



# Integracja systemów transakcyjnych

**Robert Wrembel**  
**Politechnika Poznańska**  
**Instytut Informatyki**  
Robert.Wrembel@cs.put.poznan.pl  
www.cs.put.poznan.pl/rwrembel



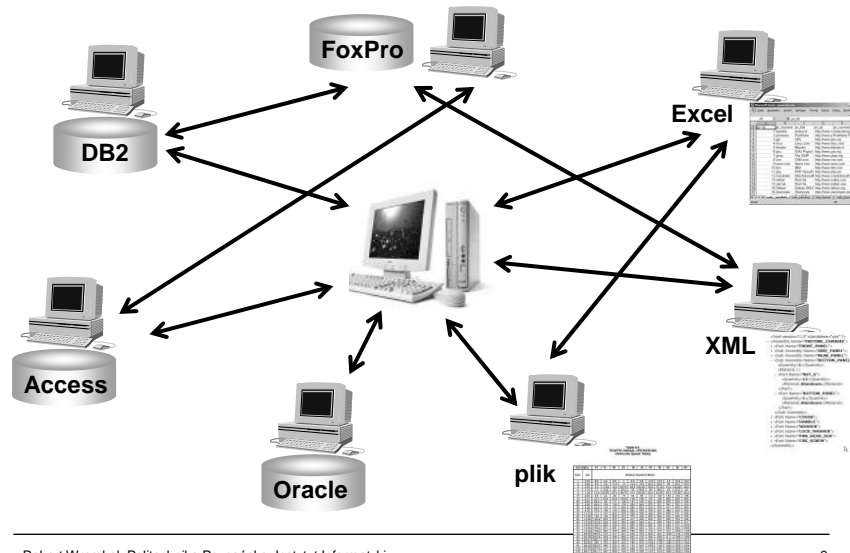
## Problematyka i architektury integracji danych

---

- ⇒ P2P
- ⇒ Systemy mediacyjne
- ⇒ Sfederowane BD
- ⇒ Systemy hurtowni danych



## Problematyka integracji danych



## Problematyka integracji danych (2)

- **Charakterystyka systemów źródłowych**
  - rozproszenie
  - heterogeniczność
- **Cele integracji**
  - systemy rozproszone
  - systemy analityczne (BI)



## Heterogeniczność źródeł

- **Różni producenci/technologie implementacyjne**
- **Różna funkcjonalność**
  - bazy danych / nie bazy danych
  - dialekty SQL
  - sposoby dostępu i przetwarzania danych
- **Różne modele danych**
  - hierarchiczne, sieciowe
  - relacyjne
  - obiektowe
  - obiektowo-relacyjne
  - wielowymiarowe
  - semistrukturalne
- **Architektury integracyjne**
  - systemy mediacyjne
  - hurtownie (magazyn) danych



## Charakterystyka źródeł danych

- **Technologia implementacyjna**
- **Funkcjonalność**
- **Konflikty na poziomie struktur danych**
  - różne reprezentacje danych (struktury)
  - **Dealer A**
    - Pojazdy (zawiera samochody osobowe + dostawcze)
  - **Dealer B**
    - Samochody\_Osobowe, Samochody\_Dostawcze



## Charakterystyka źródeł danych

---

- ⇒ **Konflikty na poziomie danych**
  - **Zduplikowane dane**
  - **Brakujące i błędne dane**
  - **Błędy wprowadzania wartości**



## Różne reprezentacje danych

---

- ⇒ **Różne modele danych w źródłach (relacyjny, obiektowy, semistrukturalny)**
- ⇒ **Różne typy danych**
  - **smallint, int, bigint, decimal (SQLServer)**
  - **smallint, int, bigint, float, real, double (DB2)**
  - **number, binary\_integer (Oracle)**
  - **znakowe typy danych o stałej i zmiennej długości**
- ⇒ **Różne ograniczenia integralnościowe**
- ⇒ **Inna reprezentacja tych samych danych**
  - **Pracownicy{NIP, imię, nazwisko, adres\_koresp}**
  - **Prac{NIP, imię\_nazw, ulica, dom, kod, miasto}**



