Introduction to field survey in Poland

Janusz Dygaszewicz

Central Statistical Office of Poland

Jerusalem, 11-14 July 2016

The 2011 population Census as the basis for introducing modern data collection methods into surveys in Poland

Mixed Model for Population and Housing Census

Mixed model - a combining data from administrative sources

(full survey covering base demographic variables – short form)

with data acquired from 20% sample survey

(long form)

The IT Census System
Census architecture

For the purposes of census design and conduction, the Central Statistical Office of Poland implemented **the IT Census System (ICS)**

The ICS integrated various technologies (from applications installed on mobile terminals, through applications managing and assisting in telephone interviews, to specialist bases, data warehouses and analytical and reporting tools)

Data collection channels in 2010 Census Round

Administrative Sources

- Including spatial data reference registers

Self-enumeration by Internet

- **CAII** – Computer Aided Internet Interview

Telephone Interview

- **CATI** - Computer Aided Telephone Interview

Interview with respondents realized by the census enumerator

- Registered on hand-held terminals with usage GPS and GIS service **CAPI** - Computer Assisted Personal Interview

CAxI channels
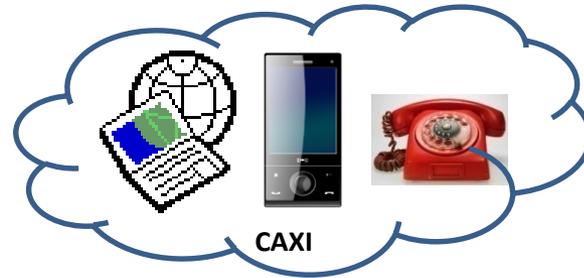
CAxI

- CAII - Computer Assisted Internet Interview,
- CAPI - Computer Assisted Personal Interview,
- CATI - Computer Assisted Telephone Interviewing.

Administrative source utilisation

Obtaining data from administrative sources for surveys

The entitlement of official statistics to use data from adminsitrative sources stem from legal provisions of

- the Act on official statistics of 29 June 1995

- the programme of statistical surveys of official statistics

Administrative sources

Data from administrative systems was used in the census:

- as a direct source of census data (personalisation of questionnaires)

- to create:

  - compilations of buildings, dwellings and persons,

  - an address-residence register

  - a sampling frame

The utilisation of administrative sources in statistics

the replacing of statistical forms

+

in the compilation of survey results

+

sampling frame quality improvement

=

better and cheaper information

Registers - data acquisition

Data Owners:

- Ministry of Finance,
- Ministry of Interior and Administration,
- Ministry of Justice,
- Agricultural Social Insurance Fund,
- National Health Fund,
- Agency for Restructuring and Modernisation of Agriculture,
- Agricultural and Food Quality Inspection,
- Agency for Geodesy and Cartography,
- State Fund for Rehabilitation of Disabled Persons,
- County Offices,
- Commune Offices,
- Regional Offices,
- Telcoms,
- Energy Suppliers,
- Office For Foreigners,
- Social Insurance Institution,
- Housing Managers,

Administrative sources

➢ The unit data obtained from registers are converted into statistical registers, simultaneously being subject to the process of cleaning, de-duplication and standardisation of data. The process is carried out in the DQS SAS environment. At the same time, metadata are collected on the quality of input data obtained from registers, the applied cleaning procedures and the final quality obtained after applying DQS procedures.

➢ The cleaned data are loaded into the Operational Microdata Base as successive logical layers corresponding to the obtained registers.

Operational Microdata Base

- **OMB was realized processes of :**
    - cleaning,

    - integration and verification of data,

    - correcting data,

    - data processing,

    - generating operational reports,

    - depersonalizing data and export of data (as Golden Record) to Analytical Base.

Golden Record generation

Integration with Census Frame and CAxI data,

Validation,

Correction,

Operational Imputation,

Transfer proper values to Golden Record,

Registers 1..n

CAxI

Golden Record
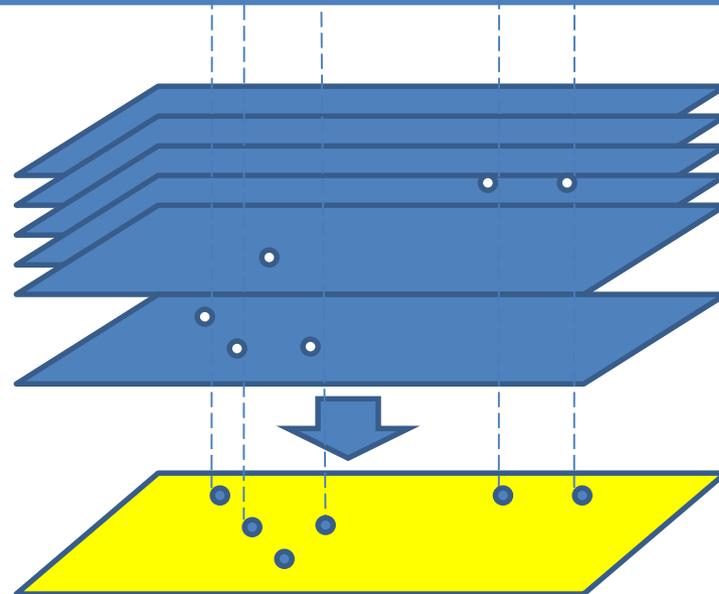
OMB Layers

AMB

On-line channels for data collection

Operational Microdata Base

Census Completeness Management

List of census
Questionarries
The control data
History of contacts with the respondent

CAII

CATI

CAPI

Architecture solution

Stage I – Preparatory works

Stage II – Data acquisition

Stage III – Results compilation

XML

Registry 1

TXT

Registry 2

Registry n

Files

ETL Tools

Statistical Files

CAXI

Questionaries

Operational Microdata Base

Golden Record

Analitycal Microdata Base

SDMX

Metadata

Metadata

Metadata

XML

Portal

Metadata server

The main objectives of AMB

Preparation and dissemination census data for statistical analysis

Dissemination of analytical and reporting platform for development of census data collected and preparation of statistical products for national and international users.

Supporting for process of creation and dissemination statistical products

Census Hub

The Metainformation System

- **<u>The Metainformation System</u>** gathered necessary metainformation describing data and census processes, including the processes indispensable to drawing up quality reports. The task of the Metainformation System was to ensure the coherent definition of statistical objects for the OMB and AMB.

Metainformation Subsystem gathered the following metadata:

- methodological,

- operational,

- systems,

- definitional,

- quality.

Model of Metainformation Systems
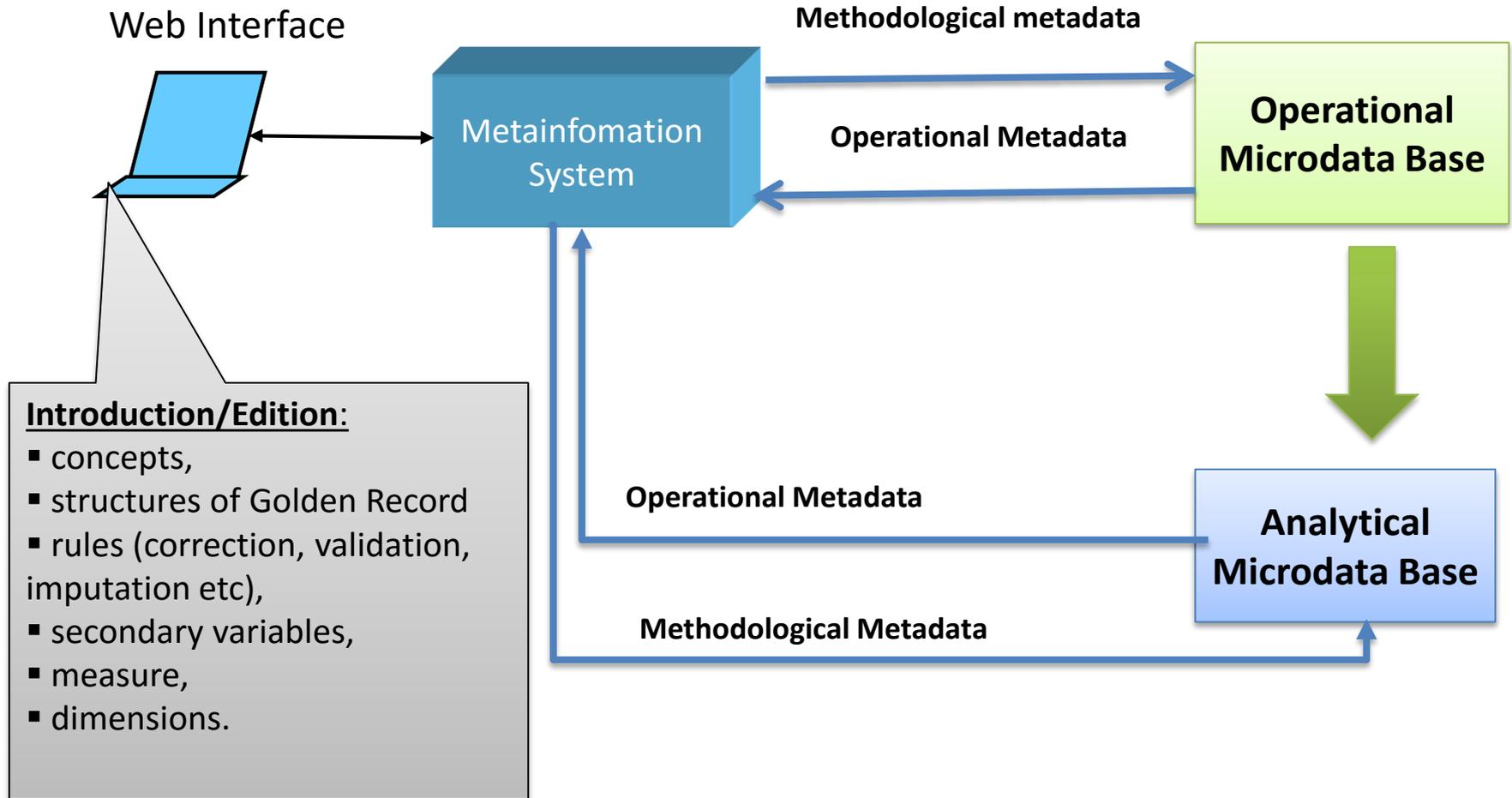
Instead of a conclusion

Census in 2002

- 180 thousands of census enumerators

- 120 mln of questionnaires

- 1 000 tons of papers

- At the end shredding census questionnaires

Census 2011

- 18 thousands of census enumerators

- 0 questionnaires

- 0 tons of papers

- **ca. 50 mln € less**

- better data
- the more reliable results
- statistical surveys in the future

National censuses were implemented in compliance with the Generic Statistical Business Process Model.

