

<i>customer-id</i>	<i>account-number</i>
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Bazy Danych i SQL

- Projektowanie baz danych (1). E-R Model.

<i>customer-id</i>	<i>account-number</i>
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 2

Etapy projektowania baz danych:

1. **Specyfikacja wymagań użytkownika** - określenie zjawisk, dostępności i użyteczności danych, ich formatu i sposobów obliczeń, cele, zakres i kontekst systemu
2. **Projektowanie koncepcyjne** - projektowanie schematu E–R bazy. Użycie modelu E–R wpływa również na realizację pozostałych faz.
3. **Specyfikacja wymagań funkcjonalnych** - dokładny opis wymagań klienta i wszystkich przyszłych użytkowników systemu
4. **Projektowanie logiczne i fizyczne**
5. **Implementacja**

<i>customer-id</i>	<i>account-number</i>
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 3

Projektowanie bazy danych:

✓ Projektowanie logicznej struktury bazy:

- **Etap I:** określenie encji i zdefiniowanie atrybutów opisujących encje
 - przyporządkowanie encji do zjawisk
 - standaryzacja nazw i formatów
 - identyfikacja źródeł danych
- **Etap II:** określenie związków między encjami
 - identyfikacja typu związków (relacji) (1-1, 1-M, N-M)
- **Etap III:** normalizacja relacji
 - obniżenie redundancji i wyeliminowanie anomalii (usuwania, wstawiania i aktualizacji)

✓ Projektowanie fizycznej struktury bazy:

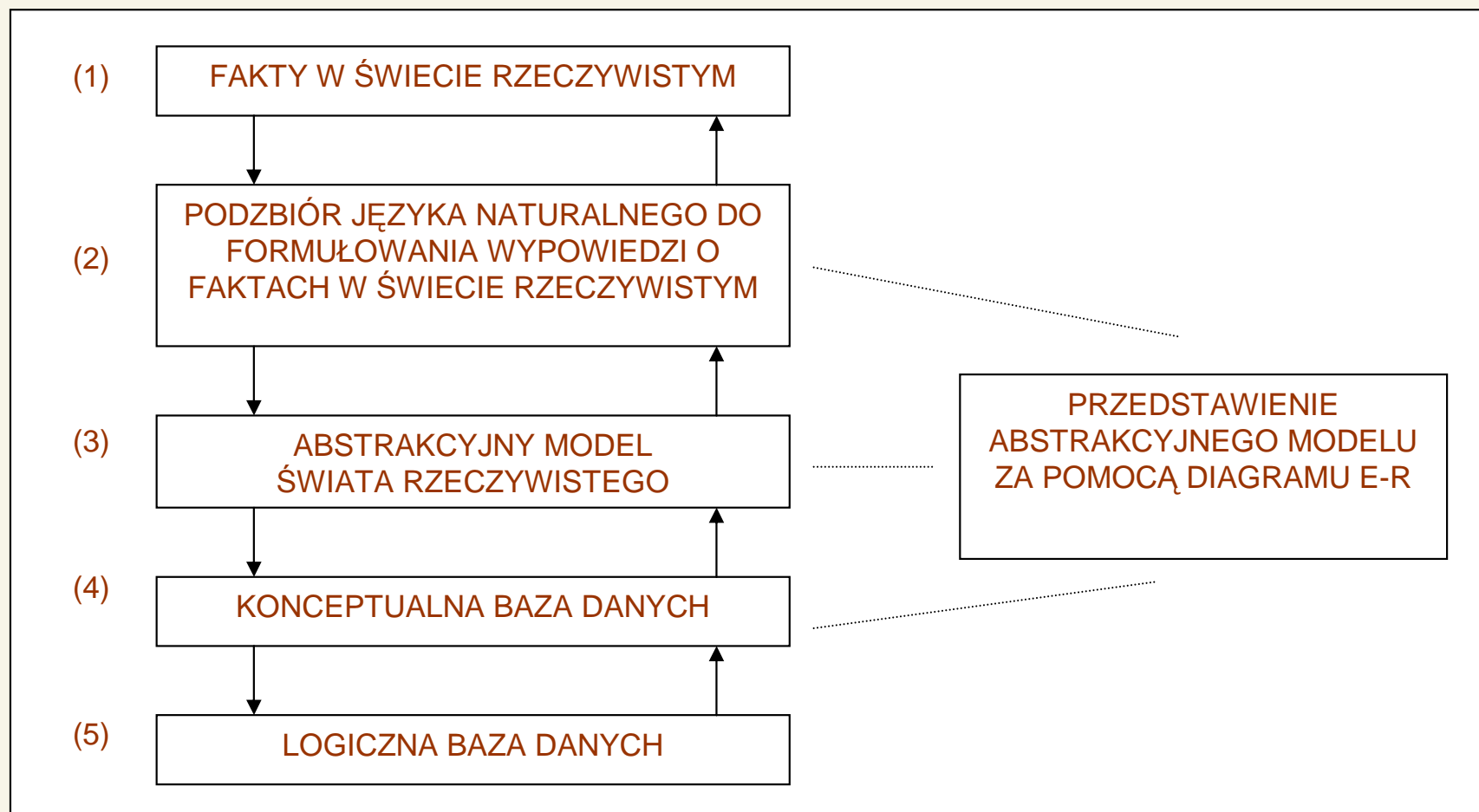
- nałożenie struktury logicznej na fizyczne urządzenia

customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 4

Projektowanie bazy danych:



<i>customer-id</i>	<i>account-number</i>
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 5

Wybór schematu relacji:

- ✓ Projekt bazy danych polega na znalezieniu właściwych schematów relacji tworzących bazę danych
- ✓ **Niewłaściwy projekt** (niewłaściwe schematy) mogą prowadzić do
 - Redundancji (powtarzania informacji).
 - Niemożności reprezentowania pewnych informacji.
 - Anomalii związanych z operowaniem danymi (głównie modyfikacją danych)
- ✓ **Cele:**
 - Unikanie redundancji danych
 - Zapewnienie reprezentowania związków między danymi
 - Zachowanie warunków integralności (umożliwienie kontroli warunków integralności podczas modyfikacji danych).

<i>customer-id</i>	<i>account-number</i>
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 6

Modelowanie konceptualne:

Model pojęciowy (konceptualny, podstawowy):

- ✓ Zapis wymagań w postaci sformalizowanej, abstrahujący od problemów implementacyjnych („co” a nie „jak”)
- ✓ Zawiera
 - modele graficzne
 - tekstowe (ale też sformalizowane) uzupełnienia modeli graficznych
- ✓ Jest ważny, ponieważ:
 - reprezentuje istotę wymagań
 - nie zależy od zmiennych możliwości implementacji
 - jest dobrą podstawą do projektowania systemów mających funkcjonować przez wiele lat

<i>customer-id</i>	<i>account-number</i>
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 7

Wymagania:

- ✓ Spełnienie wymagań użytkownika jest naszym celem działania
- ✓ Wymagania muszą więc być dokładnie znane
- ✓ Poprawne i kompletne sformułowanie wymagań na ogół nie jest proste
- ✓ Należy odróżniać *funkcje* systemu od *mechanizmów*, czyli sposobów implementacji funkcji
- ✓ Modelować dane niezależnie od przyszłej implementacji - **modelowanie E/R**
- ✓ Tworząc model pojęciowy nie można formułować zastrzeżeń co do realizowalności wymagań
- ✓ Zaprojektować dobry schemat - **normalizacja**

<i>customer-id</i>	<i>account-number</i>
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 8

Cel modelowania E-R:

- ✓ Model **E-R** służy do nieformalnego przedstawienia projektu bazy danych.
- ✓ Projekt ma postać graficzną zwaną diagramem **E-R** (entity-relationship diagram), diagramem **jednostka-związek** lub diagramem **związków encji**
- ✓ Istnieje procedura (pół)automatycznej transformacji diagramu **E-R** do konkretnej implementacji, na przykład do relacyjnej bazy danych.

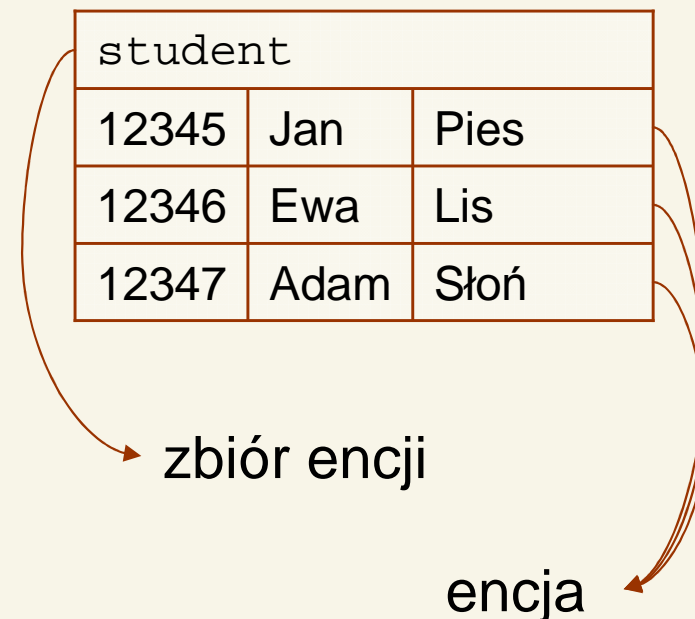
customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 9

