Methods Used in the Development of the Global Roads Open Access Data Set (gROADS), Version 1

Table of Contents

Introduction

The Global Roads Open Access Data Set, Version 1 (gROADSv1) was developed under the auspices of the CODATA Global Roads Data Development Task Group. CODATA is the Committee on Data for Science and Technology of the International Committee on Science (ICSU). The data set combines the best available public domain roads data by country into a global roads coverage, using the UN Spatial Data Infrastructure Transport (UNSDI-T) version 2 as a common data model. All country road networks have been joined topologically at the borders, and many countries have been edited for internal topology. Source data as a percentage of the total road network for each country are provided in the documentation, and users are encouraged to refer to the gROADS data set documentation for use constraints that apply to a small number of countries. Because the data are compiled from multiple sources, the dates for road network representations ranges from the 1980s to 2010 depending on the country (most countries have no confirmed date), and spatial accuracy varies. The baseline global data set was compiled by the Information Technology Outreach Services (ITOS) of the University of Georgia's Carl Vinson Institute of Government as a contribution to the Geographic Information Support Team (GIST) with funding from the U.S. Agency for International Development (USAID). Updated data for 27 countries and 6 smaller geographic entities were assembled by Columbia University's Center for International Earth Science Information Network (CIESIN), with a focus largely on developing countries with the poorest data coverage. Additional data were downloaded through the International Steering Committee for Global Mapping (ISCGM) website.

This document describes the methods used to develop the gROADSv1 data set. The document was compiled by a team including Steffen Foerster (team lead), Paola Kim-Blanco, and Alex de Sherbinin, chair of the CODATA Global Roads Data Development Task Group. We are grateful to members of the CODATA Task Group who provided suggestions and comments on the methods used to develop the gROADS data set along the way.

Data Assessment and Selection

To assess data quality and select the best data sets for inclusion in gROADS, we assessed spatial accuracy using root mean square error (RMSE), and relative completeness by calculating the total road length for each data set.

Root Mean Square Error (RMSE)

The RMSE is a commonly accepted measure of accuracy for comparing two vector data sets. As a reference layer, we used GPS track data where available (e.g. from Tracks4Africa), and where GPS data weren't available, we used high resolution imagery in Google Earth (GE), which is considered "ground truth" for our purposes (see Potere 2008). The following illustrates the steps taken to compare a sample data set to a reference data set.

To choose tie points for geo-referencing in ArcMap, first we generated a fishnet covering the globe, and centroids for each grid cell. Depending on the spatial extent of the data set, the grid cell size was either 0.5 or 1 degree. To avoid geographic bias in selection of tie points used to calculate RMSE, we used the freely available Hawths Analysis Tools for ArcGIS[1] to generate a random selection of 10 centroids per data set. We then looked for tie points in the grid cells associated with those selected centroids.

---

[1] Hawth's Anaysis Toolset was formally discontinued in Dec 2009. The most recent version is still available, but will only work with ArcGIS 9.2. The toolset has been replaced by new software called "Geospatial Modelling Environment". Both are available at http://www.spatialecology.com.

When more than one data set was available for a given country, we used the same set of randomly selected centroids to assess spatial accuracy for all of the data sets. For every grid cell associated with a selected centroid, we selected a clearly identifiable route intersection the shortest distance from the centroid (Figure 1). When the identity of the intersection across data sets was uncertain due to large deviances in the vector segments, we inspected the attribute data to assist in matching locations. When there was no intersection within the grid cell associated with the centroid, we tried to identify a distinct feature (e.g. a jog in the road) in a road segment that could be matched between data sets. If even that failed, or if there were simply no roads within the grid cell, we used the next usable road intersection or segment regardless of its distance to the centroid of that grid cell. If the next identifiable tie point was the closest tie point to a neighboring grid cell centroid, we chose the second closest tie point available. For each data set, a minimum of 10 tie points were used to calculate the RMSE. If GPSed data were not available, road layers were converted to KML and imported into Google Earth (see next section). The tie points were measured as the distance between the selected feature (e.g., intersection, jog in the road) and the same feature in the GPSed data set or imagery. The

Figure 1: Example of a half degree grid cell, its center point (centroid) and the location of the nearest junction common to both road layers

distance between tie points was measured in meters, and the offset between the road layer feature and Google Earth imagery was entered in the spreadsheet. The RMSE calculation in Excel is as follows:

=SQRT(SUM(G2^2,H2^2,I2^2,J2^2,K2^2,L2^2,M2^2, N2^2,O2^2,P2^2)/COUNT(G2:K2))

Where G2-P2 represent the cells with the offsets measured in meters. This process is covered in greater detail in the next section.
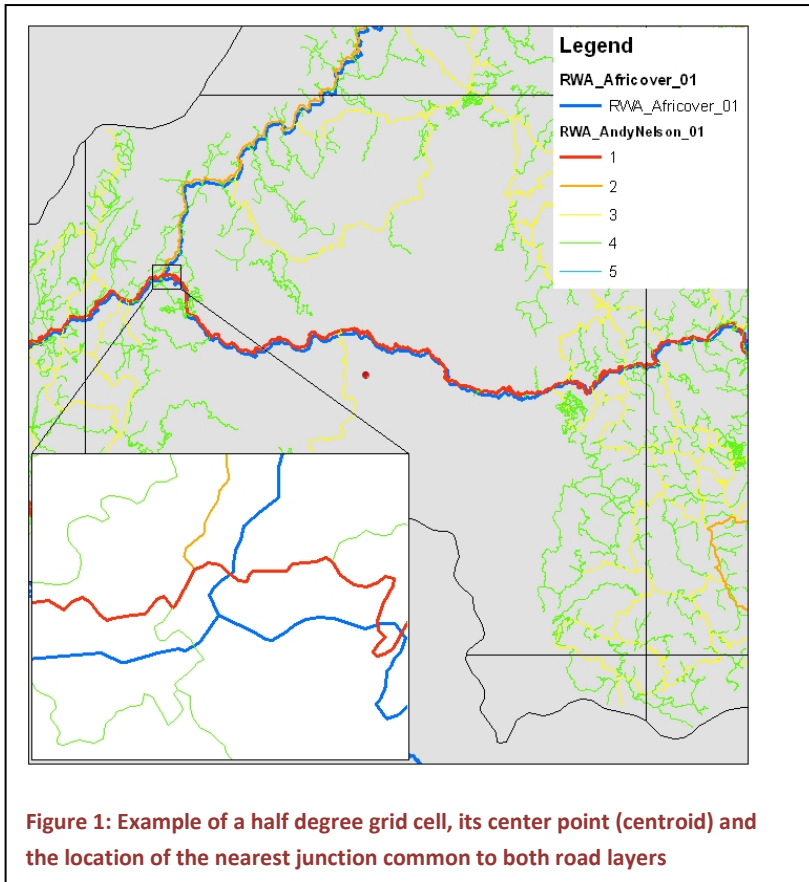
Validation against Google Earth (GE) imagery

Road layers, grid boundaries and centroids were converted from ArcGIS shapefile format to KMZ/KML file format and opened in GE version 4.3 [2]. The accuracy of the test road layer was assessed as described above, looking for an intersection closest to the centroid of a grid cell. However, roads are only clearly visible in high resolution image tiles, so when a tie point fell on a low resolution raster where roads were not clearly visible, we took the next available tie point located on a high resolution raster. Distance between tie points of the reference and sample data layer were measured with the GE line measurement tool in meters.
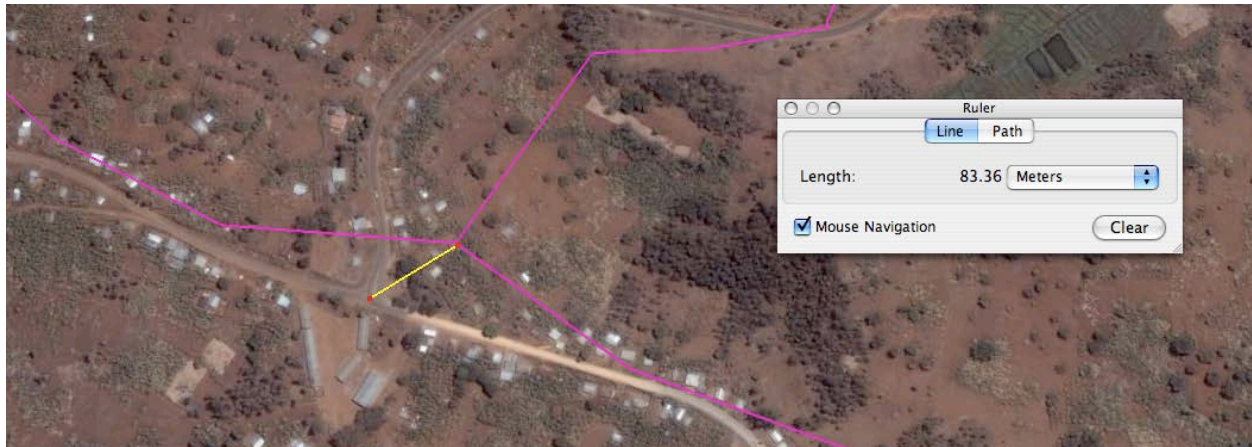
Figure 2: Example tie point for RMSE calculation using the RWA_Africover_01 road layer and GE imagery. Yellow line connects tie points.
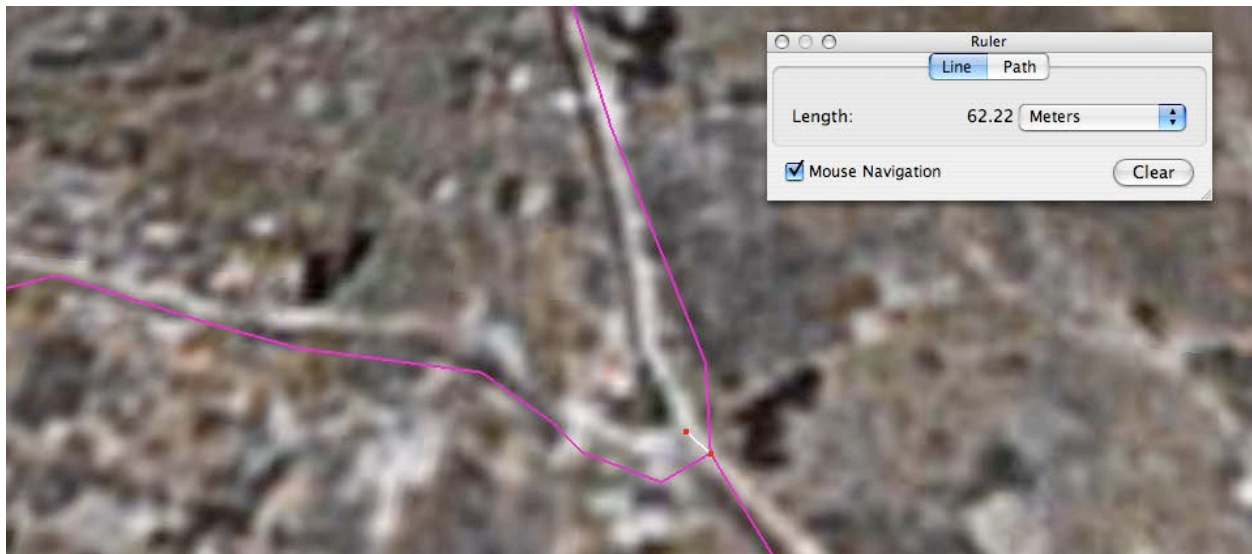
Figure 3: Example of a low resolution raster in GE and the RWA_Africover_01 road layer. Intersection is detectable in the image although exact location is difficult to pinpoint.

---

[2] Later versions can be used. Conversion to KML is done using the KML conversion tools in ArcToolBox for ArcGIS 10.x, or XTools extension in previous versions of ArcGIS. Google Earth high resolution imagery compared to Geocover (which itself has an absolute positional accuracy of about 50m) had an RMSE of about 50m (Potere, "Horizontal Positional Accuracy of Google Earth's High-Resolution Imagery Archive" in *Sensors* 2008, 8, 7973-7981).

Figure 4: Example tie point between GE imagery and RWA_AndyNelson_01 road layer. Yellow line connects tie points. Pink line is RWA_Africover_01 road layer.
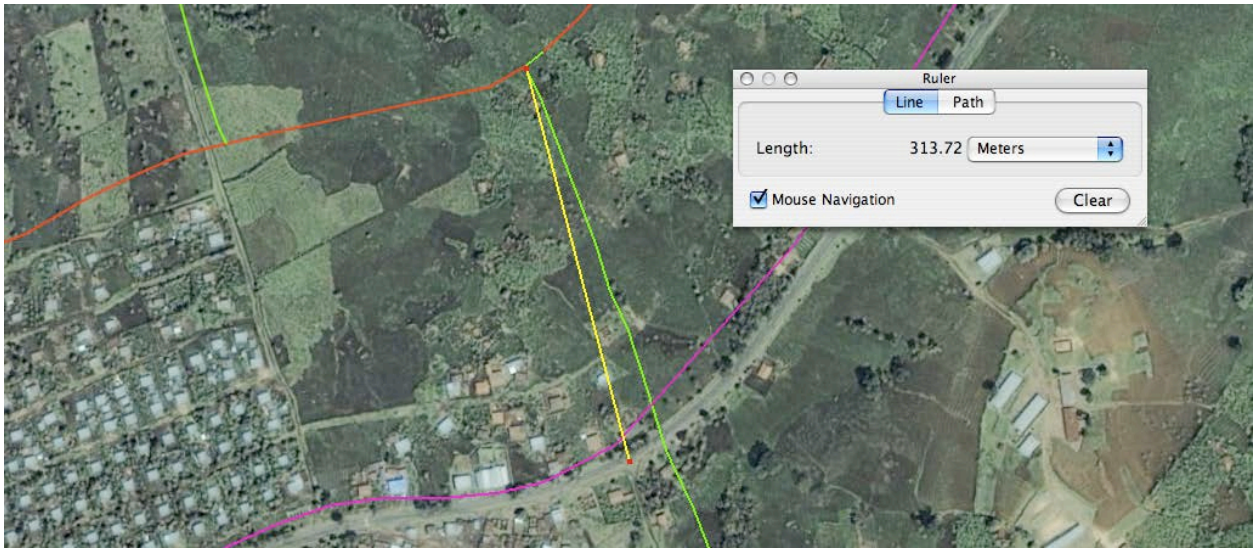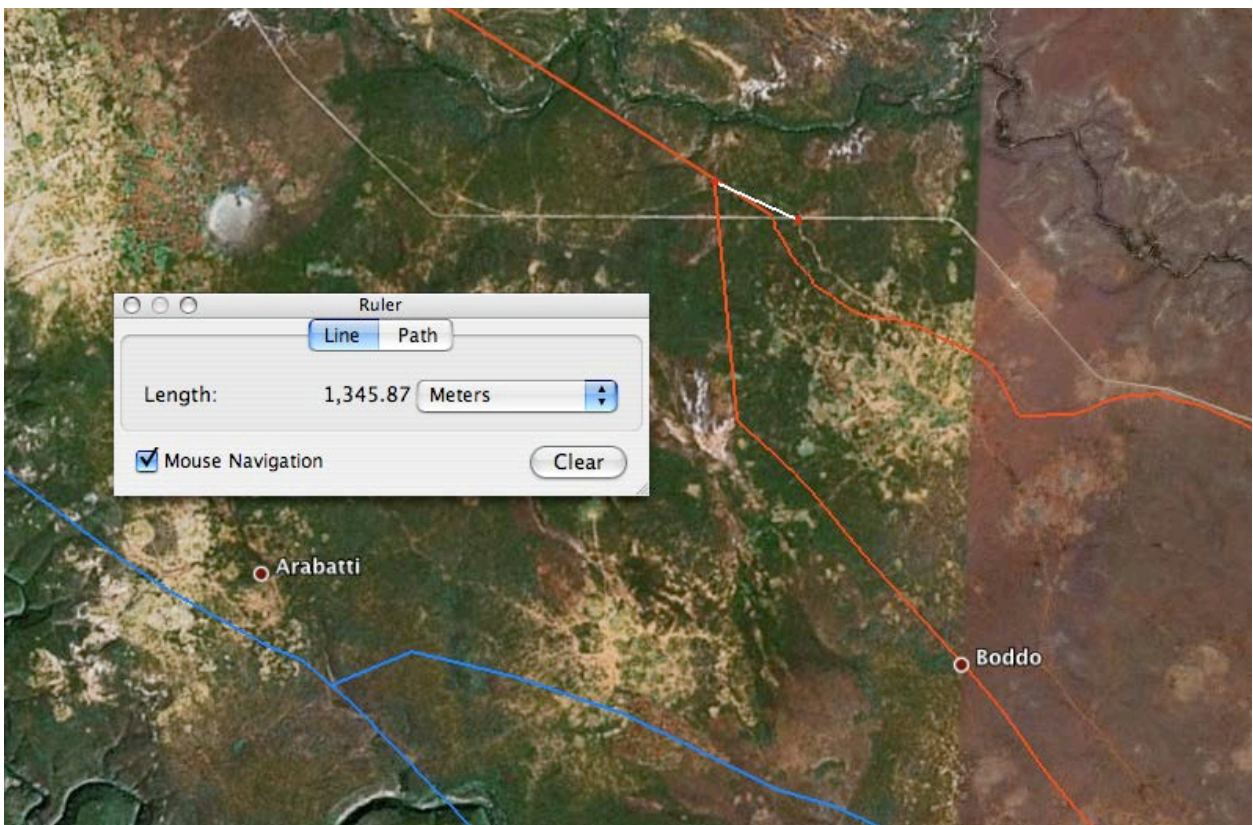