Lesson learned:

The customized GSBPM model is being implemented for current statistical surveys with additional GIS components concerning linkage of statistical data with spatial data.

Generic Statistical Business Process Model

UNECE/Eurostat/OEC

Quality Management/Metadata Management							

1 Specify Needs	2 Design	3 Build	4 Collect	5 Process	6 Analyse	7 Disseminate	8 Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.1 a Geocode frame & sample	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Build or enhance dissemination components	4.2 Set up collection	5.3 Review & validate	6.2 a Prepare spatial analysis & maps	7.2 a Manage spatial analysis & maps using GIS	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.3 Run collection	5.4 Edit & imput	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system	4.3 a Geocode collection	5.5 Derive new variables & units	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.6 Prepare business case	2.5a Design geocoding frame, sample & data collection	3.6 Test statistical business process	4.4 Finalise collection	5.6 Calculate weights	6.5 Finalise outputs	7.5 Manage user support	
	2.6 Design production system & workflow	3.7 Finalise production system		5.7 Calculate aggregates			
				5.8 Finalise data files			

The implementation of surveys using the interviewer network

Modern data collection methods used in National Censuses PSR 2010 and NSP 2011 (CAII, CATI, CAPI) are being introduced in surveys.

Legal regulations

❑ **Act of 29 June 1995** on official statistics - lays out the rules and creates a foundation for reliable, objective, professional and independent performance of statistical surveys, the results of which have the status of official statistical data, and determines the organisation and mode of conducting such surveys and the range of responsibilities associated with them.

❑ Statistical surveys conducted in a given year are specified in the form of annual programmes of statistical surveys of official statistics and are determined by the Council of Ministers by way of a regulation.

❑ Templates of reporting forms and explanations of the method of filling them in, as well as templates of statistical questionnaires and forms used in statistical surveys specified in the programme of statistical surveys of official statistics for a given year are determined by way of a Regulation of the President of the Council of Ministers (according to last amendment of the Act in 2018 it will be changed by electronic formats of collected data)

Statistical confidentiality
– a guarantee of data confidentiality

Art. 10. Identifiable microdata collected in statistical surveys are **subject to unconditional protection** This data might be used only for statistical studies, comparisons and analyses and for creating a sampling frame for statistical surveys by the President of the Central Statistical Office; making available and using this data for purposes other than specified in the Act is prohibited (statistical confidentiality)

Surveys conducted by the survey network

❑ Surveys provided for in the programme of statistical surveys of official statistics are conducted in the following scope:

- price statistics

- social statistics

- agricultural statistics

❑ Payed surveys implemented by order pursuant to Art. 21 of the Act of 29 June 1995 on official statistics for the whole country or for a given voivodship

Social surveys in 2016

| Social surveys	Frequency	Annual sample size	The annual estimated number of surveys
Labour Force Survey (individual form)	continuous/quarterly implementation	221,520 dwellings	504,321
Labour Force Survey (household form)	continuous/quarterly implementation	221,520 dwellings	239,242
Young people on the labour market	one, 2nd quarter	55,380 dwellings	39,537
Unpaid work outside the household	every 4-5 years	27,350 dwellings	62,266
The condition of households	monthly	32,400 dwellings	34,992
Expenditure on environmental protection incurred by the household by expenditure and elements of the environment	every 3 years	1,350 dwellings	1,458
A survey of household budgets (form BR-04)	quarterly	37,584 dwellings	40,590
A survey of household budgets	monthly	112,752 dwellings	121,722
A survey of fuel and energy consumption in households	every 3 years	4,700 dwellings	5,076
Participation in sports and recreational activity	every 4 years	4,700 dwellings	5,076

Social surveys in 2016 continued

Social surveys/Price surveys	Frequency	Annual sample size	The annual estimated number of surveys
Information and communication technology (households)	once a year	8,100 dwellings	8,748
Information and communication technology (individual)	once a year	8,100 dwellings	16,813
European Union Statistics on Income and Living Conditions (household)	once a year	18,000 dwellings	19,440
European Union Statistics on Income and Living Conditions (individual)	once a year	39,600 people	39,600
Access to services	cyclically	18,000 dwellings/ 39,600 people	59,040
The participation of Polish residents in travels	quarterly	75,000 dwellings	81,000
Retail price quotes	monthly	approx. 18,540 representatives	
Marketplace price quotes for major agricultural products	monthly	440 marketplaces	
Total			**1 278 921**

Agricultural surveys in 2016

Agricultural surveys	Frequency	Annual sample size	Estimated number of surveys
A survey of winter cereal yields	once a year	13,000 farms	13,000
A survey of potato yields	once a year	13,000 farms	13,000
A survey of cereal, rape and agrimony yields	once a year	18,000 farms	18,000
A survey of certain crop yields	once a year	18,000 farms	18,000
Farm Structure Survey	every 3 years	180,000 farms	180,000
Cattle, sheep and poultry stock and animal output survey	twice a year	80,000 farms	80,000
Pigs stock and animal output survey	three times a year	90,000 farms	90,000
A survey of the economic situation in farms	twice a year	51,000 farms	51,000
Total			**463,000**

Implementing CAxI methods for surveys

2012

CAPI and CATI methods

in agricultural surveys

2013

CAPI and CATI methods

in social surveys

2013

CAII method

in FSS survey

2013

CAPI method

in surveys of retail prices and in FSS survey
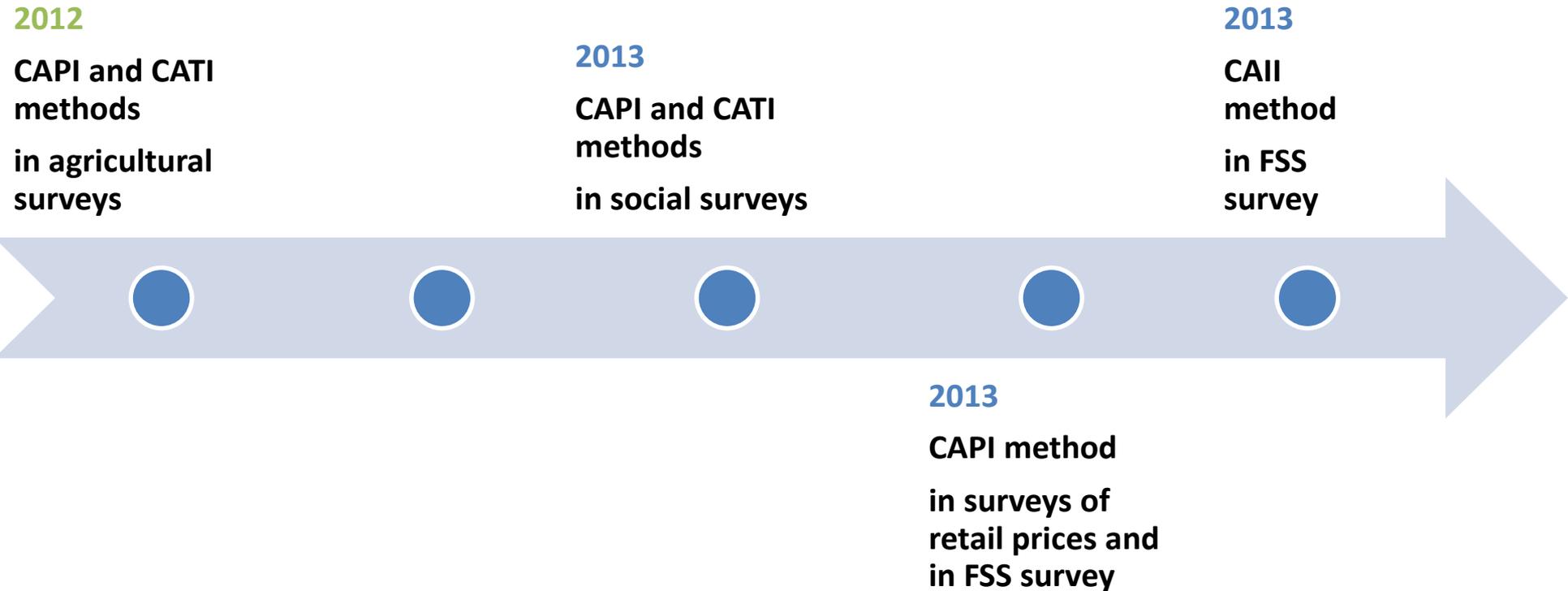
Data collection methods
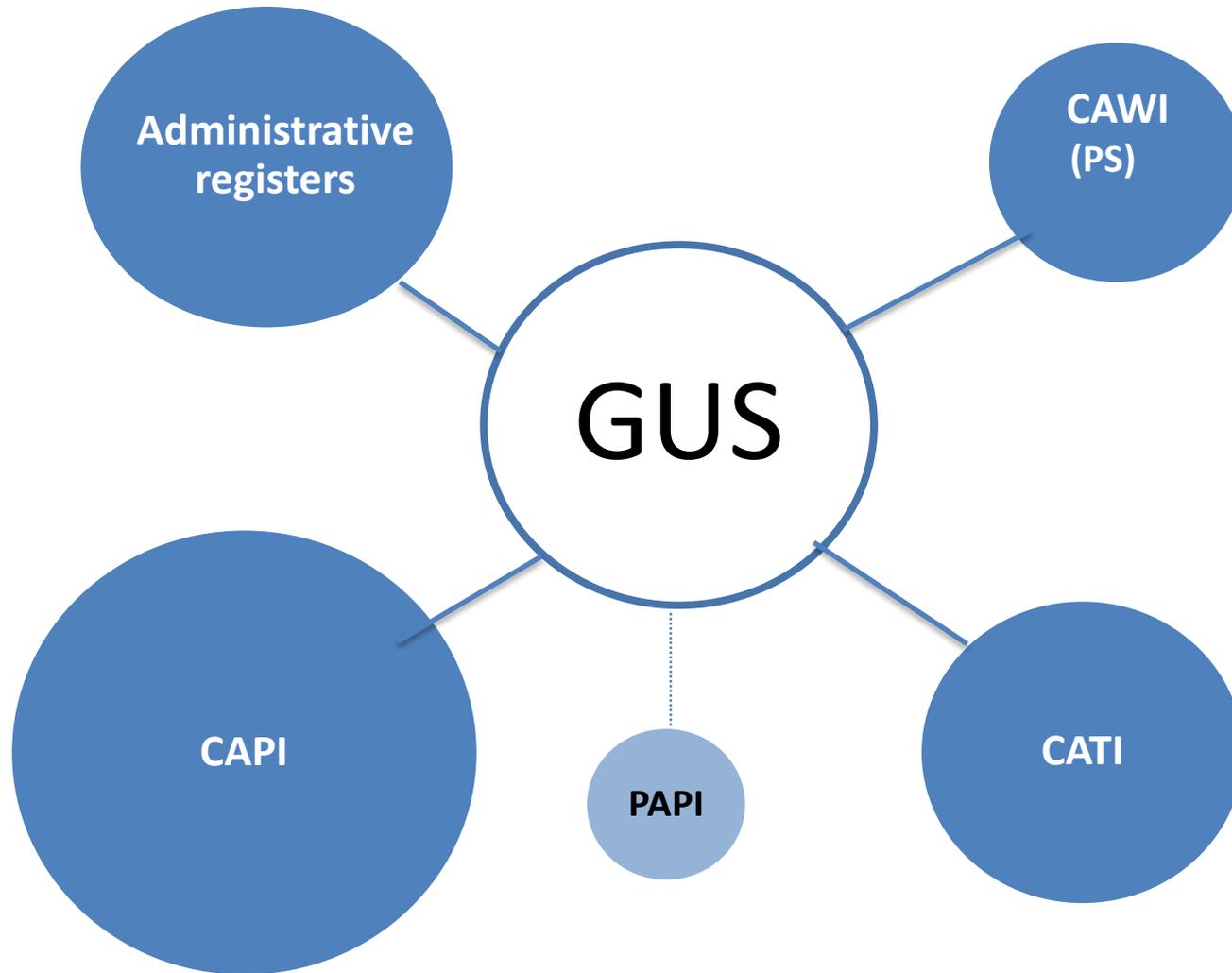
Survey network

