

Geog 469
GIS Workshop

System Requirements - Data

Outline

1. What are some principles of project management?
2. What are some fundamental issues associated with system requirements?
3. What are some issues intrinsic to spatial data?
4. What is a UML Class Diagram? How do we use it?
5. How might we choose a logical data model?

Part 1.

1. What are some principles of project management?

What is a project?

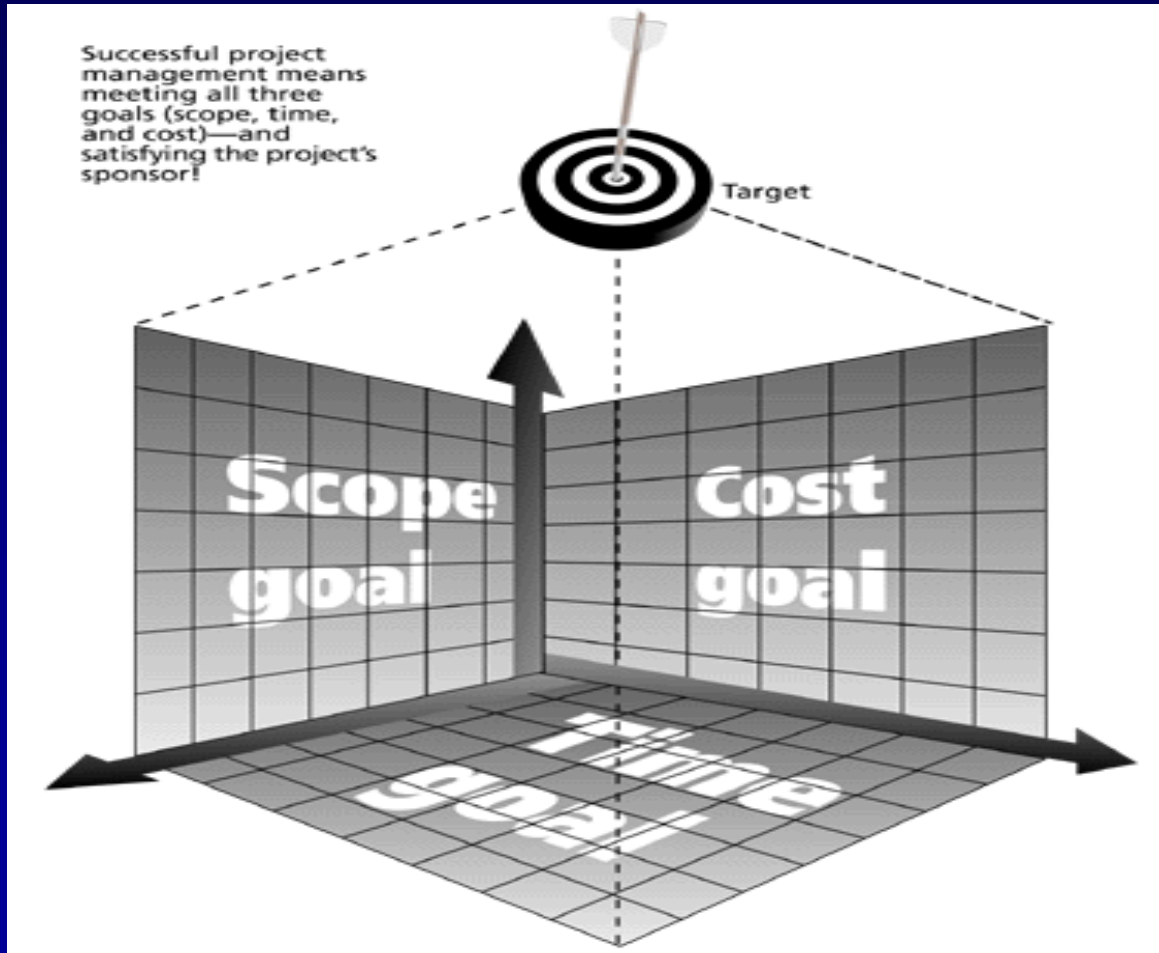
What is an application?

