

# Metadata

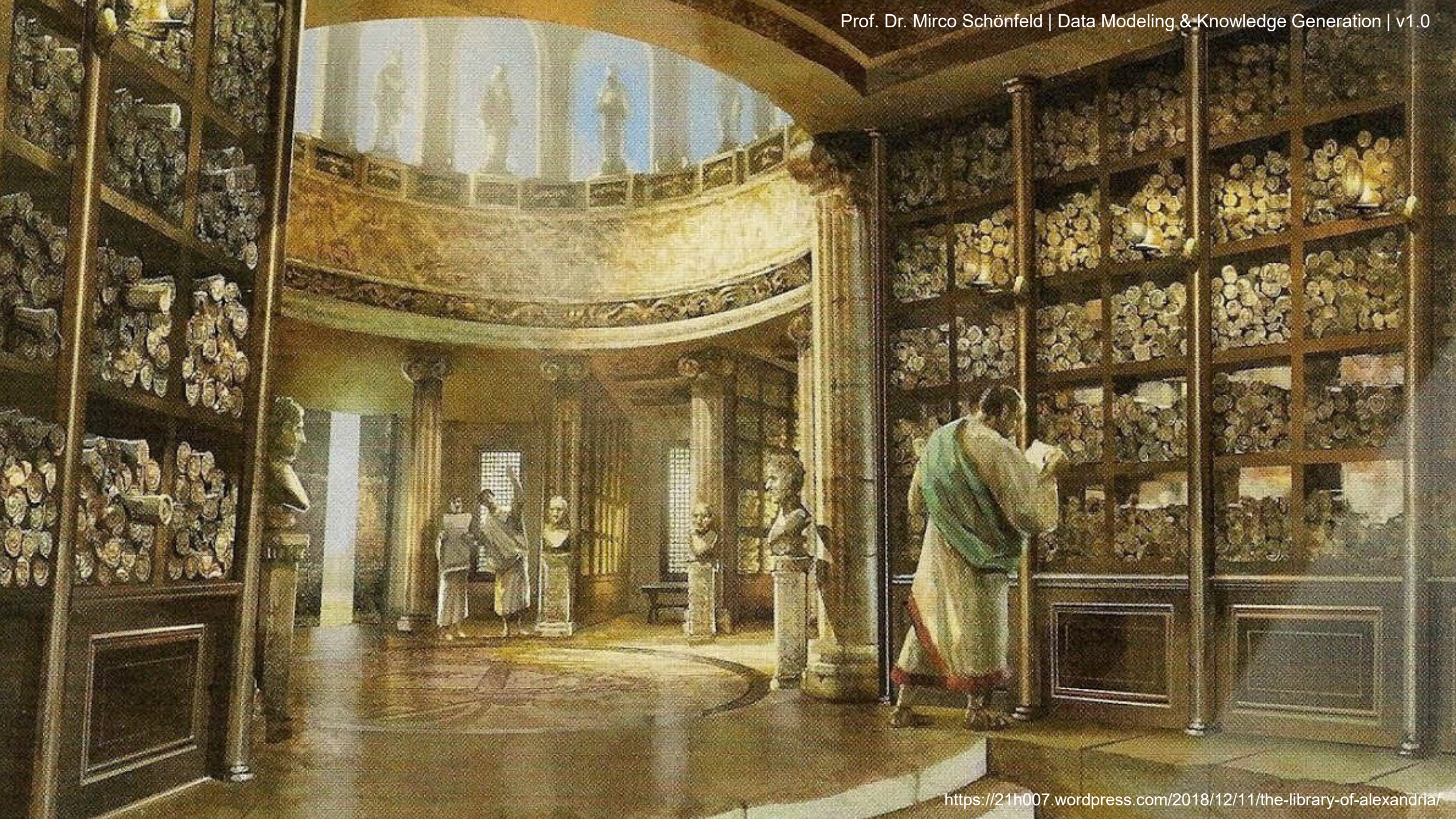
Mirco Schönfeld  
University of Bayreuth

[mirco.schoenfeld@uni-bayreuth.de](mailto:mirco.schoenfeld@uni-bayreuth.de)  
[@TWlY29](https://twitter.com/TWlY29)



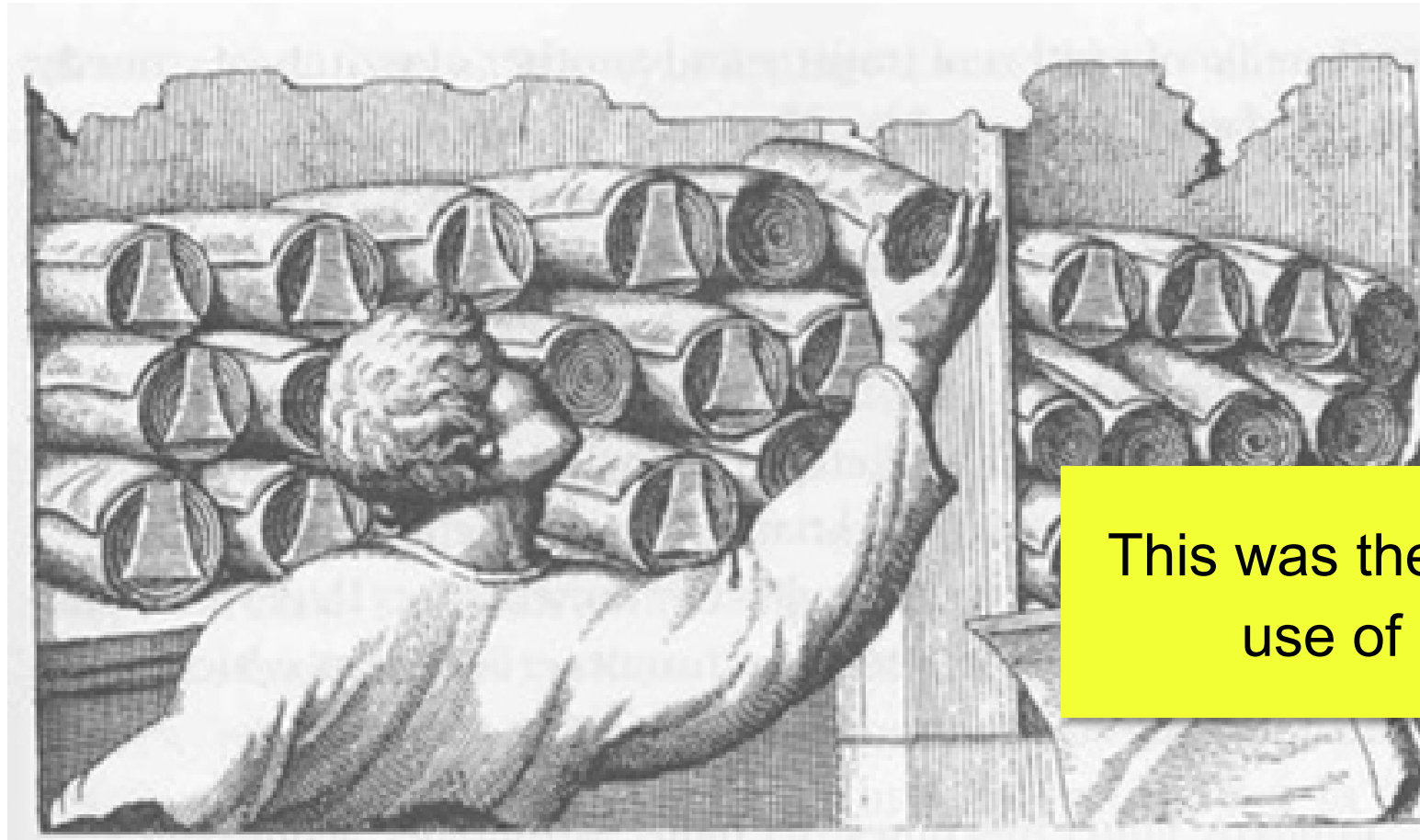






# Zenodotus

First Head Librarian of Library of Alexandria between 285-270 B.C.



This was the first recorded  
use of metadata

<http://www.systemslibrarian.co.za/first-librarian.html>



# Metadata is data about data

It describes and gives information about other data

- Machine readable
- Descriptive

Purposes of metadata:

Discovery, Management, Delivery, Access control, Use, Re-Use, Long-term preservation, ...

