# Design of Spatial Data Models

Master Program
Applied Geographic Information Technology
University of Salzburg

*Gerhard Belina*

# Before we start

**Class rules and principles**

# Scheduling

- Setup
  - Block the class in a ~bi-weekly frequency (every 2 weeks – with exceptions)
  - 3 hours class – from 15:00 – 17:00
  - Start: right at 15:00
- Class schedule
  - See table to the right
  - Bauteil 14, GI-Lecture (SC30OG1.107)
- Exam
  - 3 appointments
  - Written exam
  - Details will be provided in the classes in January

| Date | Start | End |
|---|---|---|
| 06.10.2023 | 15:00 | 17:45 |
| 20.10.2023 | 15:00 | 17:45 |
| 10.11.2023 | 15:00 | 17:45 |
| 17.11.2023 | 15:00 | 17:45 |
| 01.12.2023 | 15:00 | 17:45 |
| 15.12.2023 | 15:00 | 17:45 |
| 12.01.2024 | 14:45 | 17:45 |
| 26.01.2024 | 15:00 | 17:45 |

# Prerequisites for the class

- For this class you need to be able to:
  - work independently
  - speak and write English fluently
  - conduct research independently
  - Be critical – feel free to challenge me and your colleagues
  - Be motivated
- You should know scientific work techniques
- You should know about the basics of Geographic Information Systems and scientific modeling in general

# Some principles to be considered

- Be **on time** in order to check in for the class !!!
- Highly recommended to **join the classes** – you might miss important interactions
- Designed as an "**class-room experience**" (not as an online class)
- You need to contact me via E-Mail under:

  gerhard.belina@belina.at
- Missing a class
  - Get in contact with your colleagues to learn about the content
  - Check the content on black board
- **Be focused** – be not be distracted by chats / mails / calls / day dreaming …..
  - Distraction is a waste of time
- "hungry minds" assignments
  - Not mandatory
  - Give you the chance to practice what we discuss theoretically
  - Gives you the opportunity to reflect, if you are able to apply the theory in practice

# Before we start

***Disclaimer***

The course content has been augmented by the use of AI (mostly of type Generative Language Model) like ChatGPT / Bing Chat / Google Bard and others

# Overall course objectives

- Learn **how to** design a data model – in particular, a **spatial data model**
- Understand the **multidimensional challenge of geographic information** for the design of spatial data models
- Learn about the **standards** required for geospatial information modeling / spatial data models
- Get familiar with **tools** for data modeling (mostly via hungry minds assignment)
- Learn **data modeling notations**
- Learn the essentials of **XML** and **XML Schema** definitions
- Develop **strategies to implement data models**
- Learn the essential **concepts and underlying terminology**

Note: this is done in context of a “larger scale challenge” with regards to spatial data modelling (enterprise / nation / supranational organizations / …)

# Course Agenda

- Revisit modelling principles
- Purpose of a data model
- Understanding Spatial Data / GIS Data
- How to design a (spatial) data model
  - assessment of data requirements / layers of abstraction / data modelling techniques / ER data model / Spatial Data Base Management System
- Understanding data modeling notations
- Important standards for geospatial information modelling – a short intro
- UML (unified modeling language) and designing spatial data models by applying geospatial standards
- XML and encoding spatial data in accordance with geospatial standards
  - Geography Markup Language (GML) and the general feature model
- Spatial Databases and Spatial Data Mining

# Essential modelling principles

Design of Spatial Data Models

# Course objectives of first class

- modeling theory and their application to data modeling
- Understanding spatial data structures
- Main concepts and terminology
- The modeling stack – layers for abstraction
- Assessment of data requirements
- Identification of information entities
- Main data modelling notations

# Important concepts and terms you need to understand

# Can you explain….?

M O D E L

# Possible answers

- ?

# answers from previous class

- An abstract representation with some flaws
- Capture some aspects of reality
- Capture aspects of reality with simplicity
- Way to understand complex phenomena
- Systematic approach to understand what  is going on
- Explain things in the real world

# Scientific Model

A ***scientific model*** seeks to represent ***empirical objects***, ***phenomena***, and ***physical processes*** in a logical and objective way.