Project – focus on processing data on topic	Application – focus on repeat projects for topic
Topic-focused activity, single database	Topic-support activity, many project databases
Temporary in nature, space and time context	Ongoing, supports many projects, many contexts
Produces a unique product or service	Produces similar products or services over and over
End is determined by specific criteria of goal	Processes address goal, but are ongoing

The Triple Constraint

- Every project is constrained by its:
 - **Scope**: what's included in goal and what's not
 - **Schedule**: anticipated time to complete goals
 - **Cost**: budget allocated for project completion
- Project manager's primary duty is to balance these competing constraints, while satisfying intended business need

The Triple Constraint of Project Management



System requirements report helps you define the scope, cost, and timing constraints in advance.

Part 2.

2. What are some fundamental issues associated with system requirements?

System requirements and project management

- System requirements can be thought of as an implementation plan that recognizes and balances the triple constraints given information products identified during needs assessment
- System requirements report will be composed of
 - Data requirement
 - Software requirement
 - Hardware requirement
 - Staffing requirement
 - Timing requirement

Focus on data requirements: Components of Master Input Data List

Tomlinson (2011) figure 6.17, p. 69

1. Data identification
2. Data volume considerations
3. Data characteristics
4. Data availability and cost

How is Master Input Data List useful?

Help you recognize requirements for data, software, hardware, staffing, and timing.

Data: are data ready? Options for putting data into system...
(Buy? Outsourcing? In-house data creation?)

Software: which functions (operations) are needed to put data into system?

Hardware: workstation and server purchases are determined by data volume and computing complexity

Data Inventory form

Data inventory

Interviewer: _____ Date: _____
Name: _____ Department: _____
Phone#: _____ Division/Office: _____

Data set name: _____ Data set number: _____
Source agency: _____ Metadata? (circle) Yes No
Data format: (circle) Map Manual data file Automated data file
Air photo Imagery
Other: _____ Source data medium: _____
Digital data format: _____
Percentage available now in a digital format: _____
Total data volume: (# of pages, CDs, maps, etc.): _____
Scale of source data: Map projection/datum: _____
Photo image? Yes No
Digitizing effort (per sheet): # of polygons: _____
of lines: _____
of points: _____
Date of data capture: _____ Date of last update: _____
Percent of coverage available now: _____
Date entire coverage is available: _____
Restrictions on use: _____
Cost of data set acquisition \$: _____
Royalties on acquisition and use \$: _____

From Figure 6.18 in Tomlinson (2011) p. 72

Data availability vs. data ready

- Data availability: the date on which you receive data from its source
- Data ready: the date on which the data are in the system, processed, and ready for use in the creation of an information product
- Identify functions needed to put data available into the system (data readiness)
- Examples: to be ready for the system
 - Satellite image requires file transfer and ID creation
 - Paper map requires scanning, digitizing, topology editing, attribute editing, and so on
 - Digital data requires less work most of the time, but still might require conversion to appropriate spatial object type

Setting Priorities

- Help manage the scope of projects
- What are important and what are less important information products and data?
- Occurs due to time/cost constraints
- Decision affected by an organization's strategic plan/goals

Before estimating timing requirements, consider...

- Data input timing
 - Various options for putting data into system will affect timing
- Information product application programming timing
 - Any scripting effort, what are steps involved?
Required work hours?
- System acquisition timing
 - Some hardware should be purchased/waited
- Training and staff issues
 - Any training need?
 - Make the use of existing staff or hire new persons

Part 3.

3. What are some issues intrinsic to spatial data?

Choosing an appropriate map projection

- Intended uses of information products
 - Be aware of distortion of geometric properties (e.g. area, shape, direction, distance) in map projections
- Geographic area of interest
 - Location of standard line/parallels
- Difference in datum?
 - e.g. NAD27 vs. NAD83
- Difference in measurement units due to size of region?
 - e.g. SPC in feet vs. UTM in meters

Appropriate map scale

- The appropriate map scale can be chosen based on feature resolution and error tolerance
- Resolution: the minimum amount of detail in a feature that can be mapped at a given scale; related to data requirements for information products
- Accuracy of position: how much error can be tolerated in the data? related to cost
- p. 81 (figure 7.6, 7.7) has suggested options for the choice of map scale based on percent error in area measurement; or resulting error tolerance given a minimum area resolution and map scale

Topological editing

- Arc/info coverage
 - Stored in feature attribute table
 - Use BUILD/CLEAN command
- Shapefiles
 - Non-topological model
- Geodatabase
 - Topological rules are assigned to feature data set
 - Topological rules within/between feature class can be defined
 - Use topological editing tool; help you identify potential topological errors given predefined rules

Part 4.