Official Statistics Organisation

```
                        President of
                        the CSO

   CSO          Statistical                          Statistical
                Offices x 16                         Information
                                                     Centre

   Departments/   Branches    Divisions              Statistical
   Offices                                            Publishing
                                                      Establishment

                                                      Centre for
                                                      Statistical
                                                      Education and
                                                      Research

                                                      Central
                                                      Statistical
                                                      Library
```
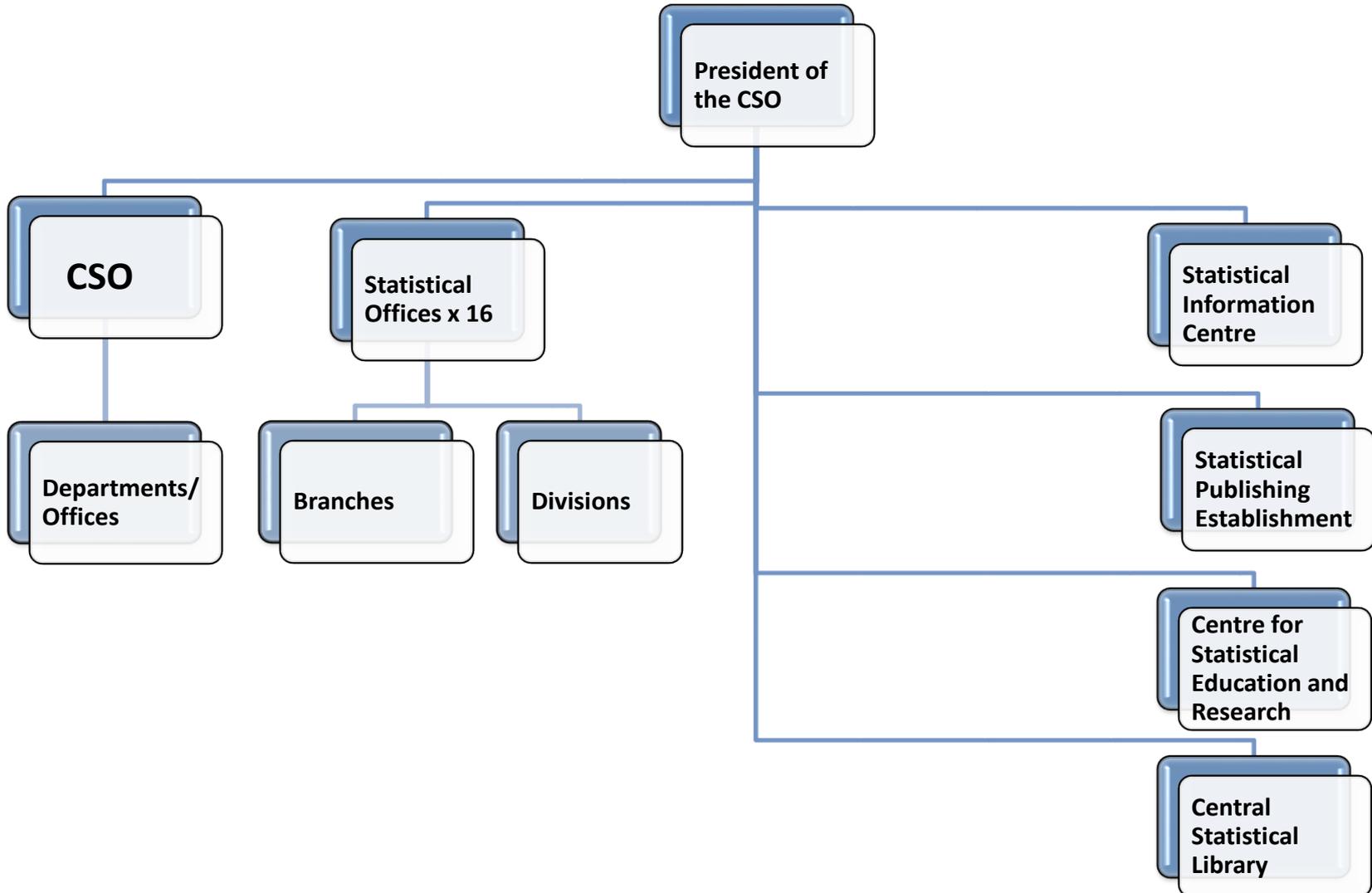
The organisation of Survey Departments
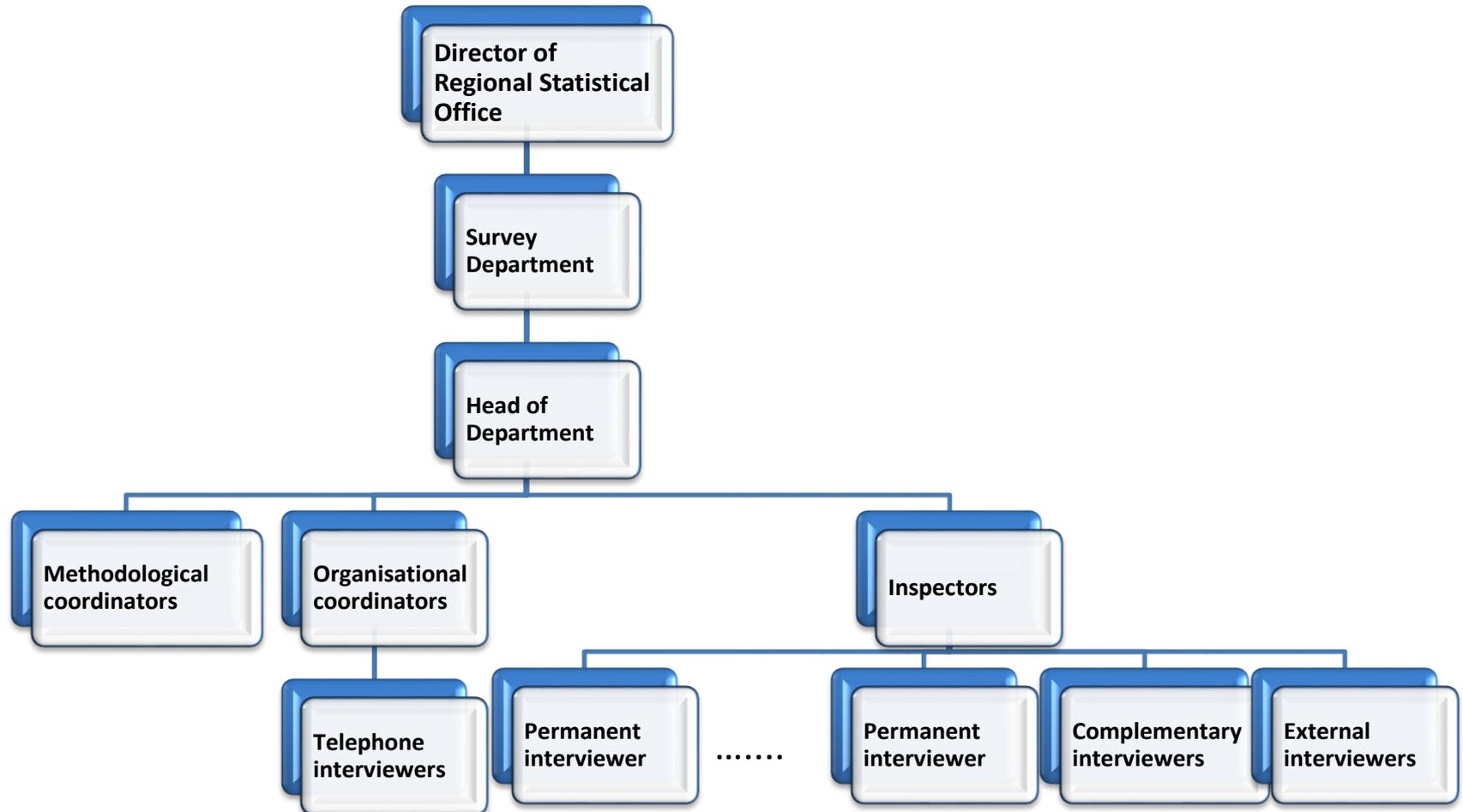
The division of interviewers

PERMANENT interviewers – persons employed by statistical offices outside civil service, who conduct filed interviews using the CAPI/PAPI methods

TELEPHONE INTERVIEWERS – temporarily engaged employees of statistical offices who conduct surveys using the CATI method

COMPLEMENTARY INTERVIEWERS – employees of statistical offices who are engaged at times of the greatest load of surveying work, when there are not enough interviewers to conduct the surveys

External interviewers – persons employed based on civil-law agreements, who conduct large-scale surveys (censuses, farm structure)

Survey Department

Head of SD – persons responsible for conducting field surveys

Methodological coordinator – a person with methodological knowledge of individual surveys, responsible for verifying the correctness of the forms completed by interviewers and telephone interviewers, and training telephone interviewers, inspectors and interviewers

