



**BRICS**  
2022 CHINA



# BIG DATA IN OFFICIAL STATISTICS

## RUSSIA'S CASE

**Federal State  
Statistics Service**  
ROSSTAT

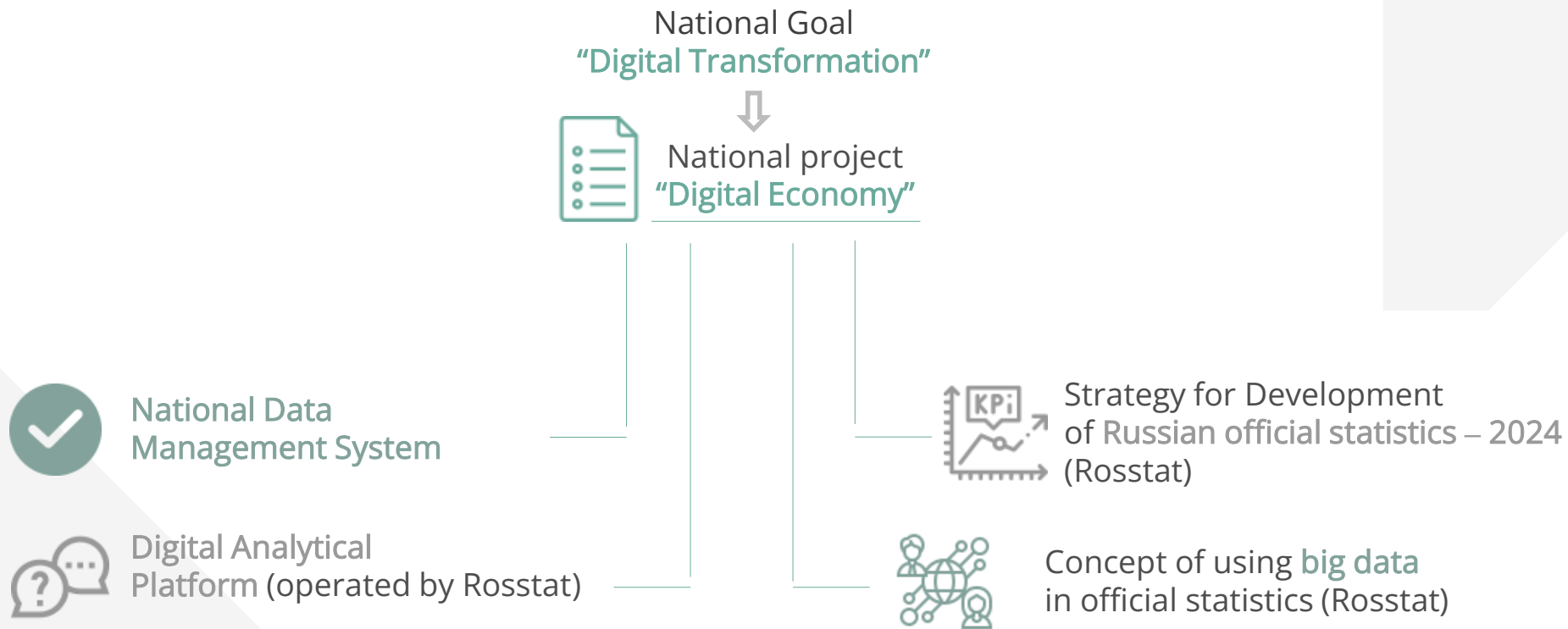


**Sergey Egorenko,**  
Deputy Head of Rosstat

The 14-th meeting of BRICS' Heads of NSO's  
October 11, 2022



# Regulatory Framework





# Main advantages of using big data

in official statistics

## Advantage 1

Reduced processing time of aggregated data, transition to a continuous surveys and calculation of official statistics

## Advantage 2

Much more detailed geographical, social and economic statistical data (increased samples, multiplied selections of activities, goods and services, geospatial information)



## Advantage 3

Increased data quality by the use of advanced technologies (artificial intelligence, neural networks) for generating official statistics, new indicators, assess activities which are not currently reported/observed

## Advantage 4

Reduced reporting burden on respondents and expenditures on official statistics for federal and regional budgets

## Advantage 5

Quick wins (current pilots by Rosstat): consumer price and retail trade statistics, population statistics, tourism statistics

# Consumer price & retail trade statistics

traditional + new data sources



## Traditional surveys



registrars  
(mobile devices)



