A review of roads data development methodologies

Global Roads Open Access Data Set (gROADS), a project of the CODATA Global Roads Data Development Working Group

Alex de Sherbinin, CIESIN, Columbia University
Taro Ubukawa, Geospatial Information Authority of Japan and ISCGM
Harlan Onsrud, University of Maine and GSDI

Presentation to the 12th Annual Meeting of the Global Spatial Data Infrastructure Association (GSDI-12), 18-22 October 2010, Singapore

International Council for Science : Committee on Data for Science and Technology

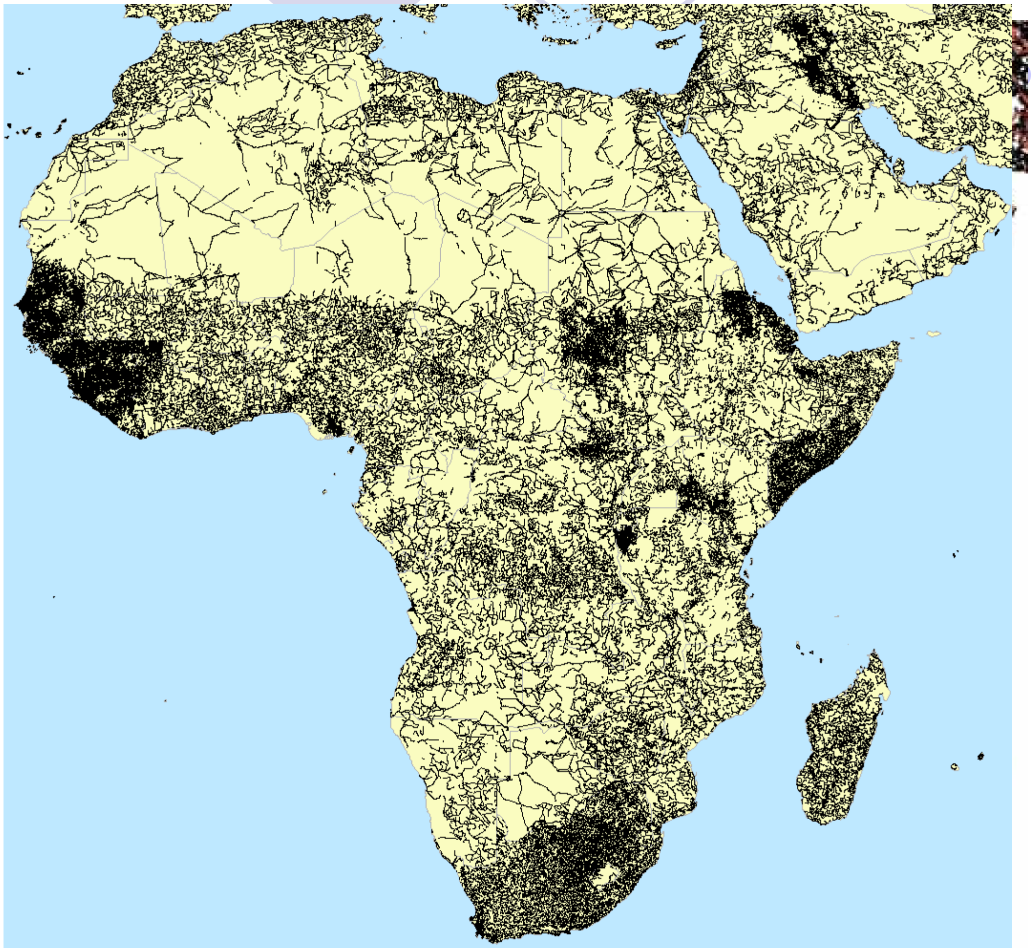
Outline

- Global Roads Open Access Data Set (gROADS)
- Methods to develop roads data
  - Summary of approach
  - Pros and cons
- Investing in open data

This presentation is based on: T. Ubukawa, A. de Sherbinin, A. Nelson, H. Onsrud, K. Payne, O. Cottray, and M. Maron. *Forthcoming.* **A review of roads data development methodologies.**

Acknowledgments: The NASA Socioeconomic Data and Applications Center underwrote the development of gROADS. CODATA paid for travel to this conference.

gROADS goal



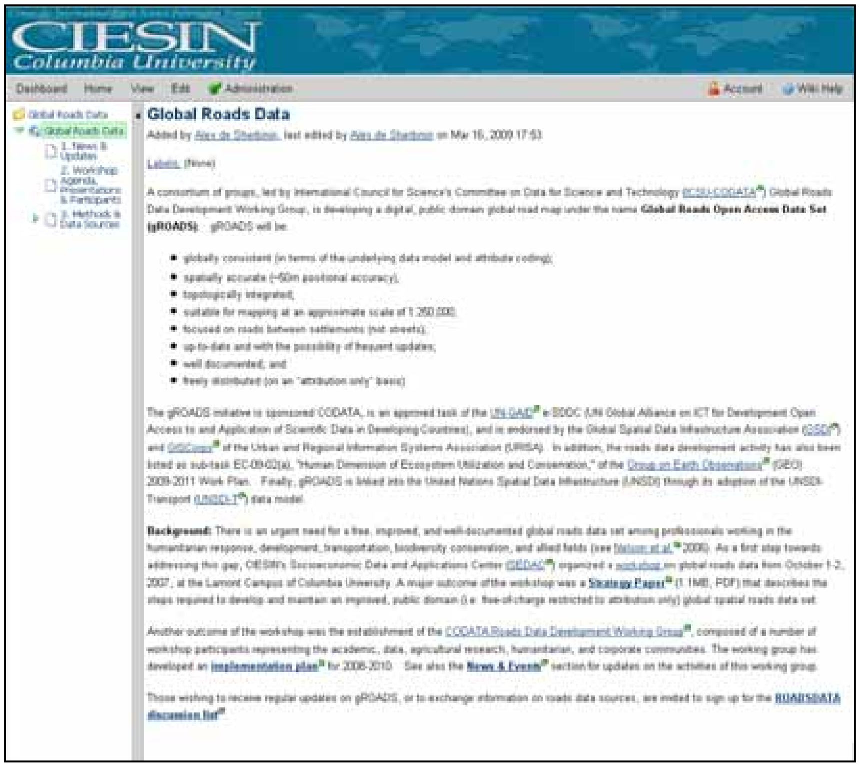
To develop a global roads open access data set (gROADS) that is:

1. globally consistent model (UNSDI-T v.2)
2. spatially accurate (~50m positional accuracy)
3. topologically integrated
4. focused on roads between settlements (not streets)
5. up-to-date and with the possibility of frequent updates
6. well documented
7. freely distributed (on attribution only basis)

Intei International Council for Science : Committee on Data for Science and Technology

gROADS releases

Visit www.groads.org


CIESIN
Columbia University

Dashboard Home View Edit Administration Account Wiki Help

Global Roads Data
Global Roads Data
1. News & Updates
2. Workshop Agenda, Presentations & Participants
3. Methods & Data Sources

Global Roads Data

Added by Alex de Sherbinin, last edited by Alex de Sherbinin on Mar 16, 2009 17:53

Labels: (None)

A consortium of groups, led by International Council for Science's Committee on Data for Science and Technology (ICSU-CODATA) Global Roads Data Development Working Group, is developing a digital, public domain global road map under the name **Global Roads Open Access Data Set (gROADS)**. gROADS will be:

- globally consistent (in terms of the underlying data model and attribute coding);
- spatially accurate (~50m positional accuracy);
- topologically integrated;
- suitable for mapping at an approximate scale of 1:250,000;
- focused on roads between settlements (not streets);
- up-to-date and with the possibility of frequent updates;
- well documented; and
- freely distributed (on an "attribution only" basis)

The gROADS initiative is sponsored CODATA, is an approved task of the UN-GAID e-SDDC (UN Global Alliance on ICT for Development Open Access to and Application of Scientific Data in Developing Countries), and is endorsed by the Global Spatial Data Infrastructure Association (GSDI) and GISCorps of the Urban and Regional Information Systems Association (URISA). In addition, the roads data development activity has also been listed as sub-task EC-09-02(a), "Human Dimension of Ecosystem Utilization and Conservation," of the Group on Earth Observations (GEO) 2009-2011 Work Plan. Finally, gROADS is linked into the United Nations Spatial Data Infrastructure (UNSDI) through its adoption of the UNSDI-Transport (UNSDI-T) data model.

Background: There is an urgent need for a free, improved, and well-documented global roads data set among professionals working in the humanitarian response, development, transportation, biodiversity conservation, and allied fields (see Nelson et al. 2006). As a first step towards addressing this gap, CIESIN's Socioeconomic Data and Applications Center (SEDAC) organized a workshop on global roads data from October 1-2, 2007, at the Lamont Campus of Columbia University. A major outcome of the workshop was a **Strategy Paper** (1.1MB, PDF) that describes the steps required to develop and maintain an improved, public domain (i.e. free-of-charge restricted to attribution only) global spatial roads data set.

Another outcome of the workshop was the establishment of the CODATA Roads Data Development Working Group, composed of a number of workshop participants representing the academic, data, agricultural research, humanitarian, and corporate communities. The working group has developed an **implementation plan** for 2008-2010. See also the **News & Events** section for updates on the activities of this working group.

