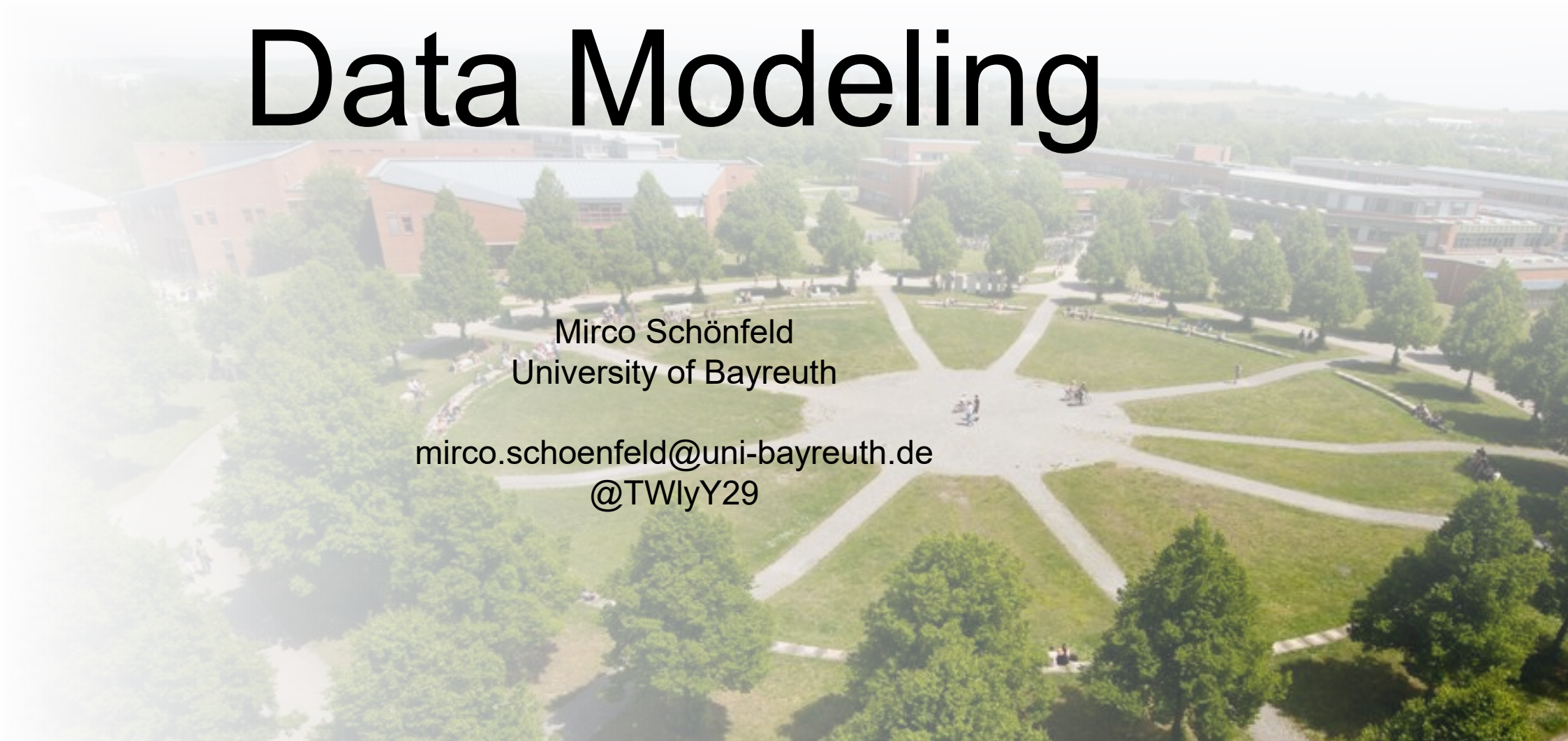
UNIVERSITÄT BAYREUTH

Data Modeling



Mirco Schönfeld
University of Bayreuth

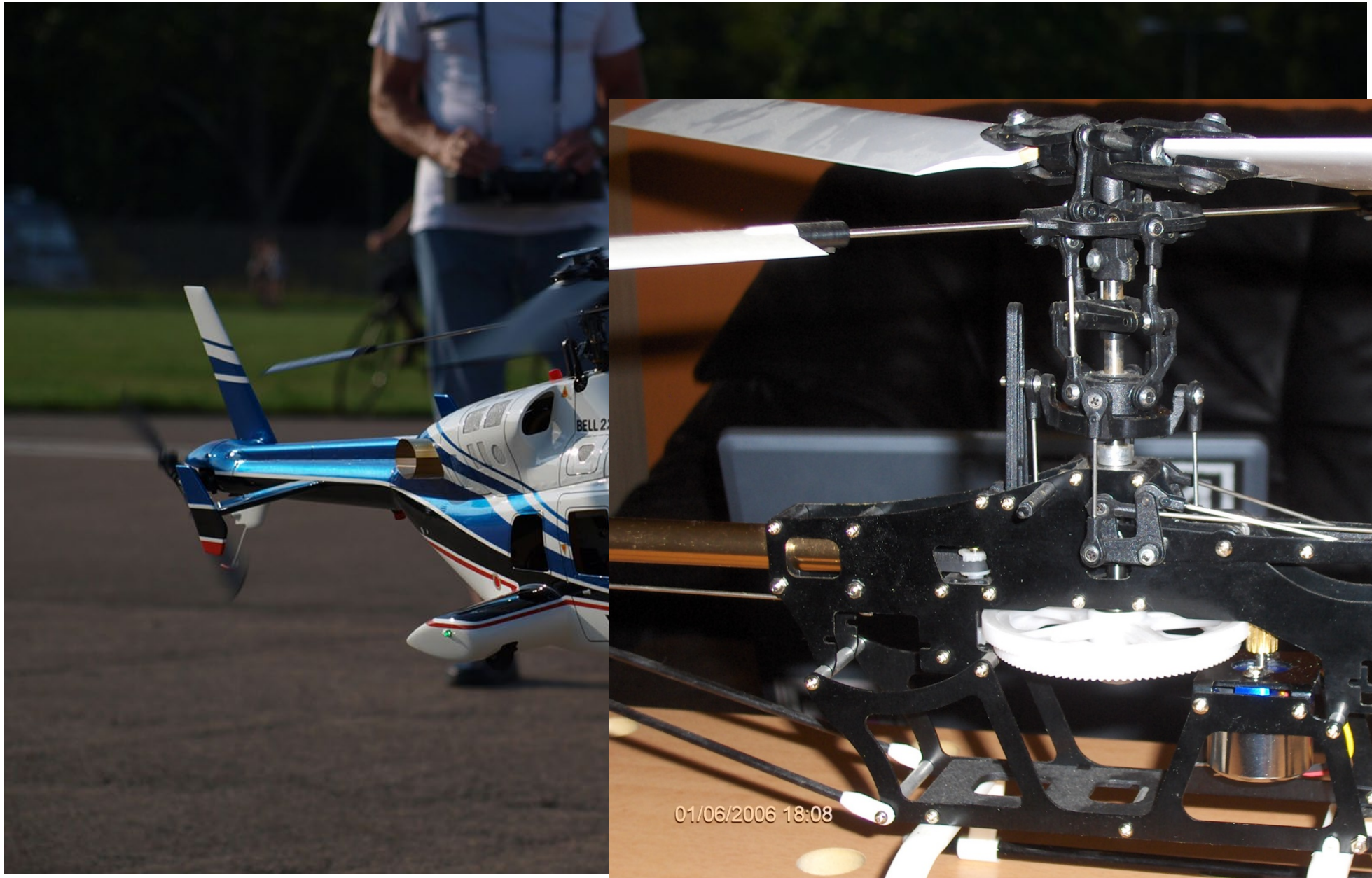
mirco.schoenfeld@uni-bayreuth.de
@TWlyY29

https://en.wikipedia.org/wiki/U.S._Army_Corps_of_Engineers_Bay_Model

https://en.wikipedia.org/wiki/U.S._Army_Corps_of_Engineers_Bay_Model

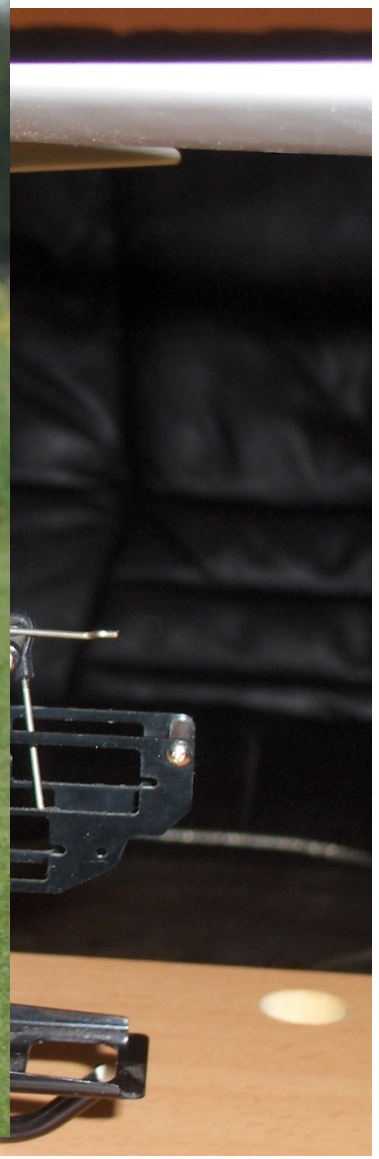
https://de.wikipedia.org/wiki/Modellhubschrauber

