

Government departments have begun opening up their datasets but we still lack a data regulator

My first column this year was a data wish list for 2022. The first wish was that we get to see survey data that have been delayed or suppressed. The second was a new law backing an empowered statistical regulator. The third was the opening up of administrative datasets collected by government departments as part of routine work.

Administrative datasets are generated using public funds but are typically withheld from the public. So I am glad to report that things appear to be changing. In an unprecedented step, the Union ministry of rural development has released data on key facilities (roads, bus stands, schools, hospitals, panchayat offices, agri-markets, etc) across 1 million rural habitations of the country. This dataset is a byproduct of India's flagship rural roads scheme, the Pradhan Mantri Gram Sadak Yojana (PMGSY).

A key goal of the PMGSY is to provide all-weather roads in the hinterlands to connect rural habitations (clusters of dwellings or village sub-units) to important sites such as schools or bus stands. The ministry used a weighting formula to prioritize roads that would link a habitation to a secondary school, hospital or a mandi (agri-market). To collect data, field engineers fanned out across India over the past few years to record the geographic coordinates of these facilities on an application developed by the Pune-based Centre for Development of Advanced Computing (C-DAC). The data on these facilities have now been released as part of the rural connectivity dataset (<https://geosadak-pmgsy.nic.in/OpenData>). It is perhaps one of the most granular geo-tagged datasets available in the public domain today. Given the paucity of rural data, this database could help researchers and private firms understand and serve rural India better. The dataset has been released under an open data licence, which means that it can be used freely by both public and private organizations.

Like any other administrative dataset, this one too poses several statistical challenges. Coverage and definitions vary across states because state-level officials were given the discretion to tailor the scheme according to the needs of each region. There could be errors in some location coordinates as well. So the data cannot be naively merged with other databases without accounting for these definitional, coverage, and quality issues.

Yet, this data release is highly promising on three counts. First, the data release has been done in an open and accessible format, which makes it easy for developers to build other applications or conduct research. The open data licence will also enable officials in other government departments to mine the data intensively without having to go through a Kafkaesque maze of approvals. The biggest beneficiary of open government data of this kind is the government itself. Despite limitations, rural connectivity data can be of immense value in framing rural policies.

Second, the ministry's data team is open about both the strengths and weaknesses of the dataset, and is keen to improve data quality. The data team is engaging with data users to make them aware of the potential uses of the dataset, context under which it was collected, and also to collect feedback, said Harsh Nisar, the lead data scientist at the ministry's data insights unit. The ministry is trying to work out a governance mechanism to incorporate public responses on deficiencies in the dataset, such as missing habitations or roads, he added.

Third, the ministry has tied up with what is perhaps India's largest open data community,

DataMeet. Started by Bengaluru-based techies S. Anand and Thejesh G.N. on 26 January 2011, DataMeet has grown into a country-wide community of data nerds today, with its membership running into the thousands. Like many other journalists, I have benefited from its high-quality discussions and pool of resources. The ministry, too, is likely to gain much from its engagement with DataMeet.

DataMeet acts as a channel of communication among data users through its mailing list, which is also used to update and upgrade its repository of open data and maps. In its early years, the group would petition ministries and departments to open up their datasets. With ministry officials now reaching out to them, life seems to have come full circle for the community. Community partners such as DataMeet can help import the geo-tagged facilities into an open map framework such as OpenStreetMap (an open-source alternative to Google Maps) for wider use, said Nisar.

The rural development ministry's example could inspire other ministries to start opening up their datasets. Involvement of the open data community in these initiatives can help improve data accessibility and quality. If all open datasets are connected via common geographic identifiers, then they could generate rich insights for both the government and private sector.

This process can become smoother over time if the government standardizes data formats and definitions across states, departments and ministries. Lack of such standardization means that a data user has to use a fair number of assumptions and adjustments to be able to use the available public datasets. This adds to the cost of doing business or research in the country, and slows down innovation. This is where an empowered data regulator such as a statutory National Statistical Commission could play a vital role by harmonizing data standards and pulling up data laggards within the government.

If only the second wish in my wish list were to come true now.

Pramit Bhattacharya is a Chennai-based journalist. His twitter handle is pramit_b

[Download](#) the App to get 14 days of unlimited access to Mint Premium absolutely free!

Log in to our website to save your bookmarks. It'll just take a moment.

Oops! Looks like you have exceeded the limit to bookmark the image. Remove some to bookmark this image.

Your session has expired, please login again.

You are now subscribed to our newsletters. In case you can't find any email from our side, please check the spam folder.

This is a subscriber only feature [Subscribe Now](#) to get daily updates on WhatsApp

END

Downloaded from **crackIAS.com**

© **Zuccess App** by crackIAS.com