Those wishing to receive regular updates on gROADS, or to exchange information on roads data sources, are invited to sign up for the **ROADSDATA discussion list**.

gROADS Catalog v.1

1. Catalog of 360+ national and regional data sets

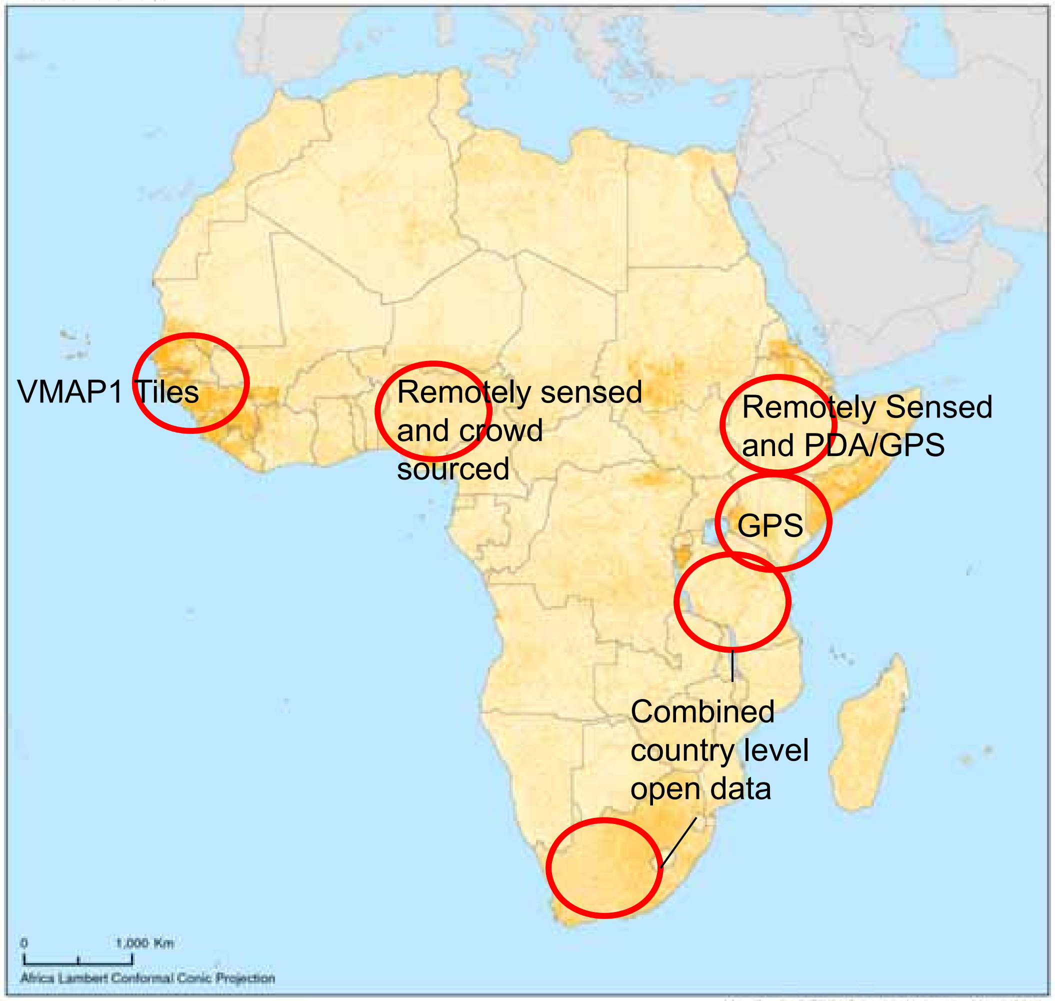
gROADS v.1 roads data set release in 2013 with data from:

1. University of Georgia's Information Technology Outreach Services (ITOS), compiled for the UN's Geographic Information Support Team (GIST)
2. Netherland's PBL Global Roads Inventory Project (GRIP)
3. Open Street Map
4. CIESIN data development activities

METHODS TO DEVELOP ROADS DATA

Global Roads Open Access Data Set, Version 1 (gROADSv1): Africa

Global Roads



Map Credit: CIESIN Columbia University, March 2013.

The Global Roads Open Access Data Set, Version 1 (gROADSv1) was developed under the auspices of the CODATA Global Roads Data Development Task Group. The data set combines the best available 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. Because the data are compiled from multiple sources, the dates for road network representations range from the 1980s to 2010, depending on the country, and spatial accuracy varies. National borders are provided for reference purposes only, and CIESIN and its sponsors do not take a position with regards to the representation of boundaries.

Center for International Earth Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY

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

1. Compiling best available public domain data
2. Remote sensing derived
   - Digitizing from moderate resolution imagery
   - Automated and semi-automated extraction
3. Field based mapping
   - PDA / GPS roads data development
   - Passive roads data collection using GPS
4. Crowd sourcing / heads up digitizing

1. COMPILING BEST AVAILABLE DATA

gROADS: Compiling best available data

- Starting with University of Georgia/ITOS GIST data
  - A combination of VMAP0, VMAP1 and national sources
- CIESIN:
  - edited topology for many countries
  - contributed new data by cleaning GPS derived data, remote sensing derived data, country data (U.S.), and Global Map (ISCGM) data
  - merged data from two or more sources for some countries (e.g., DCW and Africover for Tanzania)
  - conducted pilot projects that provided new data for Ethiopia

Steps for compiling roads data

- Evaluation of candidate data sets
  - Spatial accuracy
    - Use Google Earth imagery to calculate the RMSE
  - Coverage (total kms) and road level inclusion
  - Data restrictions and licensing such as “research-only”
- Selection of best data set according to above criteria
- Data cleaning
  - Topology cleaning
  - Connecting roads at national boundaries

Methods documentation

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

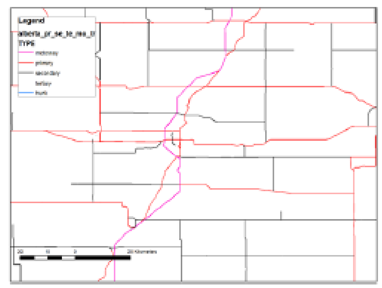
Table of Contents

Suggested citation: Center for International Earth Science Information Network (CIESIN). 2013. Methods Used in the Development of the Global Roads Open Access Data Set (gROADS), version 1. Palisades NY: NASA Socioeconomic Data and Applications Center (SEDAC).

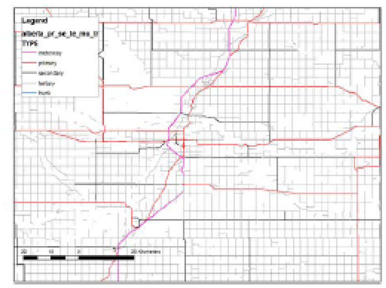
The following steps were applied to simplify the data.

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.

- 0_merged_classified
  - motorway
  - motorwaylink
  - primary
  - primarylink
  - secondary
  - secondarylink
  - trunk
  - trunklink

2. REMOTE SENSING DERIVED

WRI road mapping in Central Africa

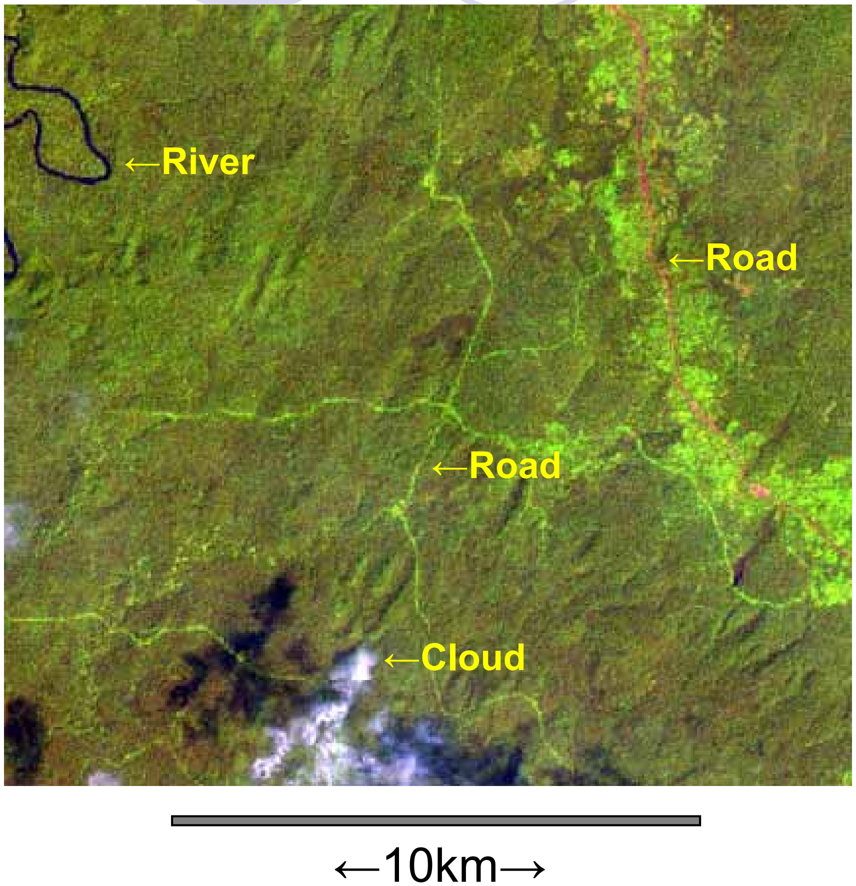
Digitization from satellite imagery

- Image types: Landsat ETM+, ASTER, ALOS-AVNIR, DMCii, SPOT 5
- Process imagery to enhance road distinction (e.g. Landsat 5-4-3 combination)
- Roads digitized manually using Erdas Imagine or ArcGIS software at 1:30,000 with 30m “snapping” tolerance
- Begin with oldest image of an area, in order to track progression
- Attributes applied to road segments from existing datsets using spatial join or according to standardized classification developed by WRI and collaborators
- Field verification of a select percentage of roads using GPS