There exists a vast amount of metadata schemes for various domains and different types of content

# Types of Metadata

## Descriptive

Mainly used for discovery & identification

*Title, Author, Abstract, Keywords, and more*

## Structural

Describe containers of data, and how compound objects are put together

*Ordering of pages that form chapters*

## Administrative

Used to manage a resource.

*Permissions, date of creation, modality of creation, ...*

## Reference

Describe quality of data. Also called technical metadata.

*Image resolution, file size, file type ...*

## Statistical

Describe statistical processes that were applied to create data

*Mainly of interest for statistical community*



# Storing Metadata

Metadata is highly structured information

Metadata can be stored

- in a separate database
- together with the actual data



# Annotating Texts



Bayreuth (German: [baɪ'ʁɔʏt], Upper Franconian: [ba(:)'raɪ̯t̩]; Bavarian: Bareid) is a medium-sized town in northern Bavaria, Germany, on the Red Main river in a valley between the Franconian Jura and the Fichtelgebirge Mountains. The town's roots date back to 1194. In the 21st century, it is the capital of Upper Franconia and has a population of 72,148 (2015). It is world-famous for its annual Bayreuth Festival, at which performances of operas by the 19th-century German composer Richard Wagner are presented.



# Annotating Texts

Bayreuth (German: [baɪ'ʁɔʏt], Upper Franconian: [ba(:)'raɪ̯t̩]; Bavarian: Bareid) is a medium-sized town in northern `<state>Bavaria</state>`, Germany, on the Red Main river in a valley between the Franconian Jura and the Fichtelgebirge Mountains. The town's roots date back to 1194. In the 21st century, it is the capital of Upper Franconia and has a population of 72,148 (2015). It is world-famous for its annual Bayreuth Festival, at which performances of operas by the 19th-century German composer Richard Wagner are presented.



# Annotating Texts

Bayreuth (German: [baɪ'ʁɔʏt], Upper Franconian: [ba(:)'raɪ̯t̩]; Bavarian: Bareid) is a medium-sized town in northern `<state>Bavaria</state>`, `<country>Germany</country>`, on the `<river>Red Main</river>` river in a valley between the Franconian Jura and the Fichtelgebirge Mountains. The town's roots date back to `<time>1194</time>`. In the 21st century, it is the capital of `<administrative_region>Upper Franconia</administrative_region>` and has a population of 72,148 (2015). It is world-famous for its annual Bayreuth Festival, at which performances of operas by the 19th-century German composer `<person>Richard Wagner</person>` are presented.

**Inline Markup** or **Inline Annotation**



# Annotating Texts

Bayreuth (German: [baɪ'ʁɔʏt], Upper Franconian: [ba(:)'raɪt̩]; Bavarian: Bareid) is a medium-sized town in northern Bavaria, Germany, on the Red Main river in a valley between the Franconian Jura and the Fichtelgebirge Mountains. The town's roots date back to **[date start]1194[date end 1194-11-09]**. In the 21st century, it is the capital of Upper Franconia and has a population of 72,148 (2015). It is world-famous for its annual Bayreuth Festival, at which performances of operas by the 19th-century German composer Richard Wagner are presented.

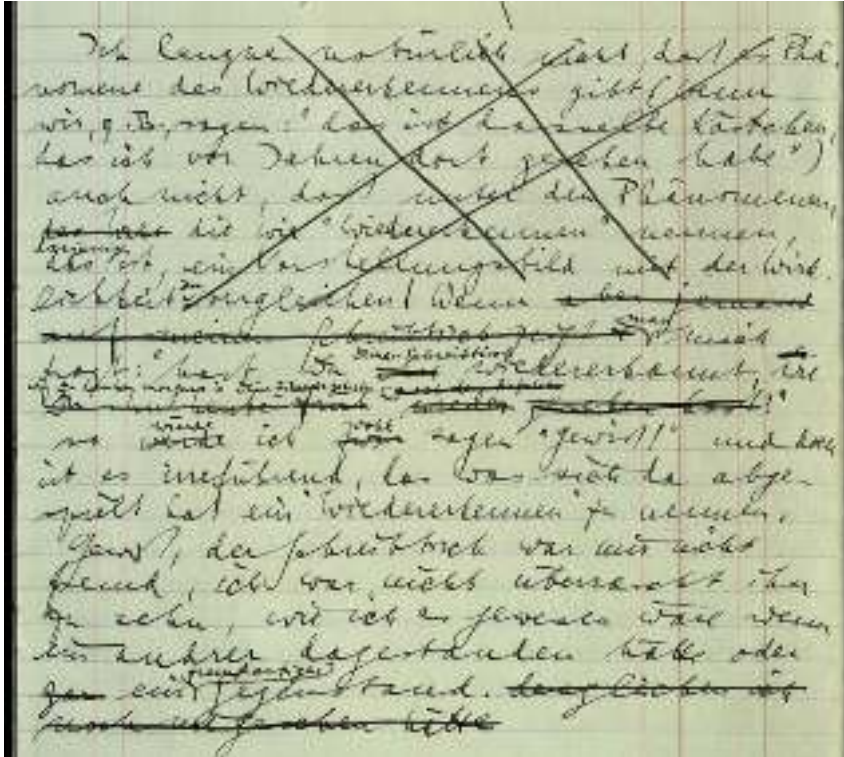


# Stand-off Annotations

Bayreuth (German: [baɪ'ʁɔʏt], **Upper Franconian**: [ba(:)'raɪt̩]; Bavarian: Bareid) is a medium-sized town in northern Bavaria, Germany, on the Red Main river in a valley between the Franconian Jura and the Fichtelgebirge Mountains. The town's roots date back to 1194. In the 21st century, it is the capital of Upper Franconia and has a population of 72,148 (2015). It is world-famous for its annual Bayreuth Festival, at which performances of operas by the 19th-century German composer Richard Wagner are presented.

<b>Start Offset</b>	<b>End Offset</b>	<b>Annotation</b>
19	27	Phonetic Spelling
31	48	Region
...	...	...

# Annotating Texts in the real world



## Text representation

Ich leugne natürlich nicht, daß es Phänomene des Wiedererkennens gibt (wenn wir, z.B., sagen: “das ist dasselbe Kästchen, das ich vor Jahren dort gesehen habe”) auch nicht, daß unter den Phänomenen, die wir “Wiedererkennen” nennen, das || dasjenige ist, ein Vorstellungsbild mit der Wirklichkeit zu vergleichen.

## Metadata annotation

```
<ab ana="abnr:5" date="field:PhilosophyOfLanguage_pub:W-PG: Part I: Sect115_date:19331214?-19331231?" date_norm="1933-12-14?-1933-12-31?" n="Ms-115,2[2]">
```

```
<seg type="publ">
```

```
<s ana="fac:Ms-115,2 abnr:5 satznr:17" n="Ms-115,2[2]_1">Ich leugne natürlich nicht, daß es Phä
```

```
<lb rend="shyphen"/>nomene des Wiedererkennens gibt
```

```
<seg type="stripped"> (w</seg>enn
```

```
<lb/> wir,
```

```
<abbr type="abb">z.B.</abbr>, sagen: “das ist dasselbe Kästchen,
```

```
<lb/> das ich vor Jahren dort gesehen habe”)
```

```
<lb/> auch nicht, daß unter den Phänomenen
```

```
<seg type="stripped">, die wir “Wiedererkennen” nennen,</seg>
```

```
<lb/>
```

```
<choice type="s">
```

```
<seg n="s_alt1">das</seg>
```

```
<seg n="s_alt2">dasjenige</seg>
```

```
</choice> ist, ein Vorstellungsbild mit der Wirk
```

```
<lb rend="shyphen"/>lichkeit zu vergleichen.
```

```
<seg type="edcom"> | </seg>
```

```
</s>
```

