NATIONAL STATISTICS OFFICE

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

ACTIVITIES

320 EMPOYEES IN THE NSO

District units
UB Statistics department
Provincial units
NSO HQ

SPS

54 questionnaires

908.7 thous.HH

227.9 thous. Enterprises

STATE ORGANIZATIONS

Agencies
Other state organizations (MB, FRC etc)
Ministries

Without statistical unit

50%

50%

With statistical unit

1. Road and Transportation
2. Defense
3. Environment and Tourism
4. Justice and Internal Affairs
5. Energy
6. Foreign Affairs
7. Finance
8. Economic Development

1. Education and Science
2. Health
3. Culture
4. Labor and social protection
5. Mining and Heavy Industry
6. Construction and Urban Development
7. Food, Agriculture and Light Industry
8. Digital Development and Communication

ONLINE SERVICE

National Data Center

2140 PERSONS IN THE PRIMARY ADMINISTRATIVE UNITS

Khoroo governors 172 persons
Soum level treasury officers 330 persons
Bag governors 1638 persons

PAHIS

Livestock census

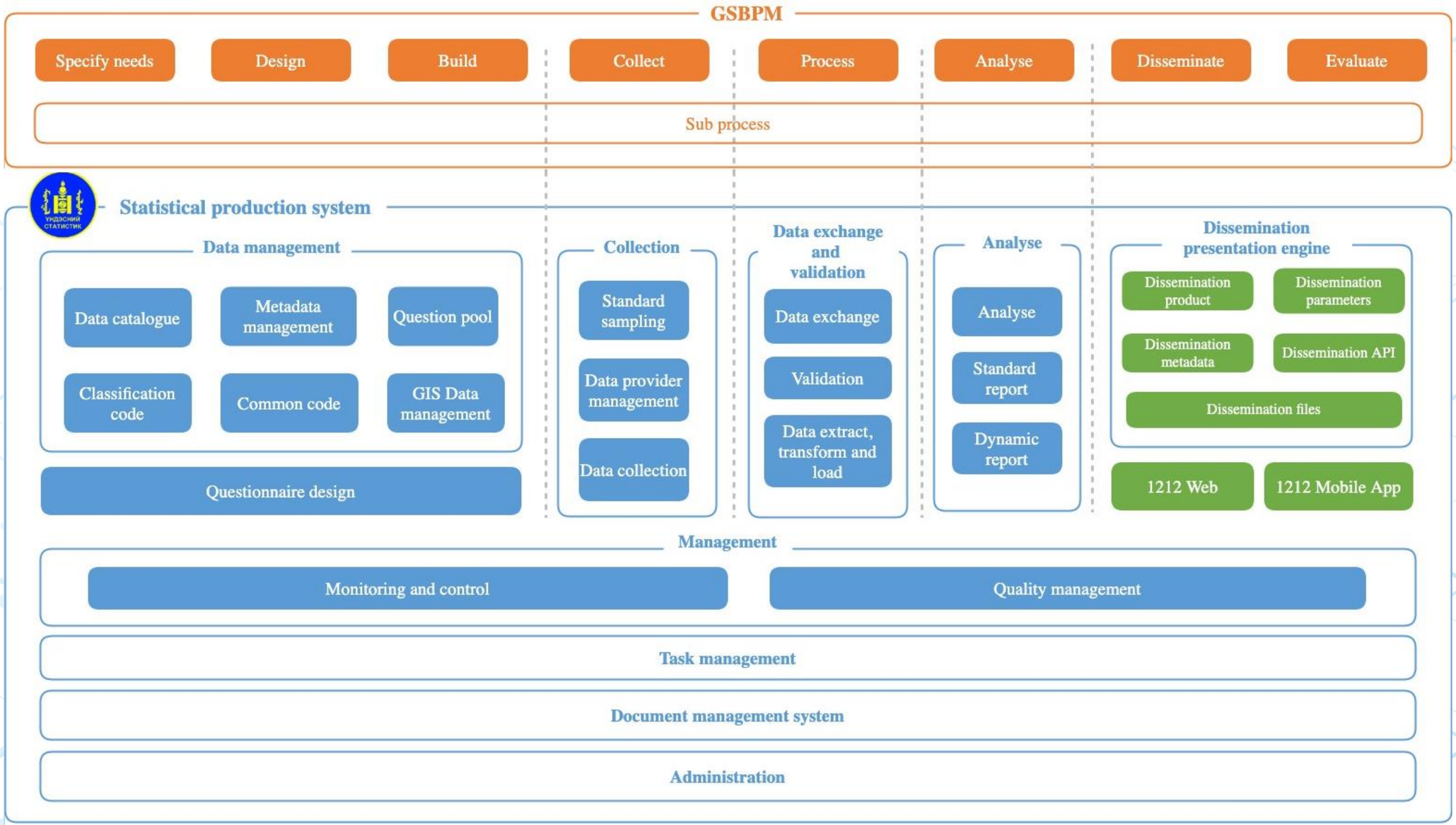
Demographic statistics

