

Semantic Web: The Basics

Kai Eckert

University of Mannheim

The Original Vision

- Berners-Lee, Hendler, Lassila: The Semantic Web, Scientific American, May 17th 2001
- Cited (and abused) extensively in literature and science marketing



Key Ideas

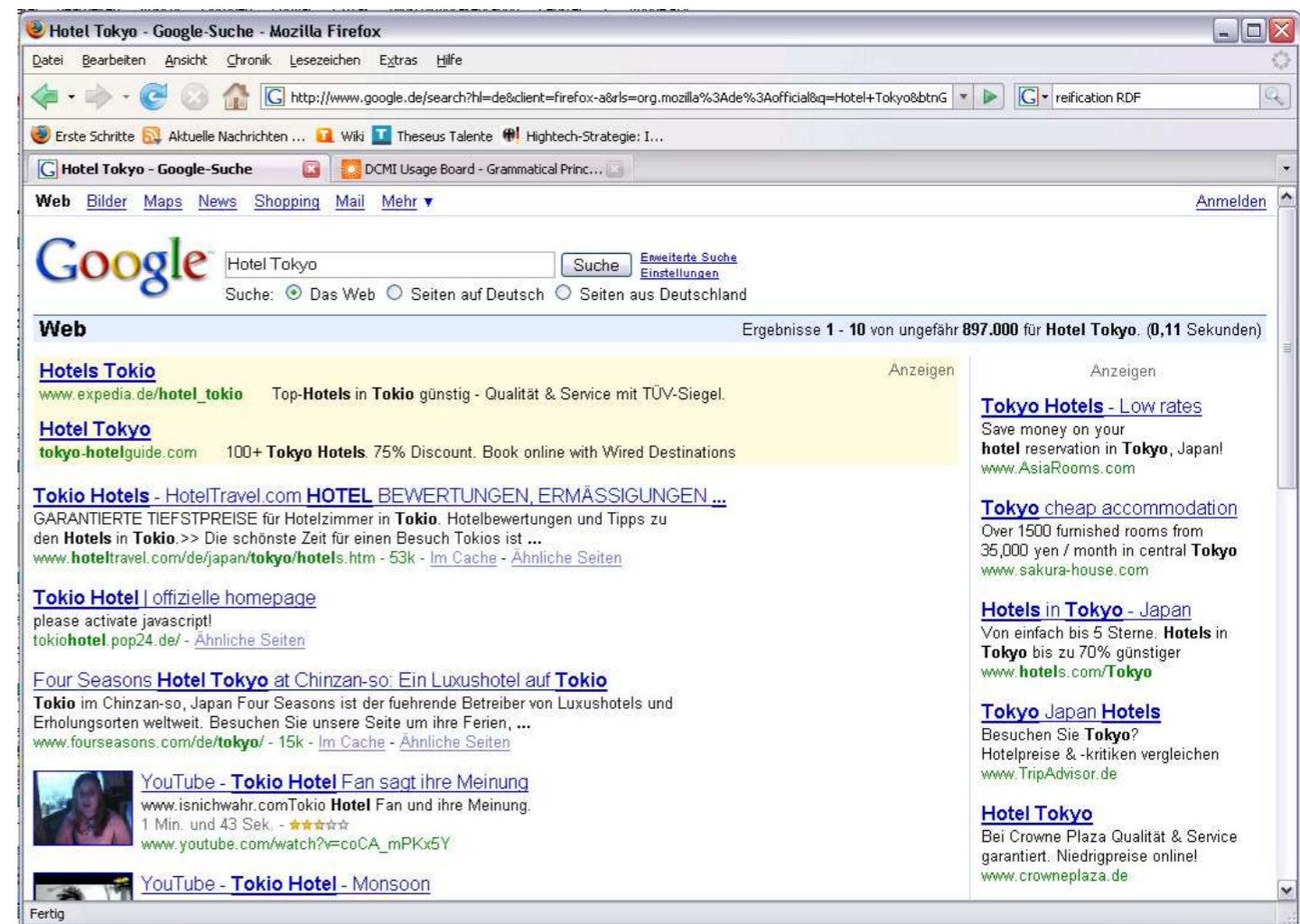
- Smart Devices
- Personal Information Agents
- Knowledge about objects, time and space
- Trusted Information

Key Technologies


- Machine-Readable Metadata
 - Based on XML
- Logic, Inference Rules and Proofs
- Ontologies
- Agent Technologies
(nowadays read: „web services“)

Why not use Google?

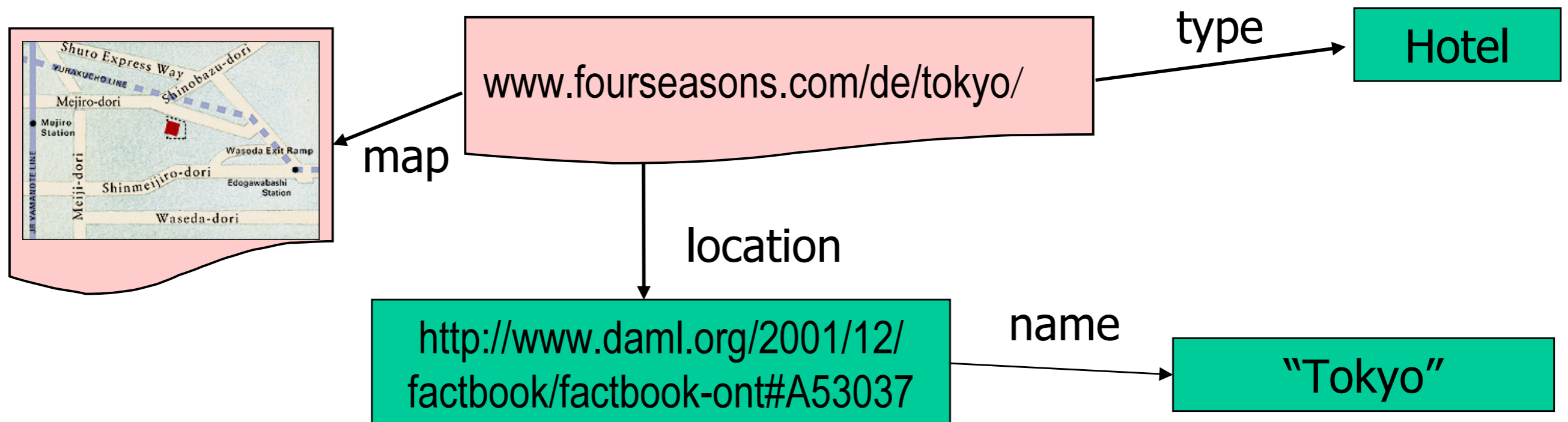
- Relevant pages instead of answers
 - Weak filtering, lots of irrelevant information
 - Biased by popularity of pages
- Hard to pose general questions
 - No way to ask about types of objects
 - Not possible to use relations



Required are:

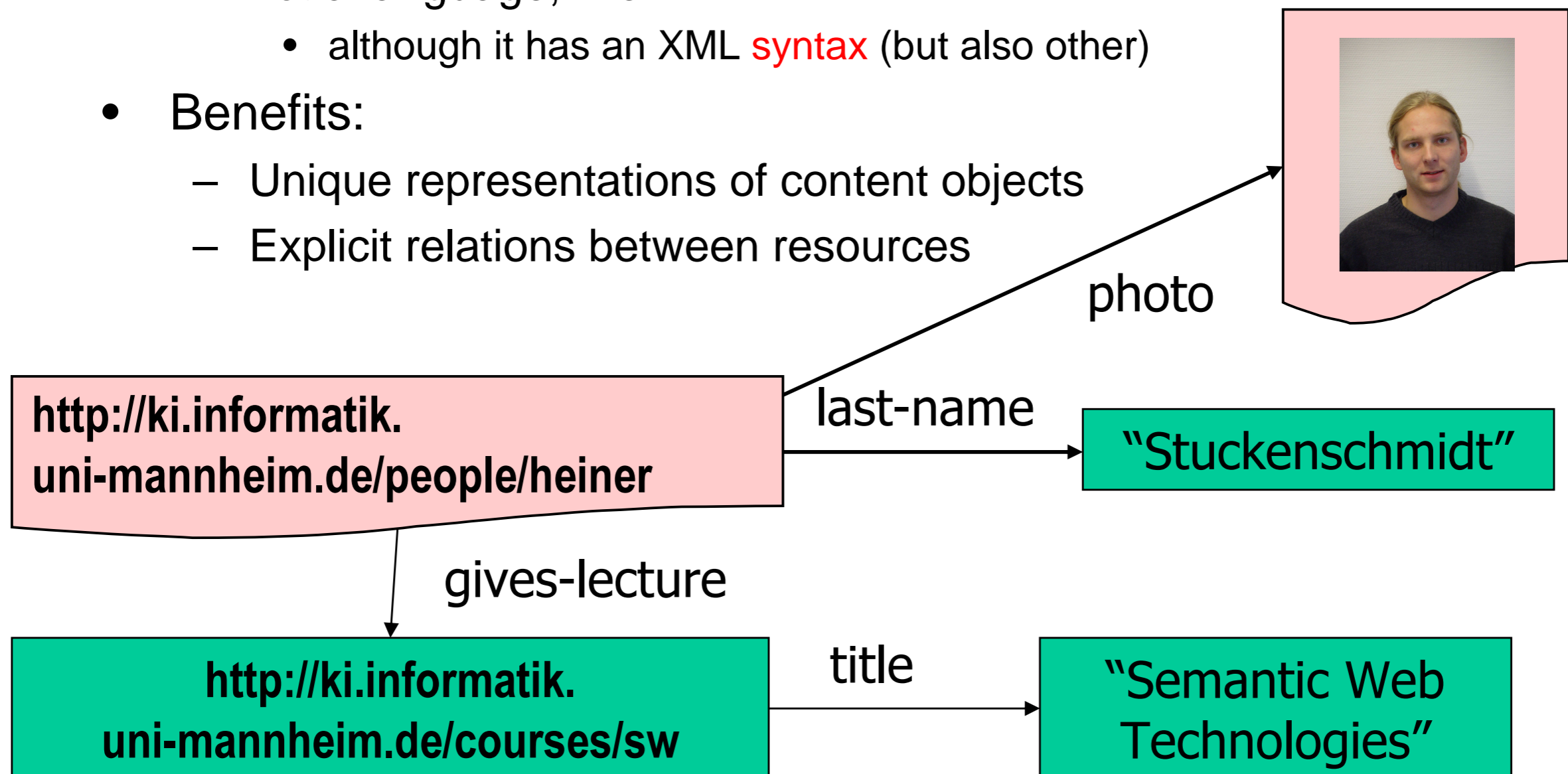
- a **standard syntax**,
 - so meta-data can be recognised as such
-  **XML, RDF, RDF Schema, OWL**
- one or more **standard vocabularies**
 - so search engines, producers and consumers all speak the same language
- lots of resources with **meta-data attached**

Structured Metadata



RDF: Resource Description Framework

- RDF is a **data model**
 - used to describe **meta-data** of a piece of data
 - not a language, like XML
 - although it has an XML **syntax** (but also other)
- Benefits:
 - Unique representations of content objects
 - Explicit relations between resources