Figure 5: Example tie point between GE imagery and ETH_CGIAR_01 road layer (red line). Blue line is ETH_ISCGM_01 road layer. Note the large deviance for the latter from what seems to be the same junction.
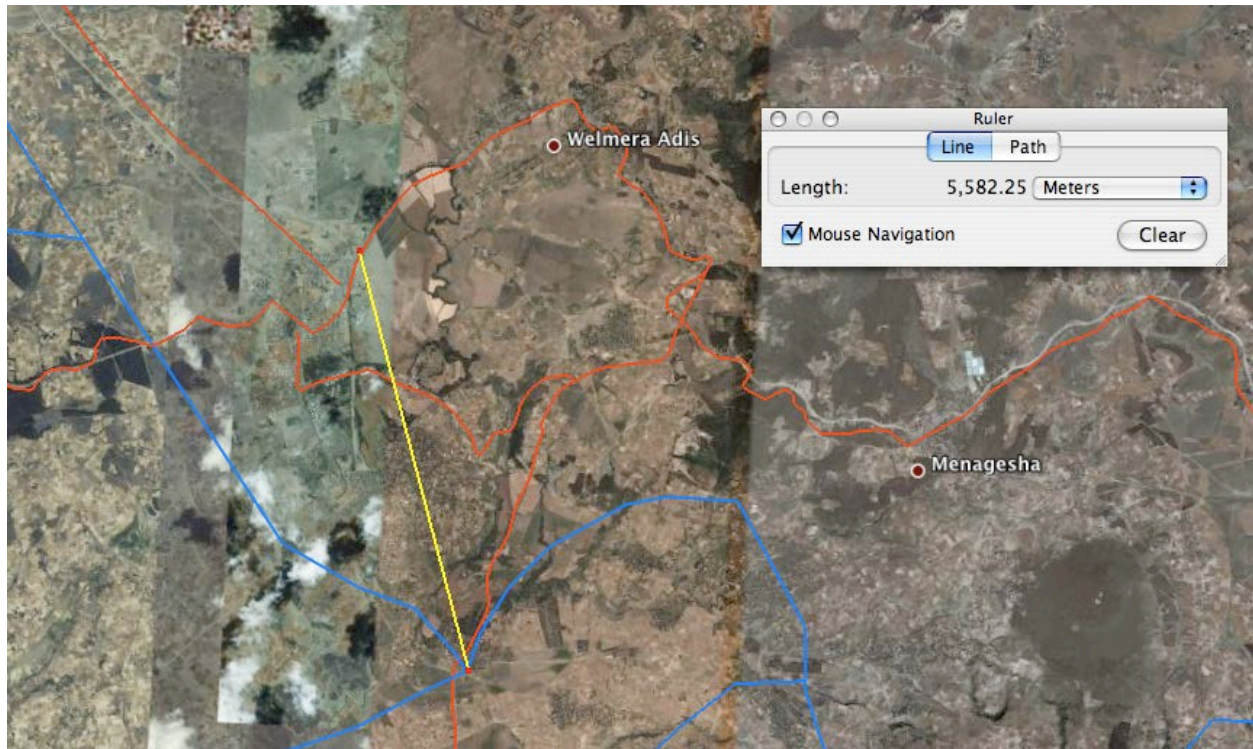
Figure 6: Example tie point between GE imagery and ETH_ISCGM_01 road layer (blue line).

Calculating total length of roads

Many data sets do not have a road length field in kilometers that permits easy calculation of total road length. There is a tool to calculate the length of each segment in a shapefile or geodatabase feature class with ArcGIS10.X.

Instructions:

1. In ArcCatalog, preview the shapefile or feature class table and add an attribute field, "Length_KM" of type double.

2. In ArcToolBox, run the "Calculate Fields" tool (Data Management Toolbox).

   a. Specify the shapefile (or feature class) as the "Input Table".

   b. Choose "Length_KM" from the "Field Name" menu.

   c. Type "!shape.length@kilometers!" in the "Expression" field.

   d. Choose "PYTHON_9.3" for "Expression Type".

   e. Click "OK".

3. The output length in kilometers will be stored in the "Length_KM" field of the attribute table of the shapefile or feature class.

At an earlier stage in the project we used an AML script (courtesy of Andrew Nelson, International Rice Research Institute) to do this. The workflow is provided here for those who are interested.

6

Before the script could be run, each data set had to be converted to ESRI coverage. The conversion can be done from ArcCatalog by right-clicking on the data set and chooseing EXPORT → Coverage… To ease the running of the script, all coverages created for the purpose of road length calculation were stored in one folder.

Instructions:

1. Before we ran the script for the first time, we did the following:

    a. Copied the length2km.exe and length2km.aml files to our local AML directory of our ArcGIS installation. (Normally, this would be C:\arcgis\arcexe9x\atool\arc\)

    b. Opened the AML file in notepad, making sure that the path on line 7 pointing to the exe file is correct

2. Copied the coverages that we wanted to assess to the local hard drive and noted their location

3. Ran ArcInfo

4. Changed the workspace to the folder where coverages were located. For example, if your files are stored in c:\roadlength\, at the command prompt, type WORKSPACE C:\ROADLENGTH\

5. Ignore any warnings that the new location is not a workspace. You can test if your workspace has been accurately set with the command WORKSPACE

6. To invoke the script, type LENGTH2KM <coverage name>, and hit ENTER

7. The output length in kilometer will be stored in a DBF file under the same location as the source coverage.

Notes: The script only works correctly on coverages that are in lat long coordinate system and not projected. If we had a data set in a projected coordinate system, we were sometimes able to calculate road length in ArcMap if the attribute information contains a SHAPE LENGTH field. If so, right-click on the column header and choose STATISTICS…, and then read the total length in the SUM field of the statistics window. In the example used for the screenshot, the data set was in UTM projection, and thus the resulting total length is given in meters instead of kilometers.
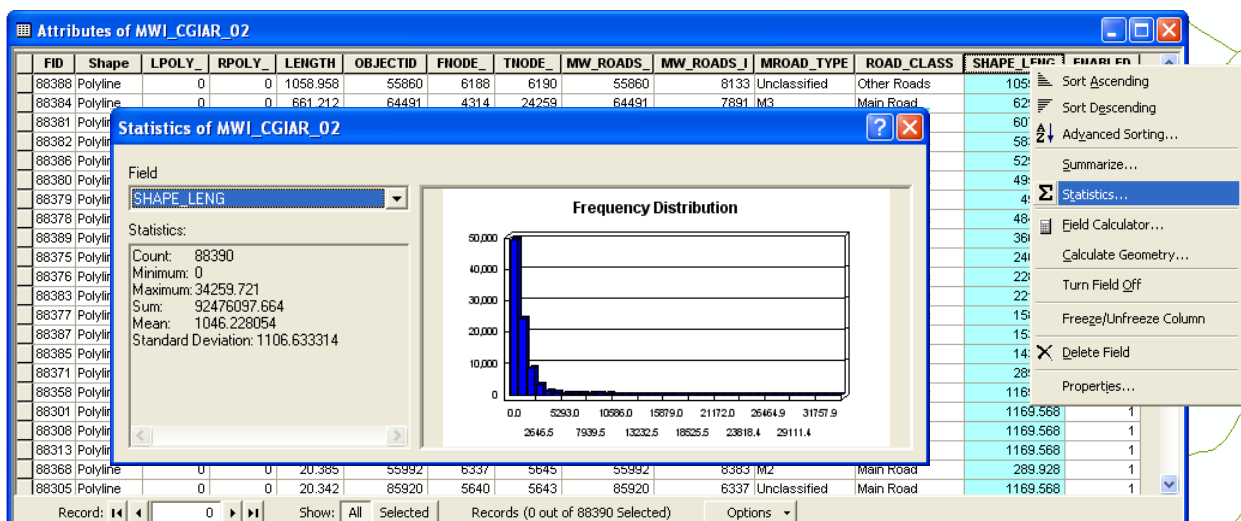
Choosing a data set for further edits

Our choice of which data sets to include in the final gROADSv1 data set was based on a set of criteria and assessments. As an initial check, we compared our data sets with other global road databases to see if a more complete data set exists that we might be able to acquire, so as not to start the extensive editing and cleaning process on a data set that is inferior to other potentially available data sets.

After we established that one of our existing data sets is likely a best candidate for inclusion in the catalog, we followed the remaining criteria in the tracking sheet. While taking into account potential licensing issues, we attempted to identify a best compromise by ranking candidate data sets on the following 4 criteria, sorted by their relative importance:

 a) extent of coverage,

 b) accuracy,

 c) completeness of attribute information

 d) quality of topology

We used the calculated total length of roads in kilometers and RMSE for the first two criteria, respectively, and we used ArcMap to check attribute information and assess data sets for topology errors.

Data ingest: Creating and populating geodatabases

Data sets were often received in shapefile format. Once a dataset was selected for use in gROADS, we stored the data in a geodatabase. The following steps outline the creation of a geodatabase and the import of shapefiles into the geodatabase.

1. Open ArcCatalog and navigate to the appropriate location in the editing folder
2. If it does not already exist, create a new country folder in the appropriate region, using the 3 letter ISO code of the country
3. Right-click on folder and choose NEW → FILE GEODATABASE
4. Name the geodatabase with the country ISO code
5. Right-click the geodatabase and choose NEW → FEATURE DATA SET…
6. Enter a name for the feature data set, using its [Source Name]. Click NEXT
7. Choose a coordinate system for the data set. Select Geographic COORDINATE SYSTEMS → WORLD → WGS 1984. Click NEXT
8. Ignore setting of the vertical coordinate system, i.e. accept default selection of <NONE>. Click NEXT
9. Accept default settings for XY Tolerance. Click FINISH
10. Right-click on the feature data set just created; choose IMPORT → FEATURE CLASS (SINGLE)…
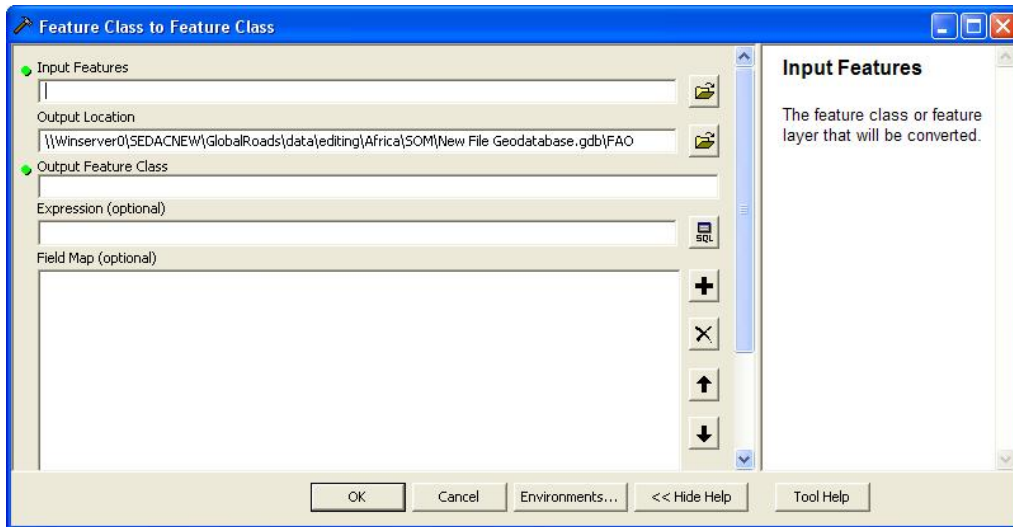
11. Select input feature, navigate to original shapefile.
12. Enter a name for the Output Feature Class. Click OK. You are done!

Data editing

Data editing is a time consuming task for any data product, and no less so for road networks. This section primarily concerns creating a consistent topological network where connectivity problems are corrected and the road network connects together seamlessly. In gROADSv1, topology was always corrected at borders between countries, and internal topology has been edited for a number of the new data sets added by CIESIN to the ITOS/GIST baseline data set (see the Introduction for explanations regarding the baseline versus the additional data). This proved to be a very time consuming process. As the CODATA Global Roads Data Development project progressed it was determined that internal topology would only be edited if the roads data were of high spatial accuracy and completeness (e.g., derived from high resolution imagery of GPS tracks). Even in these cases the cost was quite high. The level of effort required to edit the topology of a GPS-derived data set for Kenya, which had a very dense coverage, was two-person months (see Case Study: Kenya below). gROADSv1 users will need to visually assess the data or run tests described below to determine if the topology has been edited properly in the country or region for which they're interested.

Creating a topology

1. In ArcCatalog, right-click on the **Feature Dataset** inside the file geodatabase (not the Feature Class)

2. Choose NEW → TOPOLOGY

3. The NEW TOPOLOGY wizard will open. Click NEXT
   a. A feature dataset may have multiple topologies, but you can only create one topology for a given feature class in a feature dataset. If you try to create an additional topology for the same feature dataset, make sure that at least one of the feature classes in it are not already participating in an

9

existing topology, or you will receive the following error message:
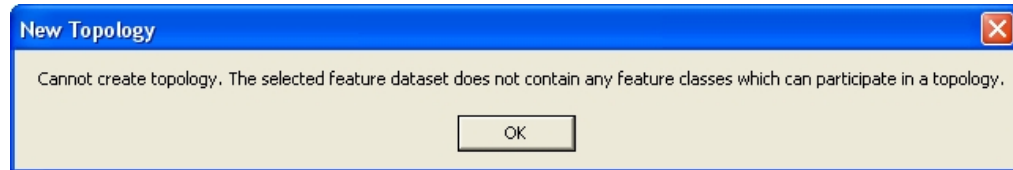
4. Enter a name for your topology. Maintain an appropriate identifier that will allow you to distinguish different topologies, i.e. keep the name of the source and the file name, and add Topology at the end, e.g. KEN_FAO_IGAD_LPI_01_Topology. Accept the default for cluster tolerance and click NEXT.
5. Select the feature classes to participate in the topology. Click NEXT
6. Accept defaults for rank numbers. Click NEXT
7. Now add topology rules one by one. Click ADD RULE…
8. Choose at least the following rules:
    a. Must not have dangles.
   i. This rule identifies any endpoints of road segments that do not meet another road segment, for example because they overshoot or stop short of another road segment (common with GPS track derived layers) that they are supposed to intersect with. However, many of these endpoints may be valid endpoints of roads not connecting to any other roads, making it necessary to carefully validate all violations turned up by this topology rule (see more details below)
    b. Must not self-overlap & Must not self-intersect
        i. Identifies artifacts in the network where road segments overlap or intersect themselves, a common problem in GPS derived data sets. Usually, those violations will need to be manually edited by removing excess vertices from the arcs.
    c. Must not intersect or touch interior.
        i. Assures that roads do not cross or meet other roads without the presence of an intersection, i.e. there need to be separate road segments on each side of an intersection.
9. After all topology rules are added, click NEXT, then click FINISH
10. Once finished creating the topology, ArcCatalog will ask if you would like to validate it. Click YES. This process can take a while, depending on the size of the data set

Editing topology in ArcMap

1. Open ArcMap, create a new empty map, save it in the folder where you stored the file geodatabase, and name it with the country ISO code
2. Click Add Data
3. Navigate to the topology within the feature data set of your file geodatabase
4. When asked if you want to import all feature classes participating in the topology, say YES
5. The topology will appear as the top-most layer in your Table of Contents
    a. By default, all types of errors are highlighted (area errors, line errors, point errors). When you are dealing with a large number of errors of different types, it may be helpful to switch off one or more of the error types. Go to Layer Properties → Symbology, and switch off the error types that you want to hide from display
6. If not already activated, activate both the EDITOR and TOPOLOGY toolbars (View → Toolbars)

7. Start an editing session (Editor → Start Editing), select the feature class to be edited (i.e. the file on which the topology was based).

