

# Geograficzny rozkład błędów w danych gromadzonych w OI UFG

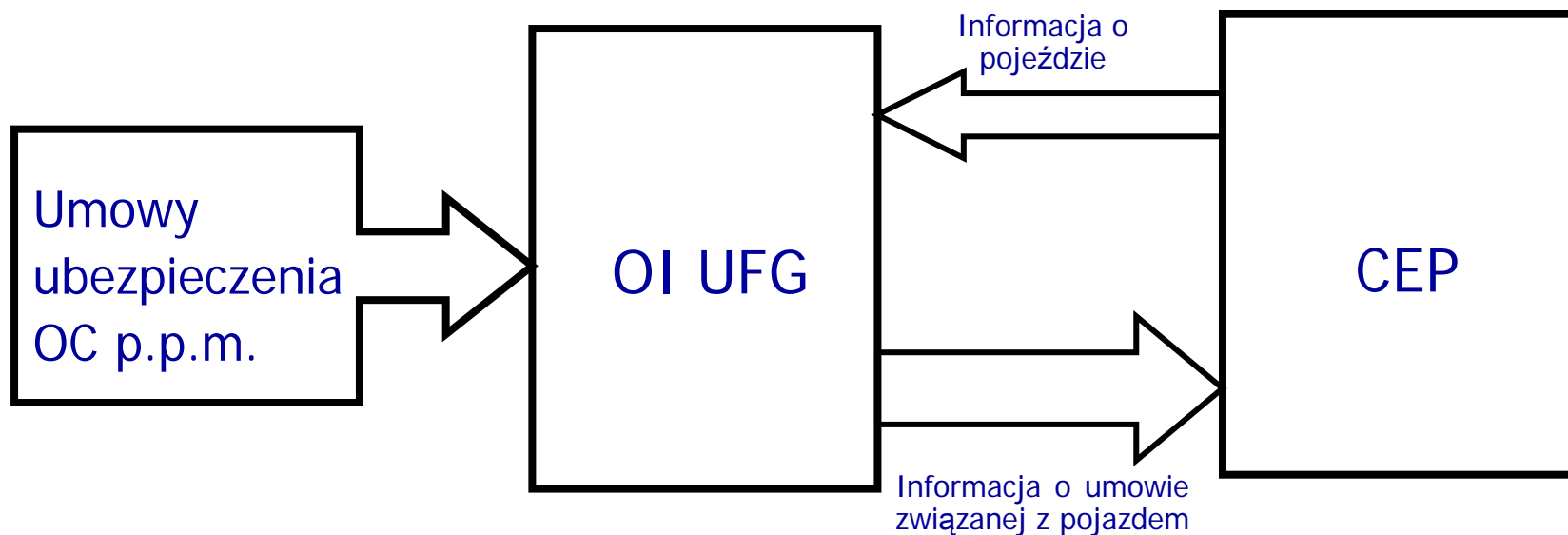
Wojciech Bijak  
Przemysław Czapliński  
Krzysztof Hrycko

Ubezpieczeniowy Fundusz Gwarancyjny

Warszawa, 07 października 2009

- I. Badanie zgodności danych UFG i CEP
- II. Błędy zidentyfikowane w przesłanych pakietach danych
- III. Unikalność numeru VIN i numeru rejestracyjnego
- IV. Mierniki jakości identyfikatorów naturalnych – Dokładność danych
- V. Standaryzacja kategorii „Rodzaj pojazdu”
- VI. Zapytanie przez portal (ZPP) o pokryciu ubezpieczeniowym w zakresie ubezpieczenia OC

## Dowiązkiwania umów z CBD OI UFG do pojazdów z CEP



## I. Identyfikatory brane pod uwagę w algorytmie

- Numer VIN
- Numer rejestracyjny
- Rodzaj pojazdu
- Marka
- Model
- Typ
- Numer silnika

