

CODATA Global Roads Data Development Working Group Implementation Plan: 2008-2010

Produced by Alex de Sherbinin and Olivier Cottray,¹
Working Group co-chairs, with contributions
by Working Group members

Introduction

The CODATA Global Roads Data Development Working Group (CODATA-Roads) seeks to develop a global roads data set 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 need for such a data set, and the reasons why commercial and military developers of roads data are unlikely to meet this need, is documented in the CODATA-Roads strategy paper,² so this document does not describe the need or potential applications of gROADS, nor does it provide details on the CODATA Roads Working Group.³ The gROADS initiative is sponsored by ICSU’s Committee on Data for Science and Technology (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). 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.⁴ Finally, gROADS is linked into the United Nations Spatial Data Infrastructure (UNSDI) through its adoption of version 2 of the UNSDI-Transport (UNSDI-T v.2) data model.⁵

CODATA-Roads will work closely with existing global road mapping efforts such as OpenStreetMap (OSM) and the Global Road Infrastructure Project (GRIP),⁶ and will

¹ Contacts: [adesherbinin AT ciesin.columbia.edu](mailto:adesherbinin@ciesin.columbia.edu) and [ocottray AT immap.org](mailto:ocottray@immap.org), respectively.

² See <http://www.ciesin.columbia.edu/confluence/roads/>

³ More information on the working group can be found on the CODATA web site: <http://www.codata.org/taskgroups/WGglobalroads/>

⁴ See http://www.earthobservations.org/geoss_imp.shtml.

⁵ See <http://www.unjlc.org/mapcenter/unsdi/unsdit-v2/>

⁶ See <http://www.openstreetmap.org> and <http://mapserver.mnp.nl/website/GRIP/>

actively seek data-sharing arrangements. The three projects are highly synergistic, yet each has a slightly different niche. For example, compared to OSM, the focus of CODATA-Roads is on roads and not intra-urban streets.⁷ This focus on roads is consistent with the data needs expressed by applied users in the humanitarian, development, and conservation communities. For gROADS, greatest effort will be made to fill data gaps where currently the OSM community is least well represented – i.e., the poorest developing countries, which are most vulnerable to humanitarian emergencies. Also, CODATA-Roads is well integrated into existing UN Spatial Data Infrastructure (SDI) initiatives, and will involve considerable collaboration with UN agencies and mechanisms such as the UN Geographic Information Working Group (UNGIWG).⁸ A major focus of collaboration will be the UN agencies responsible for humanitarian operations, in particular UNOSAT and the UN-Joint Logistics Centre (JLC), which was responsible for developing the UNSDI-T data model, and which provides GIS and logistical support to UN agencies responding to humanitarian crises.⁹

In terms of update capabilities, CODATA-Roads is unlikely to ever be able to match the size of OSM's impressive user base, which now numbers close to 70,000 contributors. OSM is the first and largest wiki mapping environment, and is dedicated to open distribution of its mapping products.¹⁰ Thus, subject to availability of funds, CODATA-Roads will seek to provide support to OSM in a number of potential areas: programming, purchase of additional servers and bandwidth, and community mapping/capacity building activities in data poor regions.

The focus of GRIP is largely roads data for the internal modeling needs of the Netherlands Environmental Assessment Agency (PBL), which runs the IMAGE and GLOBIO models. GRIP represents a mix of copyrighted commercial data and data from other sources with modest documentation and validation, and the data set is visible on the web but not yet publicly disseminated. By contrast, CODATA-Roads will carefully document all source data sets included in gROADS and seek to perform a quantitative and qualitative validation of each candidate data set. The focus is on providing a

⁷ OSM includes roads and streets, with arguably more effort going into local area street mapping. In gROADS, streets will not necessarily be excluded where they can be included at no extra cost, but for most global and regional scale analyses, and for many humanitarian applications, they are not required and their inclusion can result in large file sizes.

⁸ UNGIWG representatives were active participants at the October 2007 workshop that launched the CODATA-Roads initiative, and the Working Group was represented at the UNGIWG 9th Plenary Meeting in Vienna, Austria (4-7 November 2008).

⁹ The UN-JLC is in the process of being incorporated into the UN Logistics Cluster of the World Food Programme. Once it is incorporated CODATA Roads will seek a partnership with the Logistics Cluster. Under this partnership, roads data that are obtained and field validated by the Cluster will be cataloged, archived, and evaluated for inclusion in gROADS. The gROADS will thus be a repository for roads data, and will seek permissions for selected data sets to be included in gROADS with no restrictions. Any restrictions that may be applied will be clearly documented in gROADS documentation. The growing collection of data, in turn, will be made available to the humanitarian community at no cost.

