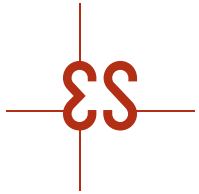


**Geocoding Statistical Business Register and it's effect on statistical dissemination - a good example how standardized address system can enhance the quality of statistical products**

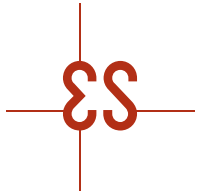
Mari-Liis Otsing, Ülle Valgma  
Statistics Estonia





# Overview

- Address Data System (ADS)
- Geocoding Statistical Business Register (SBR)
- Dissemination



# Address Data System

- The Address Data System was adopted on 1st of January 2008
- ADS has a register status
- Main goals are:
  - The creation of the central database of addresses
  - Implementation of the standard for geographical addresses

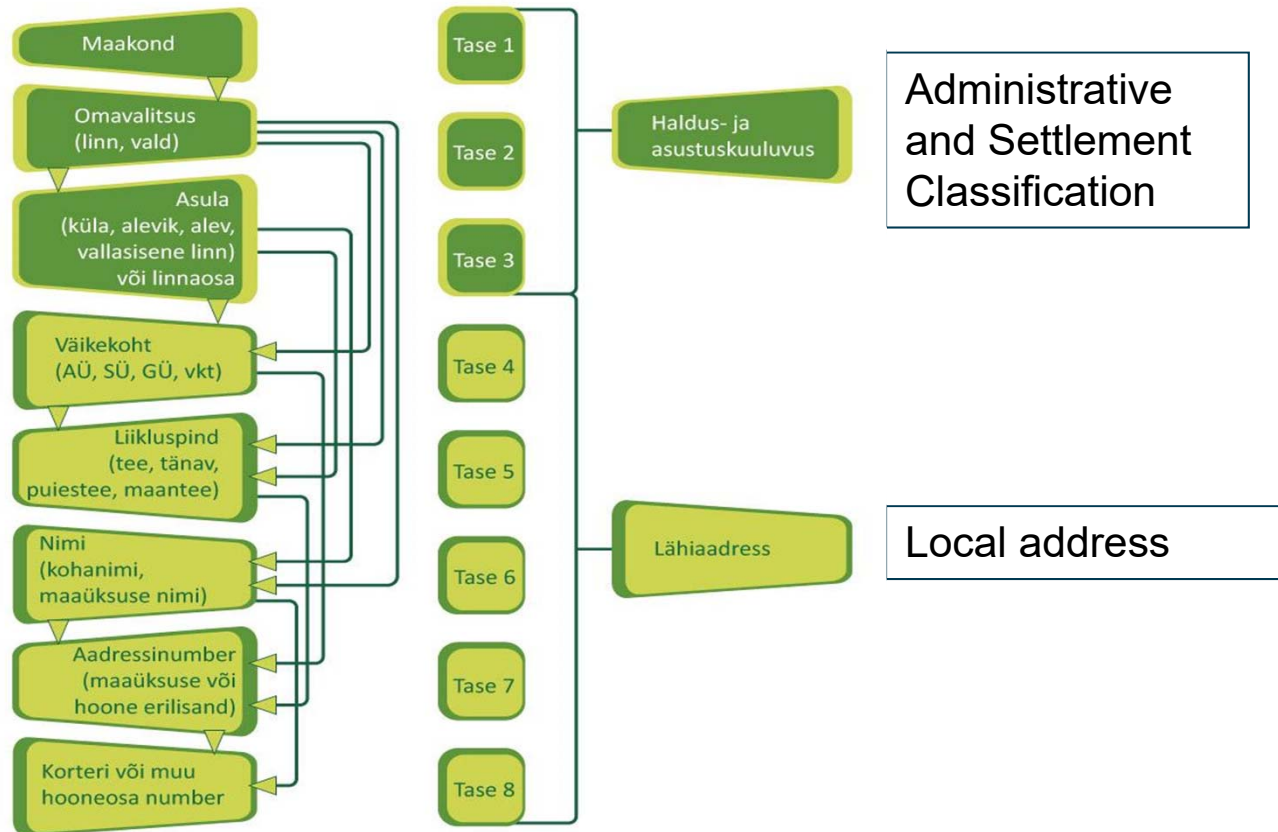


## Address standard (1)

- A normalised address is a text-form address that consist of structural elements (i.e. components) and follows business rules
- Normalisation is a process where free-form addresses will be put in a normalised format, i.e. in a certain order, with standardised spelling, punctuation and abbreviation use, and without unnecessary spaces.
  - Harju county, Tallinn, Kesklinna district, Aia street 15
  - Saare county, Valjala municipality, Jursi village, Salu
- In ADS, an address consist of 8 levels or components
- Geocoding means that the address is linked to an object in the nature and it has been attributed an address object identifier (ADS\_OID) and therefore also coordinated.

# ES Address standard (2)

1. County
2. Administrative unit (city, rural municipality)
3. Settlement unit (village, borough) or city district
4. Territorial address unit (small places, gardening cooperatives, garages)
5. Traffic surface (street, road, alley, highway, cross-street)
6. Name (place name, name of a land parcel)
7. Address number (number of a house or land parcel)
8. Part of the building (number of apartment or other part)



14.09.2017

# ES ADS services

- The Estonian Land Board as the chief processor of ADS data had developed many ways for the normalisation and geocoding of addresses:
  - X-road service
  - Mass geocoding from the table in the online application
  - Geocoding of a single item in the public online application

14.09.2017