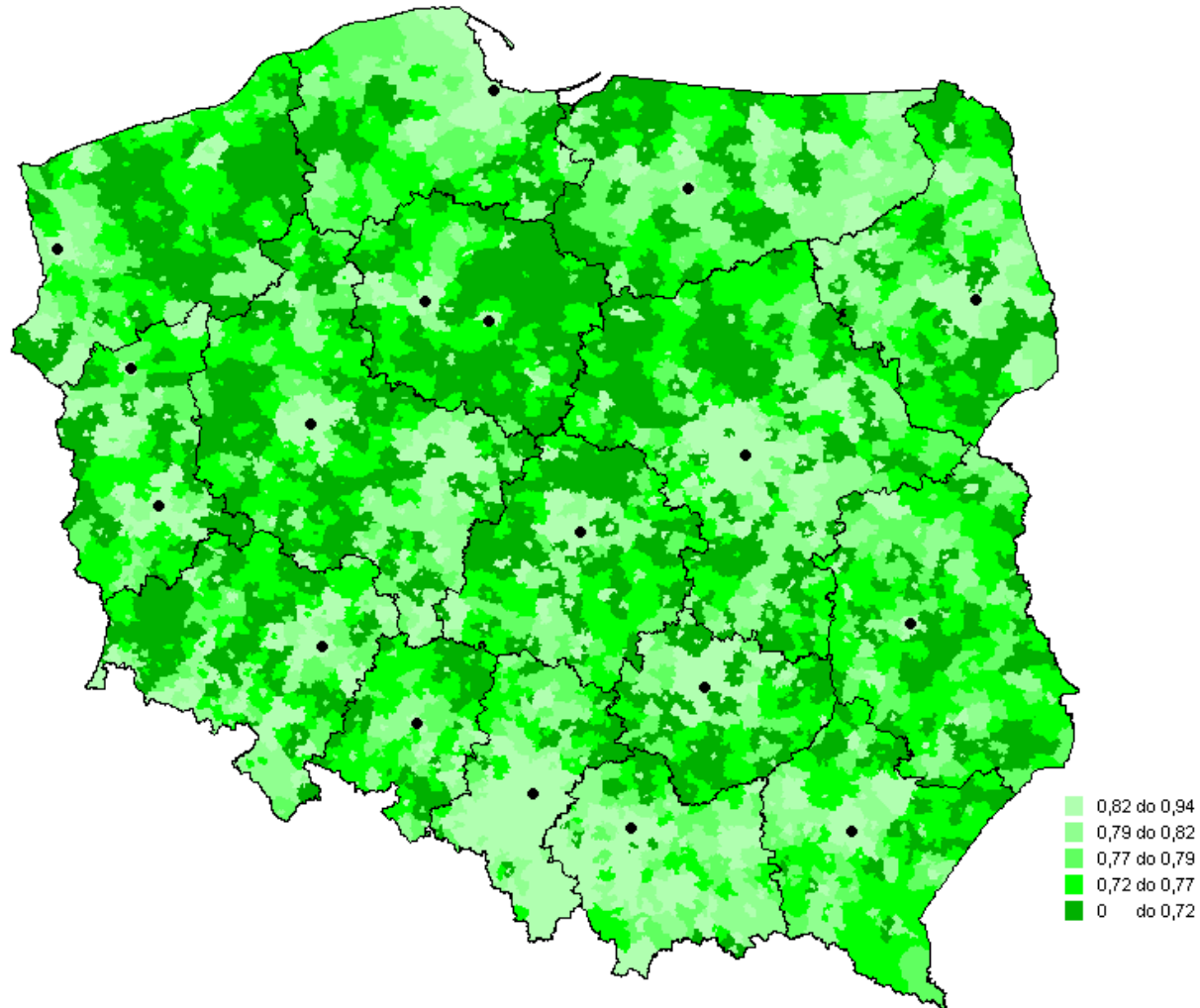
- I. Wynik działania algorytmu – komunikaty zgodne ze stanem obecnym
  - Brak dowiązania
  - Dowiązanie danych OI UFG do danych CEP
- II. Wynik działania zmodyfikowanego algorytmu – nowe komunikaty
  - Brak dowiązania
  - Dowiązanie danych OI UFG do danych CEP
    - $[a_1 a_2 a_3 a_4 a_5 a_6 a_7]$ , gdzie  $a_i = 0$  lub  $1$ , dla  $i=1, \dots, 7$

# I. Zgodność danych przekazanych przez zakłady ubezpieczeń z danymi z CEP

Zakład	Wynik
z 1	94,1%
z 2	89,3%
z 3	89,1%
z 4	87,7%
z 5	86,7%
z 6	86,5%
z 7	85,0%
z 8	84,8%
z 9	84,0%
z 10	83,4%
z 11	82,8%
z 12	81,3%

Zakład	Wynik
z 13	81,0%
z 14	78,1%
z 15	77,4%
z 16	76,5%
z 17	74,6%
z 18	74,2%
z 19	71,3%
z 20	69,9%
z 21	63,7%
z 22	63,2%
z 23	55,2%
z 24	52,8%

# I. Dowiązania pojazdów do CEP w podziale na gminy



## II. Podstawowe błędy zidentyfikowane w przesłanych pakietach danych przez zakłady ubezpieczeń (2009 r)

Błąd syntaktyczny	1 939 237	
UFG-00041	1 454 833	Pojazd wskazany w polisie nie istnieje w danych referencyjnych CEPIK
UFG-00094	1 097 138	W systemie nie istnieje odszkodowanie o identyfikatorze
UFG-00091	517 647	W systemie nie istnieje zdarzenie związane z odszkodowaniem
UFG-00029	499 653	W systemie istnieje już polisa o identyfikatorze
UFG-00030	126 498	W systemie nie istnieje polisa o identyfikatorze
UFG-00086	56 635	W systemie nie udało się odnaleźć polisy AC obowiązujące w dacie zdarzenia dla pojazdu uszkodzonego
UFG-00101	51 856	W bazie już istnieje już zmiana obowiązująca od dnia
UFG-00107	46 214	Nie można zmodyfikować danych polisy, ponieważ podjęto próbę zmiany daty wcześniejszego rozwiązania polisy
UFG-00043	39 108	Aktualizacja polisy która ma już wypełnioną datę wcześniejszego rozwiązania umowy nie jest możliwa