Organisational coordinator – a person who manages the flow of units between data collection channels, monitors the course of the surveys and managing the Call Center, receiver of clearances who creates user accounts in CORstat
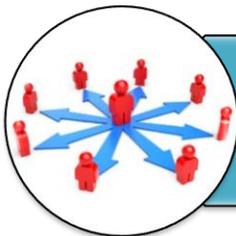
Inspector – a person who directly supervises the work of field interviewers

The number of employees at Survey Departments (FTEs) in 2015

No.	Voivodship	Number of technical coordinators	Number of organisational coordinators	Number of inspectors	Number of permanent interviewers	Total Full Time
1	2	4	5	6	7	8
1	Dolnośląskie	5	2	17	79	103
2	Kujawsko-pomorskie	5	2	13	53	73
3	Lubelskie	5	2	14	58	79
4	Lubuskie	5	1	8	33	47
5	Łódzkie	5	2	16	65	88
6	Małopolskie	5	2	16	75	98
7	Mazowieckie	6	3	26	111	146
8	Opolskie	5	1	8	38	52
9	Podkarpackie	5	2	12	49	68
10	Podlaskie	5	2	11	47	65
11	Pomorskie	5	2	15	62	84
12	Śląskie	5	4	22	86	117
13	Świętokrzyskie	5	1	6	38	50
14	Warmińsko-mazurskie	4	1	10	46	61
15	Wielkopolskie	4	2	14	73	93
16	Zachodniopomorskie	5	2	12	48	67
	Total country-wide	79	31	220	961	1 291

The task of interviewers

The task of interviewers

- conducting field surveys,
- observing statistical confidentiality rules as well as those on protection of personal data,
- transferring data from interviews to the dedicated IT systems, in line with the specified procedure,
- taking care of the quality and completeness of the collected data and the timely execution of entrusted tasks,
- improving one's qualifications in the field of operation of interview-assisting tools and technical knowledge of conducted surveys, through independent learning or participation in training courses
- taking care of the public image of official statistics

The task of telephone interviewers

The task of telephone interviewers

- conducting telephone interviews,
- providing respondents with information on the conducted survey and operating the helpline,
- arraigning surveys on the respondents' premises (if necessary),
- participating in training courses in the technical field of the surveys, techniques of phone interviewing and the operation of telephone interviewing equipment.
- independent improving of their qualifications (e-learning, m-learning).

Work tools for interviewers

hand-held mobile devices used in PSR 2010 and NSP 2011

tablets and laptops currently used in surveys

desktop computers/laptops with Internet access and appropriate software, in the case of conducting CAII surveys, with interviewer view (illustrated with an example of an e-book for household surveying)

Work tools for telephone interviewers

Workstations with a form application for conducting interviews

SIP (*Session Initiation Protocol)* Stations

Headset

Interviewer network management

The central level

- planning the interviewer network load based on the programme of the statistical surveys of official statistics,

- estimating the interviewer network load with surveys, after sampling

- annual determination of the recommended number of interviewer FTEs for individual statistical offices, and telephone interviewer positions at the national level,

- determining the needs in terms of equipping the interviewers and telephone interviewers with data-collection tools,

- coordinating the use of individual data-collection methods in surveys,

The CSO

Interviewer network management

The voivodship level

- employing interviewers, allocating interviews and interviewing areas
- allocating work to telephone interviewers
- organising and conducting technical training in the field of operating interviewing/telephone-interviewing devices and applications, at the voivodship level
- monitoring and accounting for the work of interviewers and telephone interviewers, including work-quality control
- financial settlements with interviewers
- engaging, if necessary, complementary interviewers, external interviewers and agricultural valuers

16 statistical offices

Assumptions for the calculation of the number of permanent interviewers

The following data are used for the purposes of determining the recommended number of FTEs of permanent interviewers:

- the number of interviews in surveys referred to the CAPI and PAPI methods in individual months and offices; in price surveys, the number of regions and representatives in the survey were assumed,

- the average time of conducting the interview in minutes for every survey – specified by the survey authors,

- the number of productive working days in individual months,

- the daily productive hours of the interviewer's work.

A list of surveys conducted in 2016

	January	February	March	April	May	June	July	August	September	October	November	December
BBGD	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨
BAEL	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨
KGD	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨
PKZ	🟥			🟥			🟥			🟥		
PNZ	🟥											
E-GD	🟥											
A-KR	🟥						🟥					
OS-GD	🟥	🟥	🟥									
R-ZW-S			🟥			🟥	🟥					🟥
ZD-I				🟥	🟥	🟥	🟥					
SSI				🟥	🟥							
EU-SILC					🟥	🟥	🟥					
R-ZW-B						🟥	🟥					🟥
R-SGR						🟥	🟥					
R-r-zb								🟥				
DS.-52										🟥		
R-r-pw											🟥	
Price surveys	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨
Marketplace price surveys	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨

The number of FTEs of interviewers resulting from monthly load (as calculated) in relation to the number of FTEs resulting from average annual, standardised load

■ liczba etatów ankieterów wg. struktury   ■ liczba etatów ankieterów wg. obciążeń
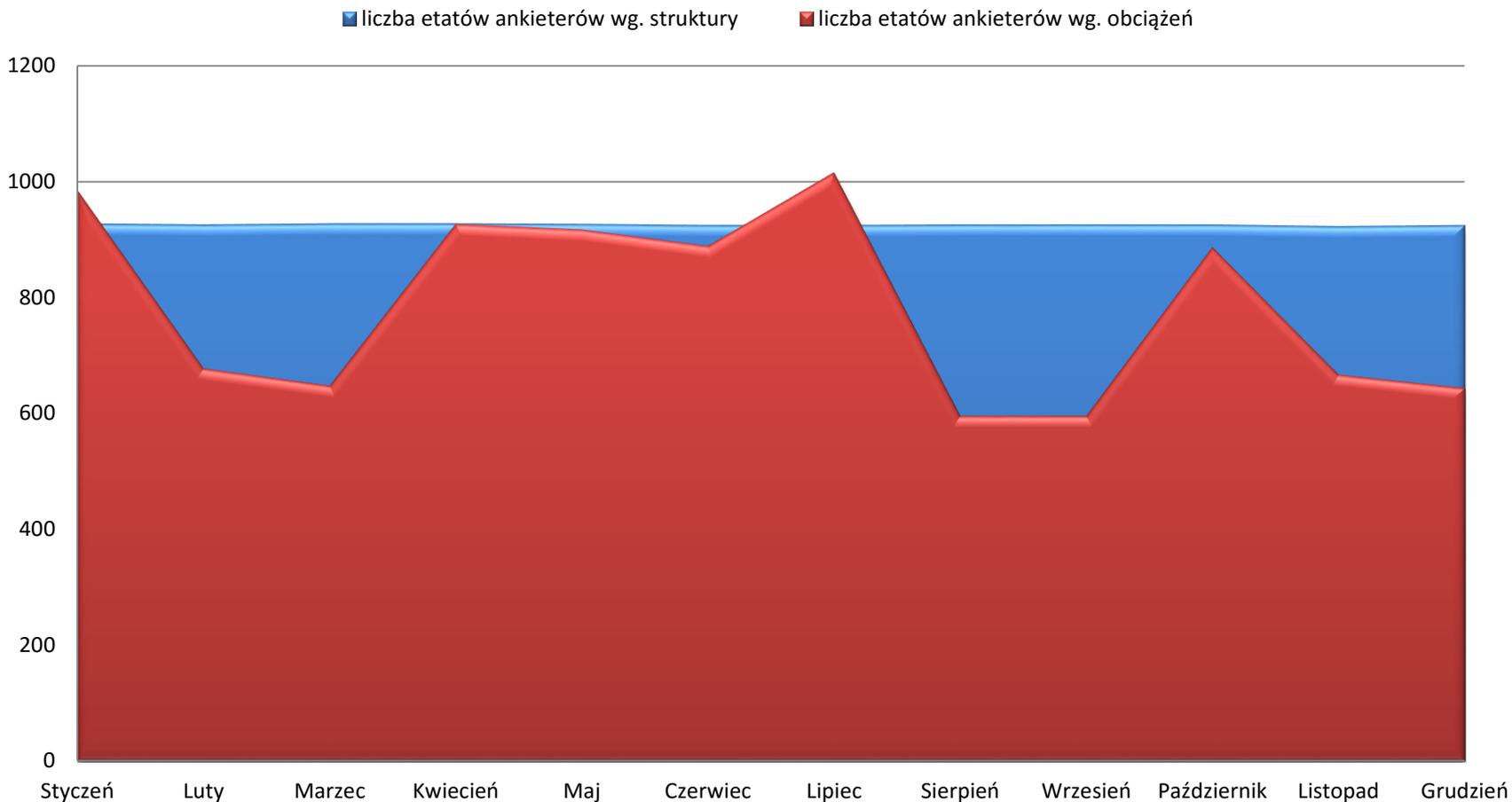
The sampling strategy as an optimisation method

❑ **Stratification**

The stratification criteria employed are:

- geographic location (voivodship, subregion),
- city/town size, type (rural or urban),
- the fact whether a given gmina is a border gmina is also taken into account in GDP surveys.

❑ **Allocation in the strata**

Two-stage allocation:

– allocation in voivodships proportionately to the square root of the number of dwellings in voivodships,

– within the voivodship, allocation proportionately to the number of dwellings in the stratum.