4. What is a UML Class Diagram?
How do we use it?

What is UML?

- Acronym for Unified Modeling Language
- Standardized efforts that combine many years of work on systems analysis and design
- Use graphic notations, i.e. diagramming language
 - More intuitive language; somewhere between natural language and programming language
- UML can support user requirements modeling, planning, system design, database design, and more; tasks are expressed through many different types (dialects) of diagrams

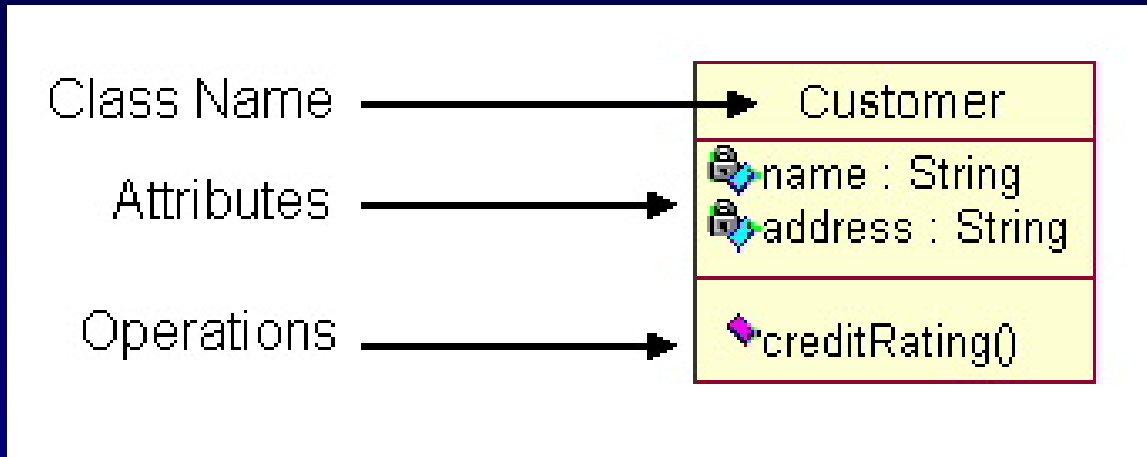
Data modeling is part of systems design

- Three data modeling contexts:
 - Conceptual data modeling
 - Logical data modeling
 - Physical data modeling
- **Class diagram** is perhaps the most important type of UML diagram; used for conceptual data modeling
- Database design is performed using data modeling class diagrams