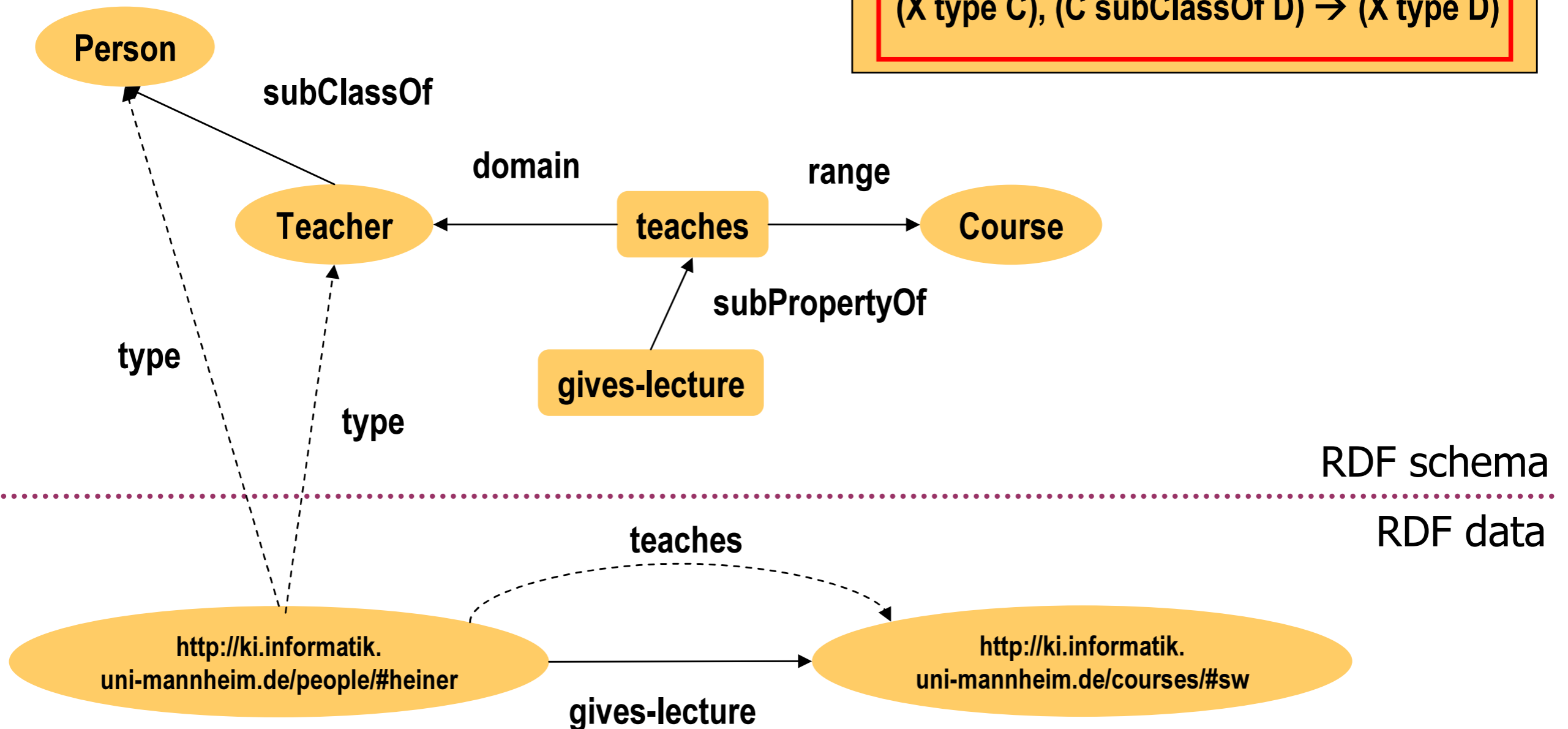


RDF Schema

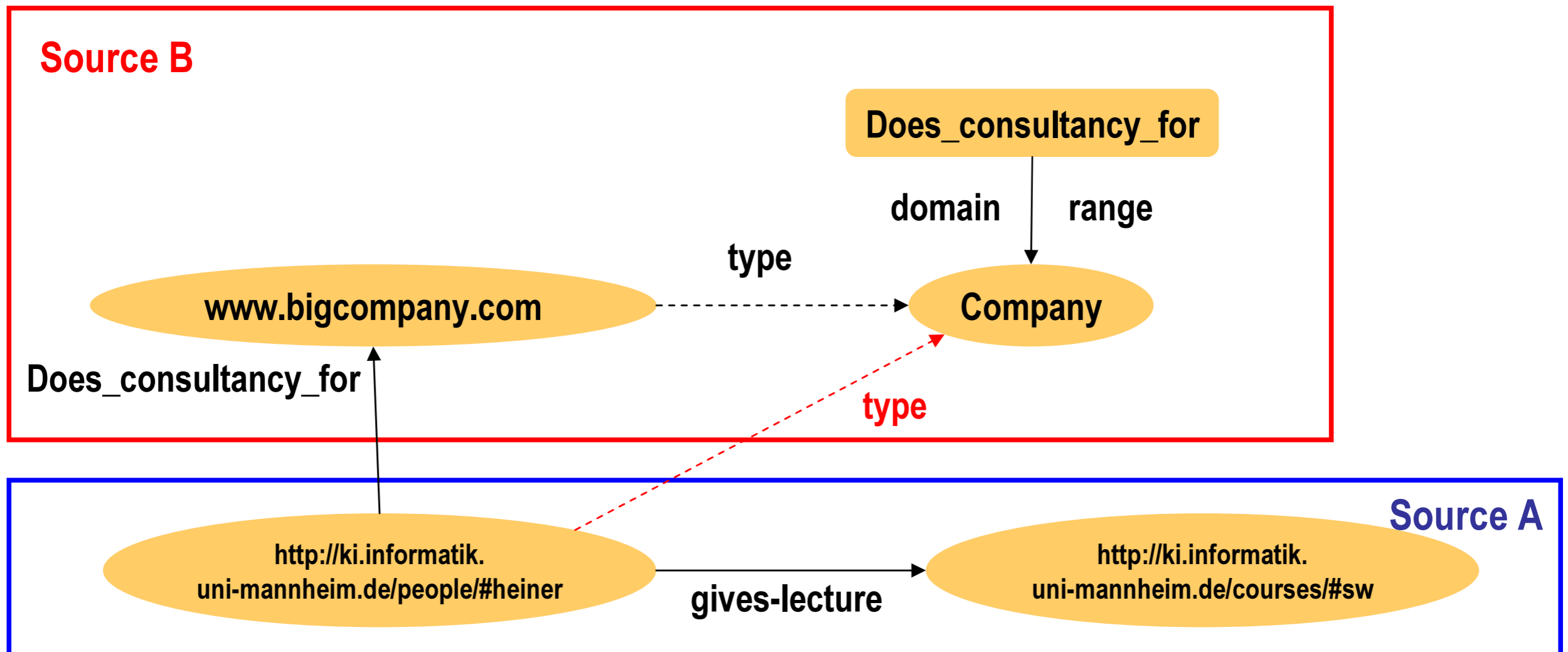
$(X R Y), (R \text{ subPropertyOf } Q) \rightarrow (X Q Y)$

$(X R Y), (R \text{ domain } C) \rightarrow (X \text{ type } C)$

$(X \text{ type } C), (C \text{ subClassOf } D) \rightarrow (X \text{ type } D)$



Problem: no semantic guarantees



Logical Reasoning about Resources

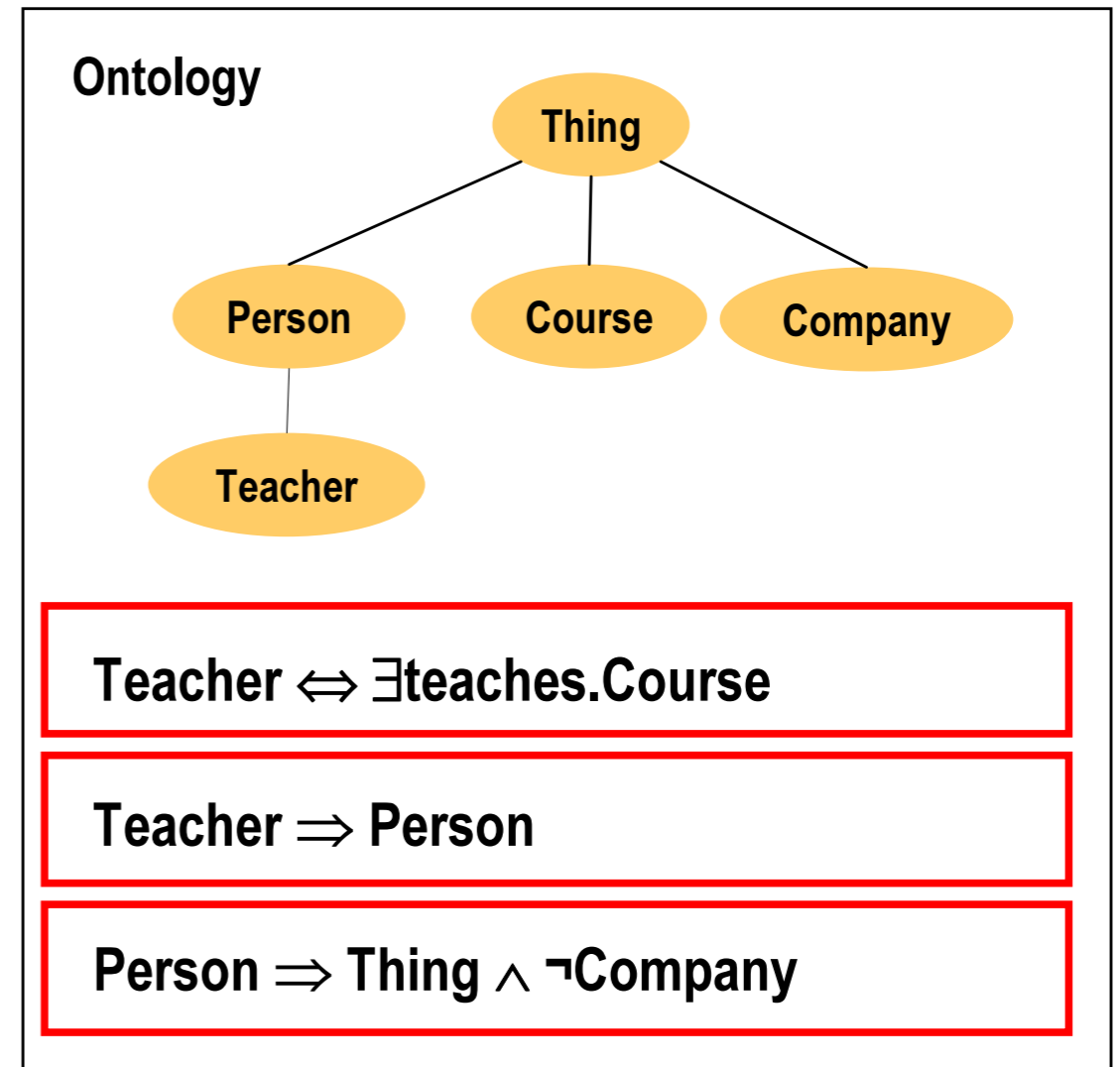
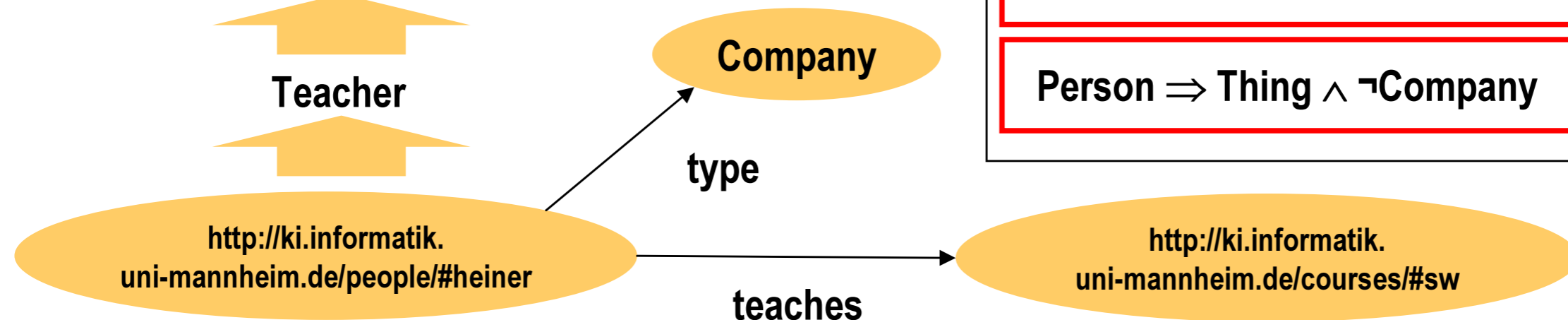
- Logical Axioms limit allowed interpretations:

$\text{Teacher} \wedge \text{Person} \wedge \text{Thing} \wedge \neg \text{Company} \wedge \text{Company}$

$\text{Teacher} \wedge \text{Person} \wedge \text{Thing} \wedge \neg \text{Company}$

$\text{Teacher} \wedge \text{Person}$

Teacher



Required are:

- a **standard syntax**,
 - so meta-data can be recognised as such



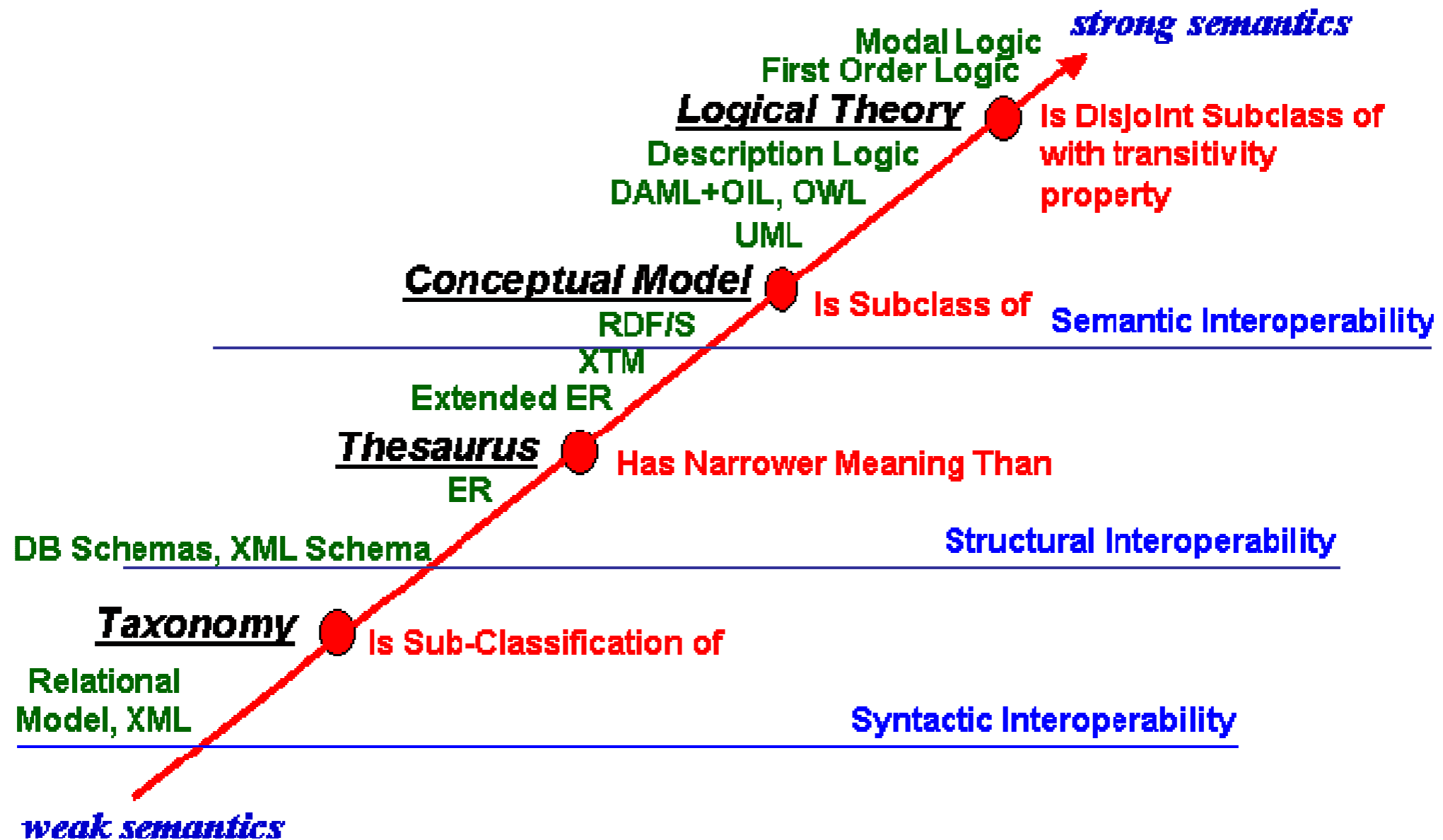
XML, RDF, RDF Schema, OWL

- one or more **standard vocabularies**
 - so search engines, producers and consumers all speak the same language
- lots of resources with **meta-data attached**

standard vocabularies (“Ontologies”)

- Identify the **key concepts** in a domain
- Identify a **vocabulary** for these concepts
- Identify **relations** between these concepts
- Make these **precise enough**
so that they can be shared between
 - humans and humans
 - **humans and machines**
 - **machines and machines**