# Annotating music

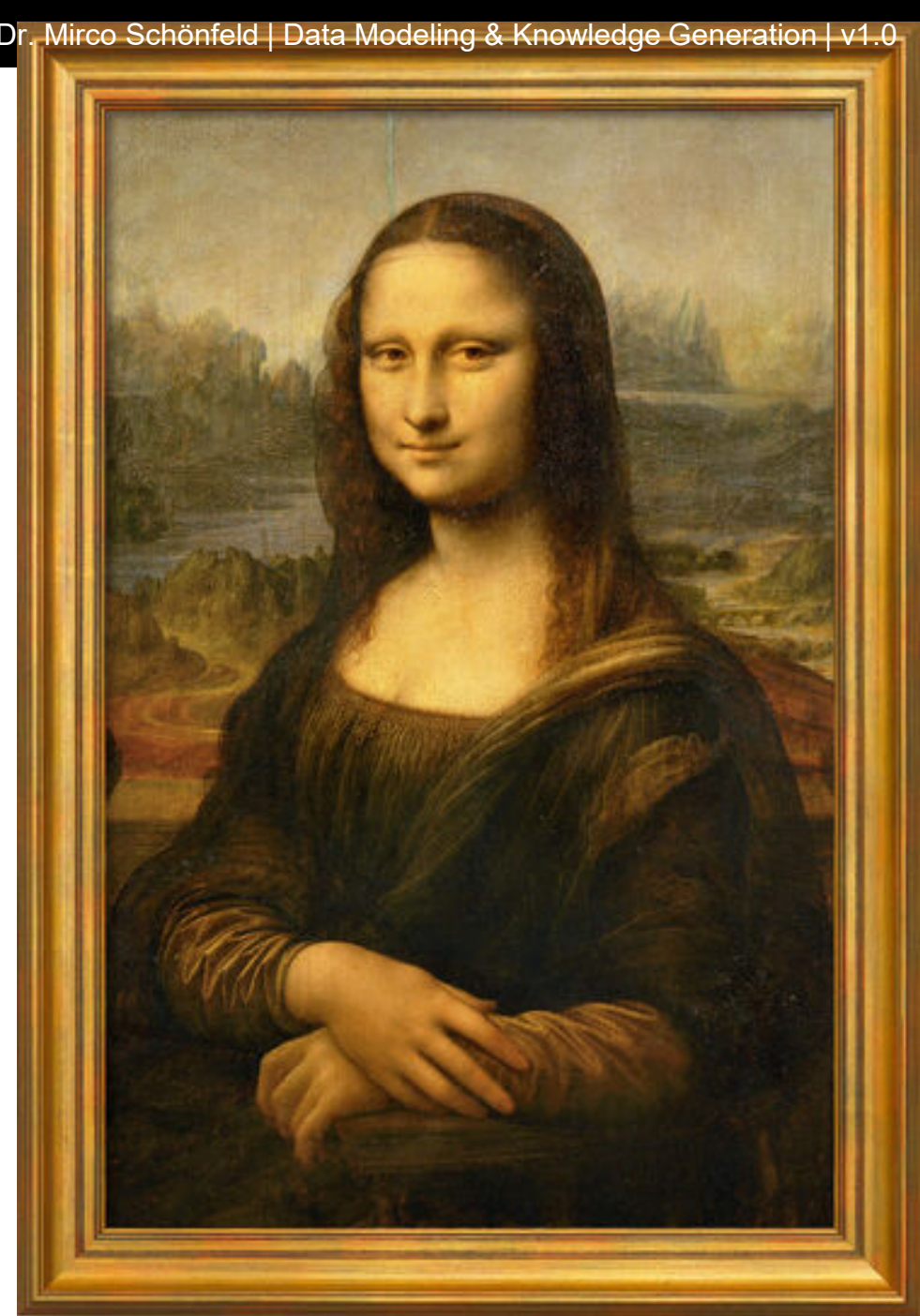


`<measure n="5">`  
`<staff n="1">`  
`<layer n="1">`  
`<note pname="c" oct="5" dur="8" dots="1" stem.dir="down" accid.ges="s">`  
`<verse n="1">`  
`<syl wordpos="1" con="d">nun</syl>`  
`</verse>`  
`</note>`  
`<note pname="b" oct="4" dur="16" stem.dir="down">`  
`<verse n="1">`  
`<syl wordpos="m" con="d">der</syl>`  
`</verse>`  
`</note>`  
`<note pname="b" oct="4" dur="16" stem.dir="up">`  
`<verse n="1">`  
`<syl wordpos="m" con="d">schö</syl>`  
`</verse>`  
`</note>`  
`<note pname="d" oct="5" dur="16" stem.dir="up">`  
`<verse n="1">`  
`<syl wordpos="t">nen</syl>`  
`</verse>`  
`</note>`

Musical score for piano and voice. The lyrics are: "Im wun - derschö - nen Mo - nat Mai, als al - le Knos - pen". The score includes piano accompaniment and a vocal line. XML annotations are overlaid on the notes, providing metadata such as pitch name, octave, duration, stem direction, and syllable information.