Class Diagrams

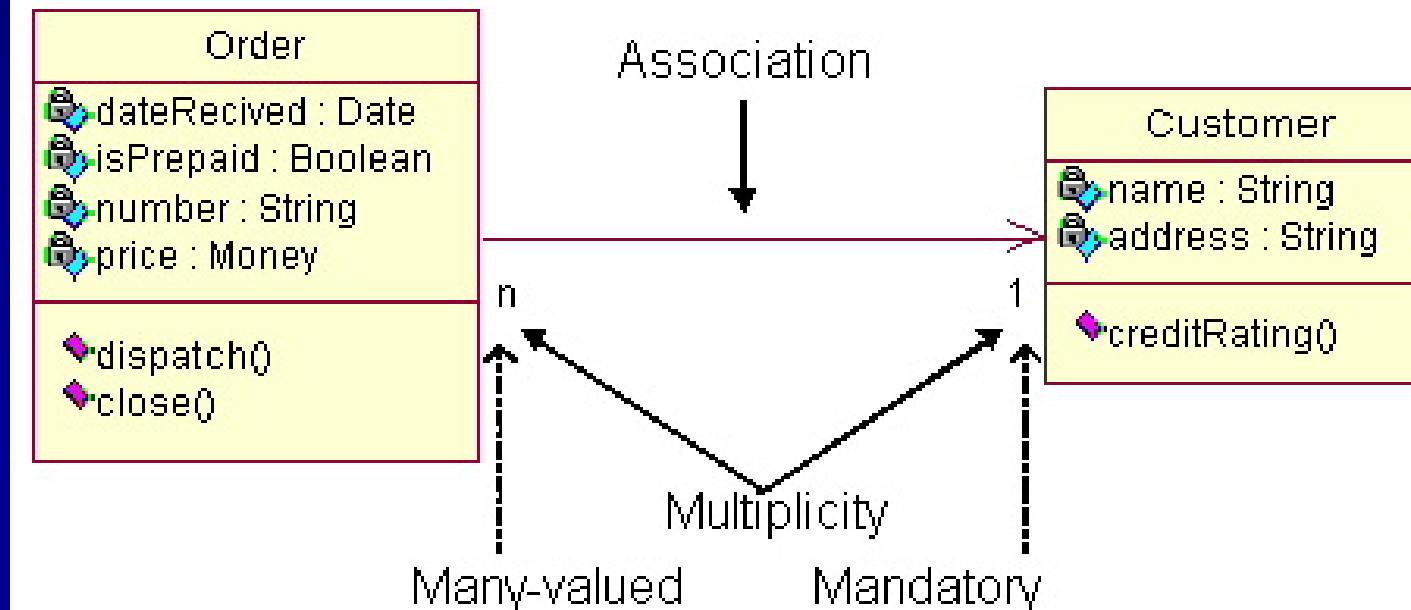
- Specify entity classes (e.g., feature classes)
- Specify relationships important to understanding problem topic (composite classes)
 - Classification (construct entity classes)
 - Association (relate entity classes)
 - Aggregation (compose entity classes)
 - Generalization (create super- and sub-classes)
- Specify operations

Class Diagrams

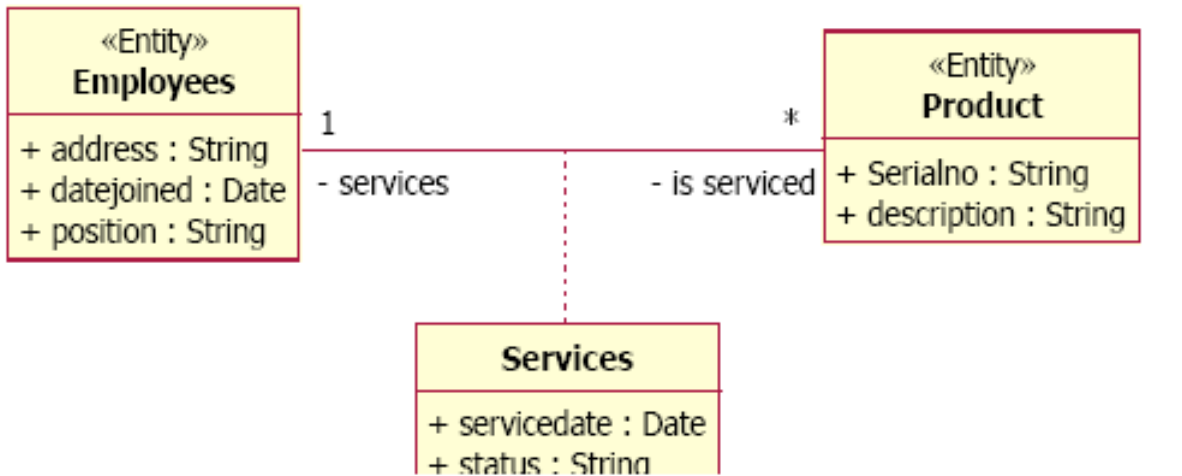


Associations

- Physical location (e.g., is next to)
- Direct actions (e.g., drives)
- Communication (e.g., talks to)
- Ownership (e.g., has)
- Satisfaction of a condition (e.g., works for)