## Różne reprezentacje danych

---

### ⇒ Homonimy

- Produkty.kod – oznacza kod produktu
- Klienci.kod – oznacza kod pocztowy

### ⇒ Synonimy

- Pacjenci.pesel
- Pacjenci.pacjentID (z wartością peselu)



## Konflikty na poziomie danych

---

### ⇒ Różne ziarno agregacji

- sprzedaż dzienna
- sprzedaż tygodniowa

### ⇒ Różne jednostki miary

- cena {PLN, EUR, USD}
- waga {kg, dkg}



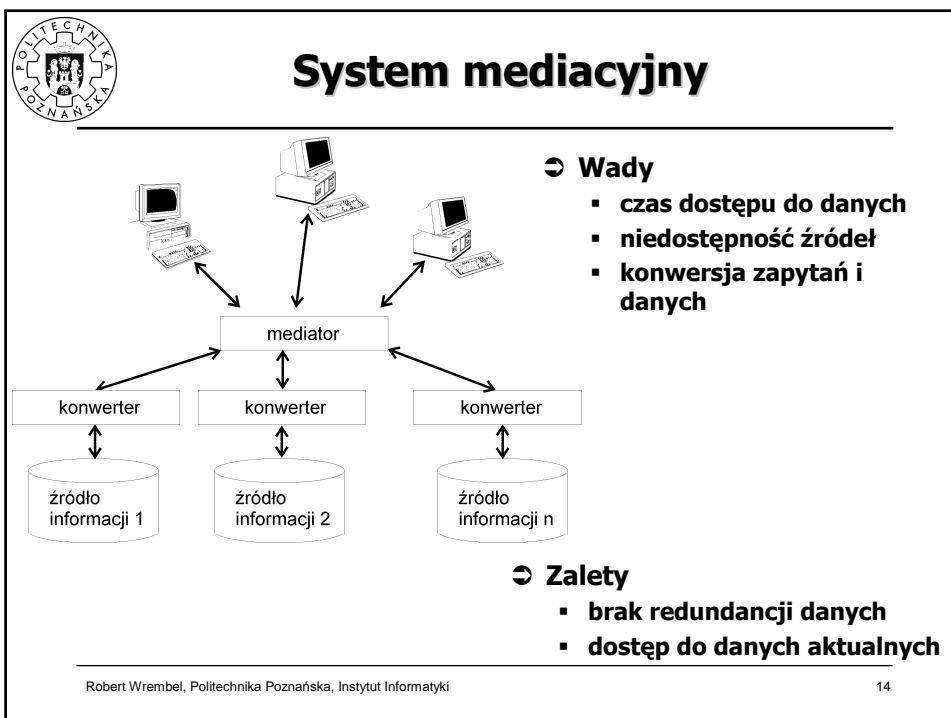
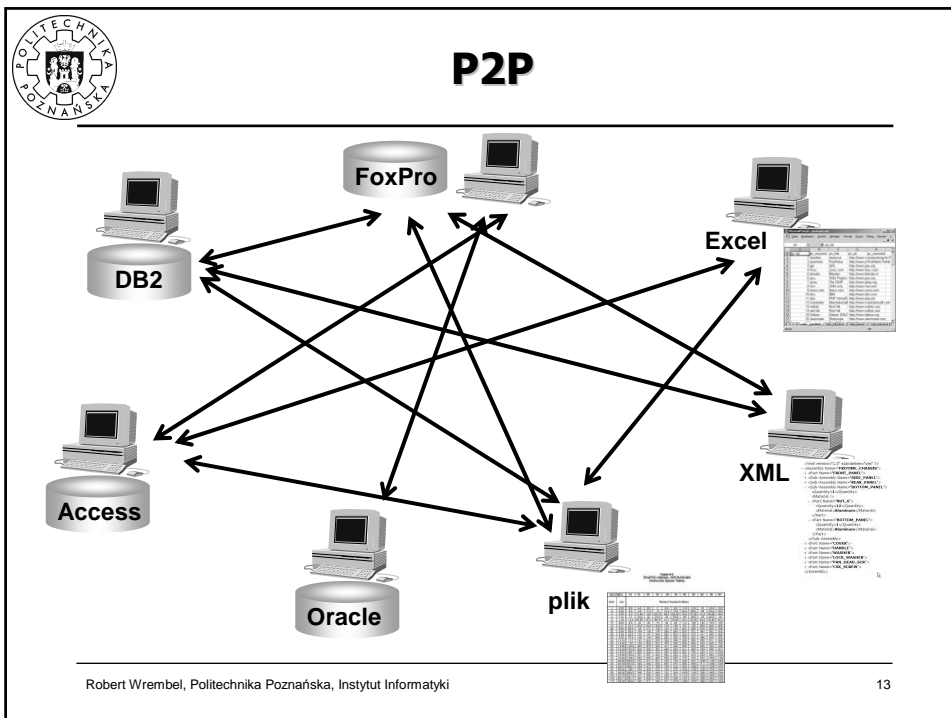
## Konflikty na poziomie danych

