



Central Statistical
Bureau of Latvia

Geocoding Population Census 2000. Lessons Learned

Dāvis Kļaviņš, GIS specialist | davis.klavins@csb.gov.lv



Central Statistical
Bureau of Latvia

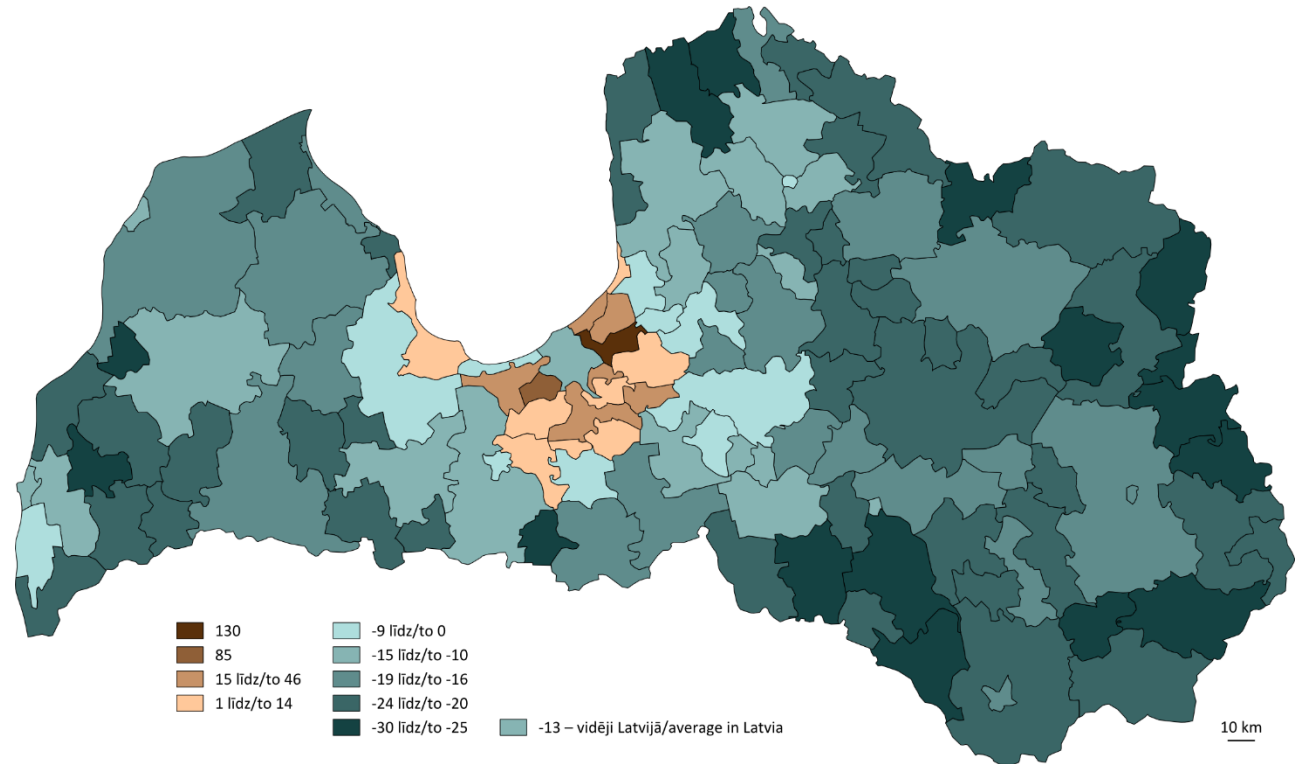
Geocoding data of Population Census 2000 is a part of Eurostat (statistical office of the European Union) supported grant project *Merging Statistics and Geospatial Information in Member States*. Project includes also analysis of migration flows and reasons for emigration and depopulation of certain areas based on the geocoded data.