Livestock, crop, industry and service

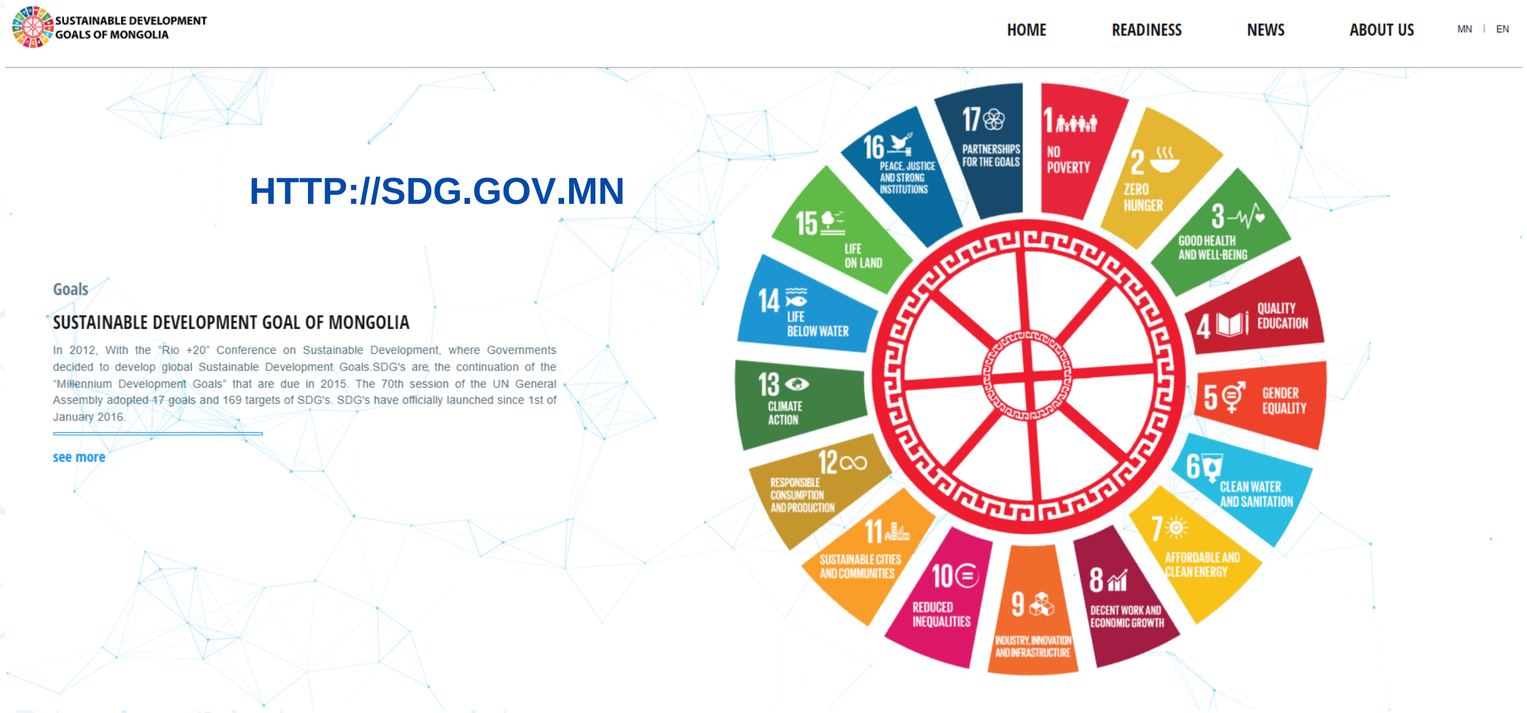
79 Methodologies

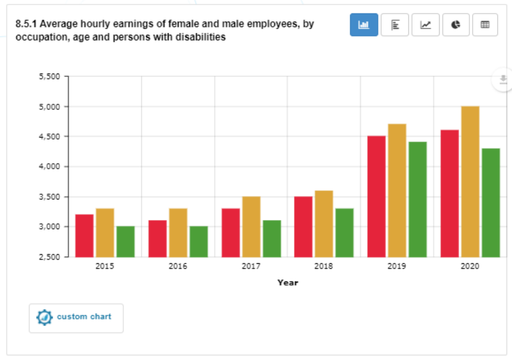
27 Classifications

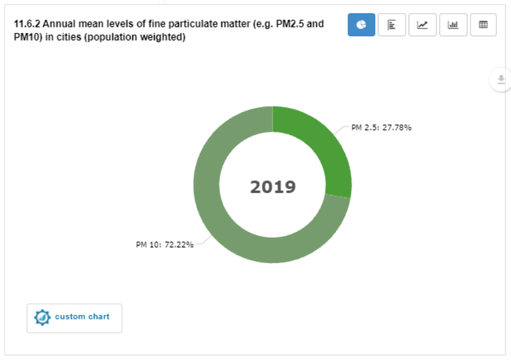
STATISTICAL PRODUCTION SYSTEM

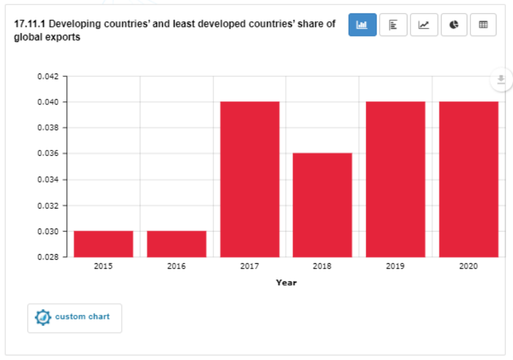


SDG IMPLEMENTATION MONITORING SYSTEM









- ✓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

GOVERMENT OF MONGOLIA

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

ADB

TA project for Government integrated database

PREPARATION WORKS

