

Government of India  
Ministry of Finance  
Department of Economic Affairs  
New Delhi-110001

Dear Honorable Finance Minister,

## RE: Pre-Budget Consultation (2018-19) with Ministry of Finance - IT Group

Thank you for inviting DataMeet to the Pre Budget Consultation with the IT Group. [DataMeet](#) is a community data enthusiasts from all over the India who come together to discuss data issues, share solutions, learn new skills, and encourage open data. We believe that a strong dedication to open data can make a huge contribution to the economy in India. The suggestions put forth here represent the collective opinions of many of us working in the private sector as well as researchers. We ask that for the 2018-2019 budget you take the following into consideration.

Data is the new oil. The new economy is based on the flow of real-time accurate data that can fuel startups, business strategy, create new markets, and higher quality services. Data can also be an incredibly powerful tool for implementing development schemes and monitoring local and nationwide progress. Data is just as valuable as any natural resource India has, and it has to be viewed as an asset, in terms of collection, management, and usability. When the [US government created their open data policy](#) they opened with this value:

“Data is a valuable national resource and a strategic asset to the U.S. Government, its partners, and the public. Managing this data as an asset and making it available, discoverable, and usable – [in a word, open](#) – not only strengthens our democracy and promotes efficiency and effectiveness in government, but also has the potential to create economic opportunity and improve citizens’ quality of life.

[Assessments](#) were made regarding the potential economic impact of the opening that data has been.:

Open weather data supports an estimated [\\$1.5 billion](#) in applications in the secondary insurance market – but much greater value comes from accurate weather predictions, which save the U.S. [more than \\$30 billion annually](#). The value of GPS data has been [estimated at \\$90 billion](#), but that number comes from an industry study and may be too high.

Mckinsey reports open data can help create [\\$3 trillion](#) a year in global economy. European Union’s study on “[Creating Value through Open Data](#)” describes EU’s intention of creating 200,000 jobs in the open data sector by 2020.

These values helped fuel the move for data to be more consumable in order to reach the most economic actors. For India’s data to have this type of value a hard look at what data exists, what data is useful, what improvements to collection and management have to be done, and ultimately the release of this



data in open and machine readable formats must be a priority. For a vibrant IT market to function multiple datasets must be available to pull from, currently private companies are functioning by collecting this data themselves, however, if the government made them available the need for silos of data collection would decrease and a data commons would be created for everyone to pull from and create more options. The availability of open data has helped fuel innovation in various parts of the world in addition there would be no Google Maps without NASA opening LANDSAT satellite imagery in public domain.

## Policies preventing the full economic impact of Government of India Data

India has been moving forward in achieving the aspirations of Digital India and development goals set forth. We are asking that the 2018/2019 budget commit to building a sustainable data ecosystem to create a strong and sustainable data economy based on open principles.

The first policy to support is the National Data Sharing and Accessibility Policy (NDSAP) that the government approved in 2012. This promotes the use of open data as [defined](#) as data that is free to be “used, re-used, and redistributed to anyone.” This initiative allowed for the 6th Pillar of Digital India - Data.Gov.in -- to be created to jump start data based start ups and transparency initiatives across the country. **We agree with the last Economic Survey that support for state and city initiatives will encourage economic growth. With the increase in Smart Cities more useful and dynamic datasets will be created in the coming years that can create whole new industries for people at every level, this policy and its implementation is the foundation for that. The Ministry of Finance can create a fund for Ministries and states who are actively fueling the data economy with open data initiatives, like Sikkim, Telegana and Surat, Gujarat**

This is also essential to curb corruption and inefficiency. For example the Ministry of Finance has [requested access](#) to PNR data to avoid fraud by Government Officials, but if this is done proactively with workflows already built into the Ministry of Railways then the Government of India doesn't have to request access and wait for the data to be released. This allows for dynamic real time turnaround of data so that implementation of schemes and policies can be monitored and evaluated in real time.

