Zastosowanie nowoczesnych metod badawczych w badaniach eksperymentalnych dotyczących ludności

Gabriela Nowakowska, Krzysztof Woźnica, Daniel Godlewski, Paweł Murawski

Departament Innowacji

Wydział Inżynierii Danych

24.03.2025 Warszawa

Wprowadzenie

- Doświadczenie w realizacji NSP 2011 i NSP 2021
- Zdobycie kompetencji i doświadczenia w standaryzacji i integracji danych
- Dostęp do rejestrów administracyjnych

Dane wejściowe

Określenie 2 typów rejestrów:

- systemy referencyjne („*rdzenie*”) – cechują się najwyższą jakością danych, kompletnością i aktualnością oraz wysokim pokryciem podmiotowym;

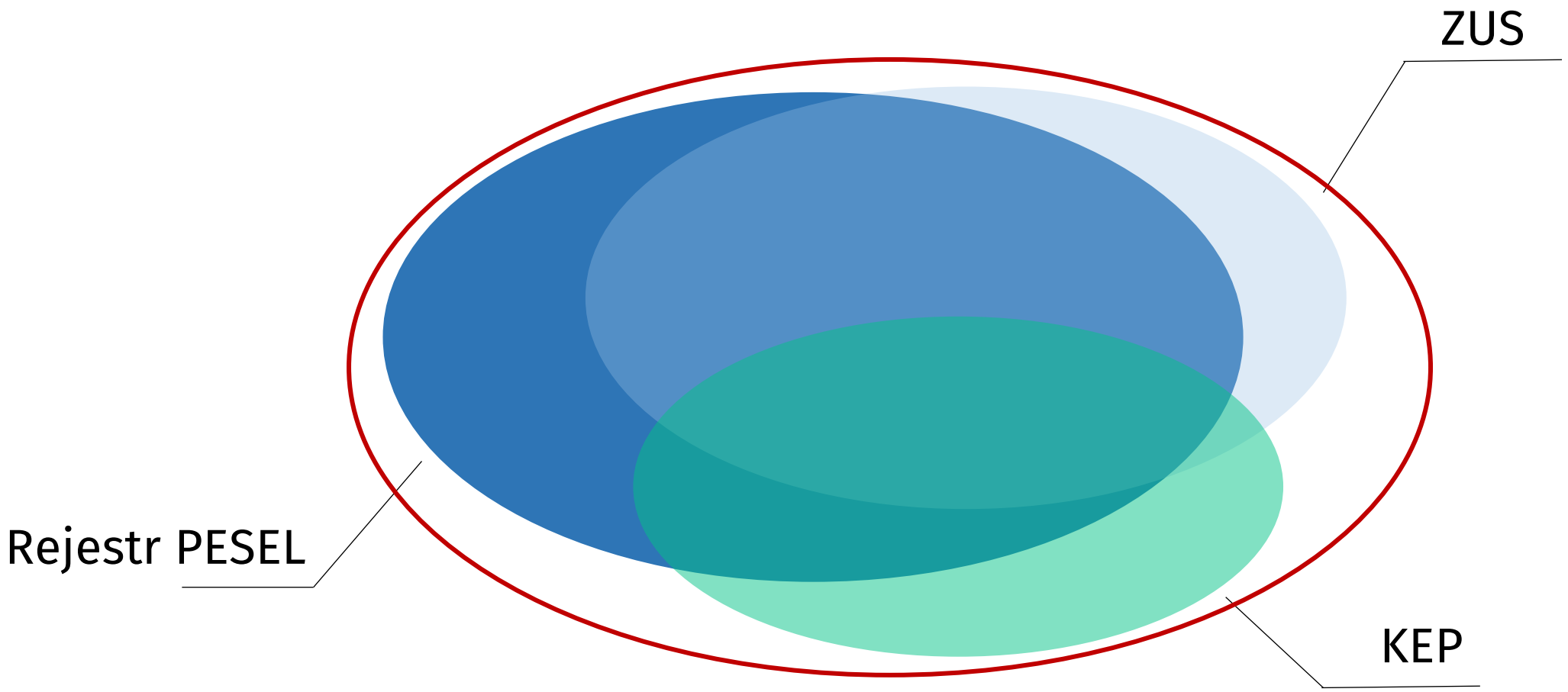
- systemy zasilające – pozostałe systemy, rejestry, bazy danych, wykazy, ewidencje.

Dane rejestrowe

Do budowy wykazu osób wytypowano następujące rejestry:

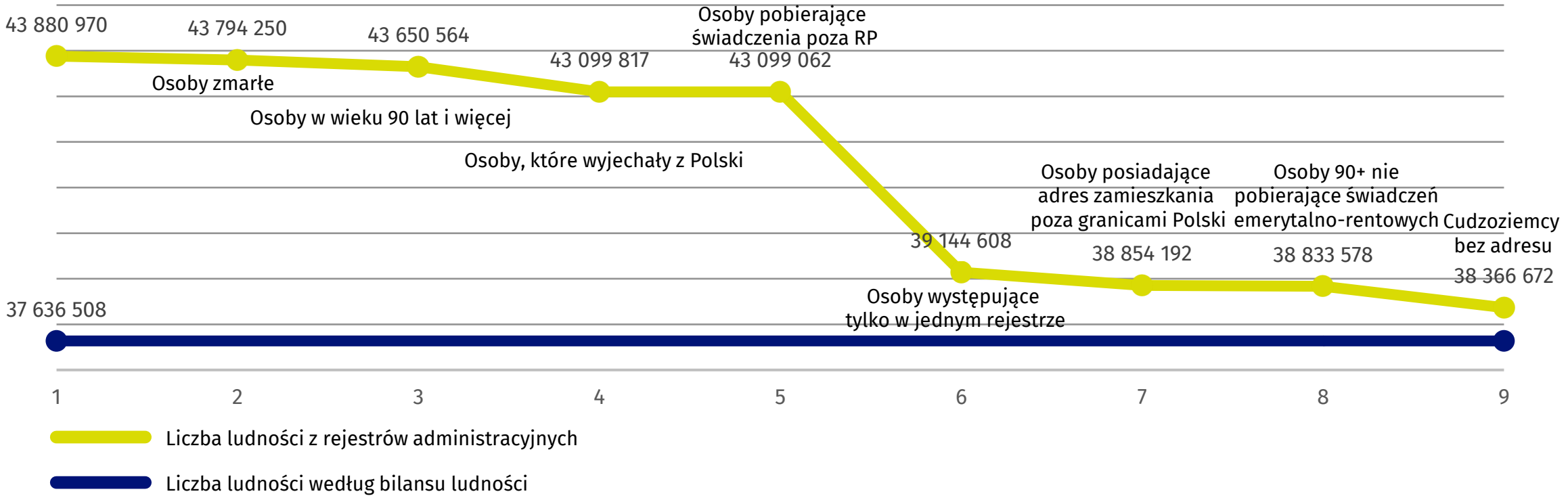
- Powszechny Elektroniczny System Ewidencji Ludności (PESEL);
  - Kancelaria Prezesa Rady Ministrów;
- Centralny Rejestr Podmiotów – Krajowa Ewidencja Podatników (KEP);
  - Ministerstwo Finansów;
- Centralny Rejestr Ubezpieczonych, Centralny Rejestr Płatników Składek;
  - Zakład Ubezpieczeń Społecznych (ZUS);
- nSIU, System obsługi emerytur i rent;
  - Kasa Rolniczego Ubezpieczenia Społecznego (KRUS);
- Centralny Wykaz Ubezpieczonych;
  - Narodowy Fundusz Zdrowia (NFZ);
- System informacyjny ARiMR, System Identyfikacji i Rejestracji Zwierząt;
  - Agencja Restrukturyzacji i Modernizacji Rolnictwa (ARiMR);
- Inne, m.in.: dane z Urzędu do spraw Cudzoziemców, Ministerstwa Rodziny, Pracy i Polityki Społecznej, Ministerstwa Edukacji Narodowej.

Integracja danych



Ślady życia

Wystąpiła znacząca różnica między statystykami bilansu ludności, a populacją z rejestrów administracyjnych.



Uzupełnianie wykazu – zmienna „obywatelstwo”

Z wyznaczonych zasobów informacyjnych dołączono kod obywatelstwa:

polskie	36 116 507
ukraińskie	1 645 743
białoruskie	176 236
gruzińskie	57 308
mołdawskie	39 279
pozostałe	331 599

Uzupełnianie wykazu o dane adresowe

O hierarchii wyboru adresu z konkretnego zbioru decydowały aktualność danych oraz referencyjność w zakresie danych adresowych. W stosunku do wszystkich adresów brano pod uwagę poziom spójności.

Przeprowadziliśmy kilka symulacji. Testowane warianty różniły się kolejnością dołączania adresu oraz liczbą źródeł.

Sprawdzanie rozbieżności względem bilansu ludności.

Wskaźniki jakości danych wyjściowych

Przypisanie adresu do osoby:

99,97%

Wskaźniki jakości danych wyjściowych

Połączenie z NOBC:

95,87%

współrzędne x,y

Liczba ludności

31 grudnia 202X	30 czerwca 202X	31 grudnia 202Y	30 czerwca 202Y
Identyfikator	Identyfikator	Identyfikator	Identyfikator
11111111111	11111111111	11111111111	11111111111
22222222222	22222222222	22222222222	22222222222
33333333333	33333333333	33333333333	33333333333
	44444444444	44444444444	44444444444
55555555555	55555555555	55555555555	55555555555
66666666666	66666666666	66666666666	
77777777777	77777777777	77777777777	77777777777
	888888888	888888888	
999999999	999999999	999999999	999999999

Porównanie z bilansem ludności

Ludność w Polsce stan na 31.12.2023 r.

Zakres różnicy	Liczba gmin
0 %	9
0,01-0,99 %	808
1-1,99 %	739
2-2,99 %	527
3-3,99 %	331
4-4,99 %	223
5-5,99 %	127
6-6,99 %	102
7-7,99 %	79
8-8,99 %	42
9-9,99 %	33
10% i więcej	151
RAZEM	3171

Gminy miejsko-wiejskie ujęte osobno.

Porównanie z bilansem ludności

Skąd tak duże rozbieżności? Cudzoziemcy?!