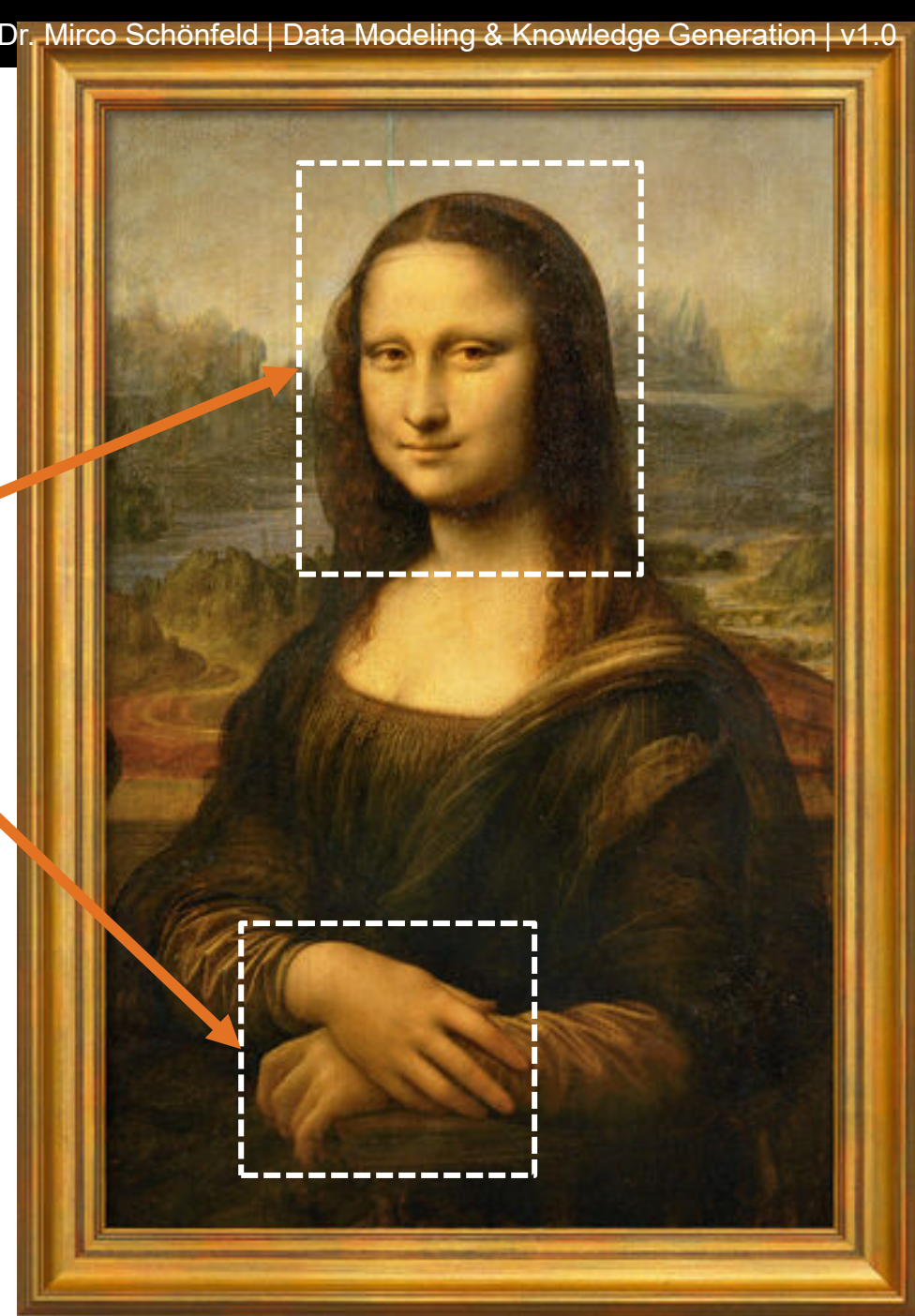
# Annotating Images



# Annotating Images

Face

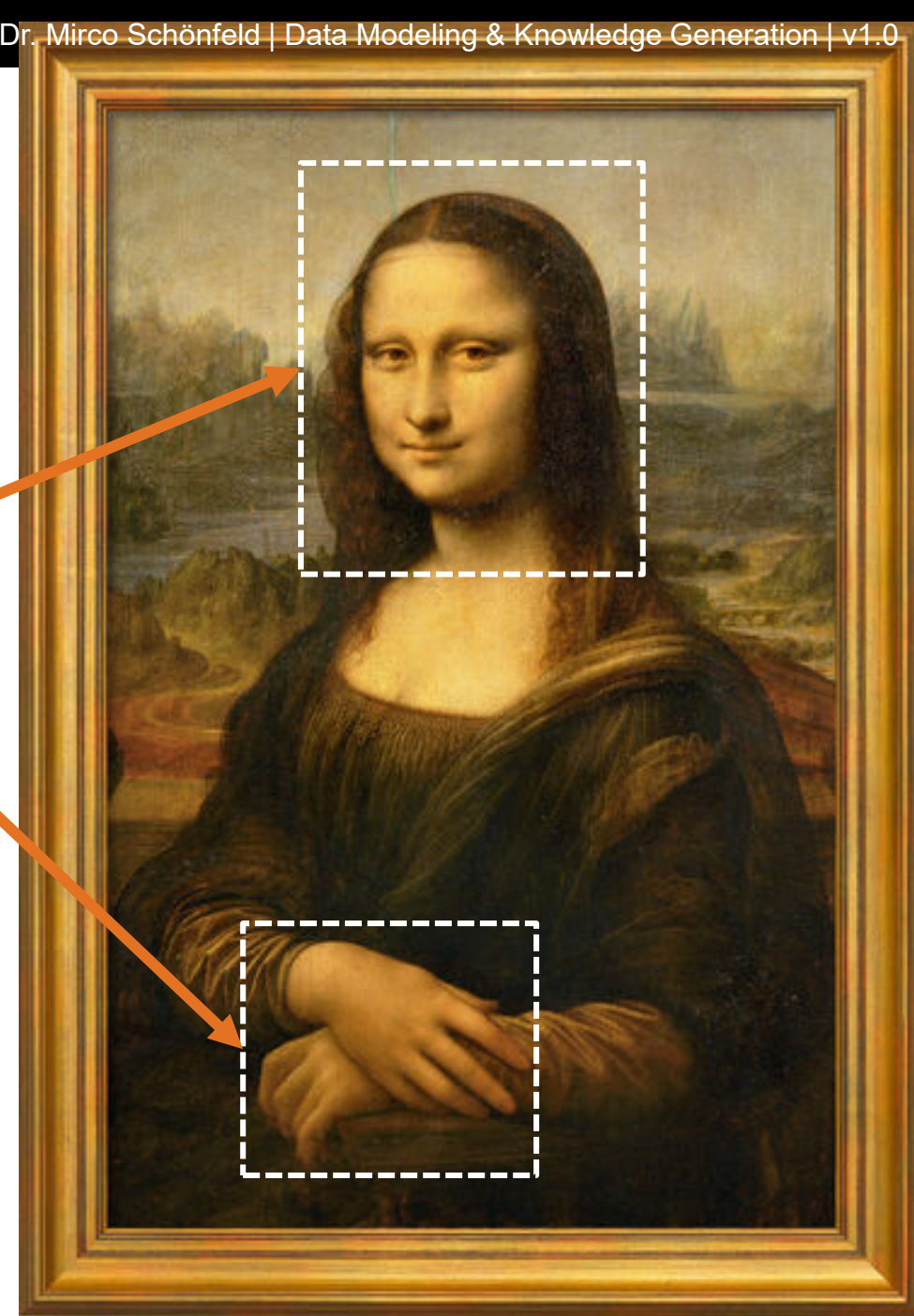
Hands



# Annotating Images

Face

Hands



ID	Upper-left-x	Upper-left-y	Lower-right-x	Lower-right-y	Annotation
124123	24	5	57	40	Face
124123	17	76	43	87	Hands

# Annotating Images

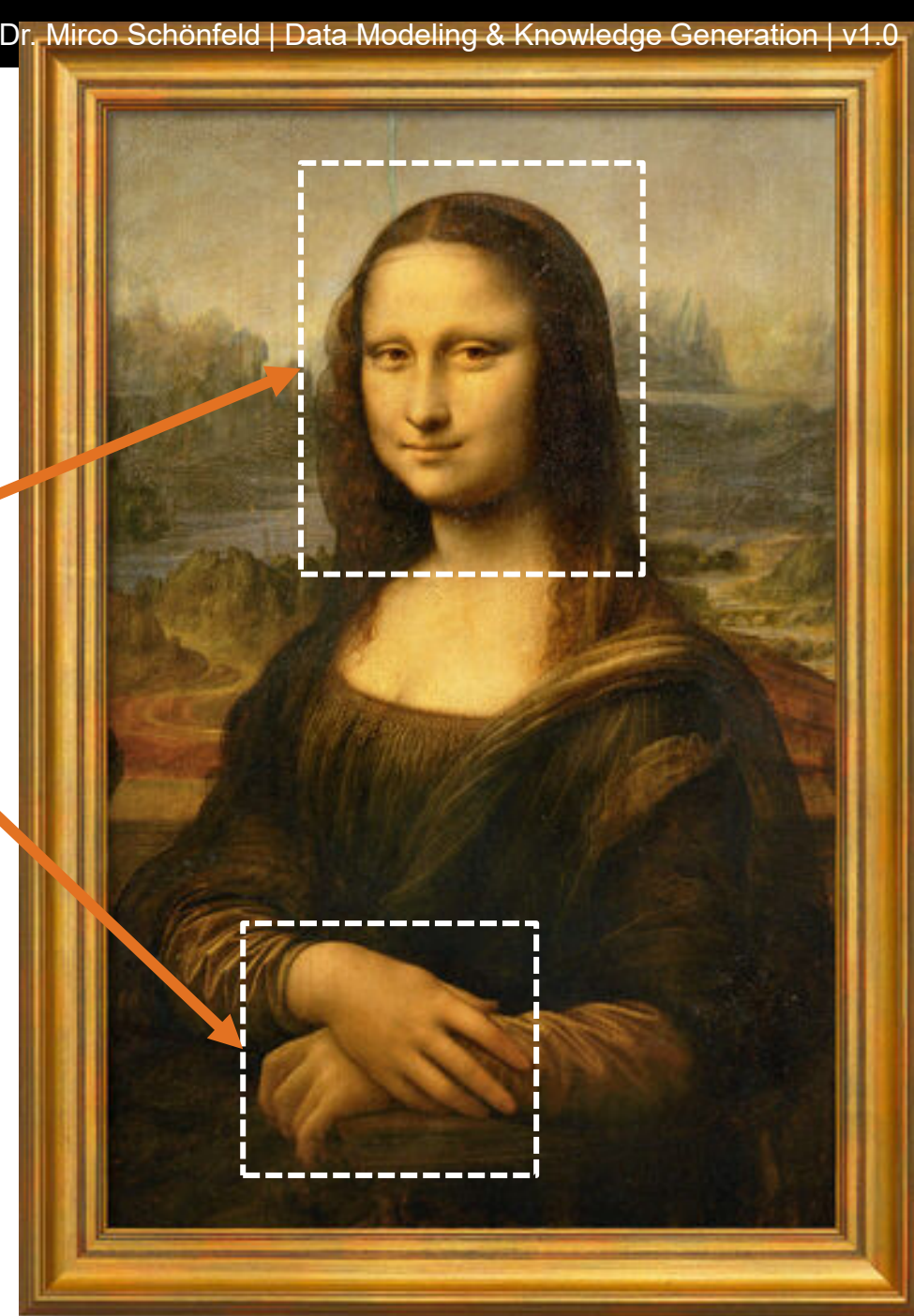
## *Unique Identifier:*

A name that is guaranteed to be unique among all other identifiers and can be used to reference this object

ID	Upper-left-x	Upper-left-y	Lower-right-x	Lower-right-y	Annotation
124123	24	5	57	40	Face
124123	17	76	43	87	Hands

Face

Hands



# Components of Formal Models

## Modeled Instance

Digital Model of a specific entity

Contains the actual entity together with all metadata and annotations related to it

## Data Model

Structure used to encode the actual data

## Metamodel

Structure of metadata and their specific organization

# Components of Formal Models

## Modeled Instance

Digital Model of a specific entity

Contains the actual entity together with all metadata and annotations related to it

