data.gov in

Open Government Data (OGD) Platform India

http://data.gov.in

DOs

- Identify and prioritize the release of datasets; categorize the type of access granted for them and publish as many high value datasets as possible.

- Contribute datasets which are in the Open List and do not fall under the Negative List.

- Ensure that the quality standards are met i.e. accuracy, free from any sort of legal issues, privacy of an individual is maintained and does not compromise with the National security.

- Ensure that the datasets being published through a workflow process are in compliance with NDSAP. Details on original source of the dataset and methodology of the data collection should be provided in metadata.

- Prepare and contribute the metadata in predefined format for the Catalogs and Resources (Datasets/Apps). The key metadata elements are Title, Description, Category, Sector/Sub-Sector, Dataset Jurisdiction, Keywords, Access Method, Reference URLs, Data Group Name, Frequency, Granularity of Data and Policy Compliance. All the metadata elements must be filled with utmost quality and ease of use.

- Pricing of data, if any, would be decided by the data owners as per the government policies.

- Ensure that data being contributed to the OGD Platform are in machine readable or in specified open data format only. The advisable formats are:

    - CSV (Comma separated Values)
    - XLS (spread sheet- Excel)
    - ODS (Open Document Formats for Spreadsheets)
    - XML (Extensive Markup Language)
    - RDF (Resources Description Framework)
    - KML (Keyhole Markup Language used for Maps)
    - GML (Geography Markup Language)
    - RSS/ATOM (Fast changing data e.g. hourly/daily)

    (Tim Berners-Lee had classified data into single star to five star categories based on formats for the data. Please refer to Annexure II in the Implementation Guidelines)

- Ensure that the data being uploaded on the OGD Platform is as complete as possible, reflecting the entirety of what is recorded about a particular subject and is de-normalized. The datasets also should be optimized by adding redundant data or by grouping data before uploading.

- Priority should be given to data whose utility is time sensitive. Real time information updates would maximize the utility the public can obtain from this information.

- Replace any Not Available, Not Reported or missing values in the data with 'NA'.

- Metadata that defines and explains the raw data should be included as well, along with formulas and explanations for how derived data was calculated.

- Keywords must be defined in data catalog to make it search friendly.

- Provide link to the reference documents (if any) or website for detailed information and explanation on the method of calculation or source of data.

- Read the process manual at https://data.gov.in/help for the step by step procedure for contribution.

- Prioritize the release of datasets and take relevant action on the basis of feedbacks and suggestions received on the OGD Platform from citizen's pertaining to the Ministry/Department.

- Ensure the correctness of login details on the OGD Platform by sending a mail to ndsap@gov.in, in case of any change.

data.gov in

Open Government Data (OGD) Platform India

http://data.gov.in

DON'Ts

- Don't contribute datasets which fall under the negative List e.g. the datasets which are confidential in nature and are in the interest of the country's security.

- Don't impose 'Terms of Service', attribution requirements, restrictions on dissemination and so on, which act as barriers to public use of data.

- Don't impose cost on the public for access of datasets, as imposing fees for access skews the pool of who is willing (or able) to access information.

- Don't publish hand written note, as it is very difficult for machines to process. Scanning text via Optical Character Recognition (OCR) results in many matching and formatting errors. Information shared in the widely used PDF format is very difficult for machines to parse. Hence, the data in these formats should be avoided.

- Data in non-Unicode formats should be avoided.

- Don't contribute datasets with any special characters (e.g. @, %, $, &, etc.) or missing values.

- Don't provide any explanation/note, including the method of calculation or source of data in data file to be attached in the web form.