Kod TERYT	Nazwa gminy	Liczba ludności metoda bilansowa	Liczba ludności metoda rejestrowa	Różnica ilościowa	Różnica procentowa	Liczba cudzoziemców	Różnica procentowa bez cudzoziemców
3018064	Mikstat	1702	4528	2826	-166%	2845	1%
1020084	Stryków	3363	5854	2491	-74%	2582	3%
1020011	Głowno	13116	21197	8081	-62%	8160	1%
0812014	Sława	4103	6028	1925	-47%	2042	3%
0215034	Jelcz-Laskowice	15004	21389	6385	-43%	7047	4%
0220024	Prusice	2144	3017	873	-41%	944	3%
3024084	Wronki	10505	14683	4178	-40%	4349	2%
3208072	Ustronie Morskie	3744	5383	1639	-44%	1898	7%

Miasta na prawach powiatu

Wykaz Osób do badań eksperymentalnych

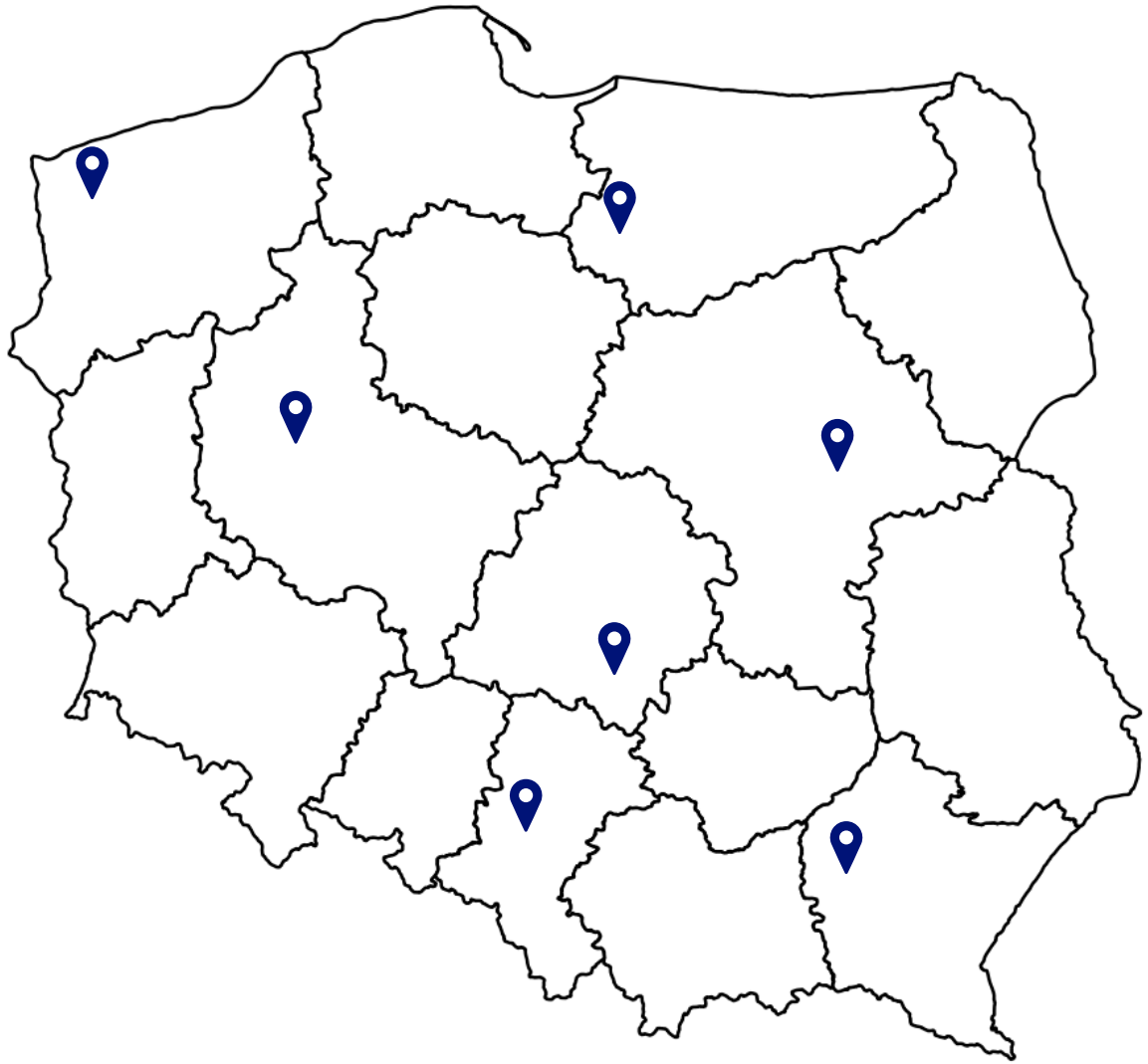
UNS/PESEL	Występowanie	Flaga śladów życia	Płeć	Wiek	Obywatelstwo	Adres	Współrzędne x,y
52015	PESEL KEP ZUS	1	1	65	PL	ul. Senatorska 12/4, /.../	487195.1458 ; 27809.2147
211	PESEL ZUS	1	2	47	PL	ul. Adama Asnyka 5, /.../	744135.4511 ; 24477.2144
9874414	PESEL KEP ZUS	1	1	22	PL	ul. Żuławska 2/29, /.../	157144.8258 ; 85874.1257
41552278	KRUS NFZ	1	2	12	PL	ul. Jana Brzechwy 89, /.../	877455.1898 ; 27111.9854
9841231	NFZ	3	2	41	PL	ul. Siennicka 57A, /.../	487195.5054 ; 28749.2441
39458159	NFZ UDSC	1	2	38	UKR	al. Józefa Becka 55/2, /.../	159197.1010 ; 18159.9189
987441	PESEL ARIMR	7	1	53	PL	ul. Akacjowa 14, /.../	218185.1718 ; 12808.7188

UNS/PESEL	Pracujący
52015	1
211	1
9874414	1
41552278	1
9841231	0
39458159	0
987441	1