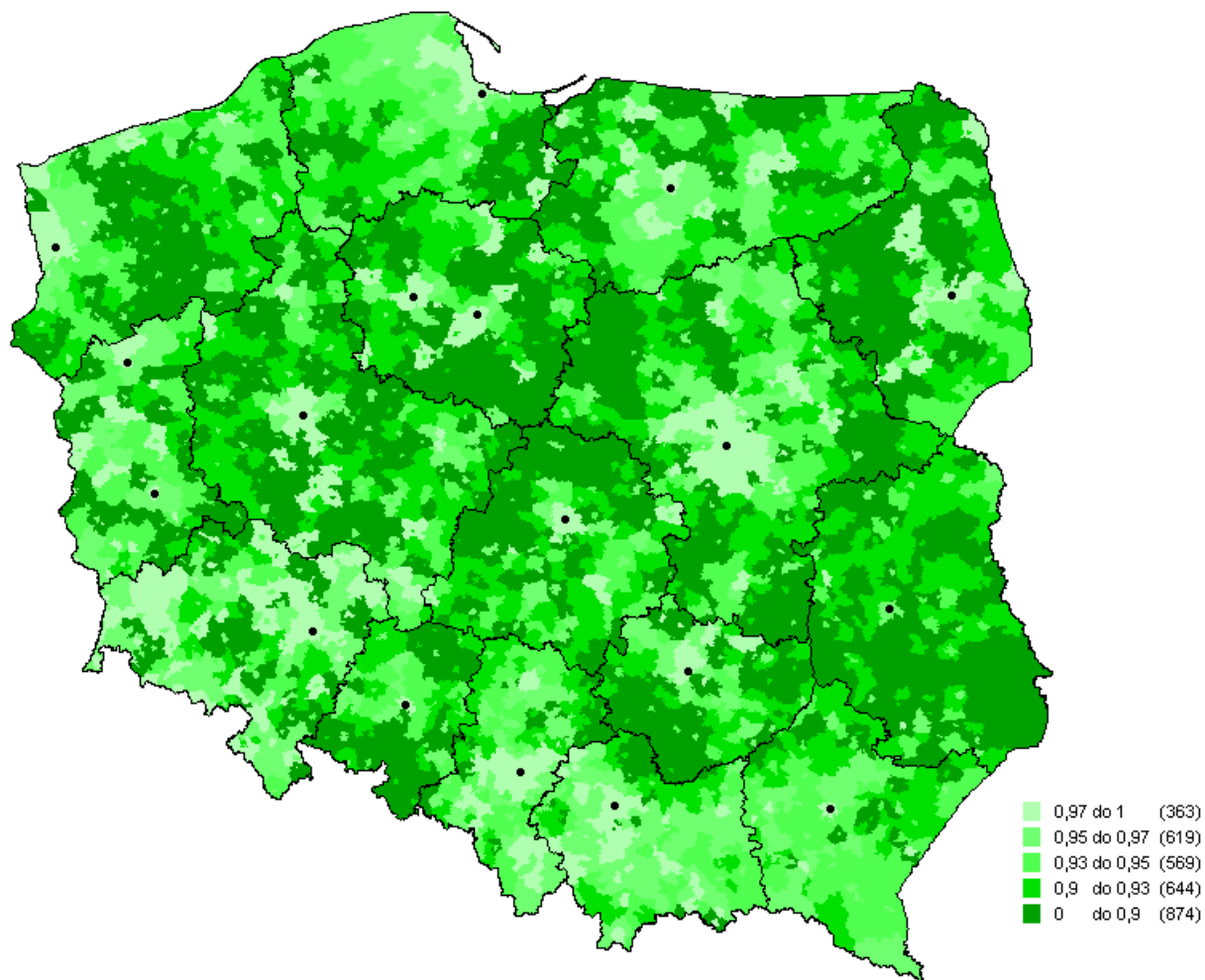
Wartości skompromitowane – wartości pola w rekordzie danych niezgodne z przyjętymi dla danego pola założeniami i standardami



### III. Przykładowe wartości skompromitowane numeru VIN

VIN		ILOSC
BD		238 879
BRAK	7 461	
000000000000000000		6 289
0		6 205
0000000000		1 611
...		1 232
-		1 033
0000	917	
XXXXXXXXXXXXXXXXXXXX		795
1		673
BN		652
X		520
00		511
XXX		476
000		382
111111111111111111		368
00000		307
....	273	

### III. Unikalność nr VIN w podziale na gminy



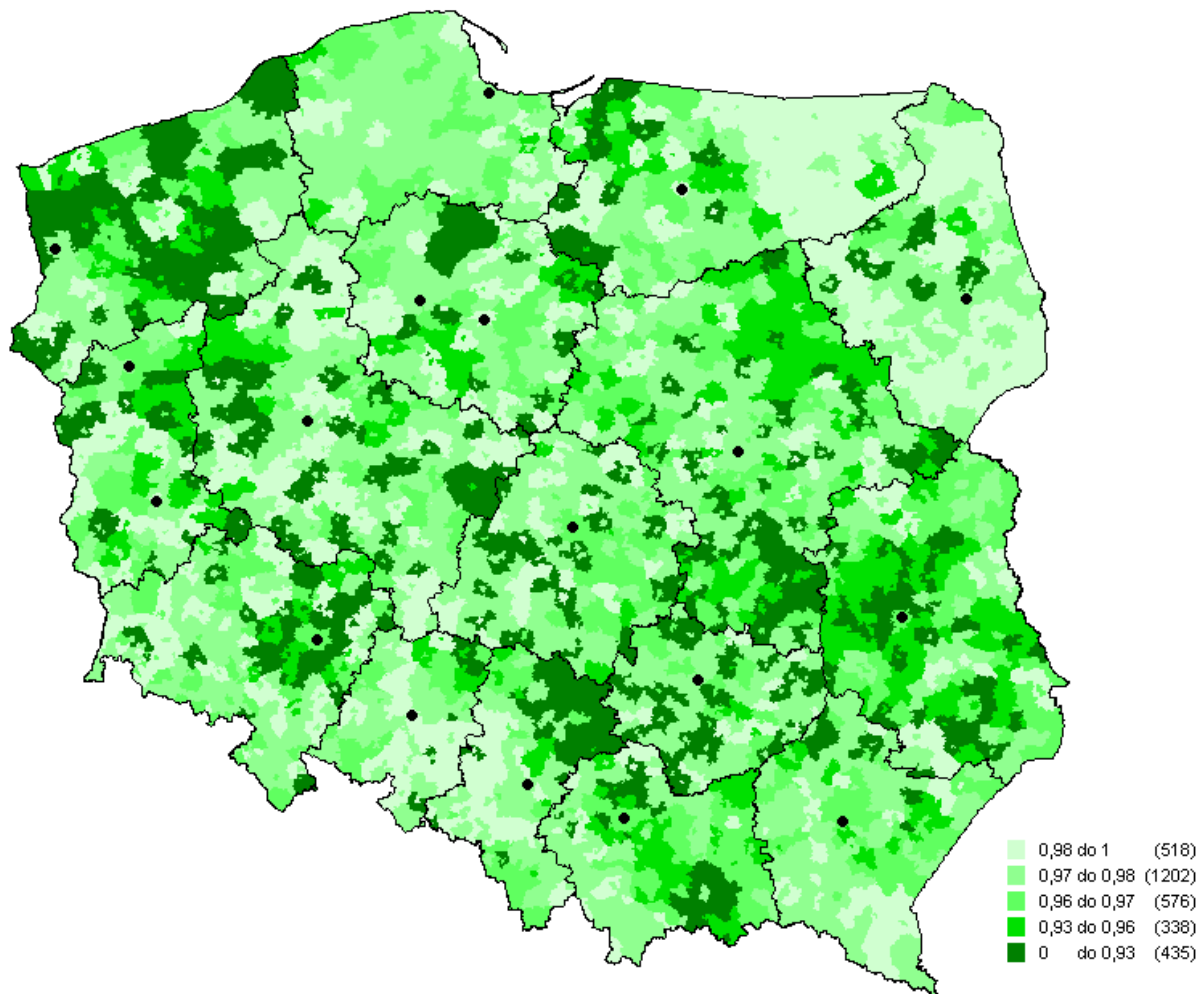
### III. Unikalność nr VIN w podziale na gminy c.d.

