



Centrālā statistikas  
pārvalde

# Registers – based Census in Latvia - transition of methods, data publishing, future challenges

Sigita Purona-Sida  
Central Statistical Bureau of Latvia  
Deputy Director of Social Statistics Department



Centrālā statistikas  
pārvalde

# Census 2021 in Latvia

## Decision made in 2012:

- Method for determining the usually resident population, methods for evaluation of census variables
- Collaboration with administrative registers
- Financial aspects
- Census 2021 will be registers – based + same reference date as for annual population, same data publication date

Census reference date – **1 January 2021**

**!!** In 2020 face-to-face interviews were suspended (Covid-19 pandemic); the decision was far-sighted and correct





Centrālā statistikas  
pārvalde



2000

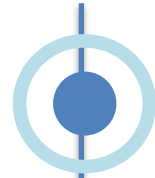




Centrālā statistikas  
pārvalde

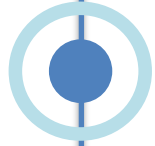


## Population estimation method



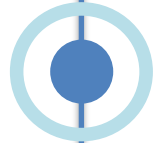
2011

The difference between the Population Register and Census data - 155 thsd or about 7 % of the resident Latvian population registered in the Population Register



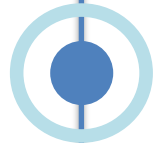
2012

The Logistic regression model has been developed to evaluate the residence status for each individual



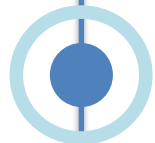
2015

Population Microcensus – to determine how precisely official statistics reflects number of population in Latvia, in cities under state jurisdiction and counties



2017 – 2018

External Migration Survey aiming to acquire additional information on migration flows and specify population estimate



2020 – .....

Development of a new population estimation method. New methodology will allow more accurate annual demography data production... according with the upcoming ESOP regulation?





## Fractional Counting

---

Pioneering experiences in Baltic countries

- Tiit & Maasing (2016)

Sign-of-Life (SoL) score  $X(k, t - 1)$  from admin data  
Residency index for person  $k \in \mathbb{P}$  (EPD) in year  $t$

$$R(k, t) = d \cdot R(k, t - 1) + g \cdot X(k, t - 1)$$

$d$  = stability rate and  $g$  = SoL rate, set heuristically...

- LCSB (2019)

Let  $z = 1$  or  $0$  indicate in population or not

Imputed census set  $\Rightarrow \{z_i : i \in \mathbb{P}\}$  for year 2011

$$q := \hat{E}(z) = \frac{\exp(\mathbf{b}^\top \mathbf{x})}{1 + \exp(\mathbf{b}^\top \mathbf{x})}$$

Logit parameter  $\mathbf{b}$  fixed since 2011

## Fractional Counting

---

More recently, SoL-logit model

- Let  $\mathbf{y}$  be a vector of binary SoL indicators of residency  
Let  $p(\mathbf{y}|z; \boldsymbol{\theta})$  be *measurement model* of  $\mathbf{y}$  given true but unobserved residency status  $z$

Simplest model:  $\mathbf{y}$  conditionally independent given  $z$

- For each EPD-person, the joint model of  $(\mathbf{y}, z)$  is

$$p(\mathbf{y}, z | \mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\beta}) = p(\mathbf{y} | z; \boldsymbol{\theta})p(z | \mathbf{x}; \boldsymbol{\beta})$$

where  $p(z | \mathbf{x}; \boldsymbol{\beta})$  is the logit model of  $z$

Model estimation by EM algorithm, yielding

$$p(z | \mathbf{y}, \mathbf{x}; \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}) = E(z | \mathbf{y}, \mathbf{x}; \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}})$$

NB. Stratified modelling necessary in reality

Still not estimates of migration *and* stock directly



Centrālā statistikas  
pārvalde



## Education

- International classification of education ISCED - 2011 was used
- Cumulative database was established using results of the 2011 Census, administrative registers and CSB surveys
- The correspondence of the acquired level of education to the person's age and to certain professions is being evaluated
- Imputations have been made for 5.8% of population aged 15 and over