Starting date 04.01.2016, duration 24 months



Central Statistical
Bureau of Latvia

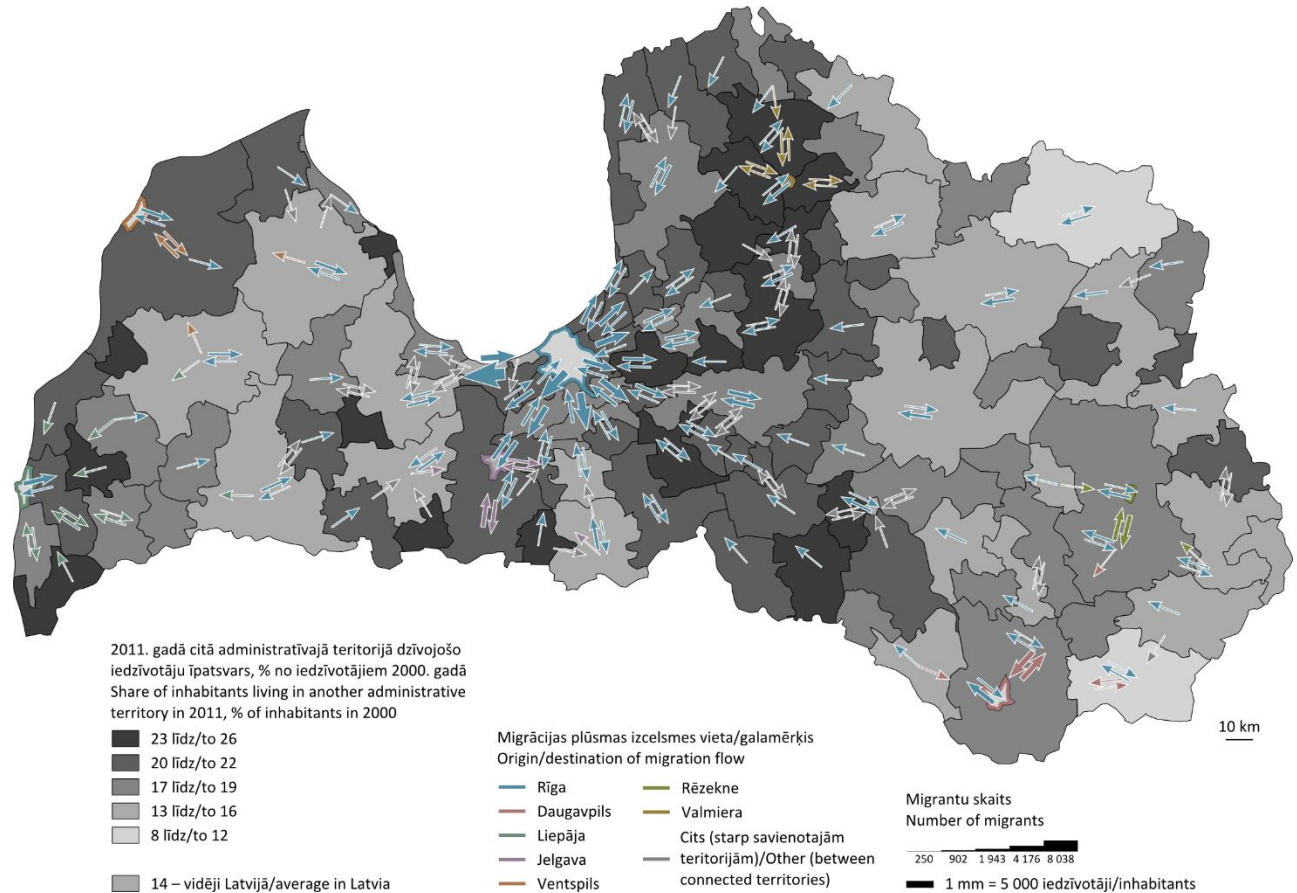
Population change in LAU 2 in 2000–2011, %





Central Statistical
Bureau of Latvia

Internal migration of population at LAU 2 level, 2000–2011





Central Statistical
Bureau of Latvia

For research on depopulation in regional breakdown, longitudinal geocoded dataset on population, employment and housing conditions is required. Such data can be obtained only in population censuses.

Data of Population Census 2011 are already geocoded, while data of Census 2000 are not, therefore their use in longitudinal research of depopulation is limited



Central Statistical
Bureau of Latvia

Geocoded data from Population Census 2000 and 2011 would provide ability for spatial planning professionals, researchers, students and other data users to analyse spatiotemporal patterns at more detailed level than at present when only data aggregated to administrative territories is available

Source codes of scripts created in the course of the project will be published and reusable



Action stages of the project

- **Geocoding Population Census 2000 dataset**
- Data publishing:
 - 1×1 km grid for Latvia and 100×100 m grid for the largest cities on population and housing indicators
 - Remotely accessible scientific use dataset containing geocoded data of Census 2000, as well as identifiers which will ensure compatibility between datasets of the Census 2000 and 2011
- Migration and depopulation data analysis using local indicators of spatial association



Central Statistical
Bureau of Latvia

Geocoding Population Census 2000 dataset

Census 2000 dataset contains information on the place of residence – address and territorial unit code

Address information (house number and street name or house name) from questionnaires was optically recognized and contains errors that must be eliminated



Central Statistical
Bureau of Latvia

Geocoding Population Census 2000 dataset

House number	Street name
/1	TARTU
11	
11	TART4
11	TARTU
2A	TA.RTU .
2A	TAR+4
2A	TART4
2A	TARTU
2A	TART-U
-2A	TARTU
5	
5	TART4
5	TARTU
5	TARTU -
5	TARTU.
ZA	TARTU



Central Statistical
Bureau of Latvia

Geocoding Population Census 2000 dataset

By linking address information of Census 2000 with the one of Population Register as of 1 March 2000 probability of equality between the both is determined

Address information of Population Register still contains format errors but much less than Census 2000 dataset



Central Statistical
Bureau of Latvia

Geocoding Population Census 2000 dataset

Probabilities were calculated in R software using fields:

1. Street/house name
2. House number
3. House block number
4. Territorial unit code



Central Statistical
Bureau of Latvia

Geocoding Population Census 2000 dataset

For each record's field, the Levenshtein distance was calculated (Levenshtein distance between two words is the minimum number of changes required to transform one word into the other). Afterwards it was divided by the longest of both words and subtracted from 1 thereby calculating the probability that the fields are equal.