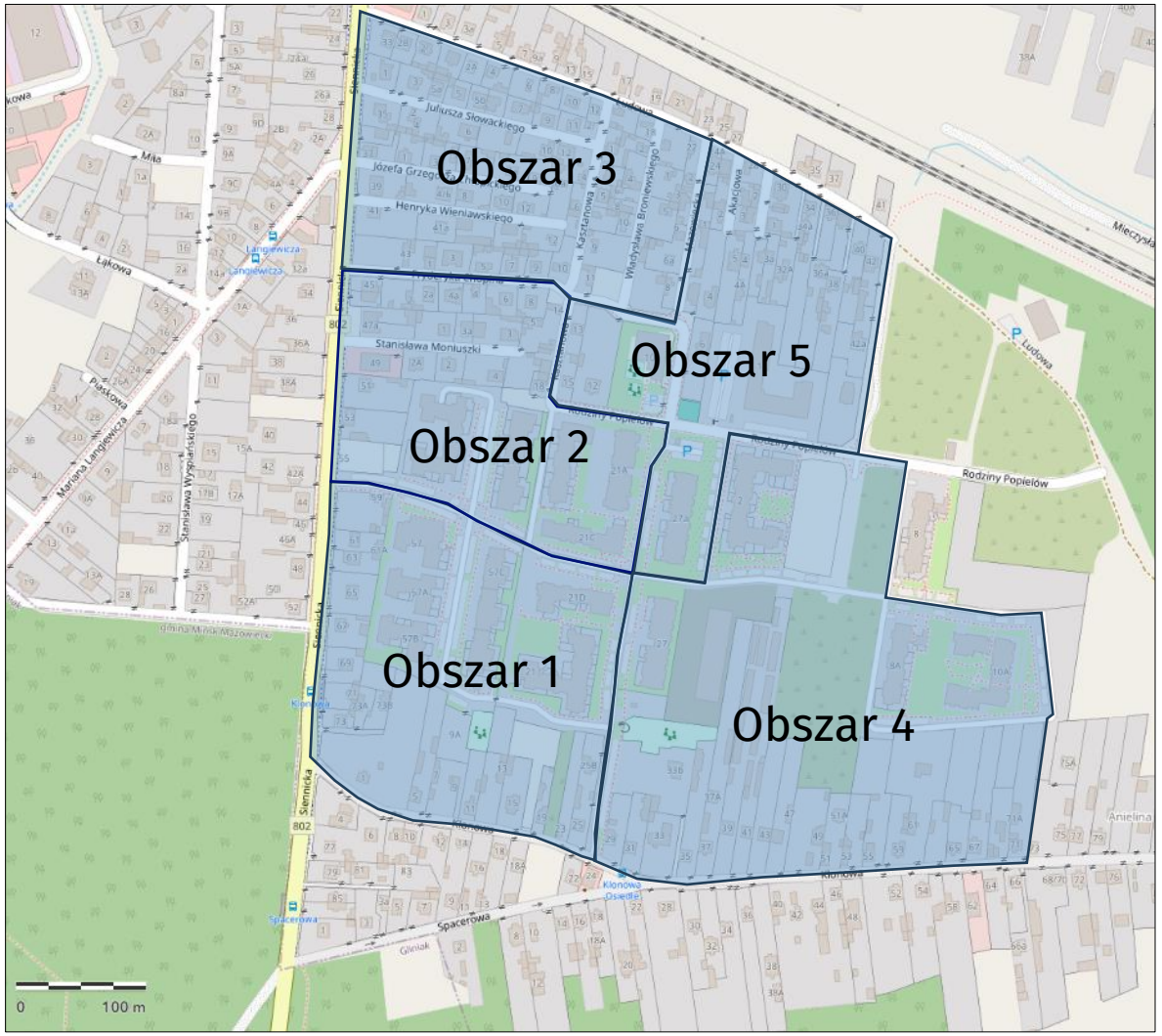
UNS/PESEL	Wykształcenie
52015	5
211	3
9874414	2
41552278	5
9841231	2
39458159	2
987441	1

Wykaz Osób do badań eksperymentalnych

	Adres	Współrzędne x,y
	ul. Senatorska 12/4, /.../	487195.1458 ; 27809.2147
	ul. Adama Asnyka 5, /.../	744135.4511 ; 24477.2144
	ul. Żuławska 2/29, /.../	157144.8258 ; 85874.1257
	ul. Jana Brzechwy 89, /.../	877455.1898 ; 27111.9854
	ul. Siennicka 57A, /.../	487195.5054 ; 28749.2441
	al. Józefa Becka 55/2, /.../	159197.1010 ; 18159.9189
	ul. Akacjowa 14, /.../	218185.1718 ; 12808.7188



