

# Disastrous Decade for Data

## The Way Forward

ARJUN KUMAR, P C MOHANAN, AMITABH KUNDU

In recent years, the Indian official data has been challenged for many reasons, most of them arising out of the perception that the government is reluctant to release unfavourable data. These doubts have been exacerbated by the controversy posed by the debate around Citizenship (Amendment) Act, National Population Register, and National Register of Citizens and the weakening of established institutions like the National Statistical Commission. The COVID-19 pandemic has further upset the data collection so that the first phase of the decennial population census and the proposed NPR exercise now stand postponed. Questions on credibility of official data and the pandemic-induced problems have come at a time when technology offers solutions to data collection, processing and dissemination. The initiatives launched by the government to improve the statistical system utilise these possibilities.

Arjun Kumar ([arjun@impriindia.org](mailto:arjun@impriindia.org)) is director, Impact and Policy Research Institute, New Delhi. P C Mohanan ([parachottimohanan@gmail.com](mailto:parachottimohanan@gmail.com)) is former acting chairperson of the National Statistical Commission and mentor, IMPRI. Amitabh Kundu ([akundu.jnu@gmail.com](mailto:akundu.jnu@gmail.com)) is distinguished fellow at the Research and Information System for Developing Countries, former member of NSC, former professor, JNU and mentor, IMPRI.

In light of the prevailing COVID-19 crisis, the Government of India has deferred the National Population Register (NPR) updating and the inception phase of Census 2021 that was to start on 1 April 2020. Given the unrelenting spread of the pandemic, it is impractical and the least advisable to send census enumerators on the field to collect and validate the information from the households. Generally, statistical exercises take a backseat during calamities. In fact, the conduct of Census 2021 and NPR was under considerable debate and criticism over the past couple of years, primarily owing to the Citizenship (Amendment) Act (CAA) enacted on 11 December 2019. These became intertwined with political overtones on the issue of the National Register of Citizens (NRC), over which protests swamped the country.

The recent years have seen a massive upsurge in demand for official data; for instance, socio-economic indicators, including those supporting the Sustainable Development Goals (SDGs), censuses and large-scale national surveys. This phenomenal rise in the need for data is balanced by the availability of big volume data in various forms, owing to the use of mobile communications, increased social media interactions and digitally enabled transactions. The use and dependence on conventional data generated by government agencies for public use and validation of other data sets has also become widespread.

### Deferment of Census

The census operations in India governed by the Census Act, 1948 are one of the largest administrative exercises undertaken anywhere in the world. The 2011 Census was the 15th census in an unbroken chain for decennial censuses beginning in 1872 and the seventh since India's independence. The only exceptions have been the withdrawal of operations in

Assam in 1981 and in Jammu and Kashmir in 1991, both due to law and order problems. The actual census now comprises two phases: first, the house-listing operations, wherein all houses (structures), irrespective of their use, are listed and the information pertaining to the state of the house, materials used, household amenities, assets, etc, is recorded. Second, population enumeration, wherein detailed information on each individual residing in the houses is collected and validated. The first phase takes about six months and the second is completed in a short period of three weeks to ensure a more accurate counting of people, taking the sunrise of 1 April as the reference time point. Unlike surveys, the census asks only a limited number of questions (the form for house-listing for the Census 2021 has 31 questions) usually on a single/limited page. The census data are protected to ensure the confidentiality regarding the respondents, and the individual information is not revealed for any purposes other than administrative and research after aggregating to specified domains.

A novelty of the 2021 Census operations is the introduction of handheld devices for recording/capturing data from the households/individuals. This has been introduced for collecting new information as well as updating the NPR. It is in keeping with the practice of shifting most data collection operations to electronic devices for allowing prompt scrutiny and transferring them to secure centralised servers. It is hoped that this will take care of the huge delays in publication of census results—a feature that has proved to be a major obstacle in their effective use for policy formulation and monitoring of the programmes.

### Politics of Data and Trust Deficit

The controversy over the 2021 Census and updating of the NPR along with the census house-listing operations do not come as a surprise, given the government's open commitment to NRC. The strong opposition to the NRC along with the census has led many researchers to worry that there will be non-cooperation from the public, which would ultimately compromise the quality of census data. Kundu and Mohanan (2020) argued on the need to go beyond

politics to address various concerns, including technical and operational questions raised by the linking of the census and NPR.<sup>1</sup> The article points that the population census, with which the NPR is tagged, suffers from a serious coverage error by missing out 2.3% of the population, as determined through the post-enumeration check in 2001 and 2011. The NPR tagged to the house-listing schedule in 2011 prepared the electronic database for only 1,180 million persons, against the census figure of 1,210 million. Considering the undercount of another 30 million in the census itself, it is evident that 60 million people were missed in this NPR exercise.