8. Activate the error inspector in the Topology toolbar

9. The error inspector window will open, making it easier to select specific types of topology errors, move between violations easily, and edit multiple violations at once if adequate
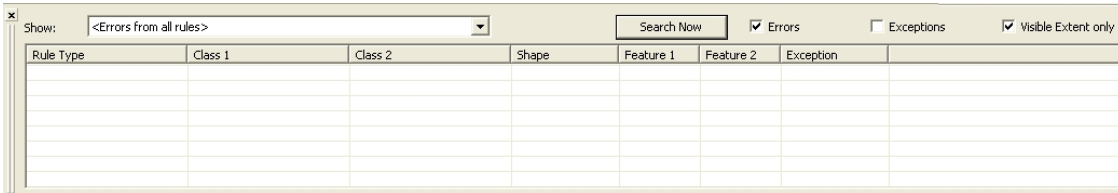
10. In the drop-down menu SHOW: you can select the type of error to search for
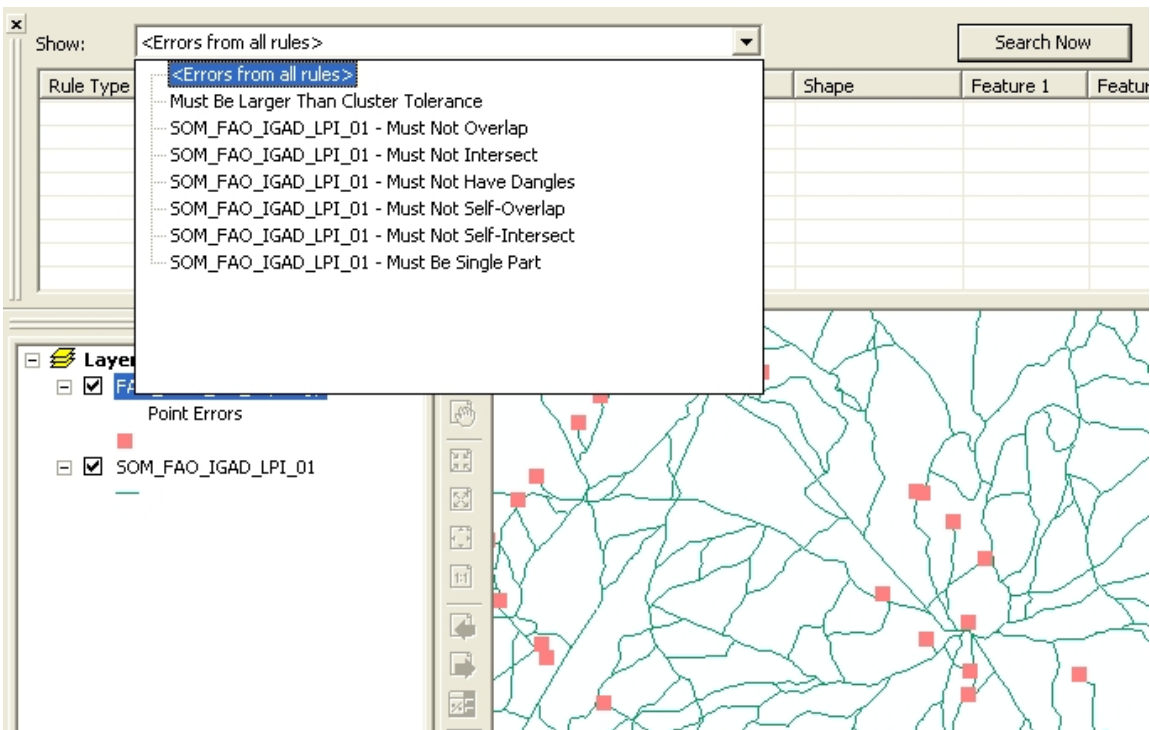
11. Zoom in to a section of the map where you want to start editing. Activate the checkbox "Visible Extent Only" in the Error Inspector window to search for violations in the visible section of the map.
    a. Although not necessary, keeping this checkbox checked makes it easier to jump from error to error, as you will stay in the same general area on the map that you had in view when you searched for errors. The error inspector will move from error to error starting on the bottom right corner of the view extent, moving left and upwards.

12. Now is a good time to open Google Earth (GE) alongside ArcMap, as you will be cross checking with GE imagery to decide on the best way to fix a given topology error, if necessary. For getting your road layer into GE, use the conversion tool in ArcToolbox (Conversion Tools → To KML). Note that only visible layers will be exported as KML, and that they will be exported exactly as they appear in ArcMap. Load the exported KML file into GE and zoom in on the section where you are about to edit the topology.

13. Click on the first violation that appears in the error inspector list. Hitting Z on the keyboard will zoom in to the selected error.

a. Note that for many error types you will have to further zoom in to inspect and correct the error, because it may not be recognizable at this zoom level

14. Compare with GE imagery to assess the error and to identify the best way to fix it. **(General guidelines on fixing topology errors are found in the section that follows.)**

15. When you are ready to make a decision on the selected error, right-click on the error in the list. A set of pre-defined options will appear for how the particular error type can be fixed.
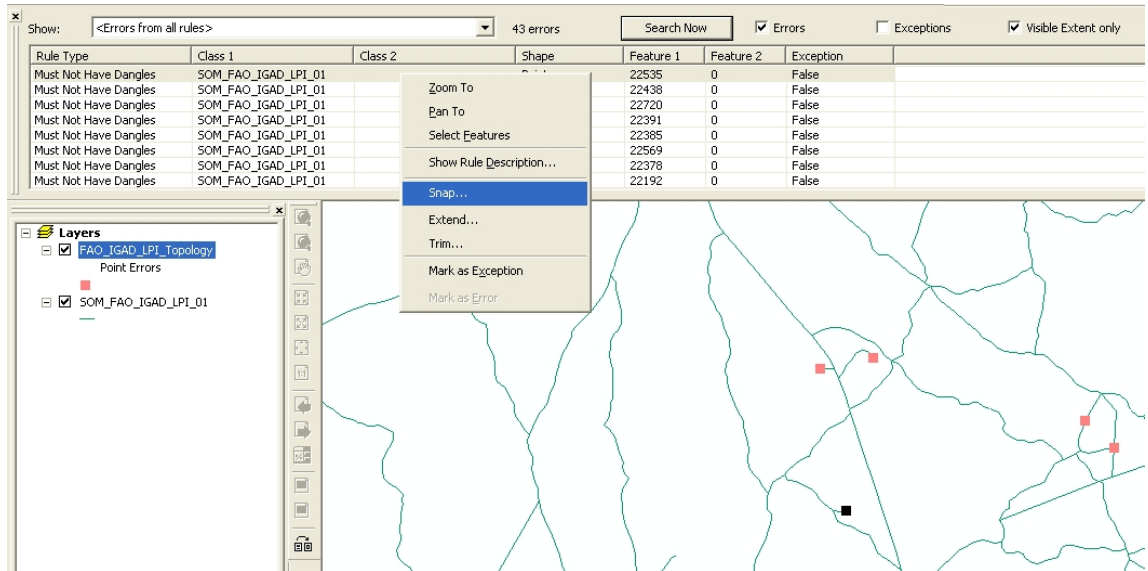
a. In the screenshot above, the error type is "Must not have dangles", for which several options exist: SNAP, EXTEND, or TRIM. Each of these requires entering a threshold distance in degrees below which to attach the dangle to. If the threshold distance is set too small, the tool may not be able to fix the error, if the distance is set to high, the connection may go to an unexpected location. Generally, when connecting to a vertex straight ahead of the dangle, EXTEND will perform better, while SNAP can be useful when connecting to a nearby vertex that is not directly ahead of the dangle. TRIM is less sensitive to threshold values, as it simply trims down vertices until it meets another arc, as long as the arc is within the threshold distance. Suggested starting values are:

   i. SNAP: 0.0001
   ii. EXTEND: 0.003
   iii. TRIM: 0.01

b. For all error types, "Mark as Exception" is available as an option to ignore the error and accept it without modifications into the final data set.

16. Note that you can only undo the last command in the topology editing tool, i.e. if you realize later that you made a mistake in fixing a violation, you will have to manually locate and edit the vertices using the regular editing tools.

17. Save your edits regularly. Note that after saving, the list of errors in the Error Inspector will be cleared, so you will lose focus on the larger section of the map that you used when searching for errors.

After having fixed all errors, re-validate your topology using the commands available in the Topology toolbar. You may get additional violations that you introduced in the process of editing the topology. For example, when snapping a dangling vertex to another road segment, the arc that you snapped to needs to be split at the intersection, or otherwise that will violate the "Must not intersect or touch interior" topology rule.

Decision rules for editing

This section addresses a number of errors encountered in data sets, and rules for how they have been addressed. It also provides some case studies addressing specific issues, including merging multiple roads data sets (Tanzania), editing detailed GPS data (Kenya), and simplifying OpenStreetMap data for use in gROADS (Canada).

Common network errors

1) Gaps in the road network as a result of unconnected segments. These gaps can be a result of two segments stopping short of each other, or one segment shooting over the other instead of connecting to a junction. Both segments have complete attribute data.
   a. Response: Segments were connected by adding one or more vertices to one of the segments, snapping to the end of the connecting segment. If both segments had the same attribute data, they were then merged. If gaps were less than about 100 m in length, they were closed with the shortest straight connecting line between the ends of the segments. If gaps were larger than that, GE imagery was used to identify the shape of the missing road segment or the location of the missing junction. Using as many visual clues as possible the gap was then closed by matching the direction and flow of the actual road as closely as possible.

2) Road segments that clearly represent roads on the ground but that do not have attribute data and are not connected to other roads in the data set. However, they are relatively close to existing segments and can be connected with little effort.
   a. Response: These roads were connected to the existing road network after referring to GE imagery, similar as described under point one. However, attribute information was left incomplete, because GE imagery is insufficient to assign accurate road types and/or road surface types.

3) Randomly appearing short road segments far removed from the network, without attribute data. These segments sometimes represent real roads visible in GE and may just have been incompletely traced, and sometimes they seem to represent technical artifacts in the network. In either case, they do not have attribute data attached to them.
   a. Response: Whether or not these isolated segments represented real roads, they were removed from the data set. In the case of a real road, the justification for this approach lies in the time and effort it would take to connect those segments to the existing network. GE imagery cannot easily be imported into ArcGIS for tracing, and manual tracing in GE followed by export as KML, conversion into SHP, and merging with the existing network would be a very time consuming process that hopefully can be replaced by automated processes at a later stage.

4) Unconnected arcs without visual confirmation of its position or link to other roads in GE but with attribute information.
   a. Response: We followed the same procedure as in point 3: removal. If this database is intended to serve as a tool to find more efficient or alternative routes for aid delivery or emergency response, there is no point to maintain unconnected arcs in the database. Additionally, it is assumed that this is not a final product and is meant to be constantly updated by users.

5) Visual inconsistency of the road's shape when comparing the data set with GE.

a. Response: Do nothing. (For accuracy, priority is given to the junctions and not to the path of the road as such.

Case study Tanzania: Merging two data sets

Paola Kim Blanco, March 2009

The point of merging different data sets into one is to enhance the road coverage and attributes in a particular country. In the case of Tanzania we used two data sets: Africover and VMAP0_AGI. The issue in Tanzania was that we had one source data set of higher spatial accuracy but limited network coverage (Africover with RMSE=216), and another data set with much lower spatial accuracy (VMAP0_AGI with RMSE= 998) but much wider network coverage. By merging the data set we sought a main network coverage of higher spatial accuracy, but we did not want to lose the great many roads contained in VMAP0.

Preliminary steps before merging: data management and SourceID.

1. Open Arc Catalog
2. Create a new folder inside "editing' subfolder with the country ISO code; hereafter TZA.
3. Copy the individual feature geodatabases (see conversion to geodatabase section above) from the data catalog into TZA.
4. Include a new field in each geodatabase attribute tables called "SourceID" (refer to the UNSDI-T framework for specifications) and place the name of the database to all the records as it was referred in the catalog (i.e. for Africover it would be TZA_Africover_01). This step can be done in Excel or Access.
   a. Open the dbf file; add a new column with the source's name and save it as dbf (dbase) table. The latter can only be done in Access, so unless certain calculations and field re-coding need to be done that require Excel formula functions, use Access to open and edit the exported attribute data, and save as a new dbf from Access.
   b. In Arc Map, add the edited table as Source Data.
   c. Right click on the original feature class, JOINS AND RELATES → JOIN… choose the fields that are used to match records between the two tables (i.e. ObjectID), and crate the join.
   d. For the join to be made permanent in the attribute information, the attribute data will have to be re-exported (DATA → EXPORT DATA…) to a shapefile, which can then be imported as a Feature Class into the geodatabase.
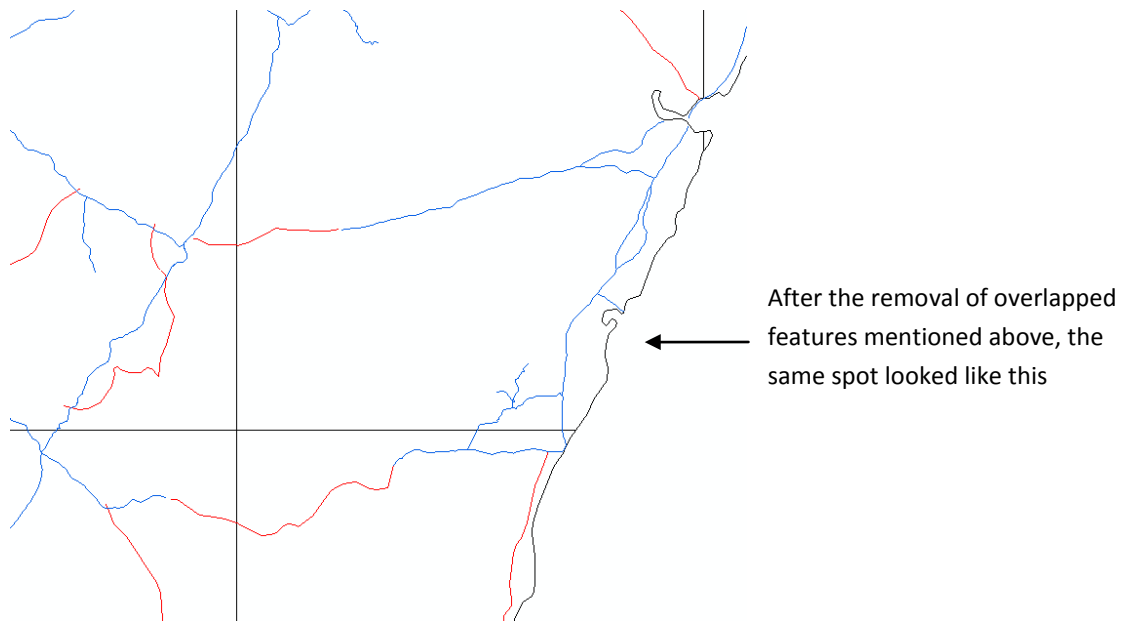
Merging:

1. Open the Arc toolbox→ data management tools → general → merge (it will open a window).
2. Pick the geodatabases (with a SourceID included) to merge from the Editing→ TZA folder as input files.
3. Select a location and name for the output file. Place the merged file in the Editing →TZA folder.
4. Click OK. Now the merged data set can be added into ArcMap for editing.

The steps followed to clean up and edit the merged data set include the removal of repeated features, editing topology and the transfer of attribute information from VMAP0_AGI to Africover.

Step1: Remove duplicate arcs

As stated before, in the case of Tanzania Africover was used as the main structure of the network and VMAP0_AGI as the complement. Therefore VMAP0_AGI's duplicate arcs were removed from the network; the latter also

included the segmentation or breakup of roads. A useful procedure was to set the symbology by SourceID so that the difference in colors could help the editor to easily identify which arc belonged to which data set.

Where both data sets overlapped, segments in red (VMAP0_AGI) were deleted and segments in blue (Africover) were maintained

Where features appeared independently (did not overlap), they were kept regardless of its data source.