Encja (zbiory encji):

- ✓ **Encja (*entity*)** - model rzeczy, osób, zjawisk, pojęć itp., o których chcemy przechowywać informacje, które mają tożsamość (są rozróżnialne)
Np. konkretna osoba, firma, zdarzenie
- ✓ **Zbiór encji (*entity set*)** – zbiór (klasa) obiektów, które są tego samego typu (mają wspólne własności), inaczej: zbiór wystąpień (instancji) tej klasy
Np. zbiór osób, zbiór firm, zbiór określonych zdarzeń



W diagramach E-R zbiory encji oznaczane są prostokątami:

student

<i>customer-id</i>	<i>account-number</i>
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 10

Atrybuty:

- ✓ Zbiór encji (i każda z nich) jest opisywany przez **zestaw atrybutów**, które odpowiadają własnościom posiadanym przez wszystkie obiekty reprezentowane przez ten zbiór. Np.:

```
student = (nr_albumu, imie, nazwisko)
```

- ✓ **Dziedzina atrybutu** – zbiór dopuszczalnych wartości, jakie może przyjmować dany atrybut na danym zbiorze encji. Dziedzina uzupełniona jest o wartość **NULL** oznaczającą, że encja nie posiada wartości atrybutu lub jest on nieznany.
- ✓ **Atrybuty:**
 - proste vs. pochodne (mogą być obliczone na podstawie innych informacji w bazie),
 - proste vs. złożone.

customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

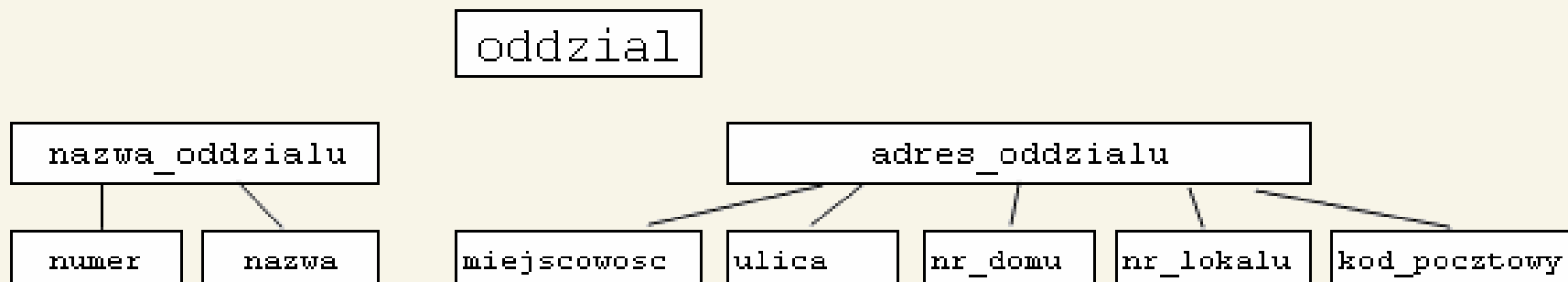
str. 11

Atrybuty proste vs. złożone (przykład):

zbiór encji:

atrybuty złożone:

atrybuty proste:



W diagramach atrybuty oznaczane są **elipsami**:

nazwa

Atrybuty złożone **elipsami z podwójną linią**:

nazwa

Atrybuty pochodne **elipsami z przerywanej linii**:

nazwa

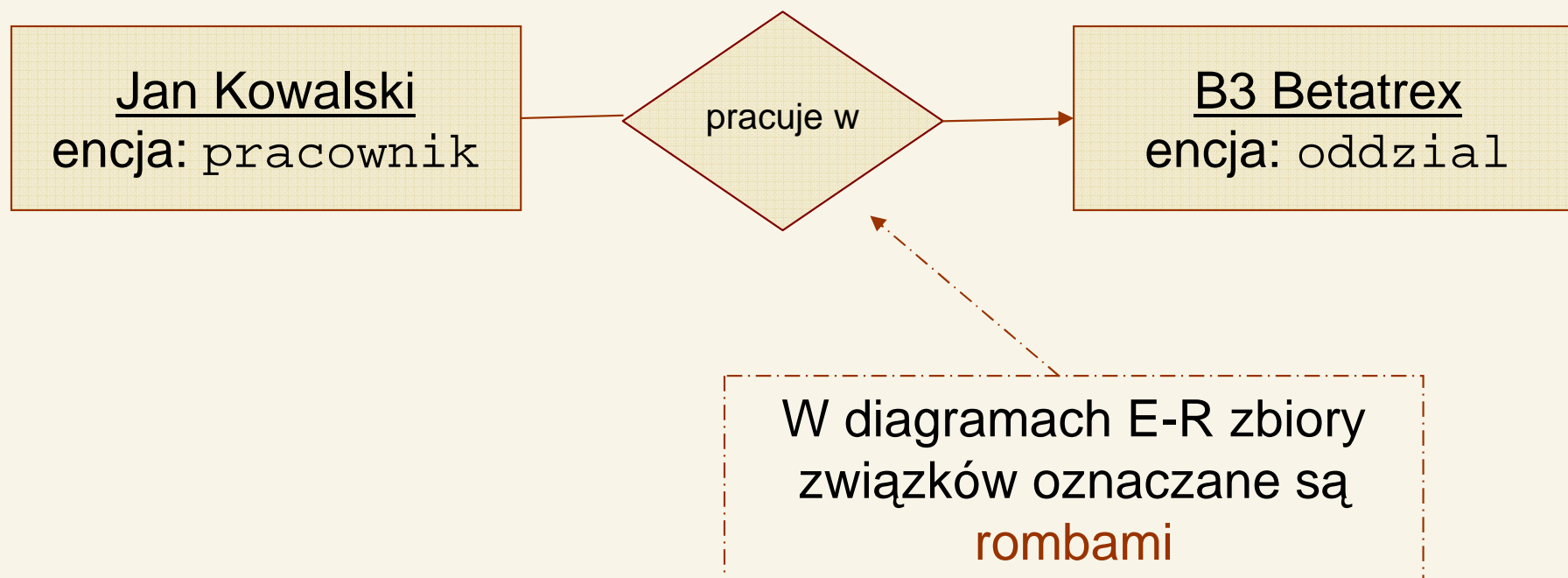
customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 12

Zbiory związków:

- ✓ **Związek** (*relationship*) reprezentuje relację pomiędzy pewnymi encjami (obiektami). Np.:

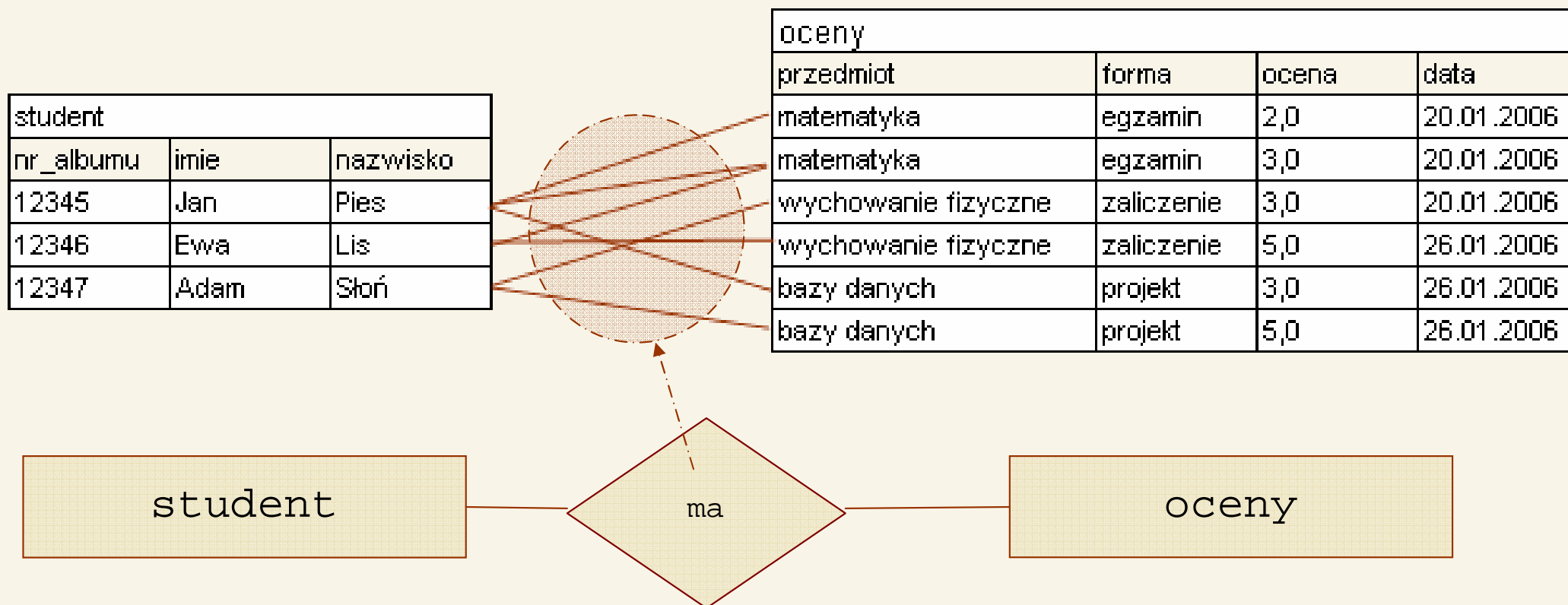


customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 13

Zbiór związków (przykład):



<i>customer-id</i>	<i>account-number</i>
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 14

Diagramy E-R – symbole podstawowe:

- ✓ **Prostokąty** – zbiory encji
- ✓ **Elipsy** – atrybuty
- ✓ **Elipsy – linia podwójna** – atrybuty złożone
- ✓ **Elipsy – linia przerywana** – atrybuty pochodne
- ✓ **Romby** – zbiory związków
- ✓ **Linie** – łączą zbiory encji ze zbiorami związków oraz atrybuty ze zbiorami encji i związków

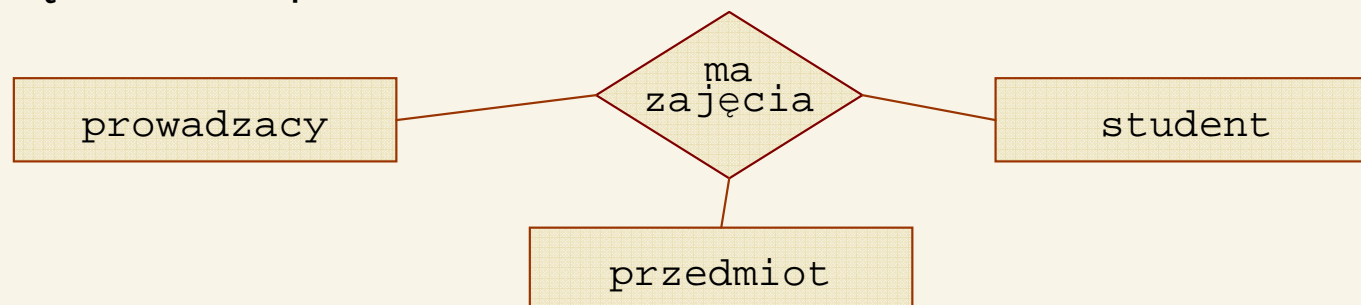
customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 15

Stopień zbioru związków:

- ✓ **Stopień zbioru związków** (*degree of a relationship set*) określony jest przez liczbę zbiorów encji, które pozostają w danym związku.
- ✓ Związek dwóch zbiorów encji nazywany jest **binarnym** (stopień = 2).
- ✓ Większość związków obserwowanych w świecie rzeczywistym to związki binarne.
- ✓ W związku może uczestniczyć pewna, większa od 2 ilość zbiorów encji – związek ***n*-arny** (stopień = *n*) – mówimy wtedy o związku wieloargumentowym lub wielocłonowym.
 - Np. zbiory encji: `przedmiot`, `przewodzący` oraz `student` mogą być połączone związkiem o stopniu = 3.



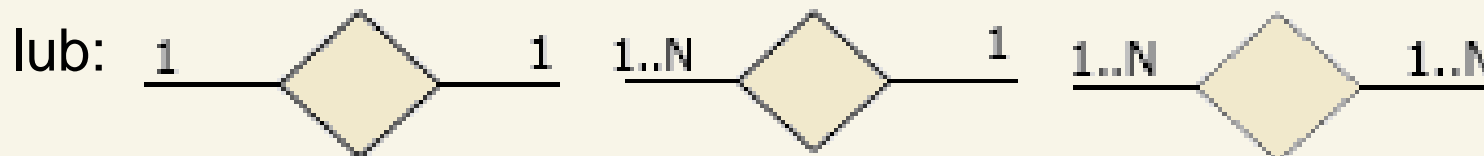
customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Stopień przyporządkowania (krotność związków):

- ✓ **Stopień przyporządkowania** zbiorów encji do związku (relacji) określony jest przez ilość encji z każdego zbioru, które są ze sobą w tym związku powiązane.
- ✓ **Rodzaje związków binarnych** ze względu na stopień przyporządkowania:
 - *jeden do jeden*
 - *jeden do wielu*
 - *wiele do wielu*
- ✓ **W diagramie E-R** rodzaje związków są oznaczane w następujący sposób:

→ (grot strzałki) oznacza stopień przyporządkowania *jeden*

— (brak grotu) – *wielu*



customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 17

Stopień uczestnictwa:

- ✓ Jeżeli każda encja zbioru **musi** brać udział w związku, wówczas jest on całkowity (uczestnictwo obowiązkowe); w przeciwnym razie – częściowy (uczestnictwo opcjonalne).
- ✓ **W diagramie E-R linie podwójne** łączą zbiory encji ze zbiorami związków w przypadku całkowitości.
- ✓ Występowanie encji pewnego zbioru (istnienie obiektu w rzeczywistości) może zależeć od występowania encji (obiektu) innego zbioru.
- ✓ Jeżeli występowanie encji *x* zależy od *y* (*is existence dependent on*), to:
 - *y* jest encją nadrzędną (dominującą) (*dominant entity*)
 - *x* jest encją podrzędną (*subordinate entity*)
- ✓ Np. jeśli encja dominująca **student** zostanie usunięta z bazy, to wszystkie wystąpienia związanych z nią encji **oceny** także muszą zostać usunięte.

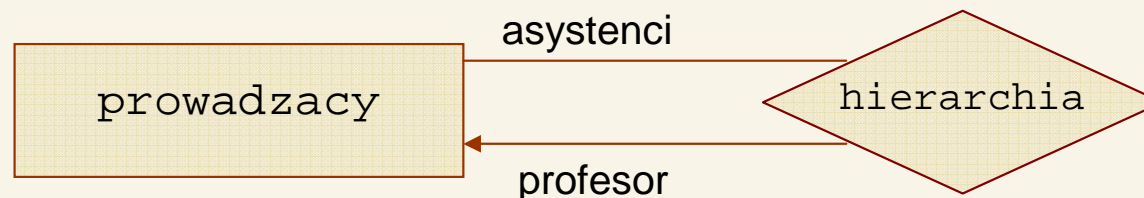
customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 18

Role:

- ✓ Zbiory encji pozostające w związku niekoniecznie muszą być różne, wówczas encje takie występują w różnych **rolach**, a związek taki nazywamy **związkiem rekurencyjnym**.
- ✓ **W diagramie E-R** role oznaczane są przez etykiety przy liniach łączących symbol encji (prostokąt) i związku (romb). Oznaczanie ról nie jest obowiązkowe, stosuje się je dla podkreślenia semantyki związku.



customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 19

Decyzje projektowe:

✓ Zbiory encji czy atrybuty

Wybór jest zależny od rodzaju modelowanego przedsięwzięcia (fragmentu rzeczywistości) oraz od znaczenia poszczególnych atrybutów.

✓ Zbiory encji czy zbiory związków

Możliwa wskazówka: związek wyznacza pewną akcję (zdarzenie), która zachodzi (zaszła) pomiędzy encjami (obiektami).

✓ Związki binarne czy związki o wyższym stopniu

Związek o stopniu większym niż 2 zawsze może być zastąpiony przez pewną liczbę związków binarnych. Jednak związki n-arne mogą lepiej oddawać semantykę relacji (jawne określenie, że chodzi o związek więcej niż 2 obiektów).

✓ Przykład:

1. Adres jako atrybut

Adres może być traktowany jako atrybut (na przykład encji `student`), wówczas jego wartością jest tekst, np: "Poznań, ul. Kwiatowa 5 m. 12".

2. Adres jako jednostka

Adres może być traktowany jako niezależna jednostka o atrybutach: `Miejscowość`, `Ulica`, `Nr Domu`, `Nr Lokalu`.

<i>customer-id</i>	<i>account-number</i>
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 20

Ograniczenia - warunki integralności

- ✓ W ramach projektowania baz danych specyfikuje się pewne ograniczenia ilościowe i zależności pomiędzy encjami lub związkami, które mają zachodzić **zawsze**.
- ✓ Ograniczenia te i zależności określają **warunki integralności** rozważnej bazy i są następnie sprawdzane i dotrzymywane.
- ✓ Do najistotniejszych warunków ogólnych, specyfikowalnych w modelu E-R należą:
 - stopień przyporządkowania zbiorów encji do związku,
 - zależność występowania encji z różnych zbiorów,
 - całkowitość vs. częściowość udziału zbioru encji w zbiorze związków.

