



Klasyfikacja danych w systemach CEP i CEK

dr inż. **Andrzej Szyszko**
Główny specjalista DEPiT MSWiA

AGENDA

- Klasyfikacja - co to jest i do czego służy?
- Odpowiedzialność administratora danych za zgromadzone dane
- Podział klasyfikacji wg. organizacji
- Źródła klasyfikacji
- Automatyczne wyznaczanie klasyfikacji z możliwością ręcznej zmiany
- Miary informacyjne
- Wymagania dla klasyfikacji



Klasyfikacja



Definicja:

informacja dodatkowa o jakości danych

Dotyczy ona uzupełnienia gromadzonych udostępnianych danych o dodatkowe informacje pozwalające na ocenę jakości otrzymanych danych przez uprawniony podmiot.

W zależności od potrzeb podmiotów, klasyfikacja będzie miała różną ogólność oraz będzie dotyczyła różnego zakresu danych.



Odpowiedzialność administratora danych za zgromadzone dane



grzywna, kara ograniczenia albo pozbawienia wolności do lat 3

- przetwarzanie danych wrażliwych bez prawa przetwarzania tych danych - art. 49 ust. 2

grzywna, kara ograniczenia albo pozbawienia wolności do lat 2

- przetwarzanie danych osobowych w zbiorze bez prawa przetwarzania tych danych - art. 49 ust. 1
- udostępnianie danych osobowych lub umożliwianie dostępu do nich osobom nieupoważnionym - art. 51 ust. 1

grzywna, kara ograniczenia albo pozbawienia wolności do roku

- przetwarzanie danych osobowych w zbiorze niezgodnie z celem utworzenia zbioru - art. 50
- naruszenie choćby nieumyślne obowiązku zabezpieczenia danych osobowych przed zabraniem przez osobę nieuprawnioną, uszkodzeniem lub zniszczeniem - art. 52



Odpowiedzialność administratora danych za zgromadzone dane



Decyzje Rady UE 2008/615/WSiSW oraz 2008/616/WSiSW:

Jeżeli podmiot jednego państwa członkowskiego dostarczył dane osobowe na mocy niniejszej decyzji, podmiot otrzymujący innego państwa członkowskiego nie może wykorzystać nieścisłości dostarczonych danych jako podstawy do uchylenia się od odpowiedzialności wobec poszkodowanego wynikającej z prawa krajowego. Jeżeli orzeczono wypłatę odszkodowania przez podmiot otrzymujący w związku z wykorzystaniem przez niego przekazanych nieścisłych danych, podmiot dostarczający dane zwraca podmiotowi otrzymującemu całą kwotę wypłaconego odszkodowania.