Initiatives of raw and real time data sharing have already started and have seen success at supporting the economy. The [Customs Department](#) publishes raw Export Import Trade data under Notification No. 18/2012-Customs (N.T) dated: 5th Mar, 2012. This portal has been used and when it was recently down due to maintenance it affected many businesses. Similarly the Ministry of Railways, while the data is not open (as defined above) but available for people to use in a limited fashion, it has created an ecosystem of private players using it to create jobs and services based economy. Players like Ixigo, Travelyaari, Travel Khaana are improving services to citizens apart from railways. We are concerned that the intention to monetize this data assets in the upcoming financial year will have consequences to these startups which can't afford to pay will. **We request the Finance Ministry to offer other Ministries support to make their data**



available in real time and open in order to promote job and economic growth under Digital India and make doing business in India easier.

When introducing new legislation, relevant data should be made public as part of the statutory process. The purpose is to ensure that there are ways to identify the efficacy of the legislation, and no separate system is required. This had been proposed as part of the Indian Financial Code (Chap 14, 78). Similar disclosures were also required under the Public Procurement Bill, 2012. **Public infrastructure related expenditure by central government should have measurable performance standards prescribed and placed on the authority website. At the site of such infrastructure activity, a board providing the link to the works should be listed. (Like for MNREGA, details have to be painted on the wall, and lists made available publicly). As well as all RTI responses should be placed in the public domain. This will also allow fast turnaround of data needed to monitor scheme implementation to course correct quickly.**

In the case data has to be monetized and cannot be made open. **We suggest a single point of sale for data through Data.Gov.in instead of each department bringing up separate portals. This is to ensure data continuity and reliability by the departments.**

**A comprehensive policy on privacy of citizen with regards to open data is very important. Such a policy can help officials open up the data without exposing private individuals information. This will make any services created with these datasets more secure.**

**We are asking that these policies be implemented according to the guidelines specified in NDSAP and by the Data.Gov.in implementation standards.** This will allow for more complete data to be available more efficiently and accessible to a wide variety of actors to use to improve governance and economic opportunities.

The Mapping Policy of India is preventing mapping initiatives taken up by the Survey of India to meet the growing need for georeferenced and geographic data. New IT companies like Zomato, OLA, Flipkart and hundreds of others are mapping routes and infrastructure in the absence of detailed Government mapping data. These companies are fueling economic growth however the lack of support from a policy puts these types of services and companies at risk. **Updating the mapping policy to allow for input from crowdsourcing or company mapping data, that is required to be reported to the Survey of India and then made available to the public in electronic formats would make the Government of India the public common for all the mapping information and allow for faster, cheaper more efficient updates to geospatial information for all actors including the government.**

**Incentives for businesses through tax breaks or the CSR protocol, who collect public good datasets, to make it open and accessible through the Data.Gov.In platform.** For instance, Infosys has a lot of water use and weather data that could be very helpful to local bodies as well as can contribute to the Central Ground Water Board and India Meteorological Department who



are in charge of monitoring and evaluation. These types of datasets can supplement government data and create better monitoring and evaluation of scheme implementation.

**Copyright for data has to be open data license by default when sharing to allow for use by commercial projects. Copyright Policy must be clarified in order to make sure data is being used rightfully and legally by parties.**

**Vendors and 3rd party data collector agreements must specify that the data collected on behalf of the public or with public funds must be owned by the Government of India and opened (taking into account privacy concerns) on the Data.Gov.in platform.**

Sustainable Development Goals calls for “for increased support for strengthening data collection and capacity building in Member States, to develop national and global baselines where they do not yet exist. We commit to addressing this gap in data collection so as to better inform the measurement of progress, in particular for those targets below which do not have clear numerical targets.” in their declaration. **It is important for India to create more data around their development goals and schemes and by making this data open, corruption and inefficiencies can be kept low for instance the Census highlighting the issues around Swachh Bharat Mission numbers in 2011.**

## Where opening data has worked

Since 2012 the NIC has been publishing data through the Data.Gov.in platform, also several Ministries have been using data and presenting it to the public for use for planning and implementation.

One of the important and not recognized areas where open data has worked is in trade, both exports and imports trade data being open has created several small scale data businesses which are providing insights to export traders to improve business. Some of these firms include [InfoDriveIndia](#), [Trade Intelligence](#), [Cybex](#), [ExIm Pulse](#).

Swachh Bharat Mission has been putting live data up for years and people have used the data to keep track of spending and toilet building, when the Census came out it was able to compare numbers and assess an issue with implementation, so that the scheme can be re-examined.

Mandi Prices was made available on Data.Gov.in and spurred a successful app that people use to buy produce. This data is being used by a larger population than the intended and is reaching more people who are interested in agricultural work.