## Economic activities

- Different reference period – November of 2020
- Determination of economic activity status
- Criteria for identification of occupation, industry, status in employment and location of the place of work
- **New product for national data users – monitoring of graduates of Latvian higher education institutions and vocational education institutions**



Centrālā statistikas  
pārvalde



## Housing

- Main sources for housing data were State Address Register and Real Estate Cadastre Information System of the State Land Service and 2011 Census
- Information about ruined, burned or demolished residential buildings and housing facilities was collected from the local governments
- Algorithms and a list of conditions was developed to supplement the missing data for several housing indicators
- A list of institutional housing has been established since 2012 – [used also for population statistics and surveys](#)



Centrālā statistikas  
pārvalde



# Dissemination of results: channels of dissemination, reach, interest from stakeholders

Academics

*Data hackathon, Academic  
breakfast*

Administrative registers, data users,  
institutions

*Census Conference*

*«Data as Value for Society in the  
Digital Age»*

International partners

Mass media, data users

*press releases (history of Census,  
announcement of events,  
change of methodology, data  
publishing, events)  
cartoons  
infographics  
video message*

*Facebook, Twitter*





Centrālā statistikas  
pārvalde



## Data publishing

28.05.2021.

Population  
Migration

30.07.2021.

Households and  
Families

24.11.2021.

Housing

15.06.2022.

Economic indicators  
Education

## Events

01.01.2021.

The critical moment of the  
census

12.05.2021.

Hackathon  
*Alternative Population and  
Housing Census*

21.05.2021.

Satellite conference  
*Academic Breakfast*

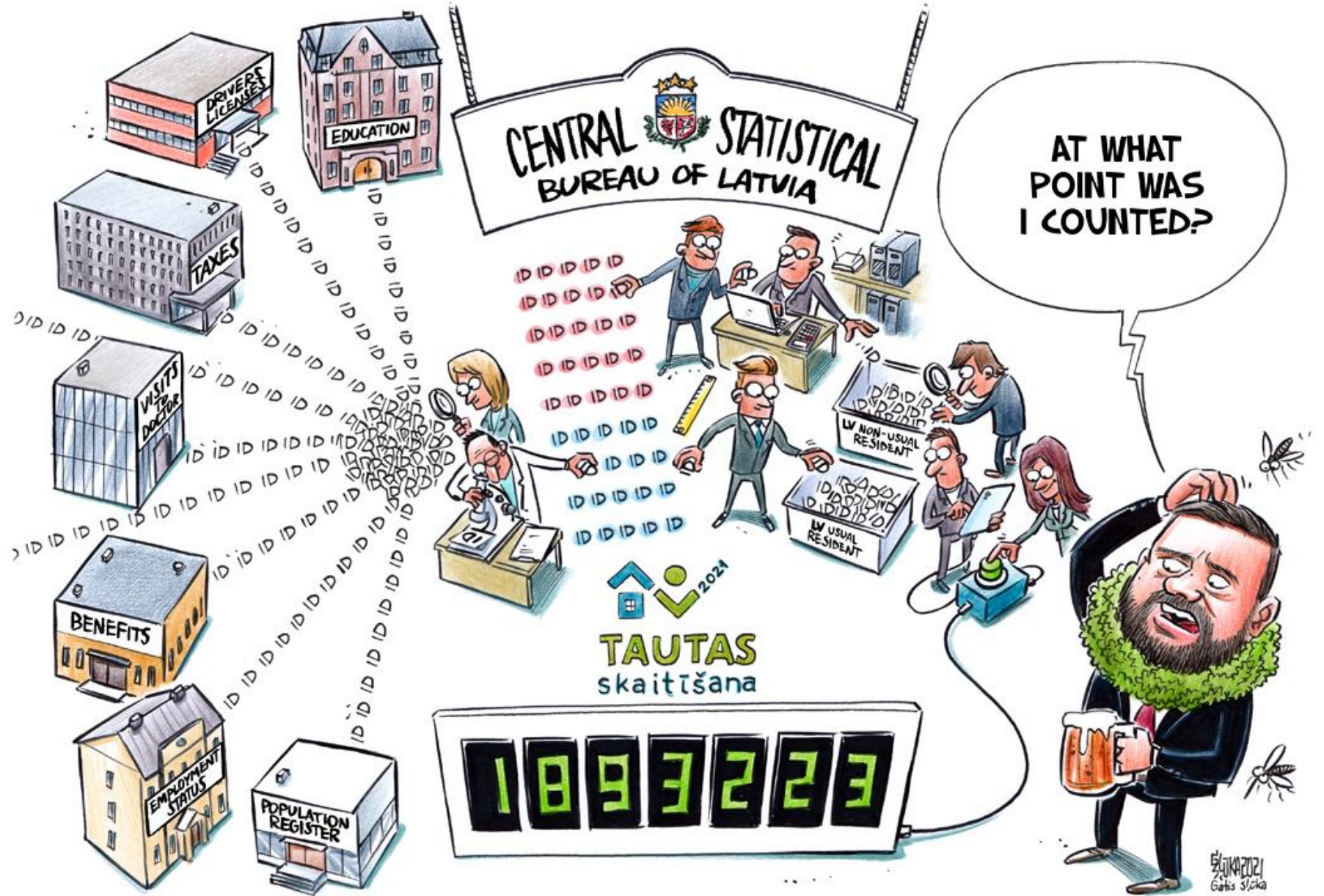
28.05.2021.