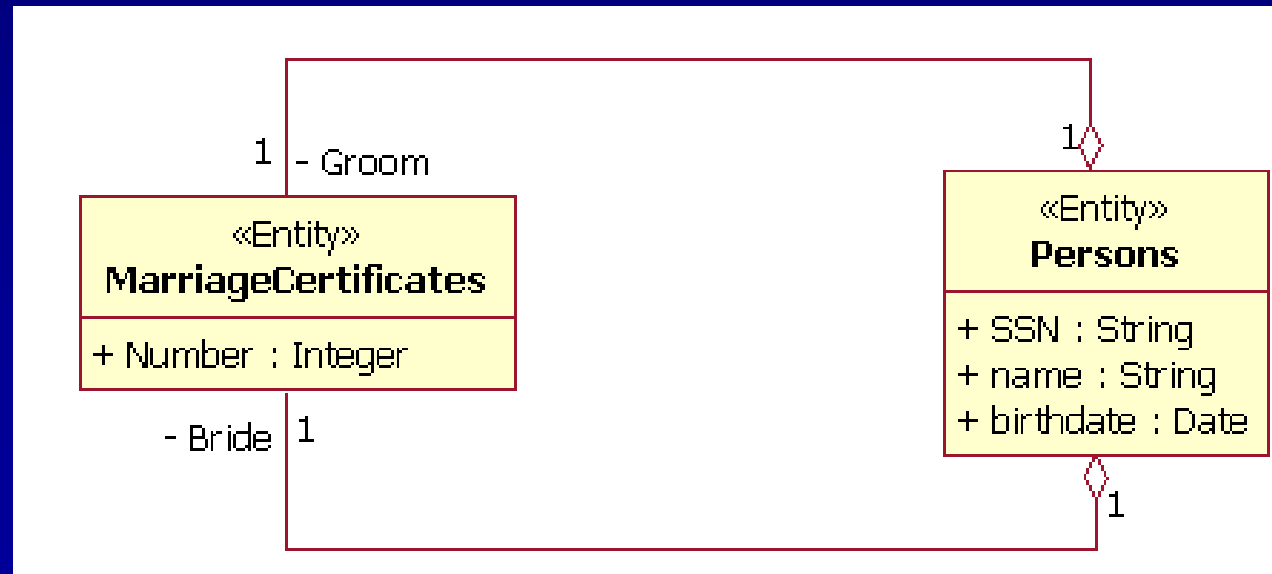


Class Diagrams



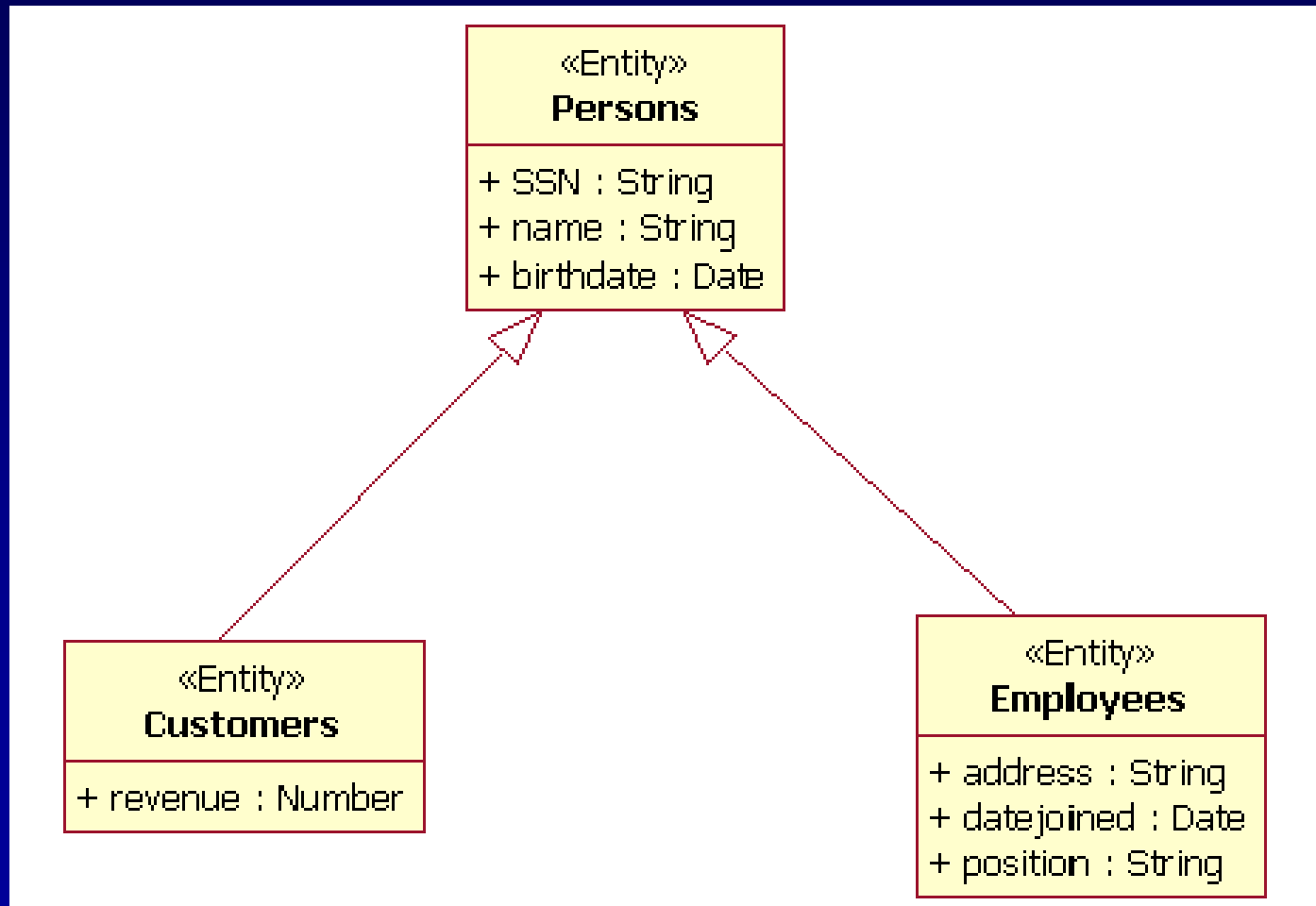
Association relationship

Aggregation relationship

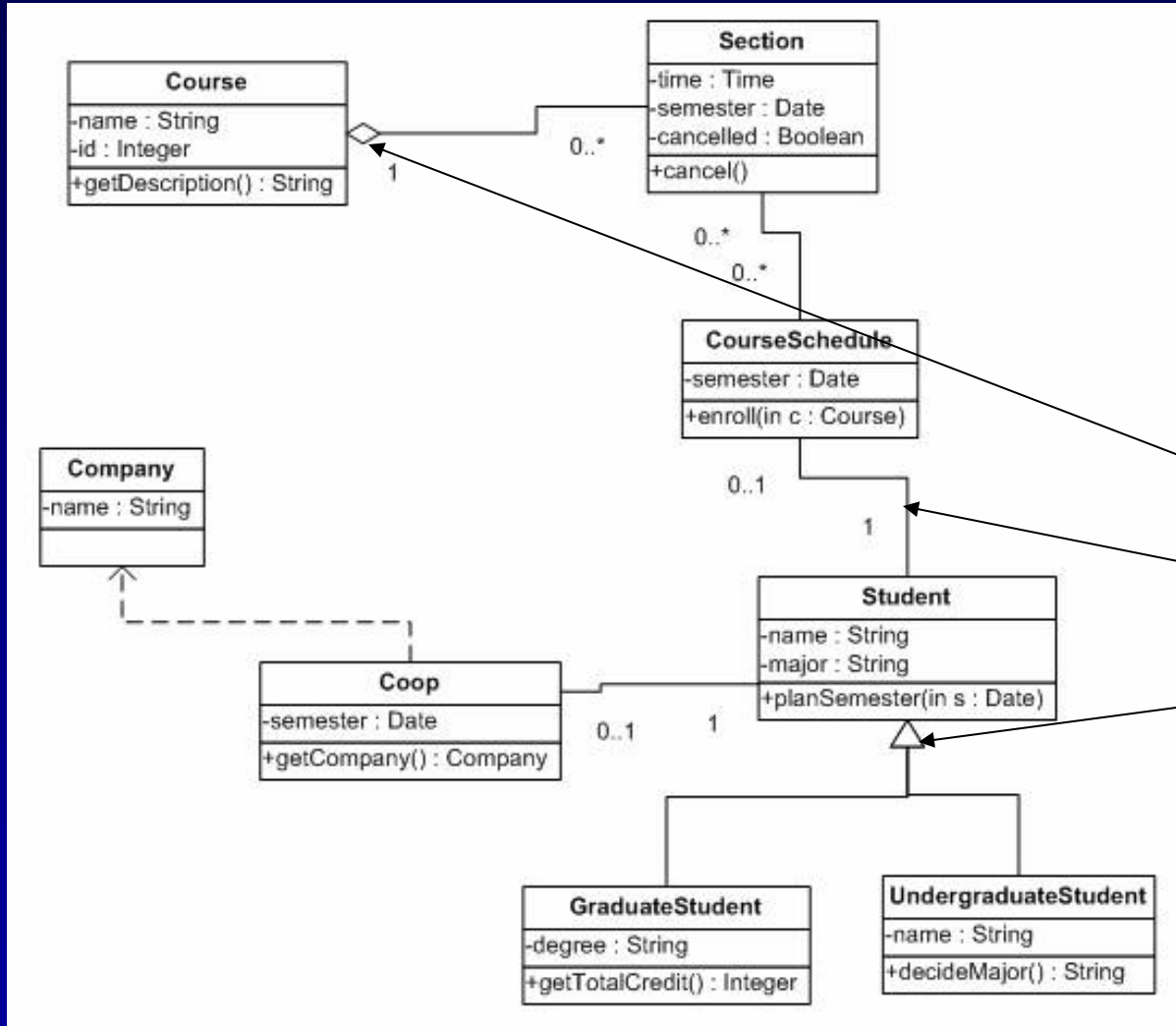


Class Diagrams

Generalization (Inheritance) relationship



Class Diagrams



Class name

Attributes

Operations

aggregation

association

generalization

UML Resources

- Software packages for class diagramming
 - Microsoft Visio or other drawing tools
 - ESRI's "Building Geodatabases with Visio"
- IBM's Rational Rose literature on design with UML
 - Introduction to UML
 - Entity-relationship modeling with UML
- UML tutorials on the web

Part 5.

5. How might we choose a logical data model?

Three types of logical data models common in GIS:

- Relational, e.g., as in ArcView shape file
- Object-oriented, e.g. as in GE Small World Systems
- Object-relational, e.g. as in Esri ArcGIS geodatabase

Compared by Tomlinson (2011) Figure 8.10. p. 107

Relational Data Model

