Survey of Open Data Practices in India

0. Introduction

Welcome to the survey of open data practices in India.

This survey is being done as part of a research study on government policies and non-government initiatives around open (government) data in India.

The study is part of a project on 'Exploring the Emerging Impacts of Open Data in Developing Countries' (or, ODDC) managed by the World Wide Web Foundation and the International Development Research Centre, Canada. The gathered information will be used for the research study mentioned above, and extracts it will also be shared online for viewing, use, modification and augmentation by other interested researchers. Thank you very much for agreeing to contribute to this survey. Below are some details to clarify the kind of information we are gathering in this survey and how it will be shared.

Data: What is data, and hence what are **data practices**, is a complicated and contextual question. For the purpose of this survey, we consider any act of using data that involves analysis, aggregation, disaggregation, visualisation (converting the data into diagrams, as in charts and maps) or comparison with other data, as **data practices**. Hence, for this survey, please do **not** consider any direct and unmodified usage of available data (such as quoting the number of babies under 5 years suffering from malnutrition from a published report) as **data practice**. Further, we are interested only in **data practices** within the contexts of advocacy, outreach, criticism and other relevant processes of communication and action. Hence, for this survey, please do **not** consider any uses of data for internal reporting (such as reports sent to donor organisations) as **data practice** of the organisation.

Data Usage: Data usage may mean any communication of the data verbally (e.g., in a public meeting or a Jan Sunwai) or visually (e.g., by creating charts and maps, or by publishing it in a report or on a poster) meant for direct consumption, as opposed to sharing of data where the actual usage of it is expected to be done by somebody else. In most cases, data usage would mean communicating the data in an aggregated and/or transformed form and not in the same granular form of the original data.

Organisation: We invited your organisation to contribute to this survey because we identified it as an important actor in India in the domain of using government data, especially data obtained under RTI act or through <http://data.gov.in/> platform, for advocacy and various other public actions. Please see if we have missed out any important organisations. Conversely, please let us know if you think that your organisation has not substantially used government data in its various exercises.

Sections of the Survey: The survey has six sections:

1.1.   Organisation: questions about the agency concerned
1.2   Organisation: questions about the agency's involvement with open data
2.1.   Accessing Data #1: questions about accessing government data
2.2.   Accessing Data #2: question about accessing data from non-gov sources
3.1.   Using Data #1: questions about usage of data
3.2.   Using Data #2: questions about linkages with other agencies for data use
4.   Collecting Data: questions regarding data created by the agency
5.   Sharing Data: questions regarding sharing of data by the agency
6.   Data Ecology: questions regarding the agency's location in the data ecology

Use and Sharing of Contributed Information: We will share all the information contributed to this survey under Creative Commons Attribution-ShareAlike 4.0 International license <https://creativecommons.org/licenses/by-sa/4.0/>. That means, the information shared by you would be publicly visible, and can be referred to and reused by anybody (including you) for further survey on similar topics, or for other purpose. If you do not want any information shared by you in this survey to not be available publicly, please mention that in the text box you will see at the very end of the survey questionnaire.

1.1. Organisation #1

Note: Additional information for the questions in italics.

1.1.1. Name and address of the organisation

1.1.2. Established in

1.1.3. How many people work with the organisation?

1.1.4. What percentage of total workers is female?

1.1.5. Please mention the topics/areas in which the organisation works on.

1.1.6. Please briefly share a few important milestone of the organisation in last five years.

1.1.7. Please mention the major sources from which the organisation receives funding.
Under no circumstances we will share this information publicly.

1.2. Organisation #2

1.2.1. Is there an organisation level mandate for working with data/ information/open data?

1.2.2. Is there a specific group within the organisation that advances the mandate?

1.2.3. Is there any support/request from the funders' side to work with data?

1.2.4. When was the organisation first introduced to the idea of "open government data," and by whom?

1.2.5. Is there an open data movement/culture in india? Does the organisation consider itself as part of the open data movement/culture in india?

1.2.6. What encourages/discourages to be part of the open data movement in india?

1.2.7. Do RTI and open data movements in India have similar or different ideological positions and practical concerns?

1.2.8. Is the organisation (relatively) more closely associated with the open data movement as opposed to the RTI movement, or vice versa?

