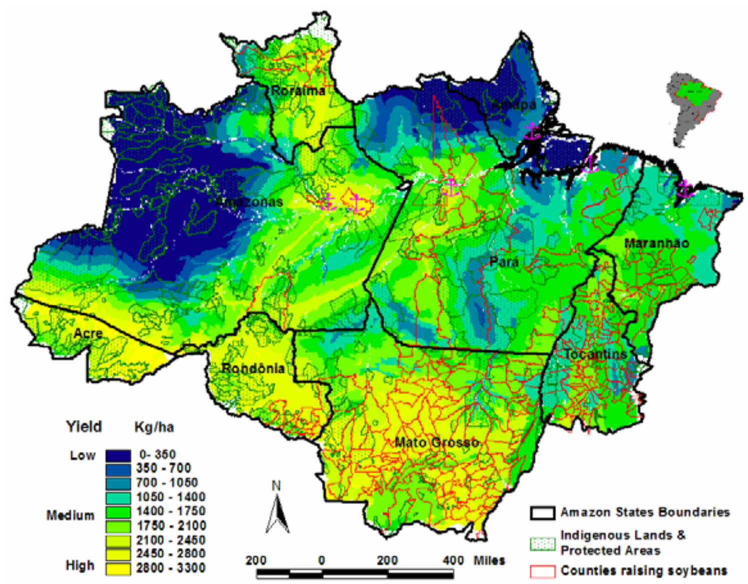
Research interest in the linkages between access (and related notions of remoteness and isolation) and rural livelihoods has spawned a growing literature on the characterization and empirical measurement of “access” and its welfare implications; including impacts on land use, enterprise mix, technology 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).



Map courtesy of Keith Alger, Conservation International

- *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.

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 (see Section IV on project management). 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.

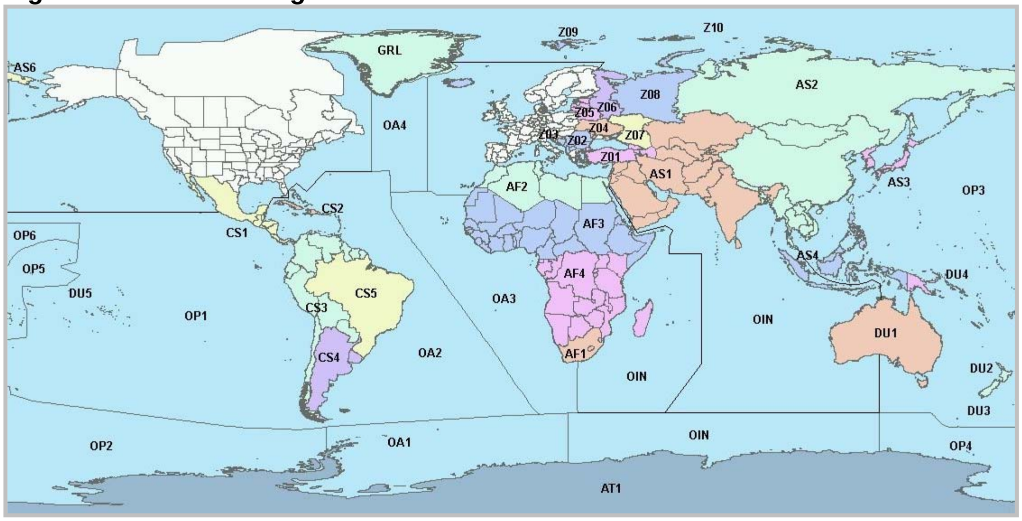
Private Sector Roads Data

Two major companies – Navteq and Tele Atlas – currently have well developed global roads databases that are primarily used for automobile navigation systems, 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 our proposed data set. 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 in Section I.B 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.

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 EU 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 future development of a global roads data set (see Section III.B below), the approach is unlikely to develop data for large parts of the developing world apart from urban areas and where wealthy tourists travel. Unlike the private sector and military-intelligence communities, the constituency for OpenStreetMap is highly committed to free and open access to data.

