PMGSY Rural Connectivity Datasets (PRCD22)

Context, Current Use and Future

Harsh Nisar
Lead Data Scientist, Data & Insights Unit
Ministry of Rural Development

harsh.nisar@nic.in
nisar.harsh@gmail.com
PMGSY Rural Connectivity Datasets

1. Roads
2. Habitations
3. Rural Infrastructure
4. PMGSY-III Proposals
5. Block Boundary
6. Location Master Data

https://geosadak-pmgsy.nic.in/opendata/
Roads

- Collected as part of District Rural Road Plans as early as 2001
- Earlier excel/tabular entry; GIS digitization started in 2015
- Many roads were not found, new roads were added, lengths differed, duplicates, multipart etc. Some of these problems still persist.
- 25,00,000+ km being released
- Live database – we wish to continuously work on improving the data and updating the release from time to time
- Digitized by each state separately either using in-house teams, local remote sensing centers or private GIS firms
- Standardization in attributes and topological checks conducted by C-DAC Pune
- Coverage and quality varies to an extent across each state.
## Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Remarks</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER_ID</td>
<td>Unique identifier for a Road</td>
<td>1670176</td>
</tr>
<tr>
<td>STATE_ID</td>
<td>Linked to Location Master</td>
<td>4</td>
</tr>
<tr>
<td>DISTRICT_ID</td>
<td>Linked to Location Master</td>
<td>385</td>
</tr>
<tr>
<td>BLOCK_ID</td>
<td>Linked Location Master</td>
<td>1038</td>
</tr>
<tr>
<td>DRRP_ROAD_</td>
<td>Road Number in our database</td>
<td>RR(VR)L185</td>
</tr>
<tr>
<td>RoadCategory</td>
<td>Village Road, MDR, SH, NH etc</td>
<td>RR(VR)</td>
</tr>
<tr>
<td>RoadName</td>
<td>Generally, __ to __ via ___</td>
<td>Lakhimpur to Kaithpara</td>
</tr>
<tr>
<td>RoadOwner</td>
<td>Department that manages the road.</td>
<td>PWD</td>
</tr>
</tbody>
</table>
What's a habitation?

- Village/ Habitation

The unit for this Programme is a *habitation* and not a revenue village or Panchayat. A habitation is a cluster of dwellings, in an area, the location of which does not change over time. Desam, Dhanis, Tolas, Majras, Hamlets etc. are commonly used terminology to describe the habitations. The population of all habitations within a radius of 500m (1.5 km of path distance in case of hills) may be clubbed together for the purpose of determining the population size. This **cluster approach** would enable provision of connectivity to a large number of habitations, particularly in the hill/ mountainous areas.

Source: [Page 3 PMGSY Operations Manual](http://example.com)
Habitations

• Collected as part of District Rural Road Plans as early as 2001
• Earlier excel/tabular entry; GIS digitization started in 2015
• 10,00,000+ GIS points
• Live database – we wish to continuously work on improving the data and updating the release from time to time
• Digitized by each state separately either using in-house teams, local remote sensing centers or private GIS firms
• Standardization in attributes and topological checks conducted by C-DAC Pune
• Coverage and quality varies to an extent across each state.
## Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Remarks</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATE_ID</td>
<td>Linked to Location Master</td>
<td>4</td>
</tr>
<tr>
<td>DISTRICT_ID</td>
<td>Linked to Location Master</td>
<td>385</td>
</tr>
<tr>
<td>BLOCK_ID</td>
<td>Linked Location Master</td>
<td>1038</td>
</tr>
<tr>
<td>HAB_ID</td>
<td>Linked to Habitation Dataset</td>
<td>356099</td>
</tr>
<tr>
<td>HAB_NAME</td>
<td>Name of the Habitation</td>
<td>Barninari Pub</td>
</tr>
<tr>
<td>TOT_POPULA</td>
<td>Population in 2011 as estimated by the field engineer</td>
<td>850</td>
</tr>
</tbody>
</table>
Rural Infrastructure

• Collected starting 2019 for the purpose of PMGSY-III
• Using in-house GEOPMGSY mobile app
• Clicking geo-tagged pictures of Educational, Health, Agro facilities as mentioned in PMGSY-III Guidelines
• 7,70,000+ facilities collected
• Field engineers, contracted staff etc were used for the data collection.
• Fixed period exercise: Inventory was frozen before proceeding with selection of roads. This data will largely not update for any state.
• Interpretations of facility definitions etc were left to the states to decide
• Generally only government infrastructure was taken but some states included private infrastructure too.