UNIVERSITÄT BAYREUTH

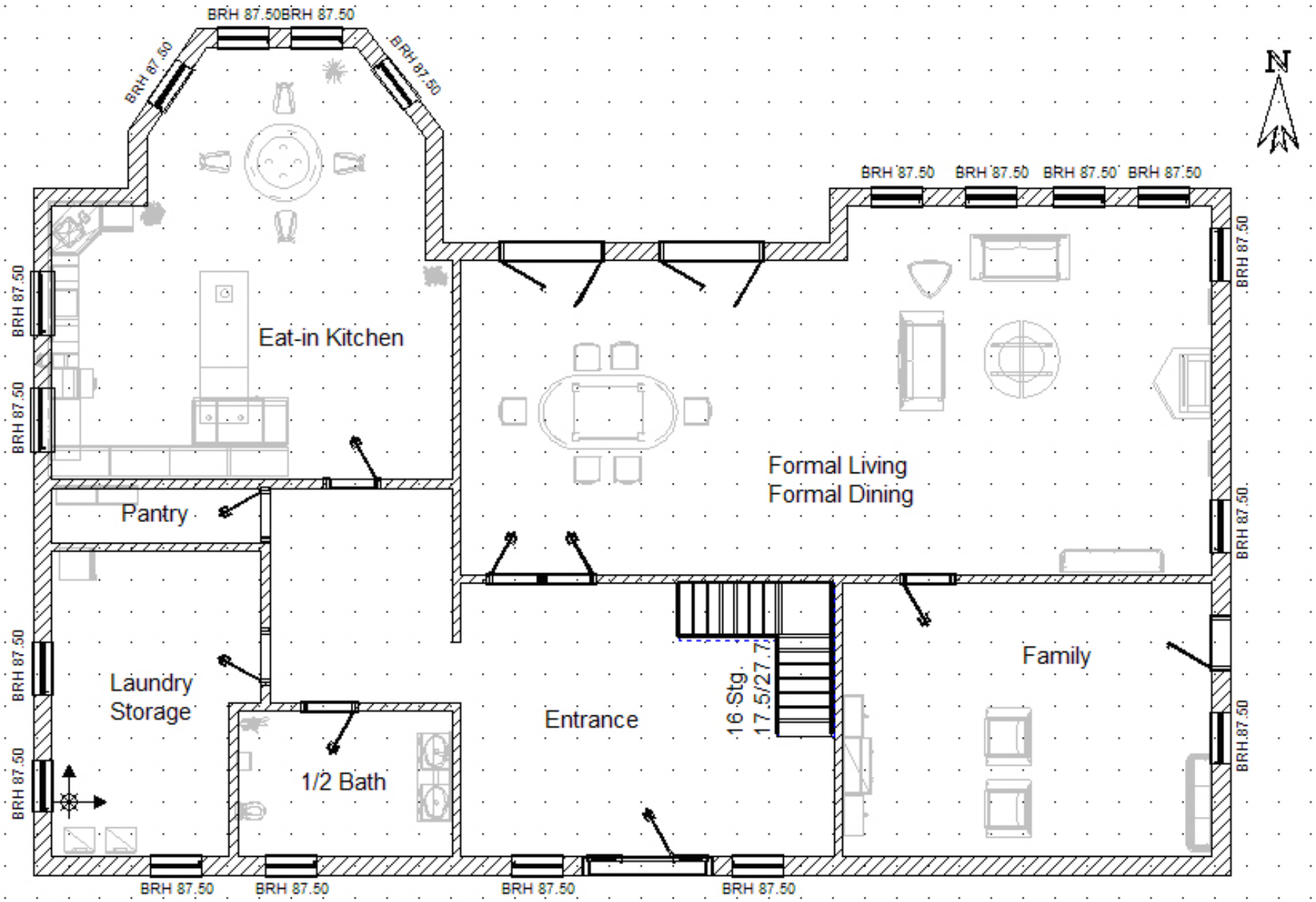


https://de.wikipedia.org/wiki/Modellhubschrauber

UNIVERSITÄT BAYREUTH

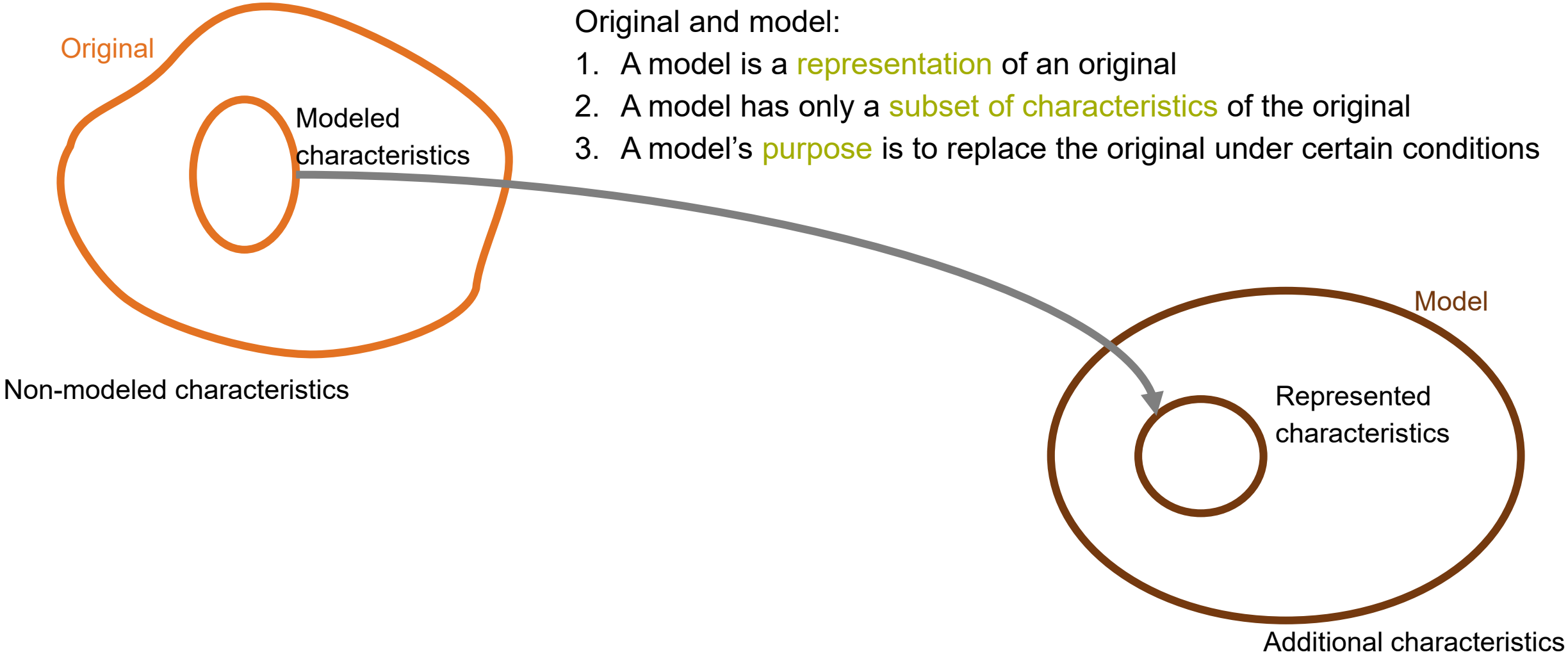


https://de.wikipedia.org/wiki/Modellhubschrauber



https://en.wikipedia.org/wiki/Floor_plan

Models in General

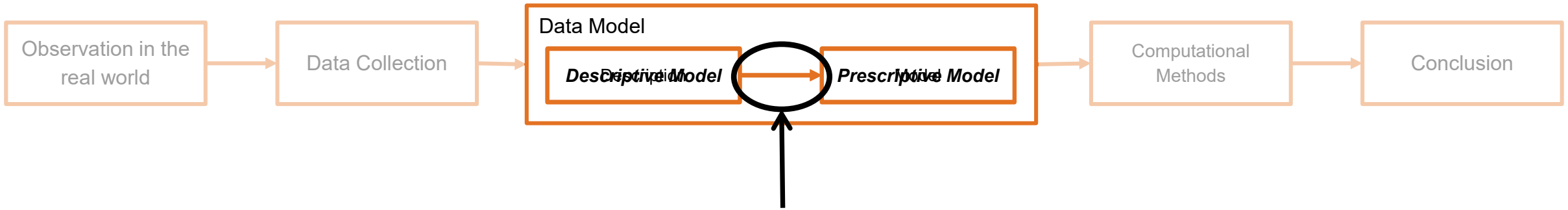


Original and model:

1. A model is a representation of an original
2. A model has only a subset of characteristics of the original
3. A model's purpose is to replace the original under certain conditions

Stachowiak, H. (1973) Allgemeine Modelltheorie, Wien, New York: Springer.
Ludewig, J. (2002). Modelle im Software Engineering - Eine Einführung und Kritik. In Modellierung in Der Praxis - Modellierung für die Praxis (pp. 7–22). Modellierung 2002. GI.
Hesse, W. (2006). More matters on (meta-)modelling: Remarks on Thomas Kühne's "matters". Software & Systems Modeling, 5 (4), 387–394.

Turning Descriptions Into Models



Descriptive Model

O▶M – “O is modeled by M”

- Original is *prior* to the model
- Model describes original

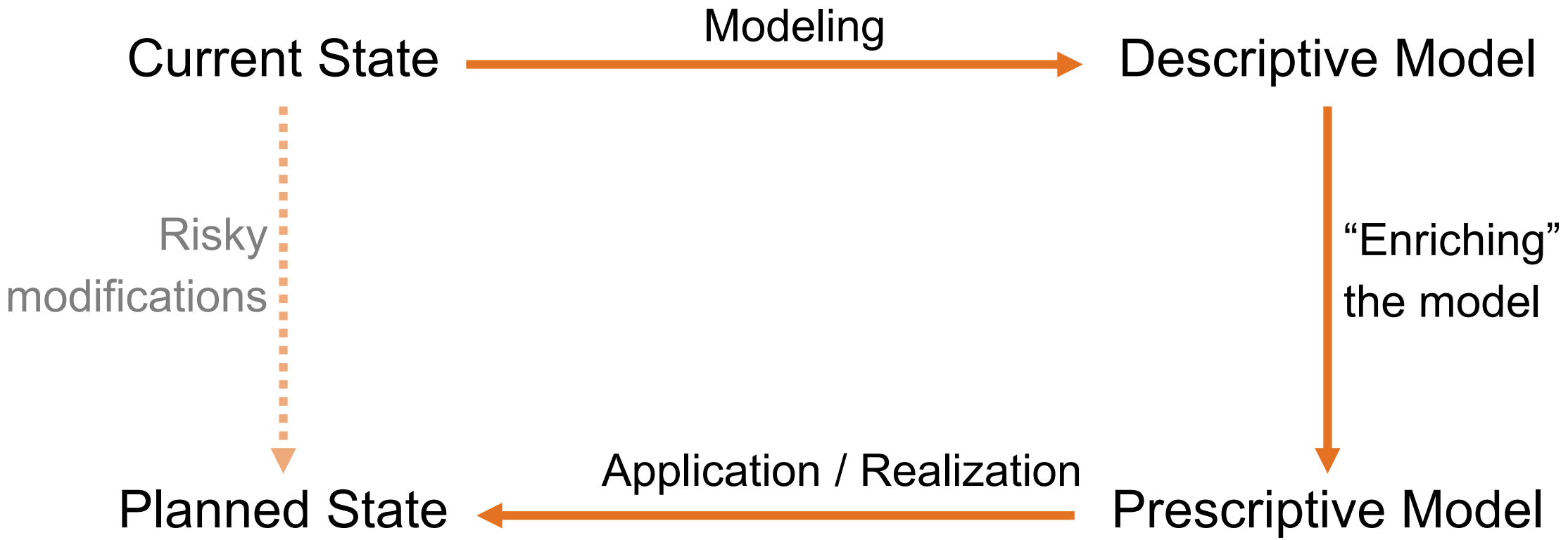
Prescriptive Model

M◀O – “M is concretised to O”

- “Original” exists *after* the model
- Model prescribes an original

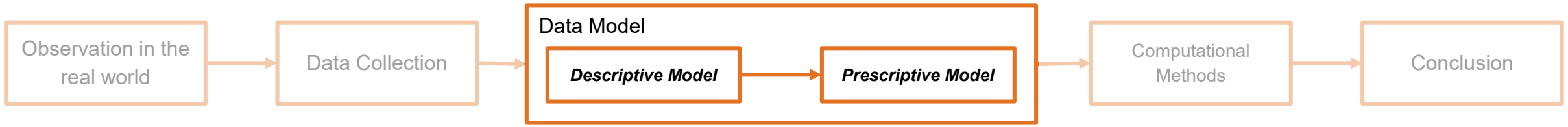
Stachowiak, H. (1973) Allgemeine Modelltheorie, Wien, New York: Springer.
Ludewig, J. (2002). Modelle im Software Engineering - Eine Einführung und Kritik. In Modellierung in Der Praxis - Modellierung für die Praxis (pp. 7–22). Modellierung 2002. GI.
Hesse, W. (2006). More matters on (meta-)modelling: Remarks on thomas kühne’s “matters”. Software & Systems Modeling, 5 (4), 387–394.

Descriptive vs. Prescriptive Models



Ludewig, J. (2002). Modelle im Software Engineering - Eine Einführung und Kritik. In Modellierung in Der Praxis - Modellierung für die Praxis (pp. 7–22). Modellierung 2002. GI.
Flanders, J. and Jannidis, F., 2018. 1 Data modeling in a digital humanities context. The Shape of Data in Digital Humanities: Modeling Texts and Text-based Resources.

Turning Descriptions Into Models



A proper data model presents an understandable meaning for data and offers a common vocabulary to talk about it.

Formal Modeling

Models for Modeling?

UNIVERSITÄT BAYREUTH

A model by is not processable by machines by itself

In order to be processable, models

- must be represented in an unambiguous and explicit language and
- must be represented in a language that represents the salient features of the model

We Need a Set of Rules

Formal models use a specific set of rules

- that explicitly and exhaustively define the model's syntax and semantics, and
- that allows the data to be processed automatically

Formal models

- are the building block for communication about data
- ensure a higher data-quality because they impose certain requirements on input data
- allow for a seamingless exchange or merging of data
  (if they follow certain standards)

Different Formal Models

UNIVERSITÄT BAYREUTH

Process Modeling	System Modeling	Data Modeling
Modeling of events in time: the amount of water passing through a river bed in a given time, the change of employment resulting from specific events, or the spread of a new scientific concept in scientific texts Also includes simulation, i.e. modeling of more complex processes	Design of software systems, usually an abstract view of the design of a piece of software	Modeling of entities: documents, events, information systems, agents, data sets

Different Formal Models

UNIVERSITÄT BAYREUTH

Process Modeling

Modeling of events in time:
the amount of water passing through a river bed in a given time, the change of employment resulting from specific events, or the spread of a new scientific concept in scientific texts