The screenshot shows the 'Address Data Geocoding Service' web application. At the top, there is a logo for 'MAA-AMET' and the title 'Address Data Geocoding Service'. Below the title, there are navigation tabs for 'Home', 'Geocoding', and 'Massgeocoding'. The 'Geocoding' tab is active. Underneath, there is a section labeled 'Output' with a progress indicator showing 'Inputfile', 'Input', 'Output', and 'Result'. The main area contains several configuration options:

- ONLY NORMALISE
- CHOOSE OBJECTS**
  - House
  - Cadaster Unit
  - Traffic area
  - County/Municipality/Settlement Unit
- CHOOSE REPRESENTATION**
  - Perform generalization, when no exact equivalent is found
  - Only one object
  - With quality control
- CHOOSE CONTENT**
  - Levels
  - ADR ID
  - ADS OID
- ASYNCHRONOUS PROCESS
- OUTPUT COLUMN LABELS

At the bottom, there are 'Back' and 'Start' buttons.



# Geocoding Statistical Business Register (SBR) (1)

- In the case of SBR addresses, the X-Road normaliser ADSnormal.v2 was used. It is slower but suitable for processing large volumes of data
- Full list of addresses was compiled from live SBR register, frozen frame database and annually made register copies starting from 1995 (565 947 unique address)
- Addresses from SBR are as follows: code of Administrative and Settlement Classification + free form text which in most cases includes a zip code
- Removed zip codes and substituting obsolete administrative and settlement unit codes with valid ones

14.09.2017



## Geocoding SBR (2)

- Address data was loaded to the X-Road normaliser input table

Example of normalisation input table

<u>ID</u>	<u>Settlement unit (can be left empty)</u>	<u>Address</u>
1215		Lääne county, Lihula rural municipality, Lihula city, TALLINNA Road 23A
1231	784	Harju county, Tallinn, HEKI Road 14-46
1332		Tartu county, Nõo rural municipality, Nõo small town, SARAPUU
45681		Vändra, Pikk street, 39-5
43713		Tartu, Puiestee street, 49

- For every item the normaliser returned the response in the hugeclob format which was decoded into a standardised address
- Then the address object id (ADS\_OID) was found for the address id with an oracle script





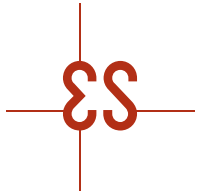
## Geocoding SBR (3)

- ADS\_OID should be as best as possible
  - For population and housing – begins with ER (apartment), MR (other rooms), EE (dwelling house) or ME (other buildings)
  - For business units – begins with EE or ME
- Treatment of non-matched addresses
  - CU (cadastre)
  - AY or LO (settlement)
  - OV (municipality)
  - MK (county)