<i>customer-id</i>	<i>account-number</i>
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 21

Klucze w zbiorze encji:

- ✓ **Nadkluczem** (*super key*) jest każdy zbiór atrybutów, którego wartości jednoznacznie identyfikują encję w zbiorze encji.
- ✓ **Kluczem kandydującym** (*candidate key*) nazywany jest nadklucz o minimalnej liczbie atrybutów w zbiorze encji.
- ✓ W danym zbiorze może występować kilka kluczy kandydujących, jeden z nich (dowolnie wybrany) nazywany jest **kluczem głównym** (*primary key*).
- ✓ **W diagramie E-R** klucze główne są oznaczane przez podkreślenie etykiety atrybutu.

customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 22

Klucze w zbiorze związków:

- ✓ Jeżeli zbiór związków posiada atrybuty o odpowiednich własnościach, na podstawie tych atrybutów można zdefiniować dla niego (jednoznacznie identyfikujące) **klucze główne** tak, jak dla zbiorów encji.
- ✓ **Złożenie** kluczy głównych zbiorów encji biorących udział w związku tworzy **nadklucz** dla zbioru związków.
- ✓ Sposób zdefiniowania **klucza głównego** zależy od stopnia związku oraz stopnia przyporządkowania poszczególnych zbiorów encji.
 - Dla binarnego związku *jeden do jeden* każdy z kluczy głównych (encji) może być kluczem głównym.
 - Dla binarnego związku *wielu do wielu* para kluczy głównych encji jest kluczem głównym.

customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 23

Zbiory słabe encji:

- ✓ Zbiór encji, który nie posiada klucza głównego (nie ma odpowiednich atrybutów), jest nazywany **zbiorem słabym encji** (*weak entity set*) w przeciwieństwie do zbioru silnego - (*strong entity set*).
- ✓ **Informacyjność** zbioru słabych encji jest uwarunkowana istnieniem innego silnego zbioru encji, który pozostaje ze słabym w związku **jeden do wielu**. Encje tego zbioru silnego dominują również nad encjami słabego w sensie zależności występowania.
- ✓ W zbiorze słabych encji występować musi wówczas **klucz częściowy** (*partial key*), który pozwala na rozróżnienie (częściowe) jego encji względem dominującego zbioru silnego.
- ✓ **Kluczem głównym** zbioru słabego encji jest kompozycja klucza głównego dominującego zbioru silnego encji i klucza częściowego zbioru słabego.

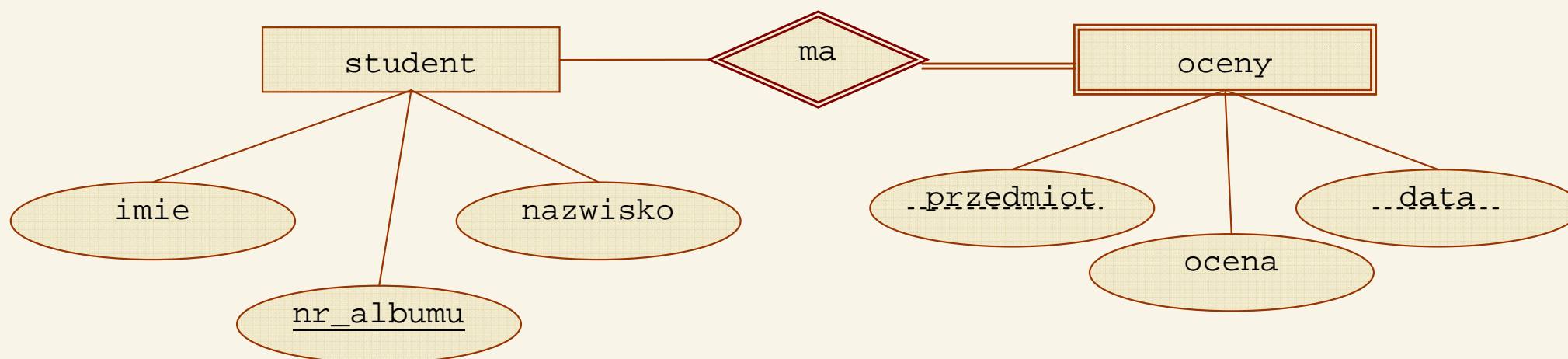
customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 24

Zbiory słabe encji (c.d.):

- ✓ W diagramie E-R **klucze częściowe** oznaczane są przez **podkreślenie** etykiety atrybutu **linią przerywaną**.
- ✓ Związki między słabą jednostką a jednostkami jej właścicielami nazywamy **związkami identyfikującymi**.
- ✓ W notacji graficznej zbiory słabe encji oraz zbiory związków identyfikujących rysujemy **podwójną linią**.