## Data Model

Structure used to encode the actual data

## Metamodel

Structure of metadata and their specific organization



ID	Upper-left-x	Upper-left-y	Lower-right-x	Lower-right-y	Annotation
124123	24	5	57	40	Face
124123	17	76	43	87	Hands

# Components of Formal Models

## Modeled Instance

Digital Model of a specific entity

Contains the actual entity together with all metadata and annotations related to it

Digitalization of the painting and the table

## Data Model

Structure used to encode the actual data

## Metamodel

Structure of metadata and their specific organization



ID	Upper-left-x	Upper-left-y	Lower-right-x	Lower-right-y	Annotation
124123	24	5	57	40	Face
124123	17	76	43	87	Hands

# Components of Formal Models

## Modeled Instance

Digital Model of a specific entity

Contains the actual entity together with all metadata and annotations related to it

Digitalization of the painting and the table

## Data Model

Structure used to encode the actual data

Format of the picture file

## Metamodel

Structure of metadata and their specific organization



ID	Upper-left-x	Upper-left-y	Lower-right-x	Lower-right-y	Annotation
124123	24	5	57	40	Face
124123	17	76	43	87	Hands



# Components of Formal Models

## Modeled Instance

Digital Model of a specific entity

Contains the actual entity together with all metadata and annotations related to it

Digitalization of the painting and the table

## Data Model

Structure used to encode the actual data

Format of the picture file

## Metamodel

Structure of metadata and their specific organization

Table and its specific structure



ID	Upper-left-x	Upper-left-y	Lower-right-x	Lower-right-y	Annotation
124123	24	5	57	40	Face
124123	17	76	43	87	Hands

# Modeling Metadata

# Metadata Models: Two Examples

```

<stage>Enter two Centinels.
<add place="margin">Now call'd <name xml:id
  <name xml:id="francisco">Francesco</name>.
</stage>
<sp who="#francisco">
  <speaker>1.</speaker>
  <l part="Y">Stand: who is that?</l>
</sp>
<sp who="#barnardo">
  <speaker>2.</speaker>
  <l part="Y">Tis I.</l>
</sp>
<sp who="#francisco">
  <speaker>1.</speaker>
  <l>O you come most carefully vpon your wat
</sp>
<sp who="#barnardo">
  <speaker>2.</speaker>
  <l>And if you meete Marcellus and Horatio,
  <l>The partners of my watch, bid them make
</sp>
<sp who="#francisco">
  <speaker>1.</speaker>
  <l part="Y">I will: See who goes there.</l>
</sp>
<stage>Enter Horatio and Marcellus.</stage>

```

Example coding of the  
First Quarto of Hamlet  
using TEI

RDF: Resource Description Framework  
Factual triplets in the form of  
(*head*, **relation**, *tail*) or  
(*subject*, **predicate**, *object*)

(Albert Einstein, **BornIn**, German Empire)  
 (Albert Einstein, **SonOf**, Hermann Einstein)  
 (Albert Einstein, **GraduateFrom**, University of Zurich)  
 (Albert Einstein, **WinnerOf**, Nobel Prize in Physics)  
 (Albert Einstein, **ExpertIn**, Physics)  
 (Nobel Prize in Physics, **AwardIn**, Physics)  
 (The theory of relativity, **TheoryOf**, Physics)  
 (Albert Einstein, **SupervisedBy**, Alfred Kleiner)  
 (Alfred Kleiner, **ProfessorOf**, University of Zurich)  
 (The theory of relativity, **ProposedBy**, Albert Einstein)  
 (Hans Albert Einstein, **SonOf**, Albert Einstein)



# Metamodels – a selection

## XML: eXtensible Markup Language

Textual format of encoding metadata. Readable for both humans *and* machines.

Consist of Tags, Elements, Attributes, ...

Hierarchical organization between elements

## RDF: Resource Description Framework

General method for conceptual description or modeling of information

Consists of subject-predicate-object triplets

Collection of triplets can be represented as semantic networks

## Relational Database:

Very old and very common

Efficient and fast systems allow for storing and processing huge databases

Consists of tables

# Metadata Standards

For specific domains or particular types of information resources, specific metadata elements exist.

Groups of such elements combined are called a *metadata schema*

Focus area	Example schema	Short description
Humanities, social sciences & linguistics	TEI	Text Encoding Initiative
Librarianship	MODS	Metadata Object Description Schema
Networked Resources	DublinCore	
	DOI	Digital Object Identifier
Music notation	MEI	Music Encoding Initiative
Ecology	EML	Ecological Metadata Language
Social Science	DDI	Data Documentation Initiative
Photographs	Exif	Exchangeable file format

# Elements of a Metadata Schema

- **Semantics**  
What do individual metadata elements mean and represent?
- **Syntax**  
How is the scheme expressed? How is metacontent formulated? Plain text, XML, RDF, Database?
- **Linear vs. Hierarchical**  
Is the schema hierarchical due to parent-child relationships between elements?
- **Granularity**  
How much detail does the schema allow to model? Low granularity provides less detailed information but can be created for lower cost.
- **Master data or reference data**  
Is a controlled vocabulary required for certain value assignments?



If both the indexer and the searcher are guided to choose the same term for the same concept, then relevant documents will be retrieved.

ISO 25964-1:2011

<https://www.iso.org/obp/ui/#iso:std:iso:25964:-1:ed-1:v1:en>

# From Metadata to Ontologies





An ontology is a way of showing the properties of a subject area and how they are related, by defining a set of concepts and categories that represent the subject.

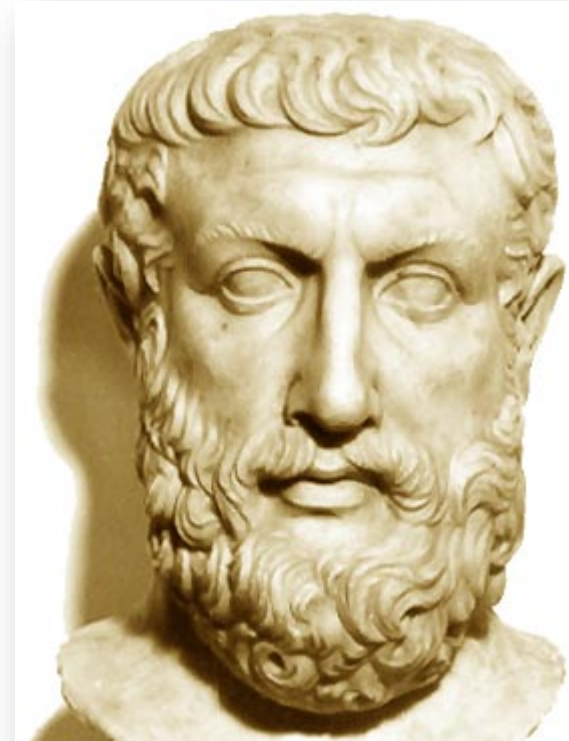
[https://en.wikipedia.org/wiki/Ontology\\_\(information\\_science\)](https://en.wikipedia.org/wiki/Ontology_(information_science))

# Ontology as the science of being

Ontology is a branch of philosophy

Studies concepts such as existence, being, becoming, and reality.

Deals with questions of how entities are grouped into basic categories and which of these entities exist on the most fundamental level.



Parmenides (born ~ 515 BC) has been considered the founder of ontology.

# Ontologies in information science

Attempt to represent entities, ideas, events, etc.  
with interdependent properties and relations,  
according to a system of categories.

Ontologies limit complexity and organize data into information & knowledge

Ontologies improve problem solving within domains



Pipe



Pipe



Pipe, Wisconsin



# Ontologies Give Structure

Ontologies define a common vocabulary for metadata.

Ontologies give us

- ...the possibility to share a common understanding of the structure of information within a specific domain
- ...the possibility to reuse domain knowledge
- ...the possibility to make domain assumptions explicit
- ...the possibility to analyze domain knowledge
- ...the possibility to separate the domain knowledge from the operational knowledge

Developing an ontology of the domain is not a goal in itself.

It is rather akin to defining a set of data and their structure for other programs to use.

# Formal Ontologies are Designed



An Ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an Ontology is a systematic account of Existence.