Rosstat retail prices  
statistics & CPI



e-reporting  
systems



Rosstat retail  
trade statistics

## Marked goods



administrative data  
(federal register)



data matrix



prices and volumes  
of the marked goods sold

## Federal tax service



cash register equipment  
of retail sellers  
(federal data base)



e-receipts



prices and volumes of goods  
and services sold

## Online-sales



web-scraping

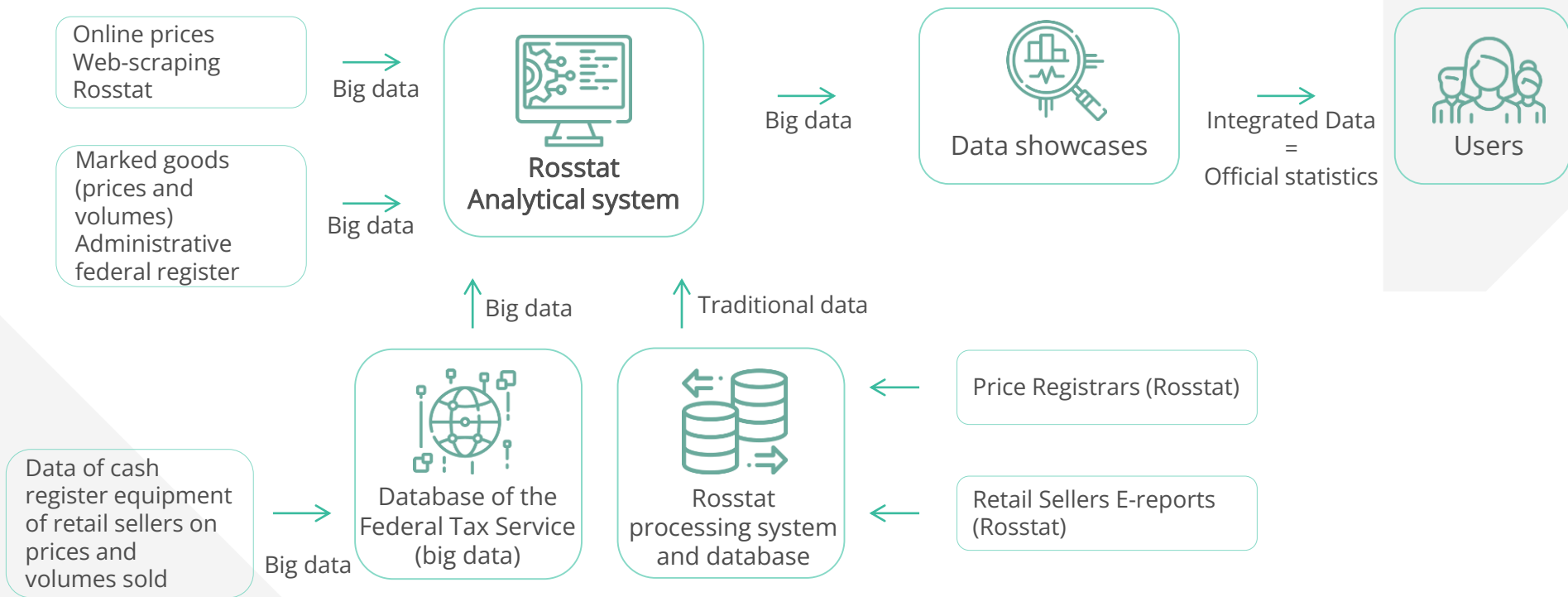


WEB API



price offers from the websites  
of retail sellers  
and service providers

# Data production process using big data



# Population statistics. Big data sources

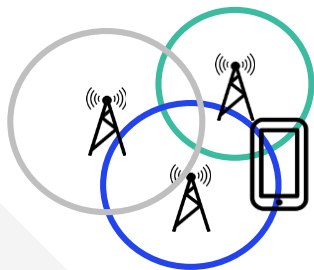


With the help of machine learning methods based on the behavioral characteristics of subscribers, data about their location is supplemented with such characteristics as gender, age, citizenship

Based on data from mobile operators

1

Subscriber triangulation

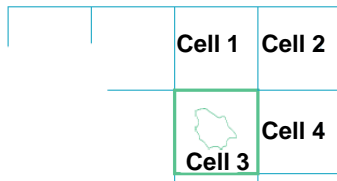


The operator registers events at stations that have a time stamp:

- calls
- messages
- internet traffic

2

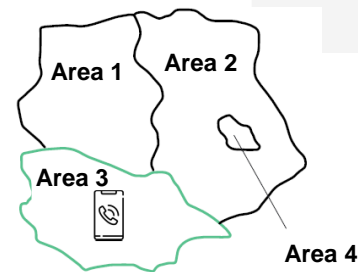
Determining the subscriber's location in a regular grid



Based on the subscriber's activity, its belonging to the basic territorial cell is determined

3

Determining the location territory



Based on the spatial affiliation of the basic territorial cell to a subject of the Russian Federation or a municipality, the territory is determined



# Formation of the **base of initial geoanalytical indicators** by operators

Unification and consolidation of operator reports into a system of linked tables with **understandable** names

Directories of territories

Directory of socio-economic segments

Population size

Population updated

Flows of pendulum migrants

Dynamics of the flows of pendulum migrants

## BASE OF INITIAL GEOANALYTICAL INDICATORS

Directory of subjects of the Russian Federation

Directory of Municipalities

Flows of pendulum migrants

- Operator 1
- Operator 2
- Operator 3

February 2020, October 2021

Flows of pendulum migrants

- Operator 1
- Operator 2
- Operator 3

February 2020, October 2021

Population size

- Operator 1
- Operator 2
- Operator 3

April 2020, October 2021

Population size

- Operator 1
- Operator 2
- Operator 3

April 2020, October 2021

Dynamics of migrant pendulum flows

- Operator 1
- Operator 2
- Operator 3

February 2020, October 2021

Dynamics of migrant pendulum flows

- Operator 1
- Operator 2
- Operator 3

February 2020, October 2021

Updated population size

- Operator 1
- Operator 2
- Operator 3

April 2020, October 2021

Updated population size

- Operator 1
- Operator 2
- Operator 3

April 2020, October 2021

Assessment of changes in administrative data for the subjects of the Russian Federation

October 2020

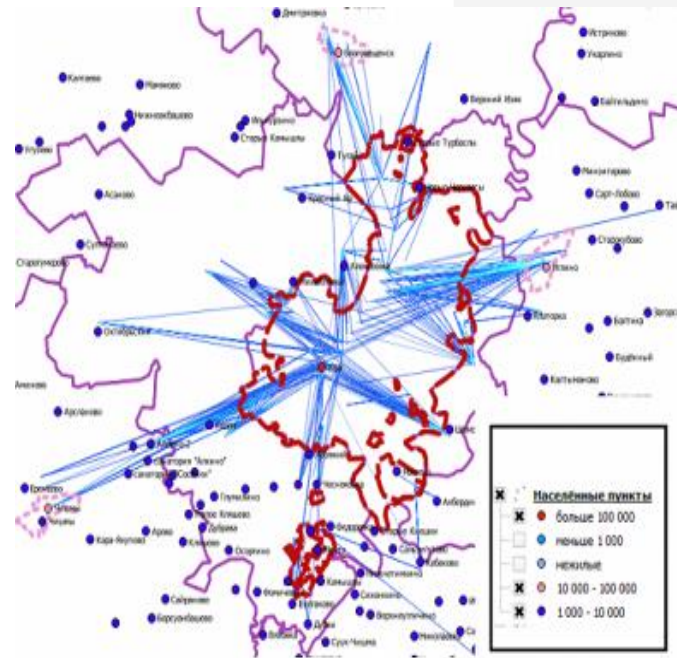
Assessment of changes based on administrative data for municipalities

October 2020

# Results of the analysis of mobile operator databases



- population density mapping
- assessment of population mobility and pendulum migration
- behavioral characteristics of the population in the field of healthcare
- behavioral characteristics of the population in the field of work
- characteristics of the population in terms of living standards, inequality and poverty
- behavioral characteristics of the population in terms of offenses
- characteristics of the behavior of the population in terms of movements for tourism
- behavioral characteristics of the population in terms of the use of mobile services (mobile) communications





# Tourism Statistics



## BD Sources



Data of mobile operators



Data of credit institutions



Data of aggregators of air tickets and railway tickets



Data of hotel aggregators



Social media data



Data of video recording systems (transport)

## Quick wins

- Unified and compatible statistics on federal and regional levels
- Fast and predictive estimates of tourists flows
- New segments included - individual accommodation facilities and “shadow services”
- Much more detailed statistics in time and areas of tourists interest, e.t.c.
- Reduced respondents reporting burden & federal budget expenditures



Thank you for  
your attention!

E-mail: [international@gks.ru](mailto:international@gks.ru)