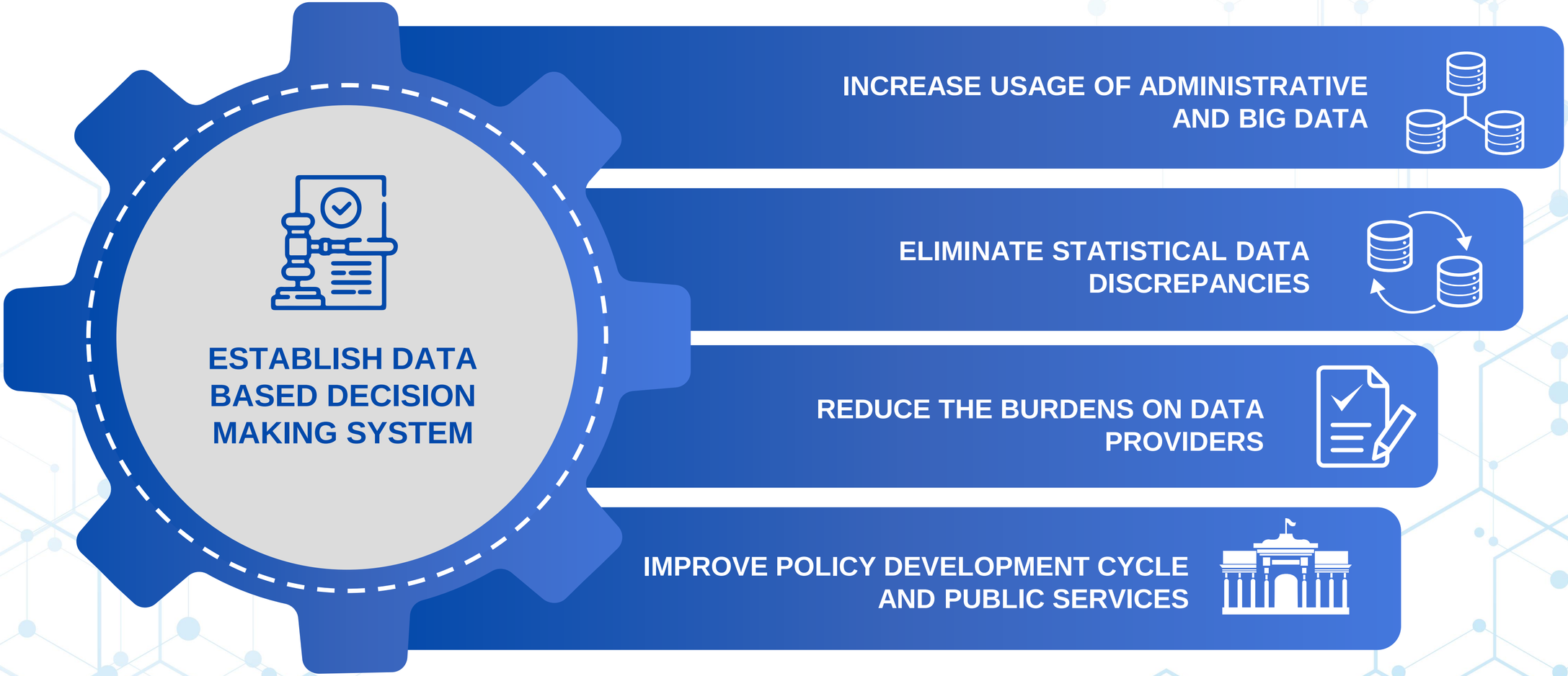
A new UNIT for data established

META data analysis - SPS

Assessment of the ADMIN databases by government organizations

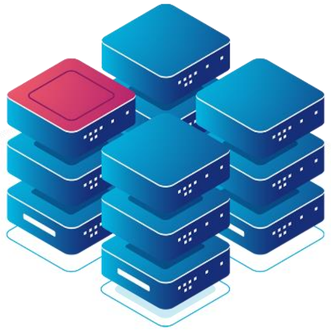
PROJECT FRAMEWORK

Goal and tasks



PROJECT FRAMEWORK

Incomplete mapping



SEPARATE DATABASE

ASSET

SOCIAL INSURANCE

HEALTH INSURANCE

BANK

GASR

EDUCATION

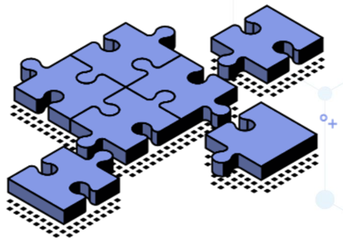
CUSTOM OFFICE

GABP /Border Protection/

TAX OFFICE

TRANSPORT

NPA



INCOMPLETE MAPPING

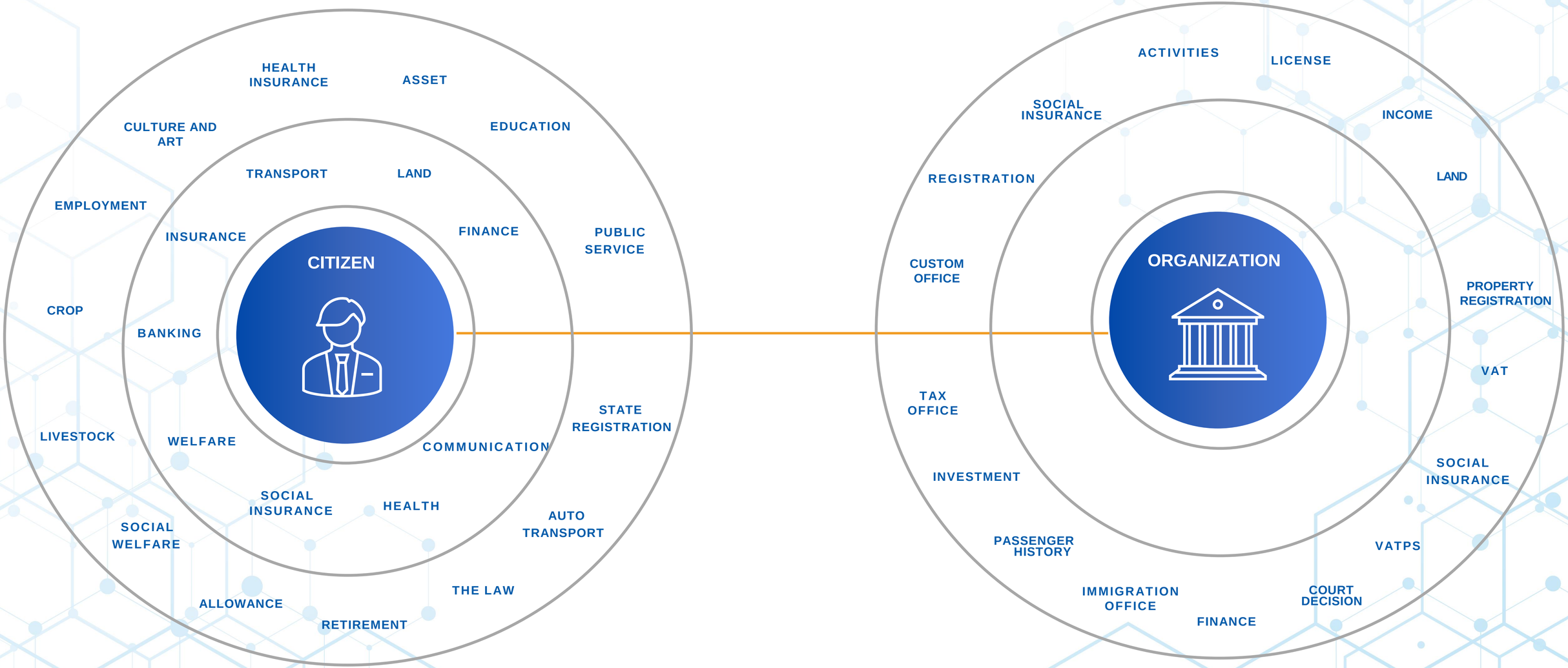