Gruber, 1995

An intentional semantic structure which encodes the implicit rules constraining the structure of a piece of reality

The aim of an ontology is to define which primitives, provided with their associated semantics, are necessary for knowledge representation in a given context



# An Ontology of Wine

Classes describe concepts in the domain

A class of wine represents all wines

Winery

Bordeaux

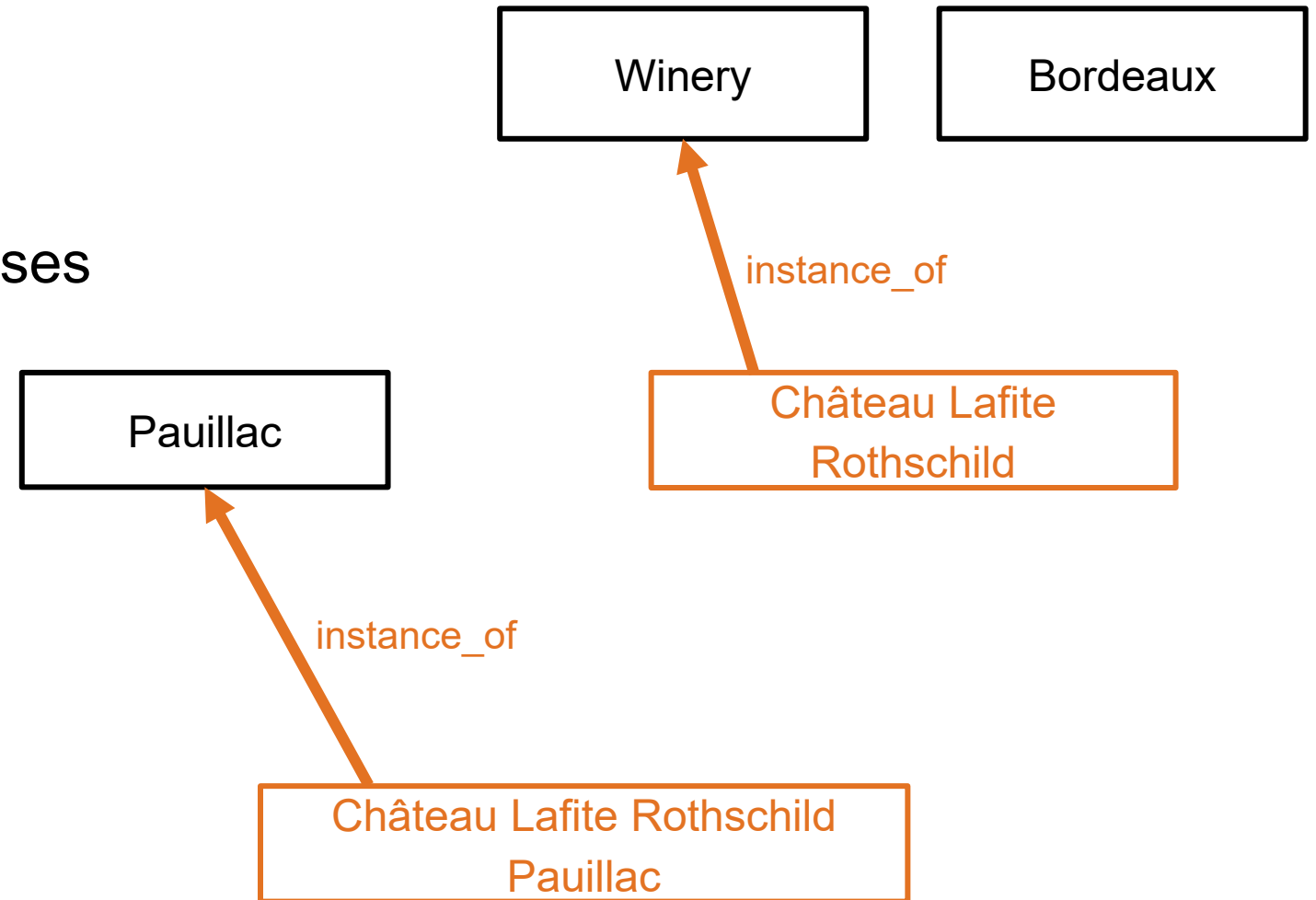
Pauillac

# An Ontology of Wine

Classes describe concepts in the domain