The sampling strategy

❑ First stage units

First stage units (FSU) are census enumeration areas or statistical units (one unit consist no more then 9 EA)

❑ 1st stage sampling

1st stage sampling is the sampling with possibility proportionate to the number of dwellings in a FSU, according to the Hartley-Rao scheme

The sampling strategy

❑ **2nd stage sampling**

The procedure of simple random sampling is used for the sampling of dwellings from a FSU

❑ **The number of dwellings sampled from a FSU**

Usually 4-6 dwellings are sampled, which will be surveyed at the same time

❑ **Rotation**

It means that one sampled dwelling is surveyed two or more times at specified intervals

Sample distribution in square root sampling

Legend:
- < 12 %
- 12 - 18 %
- 18 - 22 %
- 22 - 32 %
- 32 - 42 %
- > 42 %
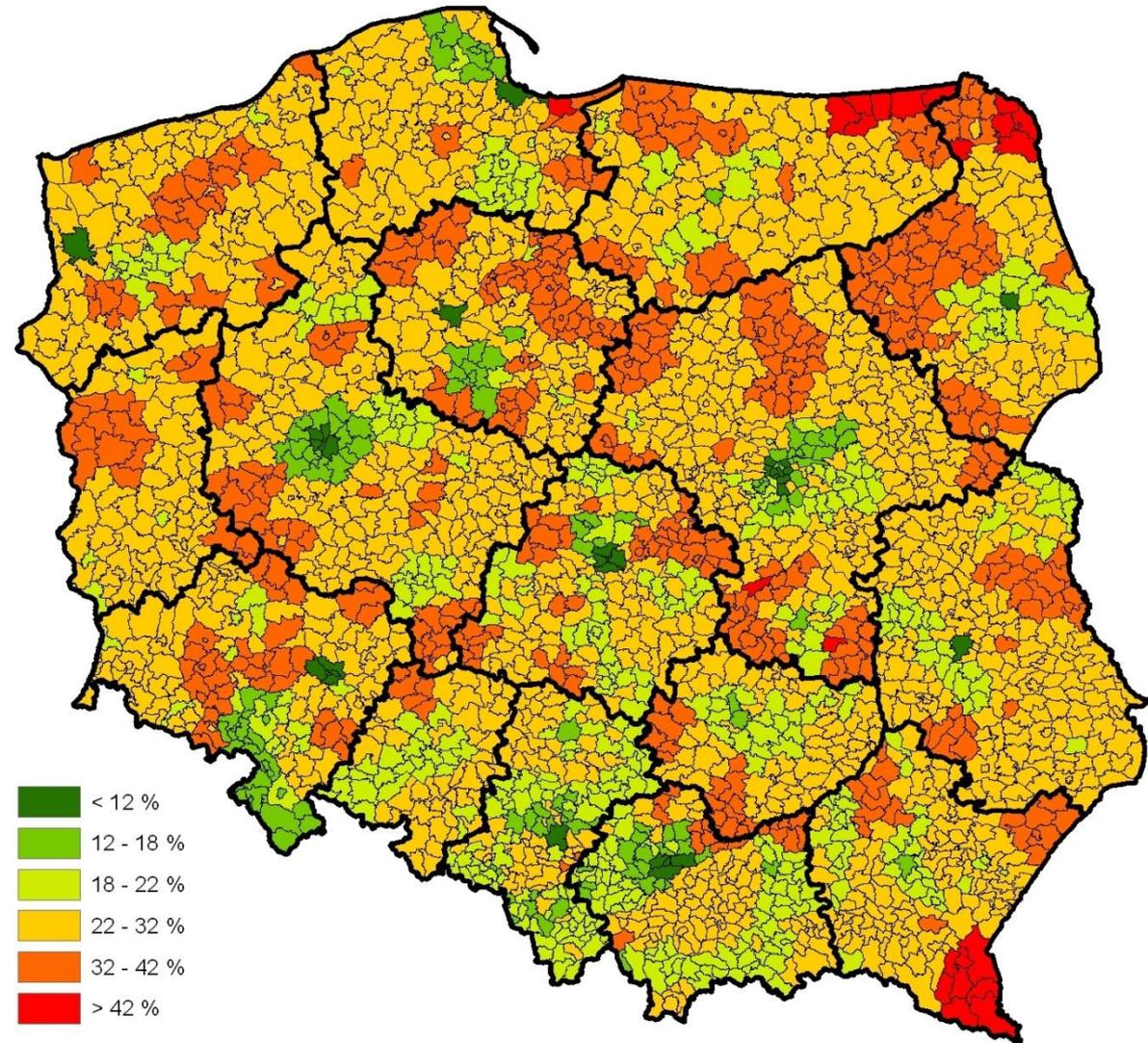
Sampling frames

Representatives surveys are conducted on the sampled samples on the basis of a sampling frame:

- Social;
- Agricultural.

The sampling frame

for social surveys (SFSS)

and its use in official statistics

SFSS

- Sampling frame for social surveys (SFFS) includes information on persons and dwellings

- Since 2015 the SFFS has constituted a basis for the creation of sampling frames for the social surveys carried out by the Central Statistical Office, as well as aided the conduction of statistical analyses.

SFSS

- The first supply of data to the Sampling frame for social surveys included the address and dwelling sampling frame created for the purposes of the 2011 National Census on the basis of the TERYT register and the PESEL population register, which included four types of structures connected to one another with unique IDs:

  - buildings,

  - dwellings,

  - persons and

  - collective accommodation establishments.

SFSS – the scope of information

- Pursuant to the Act on official statistics, the present structure of sampling frame include data on persons in respect of their:

  - name,
  - date and place of birth,
  - sex,
  - PESEL No. (PIN),
  - citizenship,
  - marital status,
  - date of marriage, date of termination of marriage,
  - education, profession, type of workplace or place of education,
  - tax identification number, title of insurance, excluding this part of the code which is subject to confidentiality arrangements,
  - **address** of residence or address of stay and/or correspondence address **including the x,y coordinates of address points,**
  - e-mail address and telephone number.

SFSS – updating

- The fundamental data sources for subjective updates of the sampling frame come from administrative sources.

- Due to the growing needs of statistical services in relation to the Sampling frame for social surveys, the scope of data for sampling frame updates is extended to include data collected from enterprises that are **electricity providers** for natural persons in Poland.

The Sampling frame for agricultural surveys

and its use in official statistics

SFAS – subjective scope

Farm – a isolated unit in technical and economic terms, with separate management (user or administrator) and conducting agricultural activities.

The sampling frame includes the farms of:

- natural persons;
- legal entities and unincorporated organisational units.

SFAS – General principles

- This sampling frame includes identification and address attributes and basic agricultural attributes.

- The fundamental function of the sampling frame is the selection of units for agricultural surveys.

- The objective of the Sampling frame for agricultural surveys (SFAS) is making it possible to generate lists of units, in order to carry out various agricultural surveys.

- The sampling frame includes farms as the sum of recorded parcels used by one legal entity, unincorporated organisational unit or a natural person. At least one record parcel (or part thereof) must be used for agricultural purposes.

SFAS – objective scope

- Farms included in the sampling frame are described using the folloing attribute groups:
  - a) identification,
  - b) Address, with x,y coordinates
  - c) telecommunication,
  - d) agricultural,
  - e) concerning activity, type economic size,
  - f) concerning reporting duties and their fulfilment,
  - g) connected with the following classifications:
    - – legal status,
    - – detailed legal form,
    - – type of ownership,
    - – economic activity,
    - – type of ownership for agricultural surveys,
    - – marking of having participated in the survey of the reason for non-response

SFAS

- The special identification of entities is facilitated by **geographic coordinates** ascribed to the residence of the agricultural holding user and the residence the farm.

- The sampling frame makes it possible to select farms for various agricultural surveys not only due to their specific type or agricultural attributes, but also due to their **location in a specific area – including irrespective of the administrative division.**

- Between statistical surveys, data included in the sampling frame allow to draw a balance sheet of the number and characteristics of farms.

SFAS – updating

To ensure that data on units for agricultural surveys are kept up to date and complete, and, consequently, accurate, the sampling frame is updated in subjective and objective terms using all available sources of information (administrative registers, Integrated Administration and Control System **(IACS)**, agricultural surveys, Big Data – satellite images).

Methods of reducing load on interviewers

Methods of reducing load on interviewers

- Diagnosing and solving problems in surveys
- Survey and variable integration
- Effective popularisation
- Requesting survey authors to reschedule the surveys
- Sampling optimisation
- Sample combination
- CAxI methods
- Use of administrative registers and big data

Methods of reducing load on interviewers – sample optimisation

Sample optimisation is carried out to reduce interviewer load, while maintaining the methodological compliance of the survey and possibly the highest accuracy of the results.

Lower interviewer load can be achieved through:

- reducing the time of interviewers' travel to the sampled addresses by accumulating addresses in a given area, especially if the interviewer is conducting several surveys at the same time
- sampling addresses near the interviewer's place of residence (if it is methodologically consistent)
- eliminating nonexistent or uninhabited addresses from the sampling frame
- eliminating regions/areas that are not easily accessible, have small population density or a significant number of uninhabited addresses, from the sampling frame
- using information on the possible availability of the respondent at working hours at a given address.

Methods of reducing load on interviewers – sample combination

- reducing the number of surveys, and thereby the number of visits paid to respondents,

- reducing load on interviewers,

- reducing the number of non-responses,

- feeding the information base for several surveys with information from a single interview,

- sampling one respondent for several surveys (carefully!).

Methods of reducing load on interviewers – the sequence of CAxI methods

In data collection organisation the sequence rules shall be introduced, in the following order: CAII, CATI, CAPI – while the CATI and CAPI methods may be parallelised .

CAII/CAWI   CATI   CAPI
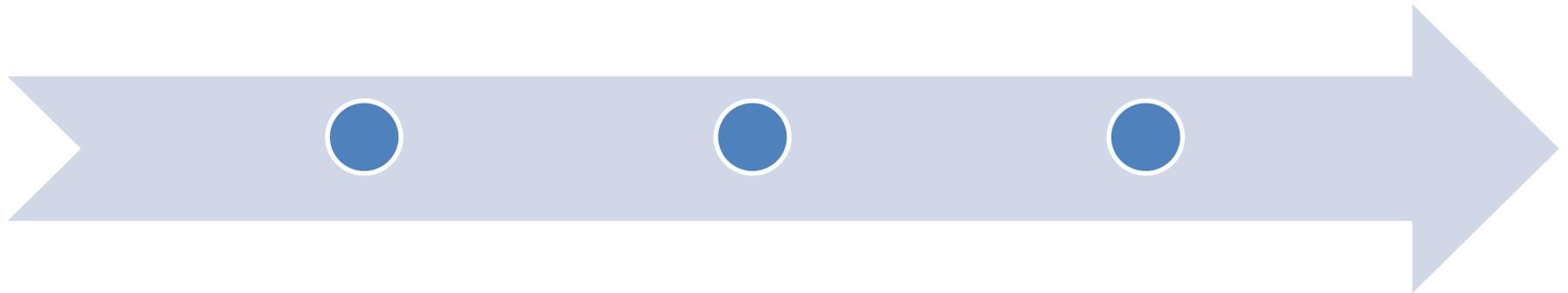
The share of CAII/CAWI in coverage improvement

The CAII/CAWI method consisting of self-enumeration can significantly impact on the better completeness of survey results, due to:

- respondents' comfort – they can complete the questionnaire at convenient time, regardless time of the day
- respondents do not have to receive the interviewers at their homes – interaction with a stranger is eliminated
- respondents do not have to answer questions from the questionnaire over the phone – also interaction with a stranger is eliminated
- respondents have full control over their answers.

The CAII/CAWI methods are very desirable from statistical point of view, as they eliminate interviewer-related costs and reduce interviewer load; however, they require extensive dissemination among respondents and making them accustomed to such a form of participation in surveys.

Organization of work on the implementing of CATI method into the surveys.

