**Meeting of the Heads
of the National Statistical Offices of
the BSEC Member States**

Moscow, Russia, June 22-23, 2016

BSEC

CURRENT AND EMERGING CHALLENGES FACING THE RUSSIAN STATISTICAL SYSTEM

CHALLENGES

Short-term and mid-term challenge

There is an understanding of the essence of a problem and how to solve it (maybe even in general terms only, not in detail). In order to create and implement the action plan, methodological guidelines should be developed, an adequate legal basis should be created, financial resources should be allocated.

Long-term challenge

A new phenomenon coming into existence is being observed. This phenomenon can have a substantial impact on official statistics (its methods, data sources, technologies, organizational structure). However, there is no clear understanding how to deal with it.

SHORT-TERM AND MID-TERM CHALLENGES - AND HOW TO MEET THEM

BSEC

♦ To implement the latest versions of ISIC and the CPC

♦ To improve the Business Register quality in strict accordance with international recommendations

♦ To use the establishment as the basic statistical unit for production accounts

♦ To capture phenomena generated by globalization

♦ To develop the production and reporting of small area statistics

LONG-TERM CHALLENGES

♦ To develop methodology and techniques to compile the Sustainable Development Goal indicators

♦ Using Big Data for Official Statistics

♦ Data Scientist vs. Statistician

♦ Modernization of higher professional education system in the field of Statistics

SHORT-TERM CHALLENGE. CLASSIFICATION OF ECONOMIC ACTIVITIES

Russia	EU	UN
ОКВЭД-2001 previous version	NACE rev.1 (1996)	ISIC rev.3
ОКВЭД-2007 currently valid version	NACE rev.1.1 (2002)	ISIC rev.3.1
ОКВЭД 2 expected to be implemented in early 2017	NACE rev.2 (2008)	ISIC rev.4

SHORT-TERM CHALLENGE. PRODUCT CLASSIFICATION

Russia	EU	UN
ОКПД – 2007 currently valid version	CPA – 2002	CPC rev.1
ОКПД 2 expected to be implemented in early 2017	CPA – 2008	CPC rev.2

As a minimum, the business register should include:

(a) Name and physical location of each enterprise;

(b) Mailing address, which may be different from its physical location;

(c) Name and address of the central office of the enterprise and establishments that are part of a multi-establishment enterprise;

(d) Kind of economic activity: description or code;

(e) Legal organization;

(f) Type of ownership

(g) Number of persons employed;

(h) Volume of sales or value of output;

(i) Source and date of information.

International Recommendations for Industrial Statistics 2008 (IRIS-2008), para 7.21

It is unlikely that statistical business registers can be satisfactorily compiled and maintained solely through surveys and the stand-alone efforts of the national statistical office.

Common examples of administrative data sources that may be used to create and support business registers include business registration systems, VAT tax systems, payroll tax systems, and records maintained by the Government for the administration of unemployment insurance, social security or other Government programmes.

(IRIS-2008, para 7.10, 7.13)

Countries are encouraged to use the establishment as a statistical unit for industrial statistics so as to ensure the homogeneity of the economic activity and its geographical distribution. However, the enterprise also can be used as the statistical unit.

(IRIS-2008, para 2.71)

The establishment is defined as the combination of activities and resources directed by a single owning or controlling entity towards the production of the most homogeneous group of goods and services, usually at one location, but sometimes over a wider area, for which separate records are available which can provide data concerning the production of these goods and services and the materials, labor and physical resources used in that production.

(IRIS-2008, para 2.64)

A principal that owns the material inputs and thereby has economic ownership of the outputs, but has the production carried out by others, is classified to the classification category that corresponds to the complete (outsourced) manufacturing activity;

The contractor in such a case is classified to section C (Manufacturing) of ISIC Rev.4, specifically to the classification category that corresponds to the manufacturing activity performed by the contractor.

(IRIS-2008, para 1.25)

There is no problem with the country-level statistics. However, there is a problem for the regional-level statistics.

With the increasing demand to produce statistics for small areas where sample surveys may be difficult to implement, administrative records represent a useful alternative.

(Guidelines on Integrated Economic Statistics OOH, 2013, para 5.98)

Traditionally statistical registers have been used as sampling frames for surveys, but they are increasingly being seen as sources of statistical data in their own right, particularly regarding data for small geographical areas, or small sub-groups of the population.

(Using Administrative and Secondary Sources for Official Statistics, UN, 2011, Section 7.2)

The SDGs require a 'data revolution' to collect and analyze disaggregated data to monitor progress.

(Achieving the global goals: why volunteers are important. - 22 March 2016)

The data revolution for development: a community of researchers and policy makers is trying to capture and strengthen old forms of traditional statistics as well as capture new forms of innovative data. By this we mean technology driven big data, but also more qualitative participatory data and perception studies.

(Exploring the role of research in the Data Revolution for development - 24 April 2015)

The users' demand for information is subject to change over time due to changes in the underlying factors:

MDGs
(60 indicators)

SDGs
(about 225 indicators)

they're broader, more complex, more numerous, they cover more topics, they're intended to be integrated and they're intended to be universal.

The data revolution is:

♦ An explosion in the volume of data, the speed with which data are produced, the number of producers of data, the dissemination of data, and the range of things on which there is data, coming from new technologies and from other sources, such as qualitative data, citizen-generated data and perceptions data.

♦ A growing demand for data from all parts of society.

(A World that Counts – November 2014)

In a nutshell, the volume of data is dramatically increasing; the users' demand for information is changing over time rapidly

Actual reported big data projects (within the scope of the Global survey on Big Data)**:**

- mobile phone data (42 projects)
- web-scraping data (31 projects)
- scanner data (23 projects)

Statistical domains for which big data was most used:

- price statistics (based on scanner data)
- tourism (mobile phone data)
- population (mobile phone data)
- transport and labor statistics (web-scraping data)

As a first step, Rosstat is focusing on price statistics.

15

"Which skills are important for your office to acquire in order to better deal with big data?":

- ▪ "methodologist on big data issues"
- ▪ "data scientist"
- ▪ "mathematical modeling specialist"

The more IT-oriented skills of "IT architecture specialist", "data visualization specialist" and "cybersecurity specialist" got less priority.

Pete Skomoroch, a Principal Data Scientist at LinkedIn:

Data scientist: better statistician than most programmers & better programmer than most statisticians.

**ICT Outsourcing vs. Insourcing:
What's best for the NSOs?**

The ICT outsourcing prevails in Rosstat nowadays. However, ICT outsourcing entails some risks. So some discussion takes place between experts.

Insourcing might work for some tasks and outsourcing for others.

The raw materials of data science are not independent data sets, but heterogeneous, unstructured data set of all kinds, – e.g., text, images, video. The data scientist will not simply analyze the data, but will look at it from many angles, with the hope of discovering new insights.

http://insidebigdata.com, May 22, 2014

Statistics is a part of data science, not the whole thing. Statistics research focuses on data collection and modeling, and there is little work on developing good questions, thinking about the shape of data, communicating results or building data products.

Hadley Wickham, IMS Bulletin, September 4, 2014

International Standard Classification of Education	Russian Classification of Professions by education

054 Mathematics and statistics
 0542 Statistics
    Actuarial science
    Probability theory
    Statistics, applied
    Survey design
    Survey sampling
	080000 Economics and management
	 080600 Statistics

THANK YOU FOR ATTENTION!

THIS PRESENTATION HAS BEEN COMPILED BY DR. IGOR ULIYANOV, ROSSTAT