- Advantages: cover large areas relatively cheaply (if images area available) and using standard methods; can get acceptable accuracy considering scale
- Disadvantages: many forested areas have heavy cloud cover and thus visible imagery difficult to acquire; misclassfication of road attributes if not field verified

Source: de Sherbinin, Yetman, and Steil. Presentation at GSDI 2010, Singapore

Landsat Imagery (5-4-3 band combination)



Some results:

- For Cameroon – 40,044 km mapped at an average accuracy of ±42m compared to GPS verification
- At first pass in 2003, nearly 10% of roads were considered to be potentially “illegal” – another pass in 2008 found suspect road building to be significantly reduced (remote monitoring and enforcement had a measurable effect?)
- Have produced roads datasets for Cameroon, CAR, Congo, DRC and Gabon

Source: de Sherbinin, Yetman, and Steil. Presentation at GSDI 2010, Singapore

Automated and semi-automated methods

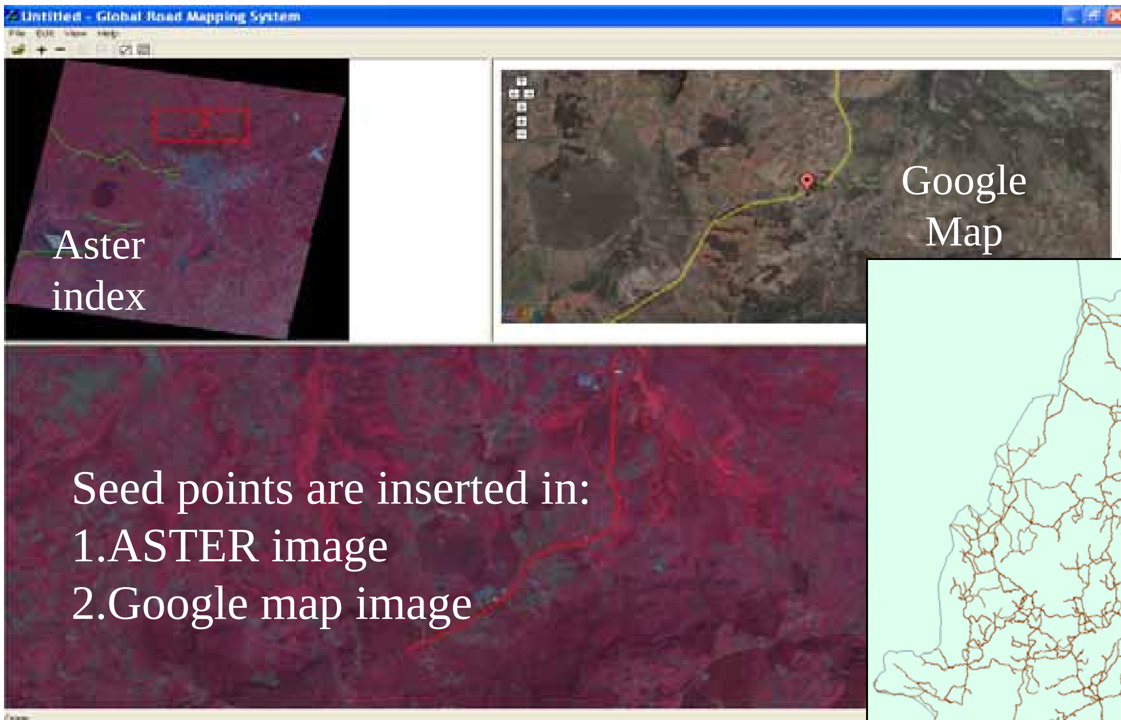
- Seeding and tracking
- Snake algorithm
- Segmentation and classification
- Multi-spectral analysis
- Edge detection

Digitizing from ASTER imagery (Ethiopia)

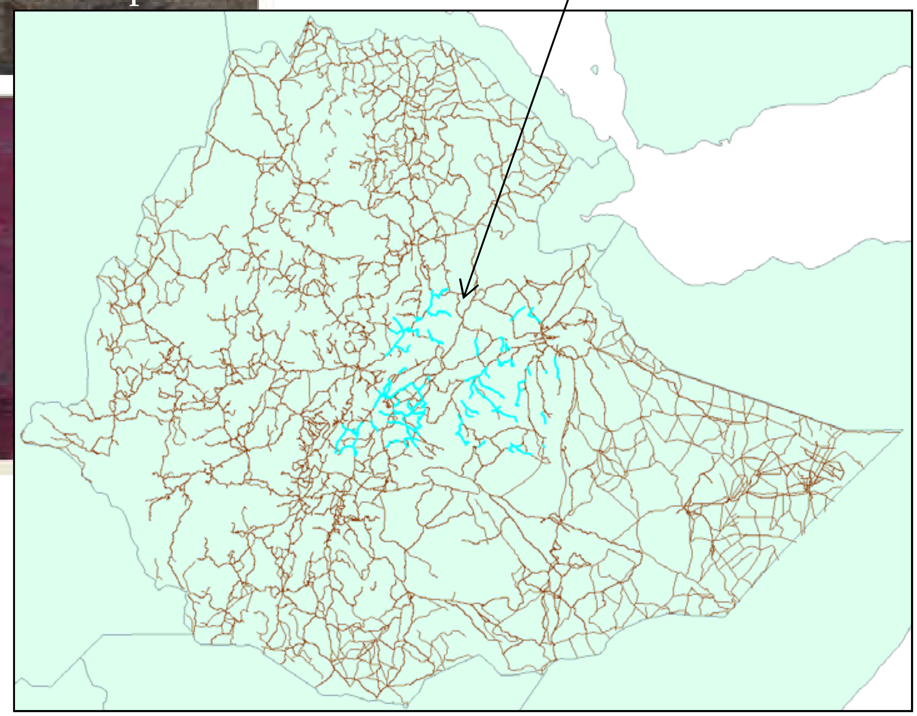
Funded by NASA-SERVIR

- Pilot tested a semi-automated road extraction tool (GRMT) developed by the Center for Spatial Information Science (CSIS) at University of Tokyo
- Used ASTER imagery (15m resolution, 60x60km footprint)
- The alpha version of the software was comparable to manual digitizing
  - line following algorithm underperforms due to similar spectral signature for roads and surrounding land covers
- Average time per scene ~8hrs (depending on rural vs. urban and number of roads)

Global Road Mapping Tool details



ASTER derived roads digitized in aqua-blue

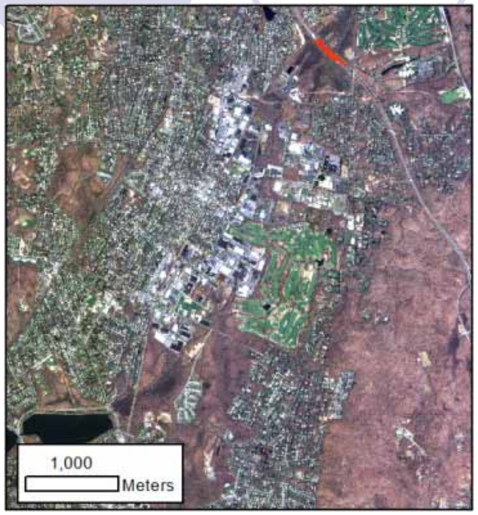


"Snake Algorithm" then uses the seed points and the image spectral signature to place the line.

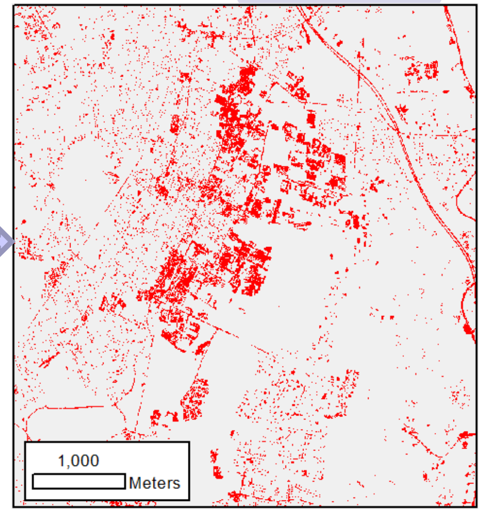
Similar to RoadTracker Commercial software