Stat Call Center in Poland

2010

2012

Signing the contract - **start building and implementation of Call Center for the purpose of censuses.**
2010-2011 - Censuses realization

The implementation of CATI into surveys

Commencement of proceedings to build the system SCC

Before 2009

STAT CALL CENTER **(SCC)** – own name of the organizational and functional structure implemented by the CSO to carry out statistical surveys by telephone contact with respondents.

The aplications and configuration of system before and during campaigns involve employees of computing center (which supports the CSO in Informatics).

Before you will implement a system ...

The organization of work

Separate room for
Call Center
"CATI Studio"
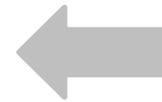
Interwiever is working
on his existing post in
the office

CATI application

The organization of data collection campaign – application

Operating panel of calls :

BIG DATA in Official Statistics

BIG DATA

These are information sets that are characterised by extensive volume, rapid growth and substantial diversity, which require new, non-classical methods and tools of storage, processing and analysis for the purposes of making decisions, discovering new phenomena and optimising processes.

Data

All Global Data in Zettabytes

1ZB = 1,126,000,000,000,000,000,000 bytes (approx)
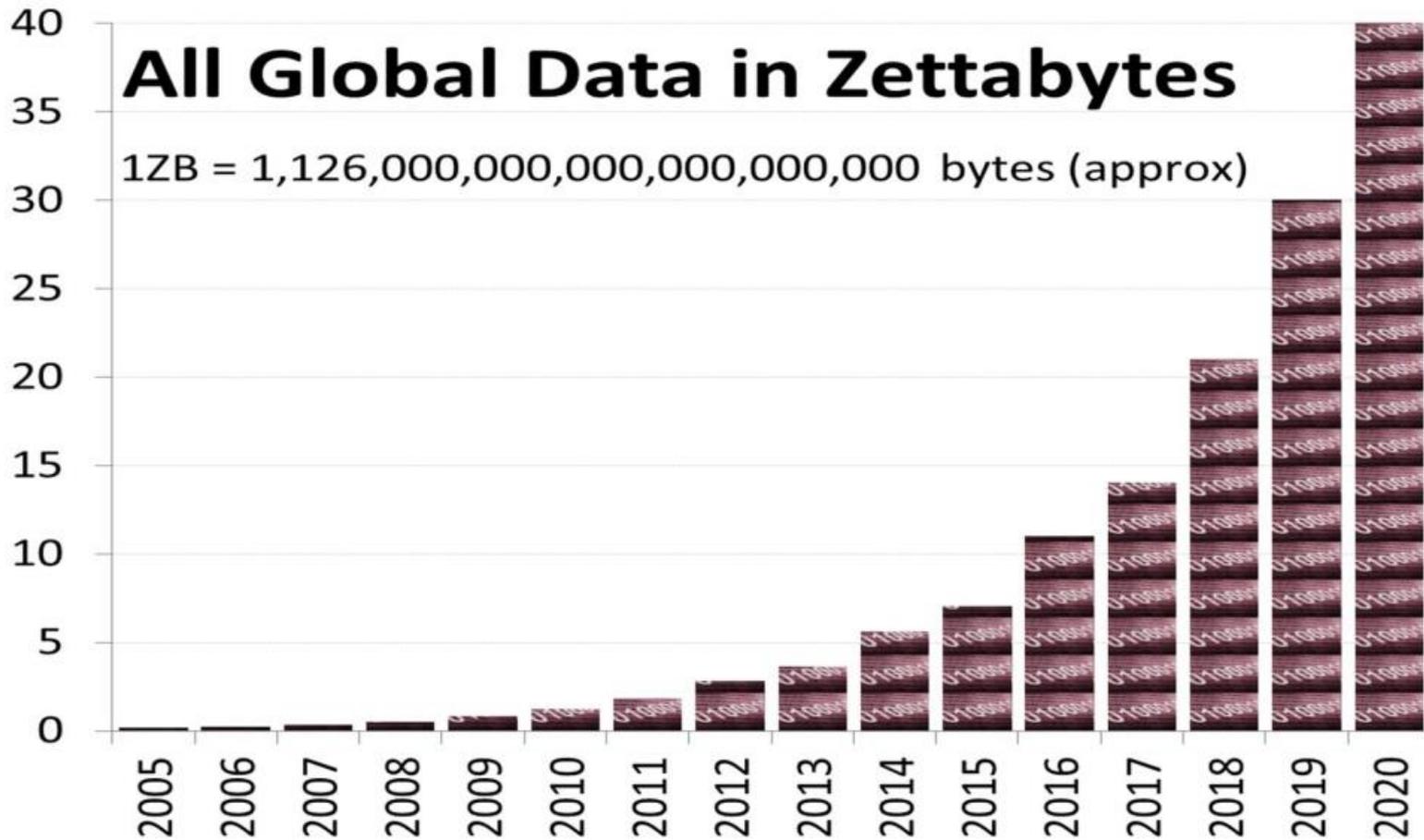
Source: UNECE

Big Data sources - EUROSTAT

Communication
- Mobile Communication
- Social Media

WWW
- Web Searches
- Analysis of websites' contents
  - Job Advertisements
  - Businesses' Websites
  - E-Commerce
  - Real estate
- Internet Traffic

Sensors
- Traffic loops
- Smart meters
- Automatic Vessel Identification System
- Satellite Images
- Webcams

Process generated data
- Reservation Systems
  - Flight Booking transactions
  - Trains
  - Hotels
- Supermarket Cashier Data
- Loyalty Cards
- Financial transactions
- Mobile Payments
- eGovernment

Crowd sourcing
- Voluntary Geographic Information websites (OpenStreetMap)
- Voluntary Information websites (Wikipedia)
- Community pictures collection

Data on mobile phone location

This is New York City's population by day ...

Traffic rank: 16th worst (46 hr./yr.)

Average daily commute: 34 min.

Hudson River

East River

Key to city maps
Population

Low   Medium   High
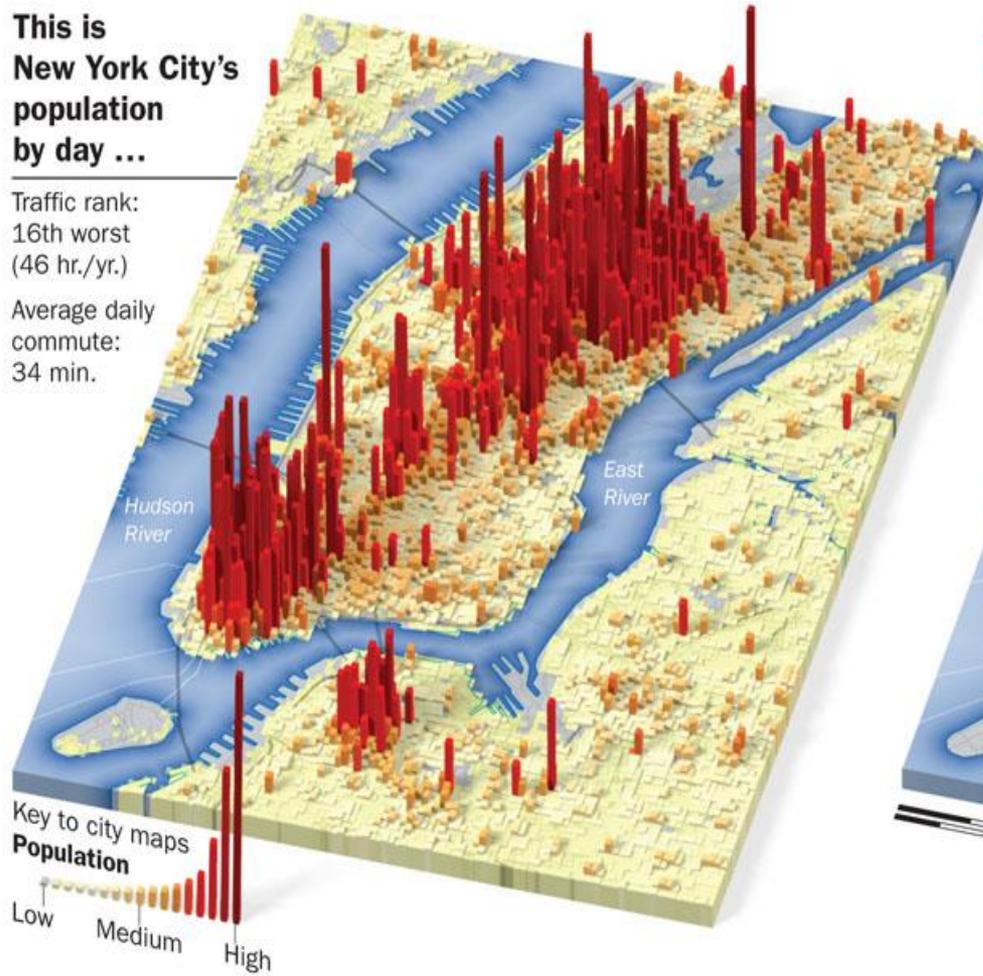
... and the population at night

