

Analiza przyczyn wypadków drogowych

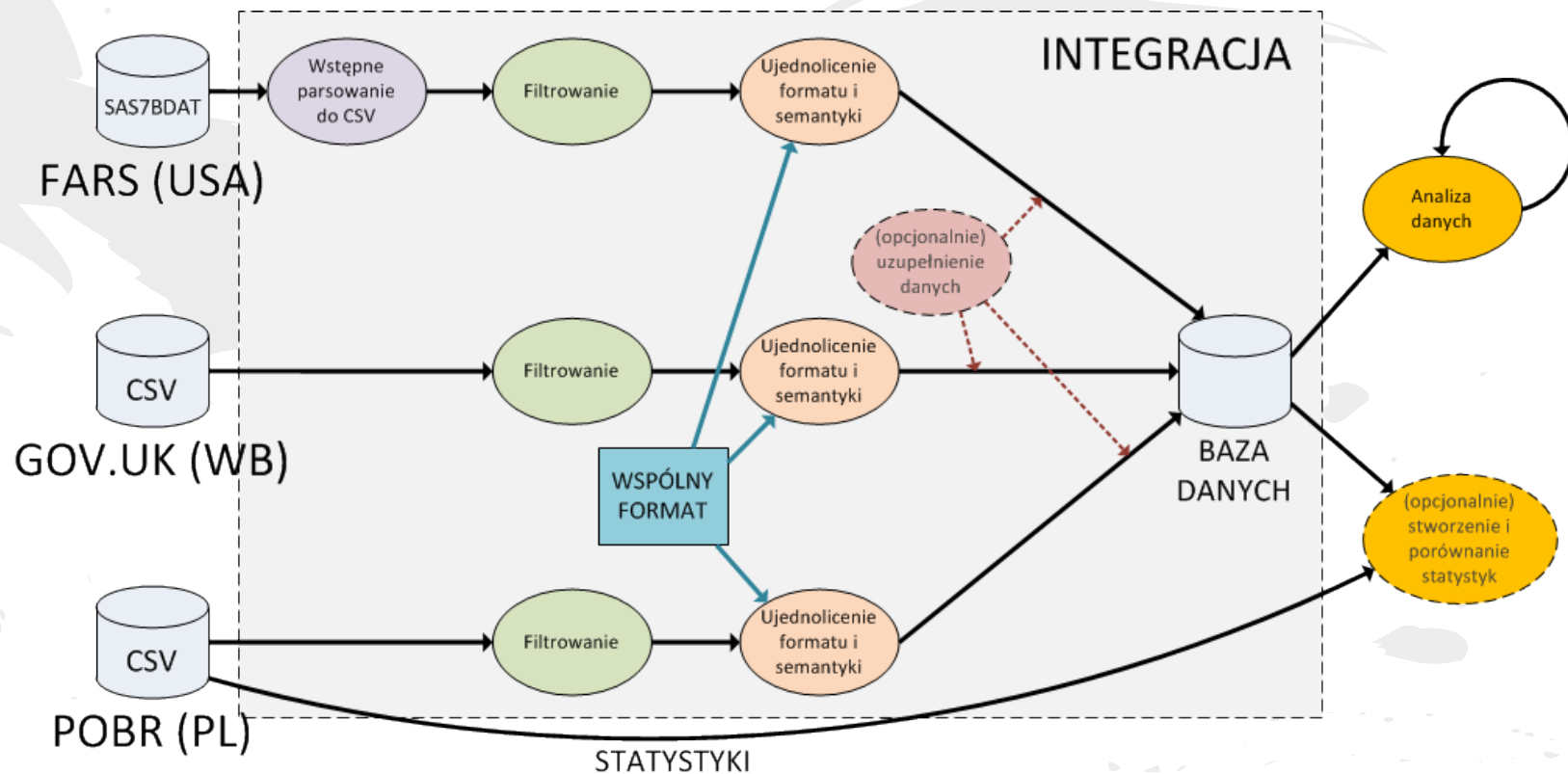
Łukasz Opióła
Beata Skiba

Temat i cel projektu

- ❑ Integracja danych o wypadkach drogowych pochodzących z różnych źródeł
- ❑ Analiza danych pod kątem przyczyn wypadków



Metodyka



Źródła danych

❑ Polska

- ❑ POBR - Polskie Obserwatorium Ruchu Drogowego, dane z policyjnego systemu SEWiK



❑ Wielka Brytania

- ❑ Dane z formularzy STATS19



GOV.UK

❑ USA

- ❑ NHTSA: FARS (Fatality Analysis Reporting System)
Wypadki śmiertelne z terenów całego USA



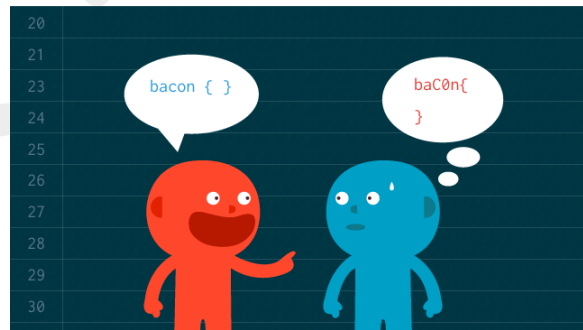
Filtrowanie danych



- ❑ Zawężenie dziedziny do danych na temat wypadków śmiertelnych
- ❑ Wybór zestawów danych z roczników, które są dostępne we wszystkich zbiorach
- ❑ Odrzucenie danych ze starych roczników jeśli ich zakres atrybutów jest zbyt wąski

Wspólny format

- ❑ Ujednolicenie formatu i semantyki danych
- ❑ Wybór atrybutów pozwalających wnioskować o przyczynach wypadków:
 - ❑ data
 - ❑ miejsce
 - ❑ warunki pogodowe
 - ❑ warunki środowiskowe
 - ❑ dane o uczestnikach (np. młodzi / pijani kierowcy)
 - ❑ dane o pojazdach (np. rocznik, systemy bezp.)



Integracja



Po przekształceniu danych do wspólnego formatu:

- ❑ (W miarę potrzeby i możliwości) uzupełnienie danych ze źródeł zewnętrznych
- ❑ Zapis do bazy danych

Analiza danych

- ❑ Analiza statystyczna
 - ❑ korelacje
 - ❑ procentowy udział atrybutów
- ❑ Wizualizacja danych
- ❑ Klasteryzacja
- ❑ Inne



Opcjonalne kierunki analizy

- ❑ Wygenerowanie zbiorczych statystyk na podstawie zgromadzonych danych
- ❑ Porównanie ze statystykami dla obszarów, dla których niedostępne są szczegółowe dane



Technologie

❑ python



- ❑ skrypty filtrujące / parsujące / konwertujące

- ❑ parser sas7bdat - github.com/openfisca/sas7bdat

❑ PostgreSQL

- ❑ relacyjna baza danych do persystencji

❑ Narzędzia do analizy danych:

- ❑ MS Excel

- ❑ R





Dziękujemy za uwagę