Semi-automated extraction with FeatureAnalyst and ArcScan

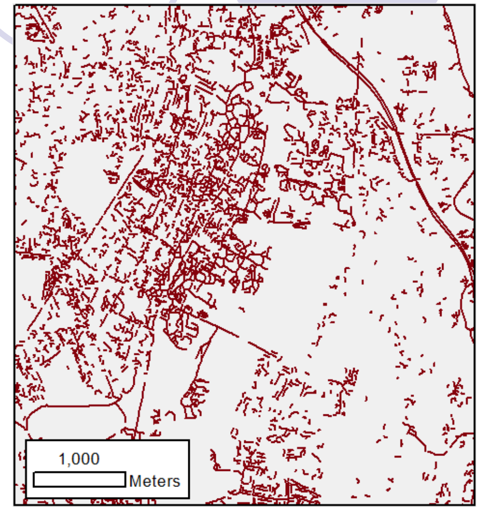
Trial with 4 m resolution OrbView3



Training data

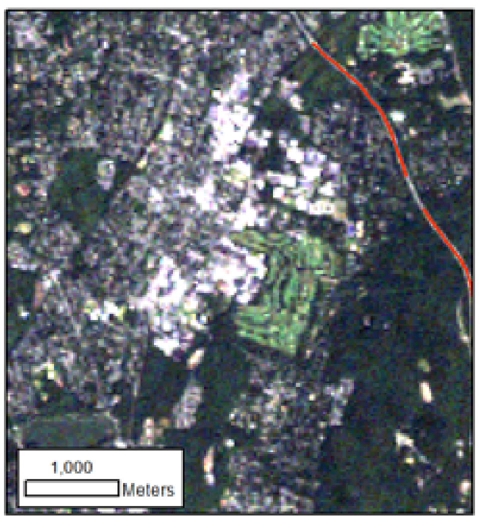


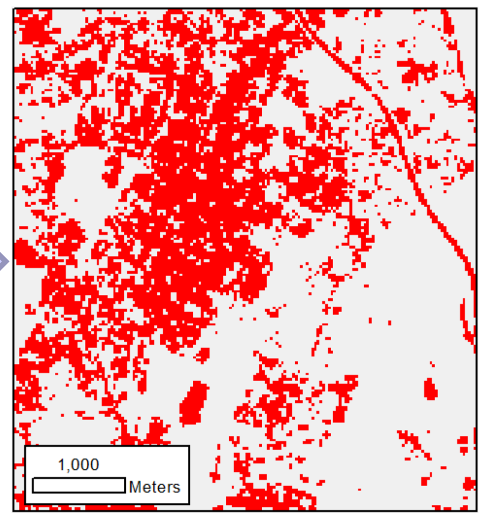
Binary raster

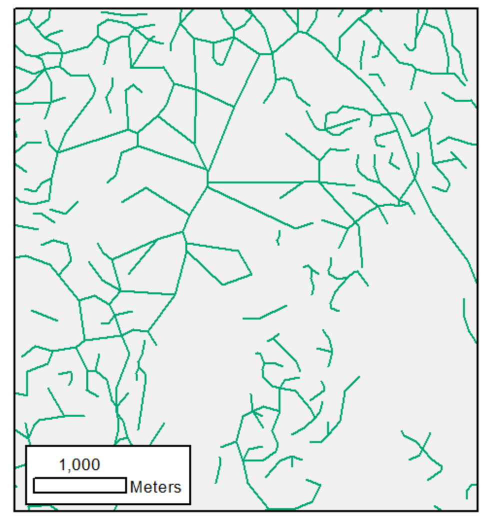


Vector

Trial with 30 m resolution Landsat







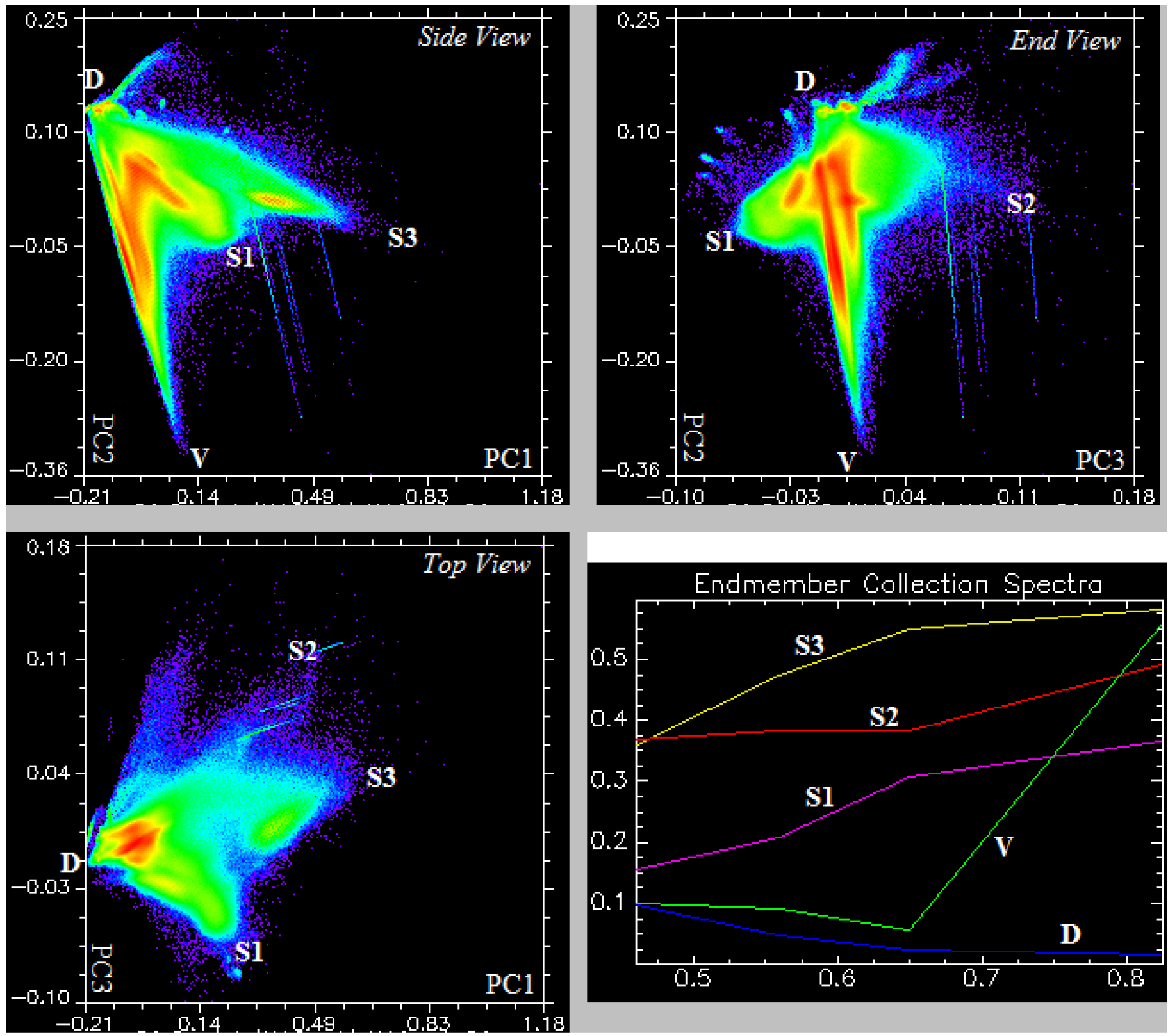
A tradeoff: resolution & footprint

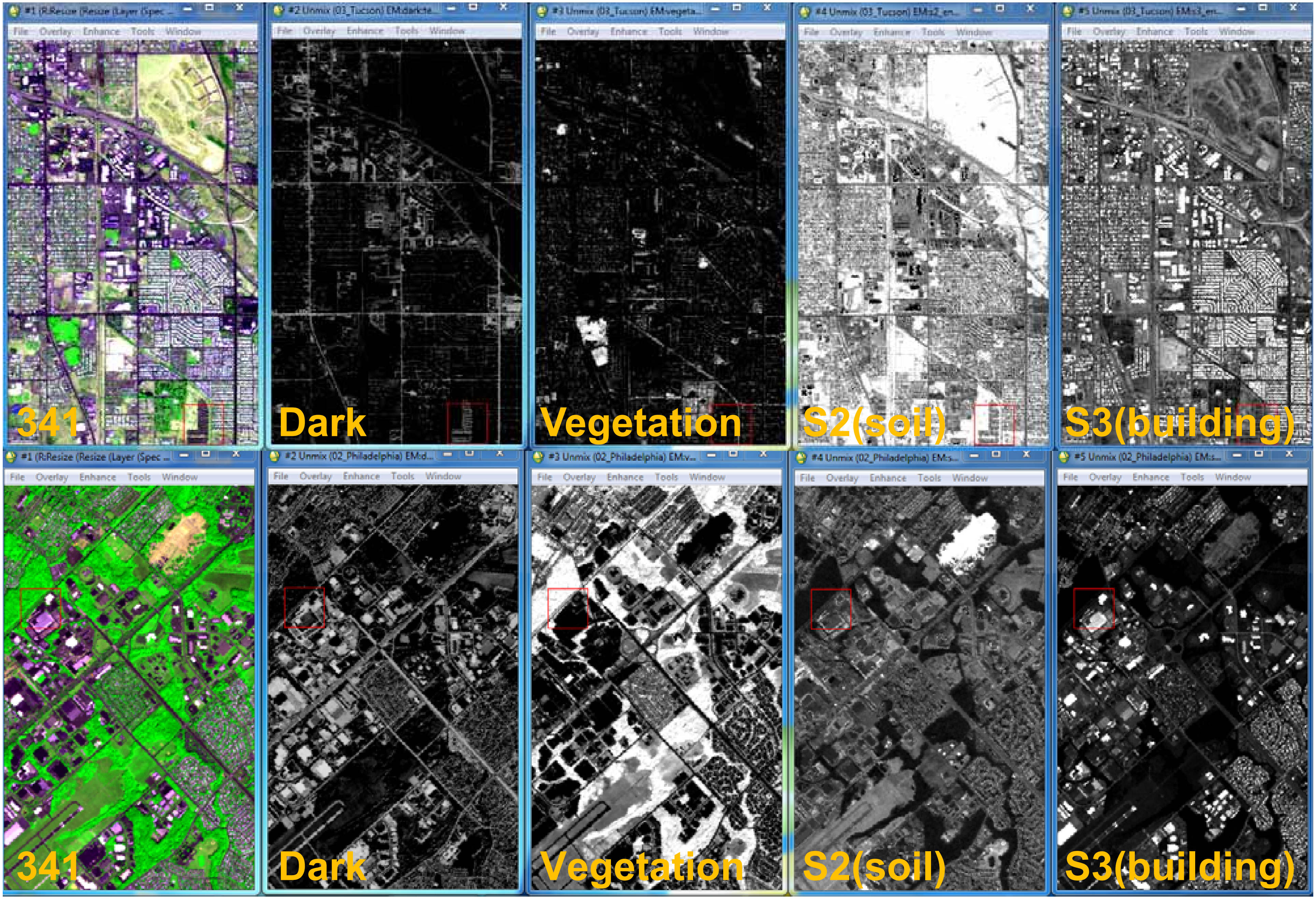
- **ALOS/AVNIR**
  - 4 bands, 10 meters spatial Res. (at nadir)
  - Swath: 70 km (at Nadir)