Hudson River

East River

1.5 km   1.5 miles

DAYTIME

NIGHTTIME

Potomac River

Potomac River
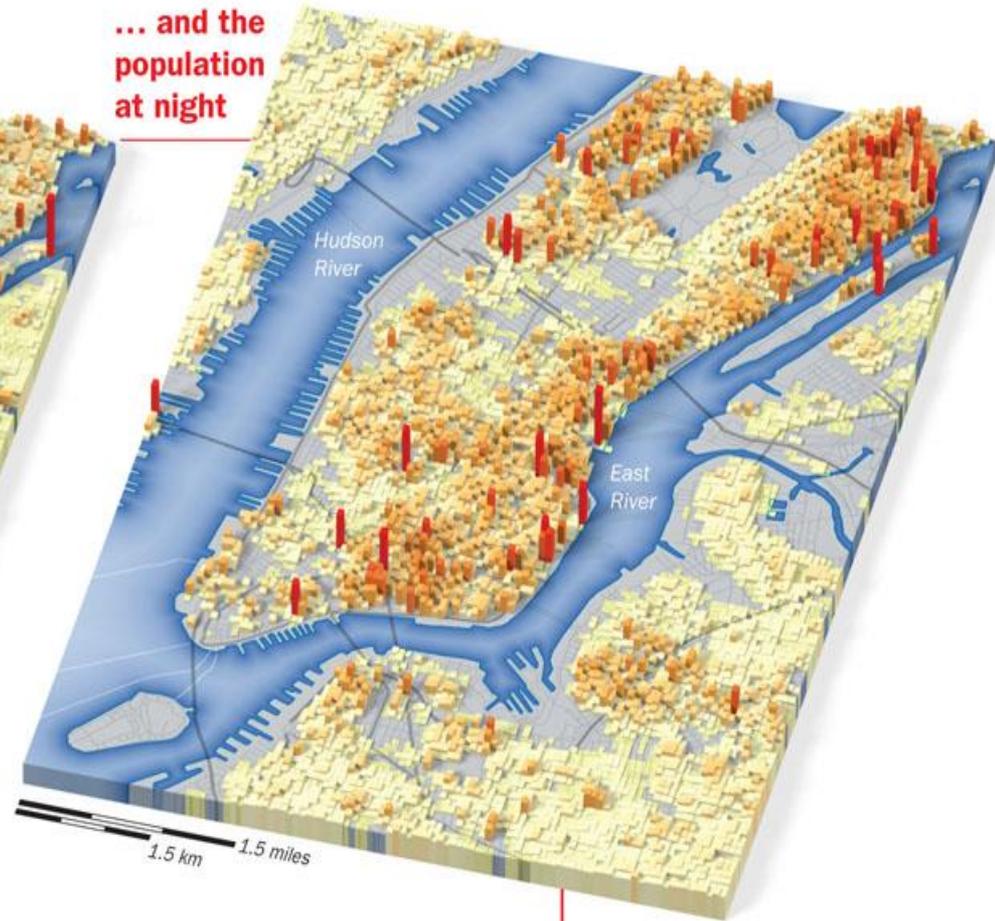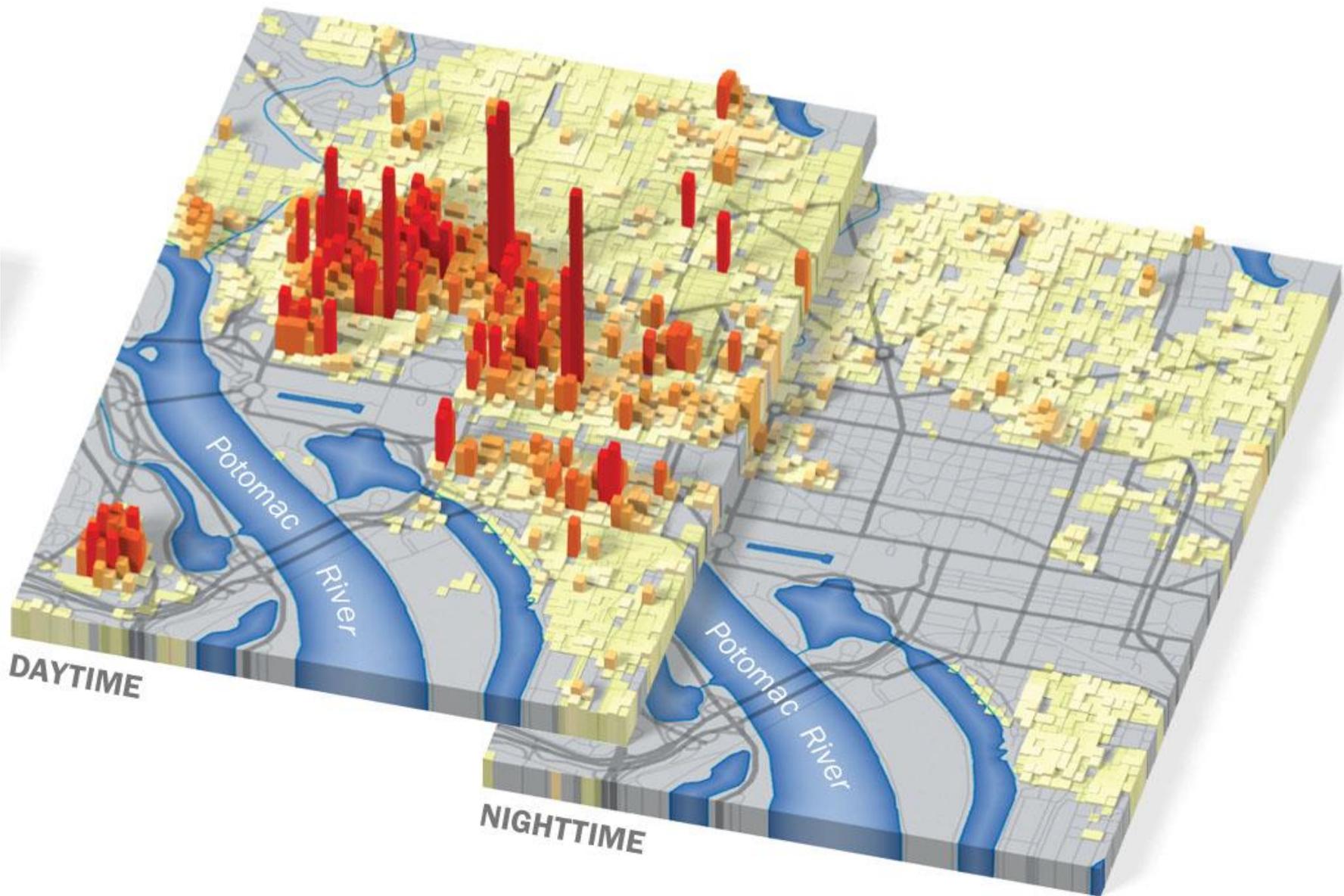
# Estonia – tourist traffic to Egypt based on mobile phone roaming
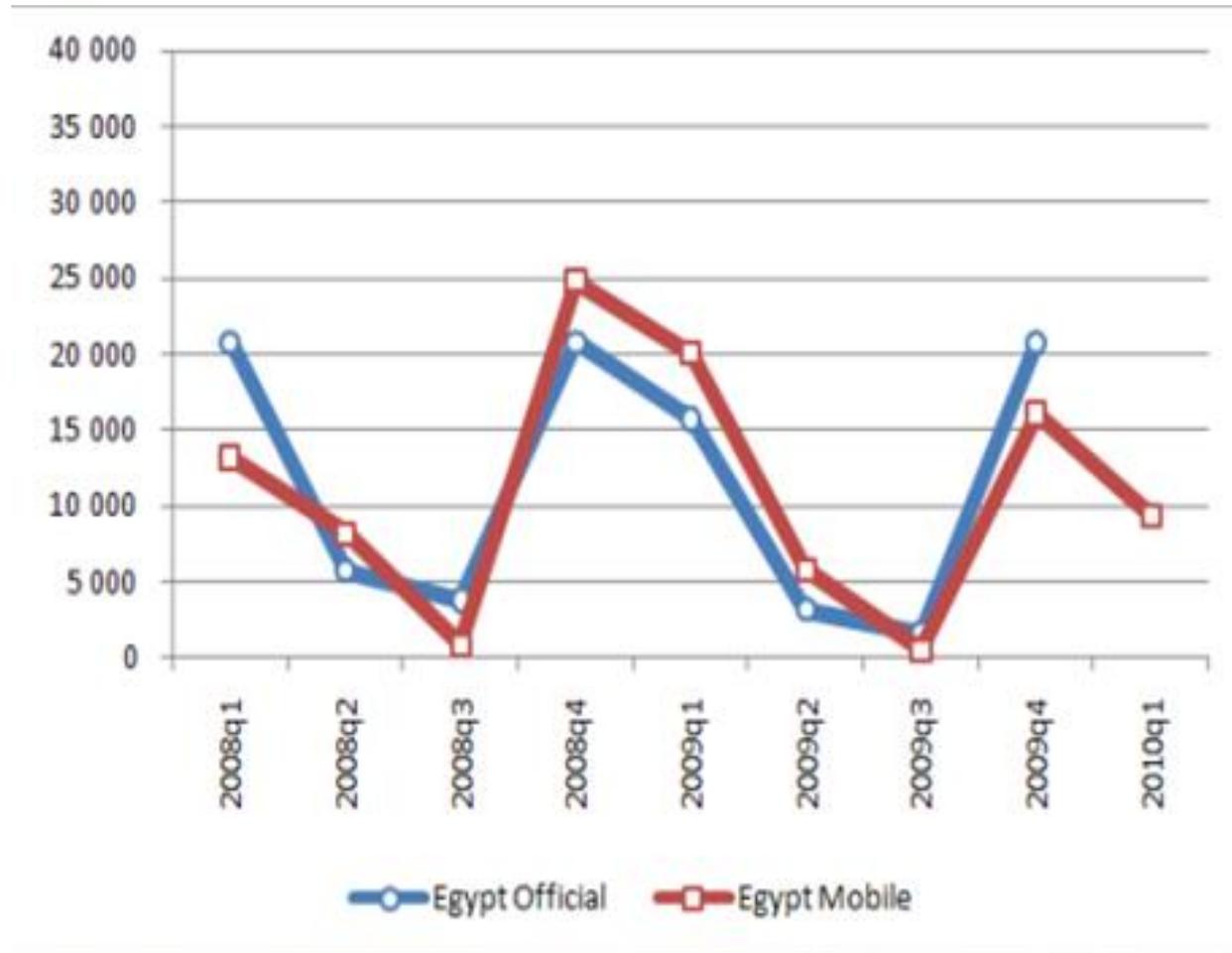
Traffic intensity sensors
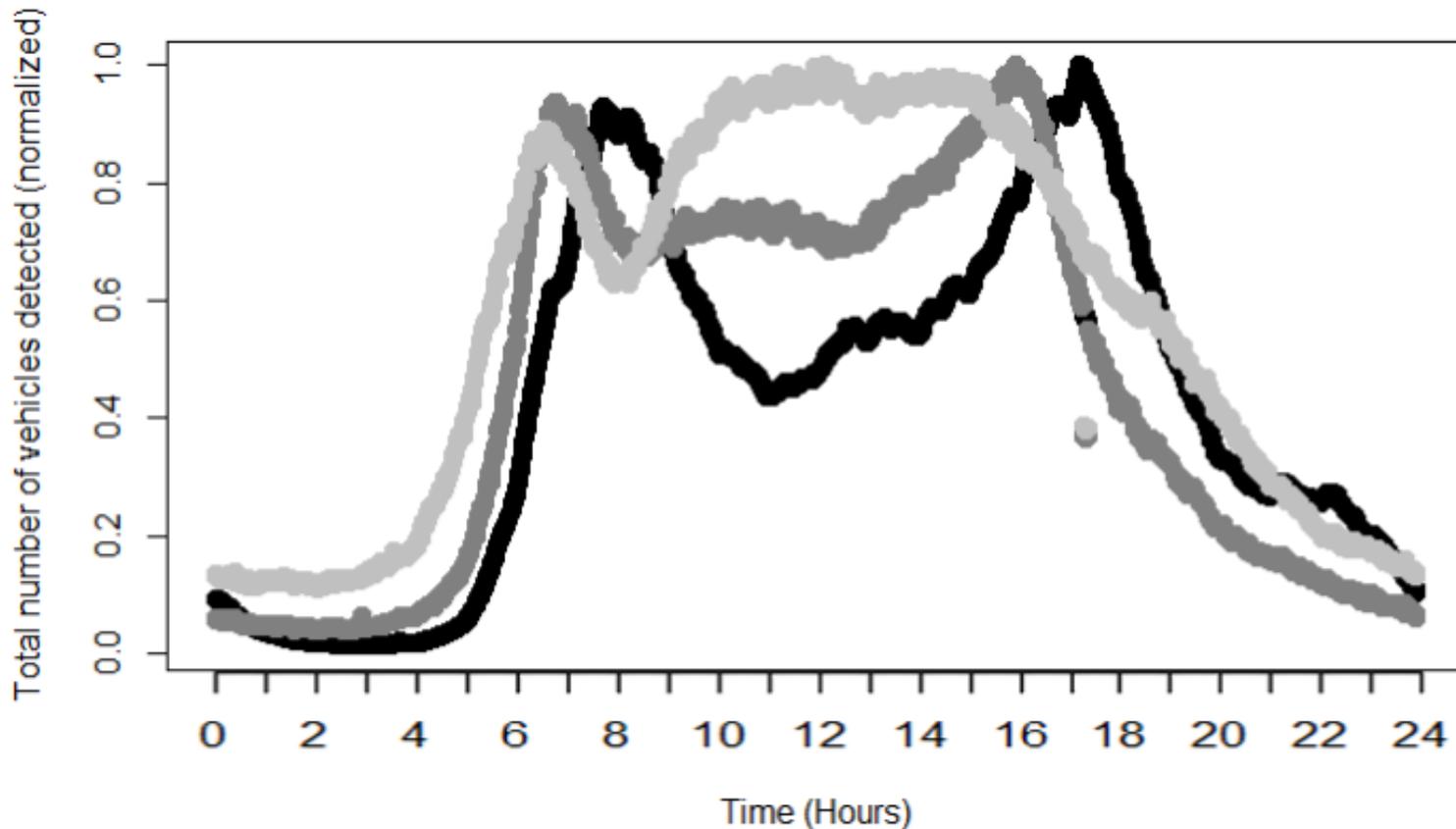
Normalized number of vehicles detected in three length categories on December 1st, 2011 after correcting for missing data. Small (<= 5.6 meter), medium-sized (>5.6 and <= 12.2 meter) and large vehicles (> 12.2 meter) are shown in black, dark grey and grey, respectively. Profiles are normalized to clearly reveal the differences in driving behaviour.