COST OF DECISION MAKING

PLANNING, OPTIMIZATION

- INCONSISTENT
- LOW UTILIZATION
- NOT STANDARDIZED

PROJECT FRAMEWORK

Complete mapping

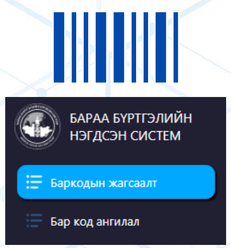


CASE STUDY – BIG DATA FOR PRICE STATISTICS

PREPARATION

Real time data

Machine learning

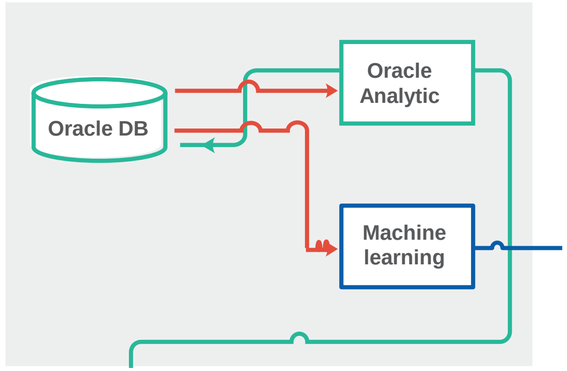


Volume: Amount of data

= 89k enterprises

= 2 million average records in a day

CALCULATION



RESULT

COMMODITIES
/days, weeks, months/

32

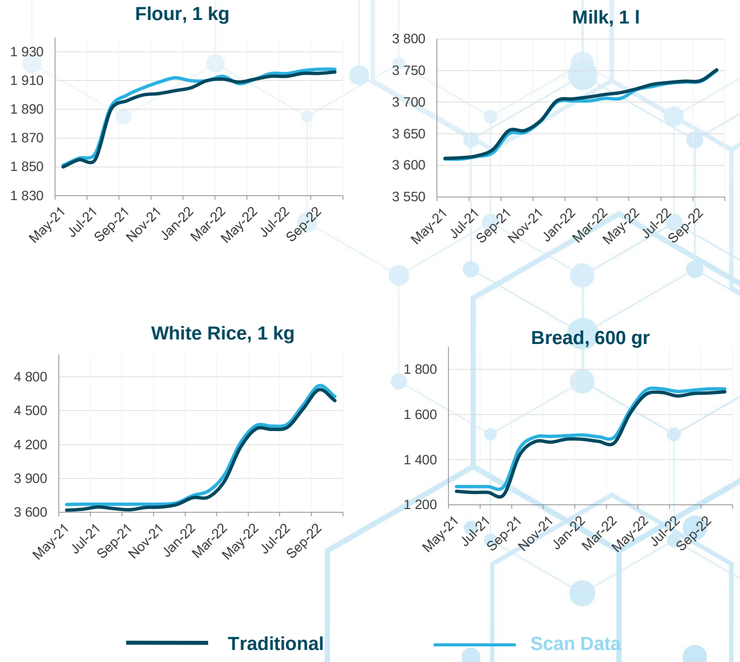
ALL SUPERMARKETS

10