Conference  
*Data as value for society in  
the digital age*



Centrālā statistikas  
pārvalde

  
**TAUTAS**  
skaitīšana







Centrālā statistikas  
pārvalde

  
**TAUTAS**  
skaitīšana





Centrālā statistikas  
pārvalde



# DATA HACKATON

<https://www.csp.gov.lv/lv/media/2144/download>

- 11 teams - both Latvian university students and already experienced data enthusiasts
- Lectures by professors from both the University of Latvia and the University of Southampton were available for hackathon participants for three days
- In turn, the members of the teams worked on specific challenges of the register-based census

How to determine whether a person's declared place of residence is also his or her actual place of residence?

What factors contribute to internal migration?

What social and demographic factors influence a person's desire to get an education?

What is the situation with the availability of comfortable housing in Latvian cities and rural areas?

What is the volume of actually rented housing in Latvia?

<https://skaties.lv/zinas/latvija/sabiedriba/nosledzies-latvija-pirmais-alternativas-tautas-skaitisanas-hakatons/>

<https://www.ba.lv/augstskola/zina/banku-augstskolas-studenti-hakatonu-iegust-centralas-statistikas-parvaldes-specialbalvu/>



Centrālā statistikas  
pārvalde



# Dissemination of results: channels of dissemination, reach, interest from stakeholders

1. Demographics and family

2. Ethnic indicators and language

3. Housing indicators and territorial aspects

*How can statistics help with data, given the situation in 2021 and new solutions for data collection (less data collection in face-to-face interviews; register data and big data; identification and use of new data sources, detailed territorial breakdown), analysis of demographic processes?*

*Data users' needs for housing data. Key needs for territorial research.*

*Data sources should be sought in the areas where data mining should be developed*