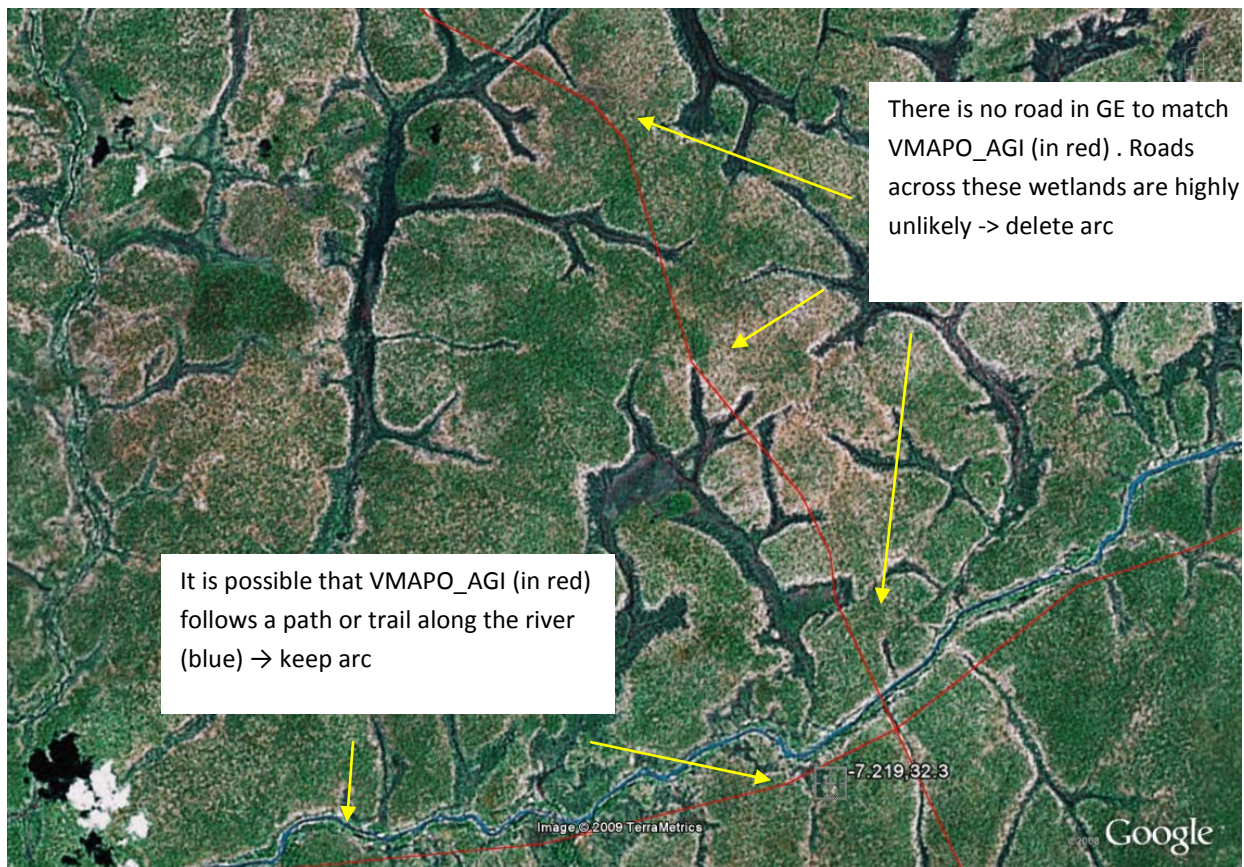After the removal of overlapped features mentioned above, the same spot looked like this
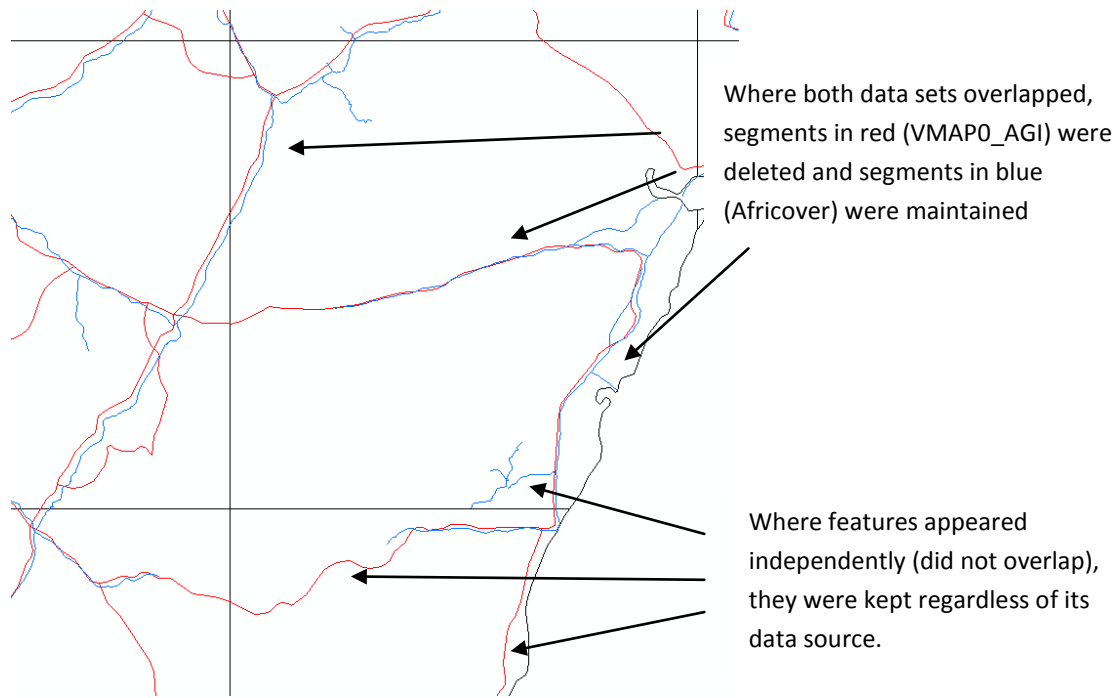
Step 2: Editing topology

A lesson from this exercise is *that when merging different data sets*, the topology edits should be made prior to the transfer of attribute information. Doing the topology first will save a considerable amount of time.

Tanzania followed the same topology rules mentioned before in this document. However, since the nature of the edits was different than previous examples, a decision tree was developed for this purpose (Figure 9). Basically all possible scenarios were covered within the decision tree; in the event of having a different situation, the decision was taken in consultation with the other team members.

Some examples that illustrate the latter are represented below.

1. In cases where VMAPO_AGI keeps record of non-existing roads in GE, arcs are deleted at discretion of the editor (see decision tree). However, it is important to note though that those deleted arcs shall be treated as exceptions and the only reason to delete them is the absence of any linear feature or slight evidence of a road's existence.
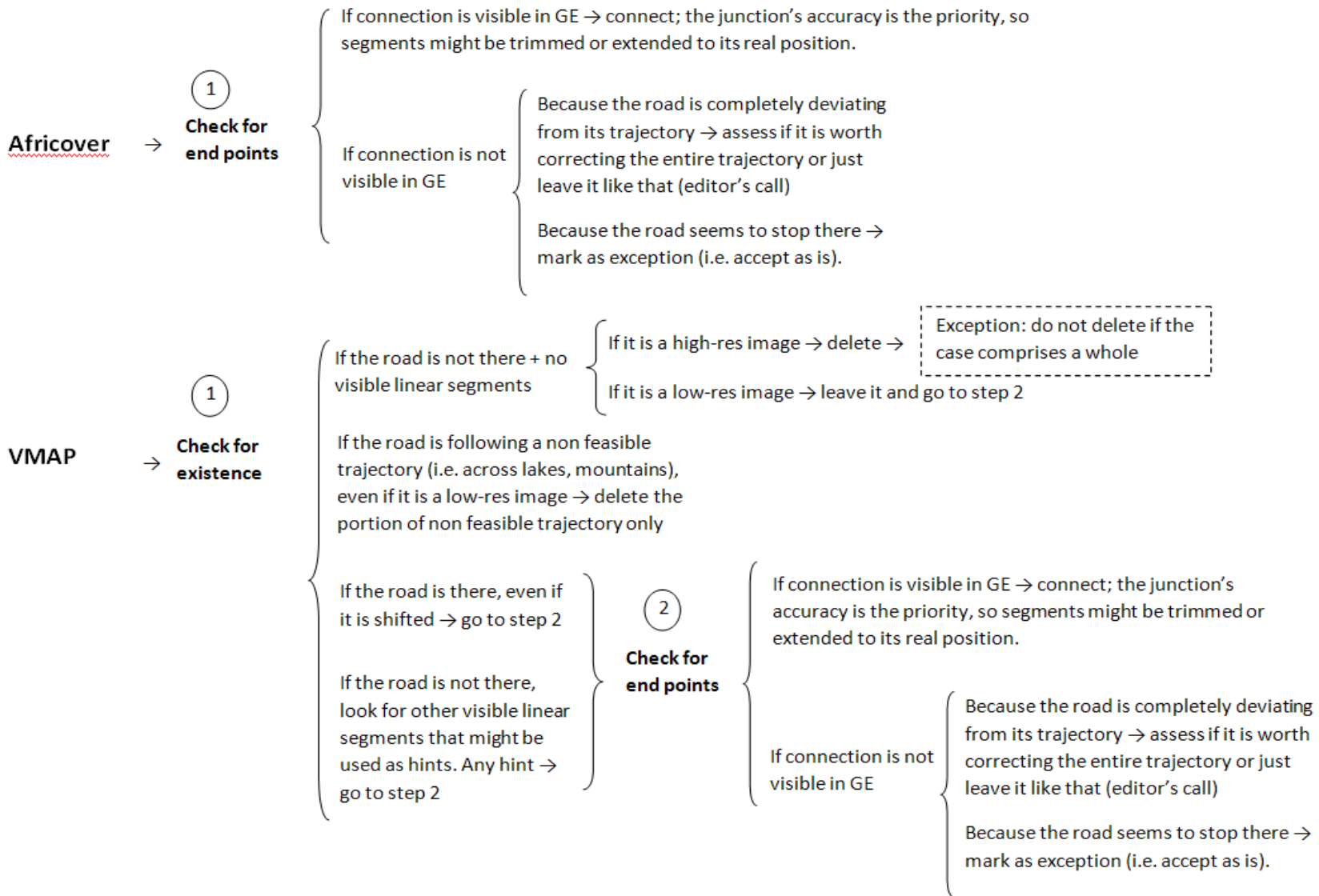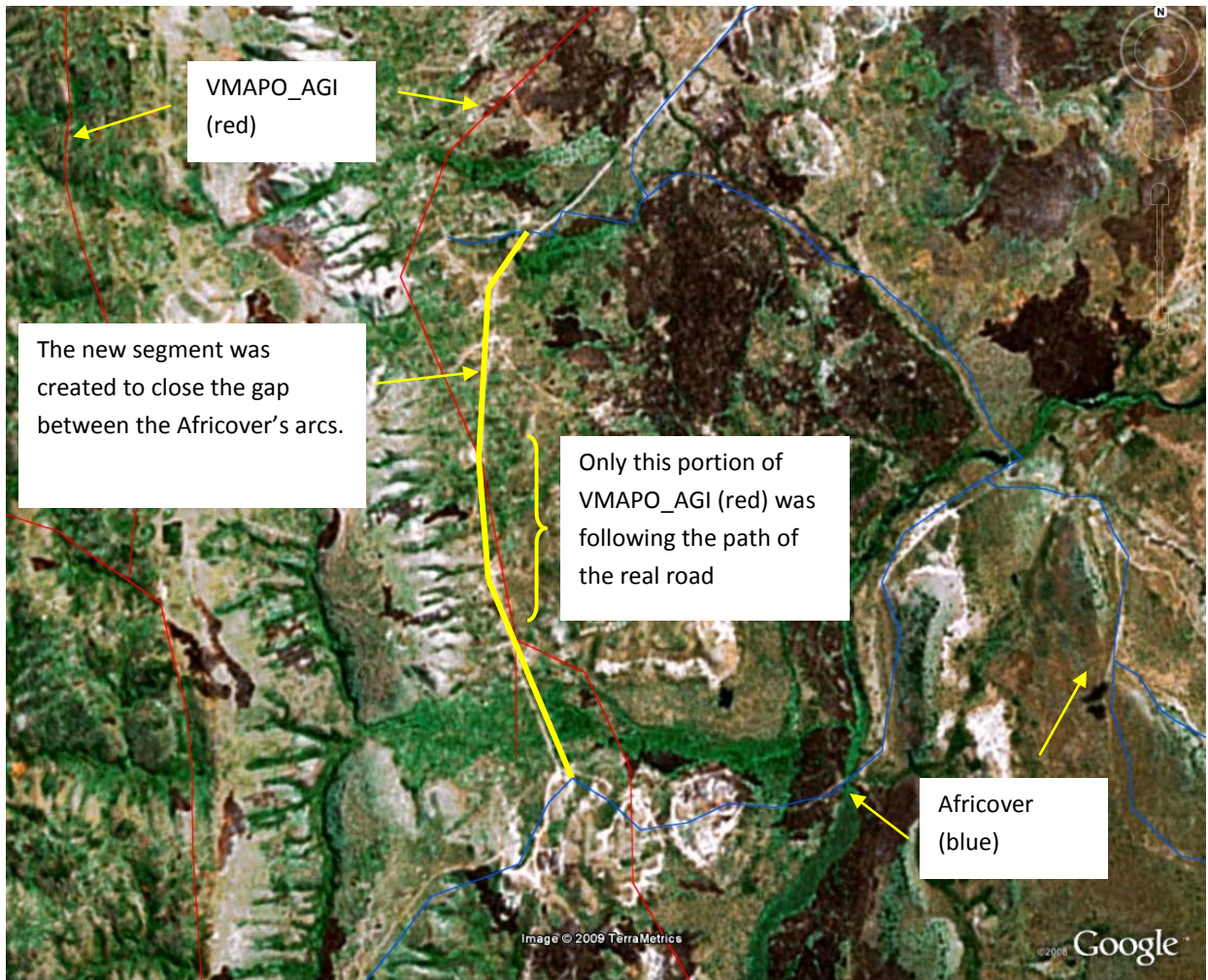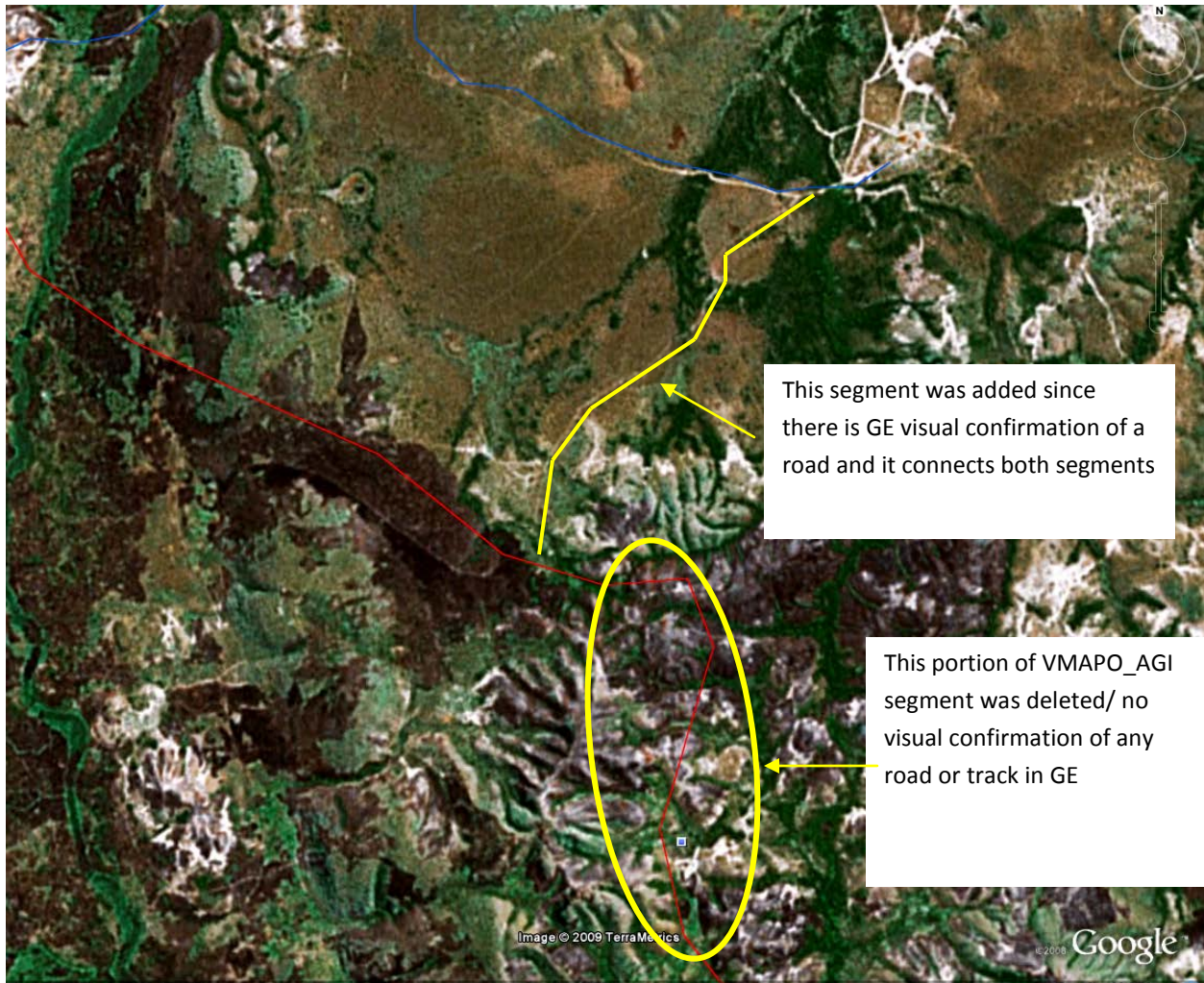
Africover → ① **Check for end points**

If connection is visible in GE → connect; the junction's accuracy is the priority, so segments might be trimmed or extended to its real position.