Earlier FAQs: http://omms.nic.in/Home/PMGSYRuralDataset/
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Remarks</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATE_ID</td>
<td>Linked to Location Master</td>
<td>4</td>
</tr>
<tr>
<td>DISTRICT_ID</td>
<td>Linked to Location Master</td>
<td>385</td>
</tr>
<tr>
<td>BLOCK_ID</td>
<td>Linked Location Master</td>
<td>1038</td>
</tr>
<tr>
<td>HAB_ID</td>
<td>Linked to Habitation Dataset</td>
<td>356099</td>
</tr>
<tr>
<td>FACILITY_I</td>
<td>Unique ID for each facility</td>
<td>899053</td>
</tr>
<tr>
<td>FAC_DESC</td>
<td>Name of the Facility as entered by the surveyor</td>
<td>Ghoruabaha Pathar Bus Stand</td>
</tr>
<tr>
<td>FAC_CATEGO</td>
<td>Larger category of the facility being released</td>
<td>Transport/Admin</td>
</tr>
<tr>
<td>FAC_SUBCAT</td>
<td>Facility sub-category (Eg. Bus stand).</td>
<td>Bus Stand</td>
</tr>
</tbody>
</table>

Will be added shortly.
PMGSY-III Proposals
“Task of connecting all eligible habitations with an all-weather road has been substantially completed, with the target date brought forward to March, 2019 from March 2022. It is now time to strengthen and widen its ambit further to include major link routes which connect habitations to agricultural and rural markets (GrAMs), high secondary schools and hospitals. Prime Minister Gram Sadak Yojana Phase III will include such linkages”

- Hon Union Finance Minister, Budget Speech, 2018

Key document is PMGSY-III Guidelines: https://pmgsy.nic.in/sites/default/files/PMGSY_III_guidelines.pdf
PMGSY-III Proposals

• 1,25,000 km is the overall target split across states
• 70,000+ km sanctioned (OMMAS)
• Significant portion of proposed alignments are digitized on GeoSADAK
• GIS Alignments of Proposals sanctioned before launch of GeoSADAK is in progress
• Proposals are drawn by Field Engineers or GIS officers in SRRDA on GeoSADAK portal [click and trace]
Pradhan Mantri Gram Sadak Yojana (PMGSY) was launched by the Government of India to provide connectivity to unconnected habitations in order to provide continuing access to economic and social services thus increasing the agricultural incomes, employment and other benefits. PMGSY-III envisages consolidation of the existing Rural Road Network launched with the aim of promoting economic development and social welfare in rural India.
### Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Remarks</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATE_ID</td>
<td>Linked to Location Master</td>
<td>4</td>
</tr>
<tr>
<td>DISTRICT_I</td>
<td>Linked to Location Master</td>
<td>385</td>
</tr>
<tr>
<td>BLOCK_ID</td>
<td>Linked Location Master</td>
<td>1038</td>
</tr>
<tr>
<td>HAB_ID</td>
<td>Linked to Habitation Dataset</td>
<td>356099</td>
</tr>
<tr>
<td>MRL_ID</td>
<td>Unique ID to link with OMMAS information</td>
<td>244300</td>
</tr>
<tr>
<td>CN_CODE</td>
<td>Unique ID to link with OMMAS information</td>
<td>657081</td>
</tr>
<tr>
<td>PROPOSED_L</td>
<td>Length of the Proposed Road</td>
<td>3.00</td>
</tr>
<tr>
<td>WORK_NAME</td>
<td>Name of the Proposed Work</td>
<td>MRL10-Jaysagar to Bartola No.1</td>
</tr>
<tr>
<td>IMS_YEAR</td>
<td>FY in which this work is sanctioned</td>
<td>2020</td>
</tr>
<tr>
<td>IMS_BATCH</td>
<td>FY against which this work is sanctioned</td>
<td>1</td>
</tr>
</tbody>
</table>

Additional information is also available on public reports in OMMAS for joining but we'll add important information here too based on feedback.

