



Adding value to the Census:

Role of Electronic data Collection
in Statistical Data Integration

INDIA

Office of the Registrar General & Census Commissioner, India
Ministry of Home Affairs
New Delhi



Layout of the Presentation

- › Meaning of Data Integration
- › What is Statistical Data Integration
- › Statistical Data Integration: The Need
- › Statistical Data Integration: The Stakeholders
- › Importance of Statistical Data Integration
- › Tools being used for Statistical Data Integration
- › Statistical Data Integration and The Census
- › Data being provided by Census
- › Data Sources that can usefully be linked with Census Data
- › NSDAP: easing the way to integration
- › Census: The present Scenario.....
- › Role of Electronic Data Collection of Census in Statistical Data Integration

Meaning of Data Integration

- › Integrating multiple data sources to produce more consistent, accurate, and useful information than that provided by any individual data source.
- › Combination of technical and business processes used to combine data from disparate sources into meaningful and valuable information.
- › Repurposes data by transforming it.
- › e.g.
consolidating and managing customer information from all available sources by a company

What is Statistical Data Integration

- › Combining data from various sources (Administrative records, Census or Survey) at unit level (individual record) or micro-level (for a small group/geographic area)
- › Linking is done based on common features of two data sources
- › Required information about the individual unit available from multiple sources is being merged to form a single database required for the purpose.

Statistical Data Integration: The Need

- › In general, No single database can serve the purpose of all the stakeholders including Government Ministries/Departments, Research Organizations etc.
- › Often, the data needed by the stakeholders are available in piecemeal from various sources, and there is no need to conduct a fresh study/ survey at all
- › Already existing data may be used in a better way to provide more meaningful information.
- › Data may be integrated either cross-sectional or over time.

Statistical Data Integration: The Stakeholders

- › Custodians of data are the data providers. These may be multiple data agencies based on the requirement.
- › Developers/Statistician assess the need and will combine data from various sources to serve the purpose of the data user
- › Data user is the Ministry/Department or any other organization in need of data.

Importance of Statistical Data Integration

- › Time effective
- › cost effective
- › Reduced duplicity of work

Tools being used for Statistical Data Integration

- Relational data base management system
- Statistical Software's
- Other Need Assessment based custom software's

Statistical Data Integration and The Census

- › Census provides many meaningful information on land and its people
- › It provides information on housing stock and amenities, population composition, literacy, occupation, migration, fertility particulars etc. by age, sex, religion, social category, disability etc.

Data being provided by Census

- › Age-wise, Sex-wise Data on:
 - Population composition
 - Educational Qualification
 - Occupation
 - Workers/Non-workers/Marginal Workers
 - Migration
 - Fertility particulars
 - Disability
- › Religion-wise and caste-wise bifurcation of above data

Data Sources that can usefully be linked with Census Data

- › National Family Health Survey (NFHS)
- › National Sample Survey (NSS)
- › Health Management Information System (HMIS)
- › Unified District Education System for Education (U-DISE)
- › Sample Registration System (SRS)
- › Birth and Death registration data
- › Crime Data

Role of Census for other Surveys/Studies

- › Used for designing of frame for other National Sample Surveys:
 - NFHS
 - NSS
 - SRS

- › Provides base for other Censuses
 - SECC

NSDAP: easing the way to integration

- › In September 1998, the Cabinet approved a proposal for a National Policy on Dissemination of Statistical Data.
- › enabled researchers to access the unit level data from different surveys and laid own certain parameters within which such data could be provided to the public such as confidentiality of individual respondents.
- › The policy laid emphasis on licensed use of unit level data and the need for creating a single point access for all official statistics.

NSDAP: easing the way to integration

Contd/.

- › The present policy goes much beyond the dissemination of statistical data and encompasses all data domains such as scientific data including geo-spatial data, administrative data besides officially compiled statistics.
- › Data will remain the property of the Agency/Department/Ministry/Entity which collected them and reside in their IT enabled facility for sharing and providing access. Access to data under this policy will not be in violation of any Acts and rules of the Government of India in force. Legal framework of this policy will be aligned with various Acts and rules covering the data.

Census: The present Scenario.....

- › At present, in India, data is collected in paper schedules.
- › The schedules are being scanned using ICR technology
- › The digitized data thus achieved is processed and then Primary Census Abstract is released comprising of all the fields having numeric entries.
- › The subjective fields are processed at later stages on the basis of their national importance.
- › Mode of collection of information for Census 2021 is under review

Role of Electronic Data Collection in Census in Statistical Data Integration

- › The task of scanning of Schedules itself took a long time, around eight months, in last Census (including time of transportation).
- › Once the data is collected electronically, no time will need to be invested in transportation of Schedules and scanning task.
- › All the data may be available just after completion of field work of the Census, thus, enabling the fast processing of the data

Role of Electronic Data Collection of Census in Statistical Data Integration

Contd/.

- › Timely availability of Census data
- › Increased usefulness due to decrease in time lag

Challenges of Statistical Data Integration

- › As per provisions of the Census Act, Unit Level data of Census may not be shared.

π

Thank You