Spectral Mixture Analysis

Three dimensional distribution

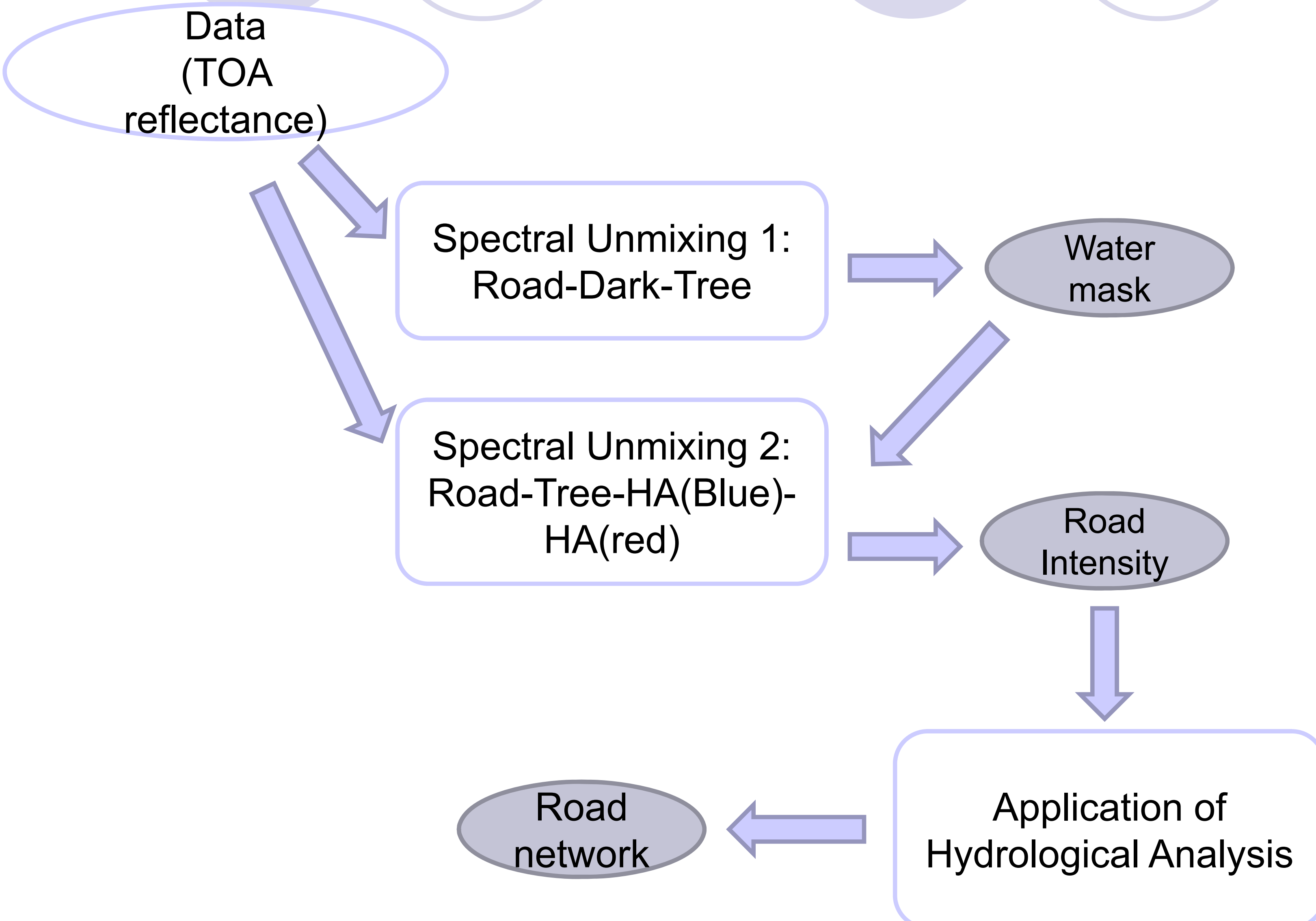
End members:
Dark, Vegetation and Substrate





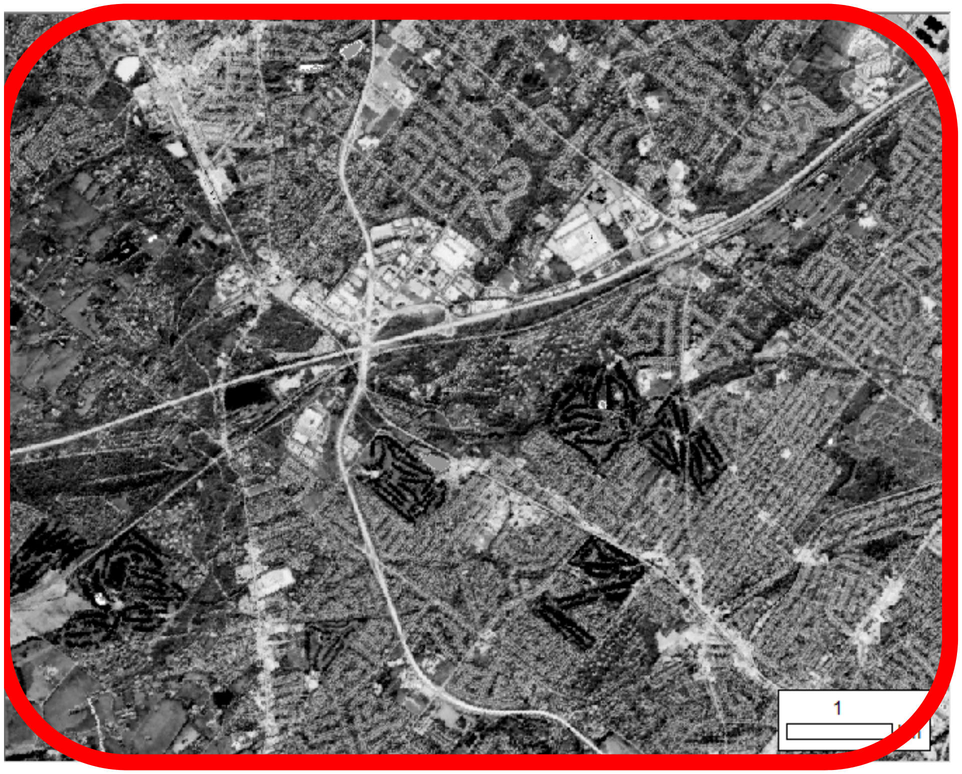
If you understand mixing, you can unmix it

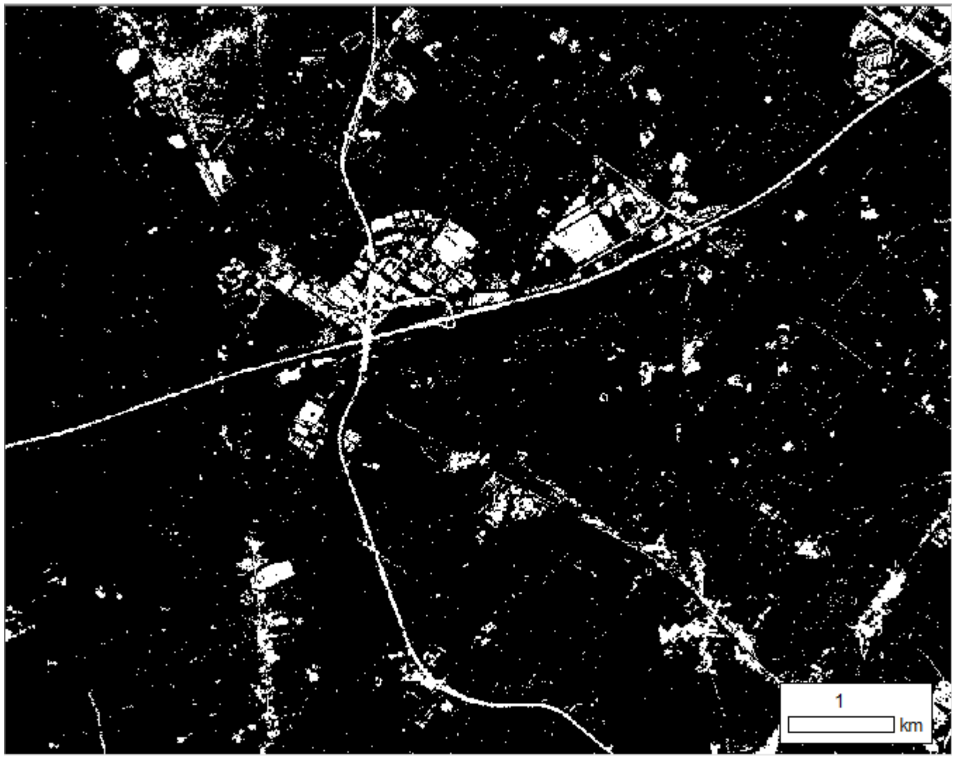
Work flow (Over view)