Wykaz Osób do badań eksperymentalnych

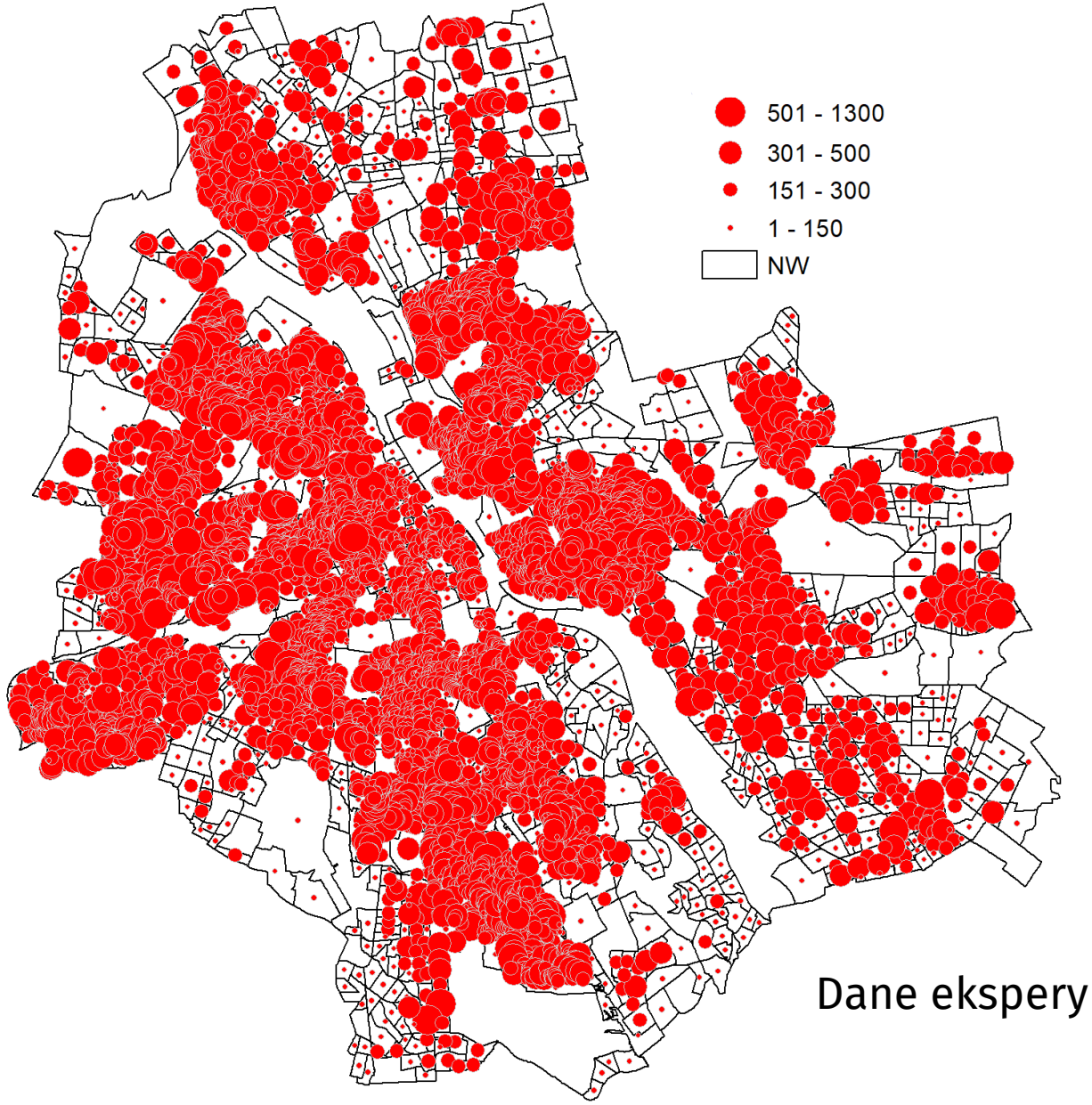


Tablica wynikowa: liczba ludności według wybranych grup wieku.

Wyszczególnienie	w wieku przedprodukcyjnym	w wieku produkcyjnym	w wieku poprodukcyjnym
Obszar 1	98	138	89
Obszar 2	63	59	41
Obszar 3	59	65	43
Obszar 4	30	36	38
Obszar 5	14	23	50