•	<b>POWIAT</b>	<b>GMINA</b>	<b>WYNIK</b>
•	lubelski	Głusk	65,2%
•	hrubieszowski	Mircze	77,2%
•	kutnowski	Oporów	77,5%
•	Wielicki	Biskupice	77,9%
•	hrubieszowski	Dołhobyczów	78,6%
•	Łowicki	Zduny	79,2%

...

•	<b>POWIAT</b>	<b>GMINA</b>	<b>WYNIK</b>
•	Wrocław		98,8%
•	Legionowo		98,8%
•	Bydgoszcz		98,8%
•	Szczecin		98,8%
•	Poznań		98,9%
•	Hel		98,9%
•	Nowy Dwór Maz.		99,1%

### III. Unikalność nr rejestracyjnego w podziale na gminy



### III. Unikalność numeru rejestracyjnego w podziale na gminy c.d.

•	<b>POWIAT</b>	<b>GMINA</b>	<b>WYNIK</b>
•	drawski	Kalisz Pomorski	80,31%
•	kozienicki	Gniewoszków	85,49%
•	trzebnicki	Oborniki Śląskie	85,59%
•	Policki	Police	85,72%
•	świecki	Bukowiec	86,22%
•	Świecki	Świekatowo	86,27%

...

•	<b>POWIAT</b>	<b>GMINA</b>	<b>WYNIK</b>
•	Giżycki	Kruklanki	99,07%
•	Giżycki	Wydminy	99,11%
•	leski	Cisna	99,13%
•	hajnowski	Czyże	99,14%
•	ełcki	Stare Juchy	99,18%



Dokładność – dane można uznać za dokładne, gdy są wolne od istotnych pomyłek, błędów i braków

Dokładność powinna być analizowana z punktu widzenia celu wykorzystania danych

Np. w przypadku raportów dotyczących ustalania pokrycia ubezpieczeniowego w zakresie AC lub OC posiadaczy pojazdów mechanicznych na określony dzień istotna jest dokładność identyfikatorów naturalnych

## Zbiory wariantów identyfikatorów

**Tabela 1.** Rozkłady identyfikatorów naturalnych związanych z umowami ubezpieczenia AC aktywnymi w dniu 1 lutego 2009 r. (fragment)

Numer rejestracyjny		Numer VIN		PESEL	
$n_i$	$M_{n_i}$	$n_i$	$M_{n_i}$	$n_i$	$M_{n_i}$
1	3 635 024	1	3 593 593	1	2 342 475
2	65 814	2	89 934	2	307 481
3	3 076	3	4 184	3	42 922
...	...	...	...	...	...
860	1	497	1	174	1
940	1	601	1	536	1
6 119	1	611	1	14 823	1

Źródło: Opracowanie własne.

**Tabela 2.** Rozkłady identyfikatorów naturalnych związanych z umowami ubezpieczenia OC p.p.m. aktywnymi w dniu 1 lutego 2009 r. (fragment)

Numer rejestracyjny		Numer VIN		PESEL	
$n_i$	$M_{n_i}$	$n_i$	$M_{n_i}$	$n_i$	$M_{n_i}$
1	13 861 866	1	13 060 051	1	6 423 362
2	772 198	2	887 064	2	2 188 244
3	52 957	3	106 279	3	718 782
...	...	...	...	...	...
7 089	1	6 578	1	2 531	1
7 716	1	7 779	1	3 535	1
54 832	1	185 496	1	123 548	1

Źródło: Opracowanie własne.



## I. Wybrane miary dokładności identyfikatorów naturalnych

$$p_1 = \frac{M_1}{N}, \quad H(p) = \sum_{i=1}^J M_i \frac{n_i}{N} \log_2 \left( \frac{n_i}{N} \right)$$

$$V(q) = \frac{\sqrt{SV(q)}}{q}, \quad H = \sum_{i=1}^N \frac{1}{N} \log_2 \left( \frac{1}{N} \right)$$

$$SV(q) = \frac{1}{M} \sum_{j=1}^J (q_j - q)^2, \quad h = \frac{H(p)}{H}$$

## IV. Dokładność danych c.d.

Wartości wybranych miar dokładności identyfikatorów naturalnych związanych z umowami ubezpieczenia OC p.p.m.

Miara	Numer rejestracyjny		Numer VIN		PESEL	
	01-02-2009	01-05-2009	01-02-2009	01-05-2009	01-02-2009	01-05-2009
$p_1$	88,091%	90,063%	82,996%	84,996%	40,891%	43,242%
$q$	6,3549E-08	6,72869E-08	6,3549E-08	6,7287E-08	6.3660E-8	6.8195E-8
$SV(q)$	0,003567	0,003752	0,011813	0,011568	0,007875	0.007591
$V(q)$	56136	55759	185889	171926	123701	111321
$H(p)$	23,908	23,825	23,908	23,825	23.905	23.806
$h$	99,115%	99,198%	98,151%	98,277%	95,579%	95,800%

## IV. Dokładność danych c.d.

Wartości wybranych miar dokładności identyfikatorów naturalnych związanych z umowami ubezpieczenia AC

Miara	Numer rejestracyjny		Numer VIN		PESEL	
	01-02-2009	01-05-2009	01-02-2009	01-05-2009	01-02-2009	01-05-2009
$p_1$	95,888%	95,925%	94,795%	94,832%	73,306%	74,730%
$q$	2,6379E-07	2,77167E-07	2,6379E-07	2,7717E-07	3.12944E-7	3.4159E-7
$SV(q)$	0,001656	0,001735	0,00029	0,00024	0,004654	0.004629
$V(q)$	6277	6259	1115	856	14870	13551
$H(p)$	21,854	21,783	21,854	21,783	21,608	21.481
$h$	99,675%	99,677%	99,727%	99,732%	98,145%	98,209%

