SCHEMA INTEGRATION AMONG DATABASES
SCHEMA INTEGRATION

• PURPOSES FOR SCHEMA / VIEW INTEGRATION
  – **NEW DATABASE DESIGN**: MERGE MULTIPLE PEOPLE’S VIEWS
  – **DATABASE REDESIGN**: MERGE MULTIPLE EXISTING DB’S TO NEW DB
  – **DATA WAREHOUSE**: CREATE INTEGRATED SCHEMA / DATABASE
  – **INTEGRATED VIEW** FOR DISTRIBUTED HETEROGENEOUS DBMS
EVOLUTION OF CONCEPTUAL DATABASE DESIGN

• TRADITIONAL DATA MODELS
  - HIERARCHICAL
  - NETWORK

• SEMANTICALLY -- RICHER DATA MODELS
  - ENTITY-RELATIONSHIP (ER) MODEL
  - SEMANTIC DATA MODEL
  - OBJECT-ORIENTED (OO) DATA MODEL
  - OBJECT-RELATIONAL (OR) DATA MODEL

• “CONCEPTUAL” DATABASE DESIGN
  - LEADS TO “CONCEPTUAL SCHEMA”
  - AIDS INTEGRATION (“SCHEMA INTEGRATION”)
PROCESS OF SCHEMA INTEGRATION

• DATA -- AN AUTONOMOUS RESOURCE
• NEED TO CAPTURE “MEANING” OF DATA
• DESIRE “ENTERPRISE-WIDE” VIEW OF DATA

TERMINOLOGY
• SCHEMA INTEGRATION
  – COMPONENT SCHEMA
  – INTEGRATED SCHEMA
• VIEW INTEGRATION
  – USER VIEW
  – CONCEPTUAL VIEW
• DATABASE INTEGRATION
  – LOCAL SCHEMA
  – GLOBAL SCHEMA

EXAMPLE
• TWO SCHEMAS DEVELOPED BY TWO GROUPS
  – BOTH FOCUS ON BOOKS/PUBLICATIONS
ENTITY-RELATION (ER) DIAGRAM – For Book Distributor

- ENTITIES
- RELATIONSHIPS
- ATTRIBUTES
**Step 1. Original schemas**

- **Book Distributor**
  - Publisher
  - Book
  - University
  - Topics

- **University Library**
  - Title
  - Code
  - Publisher
  - Research Area

**Step 2. Choose “Topics” for “keyword” (Schema 2).**

- **Book Distributor**
  - Publisher
  - Book
  - University
  - Topics

- **University Library**
  - Title
  - Code
  - Publisher
  - Research Area

**Step 3. Make Publisher into an entity (Schema 2).**

- **Book Distributor**
  - Publisher
  - Book
  - University
  - Topics

- **University Library**
  - Title
  - Code
  - Publisher
  - Research Area

**Entity and Attribute Equivalence**

**Attribute vs Entity**
Step 4. Superimposition of schema

Step 5. Creation of subset relationship
Step 6. Drop the properties of Book common to Publication
CAUSES OF DIFFERENCES

• DIFFERENT PERSPECTIVES
  – DIFFERENT NAME FOR SAME CONCEPT
  – INTERVENING STEPS
    • EMPLOYEE-DEPARTMENT
    • EMPLOYEE-PROJECT-DEPARTMENT

• EQUIVALENCE AMONG CONSTRUCTS
  – MODEL AS ATTRIBUTE OR ENTITY (E.G., PUBLISHER)

• INCOMPATIBLE DESIGN SPECIFICATIONS
  – RELATIONSHIP DIFFERENCES
    (E.G., 1:N VS. N:M)
SOME EXAMPLES

• NAMING CONFLICTS
  – **HOMONYMS**: SAME NAME, DIFFERENT CONCEPT
    E.G., EQUIPMENT
    E.G., “SIZE” OF DRESS = 1 NUMBER, “SIZE” OF PANTS = 2 NUMBERS)

  – **SYNONYMS**: SAME CONCEPT, DIFFERENT NAMES
    E.G., CLIENT, CUSTOMER, PATIENT, ...

• STRUCTURAL CONFLICTS
  – **TYPE**: ENTITY VS. ATTRIBUTE (E.G., PUBLISHER)
  – **DEPENDENCY**: 1:1 VS. N:M (E.G., HUSBAND: WIFE)
  – **KEY**: ALTERNATE WAYS TO IDENTIFY ENTITY (E.G., SS# VS. EMP#)
  – **BEHAVIORAL**: DIFFERENT INSERTION/DELETION POLICIES
    (E.G., CAN YOU HAVE DEPARTMENT WITH NO EMPLOYEES?)
EXAMPLE SCHEMAS TO BE INTEGRATED:

Departmental database

- DEPT (n: 1)
  - name
  - DEPT

- STUDENT (n: 1)
  - name
  - id
  - STUDENT

- PROFESSOR (m: 1)
  - name
  - position
  - STUDENT

- SUBJECT (m: 1)
  - sub #

Registrar's database

- COURSE (sub #: 1)
  - sub #

- STUDENT (id: 1)
  - name
  - grade

Staff Telephone directory

- STAFF
  - name
  - room #
  - ext #
  - street
  - city
  - phone #

Student Telephone directory

- STUDENT
  - name
  - year

- OFFICE ADDRESS
  - room #
  - ext #

- HOME ADDRESS
  - street
  - city
  - phone #

- TERM ADDRESS
  - street
  - city
  - phone #
INTEGRATED SCHEMA: (an exercise for the reader)
• INCREASING NEED TO PROVIDE AN INTEGRATED GLOBAL VIEW OF AN ORGANIZATION’S INFORMATION
  - AND SOMETIMES RELATED ORGANIZATIONS (CUSTOMERS / SUPPLIER)

• AN IMPORTANT STEP IS THE CREATION OF A GLOBAL SCHEMA
  - INTEGRATES SEPARATE SCHEMAS
  - CONTAINS ALL THE CRITICAL INFORMATION NEEDED