DAUGAVPIL.5 → DAUGAVPILS (distance is equal to 2)

$$Probability = 1 - \frac{2}{11} \approx 0.82$$



Central Statistical
Bureau of Latvia

Geocoding Population Census 2000 dataset

Special cases when one word fits completely into the other. Even though fields often described the same geographic feature, calculated probabilities were low since one of the fields was shortened. Therefore some adjustments were made in probability calculation for these cases.

KR BARONA → KRĪŠJĀNA BARONA

$Probability_{before} = 0.6$

$Probability_{after} \approx 0.87$



Central Statistical
Bureau of Latvia

Geocoding Population Census 2000 dataset

Probabilities were calculated for each of the four fields and then combined with weights to calculate the combined probability for each record

In several (~30) steps, addresses from Census 2000 dataset were linked between 1) historical data from State Address Register and 2) Census 2011 dataset (only addresses which doesn't link with State Address Register) using MS SQL Server 2016

Linked datasets contain coordinates to be added to the Census 2000 dataset



Central Statistical
Bureau of Latvia

Geocoding Population Census 2000 dataset

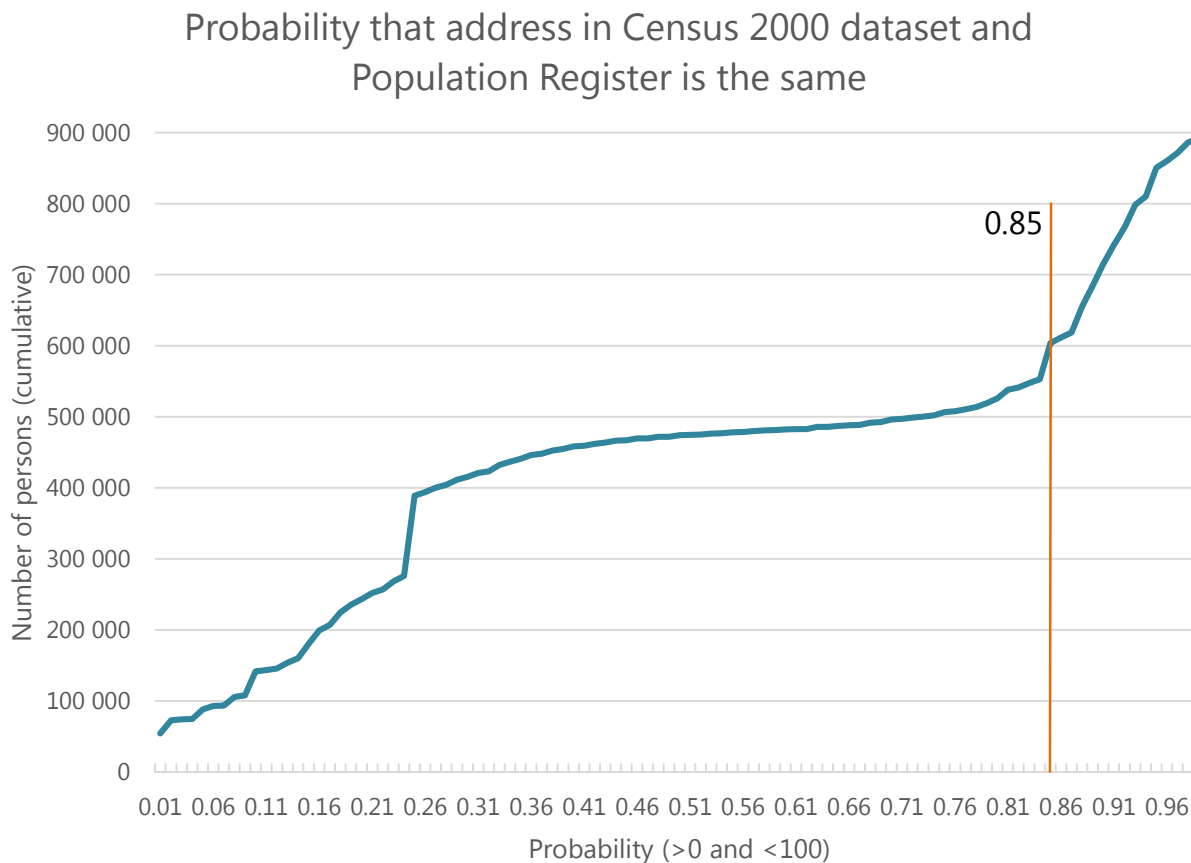
In some steps, calculated probabilities were used to substitute address information from questionnaires with the one from Population Register