All models are in simulacra, that is, ***simplified reflections*** of ***reality***, but, despite being approximations, can be extremely useful.

**Building and disputing models is fundamental to the scientific enterprise**.

Complete and true representation may be impossible, but scientific debate often concerns *which is the better model for a given task*, e.g., which is the more accurate climate model for seasonal forecasting...

(see.: http://en.wikipedia.org/wiki/Scientific_model; visited Oct. 2023)

# Properties of a scientific model

- testable assumptions
  - Models are built on a set of assumptions about how the system behaves. These **assumptions must be clearly stated and are subject to scrutiny and testing**. If the assumptions are flawed or unrealistic, the model's predictions may not align with reality.

- variability and uncertainty
  - scientific models often account for variability and uncertainty within a system. They may incorporate probability distributions or uncertainty intervals to **reflect the inherent unpredictability of some phenomena**. This allows scientists to understand the range of possible outcomes.

# Purpose of a scientific model

- Hypothesis testing / experimentation
  - Comparison of model prediction with real data

- Making predictions
  - if the model accurately represents the real-world system, it should produce predictions that align with observed phenomena. Predictive power is a crucial test of a model's validity.

# Can you explain….?

**M O D E L L I N G**

# Can you explain modeling - Interaction

Possible answers

- tbd

# Answers from previous class

- Collecting data (randomly)
- Analyzing
- Identification of common aspects
- Represent and improve data
- Re-organize and structure information
- Definition of the purpose of the model
- Development of a theory
- Process of simplication with logical rules
- Improve the accuracy of the model

# Modeling

***Modeling*** is the process of generating a model as a **conceptual representation** of some phenomenon.

Typically, a model will deal only with **some aspects of the phenomenon** in question, and two models of the same phenomenon may be essentially different…

…Such differences may be due to **differing requirements** of the model's end users, or to **conceptual or aesthetic differences** among the modelers and to contingent decisions made during the modelling process…

(see: https://en.wikipedia.org/wiki/Scientific_modelling#Generating_a_model; visited Oct. 2023)

# Modeling process

- Abstraction
- Simplification
- Purpose oriented / dependent on requirements
- Domain oriented
- Iterative (iterative refinements)

**Important to note:**

Errors, inaccuracies might be tolerable in case they do not challenge the „**fitness for purpose**“ of the model.

# Can you name….?

**Examples of models**

# You might have heard of

- Clima models
- Run off models
- Meteorological models
- Land slide models
- And many more
- Business process models
- Graph model
- Metamodel
- Digital Elevation model
- Data model
- Logcial model
- Requirements model
- Physical model
- Conceptual model

# Commonly used categories for models

- Purpose or function
  - Predictive / explanatory / descriptive
- Mathematical nature
  - Deterministic / stochastic
- Complexity
- Temporality
  - Static / dynamic
- Domain / Application
  - Meteorological models, land slide model
- …
- For the IT- Domain
  - Business Process Model, Requirements Model, Meta Model, Data Model, …

# (Spatial) Data Model

**Design of Spatial Data Models**

# Data model – concept definition

"A ***Data model*** is an abstract model that organizes *elements of data* and standardizes how they *relate* to one another and to *properties* of the real-world entities.

**A data model explicitly determines the *structure* of data.**

- The term data model is used in two distinct but closely related senses. Sometimes it refers to an abstract formalization of the objects and relationships found in a particular application domain, for example the customers, products, and orders found in a manufacturing organization.
- At other times it refers to a set of concepts used in defining such formalizations: for example concepts such as entities, attributes, relations, or tables. So the "data model" of a banking application may be defined using the entity-relationship "data model".

