



EUROPEAN CONFERENCE ON  
QUALITY IN OFFICIAL STATISTICS  
2024 ESTORIL - PORTUGAL

# The Future of Sampling Frames

Ana M. Santos, António Portugal, Carlos Marcelo, Jorge Magalhães,  
Paula Cruz, Pedro Campos, Porfírio Leitão, Sílvia Mina

Statistics Portugal • DMSI • Department of Methodology and Information Systems



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

eurostat 

The conference is partly  
financed by the European Union



EUROPEAN CONFERENCE ON  
QUALITY IN OFFICIAL STATISTICS  
2024 ESTORIL - PORTUGAL

# Contents

1. Introduction and Goals
2. Main sources of administrative data
3. Household Dynamics
4. Discussion

 INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

eurostat 

The conference is partly  
financed by the European Union



# 1. Introduction and Goals



- Need for collecting information on **individuals, families**, and households in a timely, speedy, and high-quality manner.



**Sampling frames must be highly updated, thus avoiding low error rates.**

- Errors can arise, for example, from empty dwellings, or outdated contact information.
- integration of administrative information in updating sampling frames





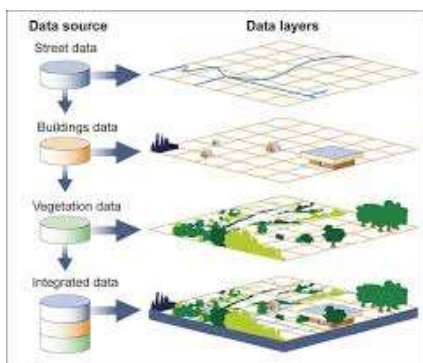
# 1. Introduction and Goals



This work reflects on how administrative sources can continuously improve information on dwellings and individuals.



Updating contact information (telephone, email, address), as well as sociodemographic information that can help in the sample selection of families and/or individuals for various statistical surveys.



Opportunities provided by geographic information systems (GIS) and their ability to help define patches of territory that visualize the dispersion of dwellings and assist in sample selection.



# 1. Introduction and Goals

- **Panel Surveys (BAInqRot)**

the creation of the sampling frame benefited from the fact that it was visually possible to represent the information for more specific strata





# 1. Introduction and Goals

What information should we use to update the sampling frames in household surveys?

What is the impact of updating the sampling frames, particularly when calculating selection probabilities?

How can we monitor the dynamics of the territory to gauge the best time and frequency for updating the sampling frames?



EUROPEAN CONFERENCE ON  
QUALITY IN OFFICIAL STATISTICS  
2024 ESTORIL - PORTUGAL



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

eurostat 

The conference is partly  
financed by the European Union

# 1. Introduction and Goals

What metrics should be used to measure the inflow and outflow of dwellings/individuals and the respective socio-demographic changes?





# 1. What the others say?

## The World Bank

### Improve sampling efficiency and coverage

- Continuous improvements to sampling frames
- adoption of innovative sampling techniques are key.
- geocoding census records
- leveraging high-resolution satellite data when the census frame is outdated,
- building an integrated sampling frame to improve coverage.

### For populations that are difficult to reach.

- oversampling
- network sampling
- while machine learning models to assist with more targeted sampling.
- responsive and adaptive sampling designs can save costs through real-time design decisions during data collection

<https://blogs.worldbank.org/en/opendata/8-technical-priorities-position-household-surveys-next-decade>





# 1. What the others say?

## Census Bureau



Sullivan, J., Genadek, K., (2024)

- Sampling frame for the **American Community Survey (ACS)**:
- **Master Address File (MAF)**, which is the Census Bureau's inventory of known housing units, group quarters (GQ), transitory locations, and selected nonresidential units in the United States.
- In addition to the MAF, the Census Bureau links individuals to domiciles in the **MAF-Auxiliary Reference File (MAF-ARF)** on a yearly basis since 2000.
- MAF updates have been regular from the **U.S. Postal Service Delivery Sequence File**, which is the Postal Service's inventory of mail delivery points
- These data are available for research purposes in a restricted-use environment.