Obtained Road Intensity (or fraction)

- Intensity (graduated)
- Binary data (using certain thresholds)



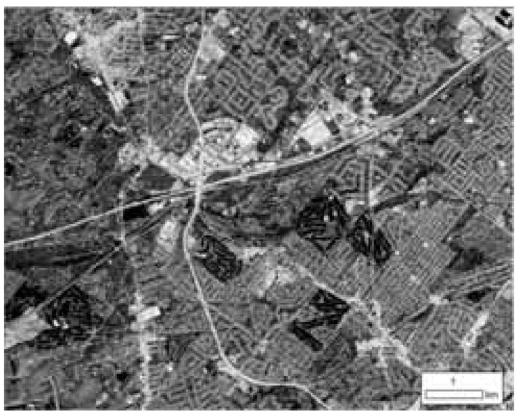


For extracting vector from raster, some methods use binary imagery (e.g. ArcScan)

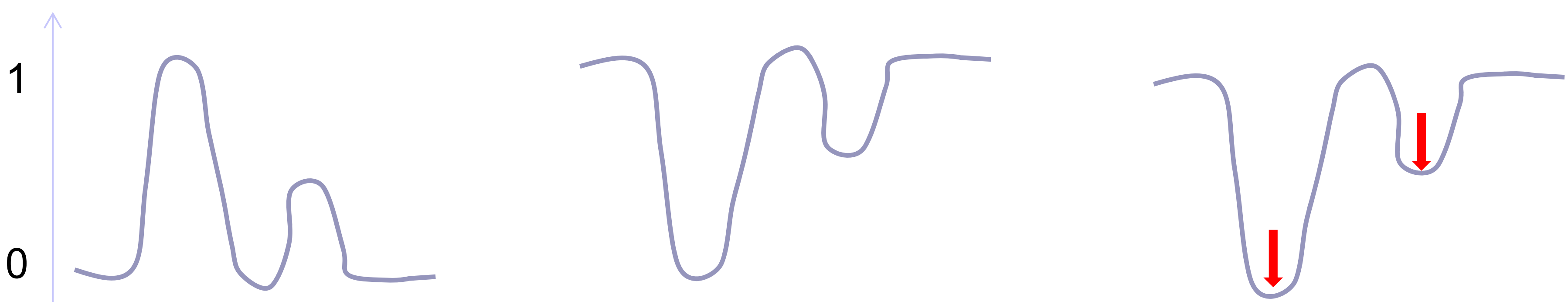
But, binary raster misses narrow roads.

Work flow 3 (hydrological method)

- Using a flow calculation

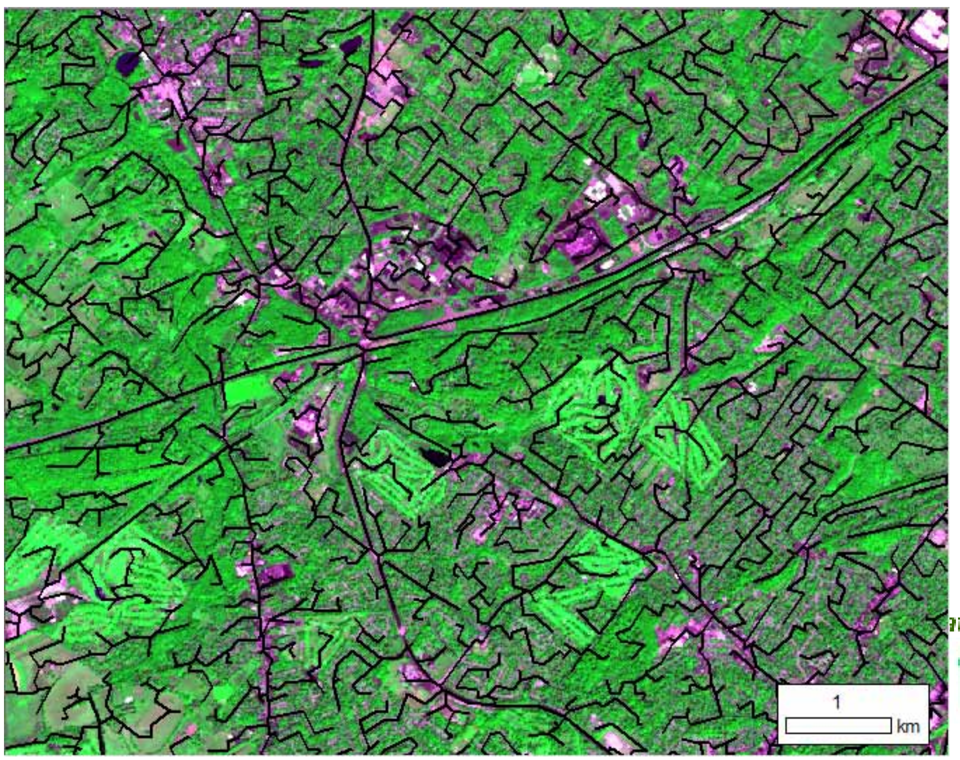


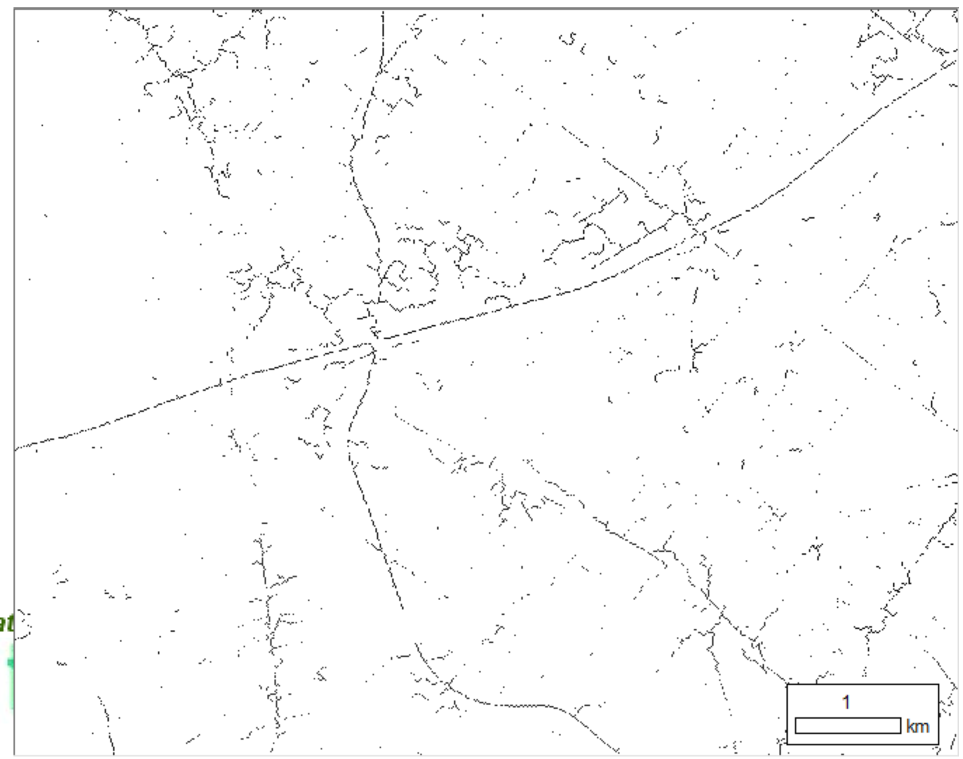
*(-1) → Opposite of intensity → Hydrological analysis → Streams



Results

- Relatively good road extraction, but also some false positives
- If the surface type is different from the original definition, roads are not detected