2.1. Accessing Data #1

2.1.1. Does the organisation access and/or use data created by Government of India? *This can include both data created by governmental processes, like financial records, or data collected by government agencies, like weather data collected by India Metereological Department.*

[Yes / No]

2.1.2. What is the typical procedure followed by the organisation to access data created by a government agency?

2.1.3. How many times has the organisation filed a RTI request in last one year? *We do not expect you to give an exact number. Please write an approximate number and **not** a range.*

2.1.4. How many datasets has the organisation downloaded from government websites (excluding <data.gov.in>) in last one year? *We do not expect you to give an exact number. Please write an approximate number and **not** a range.*

2.1.5. How many datasets has the organisation downloaded from <data.gov.in> in last one year? *We do not expect you to give an exact number. Please write an approximate number and **not** a range.*

2.1.6. The organisation has accessed government data using which of the following file formats?
*Please select as many as applicable, and if needed mention other formats in the **other** box.*

[ ] Database (MS Access, Oracle, SQL, MySQL etc.)
[ ] PDF (obtained from government website / Internet)
[ ] PDF (obtained over CD or other offline media)
[ ] Photocopy of government documents
[ ] Spatial data formats (Shapefiles, KML, etc.)
[ ] Spreadsheet (XLS, XLSX, ODS, CSV etc.)
[ ] Other:

2.2. Accessing Data #2

2.2.1. Does the organisation access and/or use data created by other non-governmental organisations?

[Yes / No]

2.2.2. If yes, please tell us names of some important NGO-s from which you have received data.
Please mention if we can share this information publicly.

2.2.3. Also, please tell us how you accessed data from these organisations, including using what file formats and under what restrictions.
Please mention if we can share this information publicly.

2.2.4. Have you ever bought data from government or non-government organisations?
*We are not considering expenses of photocopying, copying the data into CD-s etc. as costs of **buying** data.*
[ ] Bought from government agencies.
[ ] Bought from non-governmental organisations.
[ ] Bought from private data selling agencies (such as <indiastat.com>)
[ ] Never bought data/information

[ ] Other:

3.1. Using Data #1

3.1.1. Please share a few important examples of how government data was used by the organisation.
For example, to argue against the suitability of an action or policy of the government.

3.1.2. Please mention what important datasets were used by the organisation in last five years.
Please include the source of the datasets.

3.1.3. In what forms does the organisation present data for communication?
*Please select as many as applicable, and if needed mention other forms in the **other** box. **Print** and **Online** respectively indicate at how the presented data is shared, either in printed form or in online/web-based form.*

[ ] Create maps, or geographic representation (online)
[ ] Create maps, or geographic representation (print)
[ ] Infographic (online)
[ ] Infographic (print)
[ ] Publish tables of data (online)
[ ] Publish tables of data (print, including posters)
[ ] Refer to and highlight specific figures within a report, poster, brief, etc.
[ ] Other:

3.1.4. What tools do the organisation use to analyse data?
*Please select as many as applicable, and if needed mention other tools in the **other** box.*

[ ] Calculator
[ ] GIS - Commercial (ArcGIS)
[ ] GIS - Free (GRASS and Quantum GIS)
[ ] Programming Languages (Python etc.)
[ ] Spreadsheet – Commercial (MS Excel)

[ ] Spreadsheet – Free (LibreOffice Calc, Gnumeric etc.)
[ ] Statistical Tools – Commercial (SPSS, Stata, etc.)
[ ] Statistical Tools – Free (R)
[ ] Other:

3.1.5. What tools do the organisation use to visualise data?
Please select as many as applicable, and if needed mention other tools in the ***other*** *box.*

[ ] Design Softwares - Commercial (Adobe Illustrator)
[ ] Design Softwares - Free (Inkscape)
[ ] GIS - Commercial (ArcGIS)
[ ] GIS - Free (GRASS and Quantum GIS)
[ ] Programming Languages (JavaScript etc.)
[ ] Spreadsheet - Commercial (MS Excel)
[ ] Spreadsheet - Free (LibreOffice Calc, Gnumeric etc.)
[ ] Statistical Tools - Commercial (SPSS, Stata, etc.)
[ ] Statistical Tools - Free (R)
[ ] Other:

3.2. Using Data #2

3.2.1. Has the organisation collaborated with another organisation either to receive or provide support for data usage?
That is, has the organisation supported data usage (analysis, visualisation etc.) of another organisation, or asked another organisation to support its own data usage? Please check as many as applicable.

[ ] Has supported data usage of another organisation
[ ] Has not supported data usage of another organisation
[ ] Has received support from another organisation for data usage
[ ] Has not received support from another organisation for data usage
[ ] Neither received nor provided data usage support

3.2.2. Please briefly share important cases of the organisation providing data usage support to other organisation.

*Write **NA** if the organisation does not provide data usage support to other organisations.*

3.2.3. Please briefly share important cases of the organisation receiving data usage support from other organisations.
*Write **NA** if the organisation has not received data usage support from other organisations.*

3.2.4. Does the organisation feel that it can support other organisations with data usage expertise? If yes, please mention in what types of data practices the organisation has gained expertise?

3.2.5. Does the organisation feel that it would be useful to receive data usage support from another organisation? If yes, please mention in what types of data practices the organisation would like to receive support?

4. Collecting Data

4.1.   Does the organisation collect primary data for its own work, or to support work of others?
*The emphasis is on **primary data**, so please say **Yes** only if the   organisation undertakes direct surveys to collect data. For example, collecting data through a RTI request cannot be considered as collecting **primary data**.*

[Yes / No]

4.2.   If yes, the organisation collects primary data in which of the following areas? Please select as many as applicable, and if needed mention other areas in the **other** box.