SMs COVERED IN PILOT

7

Comparison of weekly commodity prices between traditional and scan data



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,

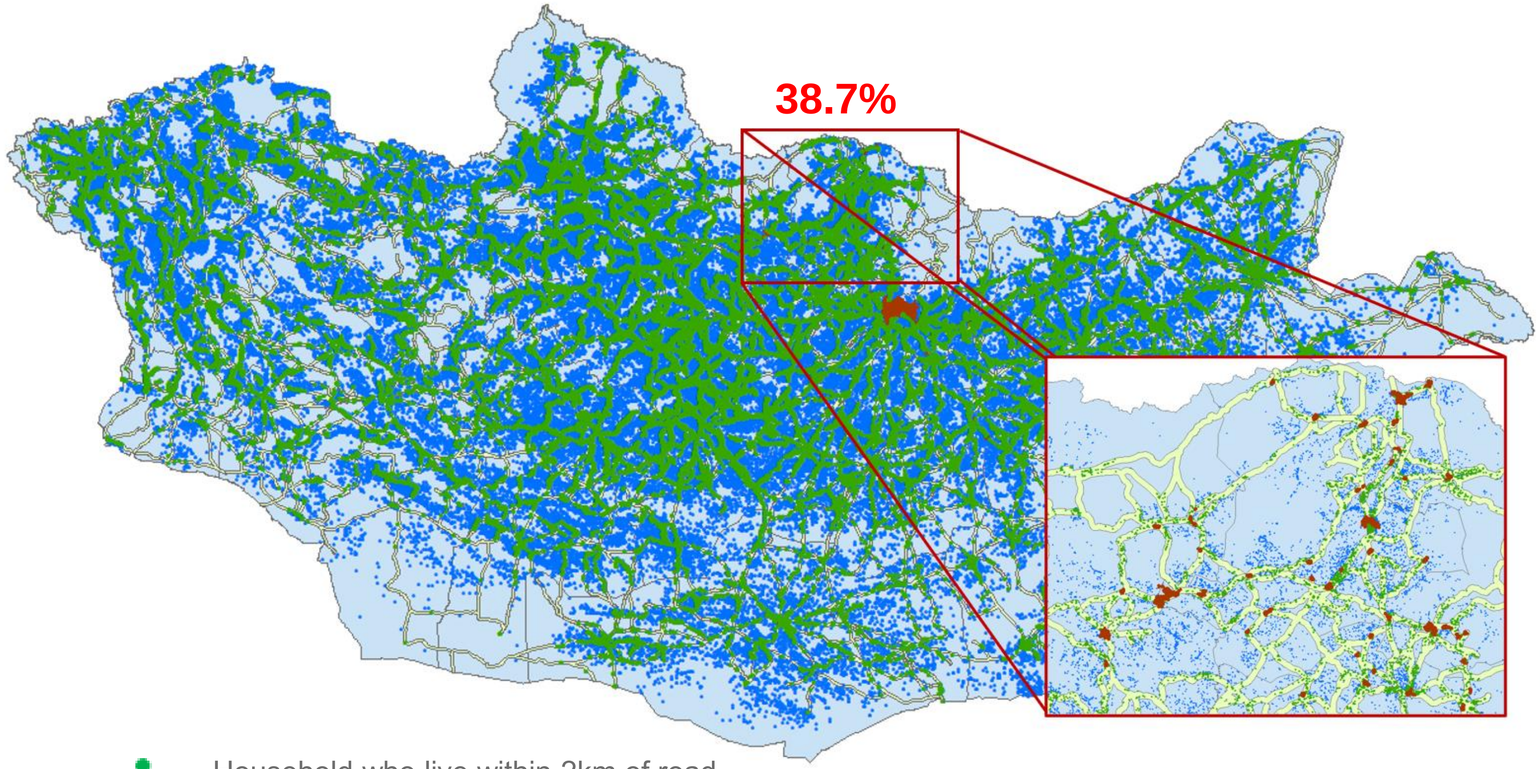
Further,
- Develop an open-source system similar to sen2agri
- Solve human resource and budget problem

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 -

Data source:
- Open street map
- Road network data (Ministry of Road and Transportation)
- Population statistics



NATIONAL
STATISTICS
OFFICE

THANK YOU FOR YOUR ATTENTION