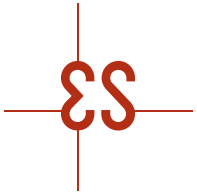
## Geocoding SBR (4)

- For addresses without automatic linking address levels were cut off starting from the end and so addresses passed through the X-Road normaliser several times
- 72% of addresses were successfully geocoded at first try
- In order to process manually, addresses were added the characteristic of belonging to a frozen frame of economic units active in 2015 and addresses which had this characteristic but were not normalised were reviewed first
- This helped to detect common errors which were then solved automatically and addresses were again sent to the X-Road normaliser
- Common errors were:
  - An unnecessary space between number and letter e.g. 4 A v.s. 4A
  - Type names 'farm', 'immovable' etc.
  - Dash instead of a missing address component
- As a result the percentage of geocoding addresses rose to 86 (98% for units active 2015)



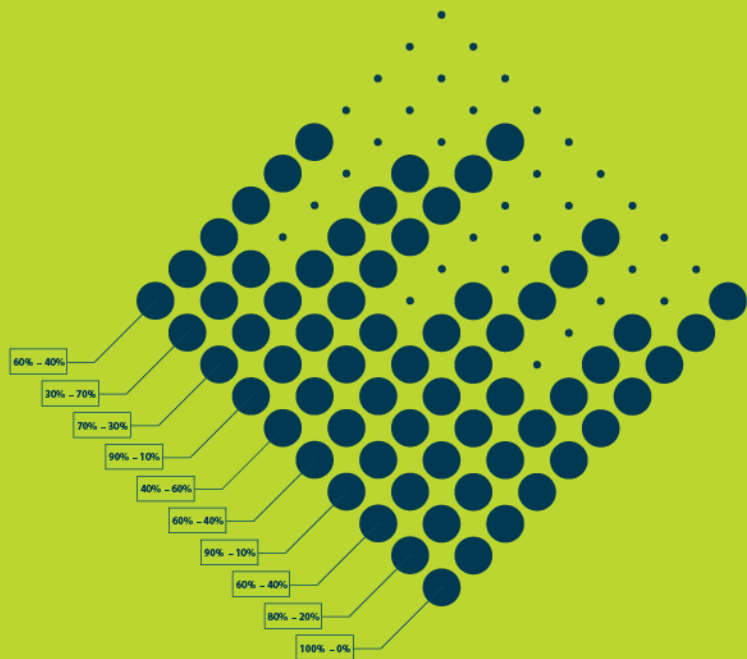
## Lessons learned

- The ADS services cannot process parallel addresses, they first need to be narrowed down to one
- The issue of what to do with apartments/rooms that are not in ADS
  - To generate fictitious codes or cut off the part of apartment/room
- Not all addresses were geocoded in the period of the grant
- The X-Road solution can be easily transferred to various databases

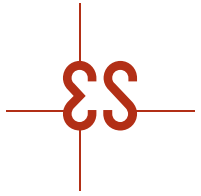


## Plans for the future

- ADS identifiers need to be updated time after time because at the moment local governments are organising addresses that do not comply with the standard in the ADS database and eliminating duplicated address objects
- To develop a solution for transferring manually found address identifiers to other databases in order to avoid duplication of work
- To continue processing of addresses that have not been geocoded

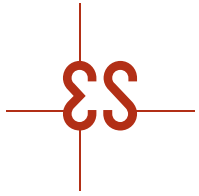


**ES** ESTE  
STATISTIKA



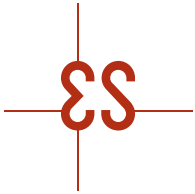
## Purpose of geocoding SBR data

- Give additional value through location
- Publish data on more detail spatial level
- Use this information in spatial analyses