Also includes simulation, i.e. modeling of more complex processes

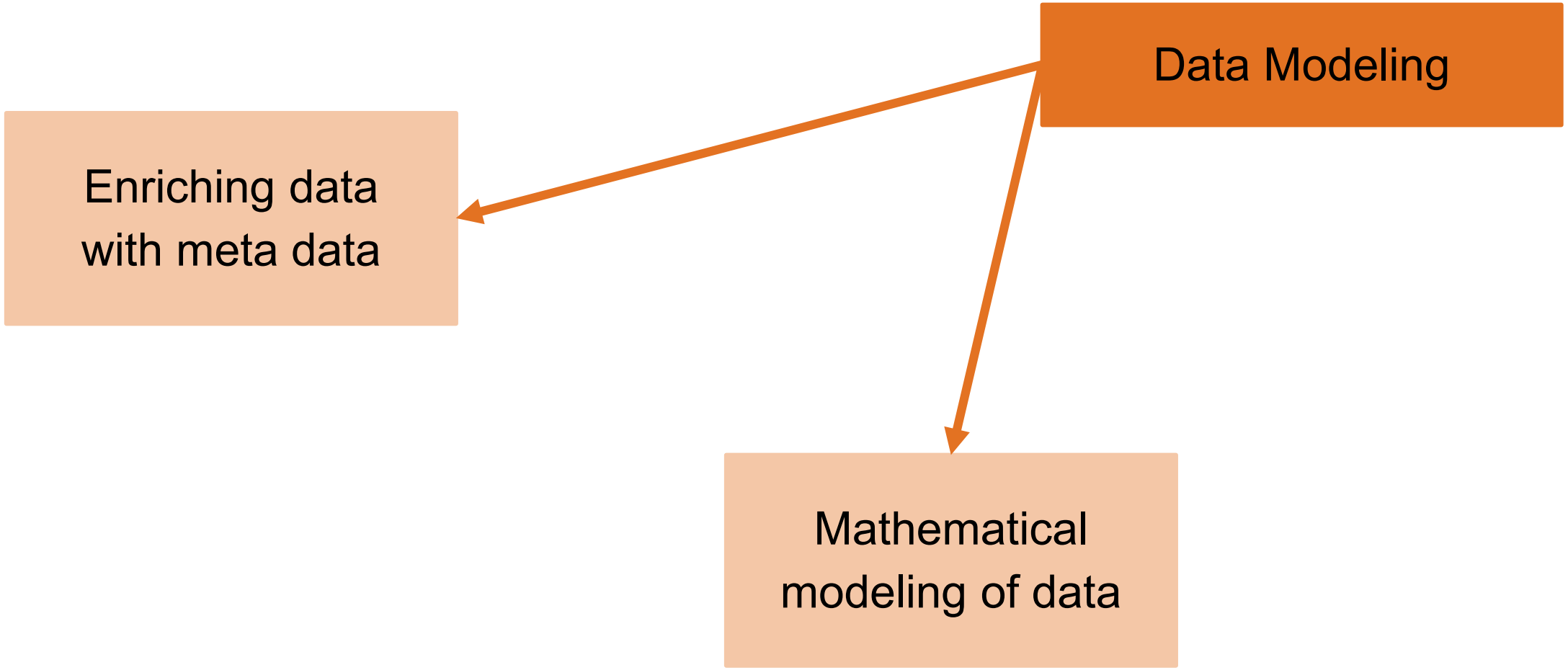
System Modeling

Design of software systems, usually an abstract view of the design of a piece of software