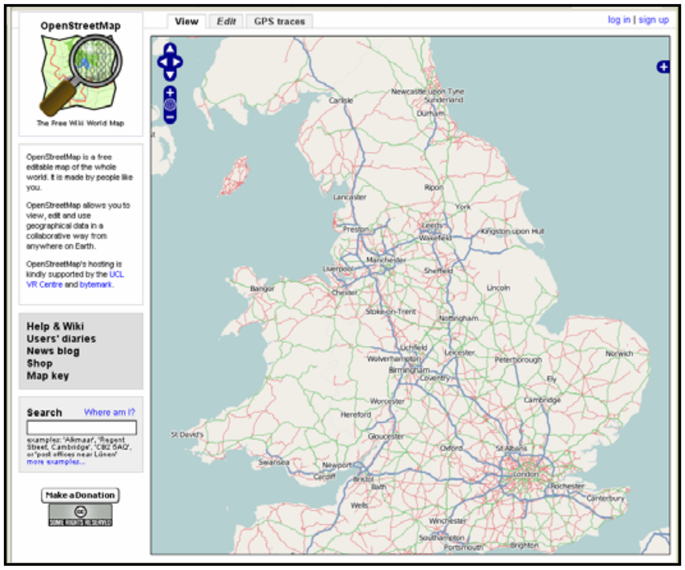


Figure 3. OpenStreetMap: Developed by a community of users in a WIKI environment, OSM represents an important approach to the current and future development of global roads data.

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.[3] 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

[3] See www.ungiwg.org and www.unjlc.org/mapcenter/unsdi

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 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© Shape Files or other common vector data exchange formats such as ESRI ARC/INFO coverages and VPF.

III. Methods

A. Base map development

A combination of the following data sets will be used to manually digitize roads and attributes according to the data model described above.

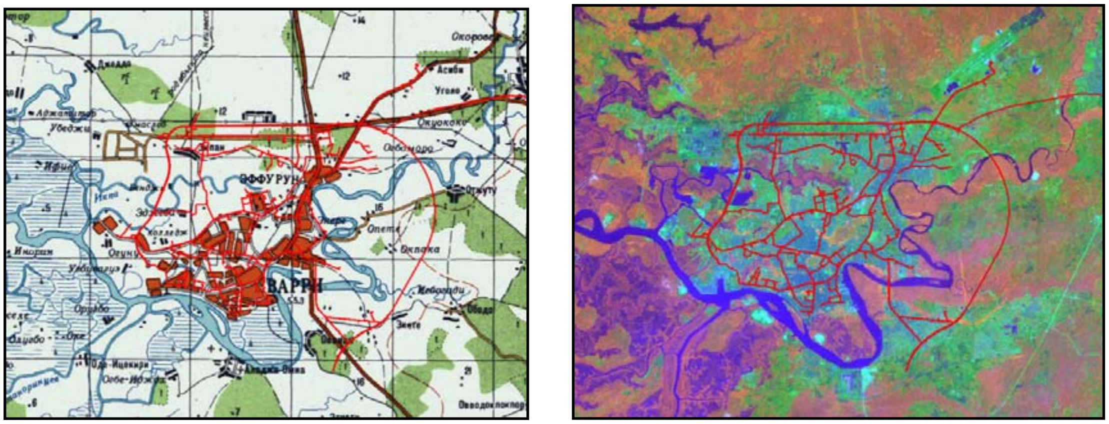
- Scanned 1:200,000 paper maps developed by the Russian military (ranging in dates from the late 1960s to the early 1980s) and the US Joint Operations Graphic (JOG) navigation maps.[4]
- Geocover[5] Landsat pansharpened 15m imagery baselined to the year 2000, which are orthorectified and are available free of charge.[6]
- GPS tracks wherever available to add the most recent routes.[7]

Roads will be manually digitized and attributes assigned according to the data model above. While digitizing, rail road networks, populated places and built up areas will be added for the sake of complete coverage of the node to node transport network in a

[4] The Russian military maps and UN JOG maps covering many parts of the world are available through UN agencies.
[5] A global mosaic is available at a cost of $1,500 from a Canadian company. The Global Land Cover Facility has the data available free of charge.
[6] Orthorectified ASTER data may be available.
[7] GPS tracks will be obtained from UN agencies or companies such as Tracks4Africa.

country. 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.