*(see: https://en.wikipedia.org/wiki/Data_model, Oct. 2023)*

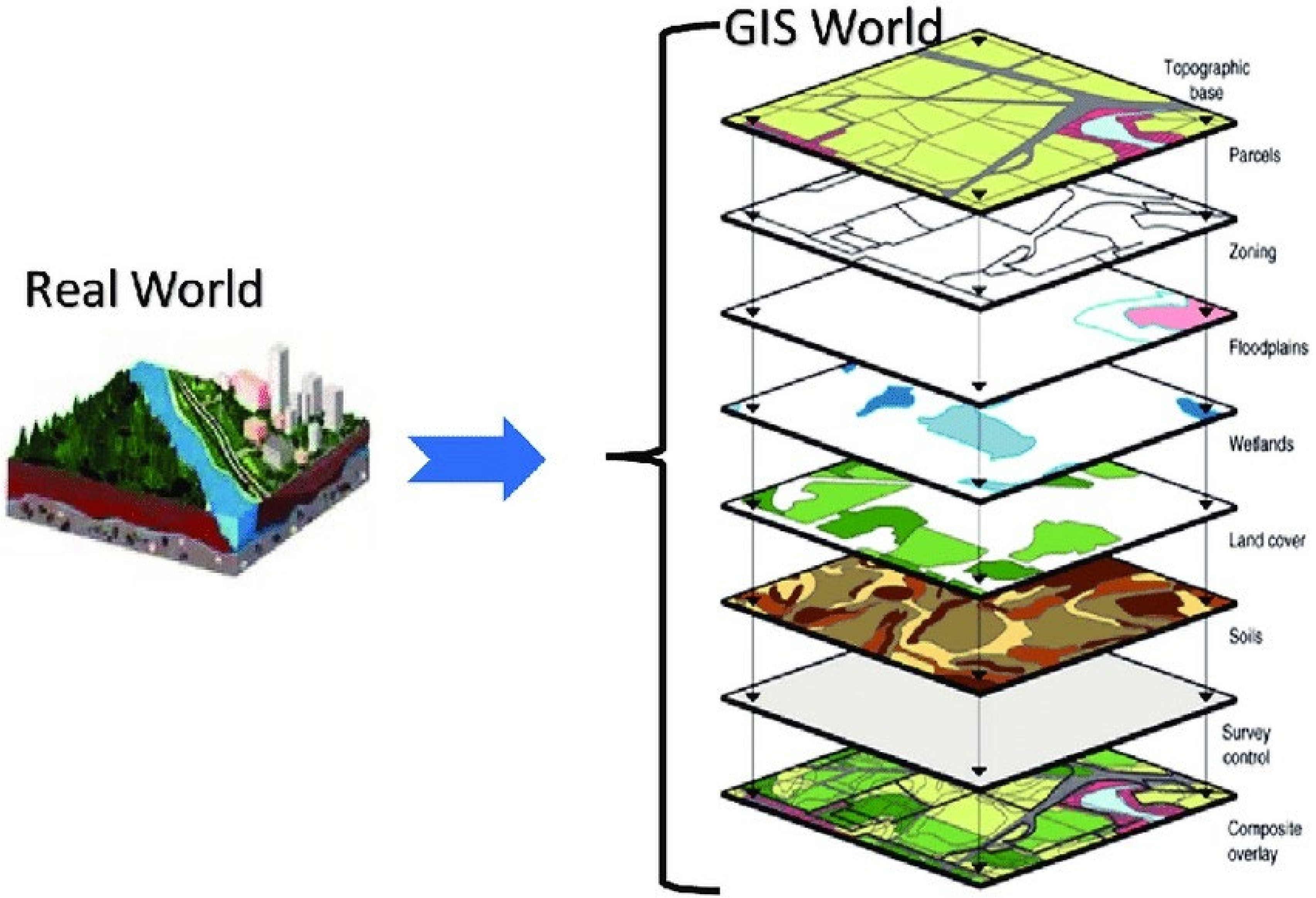
- typically used in the fields of database management, information systems, and computer science

# Why do we require a data model?

Purpose of a data model:

- Support development of information systems
- guides the way data is accessed, manipulated, and queried by information systems
- ensures the provision and analysis of data in form of
  - Reports / dashboards
  - Prognosis models
  - Enhanced data analytics / data mining
- facilitates data exchange