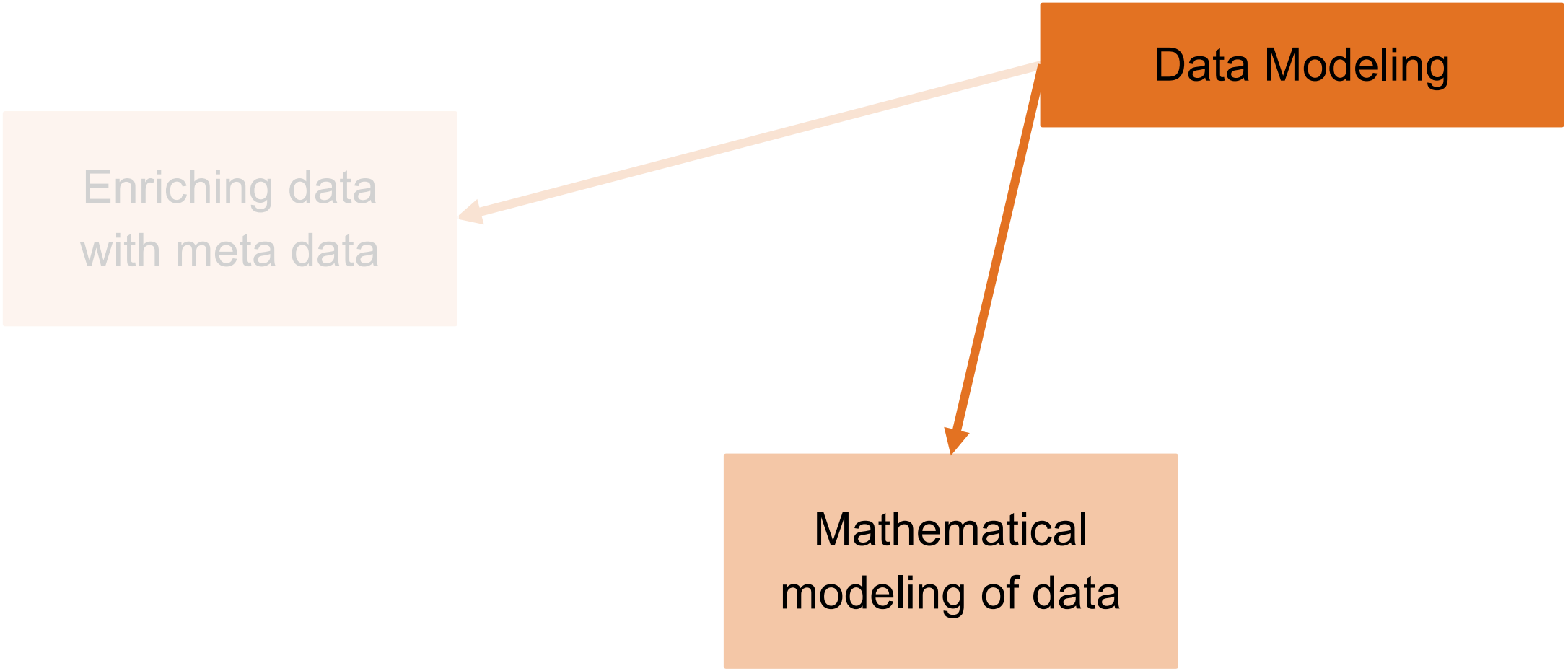
Data Modeling

Modeling of entities:
documents, events, information systems, agents, data sets

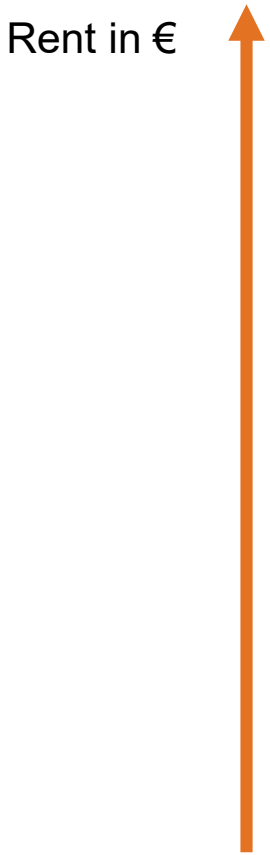
Modeling Data Formally



Modeling Data Formally

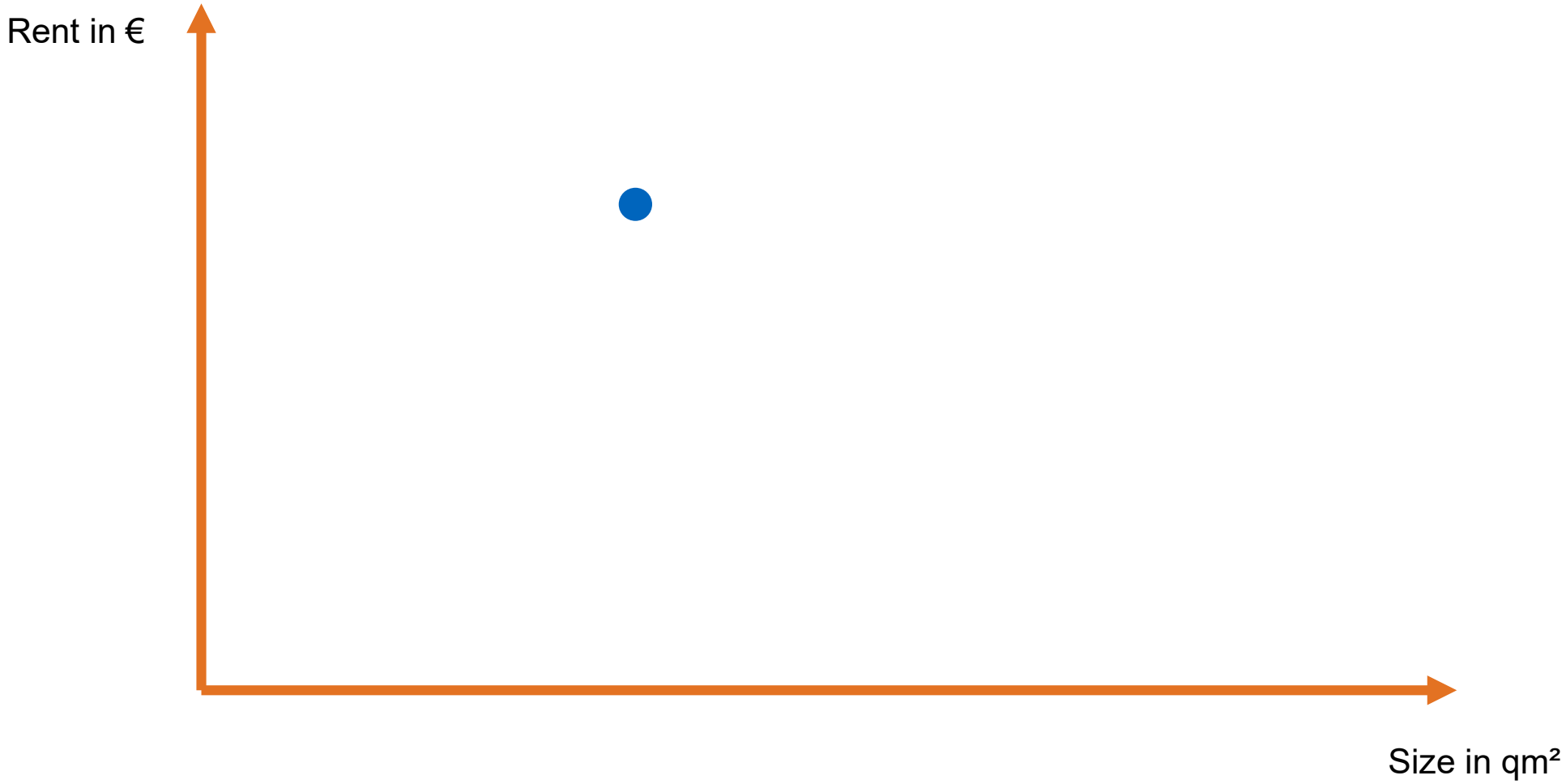


Mathematical Modeling

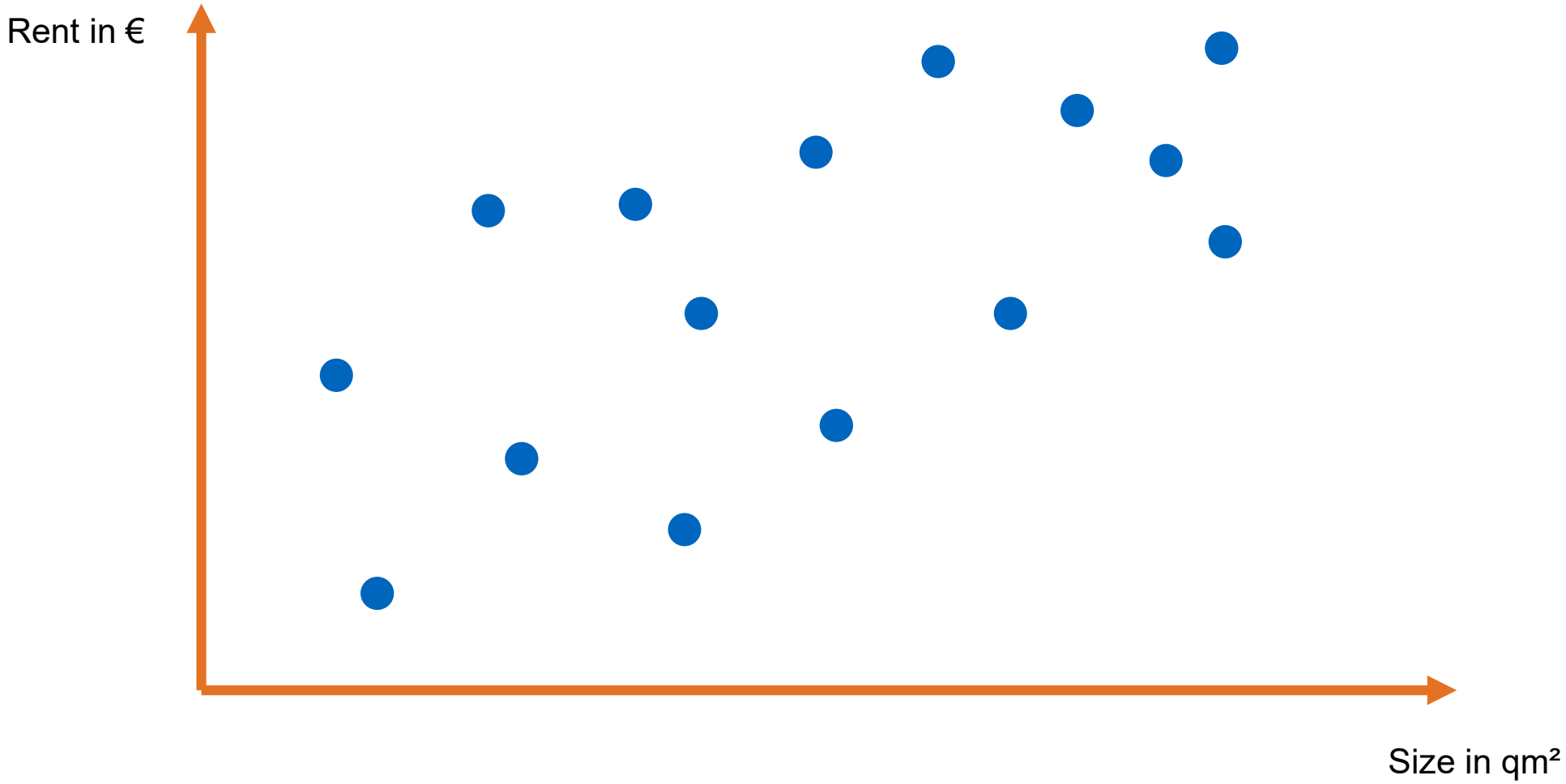


Mathematical Modeling

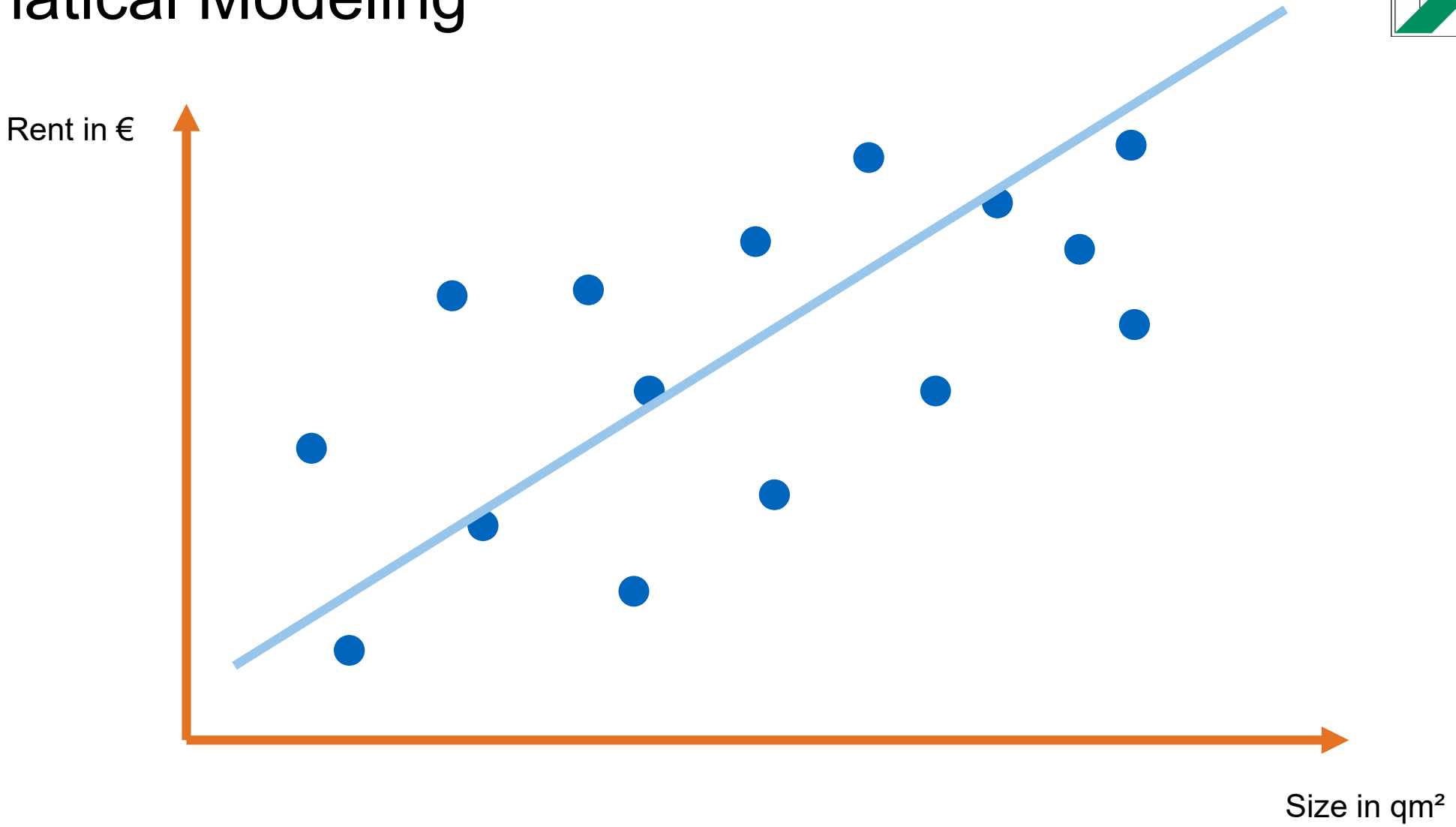
Mathematical Modeling



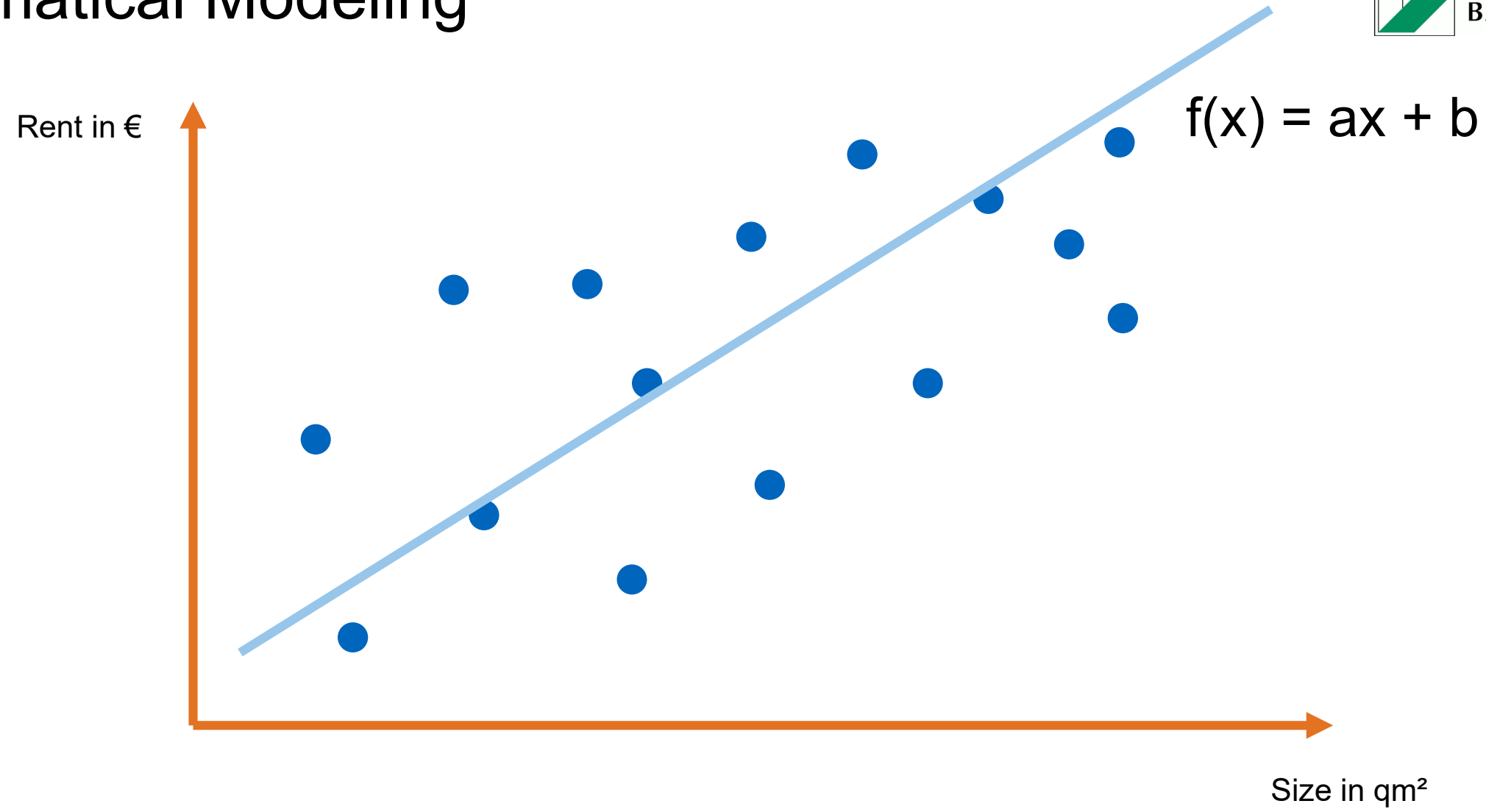
Mathematical Modeling



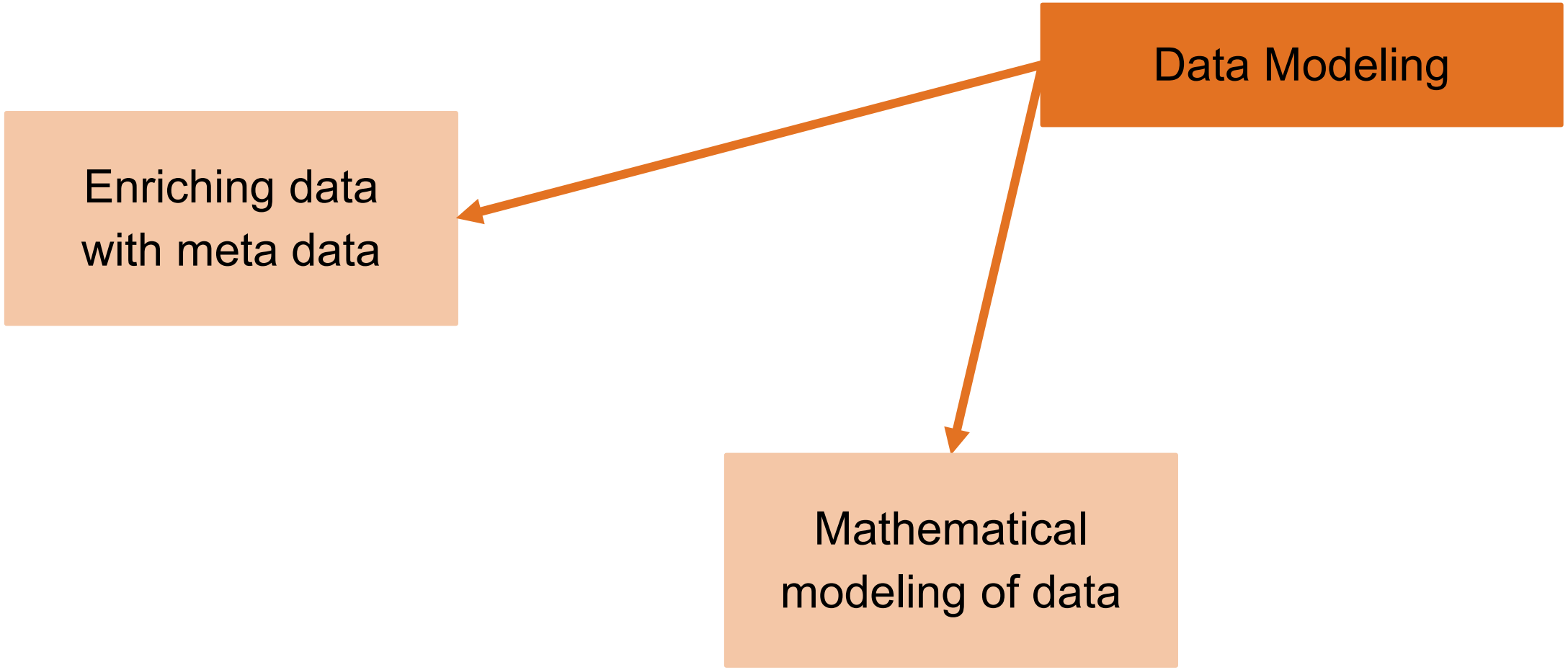
Mathematical Modeling



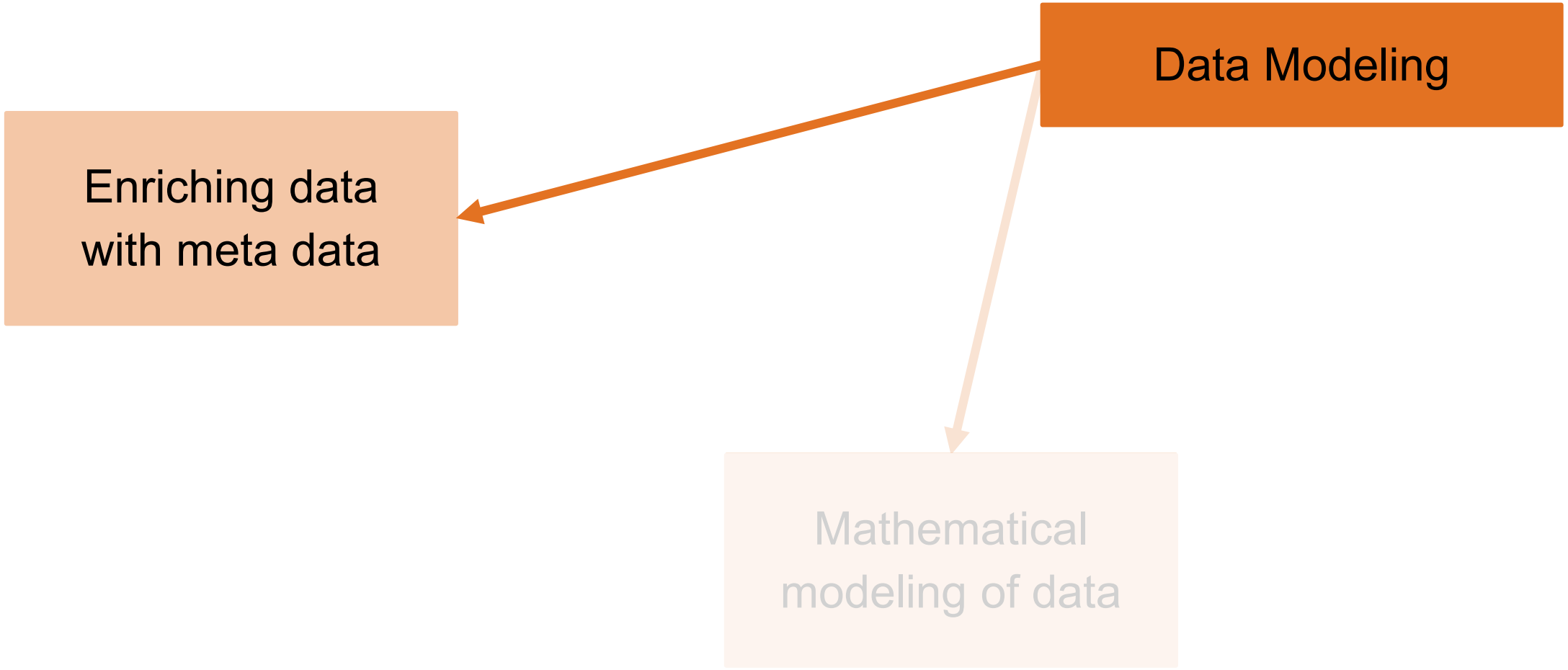
Mathematical Modeling



Modeling Data Formally



Modeling Data Formally



Data About Data

Meta data is data about modeled instances.

Meta data is used for discovery, query, and management of data

Meta data is highly structured inform

More about meta data later!

Meta data is kept

- in separate databases, or
- bundled together with the data but clearly distinguishable from it

Modeling Data…for what?

UNIVERSITÄT BAYREUTH

Research-driven Models

Modes are created for specific research questions

Operationalizing research questions and making it “tangible”

Tailored to specific question. Often useless for others.

Curation-Driven Models

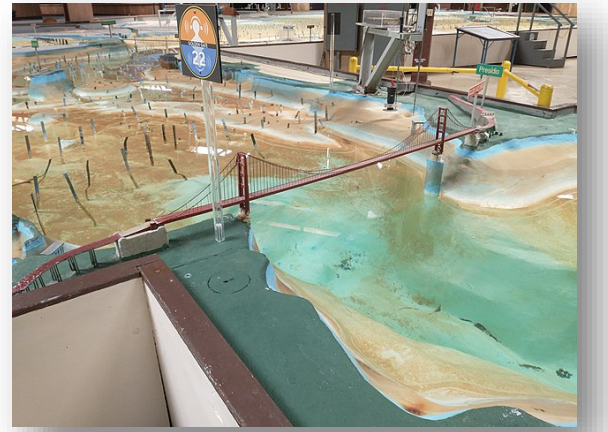
Modeling is meant to make digital resources sustainable and lasting.

Used by libraries and archives or cooperations among researchers