Standardized Vocabularies



Classifications

SYSTEMATIK
 ENTOMOLOGIE
 Spider-Web-Design

Übersicht der Systematik

Progenoten (Hypothetische Urzellen)

Prokaryonten (mit sechs Abteilungen)

- Archaeobacteria (mit zwei Abteilungen)
- Eubacteria (mit vier Abteilungen)

Eukaryonten

Plantae (Pflanzen; mit 17 Abteilungen)

- I. Schleimpilze (mit drei Abteilungen)
- II. Pilze (mit zwei Abteilungen)
- III. Flechten (Lichenes; mit zwei Untergruppierungen)
- IV. Eukaryontische Algen (mit neun Abteilungen)
- IV. Embryophyten (Grüne Landpflanzen; mit drei Abteilungen)

Animalia (Tiere; mit 30 Stämmen)

- I. Protozoa (Einzeller; mit sechs Stämmen)
- II. Metazoa (Vielzeller; mit 24 Stämmen)

Periodensystem der Elemente

Haupt-		gruppen																
I	II	III	IV	V	VI	VII	VIII					VIII						
1	H															He		
2	Li	Be															Ne	
3	Na	Mg	Nebengruppen										Al	Si	P	S	Cl	Ar
4	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
5	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
6	Cs	Ba	La-Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
7	Fr	Ra	Ac-Lr	Rf	Db	Sg	Bh	Hs	Mt									

Lanthaniden	La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
Aktiniden	Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr

Metalle (Hauptgruppen)
Metalle (Nebengruppen)
Halbmetalle
Nichtmetalle
Edelgase

Autor: Peter Maisenbacher

Folie 3



Web Directories

 open directory project
[about dmoz](#) | [add URL](#) | [help](#) | [link](#) | [editor login](#)


Search [advanced](#)

- Arts**
[Movies](#), [Television](#), [Music...](#)
- Games**
[Video Games](#), [RPGs](#), [Gambling...](#)
- Kids and Teens**
[Arts](#), [School Time](#), [Teen Life...](#)
- Reference**
[Maps](#), [Education](#), [Libraries...](#)
- Shopping**
[Autos](#), [Clothing](#), [Gifts...](#)
- World**
[Deutsch](#), [Español](#), [Français](#), [Italiano](#), [Japanese](#), [Nederla...](#)


- Business**
[Jobs](#), [Industries](#), [Investing...](#)
- Health**
[Fitness](#), [Medicine](#), [Alternativ...](#)
- News**
[Media](#), [Newspapers](#), [Weathe...](#)
- Regional**
[US](#), [Canada](#), [UK](#), [Europe...](#)
- Society**
[People](#), [Religion](#), [Issues...](#)

[Become an Editor](#) Help build the largest human-edited directory
 Copyright © 1998-2001 Netscape

2,913,781 sites - 41,165 editors - 41!




Web Images Group




The web organized by Where To Look For What You Need.™

- Arts**
[Movies](#), [Music](#), [Television](#), ...
- Business**
[Companies](#), [Finance](#), [Jobs](#), ...
- Computers**
[Internet](#), [Hardware](#), [Software](#), ...
- Games**
[Board](#), [Roleplaying](#), [Video](#), ...
- Health**
[Alternative](#), [Fitness](#), [Medicine](#), ...
- World**
[Deutsch](#), [Español](#), [Français](#), [Italiano](#), [Japanese](#), [Kore...](#)

Find  What You Need.™

"Ansel Adams" results from:

- All Results
- Articles
- Shared Pages
- News
- Web

Save  What You Want.™

» [Save a copy](#) of this page

Pages you've viewed

- Results for <search>
- Results for <news>

Start Your Search

Go Vertical. Search Made Easy.

Auto	Home Living	Sports
Cities	Money	Style
Education	Music	Tech & Games
Food	Recreation	Travel
Health		

Vertical Search

The New Simple Way to Find, Save and Share.
 LookSmart vertical search makes it easy for you to find what you need. Try a more rewarding web experience by customizing search results and content -- and get essential, not exhaustive information.

Advertiser Solutions.
 Target and reach audiences you're after.
[Learn more >](#)

Publisher Solutions.
 Own your audience and control advertiser relationships.
[Learn more >](#)

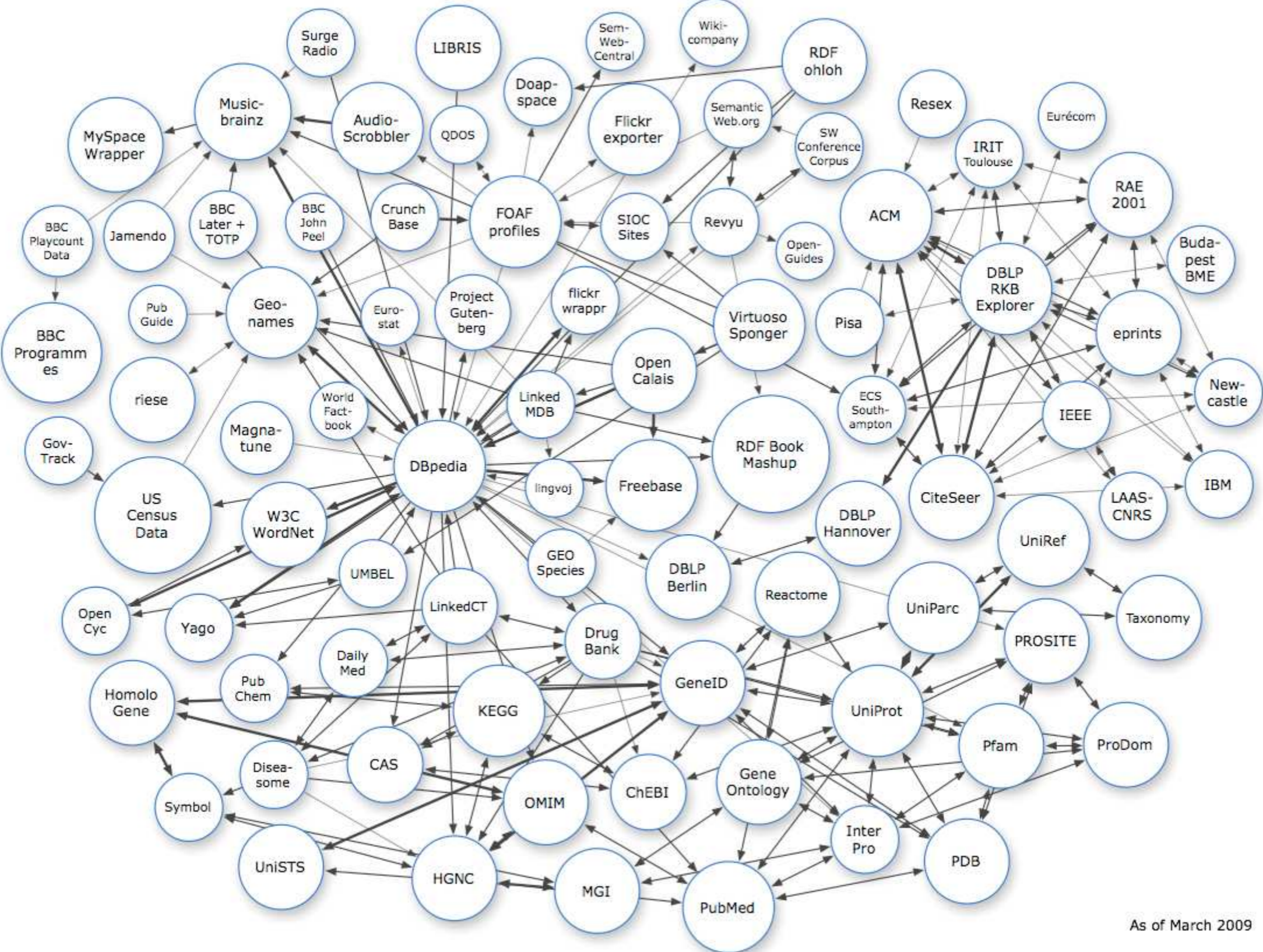
Resource Center.
 Newsletters, webinars,

wasn't until st who could success of these i said, "No one public."