The consortium will assess the cheapest ways of obtaining this data, whether through so-called "click workers" in digitizing shops in India or elsewhere, or through a sub-contract with a private company such as Georigin, Ltd., in South Africa.[8] It may be possible to develop an online software application similar to the one employed by OSM. For example, an ESRI ArcIMS custom application could be developed that would allow on-screen digitizing using the above combination of data sets with different opacity levels. If this tool can 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 (see Section IV). 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.

It must be acknowledged in advance that not all fields in the data model can be filled out based on remote digitization. For instance, road name, road and lane width, weight

[8] A preliminary estimate by Georigin 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.

restrictions, and road surface material or condition may not be easily obtained from either paper, image, or GPS sources.[9] Field validation for further data development will be a critical component of building the depth of this data set. This will be accomplished in collaboration with UN operational agencies that have an interest in seeing this data set developed for their own operational purposes.

For developed countries, we will insert Tiger data for the entire United States, and will identify other public sources of high resolution roads data for Canada, Europe, Japan, and Australia. Substantial portions of Europe may be available from OSM with the next two years. These data sets will require varying degrees of validation.

B. Future updates

From the baseline data set, we would then develop a software tool that would allow crowd sourcing of data on roads, using the OSM model of providing up-to-date orthorectified satellite data and allowing users to upload GPS tracks so that they can then be digitized. We use the term “crowd sourcing” 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.[10] If possible, this would be built on the same “OSM type” tool – only faster and more robust given that it would need to service far more users – that is described in the previous section.

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.[11] 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.

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 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

[9] Road condition can be derived from GPS tracks if one assumes that condition is directly related to travel speed.

[10] 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.

[11] The OSM WIKI allows users to enter basic attribute information, but does not currently support relational tables.

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).

IV. Project Management

Coordination will be provided by CIESIN with guidance from newly established 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 an initial list of members). CIESIN will investigate the most inexpensive means of developing the data, by either outsourcing to one or more private companies, or through distributed regional expert networks.[12] Work would be conducted to the established specifications.

Over time, this initiative may evolve to a community approach but 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 using a Creative Commons license through CIESIN servers, and on the servers of selected partners.

[12] For example, by using "click workers"; individuals who are paid for piece work. Inputs from three or more workers would be used to establish an averaged product.

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.

Name	Organization
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 Czaran	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 Praticability	RdPracticability	0	true	0
3	SeasonalPrac	Integer	4	SeasonalPrac	Difficult Season Road Praticability	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.

DomainName	RdClass	
DomainType	CodedValue	
FieldType	Integer	
MergePolicy	DefaultValue	
SplitPolicy	DefaultValue	
Description	null	
Owner	null	
Coded Values		
Name	**Code**	
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 have agreed to be members of the CODATA Working Group on Global Roads Data Development. Additional members may be added at a future date.

Name	**Organization**
Olivier Cottray (co-chair)	UN Joint Logistics Centre
Alex de Sherbinin (co-chair)	Columbia University - CIESIN
John Dann	Georigin, Ltd.
Johann Groenewald	Tracks4Africa
Timothy Haithcoat	Geographic Resources Center, University of Missouri
Glenn Hyman	CGIAR-CIAT
Andrew Nelson	EC Joint Research Centre
Harlan Onsrud	Dept of Spatial Info. Science, University of Maine
Steve Coast	OpenStreetmap

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

Hope, R. A. (2006) Evaluating water policy scenarios against the priorities of the rural poor, World Development. *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 Domain 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.

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, Alburquerque, 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

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

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