E.g., if probability of address equality between the Census 2000 dataset and Population Register is >0.85 , address from Population Register was used to link with data from State Address Register (linked fields – village name, street name and house number)



Central Statistical
Bureau of Latvia

Geocoding Population Census 2000 dataset





Central Statistical
Bureau of Latvia

Geocoding Population Census 2000 dataset

For the remaining addresses, probability of address equality between Census 2000 and 2011 datasets was determined and under several conditions used to assume that address from Census 2011 can be used

E.g., if probability of equality between house name in the Census 2000 dataset and village name in the Census 2011 dataset is ≥ 0.7 (in the Census 2000 dataset, only village name was given in the field of house name and address from Census 2011 can be used if at least one person from the household lives at the same village according to Census 2011)



Central Statistical
Bureau of Latvia

Geocoding Population Census 2000 dataset

For addresses which were unable to be linked automatically and cover ~245 000 persons (10%), probability of address equality between Census 2000 and State Address Register was determined and 10 highest probabilities for each address used for manual linking (all in case highest probability is equal for more than 10 addresses)

Number of records to be manually linked is considerably less than ~245 000 (initially ~37 000) due to grouping by address but in some cases addresses must be split into households or persons



Central Statistical
Bureau of Latvia

Geocoding Population Census 2000 dataset

For manual linking, software was programmed to ease the process. Besides addresses from State Address Register with the highest probabilities of being equal, also access to these sources are given:

- Scanned questionnaires
- All surveyed addresses of the interviewer
- Registered places of residence up to year 2016 and address according to Census 2011



Central Statistical
Bureau of Latvia

Geocoding Population Census 2000 dataset

Papildus informācija

ATK2000: 7642
 ATKV_NAME: Aglonas pagasts
 NOV_NAME: Aglonas novads
 CIEMS: RU30NA5 JAUNAGLONA
 IELA: RU30NA5 JAUNAGLONA

MAJANOS: MAJASNUM: 9
 KORPUSS: SKAITIT: 76057
 SKAITIT: 1

Piev. veids: X:
 Adreses kods: Y:
 Piezīmes: Lietotājs

Skatītāja ATKV kods

ATK	count
7642	419

Filtri

Atk2000: Ciems: Iela: Majanos: Majasnum:

Korpus: Skaitit: Rādīt apstrādātās

Adrešu saraksts

ATVK	ciems	iela	majasnos	majasnum	korpus	skaitit	adr_merged
7642	RUŠONAS JAUNAGLONA		9	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUNAGLO.		5	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUNAGLO.		13	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUNAGLO.		7	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUNAGLO.		9	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUNAGLO.		9	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUNAGLO.		5	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUNAGLO.		15	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUNAGLO.		22	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUNAGLO.		4	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUNAGLO.		5	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUNAGLO.		7	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUNAGLO.		9	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUNAGLO.		-9	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUNAGLO.			76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUNAGLO.		5	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS JAUN		9	76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	RUŠONAS			76057		1-76057ATVK.7642IELA/MĀJA.RUŠO	
7642	SKUDRU GREBEŽI			76057		1-76057ATVK.7642IELA/MĀJA.SKUDR	
7642	STRAZDI JAUNAGLONA			76057		1-76057ATVK.7642IELA/MĀJA.STRA2	
7642	TRUŠU JAUNAGLONA			76057		1-76057ATVK.7642IELA/MĀJA.TRUŠ1	
7642	7AUDZEMU		2B	76057		1-76067ATVK.7642IELA/MĀJA.7AUD2	
7642	7AUDZEMU		2B	76057		1-76067ATVK.7642IELA/MĀJA.7AUD2	
7642	7AUDZEMU		5A	76057		1-76067ATVK.7642IELA/MĀJA.7AUD2	
7642	7		6	76057		1-76067ATVK.7642IELA/MĀJA.7MINU	
7642	A BROKA		3	76057		1-76067ATVK.7642IELA/MĀJA.A BRO	
7642	A BROKA		5	76057		1-76067ATVK.7642IELA/MĀJA.A BRO	
7642	A BROKA		7	76057		1-76067ATVK.7642IELA/MĀJA.A BRO	