**Act, Rules and Gazette Notifications Related to Census, Office of the Registrar General and Census Commissioner, India**

- 1 The Census Act, 1948
- 2 The Census Rules, 1990
- 3 The Citizenship Act, 1955
- 4 The Registration of Births and Deaths Act, 1969
- 5 Rules under Section 6A of the IT (Amendment) Act, 2008, Notification No - GSR 316(E) dated 14 April 2011, regarding The Information Technology (Electronic Service Delivery) Rules, 2011—Department of IT, MoCIT
- 6 ORGI Notification dated 23 April 2010
- 7 ORGI Gazette Notification (Amendment the Census Rules 1990) dated 19 June 2018
- 8 ORGI Gazette Notification (Pre-Test Notification for Census 2021) dated 28 June 2019
- 9 ORGI Gazette Notification (Period of House listing Operation and Questions in House listing and Housing Census for Census 2021) dated 7 January 2020

Source: <https://censusindia.gov.in/2011-Act&Rules/index.html>.

Given the present level of decadal migration, roughly 12% of the people would not be found in the place where they were enumerated in 2011. The important question would then be regarding the additional value of the NPR data collected at enormous physical and financial costs. The NPR database is certainly not required for programme targeting purposes, while its utility is yet to be ascertained for any policy purpose. The large-scale exclusion that is inevitable in the NPR exercise will be a nightmare for all usual residents of the country, particularly the economically disadvantaged minorities and other vulnerable population. Given that the government has not released the existing NPR data for public scrutiny, any statement on its accuracy is a matter of conjecture. Furthermore, the apprehension that the NPR would be the base document for NRC

would have qualitative impact on the reliability of the 2021 Census data. Also, the recent migrants, who are likely to have difficulties in obtaining their identity cards or address proofs at their current place of residence, may refrain from identifying themselves as migrants.

There have been suggestions to delink the NPR exercise from the census operations. In such a situation, concerns on the loss of economy of scale, achieved in a combined exercise, would be serious. Amid the current challenges in dealing with the pandemic, the government has justifiably postponed both the census and NPR operations. Coupled with other controversies relating to major government data, this decade does not augur well for the statistical system of the country and warrants urgent response to uphold the integrity and credibility of India's statistical institutions, which have hitherto enjoyed worldwide credibility and appreciation.

**Political Economy of Official Data**

The experiences on the official data from the recent years suggest that the government itself, on occasions, has been ambivalent on various key official statistics, like employment, household consumption, or access to sanitation. On several occasions, the government has produced counter-facts from other non-traditional sources of statistics like the Micro Units Development and Refinance Agency, Jan Dhan Yojana, Employees' Provident Fund Organisation, and so on. Although, the concern over unavailability and quality of the key statistics like employment and unemployment, unorganised sector, migration, etc, have been around for several years, not much has been done to fix this lacuna. However, despite the criticisms, there is no denying that the official economic statistics continue to remain the most valued, appropriate, rich and reliable for informed research and policymaking.

Experts and researchers have unequivocally demanded that, despite the methods and results, the data from census and major surveys must be released so that further analysis could be carried out for course correction. There are apprehensions that the system involved in the elaborate process of sample design, data

collection and reporting may consider making the results congruent to the macro story by matching the numbers as an unstated requirement. This would lead to the concealing and postponement of publication of the reports and, often being released with multiple disclaimers, virtually rendering these unusable. This has been further amplified by the display of reluctance, concealing and withholding of administrative data. The dashboards of the programme management information system (MIS) data (physical and financial progress, monthly reporting) are often managed by the big private consultants as project monitoring units, having "corporate style target achievement attitude" in all the departments and ministries. This has been in trend for the past few years, and has reduced the possibility of detailed analysis or research on these data.

The era of gig economy has further accentuated the need for new forms of data. In the business-as-usual scenario, it will be difficult to adapt to the needs of changing methodology. This is because information and communication technology (ICT) will be extensively used in the days to come to bring new data sets, for example, payroll data, night lights data, GIS data, mobile phone and big data into policy domain. This calls for an urgent revamping of the entire statistical system and making it cohesive to the objectives of "Digital India."