<https://nap.nationalacademies.org/read/13387/chapter/6>

~



# 1. What the others say?

## United Nations



Sullivan, J., Genadek, K., (2024)

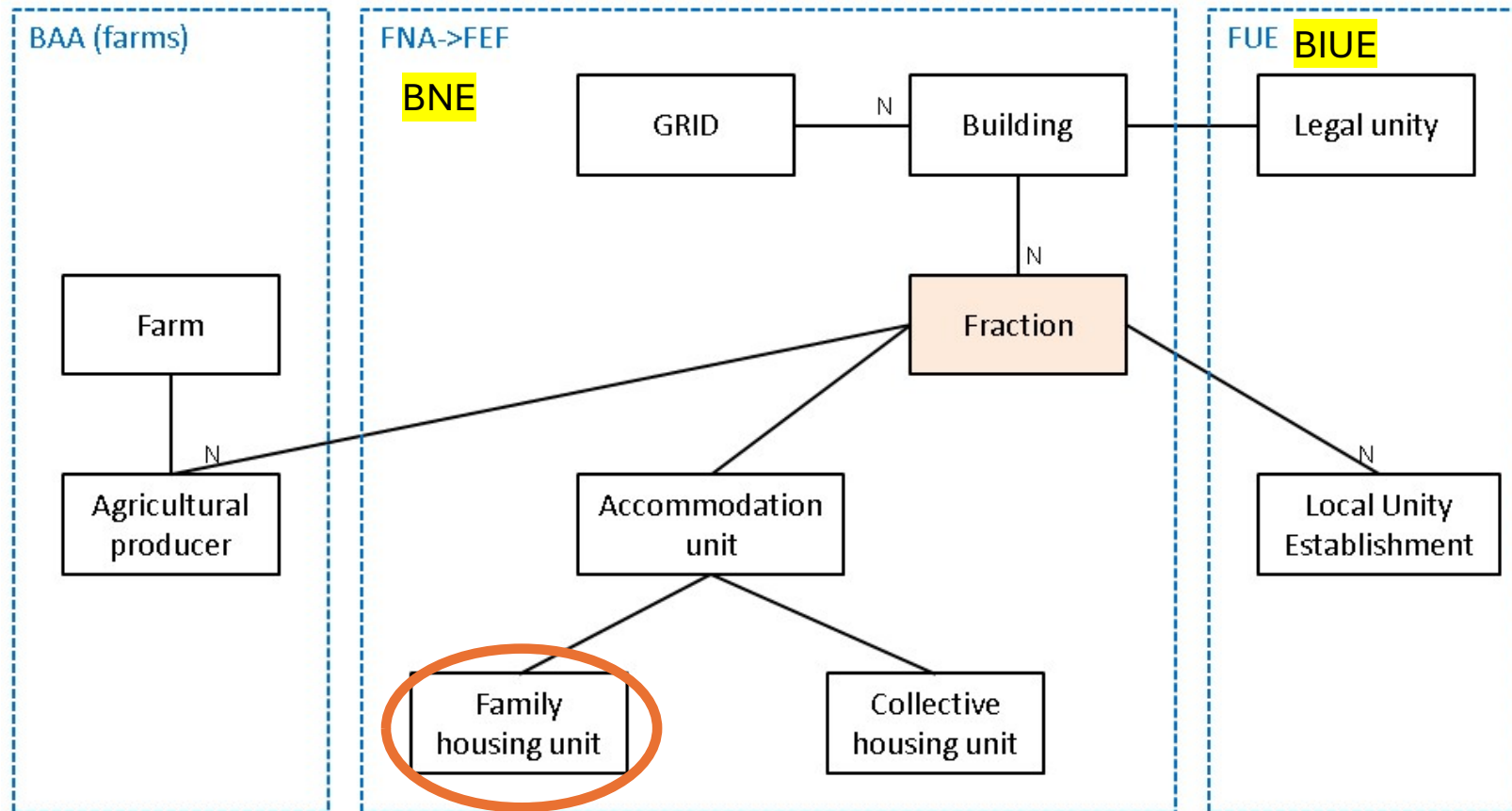
- Use high-resolution satellite imagery to create up-to-date sampling frames of households through the identification of living structures from the images (Himelein et al, 2017; Galway et al, 2012; Lowther et al, 2009).
- In addition to satellite-based approaches, practitioners have also been studying the feasibility of using MPD (Mobile Positioning Data) instead of a full household listing. In summary,