Līdzīgas adreses

Ciems_VARIS	ID	Majas_nosaukums	Iela_VARIS	Majas_Nr_VARIS	Korpus_VARIS	ATVK_VARIS	Pazīme	Varbūvbas_kart
Jaunaglona	1-76057ATVK.76...		RUŠONAS IELA	9		0604342	1	1
Jaunaglona	1-76057ATVK.76...	DZINTARIŅŠ				0604342	1	2
Jaunaglona	1-76057ATVK.76...	EZERMAJĀ				0604342	1	2
Jaunaglona	1-76057ATVK.76...	JAUNAGLONAS ...				0604342	1	2
Jaunaglona	1-76057ATVK.76...	KALNAMĀJAS				0604342	1	2
Jaunaglona	1-76057ATVK.76...	LAGŪNA				0604342	1	2
Jaunaglona	1-76057ATVK.76...	LĪGAS				0604342	1	2
Jaunaglona	1-76057ATVK.76...	SKUDRINKA				0604342	1	2
Jaunaglona	1-76057ATVK.76...	STRAZDI				0604342	1	2
Jaunaglona	1-76057ATVK.76...	VECOZOLI				0604342	1	2

Filtri (VARIS)

ATVK: Ciems: Iela: Majanos: Majasnum:

Korpus: Adreses kods:

Adrešu saraksts (VARIS)

ATVK	Ciems	Majas_nosaukums	Iela	Majas_Nr	Korpus	Adreses_kods
						Most similar address from State Address Register
	Village					Jaunaglona
	Street	RU30NA5 JAUNAGLONA				RUŠONAS IELA
	House #	9				9



Geocoding Population Census 2000 dataset

maja_num	afk_m	iel_naj	maj_num	korp_num	dzvold	skatt	adr_merged	ATVK_piek	Ciems_piek	Majas_nosaukums	lela_piek	Majas_Nr_piek	Korpuss_piek
1090116	7642	JAUDZEMU	6	8	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	6		
1090117	7642	7	6	11	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	7		
1090118	7642	7A	6	10	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	7		
1090119	7642	8	6	9	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	8		
1090120	7642	9	6	12	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	9		
1090121	7642	10	6	13	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	10		
1090122	7642	11	6	14	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	11		
1090123	7642	12	6	15	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	12		
1090124	7642	13	6	16	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	13		
1090125	7642	14	6	17	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	14		
1090126	7642	15	6	18	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	15		
1090127	7642	16	6	19	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	16		
1090128	7642	17	6	20	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	17		
1090129	7642	18	6	21	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	18		
1090130	7642	19	6	22	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	19		
1090131	7642	20	6	23	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	20		
1090132	7642	21	6	24	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	21		
1090133	7642	22	6	25	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	22		
1090134	7642	23	6	26	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	23		
1090135	7642	24	6	27	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	24		
1090136	7642	25	6	28	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	25		
1090137	7642	26	6	29	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	26		
1090138	7642	27	6	30	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	27		
1090139	7642	28	6	31	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	28		
1090140	7642	29	6	32	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	29		
1090141	7642	30	6	33	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	30		
1090142	7642	31	6	34	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	31		
1090143	7642	32	6	35	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	32		
1090144	7642	33	6	36	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	33		
1090145	7642	34	6	37	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	34		
1090146	7642	35	6	38	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	35		
1090147	7642	36	6	39	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	36		
1090148	7642	37	6	40	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	37		
1090149	7642	38	6	41	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	38		
1090150	7642	39	6	42	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	39		
1090151	7642	40	6	43	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	40		
1090152	7642	41	6	44	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	41		
1090153	7642	42	6	45	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	42		
1090154	7642	43	6	46	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	43		
1090155	7642	44	6	47	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	44		
1090156	7642	45	6	48	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	45		
1090157	7642	46	6	49	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	46		
1090158	7642	47	6	50	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	47		
1090159	7642	48	6	51	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	48		
1090160	7642	49	6	52	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	49		
1090161	7642	50	6	53	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	50		
1090162	7642	51	6	54	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	51		
1090163	7642	52	6	55	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	52		
1090164	7642	53	6	56	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	53		
1090165	7642	54	6	57	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	54		
1090166	7642	55	6	58	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	55		
1090167	7642	56	6	59	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	56		
1090168	7642	57	6	60	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	57		
1090169	7642	58	6	61	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	58		
1090170	7642	59	6	62	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	59		
1090171	7642	60	6	63	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	60		
1090172	7642	61	6	64	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	61		
1090173	7642	62	6	65	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	62		
1090174	7642	63	6	66	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	63		
1090175	7642	64	6	67	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	64		
1090176	7642	65	6	68	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	65		
1090177	7642	66	6	69	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	66		
1090178	7642	67	6	70	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	67		
1090179	7642	68	6	71	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	68		
1090180	7642	69	6	72	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	69		
1090181	7642	70	6	73	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	70		
1090182	7642	71	6	74	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	71		
1090183	7642	72	6	75	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	72		
1090184	7642	73	6	76	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	73		
1090185	7642	74	6	77	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	74		
1090186	7642	75	6	78	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	75		
1090187	7642	76	6	79	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	76		
1090188	7642	77	6	80	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	77		
1090189	7642	78	6	81	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	78		
1090190	7642	79	6	82	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	79		
1090191	7642	80	6	83	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	80		
1090192	7642	81	6	84	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	81		
1090193	7642	82	6	85	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	82		
1090194	7642	83	6	86	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	83		
1090195	7642	84	6	87	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	84		
1090196	7642	85	6	88	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	85		
1090197	7642	86	6	89	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	86		
1090198	7642	87	6	90	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	87		
1090199	7642	88	6	91	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	88		
1090200	7642	89	6	92	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	89		
1090201	7642	90	6	93	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	90		
1090202	7642	91	6	94	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	91		
1090203	7642	92	6	95	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	92		
1090204	7642	93	6	96	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	93		
1090205	7642	94	6	97	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	94		
1090206	7642	95	6	98	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	95		
1090207	7642	96	6	99	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	96		
1090208	7642	97	6	100	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	97		
1090209	7642	98	6	101	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	98		
1090210	7642	99	6	102	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	99		
1090211	7642	100	6	103	76067	1-76067ATVK	0604342	Aglona		JAUDZEMU I	100		