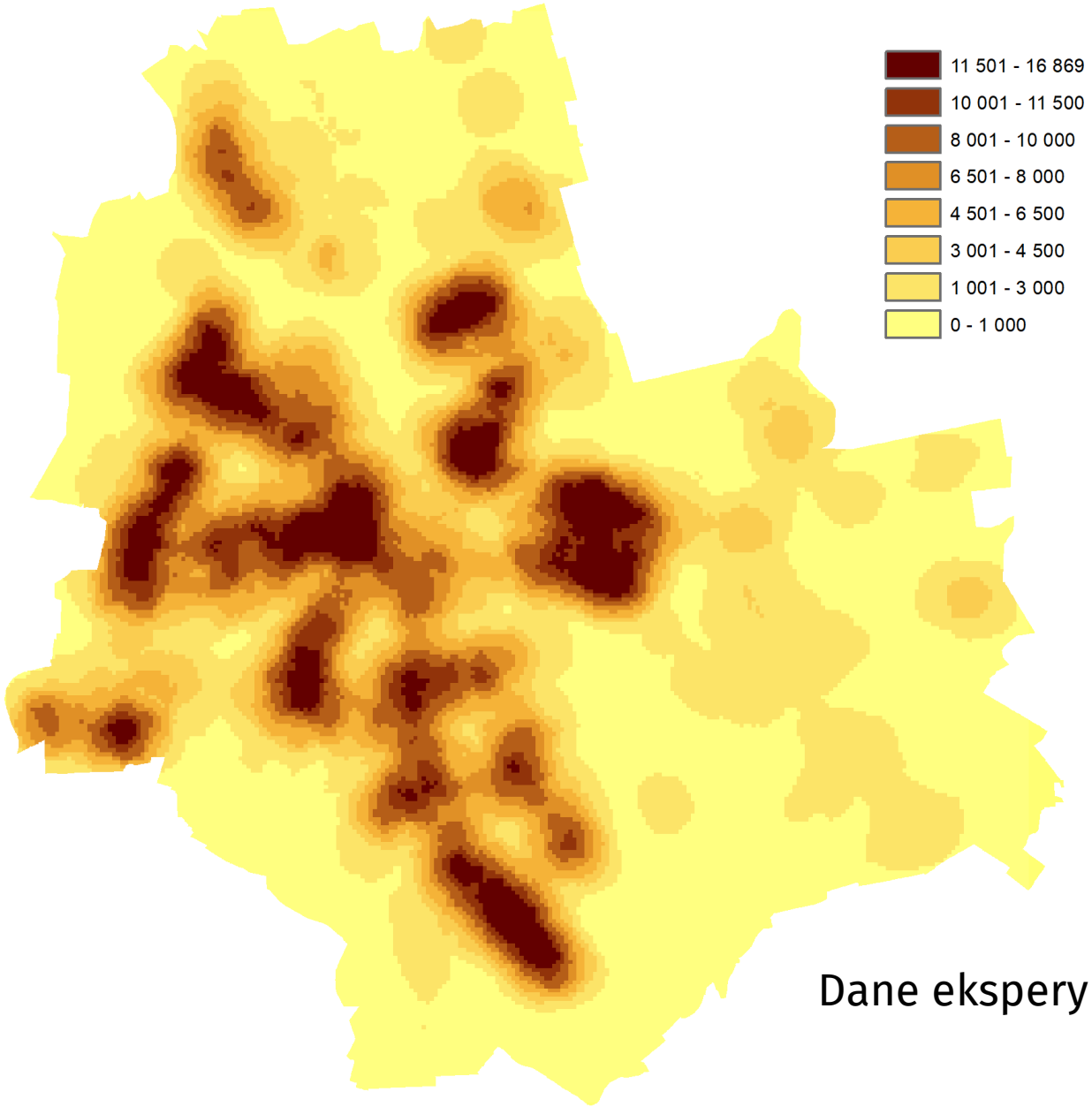
Liczba mieszkańców wg sołectw

Liczba ludności



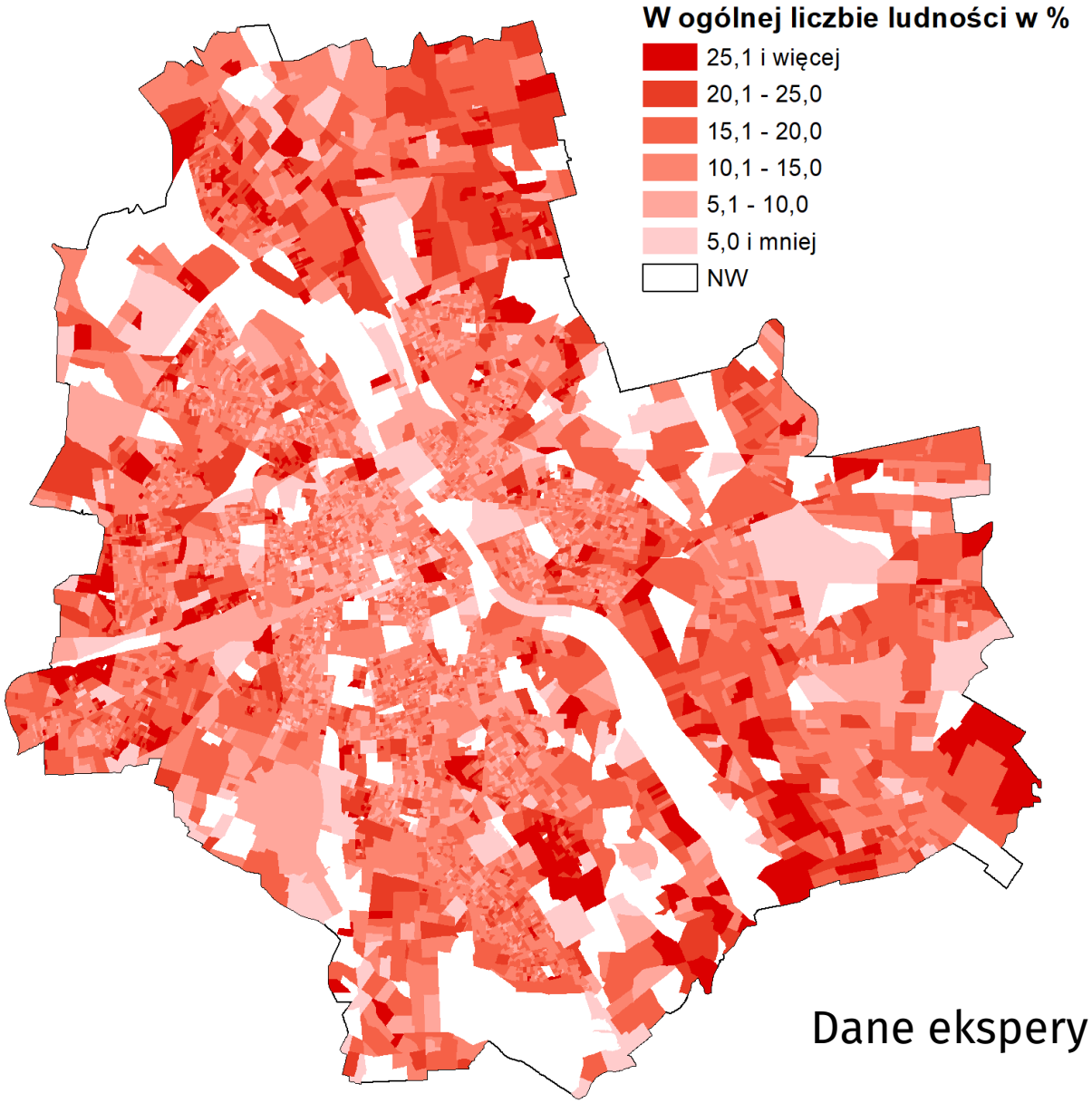
Dane eksperymentalne

Gęstość zaludnienia



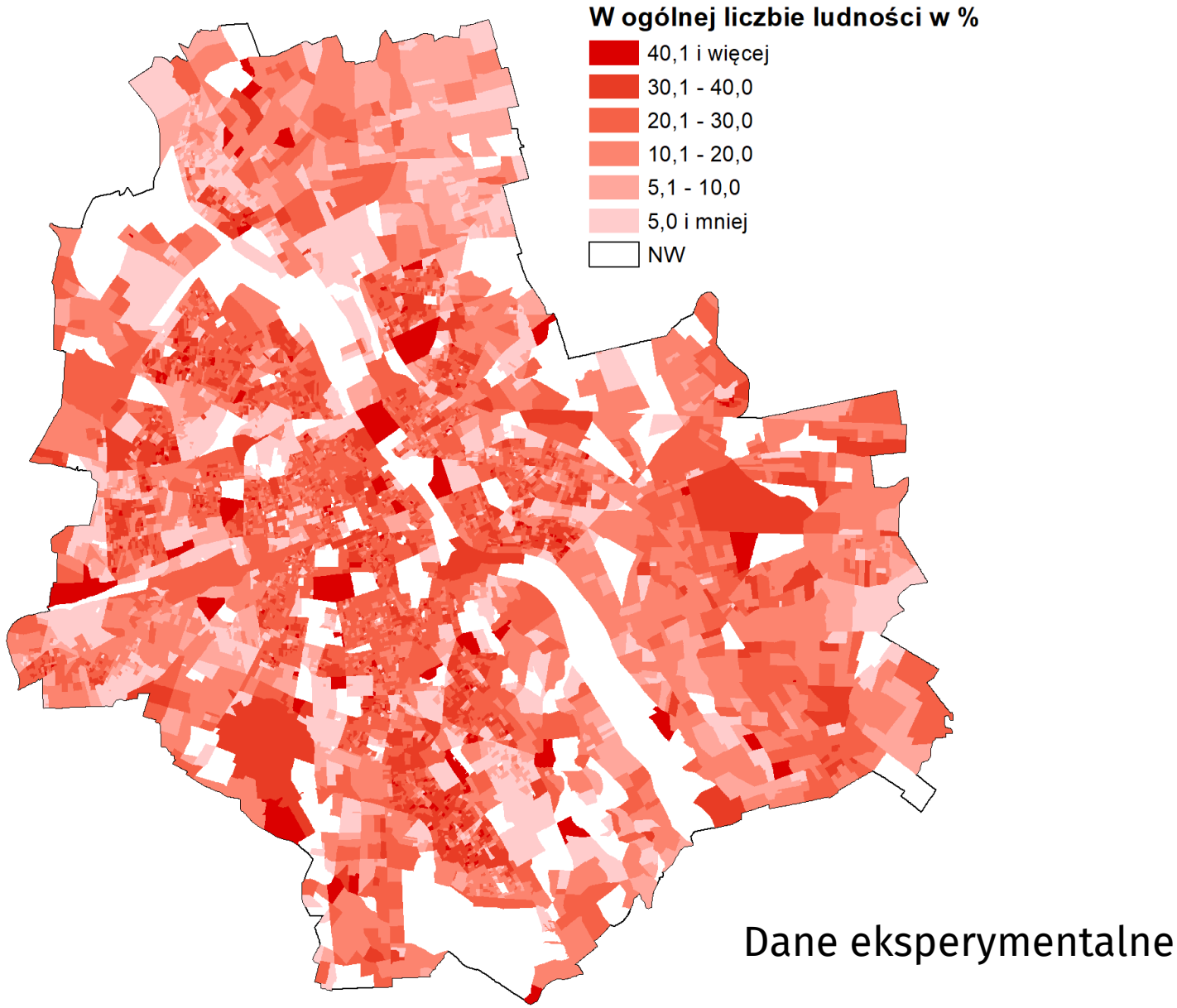
Dane eksperymentalne

Ludność w wieku 0-14 lat



Dane eksperymentalne

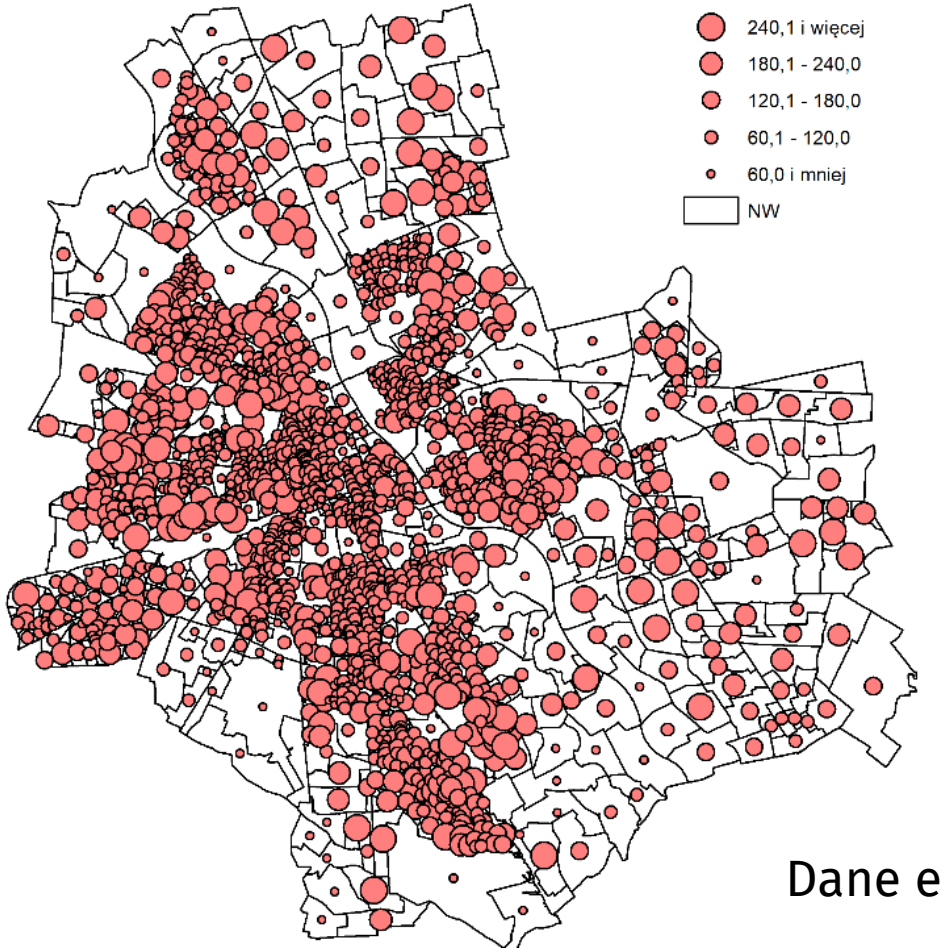
Ludność w wieku 65 lat i więcej



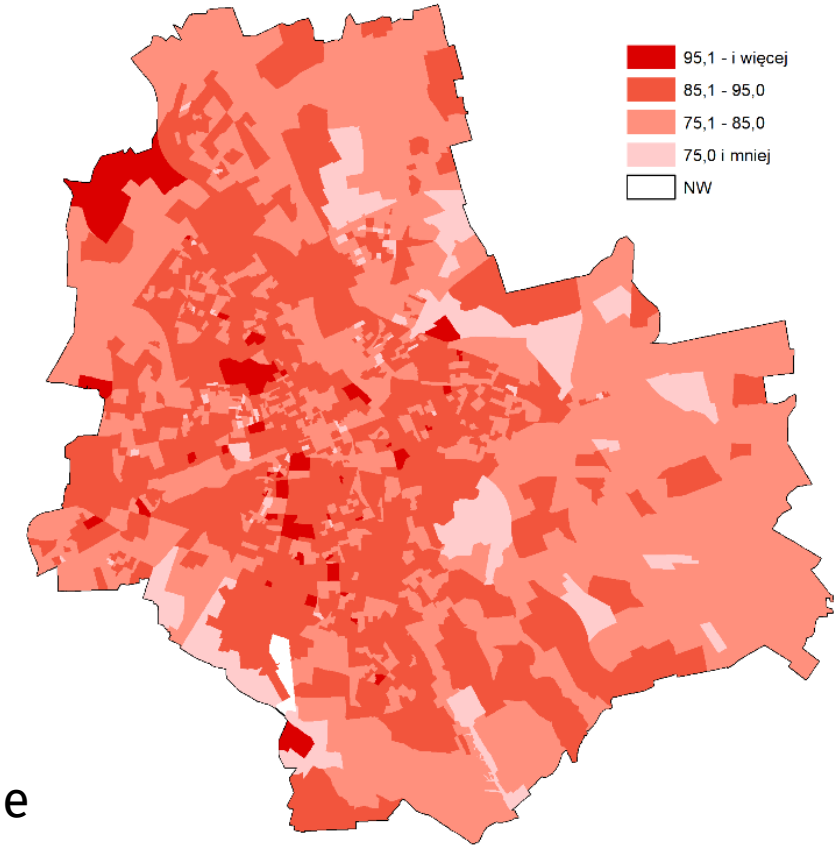
Dane eksperymentalne

Osoby prowadzące działalność gospodarczą

Ogółem



Osoby prowadzące jednoosobową działalność gospodarczą w % ogółem



Dane eksperymentalne

Wnioski

- Uzyskanie porównywalności i spójności podmiotowej w czasie
- Współdzielenie przez pracowników statystyki publicznej wyników – zmiennych dziedzinowych, operatów do badań
- Skrócenie czasu przetwarzania danych, **szybsza prezentacja wyników**, cyfrowy dostęp do danych wynikowych
- Produkcja **szerokiego spektrum danych** z różnych obszarów tematycznych, na niskich poziomach agregacji

Wnioski

- Ograniczenie zakresu pobieranych danych. **Zmniejszenie obciążenia respondentów**
- Zaspokojenie potrzeb informacyjnych użytkowników zewnętrznych poprzez tworzenie złożonych, wielodziedzinowych i wieloaspektowych opracowań statystycznych opisujących w sposób kompleksowy zjawiska społeczno-gospodarcze
- Zwiększenie wykorzystania danych administracyjnych w badaniach eksperymentalnych

Dziękuję za uwagę

Paweł Murawski
p.murawski@stat.gov.pl