<b>Atrybut</b>	niedozwolona wartość	data_ur=30.13.1970	
	brakująca wartość	NIP=null	
	błąd literowy	Poznań	
	oznaczenia symboliczne	LC3X	czerwony metalik
	skrót	Pozn., P-ń, PN	
	wielkość liter	Poznań, poznań, POZNAŃ	
	różne oznaczenia symboliczne	pleć: {0, 1}, {K, M}, {kobieta, mężczyzna}	
	format	20-03-2008, 03/20/08	
	wartości złożone	R. Wrembel, 25.06.68, Szamotuły	
	kolejność wartości	{R.Wrembel} {Wrembel R.}	
	jednoznaczność	Wenecja (Włochy), Wenecja (Polska)	



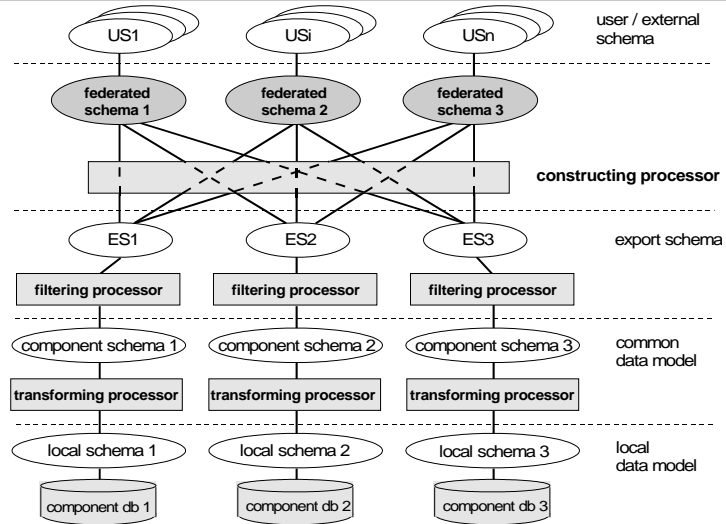
## Konflikty na poziomie danych

<b>Rekord</b>	niespełniona zależność pomiędzy wartościami atrybutów	cena_netto=100 cena_brutto=190	cena_brutto=cena_netto* 1,22
		ulica='Piotrowo' kod=62-300	
	naruszenie unikalności	{Robert Wrembel, WL8539024} {Bartosz Bębel, WL8539024}	nr dowodu osobistego powinien być unikalny
	wskazanie do nieistniejącego rekordu	{Robert Wrembel, Z20}	zespół Z20 nie istnieje
	duplikaty	{Robert Wrembel} {Wrembel Robert} {R. Wrembel}	
	sprzeczne wartości	{R. Wrembel, Szamotuły} {R. Wrembel, Poznań}	





## Sfederowane BD



## Hurtownia Danych