Wartości wybranych miar dokładności identyfikatorów naturalnych; REGON

- REGON ILOSC
- 930986308 4323 - EUROPEJSKI FUNDUSZ LEASINGOWY S.A.
- 012305049 2378 - VOLKSWAGEN LEASING POLSKA SP. Z O.O.
- 012960825 2065 - RAIFFEISEN LEASING POLSKA
- 016298263 1634 - PKO BP S.A.
- 003538527 1522 - GBG S.A/GETIN BANK SA O. W BEDZINIE
- 012015417 1456 - BEL LEASING SPOLKA ZOO
- 012527809 1437 - BRE LEASING SP. Z O.O.
- 012162265 1126 - BZ WBK FINANSE & LEASING SA
- 012213933 1021 - DAIMLER CHRYSLER SERVICES LEASING

$$H(p) = 20,685$$

$$h = 77,83$$

Klasyfikacja pojazdów w kategoriach „Rodzaj pojazdu”, „Podrodzaj pojazdu” wprowadzona jest rozporządzeniem Ministra Infrastruktury z dnia 27 września 2003 r. w sprawie szczegółowych czynności organów w sprawach związanych z dopuszczeniem pojazdu do ruchu oraz wzorów dokumentów w tych sprawach (Tabela nr.1 rozporządzenia)

### ▪ Rodzaj pojazdu

klasyfikacja obejmuje 16 kategorii pojazdów. Każda kategoria zawiera nazwę oraz kod „Rodzaju pojazdu” np. kod 03 – samochód osobowy, kod 04 – autobus, itd.

### ▪ Podrodzaj pojazdu

klasyfikacja obejmuje 60 kategorii podzielonych na grupy z których każda powiązana jest z określonym rodzajem pojazdu np.;

rodzaj	06- samochód ciężarowy
podrodzaj	01- skrzynia
	02 – furgon
	03 – wywrotka
	05 - cysterna

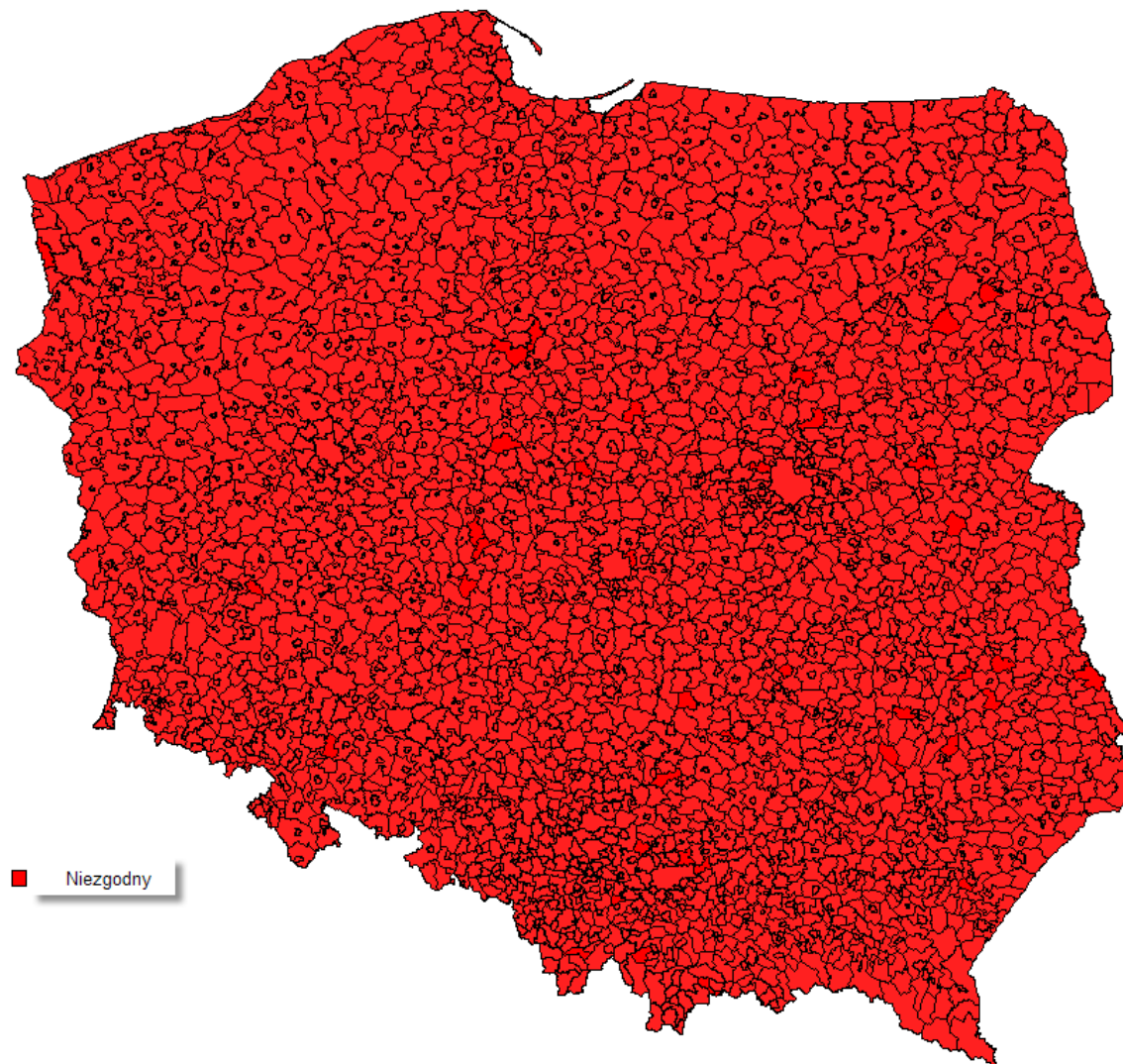
# V. Rodzaje pojazdów zgodnie z informacjami podawanymi publicznie przez niektóre zakłady ubezpieczeń