[https://unstats.un.org/wiki/display/MPDDPM/5.8.+Creating+dynam  
ic+sample+frames+for+surveys](https://unstats.un.org/wiki/display/MPDDPM/5.8.+Creating+dynam<br/>ic+sample+frames+for+surveys)



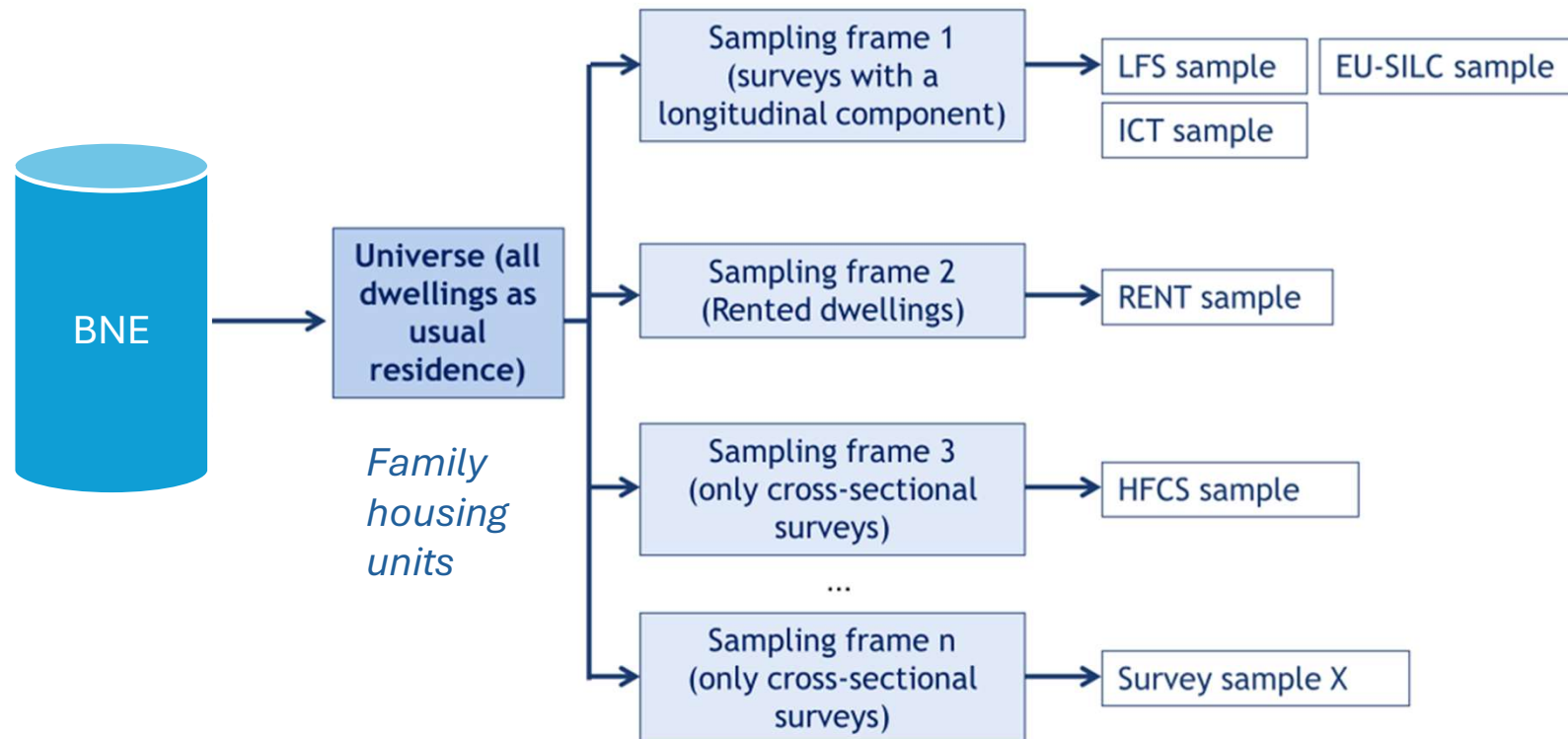
## 2. Main sources of administrative data

### Integration of Sampling Frames





## 2. Main sources of administrative data





## 2. Main sources of administrative data

### Resident Population Database (BPR)

#### Censuses with Administrative Data

- It consists of the resident population in Portugal and covers a range of geographical, demographic and socio-economic characteristics;
- It results from the integration of administrative information from various Public Administration sources.