A few Census datasets are made available (though the full Census isn't available publicly) this has allowed for many startups to take the data and make it usable and available. (I.e. [HowIndiaLives](#)) This is a prime example of building markets on Government datasets.



BHUVAN is the premier geospatial data run by ISRO. There are a dozen portals with heaps of government data to be used for natural disaster, resource management, and local planning. However because of the Mapping Policy, high resolution and street level resolution is not available which hinders use.

Sikkim, Telangana, Tamil Nadu, Orissa and Madhya Pradesh have adopted open data policies. Sikkim and Telangana have implemented it through their portal [Sikkim.Data.Gov.in](http://Sikkim.Data.Gov.in) and [Data.Telegana.Gov.In](http://Data.Telegana.Gov.In) as a way to improve access to their own data, not just for economic purposes but more specifically to improve the access to state and local level information by government officials.

[Surat, Gujarat](http://Surat.Gujarat) and [Pune, Maharashtra](http://Pune.Maharashtra) have launched local open data portals for their cities and Smart City initiatives. Local open data that is real-time and complete has the most potential for transformation, these sites can be the foundation for apps, service providers, and startups to get the data they need to launch their products instead of spending resources and money on data collection which is time consuming, expensive, and repetitive to what the government already has.

## Priority Datasets

The next big wave of innovation in India will be Artificial Intelligence (AI), once data is open and available with updated copyright law, AI can be encouraged and become a thriving Industry. Ministry of Commerce has constituted a Task Force on Artificial Intelligence for Economic Transformation this year. We would like to bring attention to the need for various open data sets in different sectors for innovation in the space of Artificial Intelligence. Given our experience the following data sets are the most sought after for startups and businesses.

1. Geographic data that is granular and always up to date - districts, villages, towns, city wards, locations of assets (schools, hospitals,), infrastructure (roads, bridges, settlements)
2. All census data from 2001,1991 in the same open formats that the 2011 data that has been published.
3. Raw Export Import trade data must be made and kept public to support the businesses who rely on it.
4. Pincodes - areas served by each pincode has to be mapped and made official. Currently Pincodes are being used for service delivery and the lack of clarity is causing multiple versions of Pincode maps to exist without really knowing what the area looks like.
5. Transport - real time, complete and accurate data across India that can communicate with each other to create multimodal solutions.
6. Health - anonymized health data (to protect privacy) - assets, personnel, drugs prescribed, diagnosis, outbreak reports, deaths with due causes.
7. Water and sanitation - what is the status water at a local village level, status of drinking water schemes, quality information, groundwater and surface water, sanitation status of villages and urban localities.



8. Agriculture related - live status of schemes for irrigation, seeds, fertilizers, pesticides, cattle
9. Rural livelihoods - live status of schemes e.g. MGNREGA
10. Resources - natural resources and the status of the resource and its contribution to the economy
11. Weather - real-time meteorological readings from stations from all over the country at a local level.
12. All Bureau of Indian Statistics standards
13. All budgetary accounts in a machine readable format
14. Real-time expenditure data. Geo-locating expenditures wherever possible. Assigning census village and town codes (or district / sub-district / state codes) to expenditures/disbursals.
15. State and local government bodies, all should have all their decisions, documents, resolutions, and minutes should be available automatically in searchable databases.
16. Doing Business Ranking related data in order to promote those principles at the local and state level.
17. The Statistical data, specifically NSSO data, this is valuable data for people.

## **In order to maximize the economic impact of this data:**

Data must be standardized, names, assets, and locations have to be given unique IDs that can be mapped from one dataset to the other.

Machine readable and open formats must be the default for the data to disseminated. Demystified data collected for administrative purposes by the Government must be explained and given proper metadata.

Bulk download of data must be available, through open APIs, past and current data is necessary for analysis and building tools of that work.

Support for the IT department, better funding for data storage, also a requirement that all data generated with public funds be owned by the Government of India.

Not only will these measures bring a huge economic impact, they will also be applauded as a way to bring more accountability and transparency to the system, as well as, create innovative ways to include citizens in the workings of Government.

Thank you again for your time and attention to these recommendations. We hope to see the next budget as a first step to really exploring the huge potential Government of India has for a wider audience.

Sincerely,  
DataMeet

Represented by Thejesh GN