# Real world representation in geographic information systems (GIS)



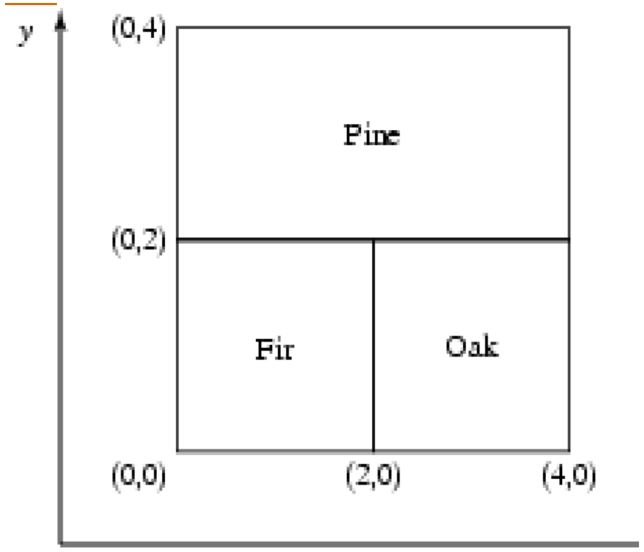
# Spatial Data Model – Spatial Data Types

A data model in geographic information systems is a mathematical construct for representing geographic objects or surfaces as data.

- the vector data model represents geography as collections of **points**, **lines**, and **polygons** -> spatial data described as objects
- the raster data model represent geography as **cell matrices** that store numeric values -> spatial data described as mathematical field
- the Graph data use nodes and edges to represent relationships between objects in space.

*(see: https://en.wikipedia.org/wiki/Data_model#Geographic_data_model , Oct. 2023 with adoptions)*

# Spatial data modeling – State Park forest cover




| Area-ID | Dominant Tree Species | Area/Boundary |
|---|---|---|
| FS1 | Pine | [(0,2),(4,2),(4,4),(0,4)] |
| FS2 | Fir | [(0,0),(2,0),(2,2),(0,2)] |
| FS3 | Oak | [(2,0),(4,0),(4,2),(2,2)] |

Field Viewpoint of Forest Stands

$$f(x,y) = \begin{cases} \text{"Pine,"} & 2 \leq x \leq 4 ; 2 < y \leq 4 \\ \text{"Fir,"} & 0 \leq x \leq 2; 0 \leq y \leq 2 \\ \text{"Oak,"} & 2 < x \leq 4; 0 \leq y \leq 2 \end{cases}$$

Object based: forest cover described as vector model in terms of polygons (using a specific semantic)

Flow based: forest cover described as raster model in terms a function

# Requirements for spatial data models

- Efficiency: The data model should be efficient in order to retrieve and process data quickly and easily
- Accuracy: The data model should accurately represent the spatial component of the data.
- Flexibility: The data model should be flexible enough to meet the requirements of a wide range of applications.
- Interoperability: The data model should be compatible with other data models and applications.

# Spatial Data Model – Spatial Data Types - interaction

Can you mention scenarios / applications – when

- modelling spatial data in form of raster is more suitable?
- modelling spatial data in form of vectors is more suitable?

Think about the requirements for the data model => what purpose should the model serve?

# Spatial Data Model – Spatial Data Types - interaction

Notes of interaction:

- tbd

# Raster / Vector – typical pro's and con's

- Raster:
  - Simple data structure
  - Efficient for most analytical operations
  - Many source data (esp. sensing data) is available in raster – easy and efficient data entry usage
  - Less compact
  - Query based analysis difficult
  - Coarse
  - Typically, only one-dimensional attribution

# Raster / Vector – typical pro's and con's

- Vector:
  - compact data structure for storing many attributes
  - Efficient combination of standard queries with spatial query operations
  - Very accurate spatial representation
  - Complex data structure
  - Overlay operations computationally intensive
  - Not suited for data with high degree of variability (e.g. digital elevation models and similar)