Summary

Characteristics of models:

1. A model is a representation of an original
2. A model has only a subset of characteristics of the original
3. A model's purpose is to replace the original under certain conditions



Descriptive Model

- Original is *prior* to the model
- Model describes original

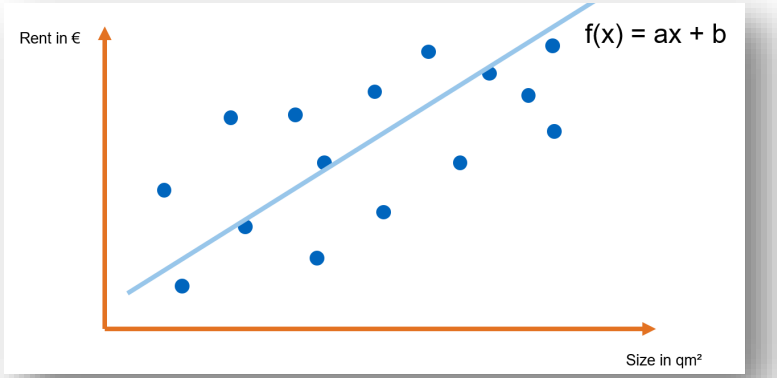
VS

Prescriptive Model

- "Original" exists *after* the model
- Model prescribes an original

Formal models use a specific set of rules

- that explicitly and exhaustively define the model's syntax and semantics, and
- that allows the data to be processed automatically



Conceptual Modeling

Why again?

UNIVERSITÄT BAYREUTH

By data modeling we try to find a translation of real-world situations to data & databases

Data models enable a user to define the data using high-level constructs without worrying about many low-level details of how data will be stored on disk

Data modeling can be viewed as series of steps with the ultimate goal of meeting a set of requirements specified by future users of the data, or people who act on their behalf.

Aim of Data Modeling

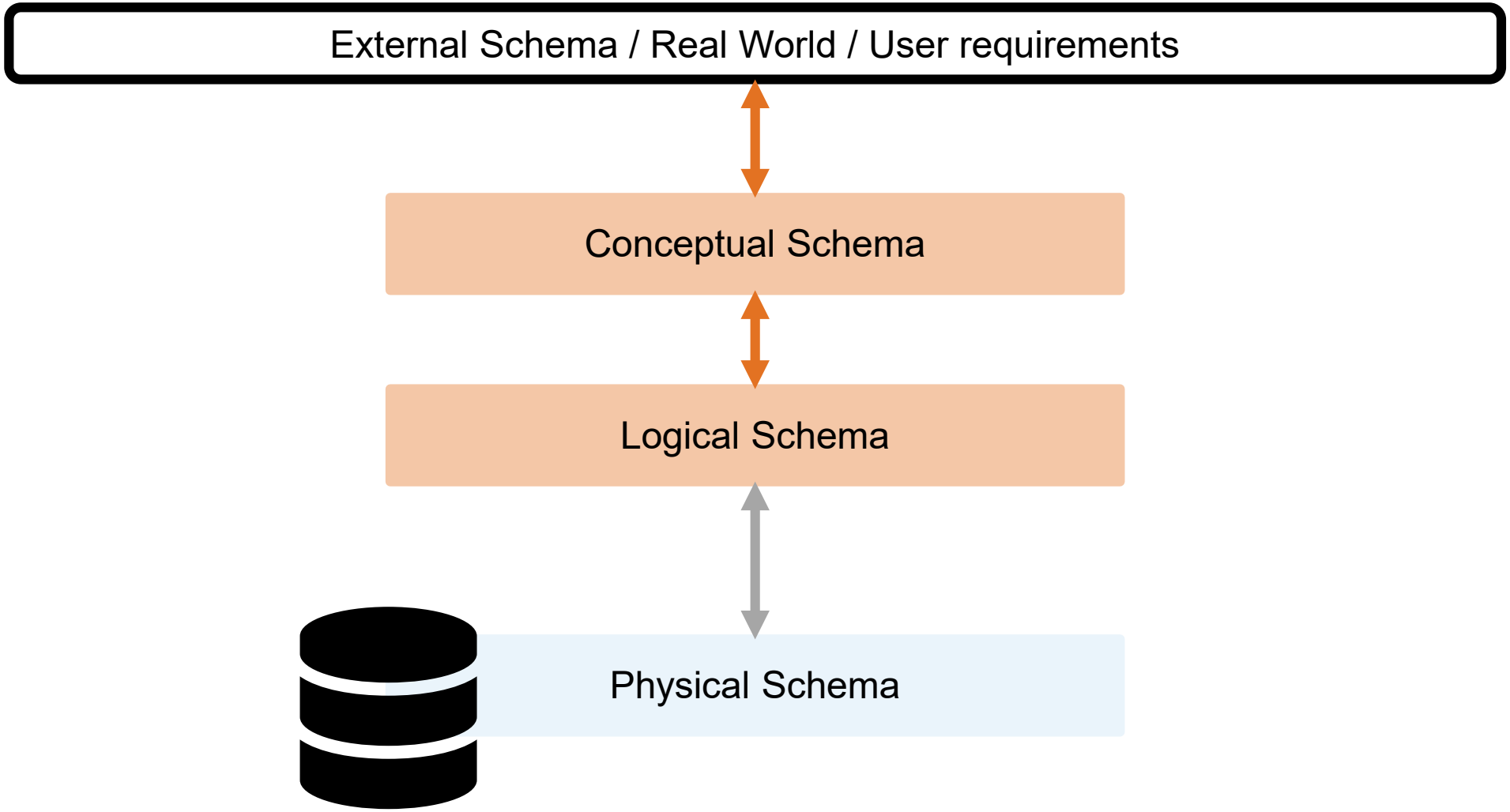
Fulfil *user requirements*!

Intended usage of a digital entity is the single most important factor determining

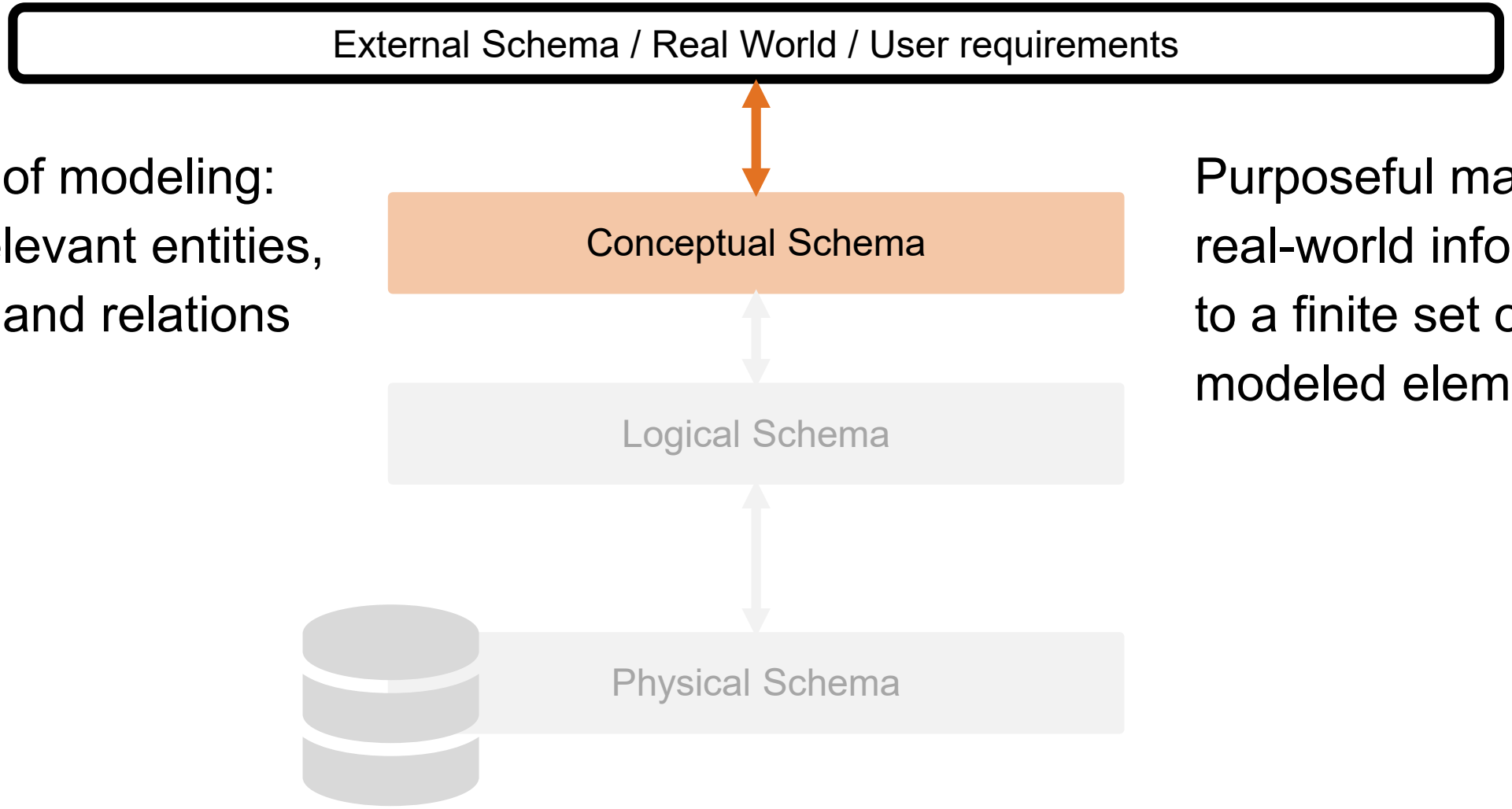
- the selection,
- the amount and depth of the annotations
- the complexity and
- richness of the data model.