If connection is not visible in GE

Because the road is completely deviating from its trajectory → assess if it is worth correcting the entire trajectory or just leave it like that (editor's call)

Because the road seems to stop there → mark as exception (i.e. accept as is).

VMAP → ① **Check for existence**

If the road is not there + no visible linear segments

If it is a high-res image → delete →

Exception: do not delete if the case comprises a whole

If it is a low-res image → leave it and go to step 2

If the road is following a non feasible trajectory (i.e. across lakes, mountains), even if it is a low-res image → delete the portion of non feasible trajectory only

If the road is there, even if it is shifted → go to step 2

If the road is not there, look for other visible linear segments that might be used as hints. Any hint → go to step 2

② **Check for end points**

If connection is visible in GE → connect; the junction's accuracy is the priority, so segments might be trimmed or extended to its real position.

If connection is not visible in GE

Because the road is completely deviating from its trajectory → assess if it is worth correcting the entire trajectory or just leave it like that (editor's call)

Because the road seems to stop there → mark as exception (i.e. accept as is).

Figure 7: Decision Tree for Topology Editing

17

2. In cases where the extension of segments is needed to close the gap between features, it will follow geographical coordinates from GE.

VMAPO_AGI (red)

The new segment was created to close the gap between the Africover's arcs.

Only this portion of VMAPO_AGI (red) was following the path of the real road

Africover (blue)

Image © 2009 TerraMetrics

Google

This segment was added since there is GE visual confirmation of a road and it connects both segments

This portion of VMAPO_AGI segment was deleted/ no visual confirmation of any road or track in GE

Editing with geographical reference

In order to have more spatial accuracy while editing the roads' topology, we followed tracks of unconnected segments with geographical coordinates in GE. In that sense, whenever we had a situation of finding a gap or an unconnected road segment in ArcMap, we manually recorded the geographical position and located it in GE by entering the x and y coordinates (in decimal degrees) into the 'fly to' menu. After visual confirmation of the road's existence and its orientation with respect to the network, we traced the road in GE adding marks in strategic points such as junctions or inflexion points and manually recorded the geographical coordinates from those marks. Back in ArcMap, we edited the segment by creating a new feature and introduced the same sequence of x and y coordinates (right click> absolute x and y or F6). We did not expect a new feature to fit into the network with a high level of precision since we know there is a margin of error between GE and the data sets; however we were flexible enough as to adjust the new feature with the rest of the roads. New segments added by this procedure were marked as "2" in the "Exs" field of the attribute table, when possible.

During this process we noticed that:

1) Unless we faced the situation of unconnected roads or gaps within the network, we referred to GE, which means that (at least in the case of Ethiopia) there is a chance of having unnoticed primary or secondary
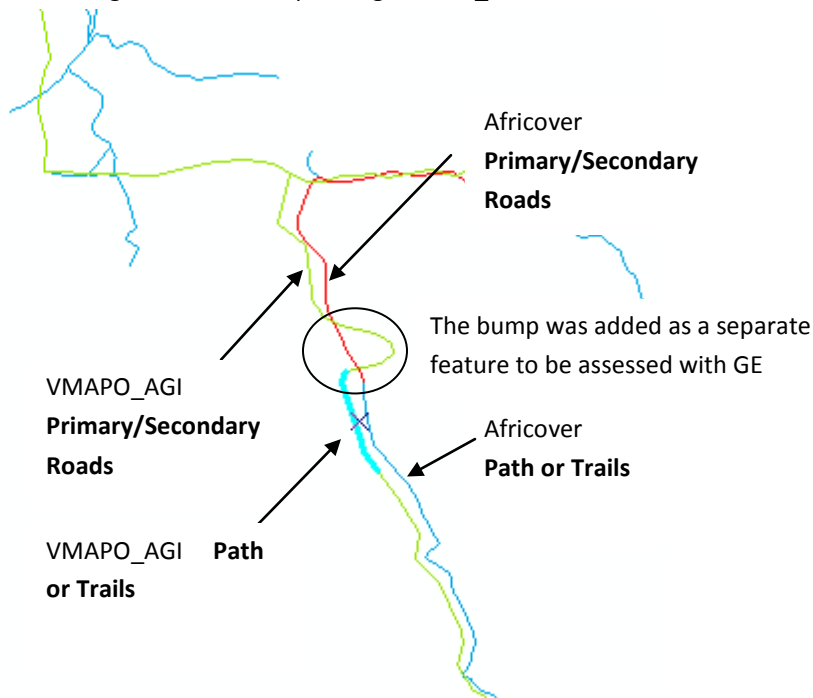
existing roads in GE. That is the problem when we work with two incompatible tools/ formats. However, the latter would not be much of a problem if we had a reliable data set as a reference, which brings out the second point;

2) In most of the editing cases of Ethiopia, if there was a gap or an unconnected segment in the network, the PBL-GRIP database (placed as a GE layer for better reference) had it solved. We are not saying that the GRIP data set is perfect but in the case of this particular country it had more coverage and completeness.

Step 3: Transfer of attribute information

The criteria followed in the allocation of attribute information was derived from the spatial positioning of segments. Thus whenever an Africover feature overlapped a VMAP0_AGI feature, it inherited its attribute information. The transfer of this information was done manually from one attribute table to the other (i.e. copy and paste). Some nuances during this stage are mentioned below:

1. One particular road is composed of several segments in both databases. VMAPO_AGI indicates a differentiation in the road's attribution: one segment of the road is categorized as 'primary/secondary road' and the rest as 'path or trails'. The point where this change takes place is spatially consistent ($\pm$ 100 m) with Africover's segmentation. The response was to transfer to each Africover segment the corresponding VMAPO_AGI attribution.

Africover
Primary/Secondary Roads

The bump was added as a separate feature to be assessed with GE

VMAPO_AGI
Primary/Secondary Roads

Africover
Path or Trails

VMAPO_AGI **Path or Trails**

However, if the Africover road was composed of one segment only or the partitions were not spatially consistent with VMAPO_AGI's segmentation, then the Africover segment was split and the corresponding attributes were allocated.

2. After a relative match between both databases in terms of morphology and spatial accuracy, at some point some inconsistencies within the road's structure and spatial position came into play. In those cases, the response was to transfer the attribute information from VMAPO_AGI to Africover only for those segments with morphological and spatial relative consistency. However, there are cases where the structure (morphology) seem to weigh more in the decision, even if the spatial accuracy was not uniformly present throughout the whole road (ie not following the $\pm$100 m rule). In those cases, the attribute information from VMAPO_AGI was alloted to the whole Africover road.

Structural inconsistency; the attribute information was not allocated to those Africover segments

The structural consistency of the road seems to weigh more than the 100 m tolerance; the attribute information was transferred
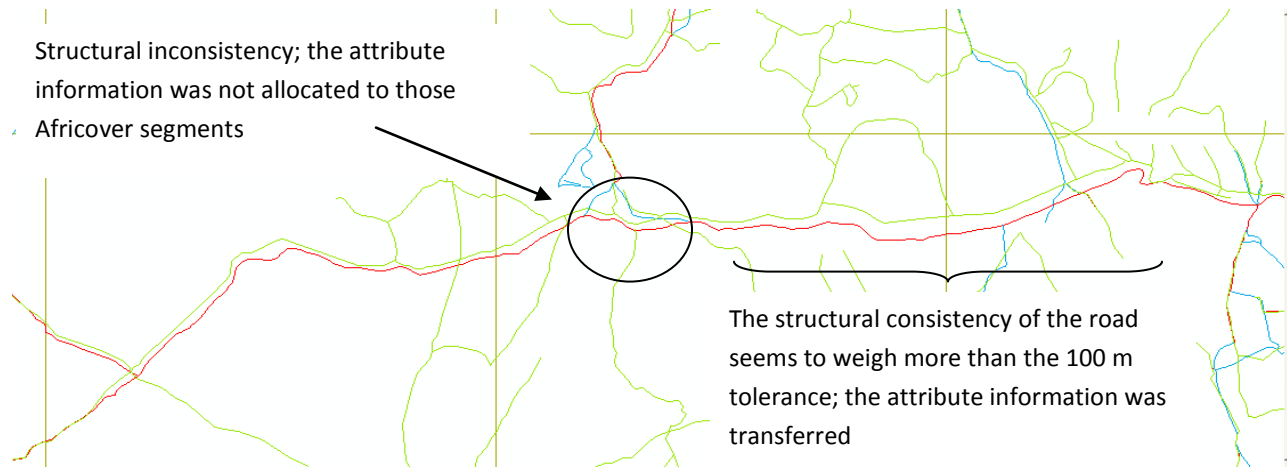
Case Study Kenya: Fixing topology for a GPS track derived data set

Steffen Foerster, November 2008

While attempting to clean up the Kenya data set, **KEN_FAO_IGAD_LPI_01**, a number of problematic issues were encountered that needed some agreed upon and sensible solutions. Many of the issues revolved around the fact that GPS track data were combined with inaccurate VMAP0 data, resulting in a worsening of an already bad topology in this data set.

For all screenshots presented below, **blue lines** represent VMAP0; **green lines** represent the candidate data set mentioned above, which was largely derived from GPS tracks but also includes some remnants of VMAP0.

Situation 1

An isolated road segment is encountered in the candidate data set, and this segment turns out to be a perfect match to a VMAP0 road. The satellite image shows that the position of the segment is off by several hundred meters compared to the real road on the ground. The missing segment is a straight line, and the true position of the road intersection can be identified on the image.

Questions:

1) Should these VMAP0 data be merged and mixed with GPS track data?
    a. Tentative answer: No (SF)
2) If yes, how should the segments become connected?
    a. As is, using the shape of the road on the ground to draw a straight line between the end of the segment and the GPS track, creating a junction at a known false location.
    b. Shifting the segment to meet the actual road on the ground, and then connecting it with the GPS track by tracing the road on the ground. (possible best solution, AMD)
    c. As is, but using the real location of the junction as an intersection, thus changing the flow of the road in the missing section.

Situation 2

An isolated road segment is encountered in the candidate data set, and this segment turns out to match to a VMAP0 road, at least for the majority of its length. However, the high resolution satellite image shows that there is no existing road anywhere near the segment that could match its location and shape.

Question: Should this segment be deleted? According to VMAP0 it may be a real road, but cannot be identified on the ground. Plus, it does not have attribute data assigned to it and is rather far removed from the nearest segment in the candidate data set.

Answer: Yes (SF)

Situation 3

A whole network of roads is encountered in the candidate data set that represent VMAP0 roads, but have no attribute data (VMAP0 attributes have been removed). Satellite imagery does not show any roads in this location, although it is possible that roads are too small to be seen at lower resolution of the example image.

The first image below shows an example from ArcMap to indicate the overlay of VMAP0 (blue) and the candidate data (green). Interestingly, in the lower right corner, there is an inconsistency where VMAP0 vertices were NOT incorporated into the candidate data set (blue line without overlay). The second image shows a smaller section of the same network in GE.

Questions:

1) Should these VMAP0 roads be merged and mixed with GPS tracks, in the absence of visible roads on the ground and the known inaccuracy of VMAP0?
   a. Tentative answer: No (SF)
2) How should we deal with the lack of attribute data in the GPS tracks (green lines without blue lines overlaid)?

Situation 5

Some roads are designated as GAPS in the candidate data set, connecting the ends of two GPS tracks. They have been assigned attribute data of the roads they connect to. Those roads were checked in GE and if no road could be identified on the ground that connects the two GPS tracks, the GAP segment was deleted. Below is an example screenshot of such a connection. The straight line connects two GPS tracks on opposite sides of a mountain range, separated by a river. No roads are visible in that area.

Situation 6

The same road has been tracked twice, on different dates, and both tracks have different attribute data. See example below. SURFCOND is "Poor" in one, and "Fair" in another. District is "Isiolo" in one and "Wajir" in another. Which attribute data shall we maintain?
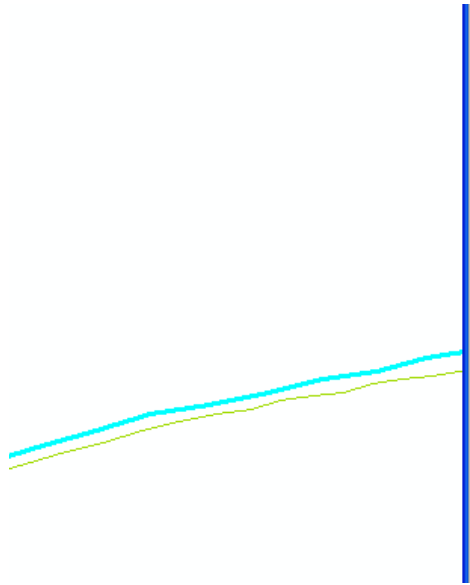
FID	445
Shape	Polyline
OBJECTID	156
SURFTYPE	Earth
NUMLANES	2
SURFCOND	Poor
TERRAIN	Flat
DISTRICT	Wajir
ROADCLASS	E
FROMDESCR	DB Isolo
TODESCR	C80 wajir
ROUGHNESS	0
VEHICLETAG	783
INVENTDATE	2/26/2004
UPDATEDATE	2/26/2004
MAX_PDOP	2.3
CORR_TYPE	Uncorrected
RCVR_TYPE	GPS Pathfinder Power
GPS_DATE	2/26/2004
GPS_TIME	03:07:12pm
FEAT_NAME	Road
DATAFILE	WajirFastComb.ssf
UNFILT_POS	1596
FILT_POS	1596
UPDATE_STA	New
GPS_LENGTH	23.981
GPS_3DLENG	23.972
AVG_HORZ_P	5.63
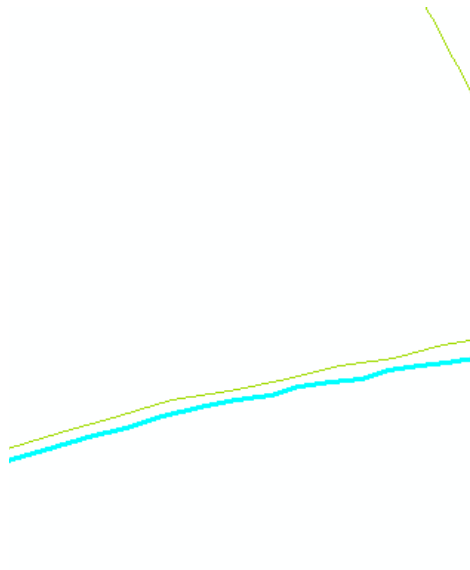
Shape	Polyline
OBJECTID	105
SURFTYPE	Earth
NUMLANES	2
SURFCOND	Fair
TERRAIN	Flat
DISTRICT	Isiolo
ROADCLASS	E
FROMDESCR	E1871
TODESCR	DB Wajir
ROUGHNESS	0
VEHICLETAG	785
INVENTDATE	5/28/2004
UPDATEDATE	5/28/2004
MAX_PDOP	2.2
CORR_TYPE	Uncorrected
RCVR_TYPE	GPS Pathfinder Power
GPS_DATE	5/28/2004
GPS_TIME	10:07:56am
FEAT_NAME	Road
DATAFILE	IsioloFastComb.ssf
UNFILT_POS	1218
FILT_POS	1218
UPDATE_STA	New
GPS_LENGTH	18.034
GPS_3DLENG	18.024
AVG_HORZ_P	5.65

Canada Case Study: Simplifying OpenStreetMap (OSM) data

Taro Ubukawa, January 2013

This section provides step-by-step instructions for simplifying OSM data downloaded in shapefile format from CloudMade. OSM provides excellent data but the level of detail may vary from city-to-city as the "crowd" provides new data. Furthermore, for the gROADS data set, our interest is in inter-settlement road networks, and not so much on street maps within towns and cities. To reduce the detail in cities, we applied the following editing methods to OSM data for Canada.