- Samochody Dostawcze
- Samochody Terenowe
- Samochody Osobowe
  
- Osobowy
- Terenowy
- Dostawczy
- Osobowo-Dostawy
  
- Osobowy
- Inne
- Składak
- Homologacja - ciężarowy na podwoziu osobowego
  
- Osobowy
- Terenowy



# V. Zgodność kategorii „Rodzaj pojazdu” z obowiązującym standardem w układzie geograficznym na podstawie danych wybranych zakładów



## VI. Zapytanie przez portal (ZPP) o pokryciu ubezpieczeniowym w zakresie ubezpieczenia OC

### □ Udostępnianie danych Zakładom Ubezpieczeń

- Portalowy dostęp do danych (identyfikacja polisy OC na dzień na stronie [www.ufg.pl](http://www.ufg.pl), odpowiedź wysyłana na wskazany adres e-mail).  
Efekt: eliminacja operatora (wąskiego gardła) w procedurze pozyskiwania informacji  
Zysk: skrócenie czasu oczekiwania na odpowiedź z 3-4 tygodni do 1-2 minut + redukcja kosztów – **realizacja 2009-06-24**



The screenshot displays the website of the Ubezpieczeniowy Fundusz Gwarancyjny (UFG). At the top, there is a banner with the UFG logo and the text 'UBEZPIECZENIOWY FUNDUSZ GWARANCYJNY'. Below the banner is a navigation menu with links: 'Strona Główna', 'O UFG', 'Akty prawne', 'Archiwum', 'Publikacje', 'Kontakt', and 'ENGLISH'. The main content area is titled 'Aktualności' and features a news item dated 23 września 2009 r. titled 'Seminarium Ośrodka Informacji 23 września 2009 r.'. A red box with white text 'Mechanizm Zapytania przez Portal "ZPP"' is overlaid on the text, with a red arrow pointing to a smaller image of a document titled 'Zapytania do Ośrodka Informacji UFG'. To the right of the news item is a login form with fields for 'Login:' and 'Hasło:', a 'Zaloguj' button, and a search bar with a 'Szukaj' button.





- I. Uruchomienie pilotażu dla 2 zakładów i PBUK - lipiec 2009
- II. Portalowy dostęp do danych dla wszystkich zakładów - sierpień 2009. Wyniki po miesiącu użytkowania systemu
- III. Publiczne uruchomienie mechanizmu - wrzesień 2009



# VI. Zapytanie przez portal (ZPP) o pokryciu ubezpieczeniowym w zakresie ubezpieczenia OC

**Identyfikacja polisy OC na dzień**

Prosimy o wypełnienie poniższych pól:

Numer rejestracyjny pojazdu \*

Numer VIN \*

Na dzień \*\*

E-mail \*\*

Przepisz kod z obrazka:

Oświadczam, iż jestem osobą uprawnioną do uzyskania informacji w oparciu o art. 104 Ustawy o ubezpieczeniach obowiązkowych, UFG i PBUK.

*\* Przynajmniej jedno z pól musi być wypełnione  
\*\* Pola zaznaczone kolorem żółtym są wymagane*

**UBEZPIECZENIOWY FUNDUSZ GWARANCYJNY**

ul. Piłska 9/11, 01-231 Warszawa, tel.: +48 (22) 631 12 30, +48 (22) 631 60 03,  
fax: +48 (22) 631 16 81, +48 (22) 631 55 73, e-mail: ufg@ufg.pl

Warszawa 10-06-2009      Znak sprawy - 02965/WIU/06/2009

**Portal Publiczny**  
pczaplinski@ufg.pl

Dot. zapytania nr 09-051414-0001 z dnia 10-06-2009.

**Zapytanie**

Cel zapytania: **Identyfikacja polisy OC w portalu publicznym**  
Treść zapytania: **Czy pojazd o numerze rejestracyjnym "WZ35520" posiadał ubezpieczenie OC w ruchu krajowym na dzień 10-06-2009?**