**Statistical Architecture**

One of the key initiatives in reforming the Indian statistical system was the setting up of the Rangarajan Commission in 2000. Though this now is a part of history, some of the outcomes of this report are reasonably fresh. The commission had recommended the establishment of a permanent National Statistical Commission (NSC) entrusted with the responsibility of formulating policies, priorities and standards on statistical matters. However, in the absence of any legislative framework, the NSC has faced challenges in bringing effective reforms. Considering this, the present government has prepared the Draft National Statistical Commission (NSC) Bill, 2019 to adhere to the Rangarajan Commission's recommendations in totality.

In 2018, the Ministry of Statistics and Programme Implementation (MOSPI) had invited public comments on draft policy called National Policy on Official Statistics.<sup>2</sup> MOSPI spearheaded a Five-Year Vision 2019–24,<sup>3</sup> for the National Statistical System and highlighted that the official statistics is a public good and an essential part of the development architecture of India. It recommends extensive uses of digital technology to provide holistic and coherent data on a real-time basis and is committed to address the existing institutional, organisational and technical challenges for policy and stronger dissemination practices for the public. The NITI Aayog released a brief National Data and Analytics Platform (NDAP) Vision Document in January 2020. Proposed to be launched in 2021, it is a flagship initiative that lays down the aim of democratising access to government data through a world-class user experience, by standardising them, providing flexible analytics and ensuring ease of their accessibility. However, in doing so, the enormity of the purpose on making the statistical architecture more open and credible must not be forgotten. This is because official data is used primarily by researchers and policy-makers, and hence, it becomes important to instil trust and confidence among them. It is important that the experts/specialists, committed to the cause of harnessing official data for the greater good are effectively utilised by the government. There is no scope to lose any further time, and the existing initiatives by the Government of India to harness ICT, such as Pro-active Governance and Timely Implementation (PRAGATI), Digital India and JAM (Jan Dhan–Aadhaar–Mobile) trinity to use MIS and dashboards, tax data, etc, needs to be leveraged, while simultaneously restoring the credibility of the statistical institutions.

### The Way Forward

With the Indian statistical system being severely underfunded and understaffed, large investments in financial and human resources are urgently needed for its empowerment. In the fight against the coronavirus under an economic lockdown, the importance of statistical data and statistical systems, and the need for a

robust ecosystem for innovation and accountability have been duly recognised. The most vulnerable and excluded sections, such as floating migrant population, workers in unorganised sector, underemployed and unemployed, worst affected by the pandemic, got excluded from any support from the government. A real-time and dynamic database/registry for unemployed, migrant and informal sector workers would help immensely to tackle the crisis in the medium term. The census is one exercise that provides migration data at district and town levels. India does not have any proper intercensal estimates of population, except the population projections attempted by experts assembled by the Registrar General of India. Organising an intercensal survey, as done in many countries, can mitigate this deficiency in the

Indian data system to a large extent. The road to improving the credibility of official statistics would take a long time and much of that depends on an effective leadership.

Disentangling the overlapping functions of various statistical bodies and creating a coherent structure and empowering the NSC are the best steps forward to ensure credible, accessible and reliable statistics. The system of official statistics in India now faces the challenge of adapting to ICT at par with global standards, and therefore, the emphasis must be on strengthening our statistical institutions to enable evidence-based policymaking, detailing out the programmes and, most importantly, their implementation. The daunting challenges concerning data posed by the CAA–NPR–NRC debate has been exacerbated by the COVID-19 pandemic.

## Centre for Development Economics Delhi School of Economics and The Econometric Society organizing

# Winter School 2020

DECEMBER 14TH-17TH

## Call for Papers:

ACCEPTING  
CONTRIBUTED  
PAPERS FOR THE  
WINTER SCHOOL  
(WS 2020)

Deadline: October 15  
For further details  
contact us at:

ph: +91-11-2700 8122

e: ws2020@econdse.org

w: <http://econdse.org/ws2020/>



Further, postponement of the first phase of 2021 Census operations and the data collection for preparing the NPR indefinitely is sure to upset the timelines for the population count in March 2021. This is because a lot of intervening activities are required to be performed after

the house-listing, before the population enumeration. The postponement of NPR and conducting the first phase of the 2021 Census work will be a logical step to ensure that the census does not suffer from a limited window during which the entire work is expected to be completed.

## NOTES

- 1 Amitabh Kundu and P C Mohanan, "Citizenship (Amendment) Act: How Do We Move Forward?" 21 March 2020, <https://www.epw.in/journal/2020/12/commentary/citizenship-amendment-act.html>.
- 2 [http://mospi.gov.in/sites/default/files/announcements/draft\\_policy\\_17may18.pdf](http://mospi.gov.in/sites/default/files/announcements/draft_policy_17may18.pdf).
- 3 <http://www.mospi.gov.in/sites/default/files/>.