elvet. This is him. ou can duplicate

isso? Go to

[Kinkade](#), [LeRoy](#)



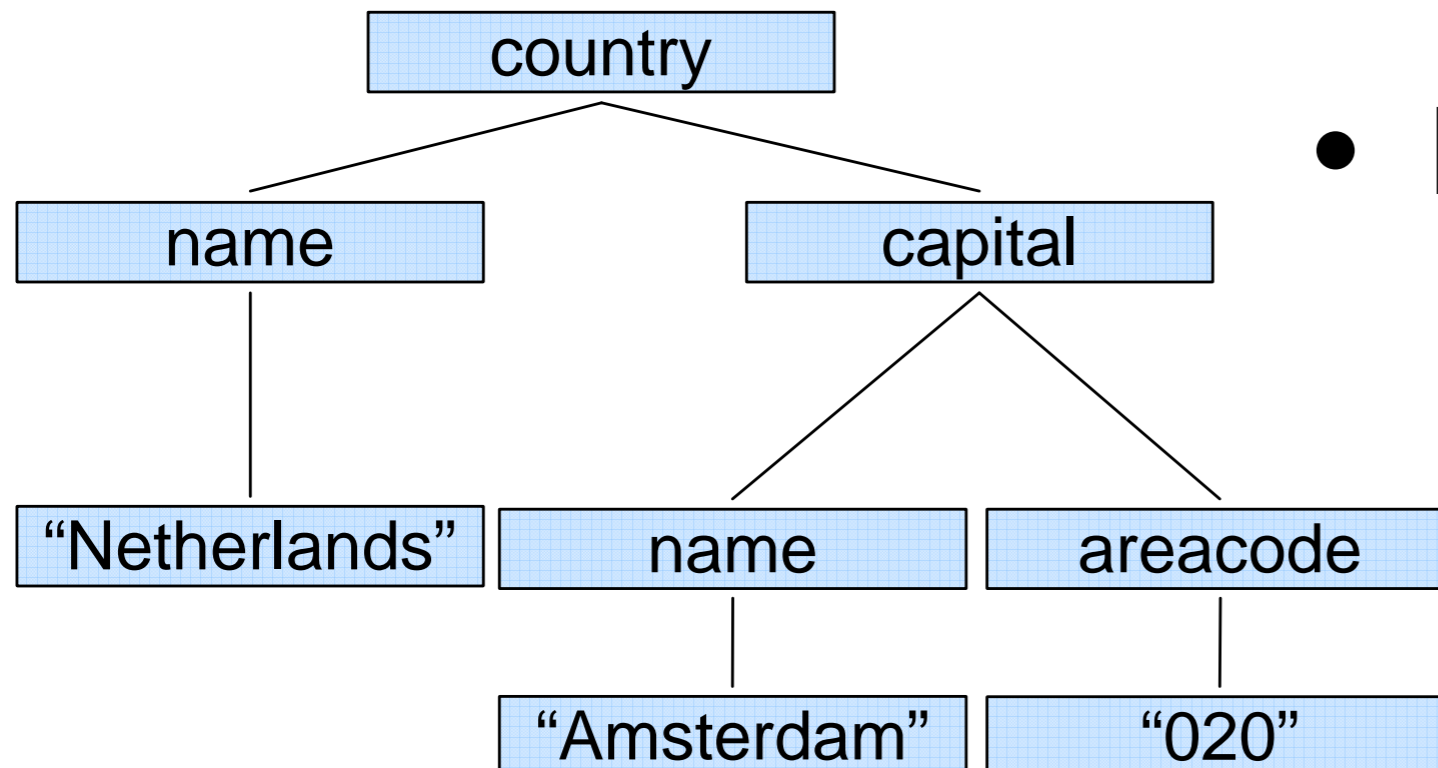
As of March 2009

RDF

- The language of the semantic web
- URIs are the words

Brief Reminder: XML

```
<country name="Netherlands">  
  <capital name="Amsterdam">  
    <areacode>020</areacode>  
  </capital>  
</country>
```