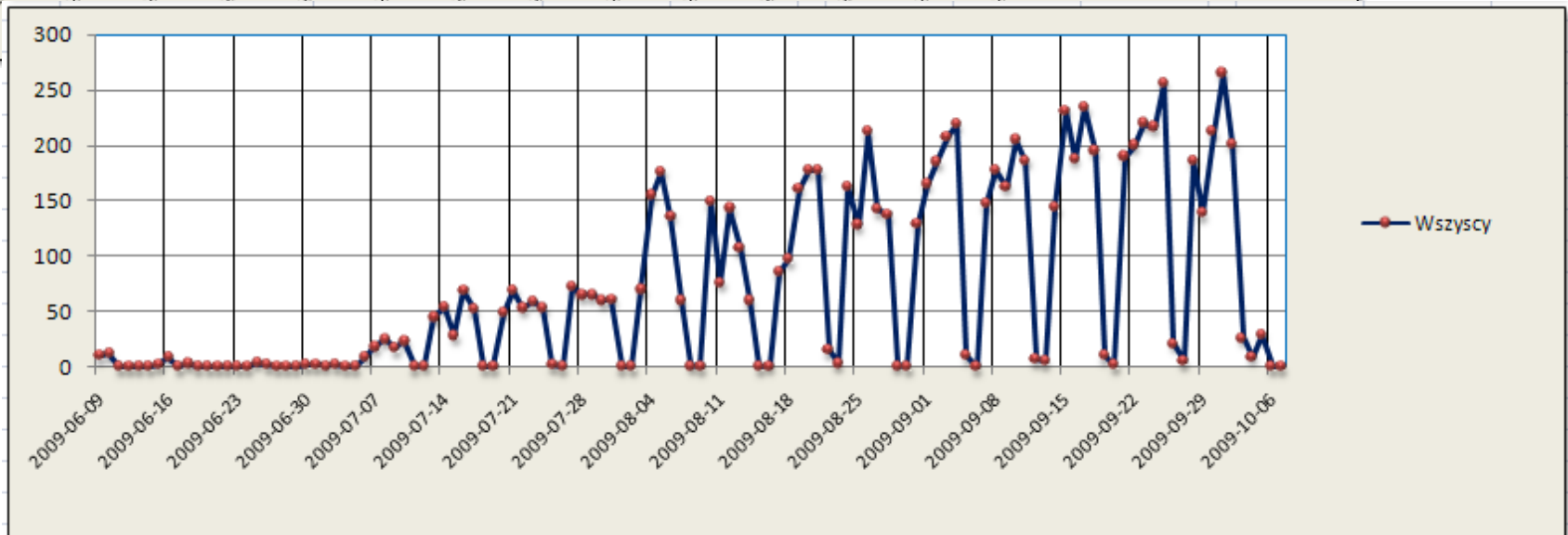
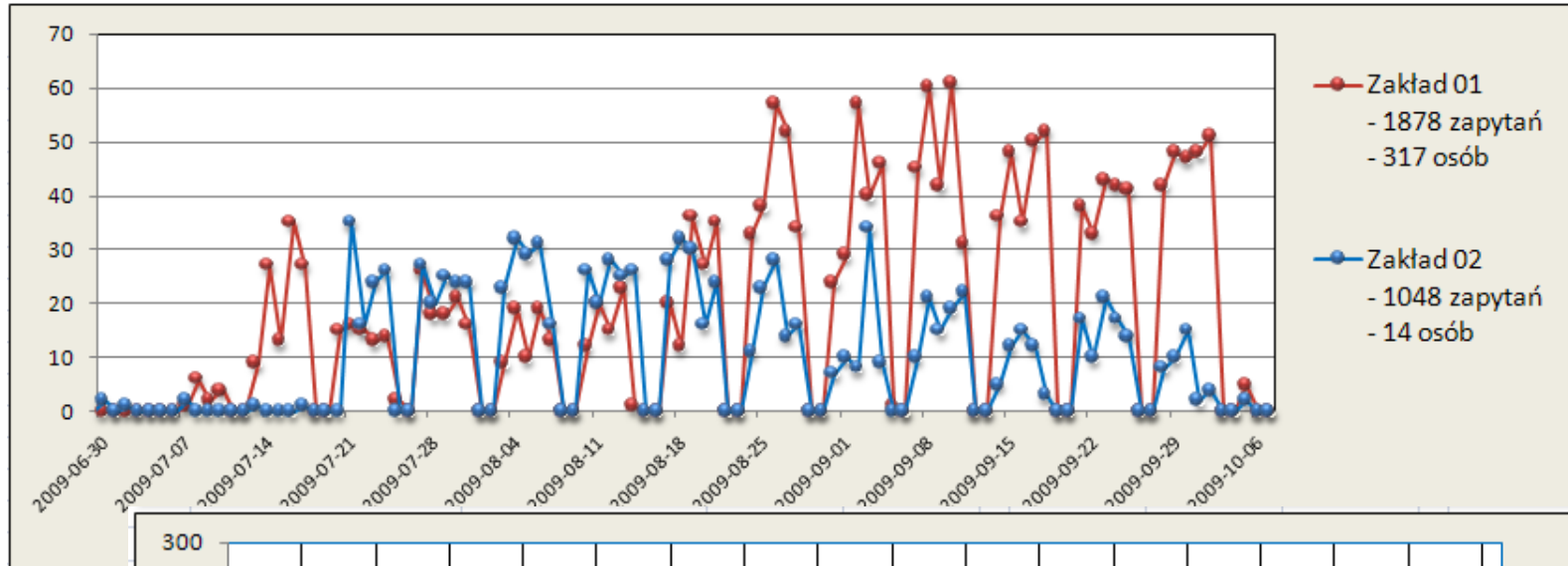
**Pojazd o podanym numerze rejestracyjnym/numerze VIN posiadał w podanym dniu następujące ubezpieczenie OC w ruchu krajowym**

**1. polisa RFC~03939500~LKRFA021 wystawiona przez "POWSZECHNY ZAKŁAD UBEZPIECZEŃ S.A."**

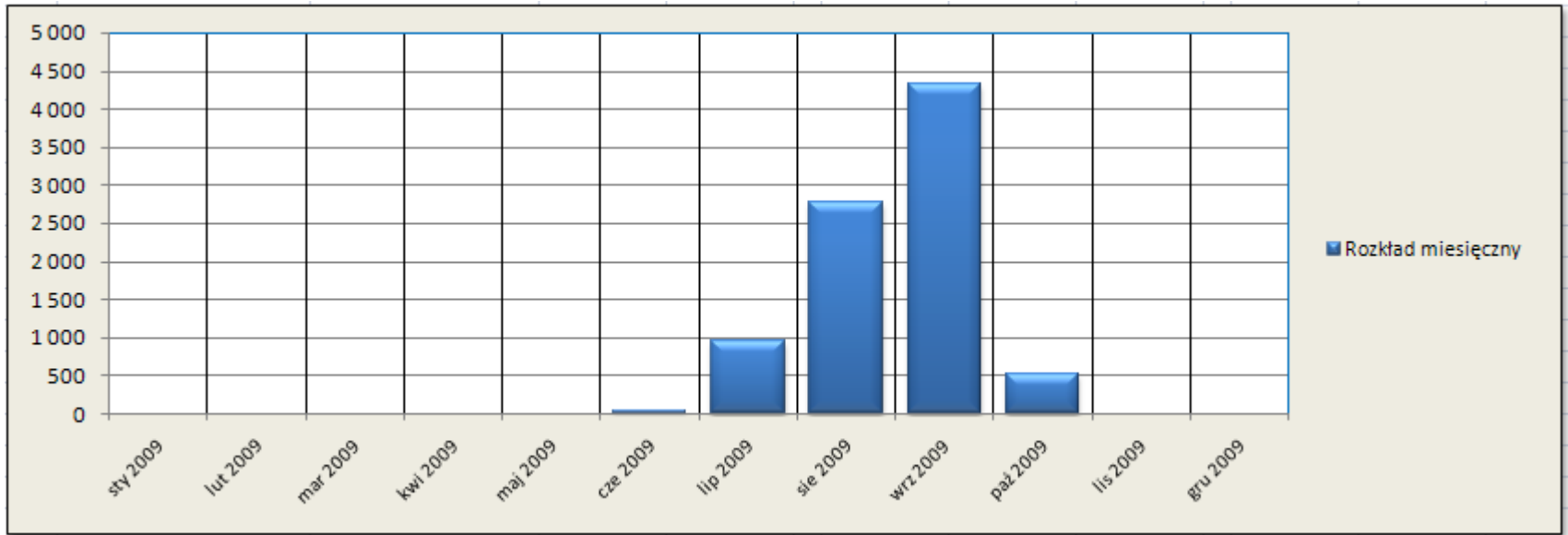
NIP 526-10-51-849      Konto: BPH PBK SA IX O/Warszawa 69 1060 2621 0000 4010 4020 3820      STRONA 1 z 1