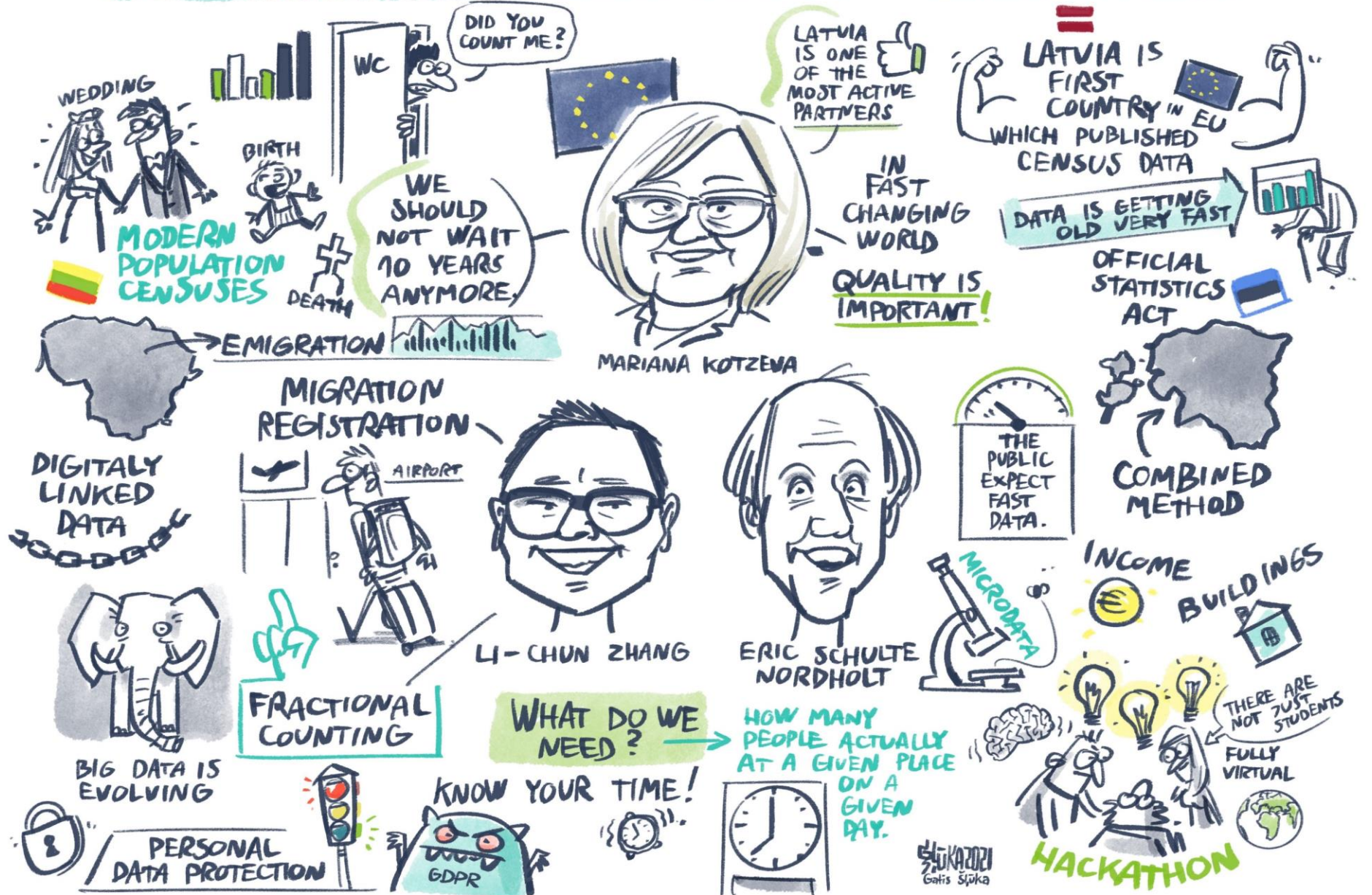
*Use of census results in research - public administration and local governments*