Clear analysis of the requirements of the digital entities
in question is an important step in data modeling.

Levels of Abstraction



Levels of Abstraction



First step of modeling:
Identify relevant entities, attributes and relations

Purposeful mapping of real-world information to a finite set of modeled elements.

The Conceptual Data Model

UNIVERSITÄT BAYREUTH

- Classification

  Fix object types (entities)
- Abstraction

  Identify relevant characteristics (attributes)
- Relations

  Describe relations between objects
- Identification

  Chose unique identifiers (keys)

Modeling a scenario

UNIVERSITÄT BAYREUTH

Modeling a scenario

Object types:
- Lectures
- Students
- Research assistants
- Professors

Attributes:
- Lectures: Lecture number
- Lectures: Titel & ECTS
- Students: Student ID
- Students: Semester
- Employees: Staff number
- Employees: Room

Relations:
- Students LISTEN to lectures
- Professors GIVE lectures
- Professors TEST students
- Assistants WORK FOR profs
- Lectures REQUIRE other lectures

Identifiers:
- Students: student ID
- Employees: Staff number

Conceptual Entities and Relations

UNIVERSITÄT BAYREUTH

An **entity-relationship model** (or **ER model**) describes interrelated things of interest in a specific domain of knowledge.

Source: wikipedia

ER models contain

- Entities

  An entity may be defined as a thing capable of an independent existence that can be uniquely identified.
  Exists either physically or logically.
  Can be thought of as nouns.
- Relations

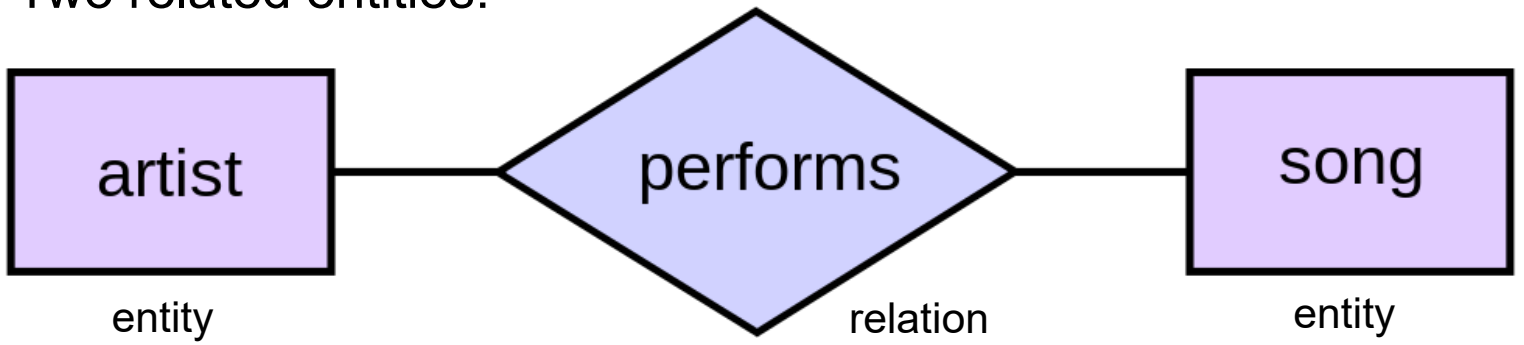
  A relationship captures how entities are related to one another.
- Attributes

  Both entities and relations can have attributes.
  Data item or property.
- Cardinality

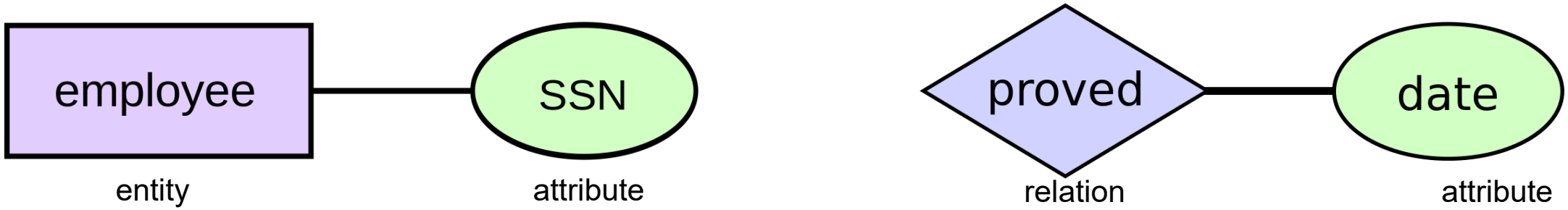
Chen, Peter (1976). "The Entity-Relationship Model - Toward a Unified View of Data". ACM Transactions on Database Systems. 1 (1): 9–36.

Entities, Relations & Attributes

Two related entities:

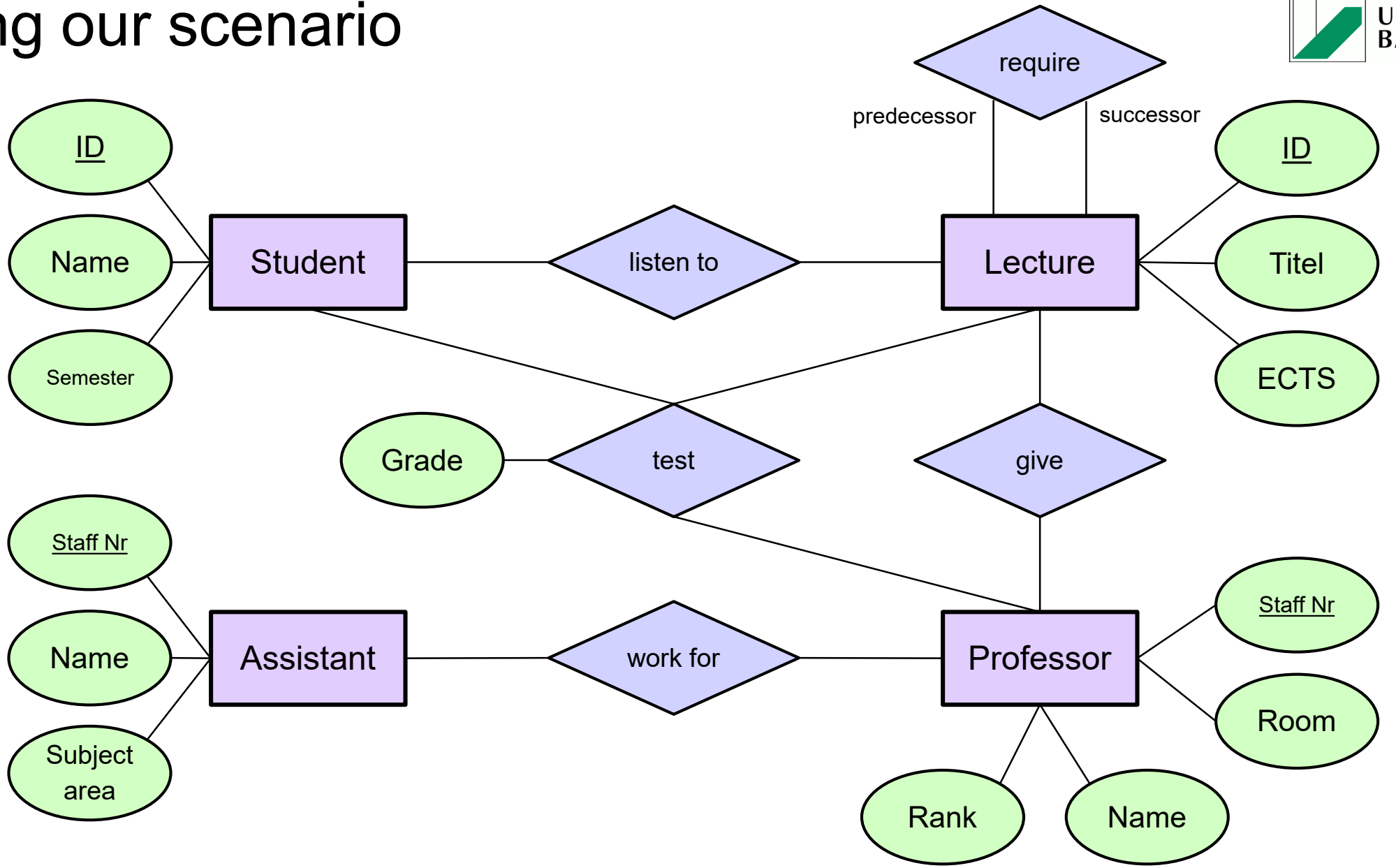


Entities and relations with attributes:

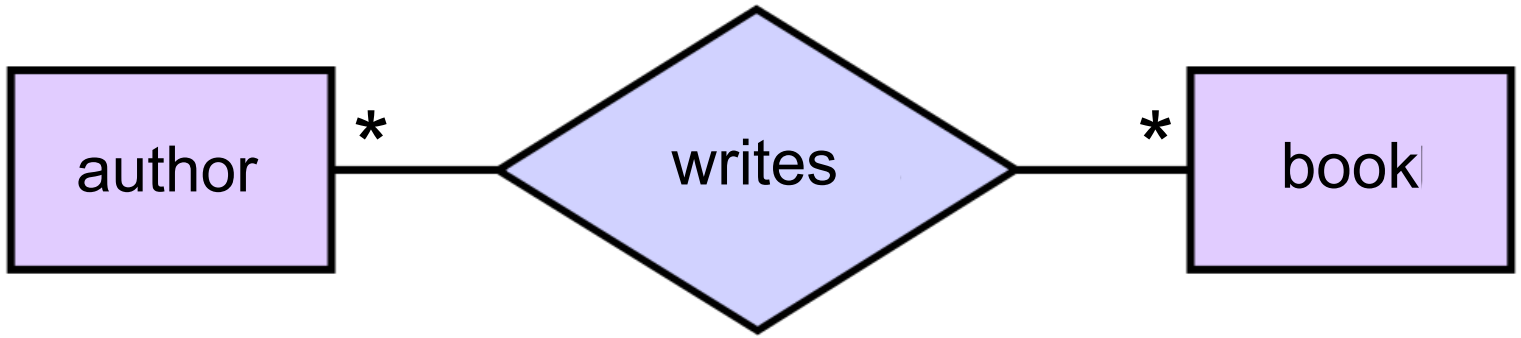


Chen, Peter (1976). "The Entity-Relationship Model - Toward a Unified View of Data". ACM Transactions on Database Systems. 1 (1): 9–36.