# VI. Zapytanie przez portal (ZPP) o pokryciu ubezpieczeniowym w zakresie ubezpieczenia OC



# VI. Zapytanie przez portal (ZPP) o pokryciu ubezpieczeniowym w zakresie ubezpieczenia OC

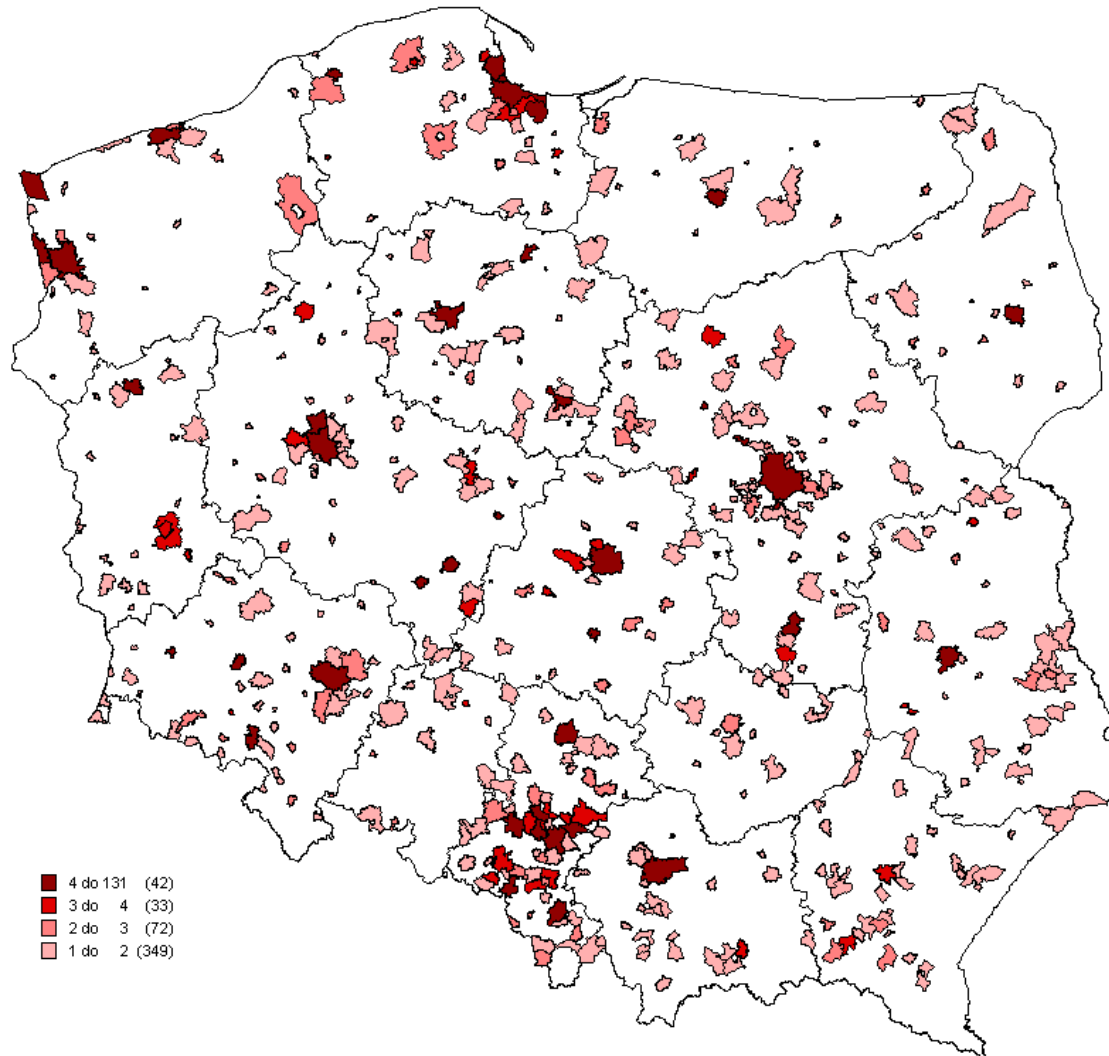


Czasy generowania odpowiedzi [s]	
Min	0
Max	399
Me	35
Sr	42

Znaleziono odpowiedź	1 148
Brak danych o OC	1 029

Liczba OC	17 000 000
Średnia szkodowość	5,00%
Ilość szkód OC	850 000
Dni roboczych	240
Zapytań dziennie	3 542
1 dzień	23:36:40
Czas zapytania	00:00:24

# VI. Zapytanie przez portal (ZPP) o pokryciu ubezpieczeniowym w zakresie ubezpieczenia OC



Dziękujemy za uwagę

- [wbijak@ufg.pl](mailto:wbijak@ufg.pl)
- [khrycko@ufg.pl](mailto:khrycko@ufg.pl)
- [pczaplinski@ufg.pl](mailto:pczaplinski@ufg.pl)