Odpowiedzialność administratora danych za zgromadzone dane



Traktat Eucaris

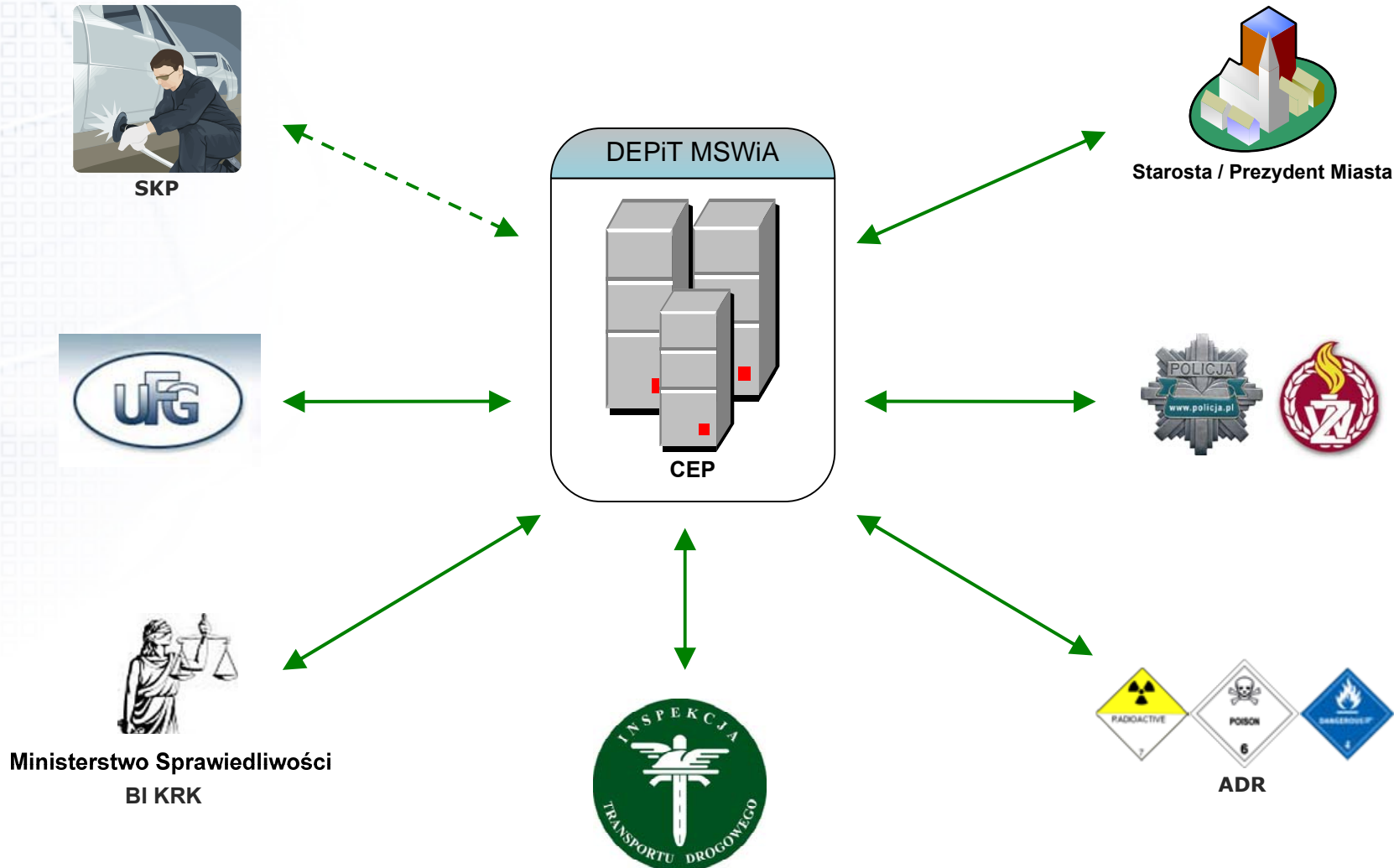
- dane poprawne?
- czy wymagana jest konieczność udzielenia odpowiedzi?
- czy zakres udzielenia informacji jest zgodny z warunkami zapytania?
- ochrona danych osobowych określona przepisami krajowymi,
- jeżeli oczywistym jest, że udostępniona informacja jest niepoprawna lub została przekazana informacja, która nie powinna być przekazana, administrator powinien o tym fakcie niezwłocznie poinformować podmiot, który otrzymał informacje,
- administrator, który otrzymał tę wiadomość powinien skasować otrzymane informacje lub je skorygować.



Odpowiedzialność administratora danych za zgromadzone dane



Źródła zasilania





Podział klasyfikacji wg. organizacji



WEWNĘTRZNA

- Szczegółowa klasyfikacja na potrzeby udostępniania danych,
- Możliwość uzyskiwania bardzo szczegółowej informacji o elementach klasyfikacji,
- Możliwość dwupoziomowego widoku klasyfikacji: skondensowanej i szczegółowej na żądanie,
- Możliwość sprawdzenia: kompletność danych, wyniki identyfikacji i weryfikacji danych przychodzących oraz innych atrybutów.



Podział klasyfikacji wg. organizacji



ZEWNĘTRZNA

- Jest uzależniona od podmiotu, od którego uzyskujemy dane,
- Będzie uzgadniana z każdym podmiotem,
- Będą to informacje zagregowane dla części procesów. Dla niektórych procesów mogą być szczegółowe,
- Problem odpowiedzialności podmiotów korzystających z ewidencji.



Źródła klasyfikacji



W SI CEPiK podczas zasilania następuje tzw. weryfikacja danych:

- poprawności formalnej przekazywanych informacji,
- niezaprzeczalności,
- kompletność danych,
- zgodności z rejestrami referencyjnymi,
- zgodność z poprzednim stanem w ewidencji,
- zgodność z warunkami formalnymi dla niektórych atrybutów.