A class of wine represents all wines

Specific wines are instances of classes



# An Ontology of Wine

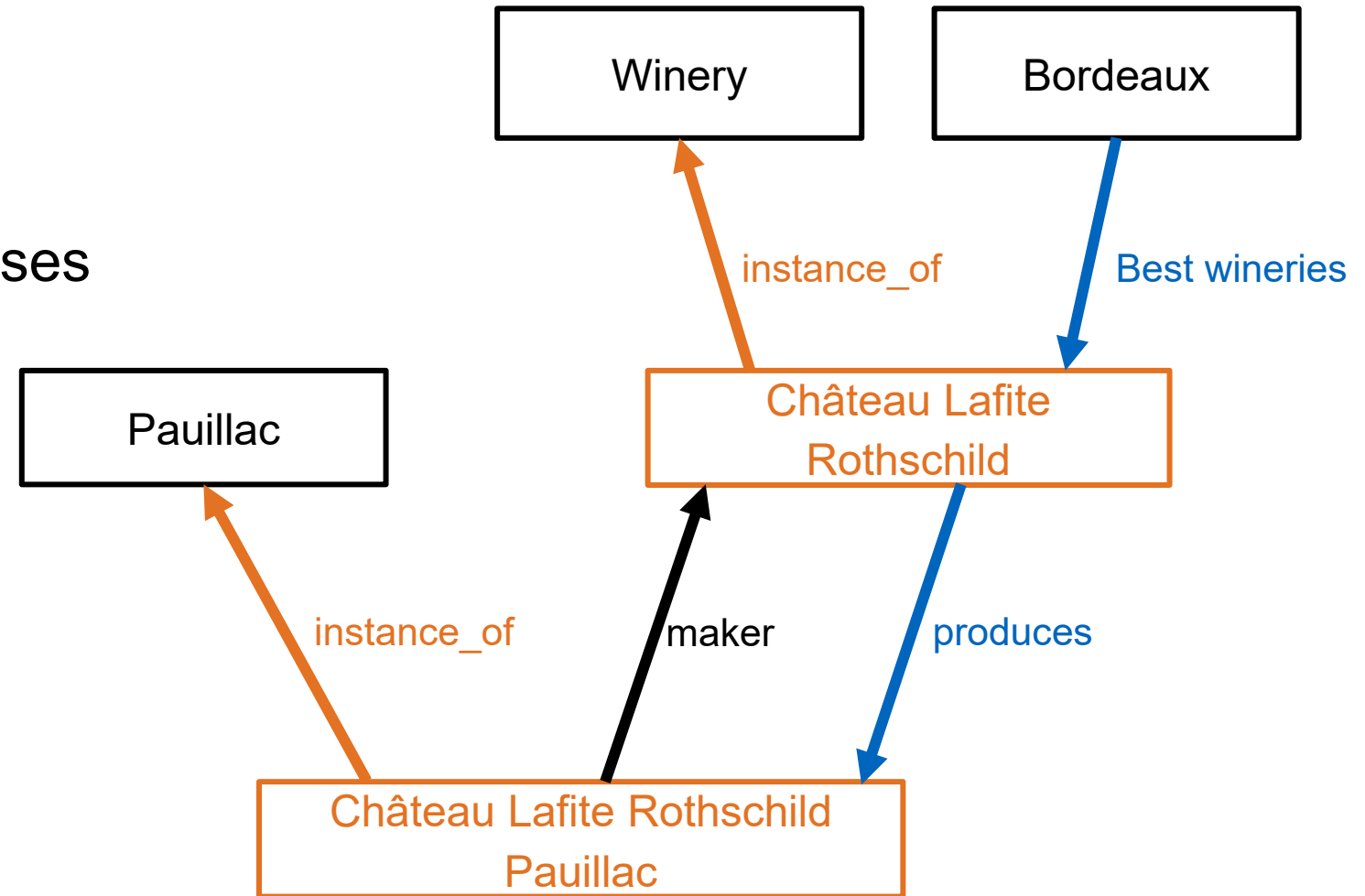
Classes describe concepts in the domain

A class of wine represents all wines

Specific wines are instances of classes

Slots describe properties of classes and instances

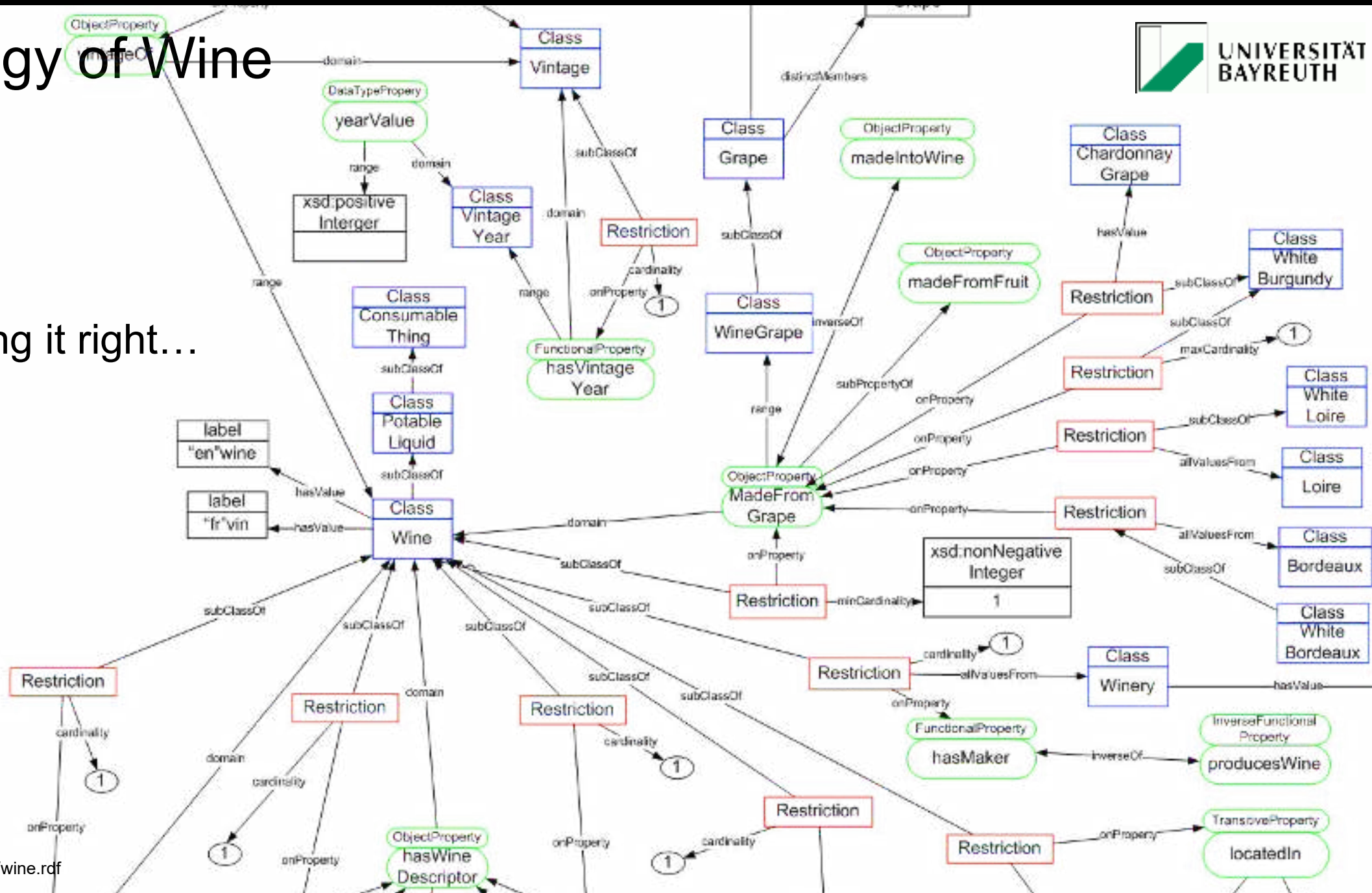
Classes can have subclasses





# An Ontology of Wine

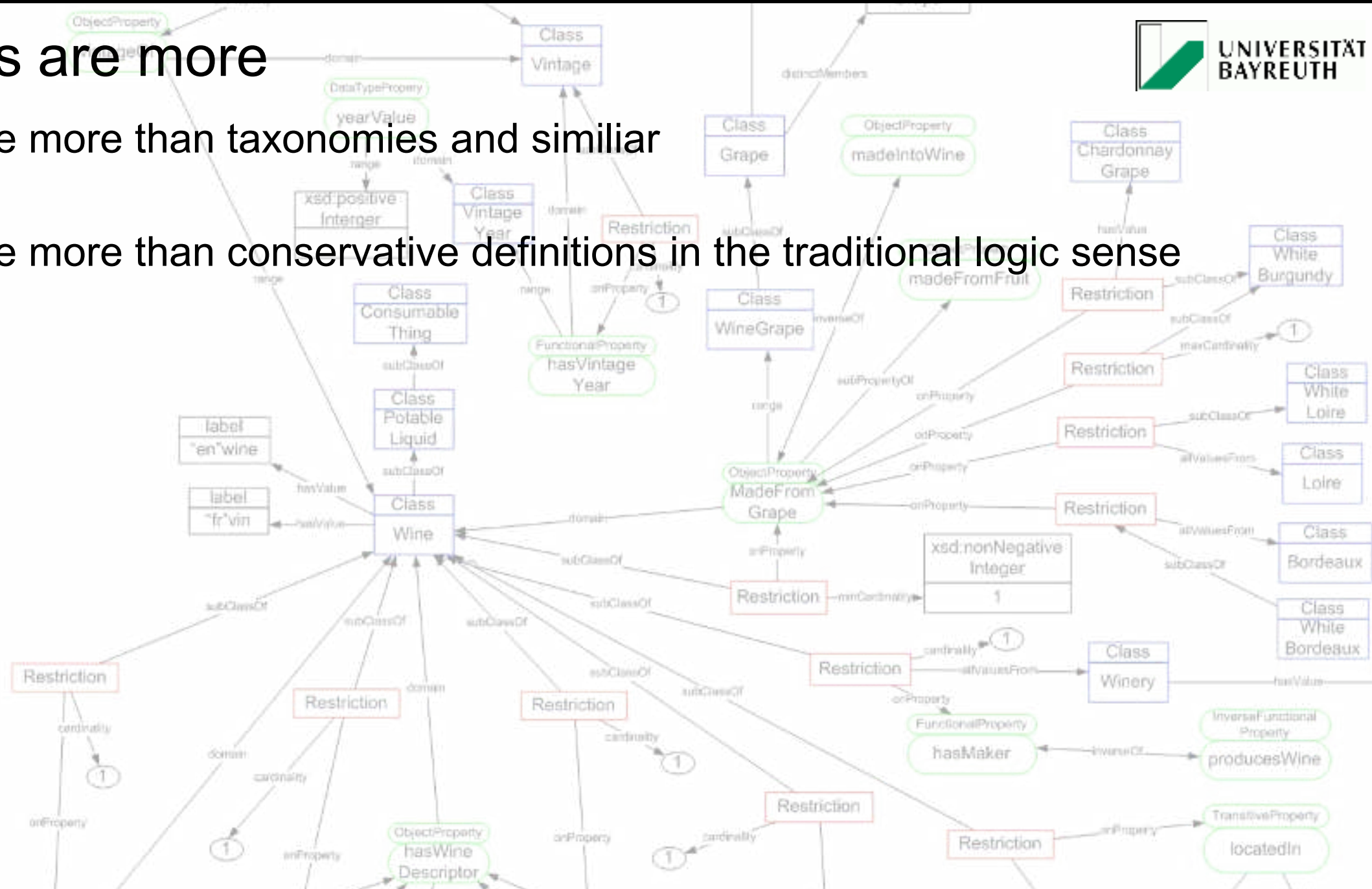
If you're doing it right...