Modeling our scenario



Cardinality



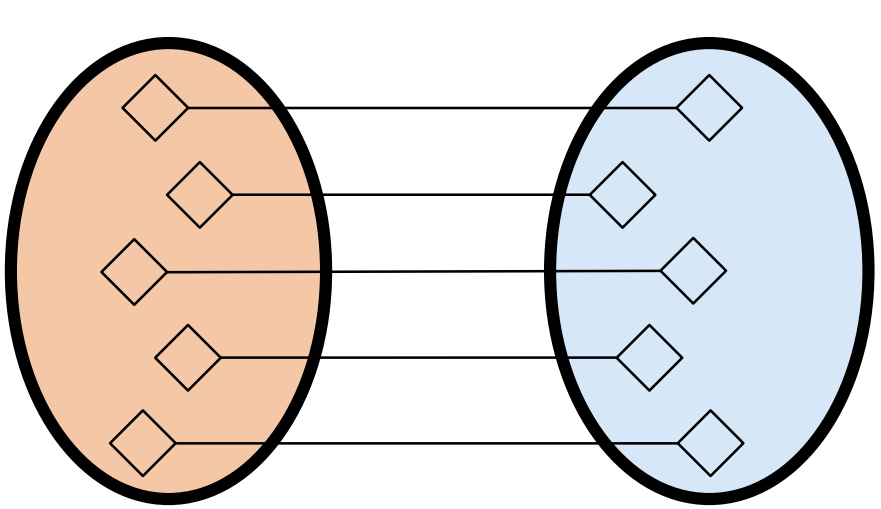
1:1

1:n

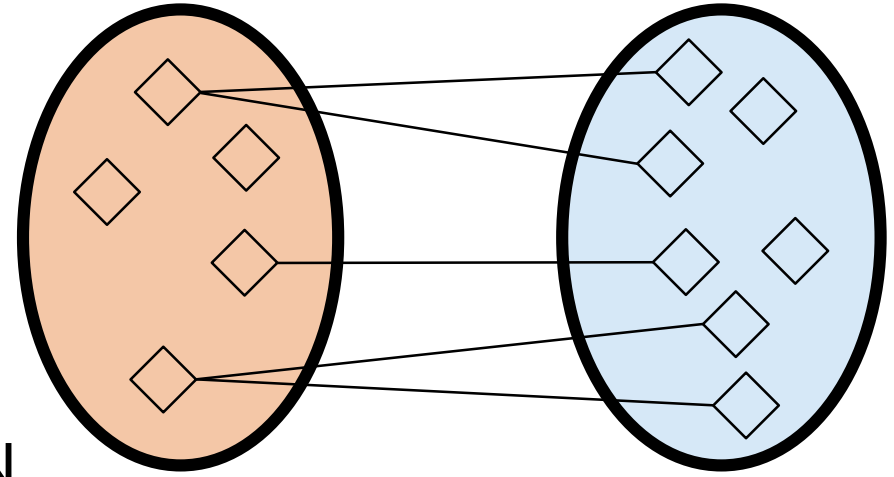
n:1

n:m

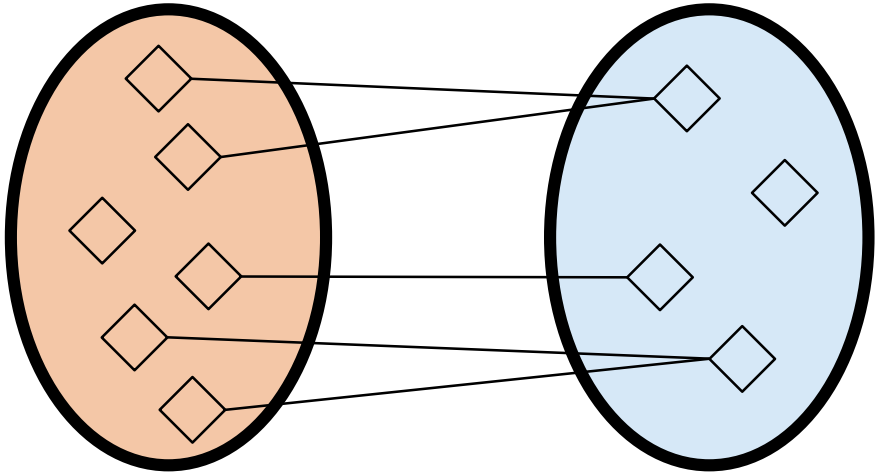
Cardinality



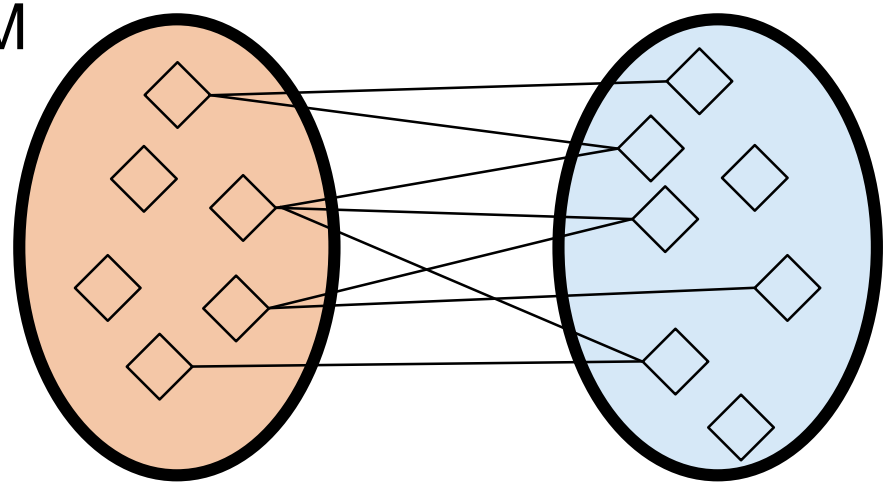
1:1



1:N

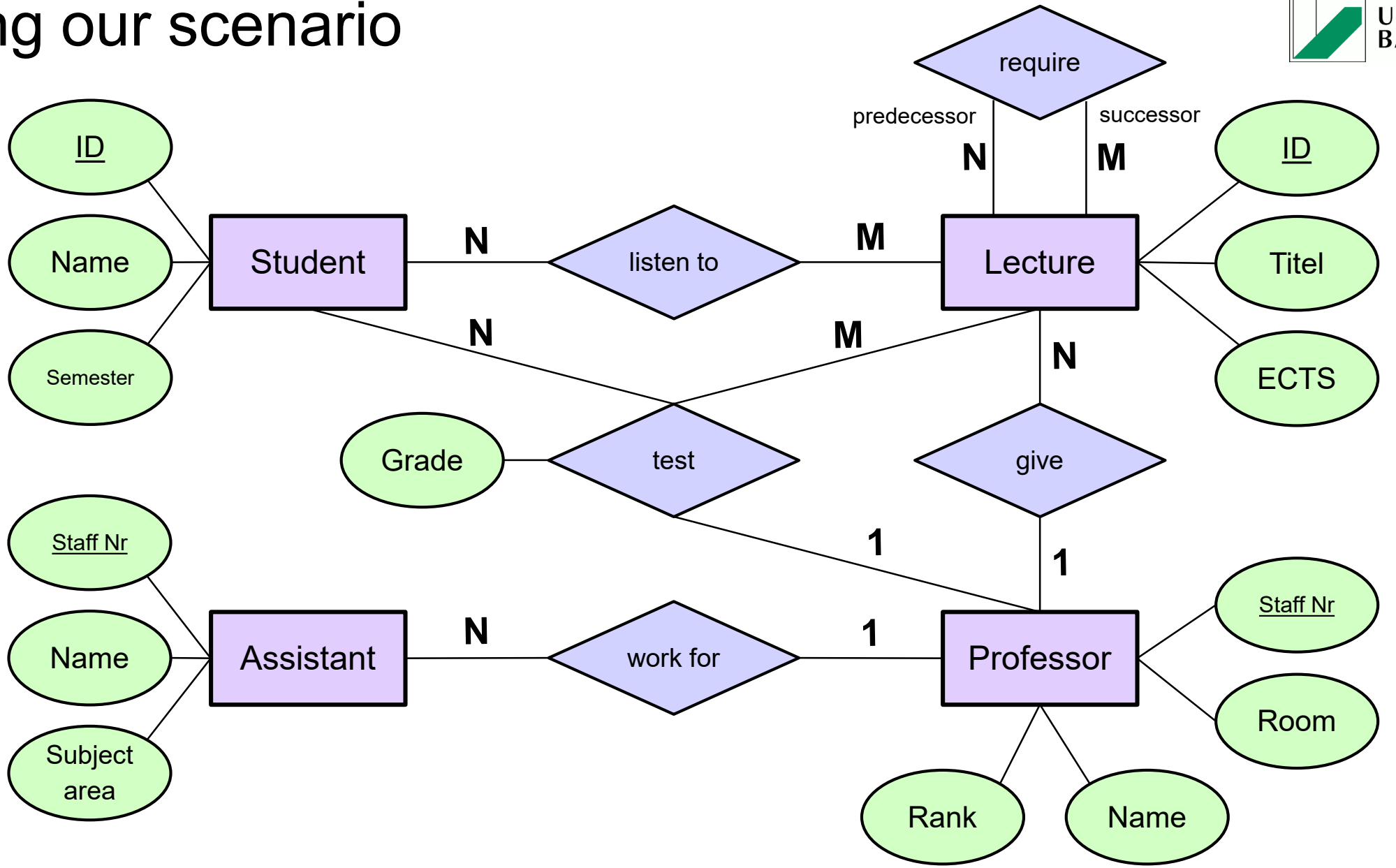


N:1



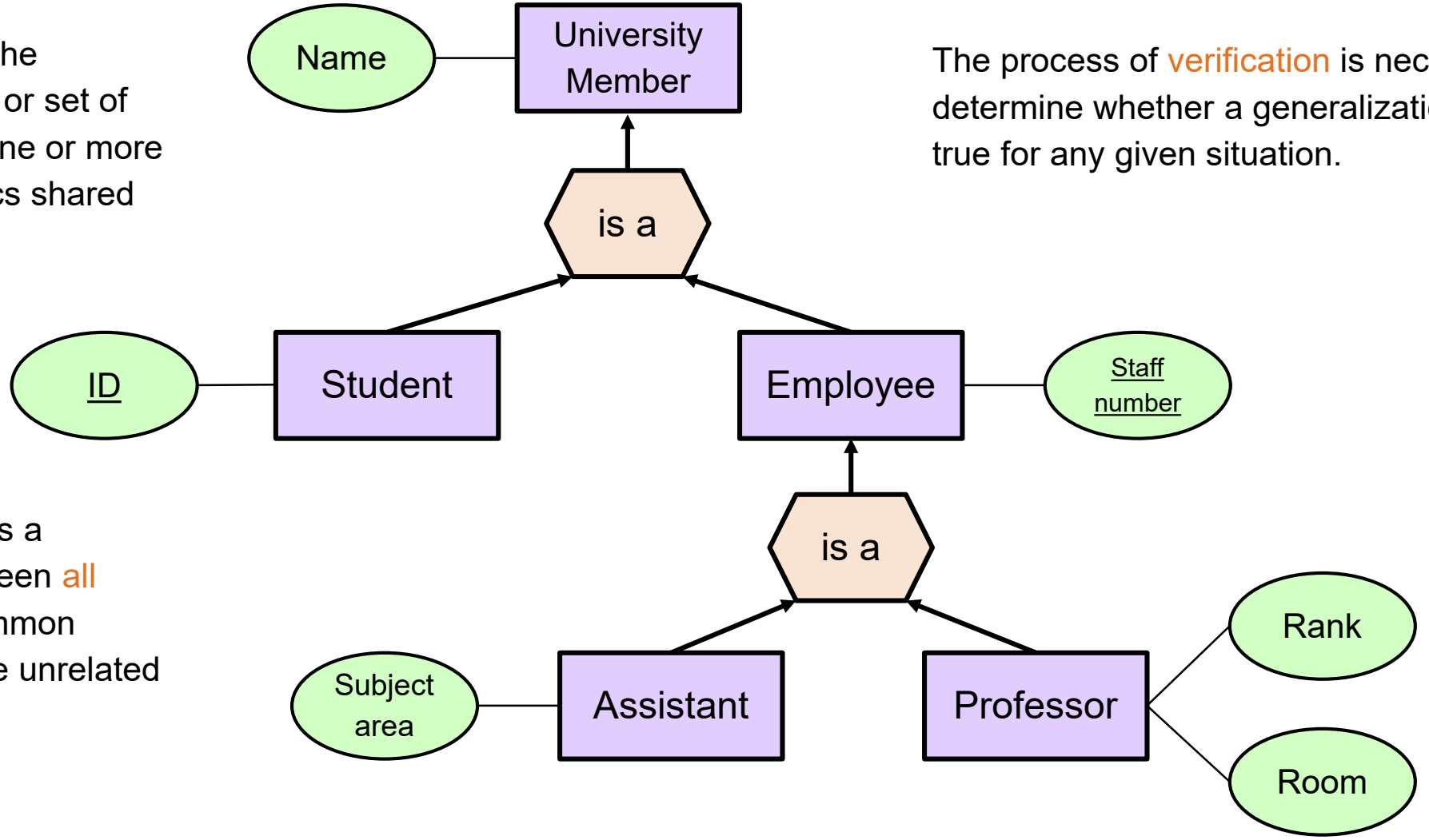
N:M

Modeling our scenario



Generalization

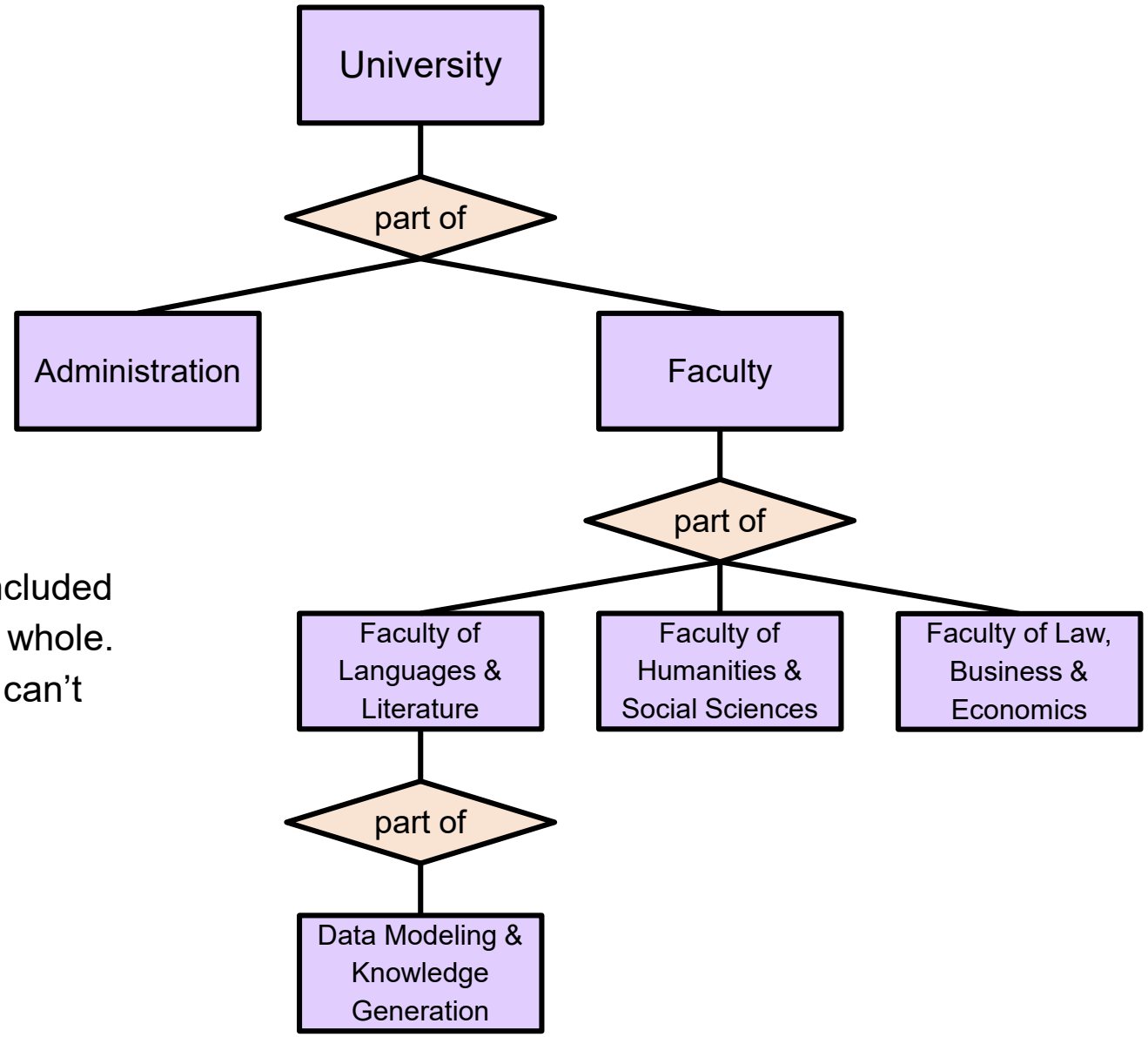
Generalizations posit the existence of a domain or set of elements, as well as one or more common characteristics shared by those elements



The process of verification is necessary to determine whether a generalization holds true for any given situation.

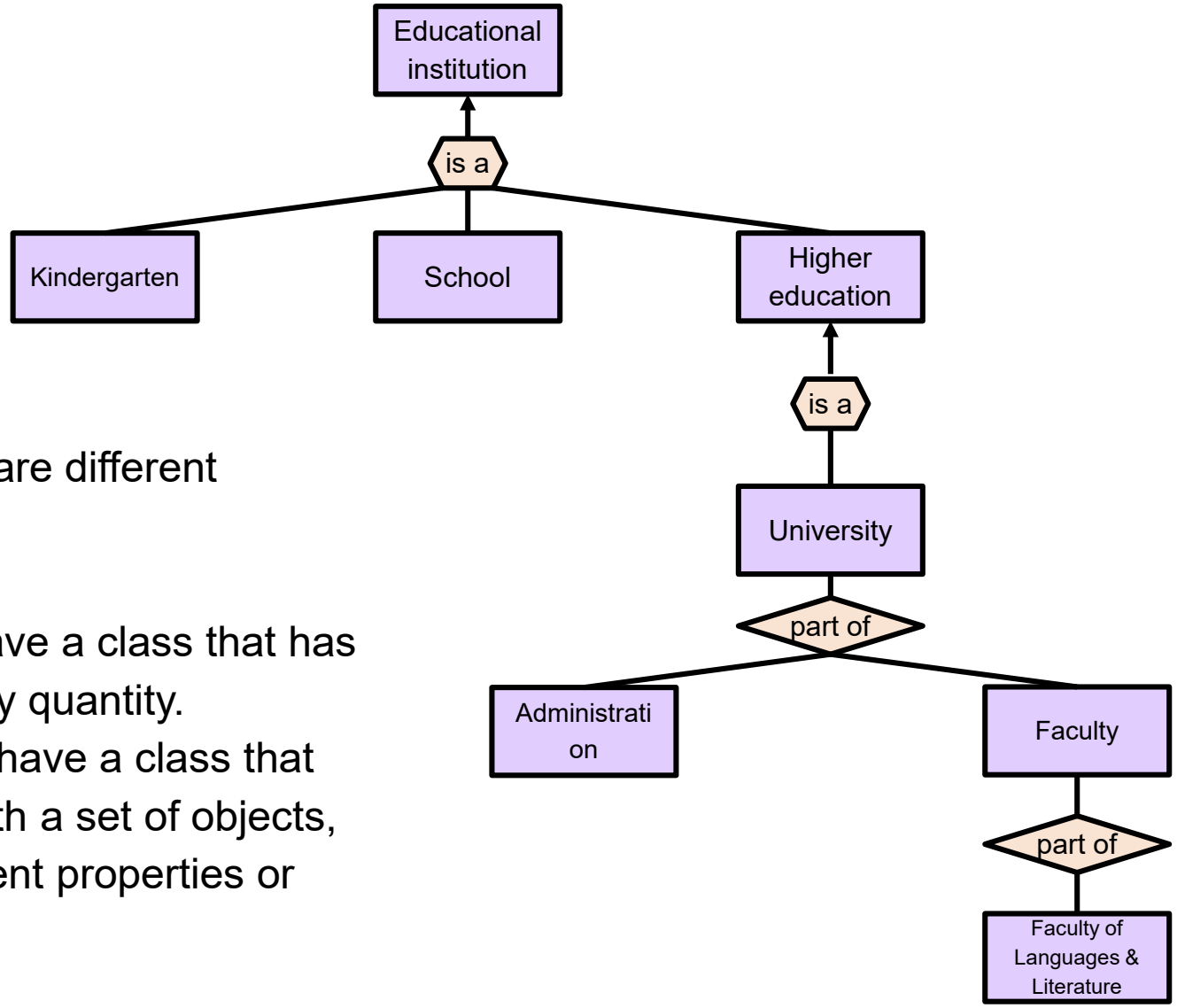
Generalization requires a common relation between all parts. Without this common relation, parts might be unrelated

Composition a.k.a. Composed Aggregation