Social Media

A consumer sentiment and attitude survey in the Netherlands based on social media communications.
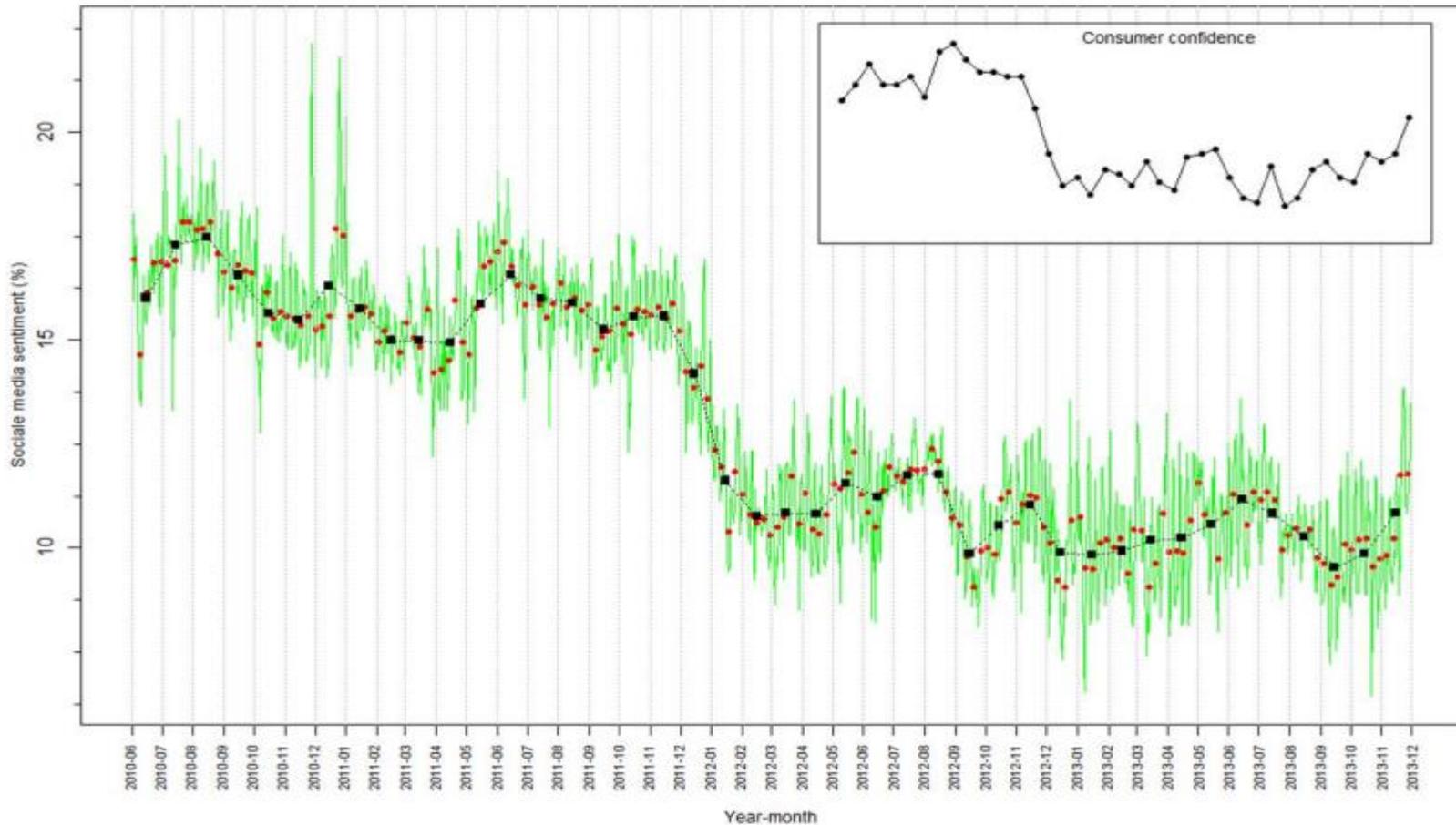
Challenges

|		
Law	Data security	Privacy
Ethics	Competences	Methods
Technologies	Quality	Access to data

Field of interest of Polish official statistics
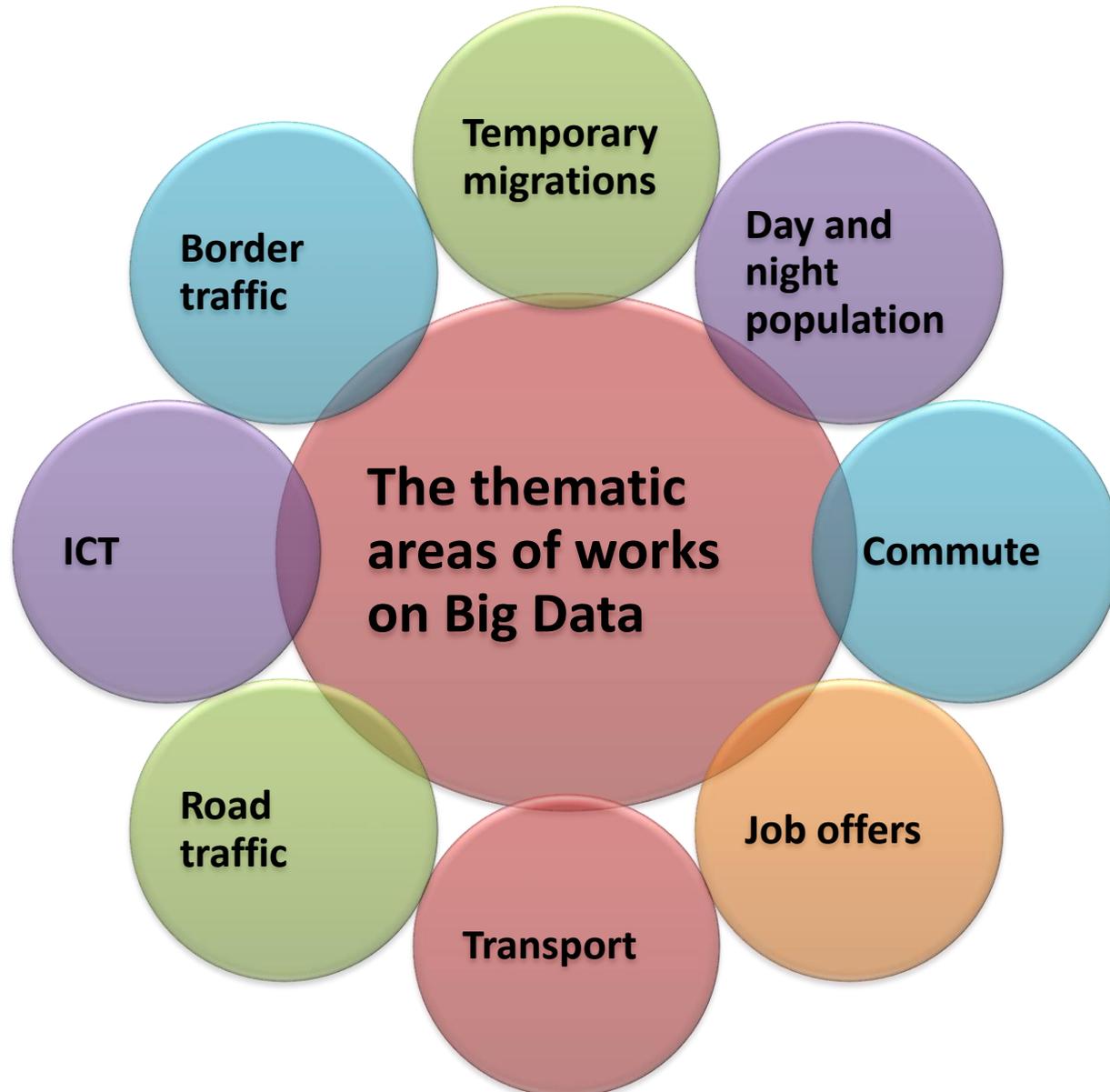
„Better data, better lives"

„Big data needs official statistics as much as official statistics need big data"

The second International Conference on Big Data for Official Statistics
Abu Dhabi from 20 to 22 October 2015