[ ] Agriculture
[ ] Base line / household surveys
[ ] Budget / financial data
[ ] Children's issues
[ ] Civil liberties
[ ] Cultural artifacts / buildings

[ ] Education
[ ] Election / electoral reform
[ ] Employment and workplace
[ ] Energy
[ ] Environment / ecology
[ ] Land and landed resources
[ ] Performance of government agencies
[ ] Police / military reform
[ ] Public opinion and issues
[ ] Water and sanitation
[ ] Women's and gender issues
[ ] Other:

4.3. Also, please mention a few important datasets that the organisation has collected through direct surveys.

4.4. And what tools among the following are used by the organisation to collect primary data?
*Please select as many as applicable, and if needed mention other tools in the **other** box.*

[ ] Digital device-based data collection (data entered directly to a digital device by trained surveyors)
[ ] Mapping by informants, or participatory mapping
[ ] Mapping by trained surveyors
[ ] Mobile phone-based data collection (data entered by informants, sent back to the organisation as SMS)
[ ] Mobile phone-based data collection (data entered by trained surveyors)
[ ] Paper questionnaire (data entered directly by trained surveyors)
[ ] Paper questionnaire (shared with informant and data entered by informant)
[ ] Web-based data collection (informants submit data through a web-based questionnaire), or crowdsourcing
[ ] Other:

5. Sharing Data

5.1.   Does the organisation share data with any of the following agency types?
*By **share** we do not mean sharing of data derivatives (like reports, posters, infographics etc.) but sharing of **raw data**. Please select as many as applicable, and if needed mention other organisation types in the **other** box.*

   [ ] Academic institutions
   [ ] Governmental bodies
   [ ] Media agencies
   [ ] Non-governmental organisations (advocacy)
   [ ] Non-governmental organisations (research)
   [ ] Political organisations
   [ ] Residents' bodies (say Resident's Welfare Associations)
   [ ] Other:

5.2.   If the organisation shares **raw data** with any government bodies, please share names of a few important ones among them.
Please mention if we can share this information publicly.

5.3.   If the organisation shares *raw data* with any non-governmental (in a wide sense) organisations, please share names of a few important ones among them.
Please mention if we can share this information publicly.

5.4.   What difficulties do the organisation face in sharing data with other organisations?
*Please select as many as applicable, and if needed mention other difficulties in the **Other** box.*

   [ ] Ambiguity about the usefulness of the data to any other organisation/individual
   [ ] Data sharing is not an organisational priority, as other actions exhaust limited resources
   [ ] Insufficient resources to store and/or share data (say, lack of resources to maintain data archive and share it online)

[ ] Restrictive licensing / copyright associated with the data
[ ] Sensitive nature of the data
[ ] Unclear licensing / copyright associated with the data
[ ] Unknown / informal source of the data
[ ] Other:

5.5.   The organisation shares data using which of the following file formats?
*Please select as many as applicable, and if needed mention other formats in the **other** box.*

[ ] Database (MS Access, Oracle, SQL, MySQL etc.)
[ ] PDF (obtained from government website / Internet)
[ ] PDF (obtained over CD or other offline media)
[ ] Photocopy of government documents
[ ] Spatial data formats (Shapefiles, KML, etc.)
[ ] Spreadsheet (XLS, XLSX, ODS, CSV etc.)
[ ] Other:

6. Data Ecology

6.1.   How does the organisation see its role in relation to other organisations working with (either producing, sharing or using) data ?
For example, whether the organisation leads usage of data in its sector, or perhaps it focuses on gathering data to support works of other organisations. It can of course be a combination of multiple roles.

6.2.   Please mention a few other organisations **in your state** with which the organisation shares relationships related to data usage. Please also mention the nature of these relationships , such as "receiving data" and "supporting data analysis."
Please mention as many organisations as possible.

6.3.   Please mention a few other organisations **outside your state but within India** with which the organisation shares relationships related to data usage. Please also mention the nature of these relationships, such as "receiving data" and "supporting data analysis."

Please mention as many organisations as possible.

6.4.   Please mention a few other organisations **outside India** with which the
organisation shares relationships related to data usage. Please also mention
the nature of these relationships , such as "receiving data" and "supporting
data analysis."
Please mention as many organisations as possible.

7. Final Details

7.1.   Name of the person who has filled up the questionnaire.

7.2.   Designation of the person who has filled up the questionnaire.

7.3.   E-Mail of the person who has filled up the questionnaire.

7.4.   Would you want any information shared by you in this survey to not be
available for public viewing?
*Please mention the question concerned, and please mention the reason (for not
making the information public).*

I am grateful to you for taking out your valuable time to contribute to this survey.

Please feel free to write regarding any clarifications and suggestions.

Sincerely,

Sumandro Chattapadhyay
<http://ajantriks.net/>
<http://ajantriks.github.io/oddc/>