Data were downloaded from CloudMade (http://downloads.cloudmade.com/) in September 2012 as 13 regional zip archives representing the ten Provinces and three Territories of Canada. According to the CloudMade website, the last maps update was 13 December 2011.

13 regions: Alberta, British Columbia, Manitoba, New Brunswick,
Newfoundland and Labrador, Northwest Territories, Nova Scotia, Nunavut,
Ontario, Prince Edward Island, Quebec, Saskatchewan, Yukon

Each regional zip archive includes seven shapefile data sets: administrative, coastline, highway, location, natural, point of interest (poi), and water. The "highway" data set includes several classes of roads identified by the TYPE attribute.

Road data abstract
Data format: Esri shapefile
File names: XXX_highway (XXX: region name)
Attributes: TYPE, NAME, ONEWAY, LANES

Road types are not exactly the same among the regions. The following are the types of roads observed in Alberta Province: abandoned, bridleway, closed, construction, closing, cycleway, footway, living_street, motorway, motorway_link, path, pedestrian, primary, primary_link, proposed, raceway, residential, road, secondary, secondary_link, track, traffic, traffic_signals, trunk, trunk_link, turning_circle, and unclassified.

Because the OSM data set is prepared for large map scale applications focused on small geographic areas (see http://wiki.openstreetmap.org/wiki/Slippy_map_tilenames#Resolution_and_Scale), there were two major issues that needed to be solved for incorporation into gROADS. The first issue is that many routes are acquired as double line (e.g. divided highway), and the second is that junctions are described in detail. The figures below illustrate this. In its first release, gROADS does not represent divided highways using two vectors, or provide directional information.

The following steps were applied to simplify the data.

**Step 1: Extraction of Road segments based on their TYPE attribute**

Roads with the following TYPE attributes were extracted from the regional highway shapefiles: "primary", "secondary", "motorway", "trunk" (see figure below left). In addition, "primary_link", "secondary_link", "motorway_link", "trunk_link" segments were also extracted for reference during topology editing. We decided not to extract "tertiary" roads at this time since comparison of the OSM data with other data sets revealed that the OSM data set has a much greater density of roads than the others (see figures below).
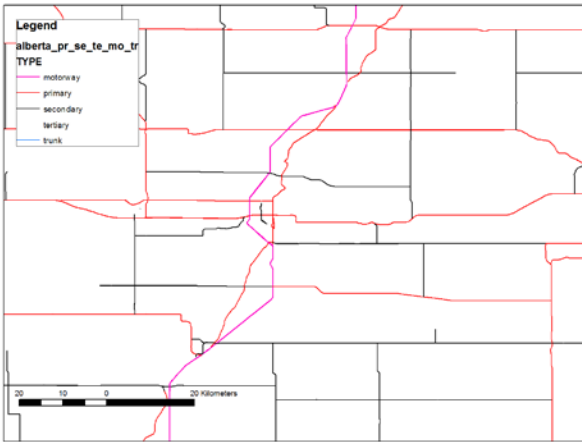
OSM extracted roads– without "tertiary"
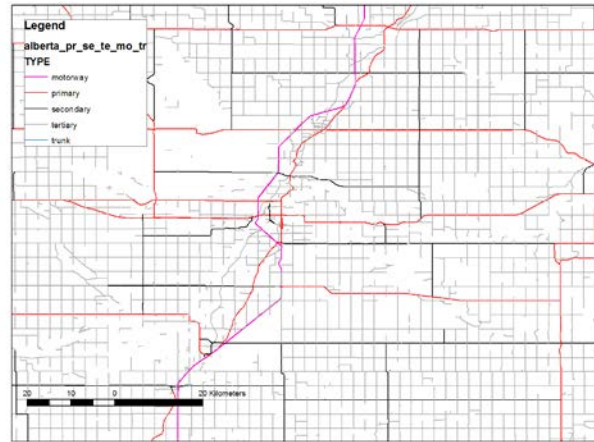
OSM extracted roads – with "tertiary"
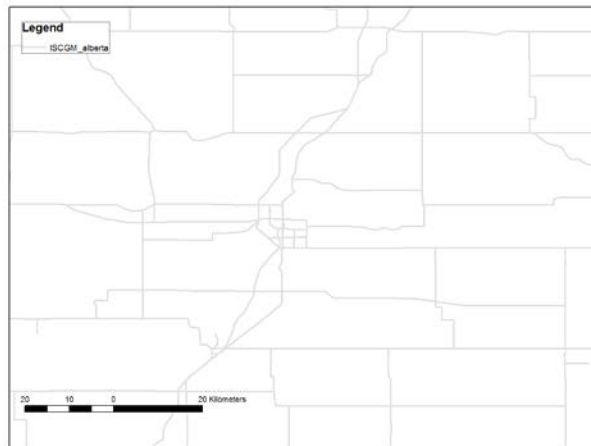
ISCGM data set for the same site

Following this extraction by TYPE, the regional roads data were merged into a national shapefile. Each road TYPE was then exported to a separate shapefile in preparation for the next processing step.

Step 2: Generalizing Roads (Combining double lines)

In order to merge double lines, the roads are generalized by converting into the ESRI coverage data format. During conversion to coverage, segments and topologies of the data set are cleaned within a specified tolerance. (Note: there is now a tool named "Collapse Dual Lines to Centerline" in ArcGIS. This tool seems to be used for this process; however, we found that this tool will only generalize lines if the distance separating them remains smaller than the specified "Maximum Width" for the entire length of the lines. If the distance separating the lines exceeds the "Maximum Width" at even one location, the tool will fail to generalize the lines.)

If you merge all road classes (TYPES) before generalizing the double lines, the upper class roads can be replaced with lower class roads as shown in the figures below. This results in a disconnection of the upper class routes.
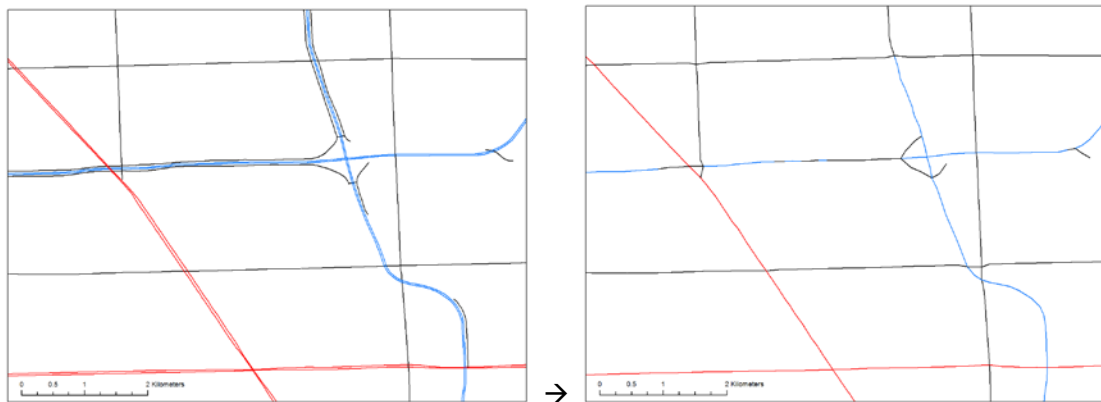
Color scheme: Primary = red; Secondary = black; Mortorway = blue; Trunk = green

Therefore, each data set, "primary", "secondary", "motorway", and "trunk", was separately converted into a coverage file to generalize the double lines. The XY tolerance was set to 0.0005 decimal degrees (Note: 0.0083333 dd is about 1km at equator, so this is about 100m scale). Then, the coverage files were converted into shapefile, which were imported into file geodatabases for topology editing.

The following step-by-step methods in ArcGIS are provided for those wishing more information.

    1. Open ArcToolbox in ArcMap → Conversion Tools → To Coverage → Feature Class To Coverage

    2. Specify the input file name (shapefile), chose type "arc", enter 0.0005 (decimal degree) for the XY tolerance. Click OK.

    3. After the output coverage file is added to the ArcMap table of contents, right click the "arc" class, select "Data" → "export Data", and then save it as a shapefile.
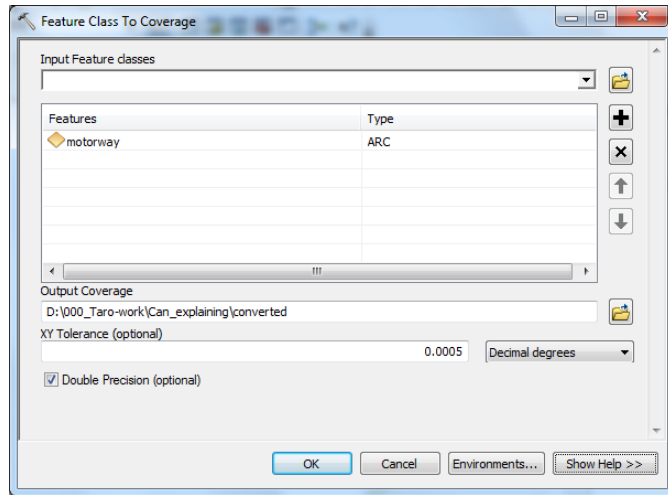
Feature class (left) converted to coverage then shapefile (right), collapsing double lines in the process.

 →

 →
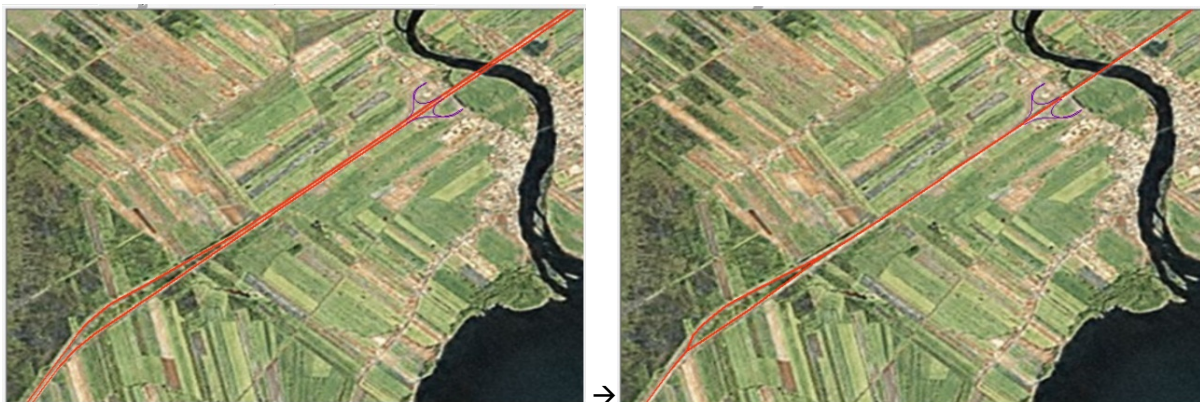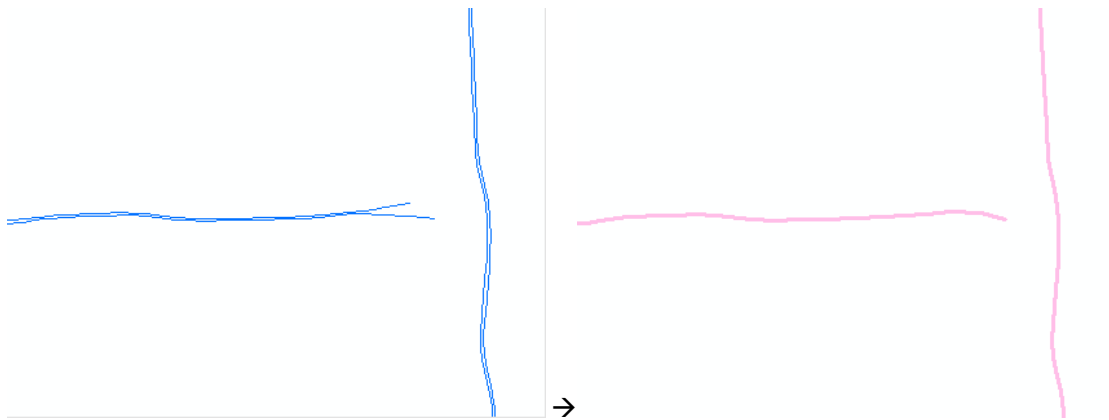
The following step-by-step is for collapsing dual lines to center lines:

    1. Open ArcToolbox → Cartography Tools → Generalization → Collapse Dual Lines To Centerline

2. Specify the input file name (shapefile), set the maximum width for collapsing for 0.0005 (decimal degree). Click OK
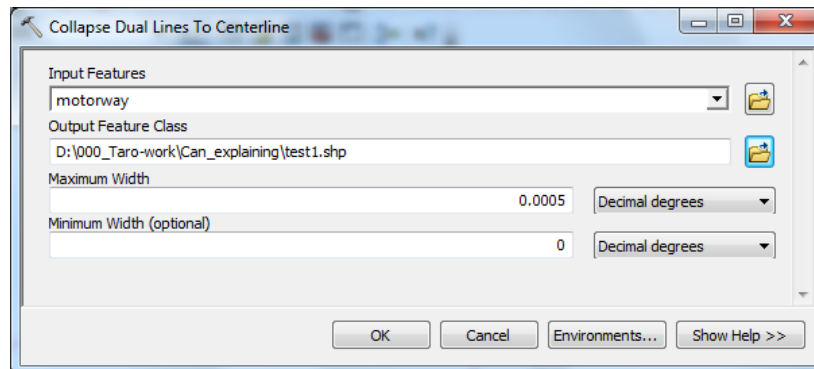
Step 3: Topology Checking

Before merging all generalized road classes, we built a topology for each class and checked the connections at road junctions. Where the lines were disconnected, resulting in dangles, we connected them manually, using our "XXX_link" data (see above) for guidance. We ignored dangles that were not the result of a broken junction. Connecting the road segments manually was time consuming and future work would benefit from development of a more efficient method.
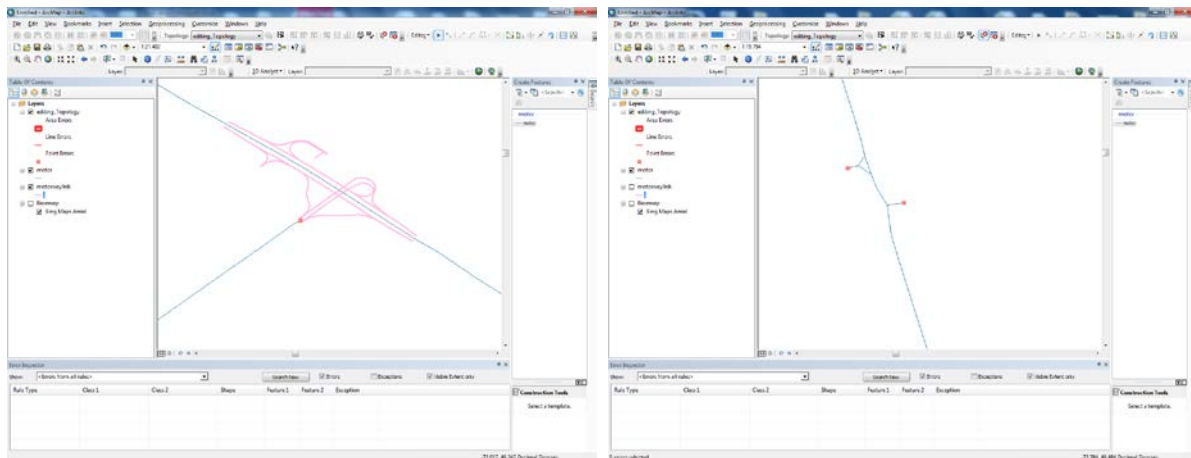
Connected dangle   Ignored dangles
                                        Pink: link segments, Orange: topology error

In some cases, there was a disconnection in the OSM data, while other maps indicated a connection. We judged the connections based solely on the OSM data.
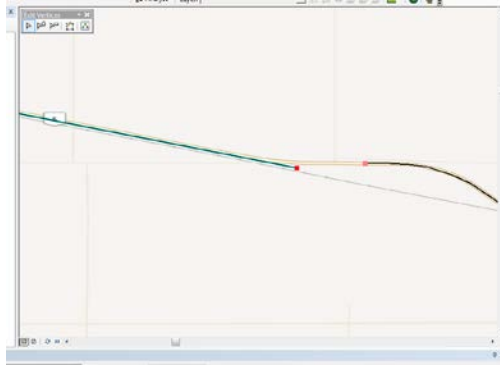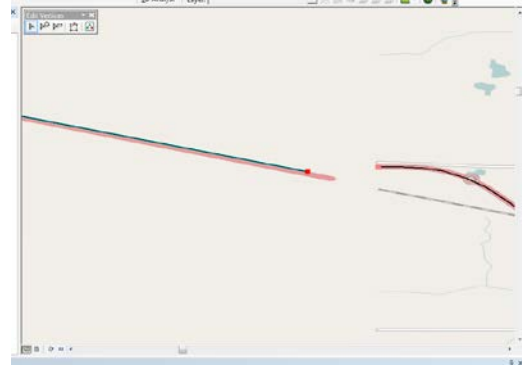
| Bing (connected) | OSM (not connected) |

Step 4: Combining all classes

The topology edited data were exported from feature classes to shapefiles and then merged into a single national roads shapefile. In order to preserve roads of each class during the merge, we used a small tolerance of 0.00001 decimal degrees (ca. 1-10 m order).