# Observations about accuracy / precision

## Conversions and errors:

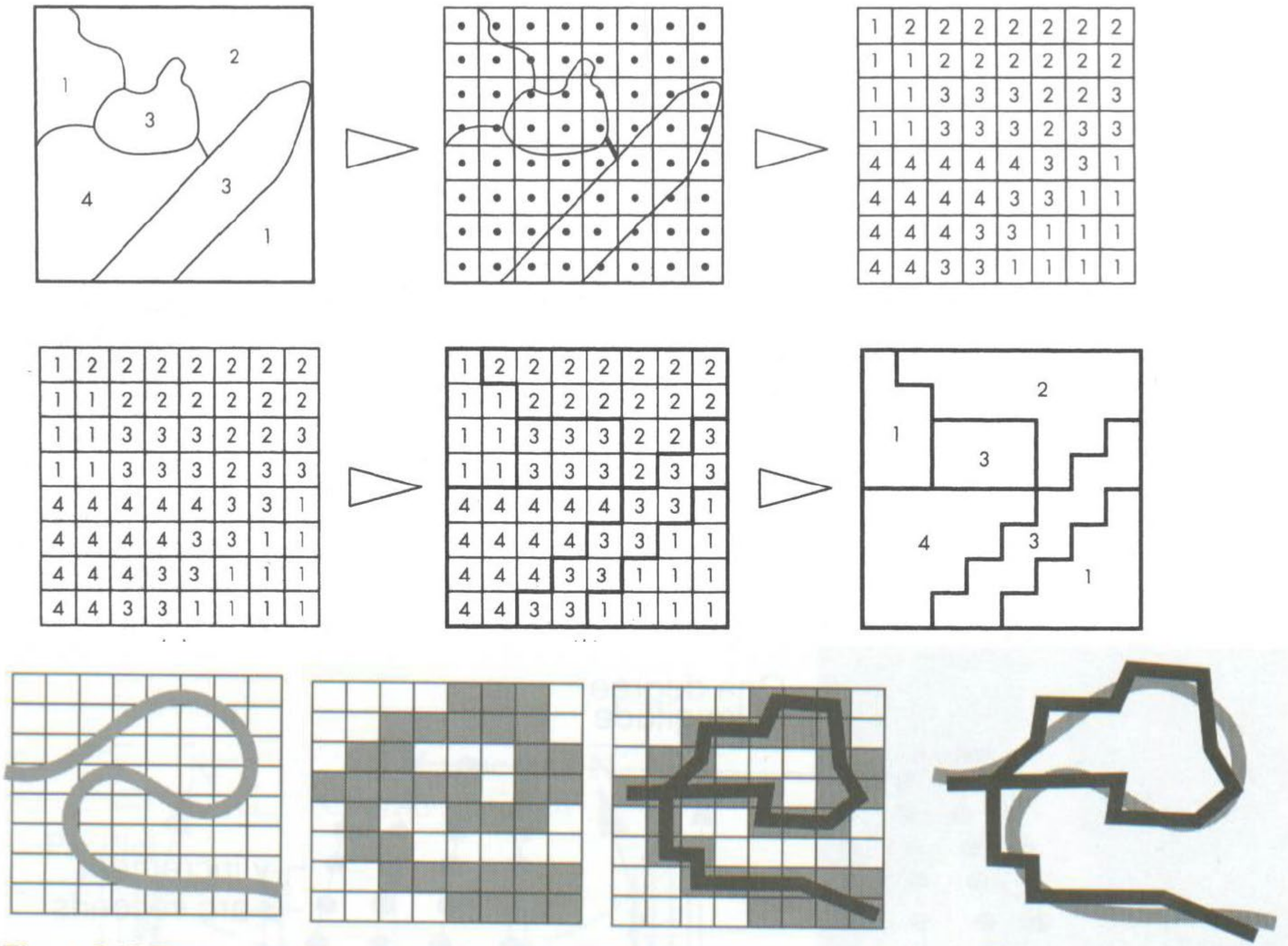


**Figure 3.19** Errors caused by exchanging data between raster and vector formats. The original (gray) river after raster-to-vector conversion appears to connect the loop back.