These are only sanctioned proposals, doesn't mean construction would have started (tendering etc).
Location Master Data

- Master data for State_ID, District_ID and Block_ID in the shapefiles
- Not to be considered as India’s latest or official administrative division dataset
- This dataset lags the actual administrative splits on the ground because of the effort involved in moving existing processes on the IT system to new geographies
- This dataset provides location information regarding the remaining shapefiles

<table>
<thead>
<tr>
<th>STATE_ID</th>
<th>STATE_NAME</th>
<th>DISTRICT_ID</th>
<th>DISTRICT_NAME</th>
<th>BLOCK_ID</th>
<th>BLOCK_NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Andhra Pradesh</td>
<td>528</td>
<td>Srikakulam</td>
<td>2225</td>
<td>Hiramandalam</td>
</tr>
</tbody>
</table>
Current Uses of Data
**Trace Maps**

Multi-objective transport planning algorithm

**Basic Principle**
Each habitation should have access to nearest school, hospital, market and administrative buildings

- Identify nearest facilities of every kind and draw the route from the habitation to each of the facilities
  
  Eg. Routes from Habitation A to nearest high school, hospital, agricultural market etc

- 22 such routes for every habitation – termed as its routes of fundamental access

- Repeat it for 1 million habitations ie 22 million fundamental routes

- Identify the roads which are part of these 22 million routes and calculate importance based on the destination facilities and population of habitation using the route

[Paper] Algorithmic Rural Road Planning in India: Constrained Capacities and Choices in Public Sector

**Inputs:**
Geo-tagged facilities
Habitation & rural road shapefiles

**Outputs**
Trace Map PDF
Trace Rank CSV
Python Console

Use face to access QGIS API interface or Type help(face) for more info.
Open Call

• Researchers, Civil Society, Students, Startups, GIS Industry
• Explore different innovative uses of these datasets
• Establish feedback loop so that we are able to improve the datasets we’ve released
• Feedback on attributes, data quality, missing data etc welcome
• While PMGSY engineers will update the data, how do we incorporate civil feedback on data back into our database with necessary checks and balances?
Geo-tagging Galore

- Many scheme are requesting geo-tagged photographs to be collected at the point of delivery
- SBM toilets, PMAY-G houses, Ujjwala, NREGA assets etc.
- If we combine PMGSY habitations with NREGA assets, you can get both micro and macro sense of how welfare is being distributed
• Spatial Queries (buffers etc)
• Joining things based on lat-long versus MIS
• Service Area Analysis (how far are services to each habitation)
• QGIS, Google Earth Engine, GeoPandas
• https://www.qgistutorials.com/en/
• https://spatialthoughts.com/
Caution

- In many states, villages have been marked as habitations
  - You’ll get only one point per village
- Population entered can be wrong or double counted
- Many habitations may now be in urban
- Facilities aren’t exhaustive and fixed in time (wouldn’t rely on it to make service delivery claims)
- If you are researching a scheme, worth writing to them or visiting their office once. Almost all the data has heavy context and assumptions behind them and can’t be taken as-is.
Gratitude

• Field Engineers on the ground who’ve travelled 100s of kilometers to geotag various assets for PMGSY-III
• GIS professionals across states who’ve painstakingly cleaned and created habitation and road datasets
• State ITNOs who led the digitization efforts in the state and embraced GIS even though it’s a different field altogether.
• C-DAC, Pune for standardizing, checking quality and curating the datasets nationally
Getting Started

1. Download QGIS
   (https://www.qgis.org/en/site/forusers/download.html)

2. Download PMGSY Datasets (https://geosadak-pmgsy.nic.in/opendata/)

3. 🌟
Thank You