- Developed based on mathematical simplicity
- Relation is a collection of attributes
- Simple name for relation implementation is “table”
- Relationships between tables are created through common attributes using a table join
 - Logical linkage: which keys needed?
- Common logical model in GIS as well as hybrid form (file for features + relational table for attribute): e.g. ArcGIS shapefile data model

Object-oriented (OO) Data Model

- A collection of object (complex, natural)
- Object has attributes and operations
- Class is a collection of similar objects
- Database can be built using class hierarchy (e.g. some attributes of city database can inherit from supper class such as spatial objects)
- Model complex relationships
 - Association: any relationship
 - Aggregation: weak form of whole-part relation
 - Composition: strong form of whole-part relation
 - Generalization: (e.g. secretary and boss are employees)
- For example, as in GE Small World Systems

Object-relational Data Model

- Hybrid of relational model and OO model, but in single model
- Extend relational data model so that advantages of OO (e.g. complex relationships) can be incorporated
- Maintain table with abstract data type
- Can use SQL
- For example, as in Geodatabase data model

Choosing a logical data model

- Relational data model can support processing that involves a simple conceptual design; relationships are derived at processing, commonly not stored; time consuming
- Problems requiring handling of complex relationships, integrity and versioning, OO or OR will work better because of stored relationships (e.g. facilities management network)
- Tomlinson (2011) p. 110 figure 8.11 presents a table suggesting which logical data model is appropriate for a given data modeling situation
- What are the implications for data modeling in your problem statement needs assessment? That is, why require two assignments be completed simultaneously?