3. FIELD BASED MAPPING

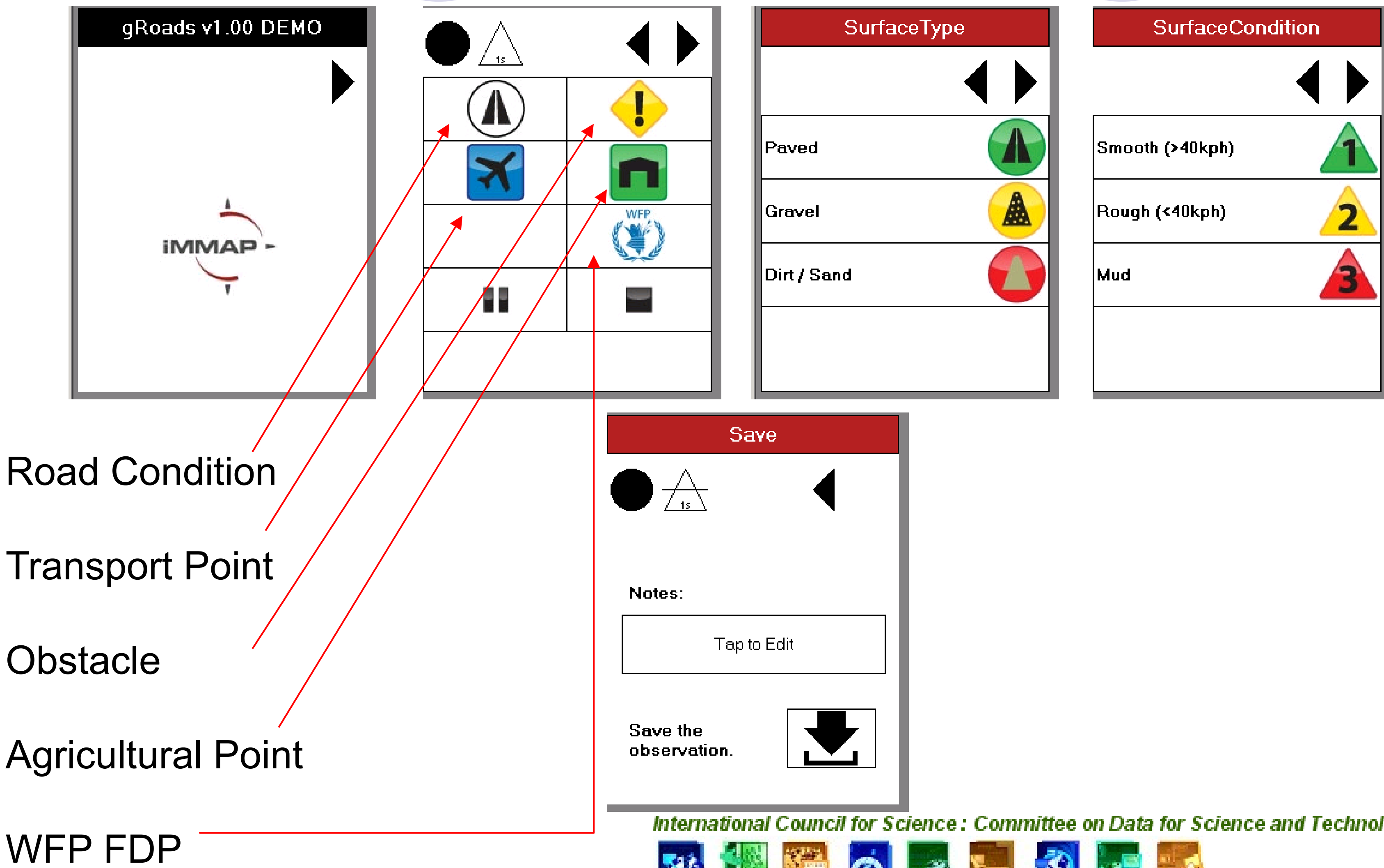
Using GPS enabled PDAs (Ethiopia)

Funded by Gates Foundation/AGCommons

- Output 1: A UNSDI-T compliant roads data set from GPS tracks, along with agricultural features of interest
- Output 2: Software for a PDA tool that includes all fields of the UNSDI-T data model (based on *Cybertracker*) ⟶

- Approach:
  - Incidental data collection: Engage third parties (WFP field teams) who are conducting missions to hard-to-reach places to collect data
  - Active data collection: Pay truck drivers or higher cars to collect data for regions that are missing (not done)
  - Incorporate data from third party sources where possible
- IMMAP hired a local representative to train WFP staff and manage data collection
- RCMRD collaborated on field campaigns
- CIESIN completed data cleaning and compilation

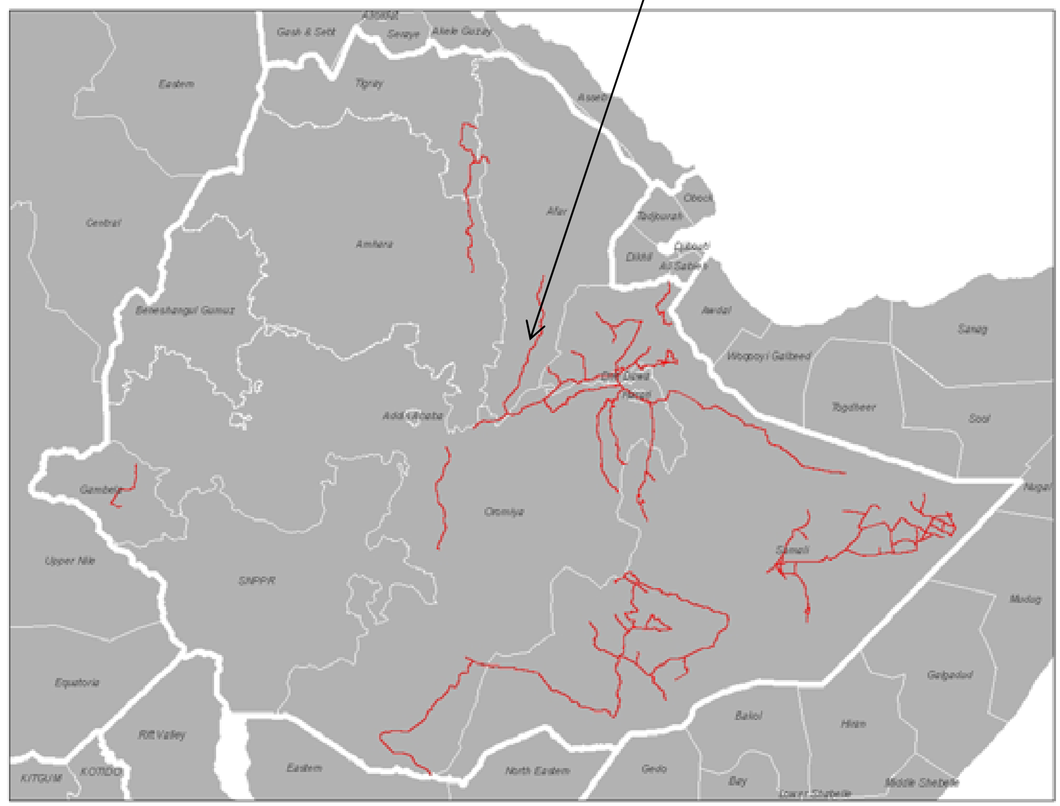
Customized Cybertracker Tool



Challenges ecountered

- Resistance on the part of WFP field teams to additional duties
- Redundancy in routes covered
  - no incentive to take longer and less secure routes in order to cover additional roads
- Coding of roads traveled more than once was sometimes inconsistent
- Collection using truckers not possible because of illiteracy