Algorytm

Wynikiem ww. sprawdzeń jest tworzenie tzw. *wektora weryfikacji*, który zawiera informacje z przebiegu procesu zasilania dla każdego komunikatu przychodzącego do SI CEPiK.



Źródła klasyfikacji



	Źródło danych					Rodzaj weryfikacji dla atrybutu					
	Policja		UFG	SKP	Pojazd	A	B	C	D	E	F
Badane atrybuty	Pojazd	Dokument									
1. Marka	+	+	+	+	+	+	+	+	-	-	+
2. Model	+	+	+	+	+	+	+	+	-	-	+
3. Rodzaj	+	+	+	+	+	+	+	+	-	-	+
4. Rok produkcji	+	+	+	+	+	+	+	+	-	-	+
5. Numer rejestracyjny	+	+	+	+	+	+	+	+	-	-	+
6. Nr dowodu rejestracyjnego/pozwolenia czasowego	-	+	+	+	+	+	-	+	-	-	+
7. VIN	+	+	+	+	+	+	+	+	-	-	+
8. Numer silnika	+	+	+	+	+	-	-	+	-	-	+
9. PESEL	-	-	-	-	+	+	-	+	-	+	-
10. REGON	-	-	-	-	+	+	-	+	-	+	-
11. Imię i nazwisko	-	-	-	-	+	+	+	-	-	+	-
12. Nazwa	-	-	-	-	+	+	-	-	-	-	-
13. Data pierwszej rejestracji w kraju	-	-	-	-	+	-	+	-	-	-	+
14. Data pierwszej rejestracji	-	-	-	-	+	-	+	-	-	-	+
15. Data wpisania do ewidencji	-	-	-	-	+	-	+	-	-	-	-
16. Liczba miejsc siedzących	-	-	-	-	+	-	+	-	-	-	-



Źródła klasyfikacji



Rodzaj weryfikacji dla źródła danych					
A. Sprawdzenie kompletności	+	+	+	+	+
B. Poprawność formalna	+	+	+	+	+
C. Unikalność	-	-	-	-	+
D. Zgodność z ITS	-	-	-	-	+
E. Zgodność z rejestrem PESEL/REGON	-	-	-	-	+
F. Zgodność z alternatywnymi źródłami danych (porównanie dotyczy zgodności z danymi z SI Pojazd)	+	+	+	+	+



Źródła klasyfikacji



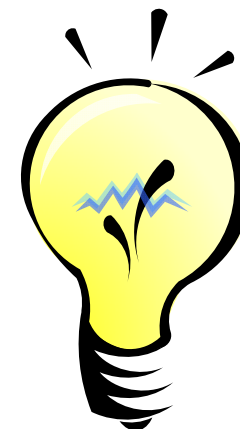
- Niezgodności w danych obiektu,
- Na podstawie wewnętrznych mechanizmów kontroli jakości w centralnych ewidencjach, tzw. demonów, wykrywane są niezgodności dla obiektów,
- Informacje o tych niezgodnościach stanowią element klasyfikacji jakości danych (większa liczba niezgodności powoduje zmniejszenie wiarygodności).



Automatyczne wyznaczanie klasyfikacji



- Sposób automatyczny, w oparciu o reguły i algorytmy,
- Parametryzacja i wersjonowanie reguł i algorytmów,
- Wykorzystanie wiedzy dla różnych działań





Miary informacyjne



- Uzupełnieniem klasyfikacji danych dla obiektów czy atrybutów są miary informacyjne dot. całej ewidencji,
- Obejmują one m.in. opóźnienia w zasilaniu ewidencji.





Wymagania dla klasyfikacji



- Przepisy międzynarodowe wymagają aby dane, które są udostępniane, miały sklasyfikowany poziom wiarygodności i jakości,
- Dla Decyzji Rady UE 2008/615/WSiSW (Prüm), będzie wymagana także informacja o jakości, wiarygodności, poprawności udostępnianych informacji (w ramach systemu EUCARIS), nawet na poziomie atrybutów,
- Decyzja Rady UE 2006/960/WSiSW w sprawie współpracy policyjnej, wymaga określenia poziomu wiarygodności danych.



DZIĘKUJĘ ZA UWAGĘ

andrzej.szyszko@poczta.mswia.gov.pl

tel. +22 60 28 409