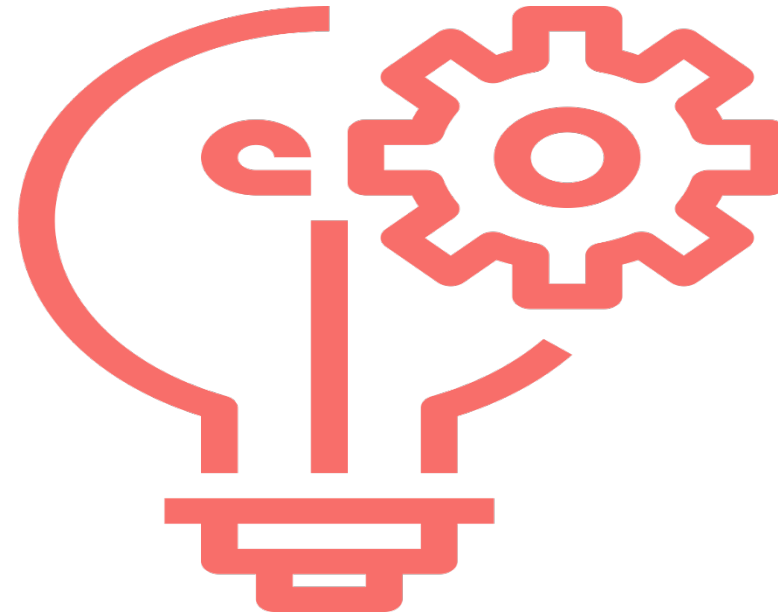
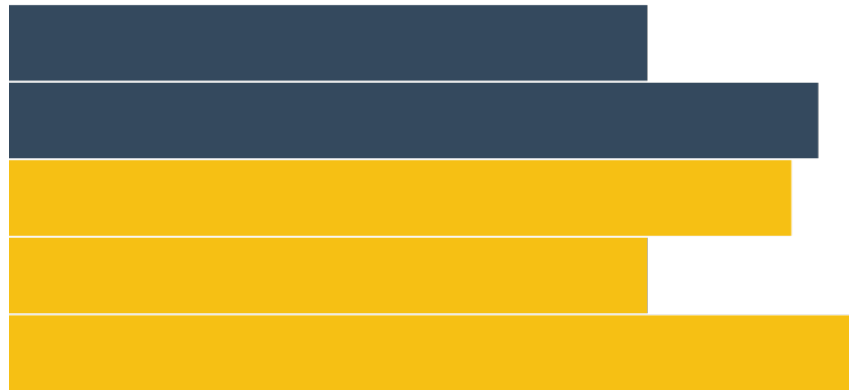


## GDPR regulation: key issues in producing population statistics



Lavinia Elena Balteanu  
Director - Population and external migration unit  
National Institute of Statistics

- Population statistics: from old to new
- New challenges, new solutions
- Conclusions

## Legal base in computing population statistics

- Inter-census population statistics:
  - no 862/2007 on statistics on migration
  - no 1260/2013 and no 205/2014 on European demographic statistics
- Census population and housing statistics:
  - no 763/2008 on population and housing censuses
- Requesting to measure resident population and international migration according to concept of usual residence and presence/absence for a length of at least 12 months on the MS territory

- Main problems/difficulties
  - Immigrants + emigrants for Romanian citizens = data available only for definitive migration
  - Immigration for other citizenship = data available mainly for those outside EU
  - Lack of data sources for measuring real size of migration, especially of emigration
  - Lack of data sources for measuring the usual resident population
- **Solution:** estimation based on combination of methods and data sources:
  - Econometric model based on small-area estimation techniques (SAE), which allows to estimate the structure of migrants at national, regional and county level
  - Annual flows of migrants for the 2002+ provided by ISTAT-Italy and INE-Spain based on Trilateral Agreement of Romania, Italy and Spain, facilitated by Eurostat
  - Mirror international migration (Eurostat database and data from other National Statistical Institutes)

- Strengths:

- Use of combination of data sources allowed considerable improvement of estimates on E and I (confirmed by statistics published annually by the MS)
- Very detailed data series on migrant flows from ISTAT and INE ensures good estimation in terms of size and structure by requested variables of the annual flows for international migration

- Weaknesses

- Circular migration
  - Returnees (return of migrants back in Romania)
  - Romanian emigrants leaving from Italy and Spain and going to another country (not Romania)
- Estimates at low (NUTS5) territorial level are not available

## Moving step by step to a system based on administrative data sources

### Why?

- The new approach at EU level of moving more and more to registered data underlined by
  - The new guidelines (under discussion) on how to move to a concept of population mainly based on registered data (record data at individual level)
  - The provisions of **REGULATION (EC) No 223/2009 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 March 2009 on European statistics** amended by Regulation (EU) 2015/759 of the European Parliament and of the Council of 29 April 2015 stating:  
*“In order to reduce the burden on respondents, the NSIs, other national authorities ...and the Commission (Eurostat) shall have the right to access and use, promptly and free of charge, all administrative records and to integrate those administrative records with statistics” (Article 17 a)*
- Lack of data in producing population statistics according to the EU requirements
- Moving from an old fashioned census to a new one based on administrative data sources

## How?

- Computing the Romanian population using administrative sources. For this purpose an algorithm (methodology) was already developed which investigates, in cascade, at **individual record level**, several administrative sources
- Designing of database population based on linking primary data coming from different administrative data bases (STATPOP)
- Signing agreements with institutions managing administrative sources for building and feeding STATPOP

## Using registered data in the context of GDPR provisions

- *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data and on the free movement of such data and repealing Directive 95 / 46 / EC (GDPR)*
- According to *GDPR* and to the *Handbook on European data protection law 2018*:
  - Data are personal data if they relate to an identified or identifiable person, the 'data subject'
  - Data are anonymised if they no longer relate to an identified or identifiable individual
  - The principles and rules of data protection do not apply to anonymised information



## Anonymisation algorithm

- Taking into account the GDPR provisions, an anonymization algorithm was developed in order to be applied at data administrative owner level before transferring data to INS
- Consisting in encryption of the personal identification number (PIN) of each individual by following certain steps
  - Extracting, in a distinct variable, the first 7 characters in the PIN, those corresponding to gender and date of birth
  - Creating an identifier (called unique statistical identifier - ISU) made by applying a HASH function to the PIN
  - Using one of the following: Oracle, function DBMS\_CRYPTO.Hash(), implemented implicitly *or* SQLServer, function hashbytes(), implemented implicitly *or* Other database management systems, function digest(), available for free by using the package `digest` from R ([www.r-project.org](http://www.r-project.org)).
  - ISU validation: 40 characters
  - Testing for at least 5 cases before sending data

- The resulting identifier is **irreversible** in the sense that no other procedure can be applied to recompose the PIN from which it originates, i.e. the data transferred to the INS are anonymous
- In this situation, as provided for in *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data and on the free movement of such data and repealing Directive 95 / 46 / EC - Recital 26:*

*“The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. **This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes**”.*

- In the context of the new European approach of computing population statistics mainly based on registered data, a good inter-institutional cooperation is essential
- It is obvious that even if the main responsible in producing statistics is the statistical office in most of the cases, it is actually a combined effort of the statistical office and the administrative data owner



CONFERENCE  
BUCHAREST, 8-9 APRIL 2019



NATIONAL  
INSTITUTE OF  
STATISTICS

*THANK YOU  
FOR YOUR ATTENTION!*