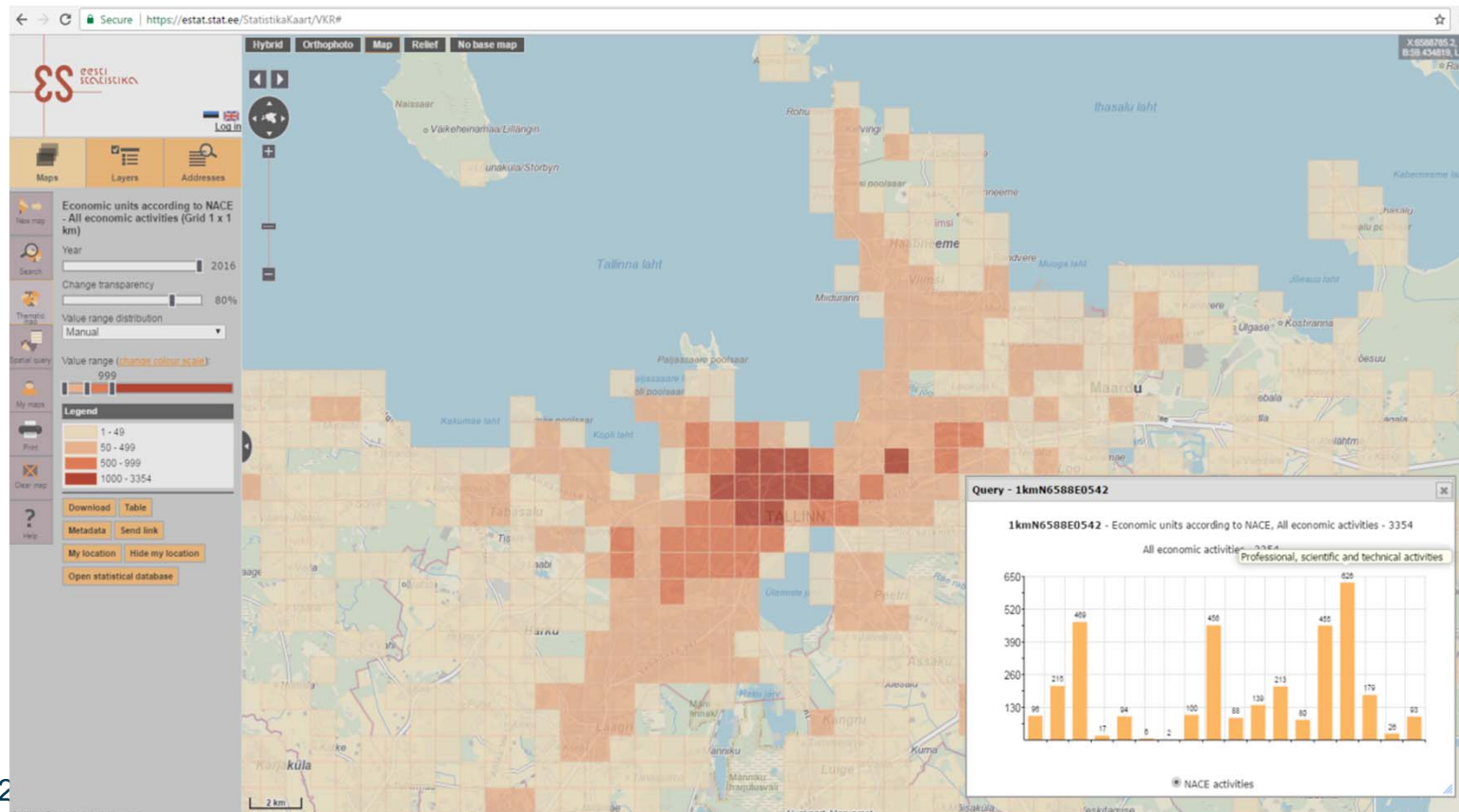


## Steps before disseminating

- Point based data – aggregation to any kind of area
- Grids – same spatial dimension for different statistics, constant in time
- Confidentiality

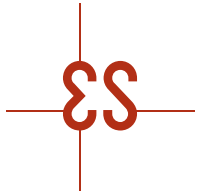


# Map application: kaart.stat.ee



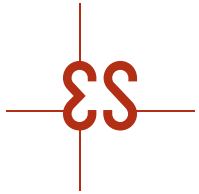
14.09.2





# Functionalities of map application

- View
- Study
- Share
- Download
- Address search
- Metadata
- Spatial query



## Conclusion

- Address standard enables geocoding addresses effectively
- Map application enables publishing statistics about uncommon areas, visualizing regional statistics, and making spatial queries

**Thank you for listening!**