Within all surveyed addresses of the interviewer – previous and next household is at the same street



Central Statistical
Bureau of Latvia

Geocoding Population Census 2000 dataset

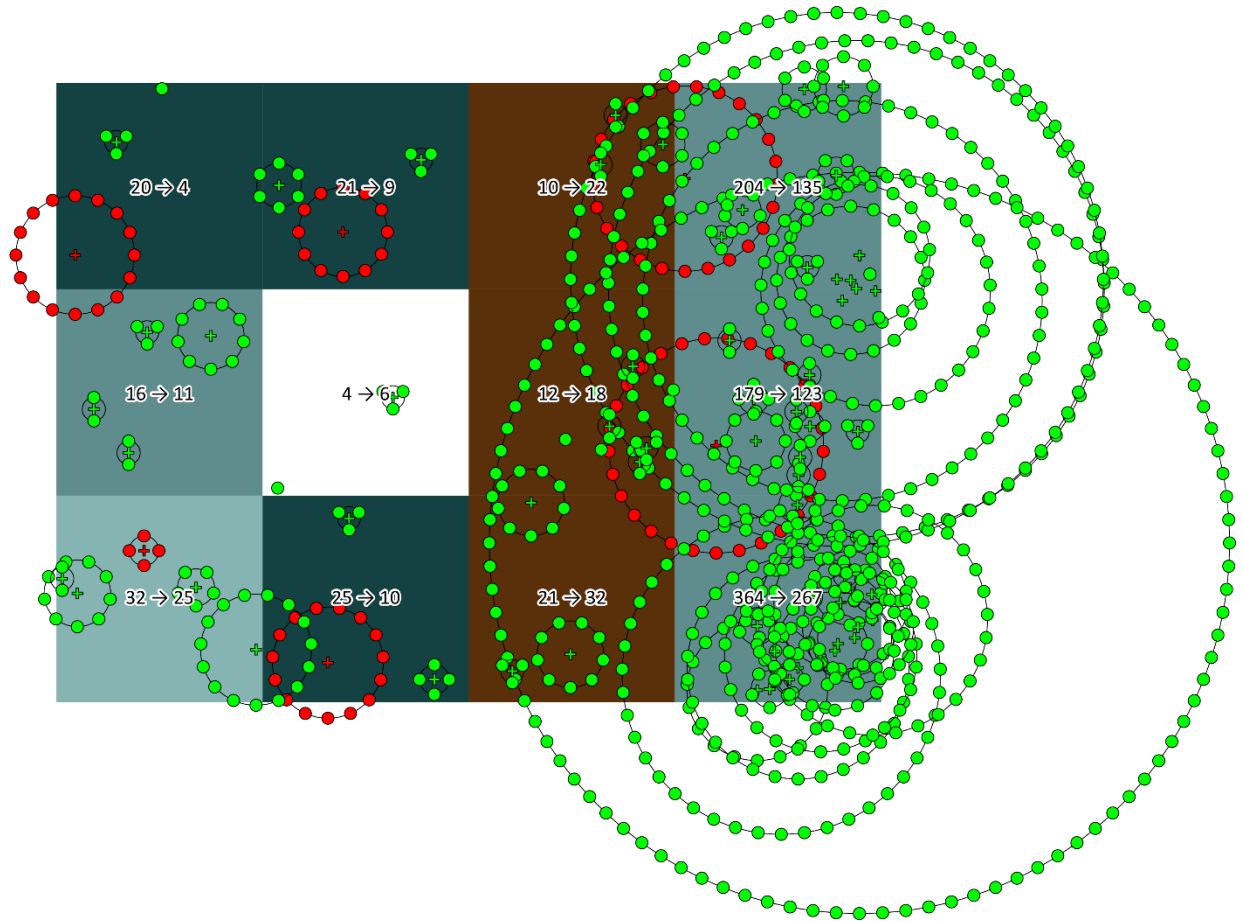
	Number of persons	% of population
Automatically linked	2 081 263	88
Manually linked (incl. automatically linked without coordinates, more than one address automatically linked and corrected errors in automatic linking)	295 931	12