- Syntax:
 - Elements,
 - Attributes
 - Data
- Data model:
 - Tree Structure

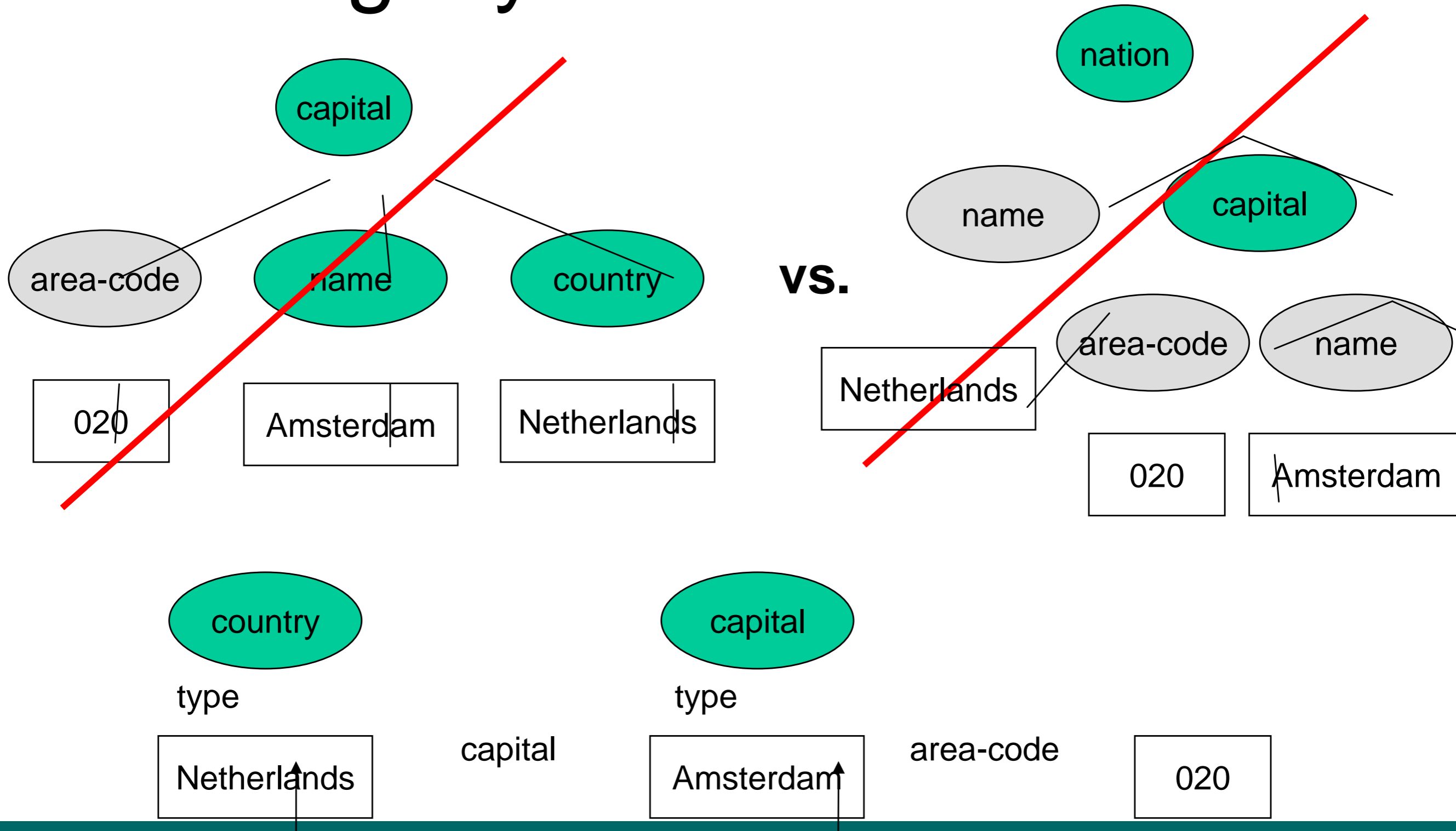
'Semantic' Problems of XML

- No commitment on a conceptual model:
 - is **country**:
 - An Object?
 - A Class?
 - A Relation?
 - What is the meaning of nesting?
- No commitment to a vocabulary
 - Is **country** the same as **nation** ?

```
<capital areacode="020">  
  <name>Amsterdam</name>  
  <country>Netherlands</country>  
</capital>
```

```
<nation name="Netherlands">  
  <capital name="Amsterdam">  
    <areacode>020</areacode>  
  </capital>  
</nation>
```

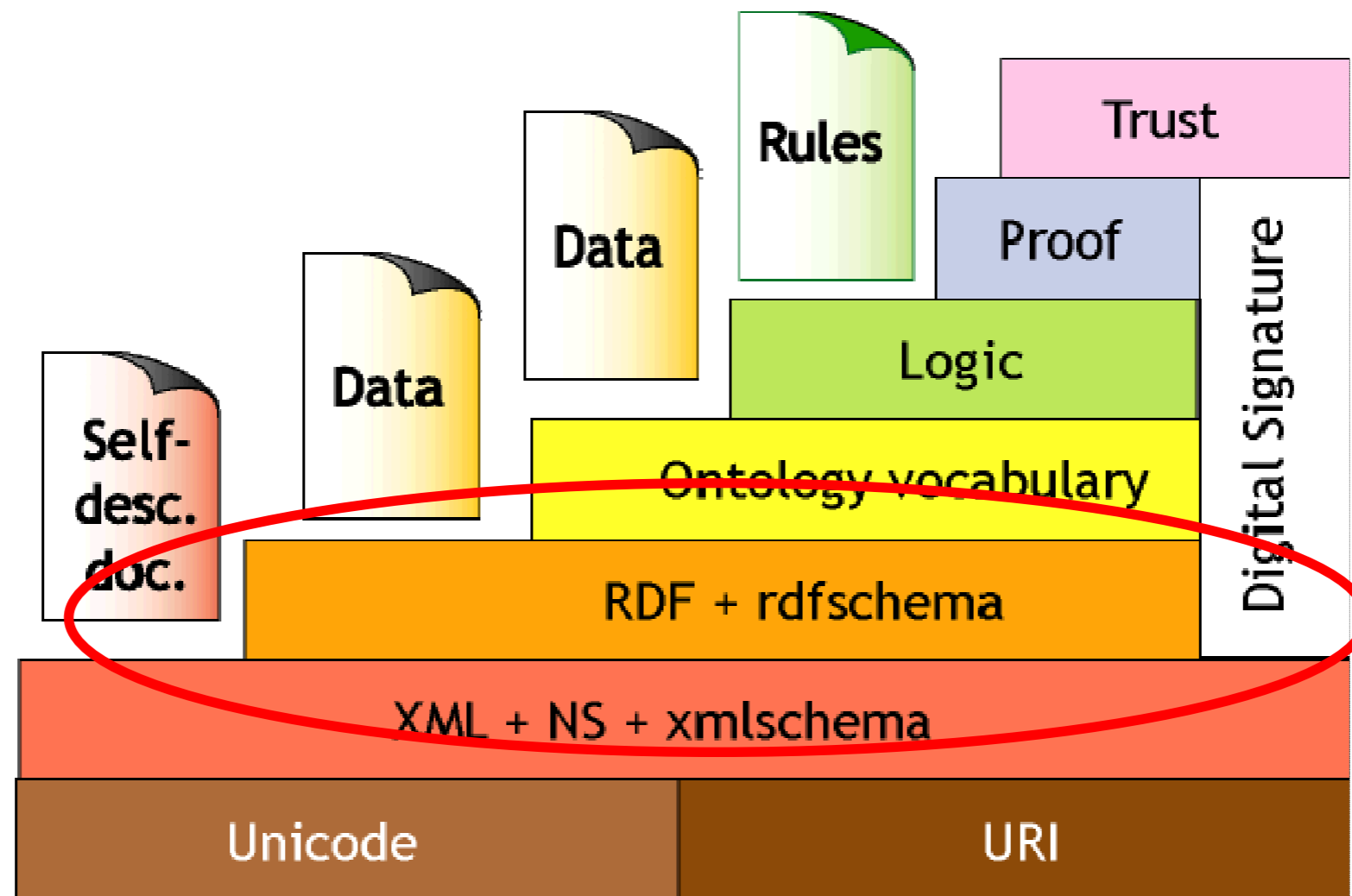
Ambiguity in XML Structures



What is RDF?

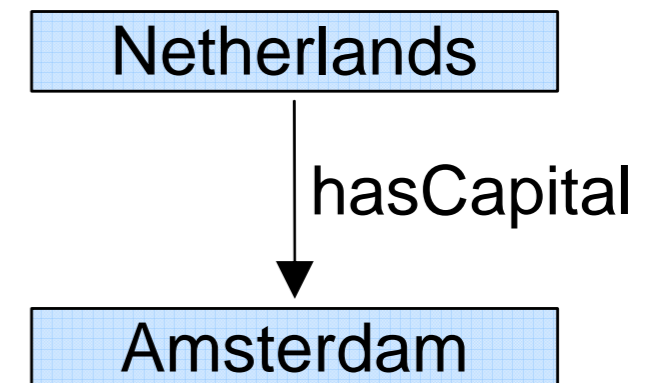
- RDF
 - “Resource Description Framework”
 - W3C Recommendation
(<http://www.w3.org/RDF>)
- RDF is a data model
 - For representing **Metadata**
(data about information resources)
 - Providing a unique, machine-readable representation of a conceptual structure

Role of RDF on the semantic web



RDF in Detail: The Data Model

- *An RDF Model is a set of statements*
- *statements are triple of the form (subject, predicate, object):*
 - (Netherlands, hasCapital, Amsterdam)