¹⁰ Contributors to OSM agree to have the data distributed via a Creative Commons Share Alike license, which requires that all derivative products be distributed under the same license. The OSM is exploring adoption of the Open Data License (ODL), but this is subject to debate within the OSM community.

publically disseminated data set without restrictions for applied, research or commercial users.¹¹

The project has two major components which will be carried out roughly in succession, though elements may be carried out simultaneously. The **first step** is to compile a composite global road coverage based upon best available sources. The **second step** is to develop a wiki mapping environment that permits the wider community, including national and international agencies, to edit the base map, thereby contributing new data and serving to keep gROADS up-to-date.¹² The implementation plan for the remainder of 2008 through 2010 is outlined below. The work will be coordinated by the Center for International Earth Science Information Network (CIESIN) of the Earth Institute at Columbia University under the auspices of CODATA-Roads and UN-GAID e-SDDC.

I. Compiling the baseline gROADS

1. New Roads Data Development Using Data Extraction Methods (Sept 2008-ongoing)

CODATA-Roads will assess, and where appropriate, coordinate, the work of a number of groups that are working on the semi-automated extraction of roads data from orthorectified satellite imagery or digitized topographic maps. Wherever possible, these data streams will be integrated into gROADS. There are a number of such initiatives underway at present:

- A. CIESIN is coordinating a collaborative project involving the Center for Spatial Information Science of the University of Tokyo, Japan's National Institute of Advanced Industrial Science and Technology (AIST), and the Regional Center for Mapping and Rural Development (RCMRD) in East Africa.¹³ The project is pilot testing the use of a software package developed by the University of Tokyo to semi-automate the extraction of roads data from satellite imagery, focusing on the highlands of Ethiopia, where accurate roads data are scarce. Orthorectified mosaics of ASTER imagery are being provided by AIST. The road segments will be provided to CIESIN, which will add attribute information and ensure proper topology. The staff time for this exercise is being measured so as to be able to compare it to manual digitization. (*Lead: WG co-chair A. de Sherbinin*)
- B. Information Management and Mine Action Programs (iMMAP) in collaboration with the Information Technology Outreach Services (ITOS) of the University of Georgia, is testing a number of commercial off-the-shelf

¹¹ No restrictions will be applied to third party users who may wish to develop value-added products from the roads data set provided they do not seek to sell the original data.

¹² Unlike OSM, the editing infrastructure of gROADS will be built on Open Geospatial Consortium (OGC) specifications and open standards, and the product will be distributed via the Internet in standard formats (e.g., ESRI shapefile, etc.). See Section II below.

¹³ This work is being funded by the NASA SERVIR Project through a sub-contract with the Centro del Agua del Trópico Húmedo para América Latina y El Caribe (CATHALAC).

(COTS) software packages such as Feature Analyst with RoadTracker (Visual Learning Systems), Cognition Network Technology (Definiens), and Objective (ERDAS) to extract road features from topographical maps of North Korea. They are also comparing the time and effort to manual digitizing. (*Lead: WG co-chair O. Cottray*)

- C. The Institute for Remote Sensing Applications (IRSA) of the Chinese Academy of Sciences (CAS) is dedicating a graduate student to assist with the CODATA-roads, and also has developed methods for semi-automated roads extraction from satellite imagery. (*Lead: WG member Jinnian Wang*)
- D. As a contribution to gROADS via iMMAP, UNOSAT has agreed to submit a data needs request for two or three high priority countries under the rubric of the RESPOND project.¹⁴ The needs request will be channeled by RESPOND to companies able to digitize roads data according to gROADS specifications from satellite imagery, with a clear proviso that the data be available free and clear of any restrictions for integration into gROADS. (*Lead: WG co-chair O. Cottray*)
- E. CODATA-Roads is exploring a collaboration with GISCorps to develop data for hazard prone countries.¹⁵ Initial indications suggest that they would be interested in helping to digitize new areas using manual and semi-automated techniques. (*Lead: WG co-chair O. Cottray*)

2. Cataloging existing national and regional roads data sets (June 2008-ongoing)

With funding from the NASA Socioeconomic Data and Applications Center run by CIESIN, initial work has begun to catalog data sets for developing countries. As of October 2008, more than 140 national-level data sets have been identified. For each data set, information is compiled on the source (originator and distributor), year of publication, permissions (i.e., restrictions on use), length of the road network, spatial accuracy (see Validation section below), and attribute information (e.g. road type or quality coding), among others.

Larger collections of data, such as OSM data and any public domain data collected by GRIP, will also be included in this catalog and evaluated for possible inclusion in gROADS.

Lead: WG co-chair A. de Sherbinin

3. Validation of existing roads data (June 2008-ongoing)

Validation of data will be critical to ensure a minimal quality control for each data set that is incorporated into gROADS. In terms of spatial accuracy, CIESIN has developed the following procedures to evaluate each “candidate” roads data set against a reference

¹⁴ See <http://www.respond-int.org/respondlive/>.