Central Statistical
Bureau of Latvia

Interpolating household coordinates if only village name given in questionnaires

- Census 2000 person
 - Census 2000 person assigned to village
- 20 → 4 Usually resident population change; 2000–2011
- Usually resident population change; 2000–2011 (%)
- >45 or uninhabited in 2000
 - 36 to 45
 - 26 to 35
 - 16 to 25
 - 6 to 15
 - -5 to 5 or number of population is between 1 and 10 in at least one year and <19 in 2000 or <15 in 2011
 - -15 to -6
 - -25 to -16
 - -35 to -26
 - -45 to -36
 - <-45 or uninhabited in 2011





Central Statistical
Bureau of Latvia

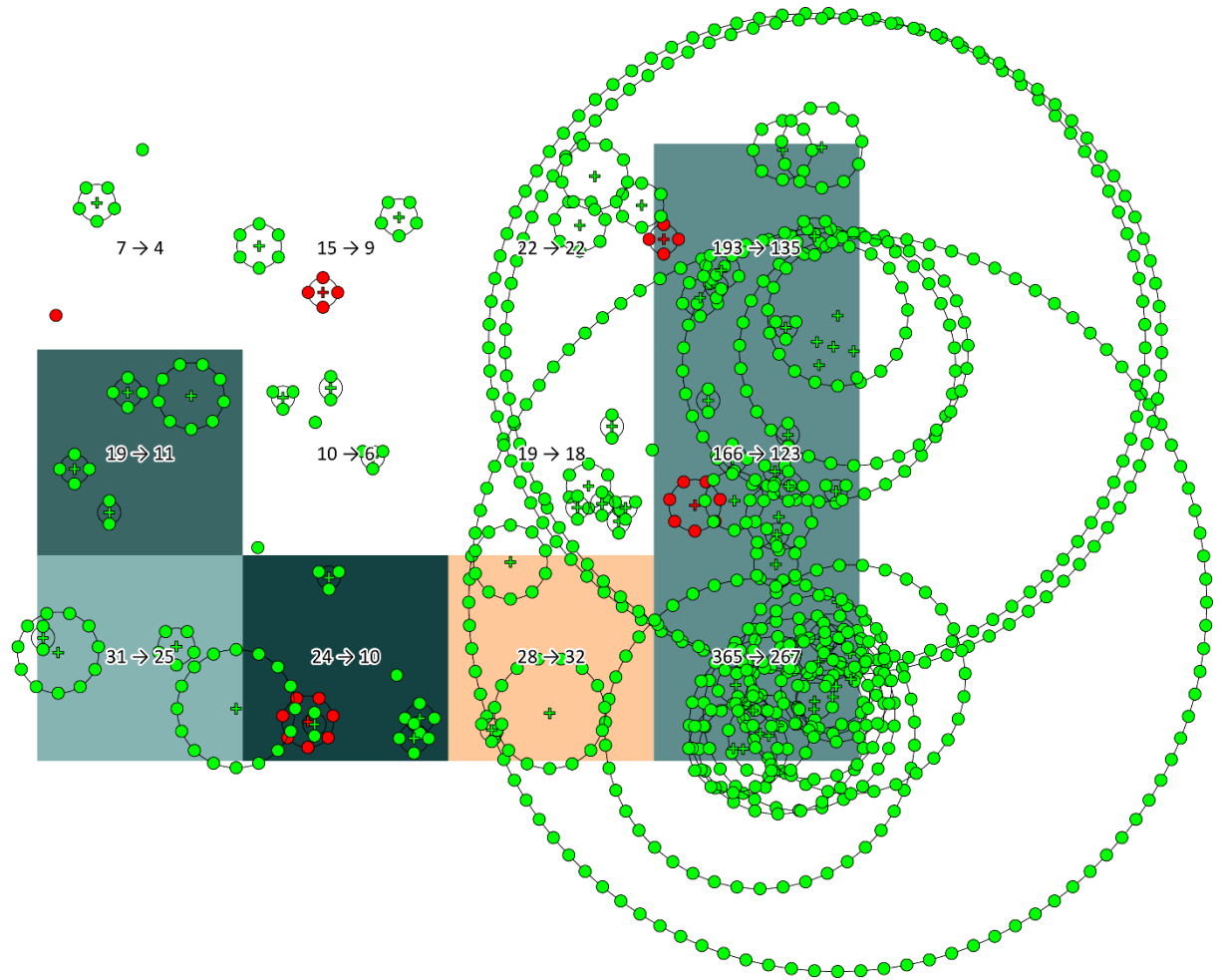
Interpolating household coordinates if only village name given in questionnaires