# Ontologies are more

## Ontologies are more than taxonomies and similar

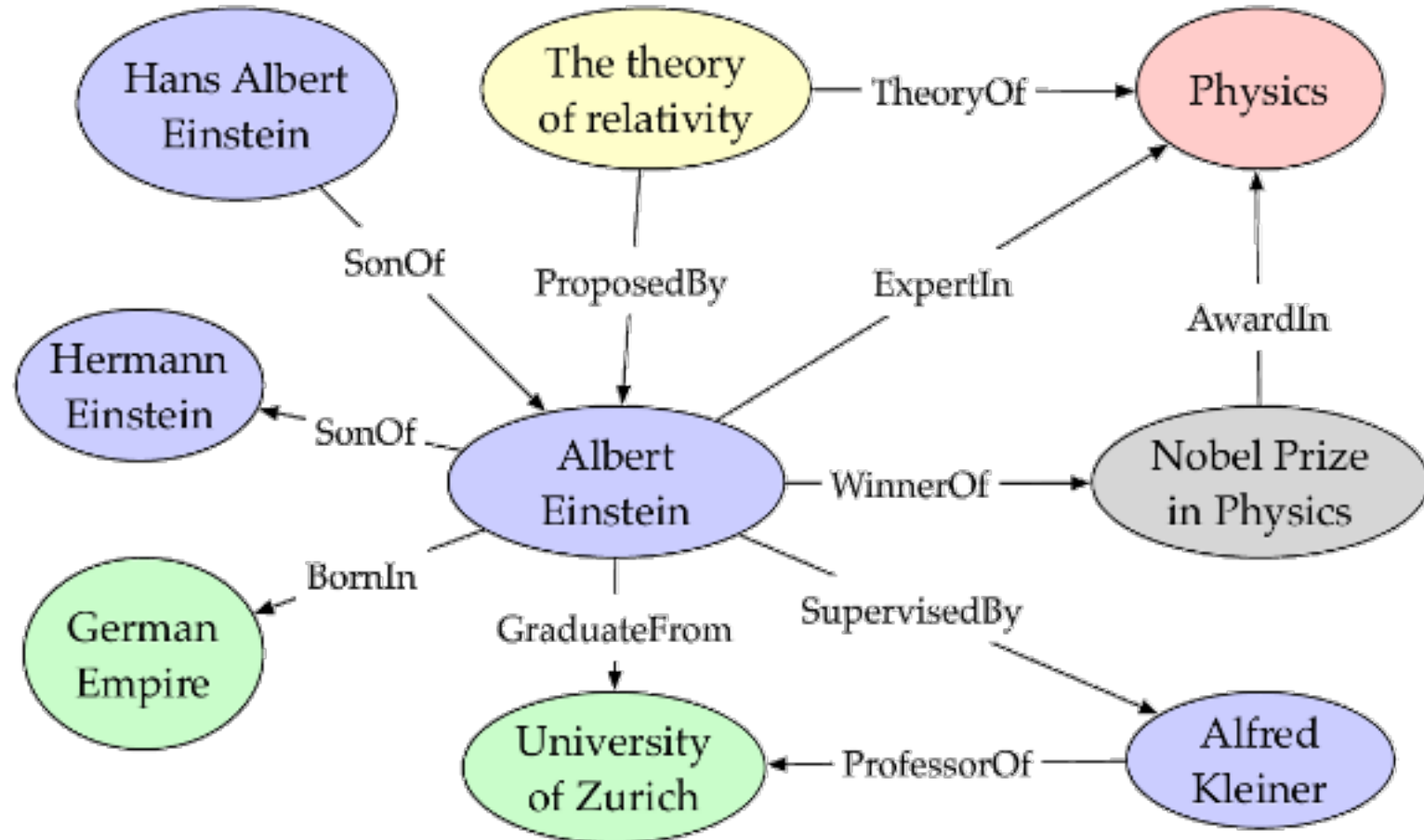
## Ontologies are more than conservative definitions in the traditional logic sense



# Towards a Knowledge Graph

Using RDF as part of an ontology allows representing the factual triplets as a graph.

- (Albert Einstein, **BornIn**, German Empire)
- (Albert Einstein, **SonOf**, Hermann Einstein)
- (Albert Einstein, **GraduateFrom**, University of Zurich)
- (Albert Einstein, **WinnerOf**, Nobel Prize in Physics)
- (Albert Einstein, **ExpertIn**, Physics)
- (Nobel Prize in Physics, **AwardIn**, Physics)
- (The theory of relativity, **TheoryOf**, Physics)
- (Albert Einstein, **SupervisedBy**, Alfred Kleiner)
- (Alfred Kleiner, **ProfessorOf**, University of Zurich)
- (The theory of relativity, **ProposedBy**, Albert Einstein)
- (Hans Albert Einstein, **SonOf**, Albert Einstein)



# The Knowledge Graph is More Than That

a.k.a. Semantic Network

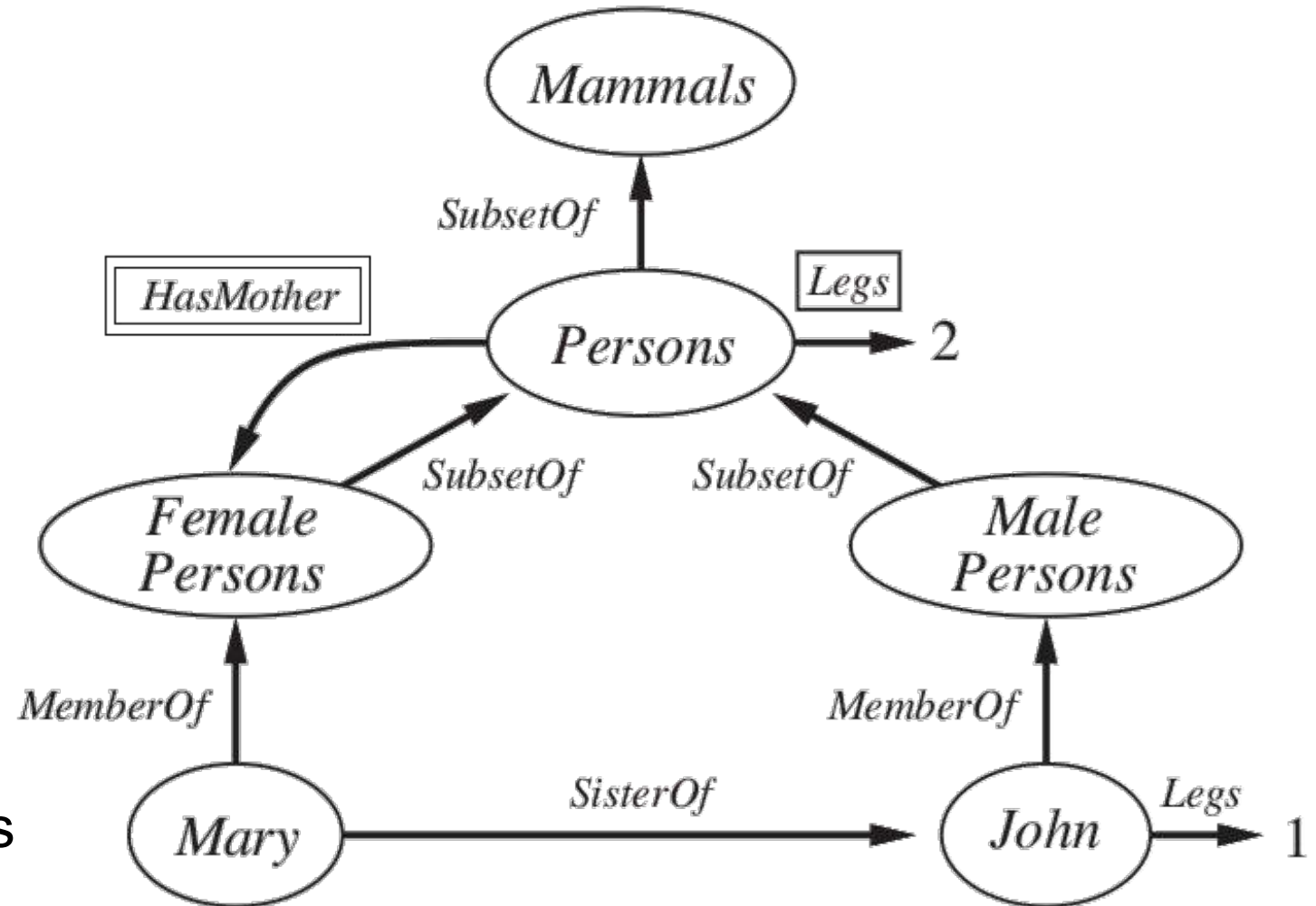
Important class of representation of knowledge

Origin: Charles Peirce “Existential Graphs”

Entities can be concepts  
as well as real-world objects

Basically classes and instances  
of ontologies combined.

Knowledge graphs are key to  
building large & powerful AI systems



Peirce, C. S., 1909. *Existential graphs*. Unpublished manuscript; reprinted in (Buchler 1955).

Russell, S. and Norvig, P., 2002. *Artificial intelligence: a modern approach*. New Jersey: Pearson Education.

Markman, A.B., 2013. *Knowledge representation*. Psychology Press.

“ If content is king,  
then metadata is queen.

Somebody on the Internet.

Thanks.

[mirco.schoenfeld@uni-bayreuth.de](mailto:mirco.schoenfeld@uni-bayreuth.de)