Centrālā statistikas pārvalde

# DATA AS VALUE FOR SOCIETY IN THE DIGITAL AGE



GAUKAZI  
Gatis Šķuka



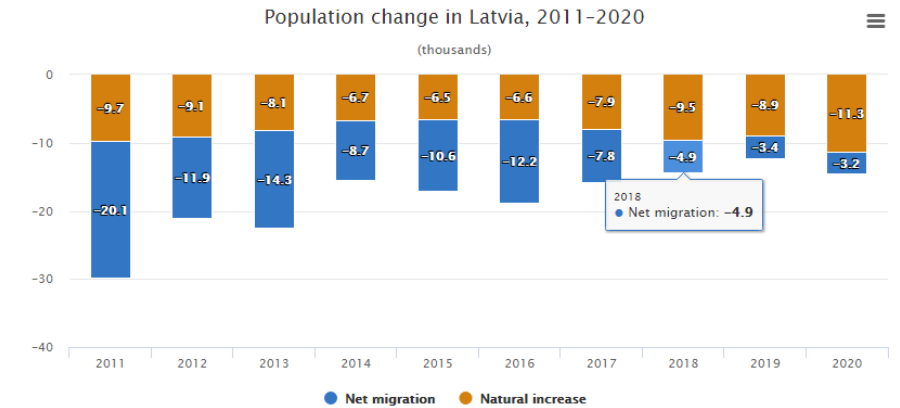
Centrālā statistikas  
pārvalde



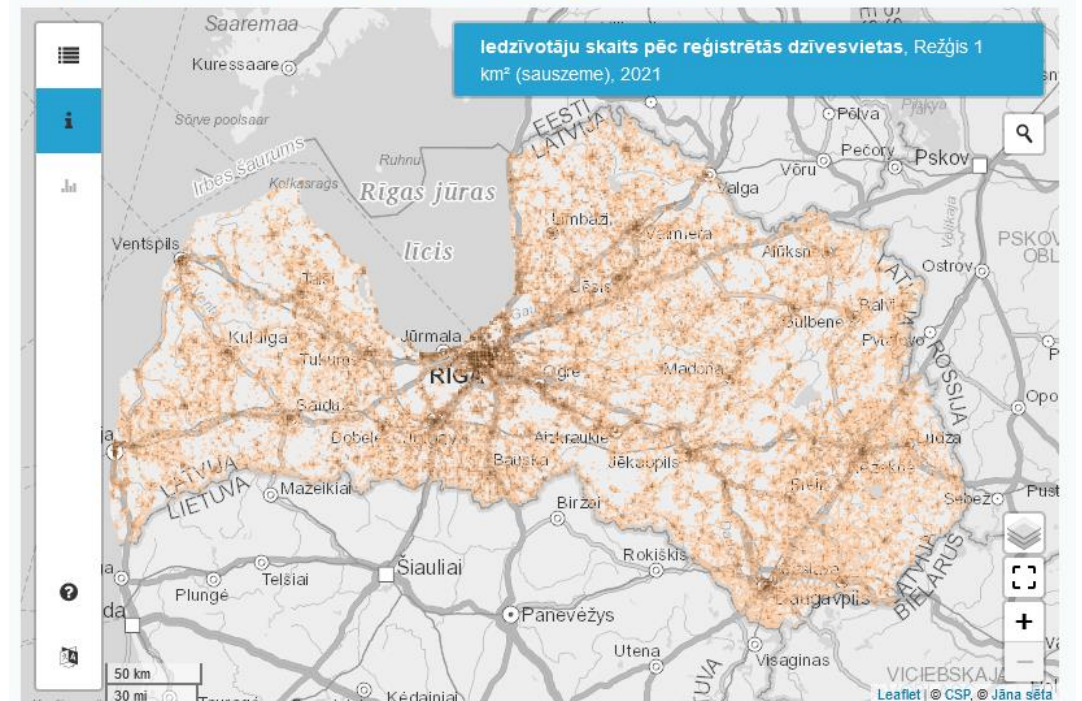
# Dissemination of Census results

- Official Statistics Portal:
  - interactive graphs
  - interactive maps
- EU grid tables only as data table in open data source
  - Do not show additional information if there are less than 10 persons in the grid
  - In the Latvian grid publishing information only as a percentage

<https://geo.stat.gov.lv/stage2/#lang=lv>



ledzīvotāju skaits 1 km<sup>2</sup> režģa šūnās 2021. gada sākumā







Centrālā statistikas  
pārvalde



## Advantages of registers-based census

- Data are available faster
- More accurate, more relevant data according to the classification
- Reduced face-to-face interviewing
- Reduced respondents' burden
- Smoother discussions on the budget



Statistics Latvia  @CSB\_Latvia · May 28, 2021

Latvia is the first EU member state having produced validated Census 2021 data, pointed Director General of @EU\_Eurostat @MarianaKotzeva

[csp.gov.lv/lv/2021-gada-t...](https://csp.gov.lv/lv/2021-gada-t...)

#TS2021

“

*Latvia is the first EU  
member state having  
produced validated  
Census 2021 data*





Centrālā statistikas  
pārvalde



## Challenges of register-based census

- Further development of national administrative registers
- Strengthening existing and establishing new partnerships





Centrālā statistikas  
pārvalde