¹⁵ See <http://www.giscorps.org/>

layer such as GPS tracks or high to moderately high resolution orthorectified satellite imagery (e.g., in Google Earth).¹⁶ First, a half-degree mesh is overlaid in a GIS environment and the road intersection or other easily distinguished feature (e.g. a sharp turn) that is closest to each centroid is identified. This ensures a semi-random selection procedure. Second, tie points are identified between the candidate layer and the reference layer and the distance is measured in meters. Finally, the root mean square error (RMSE) is calculated from the distance between the candidate and reference layer for each tie point.

In terms of completeness, an algorithm developed by Uwe Deichmann at the World Bank will be used to assess the length of the road network of each candidate data set against a baseline such as VMAP0. Additional approaches will be evaluated, such as those implemented by UN-JLC for roads data in Haiti.

In addition, visual inspection will be conducted to identify missing roads, and to get a better feel for the overall spatial accuracy and completeness. For each country, the most complete and spatially accurate roads layer will be identified, and permissions will be sought from the data set originators (see Permissions section below).

Lead: WG co-chair A. de Sherbinin

4. Formalizing the data commons language under which to disseminate gROADS

An important step, prior to seeking permissions from data providers, will be the formalization of the licensing arrangement under which gROADS will be distributed. Two candidates being considered are the Open Database License and the Public Domain Dedication and License (PDDL). However, it is equally possible that the working group will opt not to apply any kind of license at all, but rather employ some kind of quality assurance trademark along with language that makes it clear that the data are dedicated to the public commons but that the originators seek attribution for any published works that result from its use. Candidate language will be circulated to the WG for consideration by November 2008, and a decision will be made by December 2008.

Lead: WG member H. Onsrud

5. Seeking permissions on existing data (January 2009-ongoing)

Where national or regional data are not already in the public domain (e.g. US Tiger Line files), reasonable efforts will be made to seek permissions from the data set originators. An email and/or letter requesting permissions will be mailed to the agency or individual who, as best as can be determined, produced the original data set. If permissions are denied, the data set will remain in the catalog and may be available for use solely for

¹⁶ Tracks4Africa has agreed to provide GPS tracks for a set of nine East African countries for validation purposes.

humanitarian crises or research purposes. Emphasis will be placed on the need for open access data, and every effort will be made to convince data holders of the benefits of providing their data free through gROADS (such as the potential for their data to be kept up to date through the contributions of a wider community).

Lead: WG co-chair A. de Sherbinin

6. Compile the candidate layers into a composite data set (January 2009-ongoing)

Those layers for which permissions can be obtained will be integrated into gROADS using a consistent data model (the UNSDI-T), and roads will be connected at the border with roads from adjacent countries.

Lead: WG co-chair A. de Sherbinin

7. Release of version 1 of gROADS (December 2009-June 2010)

Version 1 of gROADS will be released in the first half of 2010. Subsequent updates will be released on a 2-3 year rolling basis. The data will be released free-of-charge via the SEDAC web site, and may be mirrored by other hosts.

Lead: WG co-chair A. de Sherbinin

II. Developing the wiki mapping environment

It is possible that CODATA-Roads will utilize the system developed by ITHACA, which is Web Feature Service-Transactional (WFS-T) compatible and which relies on a thick client (ESRI's ArcGIS) for the editing environment, though viewing is through any web browser.¹⁷ Within the UN system, UNHCR and WHO/INTERSOS are also experimenting with WFS-T functionality, and these will be evaluated, along with COTS technology such as ERDAS's Wikimapia. Ideally, the working group would like to develop or adapt a fully OGC-compatible environment using open source software. However, to do so will require external funding.

1. Develop specifications for the mapping environment (January 2009-June 2009)

The basic specifications for the mapping environment will conform to the OGC's WFS-T specification. However, additional functionality may be required in order to customize this specification for use by the CODATA-Roads initiative. This work will be dependent

¹⁷ This platform is accessible directly through ArcSDE/Oracle and PostGreSQL.

on further funding, and may be carried out by the OpenGEO of the Open Planning Project.¹⁸

Lead: Coordination by Greg Yetman (CIESIN)

2. Programmers to develop the software (June 2009-March 2010)

This will depend fundamentally on the developments mentioned above. Options include one dedicated development team (e.g. OpenGEO programmers) vs. coordinating a consortium of developers. The tool will be UNSDI-T compliant.

Lead: Coordination by Greg Yetman (CIESIN)

3. Building a Community of users and contributors (January 2010 – ongoing)

The CODATA WG will be pro-active in generating and maintaining a community of contributors to ensure the continuous up-dating of gROADS. The approach will be based on the OSM experience but will target mainly agencies and professionals with a direct involvement and stake in the availability of quality roads data.

This effort will require 2 to 3 regional awareness and familiarization workshops per year as well as a communications campaign to reach out to potential partners within the humanitarian, development and conservation communities. This too is contingent on additional funding.

Lead: Nicolas Chavent (SDI Consultant)

¹⁸ See <http://topp.openplans.org/our-work/opengeo/>.