To merge all classes we used the following steps:

1. Open ArcToolbox → Data Management Tools → General → Merge

2. Specify all of the road class shapefiles as Input data, then click "Environments" to set the XY tolerance to 0.0005 (decimal degree). Click OK
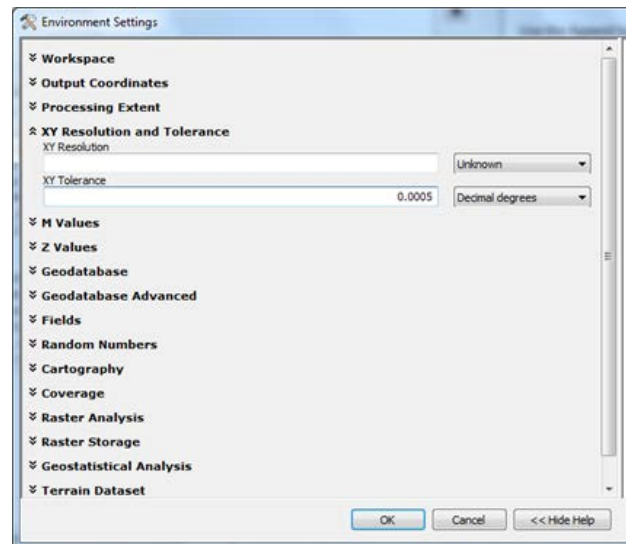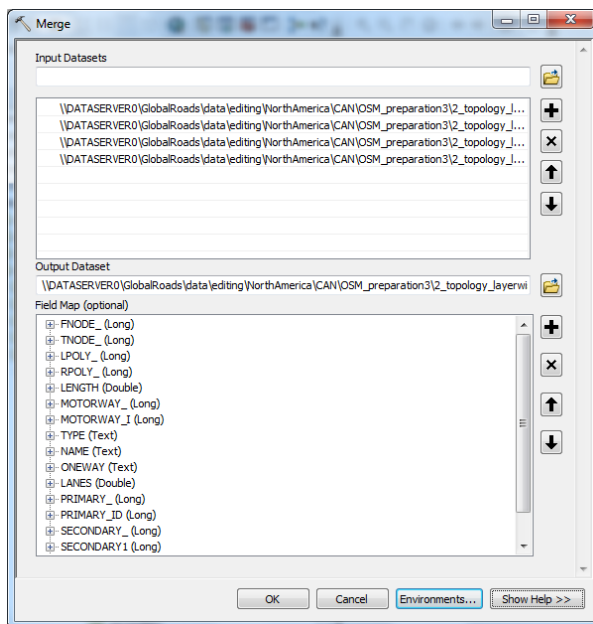
Then, we built topology and checked for dangles between road classes. These were corrected manually. For all of Canada, it took about 10 hours for the steps 2.3 and 2.4.
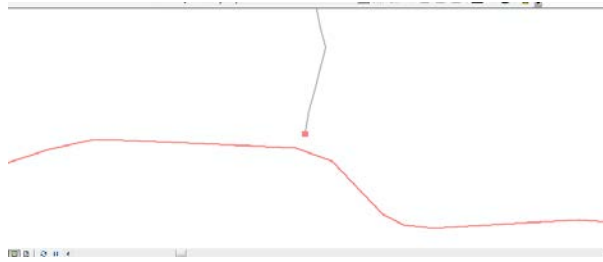
Above: Example of disconnection

Step 5: Attribute translation

The following table outlines the attribute transform from the Canadian OSM data to the UNSDI-T model. The classification of roads may differ by country by country, so we referred to several websites and consulted a Canadian colleague.

OSM Attributes	Translated attributes	Note
Type - Primary - Secondary - Motorway - Trunk	**FClass** - Primary - Secondary - Highway - Highway	Although all classes can be highway, the motorway and trunk classes are always interpreted as highway.
Road name - (Actual Name)	**RteName** - (Actual Name)	
Lanes	(NumLanes) n.a.	As the double lines are merged, original "lane" attributes cannot be transferred.
ONEWAY	n.a.	

Other general rules and guidelines

There are cases of unconnected roads with defined information in the attribute table and a certain level of visual confirmation in GE, but not enough as to accurately connect them to the network. The main uncertainty comes when we are trying to define the geographical position of the road and/or junction. As we can see from the GE image below, we have an unconnected road at the bottom-left side which can be further traced by pins 13, 14 and 15. However, after pin 15 the low resolution of the image makes it difficult to locate the next pin. The distance from pin 15 to the nearest road is more than 6 km.
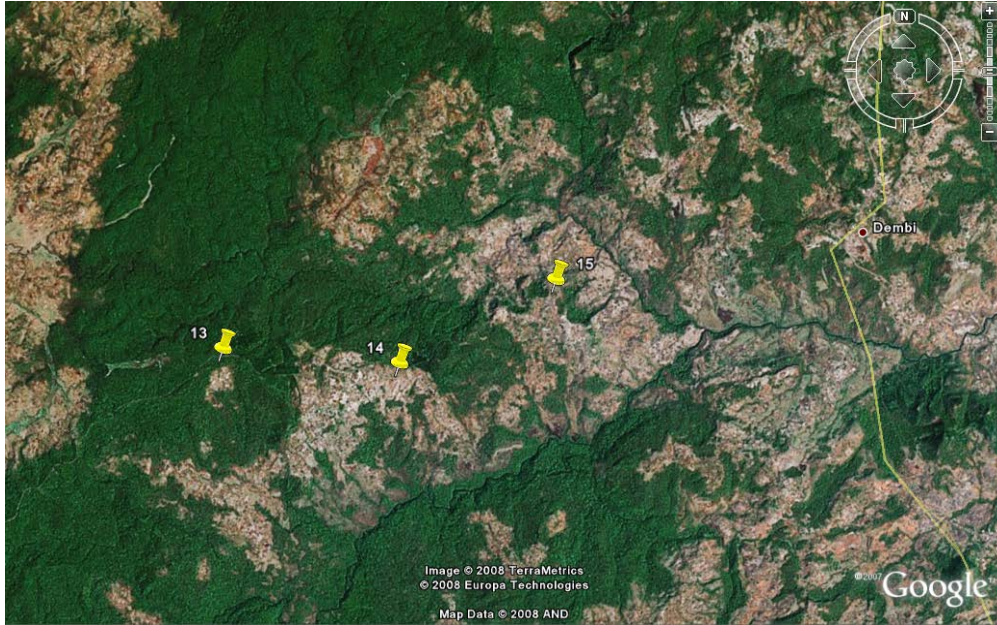
Question:

1) Should we remove this road from the data set?
    a. No; however joining it to the network by a straight line is not a good idea either (it is more than 1 km). The placement of the junction would be highly speculative because GE is not helping much this time (PKB).

Network Validation

For some of the data sets in gROADS v1, we ran a network validation. This involved creating and building a network in ArcCatalog, and creating a route with driving directions in ArcMap.

Creating the network data set:

1. In ArcCatalog, right click on the Feature Data set, select NEW → NETWORK DATA SET...
    a. If you get an error message, make sure all instances of the data set are closed in ArcMap
2. Enter a name for the network data set, keep the default, which is [Feature Data set Name]_ND
3. Select the Feature Classes that will participate in the network data set
4. Skip the CONNECTIVITY settings
5. Skip the next dialog window asking for modification of connectivity based on elevation data
6. In the next dialog box, make sure YES is selected for modeling turns
7. In the next dialog box asking to specify attributes, click NEXT. You will be asked whether you want a cost attribute based on the SHAPE LENGTH (there needs to be at least one attribute specified), click YES
8. In the next dialog box, select YES for setting driving directions, then click NEXT

9. Select a field used to identify road segments as shown in the screenshot under PRIMARY
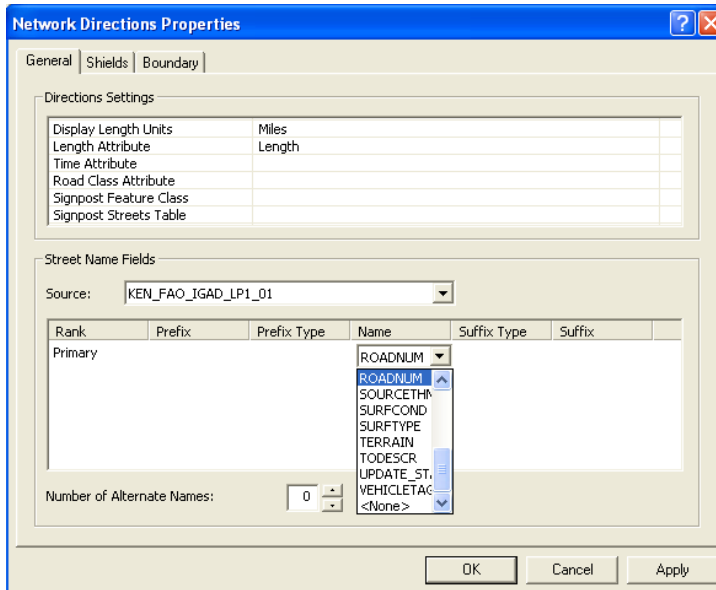

10. Confirm with OK
11. Click NEXT and FINISH creating your network data set
12. When asked, confirm for your data set to be built after it has been created. This can take time

Running a sample network analysis:

1. In ArcMap, open your network data set
2. Activate the Network Analyst Toolbar. You may have to activate the Network Analyst extension first under TOOLS → EXTENSTIONS → Network Analyst
3. In the Network Analyst Toolbar, choose NEW ROUTE

4. Go to NETWORK ANALYST → OPTIONS, and set snapping to positions along the network by selecting the appropriate checkbox (see screenshot). Deactivate the OFFSET checkbox
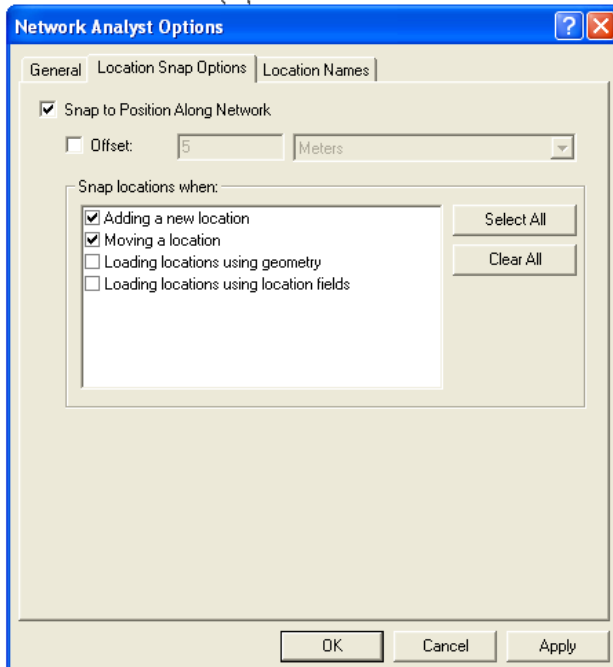
5. Click on the CREATE NETWORK LOCATION TOOL button in the Network Analyst Toolbar
6. Select at least two locations on your network (space them out a bit for a challenge)
7. Click the SOLVE button on the Network Analyst Toolbar
8. If you are receiving an error that the network cannot be solved, this is likely due to remaining topology errors. Go back to topology editing, make sure you have applied all appropriate rules, and repeat the process
9. If everything works as planned, ArcMap will display the shortest connection between the two points, and you can obtain driving directions by clicking on the DIRECTIONS button
10. You are done

ETL – Extract, Transform, Load

As a final step, we transformed the edited and cleaned version of a data set so that its attribute information conforms to the UNSDIT Version 2 data model (Annex III). The process, also called Extract – Transform – Load (ETL), involves getting the attribute information into Excel, re-coding attributes to the UNSDIT data model, and loading the new attribute information into the global data set. The steps are described below.

Note: As the project progressed, ETL was generally executed prior to topology editing. Doing ETL last makes sense for cases in which two data sets are being merged (e.g. the Tanzania case study detailed above), thus requiring internal topology editing and attribute transfer from one data set to the other. In cases requiring topology editing only along the country borders, it makes more sense to edit topology after ETL.

Extracting the attribute information:

1. Before you do anything else, back-up your data set!!! Create a back-up copy of the geodatabase, save it in the same folder in the format <ISO Code>_Backup_<MMDDYYYY>
2. Open the final edited feature class in ArcMap
3. Open attribute table, and export to DBF, save in same location as geodatabase, use file convention <Feature class name>_Attributes.dbf

Transforming the data model:

1. Open the table in Excel
2. Copy all data to a second worksheet, and call it TRANSFORMATION
3. Next, add a worksheet FIELD MAPPING where you document the re-mapping of attribute fields to the UNSDIT model
   a. The layout of the FIELD MAPPING sheet should look like this

	A	B	C	D	E
1	Dataset Field Name	UNSDIT_V2 Name	Dataset Categories	UNSDIT_V2 Categories	UNSDIT_V2 Codes
2	SURFTYPE	SrfTpe	Gravel	Gravel	2
3	SURFTYPE	SrfTpe	Earth	Dirt/Sand	3
4	SURFTYPE	SrfTpe	Surface Dressing	Paved	1
5	SURFTYPE	SrfTpe	Sand	Dirt/Sand	3
6	SURFTYPE	SrfTpe	Premix	Gravel	2
7	SURFTYPE	SrfTpe	GAP	Unspecified	0
8	SURFTYPE	SrfTpe	Concrete Block	Paved	1
9	SURFTYPE	SrfTpe	Set Stone	Paved	1
10	SURFTYPE	SrfTpe	Concrete (Jt-Plain)	Paved	1
11	SURFTYPE	SrfTpe		Unspecified	0
12	SURFCOND	SrfCond	Fair		
13	SURFCOND	SrfCond	Good		
14	SURFCOND	SrfCond	Poor		

4. To identify unique categories for a given field in the attribute table, use the ADVANCED FILTER in Excel as follows:
   a. On the DATA – SORT & FILTER tab, click on ADVANCED
   b. In the dialog box, keep FILTER THE LIST, IN PLACE activated
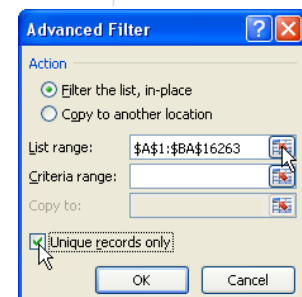
    c.   Click on the button next to LIST RANGE (see screenshot to the right) and in the worksheet, click on the column header of the column that you want to filter, and return to the dialog box by clicking the button in the Advanced Filter dialog (see screenshot below)
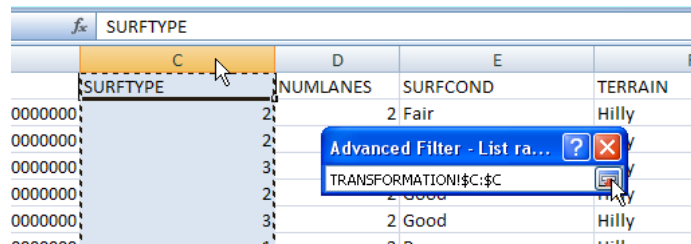
    d.   Click OK, and only your unique values in the column will be displayed
    e.   Copy those values onto the FIELD MAPPING worksheet, and decide how to map them to the UNSDIT_V2 data model fields

5.   When you have mapped all attributes of a given field to their equivalent in the UNSDIT model, as far as possible, use the FIND & REPLACE tool in Excel to recode your fields on the TRANSFORMATION worksheet accordingly

6.   Also include fields on this sheet for which you do not have to recode any values but where the name of the field needs to be changed

7.   All other fields that have no equivalent in the UNSDIT data model will be deleted. You can copy a list of those fields onto a separate worksheet for DELETED FIELDS.