<i>customer-id</i>	<i>account-number</i>
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 25

Diagramy E-R → tabele

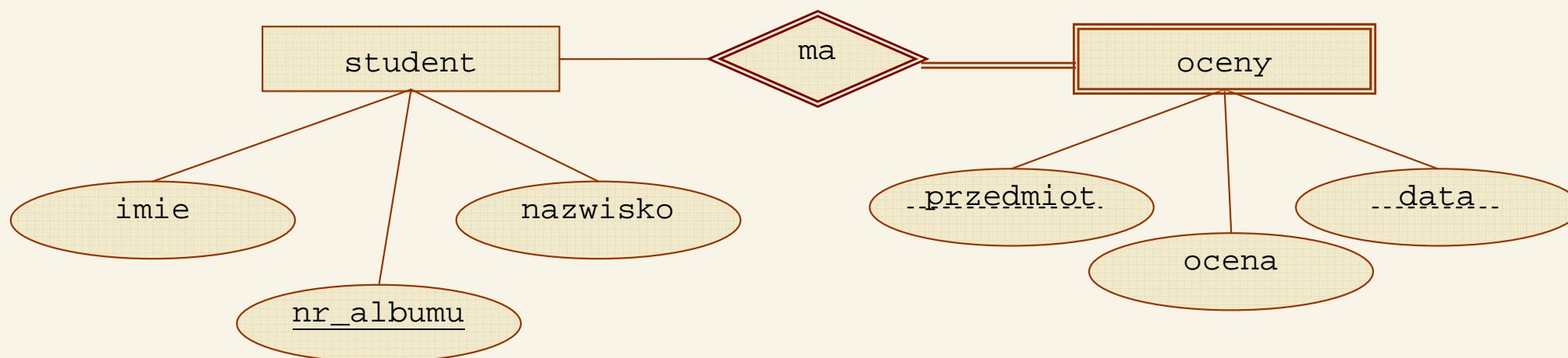
- ✓ Opis bazy danych za pomocą diagramu E-R może być przekształcony w opis w postaci kolekcji tabel.
- ✓ Istnienie kluczy głównych pozwala na jednoznaczne wyrażenie zbiorów encji i zbiorów związków w formie tabel.
- ✓ **Zasady ogólne**
 - Każdemu zbiorowi encji odpowiada tabela (tabela przyjmuje nazwę zbioru encji).
 - Każdemu zbiorowi związków (jeżeli ten posiada atrybuty) odpowiada tabela (tabela przyjmuje nazwę zbioru związków).
 - Kolumny tabel reprezentują atrybuty opisujące odpowiednio encje lub związki.
 - Wiersze tabel odpowiadają encją lub związkom
- ✓ **Zbiór silny encji** jest reprezentowany bezpośrednio
- ✓ Dla **zbioru słabego encji** tabela musi zawierać dodatkowo atrybuty stanowiące klucz główny tabeli nadrzędnej.

customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Kraków, 2006

str. 26

Diagramy E-R → tabele (przykład)



student		
nr_albumu	imie	nazwisko
12345	Jan	Pies
12346	Ewa	Lis
12347	Adam	Słoń

oceny			
nr_albumu	przedmiot	ocena	data
12345	matematyka	2,0	20.01.2006
12346	matematyka	3,0	20.01.2006
12347	wychowanie fizyczne	3,0	20.01.2006
12346	bazy danych	3,0	26.01.2006
12347	bazy danych	5,0	26.01.2006

customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

Diagramy E-R → tabele (c.d.)

- ✓ Jeżeli **zbiór związków** nie ma własnego klucza głównego, wówczas tabela musi dodatkowo zawierać odpowiednią do własności kombinację kluczy głównych swoich zbiorów encji.
 - Np. Związek *wielu do wielu* jest reprezentowany przy pomocy tabeli o dodatkowych kolumnach odpowiadających kluczom głównym zbiorów encji biorących udział w związku.



student		
nr_albumu	imie	nazwisko
12345	Jan	Pies
12346	Ewa	Lis
12347	Adam	Słoń

ma_zajecia	
nr_albumu	ID_przedmiot
12345	1
12345	3
12346	1
12347	1
12346	4
12347	4
12345	4

przedmioty				
ID_przedmiot	rodzaj_studiow	przedmiot	prowadzacy_pesel	forma
1	stacjonarne	matematyka	29035934567	egzamin
2	zaoczne	matematyka	15057698456	egzamin
3	podyplomowe	bazy danych	31169675846	zaliczenie
4	stacjonarne	ekonomia	31169675846	zaliczenie