MONGOLIAN EXPERIENCE IN USING BIG DATA FOR SDG AND PRODUCTION OF STATISTICS

7TH INTERNATIONAL CONFERENCE ON BIG DATA AND DATA SCIENCE FOR OFFICIAL STATISTICS

8 NOVEMBER 2022
CONTENT

1. ACTIVITIES

2. DATA GOVERNANCE

3. USE OF NON-TRADITIONAL DATA
STATISTICAL PRODUCTION SYSTEM

GSBPM

Specify needs  Design  Build  Collect  Process  Analyse  Disseminate  Evaluate

Sub process

Statistical production system

Data management
- Data catalogue
- Metadata management
- Question pool
- Classification code
- Common code
- GIS Data management

Collection
- Standard sampling
- Data provider management
- Data collection
- Data exchange and validation
- Data exchange
- Validation
- Data extract, transform and load

Analyse
- Analyse
- Standard report
- Dynamic report

Dissemination presentation engine
- Dissemination product
- Dissemination parameters
- Dissemination metadata
- Dissemination API
- Dissemination files
- 1212 Web
- 1212 Mobile App

Questionnaire design

Management
- Monitoring and control
- Quality management
- Task management
- Document management system
- Administration
✓ Monitoring and evaluation of the progress and implementation of the SDGs

✓ Establishment of information flows

✓ Facilitation of the exchange of data between government agencies.
DATA GOVERNANCE

POLICY & LEGAL ENVIRONMENT

VISION 2050 long-term state policy of Mongolia:

Mid term Strategic Plan of NSO /2021-2025/

Improvement of Legal Framework /2022.05.01/

Speaker of the Parliament /2021.05/

Government integrated database

INTER AGENCY COOPERATION

MINISTRY OF DIGITAL DEVELOPMENT AND COMMUNICATION 2022, established

National data strategy

National meta data platform

SUPPORT FROM INT’L ORGANIZATION

PARIS 21

Improve the Statistics law

TA project for Government integrated database

PREPARATION WORKS

A new UNIT for data established

META data analysis - SPS

Assessment of the ADMIN databases by government organizations
INCREASE USAGE OF ADMINISTRATIVE AND BIG DATA

REDUCE THE BURDENS ON DATA PROVIDERS

ELIMINATE STATISTICAL DATA DISCREPANCIES

IMPROVE POLICY DEVELOPMENT CYCLE AND PUBLIC SERVICES

ESTABLISH DATA BASED DECISION MAKING SYSTEM
PROJECT FRAMEWORK

- ASSET
- BANK
- EDUCATION
- TAX OFFICE
- TRANSPORT
- NPA
- SOCIAL INSURANCE
- HEALTH INSURANCE
- GASR
- CUSTOM OFFICE
- GABP /Border Protection/

SEPARATE DATABASE

INCONSISTENT
LOW UTILIZATION
NOT STANDARDIZED

INCOMPLETE MAPPING

Incomplete mapping

COST OF DECISION MAKING

PLANNING, OPTIMIZATION
CASE STUDY – BIG DATA FOR PRICE STATISTICS

**PREPARATION**

- Real time data
- Machine learning

**CALCULATION**

- Volume: Amount of data
  - 89k enterprises
  - 2 million average records in a day

- Machine learning
- Oracle DB
- Oracle Analytic

**RESULT**

- Comparison of weekly commodity prices between traditional and scan data

**COMMODITIES**

- Flour, 1 kg
- Milk, 1 l
- White Rice, 1 kg
- Bread, 600 gr

**Advantages:**
- Reducing the financial burden
- Coverage increased: more than 80 thousand enterprises scanning data
- Data quality: Real time prices and data precision
- Reducing the burden on statistical surveys (interviewers and respondents)

**Disadvantages:**
- Not fully collected consumer goods in CPI basket
- Data storage and big data analyzing infrastructure are required
- To access to and collect data
EXPERIENCE IN CROP DATA USING RS TECHNIQUES

Project implementation plan and methodology

01 STATISTICAL DATA
- Sown/harvested area
- Harvest by crop type
- Cadastral registration
- On-site study data

02 SATELLITE IMAGE
- Sentinel 2 image
- Multispectral drone image
- Spectroradiometer measurement

03 ANALYSIS
- Random forest
  - Train/test dataset
  - Accuracy assessment
- Spectral angle mapper (SAM)
  - To process dataset of field measurement

04 RESULT
- Overall accuracy - 76.9%
- Kappa coefficient - 63.2%
  - By wheat, rapeseed, potato
- Overall accuracy - 92.9%
- Kappa coefficient - 89%
  - By cabbage, beetroot, oats

05 TO INTRODUCE/ IMPLEMENT
As for the methodology, our experience shows:
- Apply the SAM method for crop type classification,
- Apply the random forest for crop yield estimation,
Further,
- Develop an open-source system similar to sen2agri
- Solve human resource and budget problem
USE OF NON-TRADITIONAL DATA

Utilization of geospatial data and statistical database for production of SDG indicator
(Experimental estimation, ADB-NSO Project, 2019)

Indicator 9.1.1. Proportion of the rural population who live within 2 km of an all-season road - 38.7%

Data source:
• Open street map
• Road network data (Ministry of Road and Transportation)
• Population statistics
THANK YOU FOR YOUR ATTENTION