Loading the transformed data model:

1.   Save the final version of the transformed data model in Excel format.

2.   As Excel cannot save as DBF file format, we need to import the file into Access, then save it as a DBF file in Access. To do this, create an empty database and import the Excel spreadsheet you created. Take care not to add a PRIMARY KEY in Access, as this would add an additional field to your table. Then EXPORT… using the naming convention <Feature Class Name>_UNSDIT.dbf

3.   Make sure you have XTOOLS for ArcGIS installed

4.   Open the feature class in ArcMap

5.   Create a join with the DBF file that contains the UNSDIT data model, based on the OBJECTID

6.   Check the number of records in the resulting attribute table to verify that that the file has been joined correctly, it should have exactly the same number of records as the original file

7.   Export the data into a new shapefile, add the shapefile to the Map document

8.   In XTOOLS Pro, go to TABLE OPERATIONS → MultiDelete Fields

9.   Select all fields from the original data set and delete. This process can take a while.

10. If some fields in the original attribute table had the same names as in the UNSDIT data model, the fields that you joined would have been renamed during the JOIN operation (e.g. NUMLANES_1 from NUMLANES). At this point, check for any renamed fields and adjust the names back to the UNSDIT name using XTOOLS Pro → TABLE OPERATIONS → TABLE RESTRUCTURE, which will create a new shapefile that you can name with the extension _final.shp to the original file name. If you get an error message by XTOOLS, try a shorter file name.

11. If all worked well, you should now have a shapefile that fully conforms to the UNSDIT standard!

12. In ArcCatalog, navigate to your Feature Data set in your Geodatabase

13. Import the shapefile as outlined on page 8. Name the final version of your feature class <Data set Name>_UNSDIT_FINAL.

<u>Notes:</u>

In the process of the ETL as described above, you have generated quite a few files, most of which are no longer needed once you have arrived at the final data model. As a review, here is a list of files generated using the example of the KEN data set, and what you should do with them:

- KEN_FAO_IGAD_LPI_01_exdited.dbf (exported original attributes) → DELETE
- KEN_FAO_IGAD_LPI_01_UNSDIT.dbf (remapped attribute table from EXCEL via Access Export) → DELETE
- KEN_UNSDIT.shp (shapefile result of the original JOIN operation, with changed duplicate field names) → DELETE
- KEN_UNSDIT_final.shp → DELETE after it has been successfully imported into the geodatabase
- Transformation.accdb (access database for turning Excel data into DBF format) → DELETE
- **KEN_FAO_IGAD_LPI_01_edited.xls** (the excel file where attributes were remapped) → KEEP (this is the most important file to keep as it shows what you did and provided both original and transformed attribute data)
- KEN_Backup_03182009.gdb (backed up database before ETL) → KEEP as a backup in backup folder

References

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

Annex I. ISO3 version 10 country codes

All file naming in gROADS was done using International Standards Organization three letter country codes (ISO3 version 10). The followign is a reference list of the codes that were in existence during the data development process.

ABW	*Aruba*		*CHN*	*China*		*GLP*	*Guadeloupe*
AFG	*Afghanistan*		*CIV*	*Côte d'Ivoire*		*GMB*	*Gambia*
AGO	*Angola*		*CMR*	*Cameroon*		*GNB*	*Guinea-Bissau*
AIA	*Anguilla*		*COD*	*Dem. Rep. Congo*		*GNQ*	*Equ. Guinea*
ALB	*Albania*		*COG*	*Congo*		*GRC*	*Greece*
AND	*Andorra*		*COK*	*Cook Islands*		*GRD*	*Grenada*
ANT	*Netherland Antilles*		*COL*	*Colombia*		*GRL*	*Greenland*
ARE	*United Arab Emirates*		*COM*	*Comoros*		*GTM*	*Guatemala*
ARG	*Argentina*		*CPV*	*Cape Verde*		*GUF*	*French Guiana*
ARM	*Armenia*		*CRI*	*Costa Rica*		*GUM*	*Guam*
ASM	*American Samoa*		*CUB*	*Cuba*		*GUY*	*Guyana*
ATA	*Antarctica*		*CXR*	*Christmas Island*		*HKG*	*Hong Kong*
ATF	*French Southern Terr.*		*CYM*	*Cayman Islands*		*HMD*	*Heard & McDonald Isl.*
ATG	*Antigua and Barbuda*		*CYP*	*Cyprus*		*HND*	*Honduras*
AUS	*Australia*		*CZE*	*Czech Rep.*		*HRV*	*Croatia*
AUT	*Austria*		*DEU*	*Germany*		*HTI*	*Haiti*
AZE	*Azerbaijan*		*DJI*	*Djibouti*		*HUN*	*Hungary*
BDI	*Burundi*		*DMA*	*Dominica*		*IDN*	*Indonesia*
BEL	*Belgium*		*DNK*	*Denmark*		*IMN*	*Isle of Man*
BEN	*Benin*		*DOM*	*Dominican Rep.*		*IND*	*India*
BFA	*Burkina Faso*		*DZA*	*Algeria*		*IOT*	*Indian Ocean Territory*
BGD	*Bangladesh*		*ECU*	*Ecuador*		*IRL*	*Ireland*
BGR	*Bulgaria*		*EGY*	*Egypt*		*IRN*	*Iran*
BHR	*Bahrain*		*ERI*	*Eritrea*		*IRQ*	*Iraq*
BHS	*Bahamas*		*ESH*	*Western Sahara*		*ISL*	*Iceland*
BIH	*Bosnia & Herzegovina*		*ESP*	*Spain*		*ISR*	*Israel*
BLR	*Belarus*		*EST*	*Estonia*		*ITA*	*Italy*
BLZ	*Belize*		*ETH*	*Ethiopia*		*JAM*	*Jamaica*
BMU	*Bermuda*		*FIN*	*Finland*		*JOR*	*Jordan*
BOL	*Bolivia*		*FJI*	*Fiji*		*JPN*	*Japan*
BRA	*Brazil*		*FLK*	*Falkland Islands*		*KAZ*	*Kazakhstan*
BRB	*Barbados*		*FRA*	*France*		*KEN*	*Kenya*
BRN	*Brunei Darussalam*		*FRO*	*Faeroe Islands*		*KGZ*	*Kyrgyzstan*
BTN	*Bhutan*		*FSM*	*Micronesia*		*KHM*	*Cambodia*
BVT	*Bouvet Island*		*GAB*	*Gabon*		*KIR*	*Kiribati*
BWA	*Botswana*		*GBR*	*United Kingdom*		*KNA*	*St. Kitts & Nevis*
CAF	*Central African Rep.*		*GEO*	*Georgia*		*KOR*	*South Korea*
CAN	*Canada*		*GHA*	*Ghana*		*KWT*	*Kuwait*
CCK	*Cocos Islands*		*GIB*	*Gibraltar*		*LAO*	*Laos*
CHE	*Switzerland*		*GIN*	*Guinea*		*LBN*	*Lebanon*
CHL	*Chile*						

Code	Country	Code	Country	Code	Country
LBR	Liberia	NPL	Nepal	SVN	Slovenia
LBY	Libya	NRU	Nauru	SWE	Sweden
LCA	St. Lucia	NZL	New Zealand	SWZ	Swaziland
LIE	Liechtenstein	OMN	Oman	SYC	Seychelles
LKA	Sri Lanka	PAK	Pakistan	SYR	Syria
LSO	Lesotho	PAN	Panama	TCA	Turks & Caicos Islands
LTU	Lithuania	PCN	Pitcairn	TCD	Chad
LUX	Luxembourg	PER	Peru	TGO	Togo
LVA	Latvia	PHL	Philippines	THA	Thailand
MAC	Macao	PLW	Palau	TJK	Tajikistan
MAR	Morocco	PNG	Papua New Guinea	TKL	Tokelau
MCO	Monaco	POL	Poland	TKM	Turkmenistan
MDA	Moldova	PRI	Puerto Rico	TLS	East Timor
MDG	Madagascar	PRK	North Korea	TON	Tonga
MDV	Maldives	PRT	Portugal	TTO	Trinidad & Tobago
MEX	Mexico	PRY	Paraguay	TUN	Tunisia
MHL	Marshall Isl.	PSE	Occ. Palestinian Terr.	TUR	Turkey
MKD	Macedonia	PYF	French Polynesia	TUV	Tuvalu
MLI	Mali	QAT	Qatar	TWN	Taiwan
MLT	Malta	REU	Réunion	TZA	Tanzania
MMR	Myanmar	ROU	Romania	UGA	Uganda
MNG	Mongolia	RUS	Russia	UKR	Ukraine
MNP	Northern Mariana Isl.	RWA	Rwanda	UMI	US Minor Outlying Isl.
MOZ	Mozambique	SAU	Saudi Arabia	URY	Uruguay
MRT	Mauritania	SCG	Serbia & Montenegro	USA	United States
MSR	Montserrat	SDN	Sudan	UZB	Uzbekistan
MTQ	Martinique	SEN	Senegal	VAT	Holy See
MUS	Mauritius	SGP	Singapore	VCT	St. Vincent & the Gren.
MWI	Malawi	SGS	S. Georgia/ Sandwich I.	VEN	Venezuela
MYS	Malaysia	SHN	St. Helena	VGB	British Virgin Islands
MYT	Mayotte	SJM	Svalbard/Jan Mayen I.	VIR	US Virgin Islands
NAM	Namibia	SLB	Solomon Islands	VNM	Viet Nam
NCL	New Caledonia	SLE	Sierra Leone	VUT	Vanuatu
NER	Niger	SLV	El Salvador	WLF	Wallis & Futuna Isl.
NFK	Norfolk Island	SMR	San Marino	WSM	Samoa
NGA	Nigeria	SOM	Somalia	YEM	Yemen
NIC	Nicaragua	SPM	St. Pierre & Miquelon	ZAF	South Africa
NIU	Niue	STP	Sao Tome & Principe	ZMB	Zambia
NLD	Netherlands	SUR	Suriname	ZWE	Zimbabwe
NOR	Norway	SVK	Slovakia		

Annex II. Catalog fields of for the CODATA-Roads data set catalog

A catalog of country and regional roads data sets, the "CODATA Catalog of Roads, v1 (1976–2008)", is available for download from http://sedac.ciesin.columbia.edu/data/set/groads-codata-roads-catalog-v1. The following are the fields it contains.

Field	Description
Extent	Bounding box in lat-lon.
Coverage	Country or countries covered (relational table pull-down list)
ISO3	ISO3 code(s)
Region	UN region within which the data lie (relational table pull-down list)
Subregion	UN sub-region within which the data lie (relational table pull-down list)
Title	Data set title
Originator	Institution that created the data
Distributor	Institution distributing the data
Contact1	Contact person, organization, address, phone number, and email
Contact2	Contact organization
Contact3	Contact organization address & phone number
Contact4	Contact person email
Source	Source data upon which the data set creator relied to create the data set
Method	Method of data development
Year1	Year of creation
Year2	Approximate year (or range of years) of road network portrayal
URL	URL for the data set (if available online)
MetadataURL	Online link to metadata or text file with data about the data
Metadata	Link to document containing the metadata on SEDAC servers.
Constraints1	Coding of constraints for use (e.g., scientific/research use only, noncommercial use only)
Constraints2	Coding of constraints for redistribution
Permissions1	Permission to redistribution obtained? (yes/no/status of process)
Permissions2	Link to original PDF of the permissions document, email, or letter granting redistribution rights.
Restrict2	Original license text, or link to the document in which it is contained.
TotalKm	Total kilometers of road network coverage by country.
PctDCW	Road length as a percent of DCW/VMAP0.

PctIRF	Road length as a percent of IRF statistics by road class.
Class	Road classification system employed by the data set
ClassLevel	Lowest class level based on a standard classification system we'll need to borrow or develop (e.g. Motorways, limited access divided highways; Highways, main or national roads; Secondary, regional roads; Urban streets; Unimproved roads, logging roads and tracks).
Attributes	Other road attribute information such as road condition, weight restrictions, surface type, etc., that are provided with the data set.
Horizontal	Approximate horizontal accuracy in Root Mean Square Error (RMSE) compared to orthorectified satellite imagery or GPS tracks
Language	Original language of the data set
OrigFormat	Orginal format of data set (e.g. Shapefile, coverage, VPF, paper map)
Size	Size in MB of the data file
EPSG	EPSG code for defining the projection, i.e. EPSG 4326 is lat lon, WGS 1984 (see http://spatialreference.org/ref/epsg/)
Scale	Nominal scale of the data

Annex III. UNSDIT V.2 Database Model

gROADSv1 used the UN Spatial Data Infrastructure Transport (UNSDI-T) version 2 as a common data model. The following are the field codes and full names, with sample coding. It should be mentioned that owing to the variability in source data, many of the fields could not be properly coded.

Field	Full Name	Type	Coding
OBJECTID		Object ID	
Shape		Geometry	
SourceID	Source ID	String	
Picture	Picture	Blob	
Exs	Existence Category	Long Integer	1=Definite, 2=Doubtful, 0=Unspecified
Notes	Notes	String	
RoadID	Road ID	String	
ONme	Official Road Name	String	
RteNme	Route Name	String	
NtlClass	National Inventory Road Class	String	
FClass	Functional Class	Long Integer	1=Highway, 2=Primary, 3=Secondary, 4=Tertiary, 5=Local/ Urban, 6=Trail, 7=Private, 0=Unspecified
Crgway	Carriageways	Long Integer	1=Single, 2=Dual, 0=Unspecified
NumLanes	Number of lanes	Long Integer	
LneWidthM	Lane Width (m)	Double	
RdWidthM	Road Width (m)	Double	
AxleLoadMT	Maximum Axle Loading (MT)	Double	
TotLoadMT	Maxium Total Loading (MT)	Double	
SrfTpe	Surface Type	Long Integer	1=Paved, 2=Gravel, 3=Dirt/Sand, 4=Steel, 5=Wood, 6=Grass, 0=Unspecified
SrfCond	Surface Condition	Long Integer	1=Rough (<40kph), 2=Smooth (>40kph), 3=Snow/Ice, 4=Mud, 0=Unspecified
SrfPrep	Surface Preparation	Long Integer	1=Natural Compaction, 2=Traffic Compaction, 3=Engineered Compaction, 4=Uncompacted, 0=Unspecified

Field	Full Name	Type	Coding
IsSeasonal	Affected by Season	Long Integer	1=Yes, 2=No, 0=Unspecified
CurntPrac	Current Road Practicability	Long Integer	1=Non-motorized, 2=Motorbike, 3=4WD <3.5MT, 4=Light Truck <10MT, 5=Heavy Truck <20MT, 6 Truck + Trailer >20MT, 0=Unspecified
GdWthrPrac	Good Weather Road Practicability	Long Integer	1=Non-motorized, 2=Motorbike, 3=4WD <3.5MT, 4=Light Truck <10MT, 5=Heavy Truck <20MT, 6 Truck + Trailer >20MT, 0=Unspecified
BdWthrPrac	Bad Weather Road Practicability	Long Integer	1=Non-motorized, 2=Motorbike, 3=4WD <3.5MT, 4=Light Truck <10MT, 5=Heavy Truck <20MT, 6 Truck + Trailer >20MT, 0=Unspecified
SpeedLimit	Speed Limit (Km/hr)	Long Integer	
CurntSpeed	Current Average Speed	Long Integer	
GnralSpeed	General Average Speed	Long Integer	
IsUndrCstr	Is under Construction / Repairs	Long Integer	1=Yes, 2=No, 0=Unspecified
CstWrkETC	Construction Work Est. Completion Date	Date	
GradDeg	Gradient (degrees)	Double	
Sec	Road Security Category	Long Integer	1=Category A (low risk), 2=Category B (low to medium risk), 3=Category C (medium to high risk), 4=Category D (high risk), 5=Category E (critical risk), 0=Unspecified
HasShouldr	Has Shoulder	Long Integer	1=Yes, 2=No, 0=Unspecified
HasSidewalk	Has Sidewalk	Long Integer	1=Yes, 2=No, 0=Unspecified
DrivSide	Driving Side	Long Integer	1=Left, 2=Right, 0=Unspecified
IsElevated	Is elevated / suspended above ground or water surface	Long Integer	1=Yes, 2=No, 0=Unspecified
HasMedian	Has Median	Long Integer	1=Yes, 2=No, 0=Unspecified
OpStatus	Operational Status	Long Integer	1=Open, 2=Restricted, 3=Closed, 4=Abandonned/Disused, 0=Unspecified
Shape_Length	Length of segment	Double	
Length_KM	Length of segment in kilometers	Double	