- Census 2000 person
- Census 2000 person assigned to village

20 → 4 Usually resident population change; 2000–2011

Usually resident population change; 2000–2011 (%)

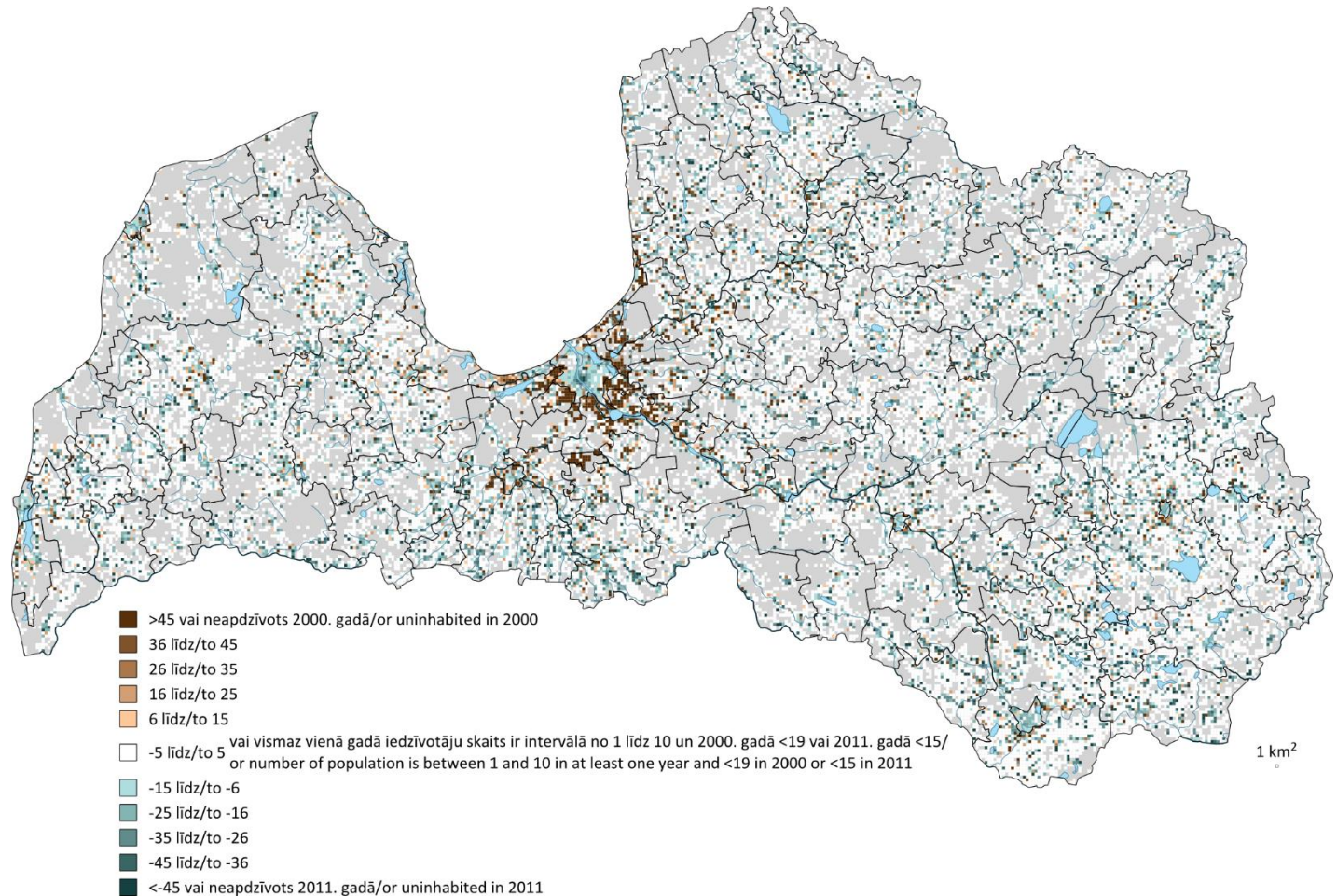
- >45 or uninhabited in 2000
- 36 to 45
- 26 to 35
- 16 to 25
- 6 to 15
- -5 to 5 or number of population is between 1 and 10 in at least one year and <19 in 2000 or <15 in 2011
- -15 to -6
- -25 to -16
- -35 to -26
- -45 to -36
- <-45 or uninhabited in 2011





Central Statistical
Bureau of Latvia

Preliminary results: change of usually resident population, 2000–2011





Central Statistical
Bureau of Latvia