In a composition, parts are physically included in the whole. Parts live and die with the whole. A part can only belong to a whole but it can't exist on its own.

Generalization & Composition



Generalization and composition are different concepts:

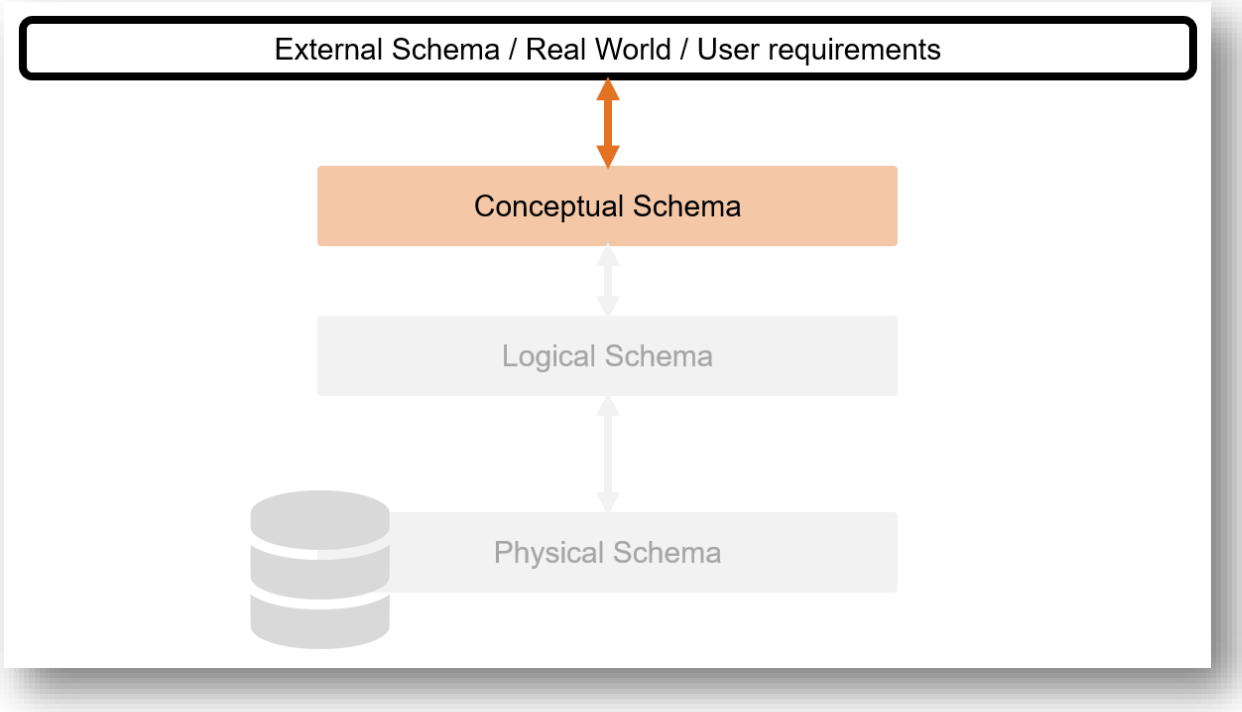
- Use composition when you have a class that has a set of another objects, in any quantity.
- Use generalization when you have a class that shares common properties with a set of objects, but can also have other different properties or behavior.

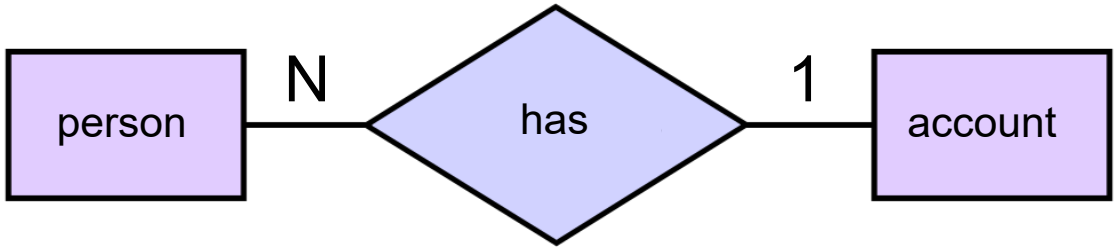
Summary

UNIVERSITÄT BAYREUTH

By data modeling we try to find a translation of real-world situations to data & databases

Clear analysis of the requirements is an important step in data modeling.





An **entity-relationship model** describes interrelated things of interest in a specific domain of knowledge.

Thanks.

mirco.schoenfeld@uni-bayreuth.de