PDA derived roads in red



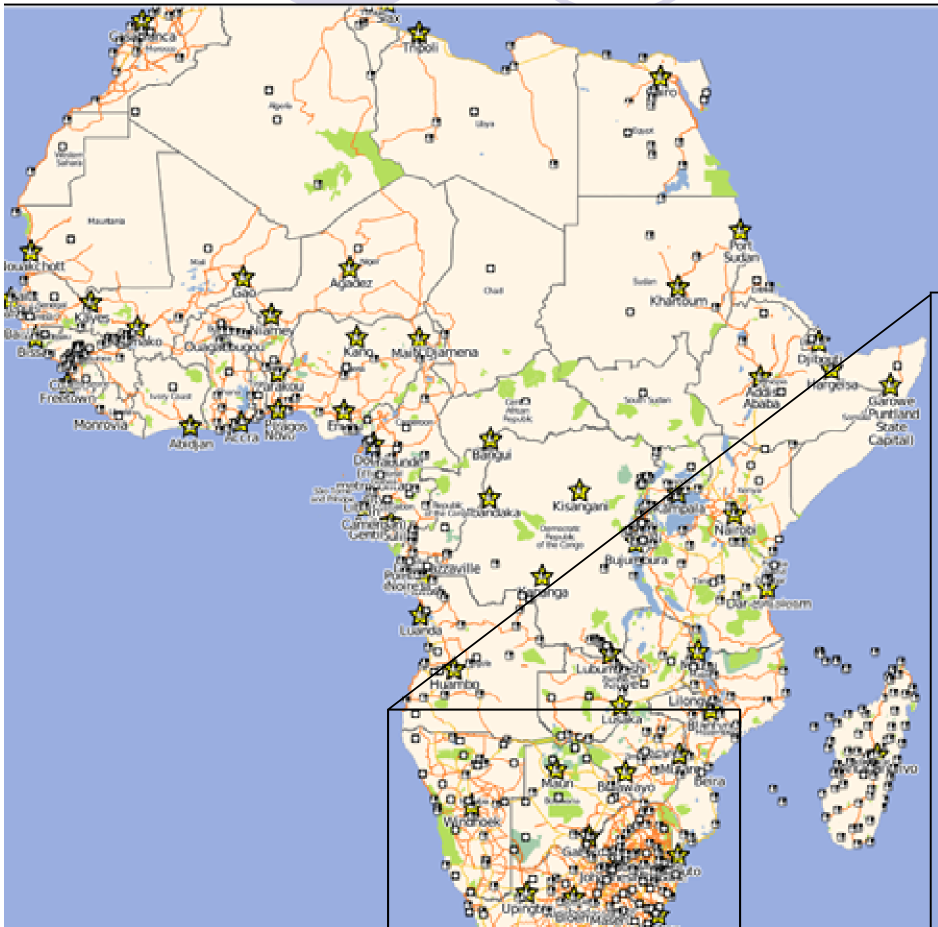
Passive Collection with GPS

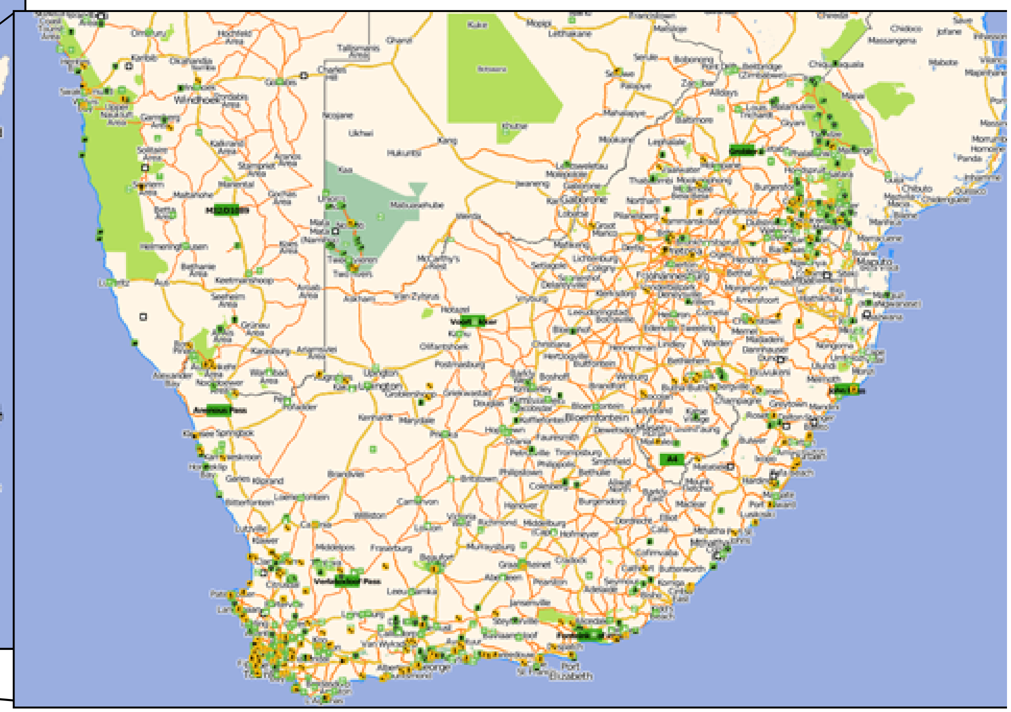
- Approach pioneered by Tracks4Africa
- Method: Amass large quantities of GPS tracks from recreational travelers and assimilate these tracks to:
  - Create road centerlines by averaging tracks
  - Infer road quality from average travel speeds

- Advantages: low cost, spatially accurate data, with features of interest to recreational travelers
- Disadvantages: Additional attribute information (road name, road surface type, etc.) not generally collected

Tracks4Africa

In the year 2005 Tracks4Africa started to sell T4A GPS Maps to people outside the Tracks4Africa community of travelers who contributes the data. The company found a unique balance between crowd sourced data, community driven development of our products and a sustainable commercial model.





4. CROWD SOURCING / HEADS UP DIGITIZING

Crowd sourcing

Open Street Map

- 900,000+ registered users
- Roads manually digitized from:
  - GPS tracks collected by volunteers
  - Base imagery from Bing
- Humanitarian OSM Team
  - Mobilized quick response for Haiti in Jan 2010

Google Map Maker

- Unknown number of active volunteers
- Roads manually digitized from Google Maps base imagery

Crowd sourcing challenges

- Quality control
  - OSM has relies on community norms
  - Map Maker has volunteer moderators who review edits and provide feedback to other volunteers
    - One's "trust" within the program determines “weight” of moderation
- Licensing
  - OSM adopted the CC Share Alike license and migrated to Open DB License (a viral license)
    - Extensive documentation on permitted and unpermitted uses (http://wiki.openstreetmap.org/wiki/OpenStreetMap_License)
  - Map Maker is for non-commercial use only
- Depends on orthorectified high res imagery
  - Favors more urbanized regions where high res imagery exist
  - Imagery themselves have positional inaccuracies

Positional accuracy by provider*

Data set	Points # (Scene #)	RMSE (meters)	Mean Error (meters)	SD (meters)	Range (meters)
Google	140 (10)	**8.2**	7.0	4.2	(0.5-20.1)
Bing	137 (10)	**7.9**	7.0	3.6	(1.6-22.1)
OSM	116 (10)	**11.1**	8.8	6.9	(0.2-55.1)

* Relative to ALOS/PRISM (1b2),which expected accuracy is 6.1 m (orthorectified)

Source: Ubukawa, 2013. Available at www.groads.org news page.

INVESTMENTS IN OPEN DATA

Two Models

Old School

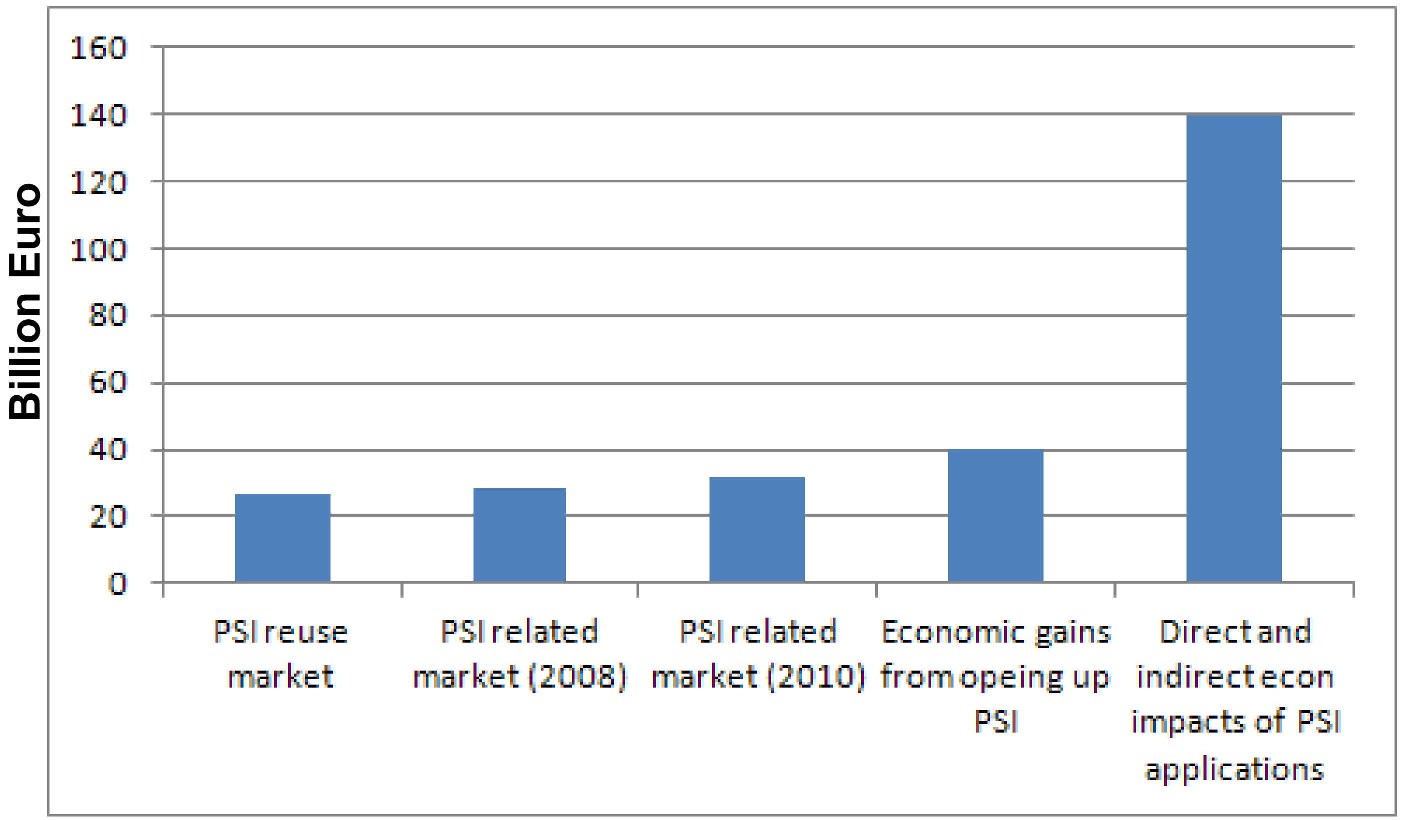
- Information has a price
- Information is power
- User pays for data
- User unable to redistribute value-added products

Open Data

- Information is provided free of charge and without copyright restriction
- Society is better informed
- Research is improved
- Lower costs to industry
- Information sector is spawned and grows
- Taxes on this sector fund data creation

Investing in Open Data

Economic benefits of open public sector information (PSI) in the EU27



Source: Vickery (2011), “Review of Recent Studies on PSI Re-Use and Related market Developments”

International Council for Science : Committee on Data for Science and Technology

Want to learn more or help out?

- Visit www.groads.org
- Sign up for the ROADSDATA discussion list to exchange information on data sets
- Send us your data!
  - Contact Alex de Sherbinin at amd155@columbia.edu