## 3. Household Dynamics

### An important issue for Sampling Frames...

Dwellings can be considered as:

- **With** Occupying Households
- **Without** Occupying Households (Vacant or under construction)





# 3. Household Dynamics

## ISSUES THAT MATTER



how many dwellings were demolished?



how many dwellings were built?



What is the form of occupation of a dwelling?

Occupation status: occupied/non-occupied

### Births and deaths of dwellings/buildings

Can be taken from the Tax Authority (IMI) and SIOU.

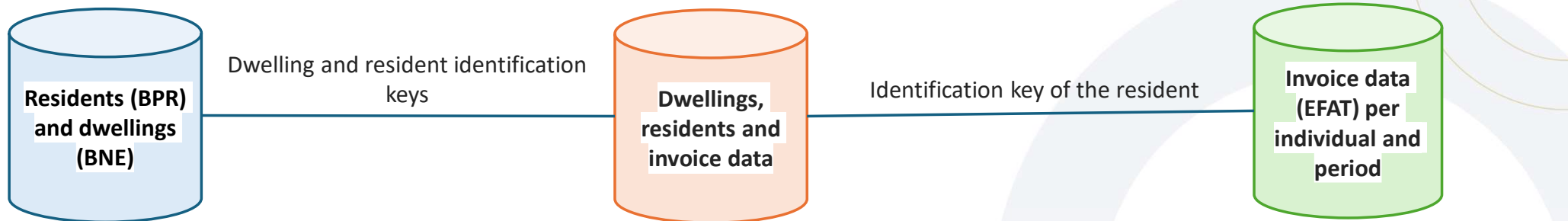
### Occupation

can be taken from BPR and indications of residence (ex. e-Fatura).

Ex. based on VAT Number of the service provider to one of the individuals in the dwelling (NIF vs COD-ALOJ) and associate electricity and water consumption



# 3. Household Dynamics



- BNE dwelling identification key (internal)
- Resident identification (encrypted VAT number in BPR) and associated dwelling based on previous matching between BNE and BPR
- Origin of the dwelling (Census 2011, 2021, etc.)
- Dwelling status and date of last update on BNE (main, secondary, etc.)
- Dwelling location (address, complete postal code and parish), as registered in BNE and BPR (and degree of address similarity)

- Dwelling identification and characterization (BNE and BPR)
- Resident identification (EFAT and BPR)
- Household's expenditure on water and energy

- Reference period (year, month)
- Encrypted VAT number of the consumer
- Supplier's issuing VAT number
- Supplier's issuing main location (Statistics Portugal Business Register - BDIFUE)
- Average taxable amount on energy and water



## 4. Discussion

- Update sampling frames
- Use appropriate metrics to measure:
  - the inflow and outflow of dwellings/individuals
  - The socio-demographic changes (occupation status)





## 4. Discussion

### References

- Galway, L. P., Bell, N., Sae, A. S., Hagopian, A., Burnham, G., Flaxman, A., ... & Takaro, T. K. (2012). A two-stage cluster sampling method using gridded population data, a GIS, and Google Earth<sup>TM</sup> imagery in a population-based mortality survey in Iraq. *International journal of health geographics*, 11(1), 1-9.
- Himelein, K., Eckman, S., Murray, S., & Bauer, J. (2017). Alternatives to full listing for second stage sampling: methods and implications. *Statistical Journal of the IAOS*, 33(3), 701-718.
- Lowther, S. A., Curriero, F. C., Shields, T., Ahmed, S., Monze, M., & Moss, W. J. (2009). Feasibility of satellite image-based sampling for a health survey among urban townships of Lusaka, Zambia. *Tropical Medicine & International Health*, 14(1), 70-78.
- Sullivan, J., Genadek, K., (2024) Associate Directorate for Economic Programs, U.S. Census Bureau, Working Paper Number ADEP-WP-2024-01



# EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS 2024 ESTORIL - PORTUGAL