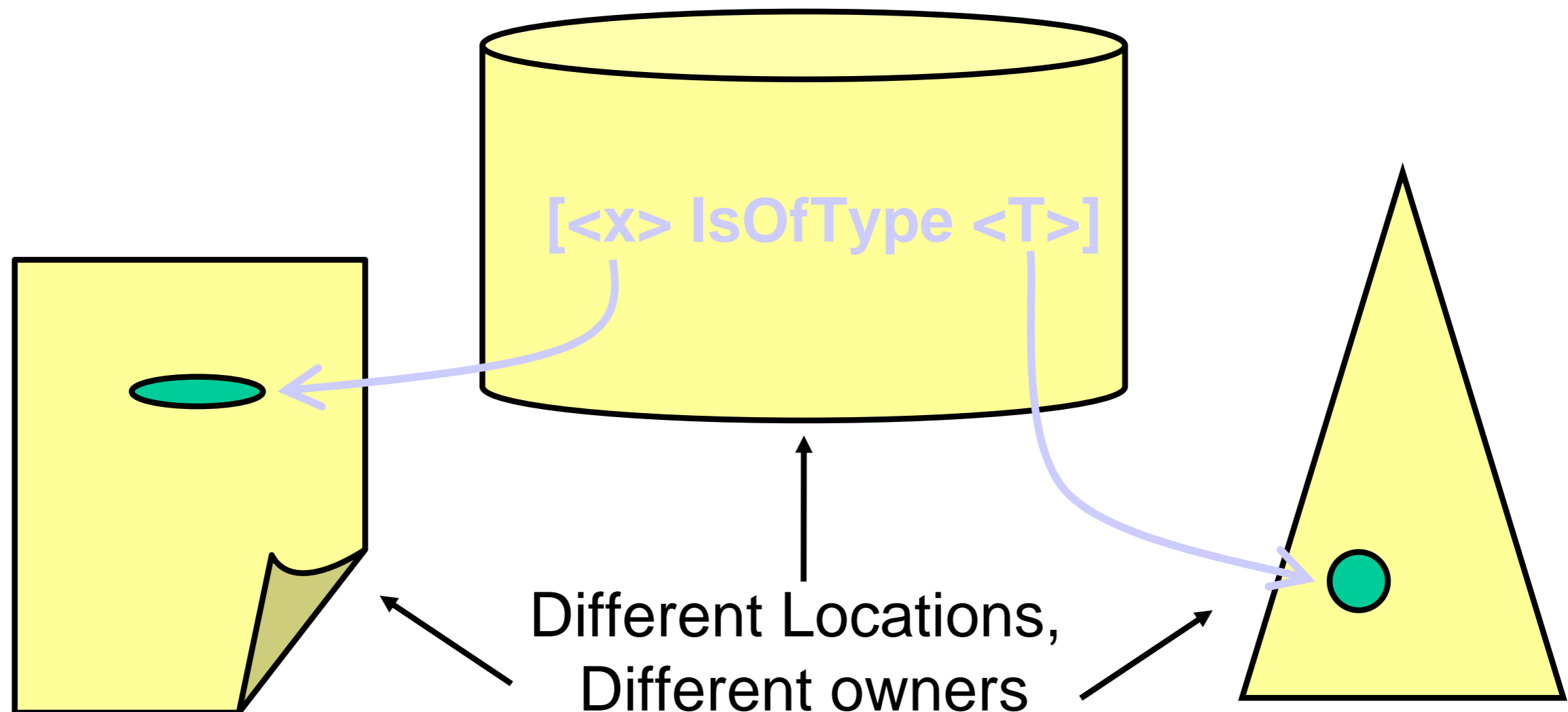
- *statements describe resources (Information Objects)*
- *A resource can be anything that has a URI:*
 - *A (part of a) Document, an image, a web page, ...*
<http://www.cs.vu.nl/index.html>
 - *A book in a library:*
 - [isbn://5031-4444-3333](http://www.isbn.org/5031-4444-3333)

What's in a triple

- There are different kinds of things in a triple:
 - URIs (= objects in the domain)
 - Literals (= Data)
 - Blank Nodes (= anonymous objects)
- There are limitations on the use in a triple
 - Subject: URI or Blank
 - Predicate: URI
 - Objects: URI, Literal or Blank

The beauty of URIs

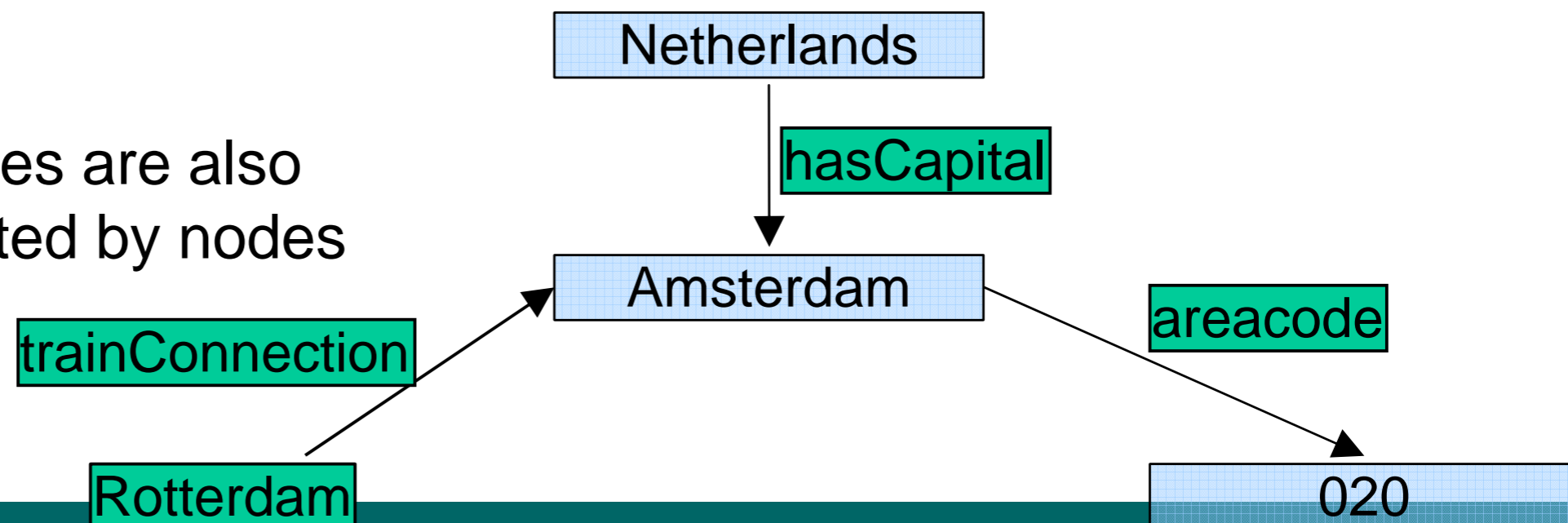
- Statements about information resources can be made independently of the resources themselves:



Statements and Graphs

- The subject of one statement can be the object (or even the predicate!) of another statement
- Sets of statements can be drawn as labelled graphs
 - Subjects/Objects are nodes
 - Predicates are directed edges (and nodes at the same time)

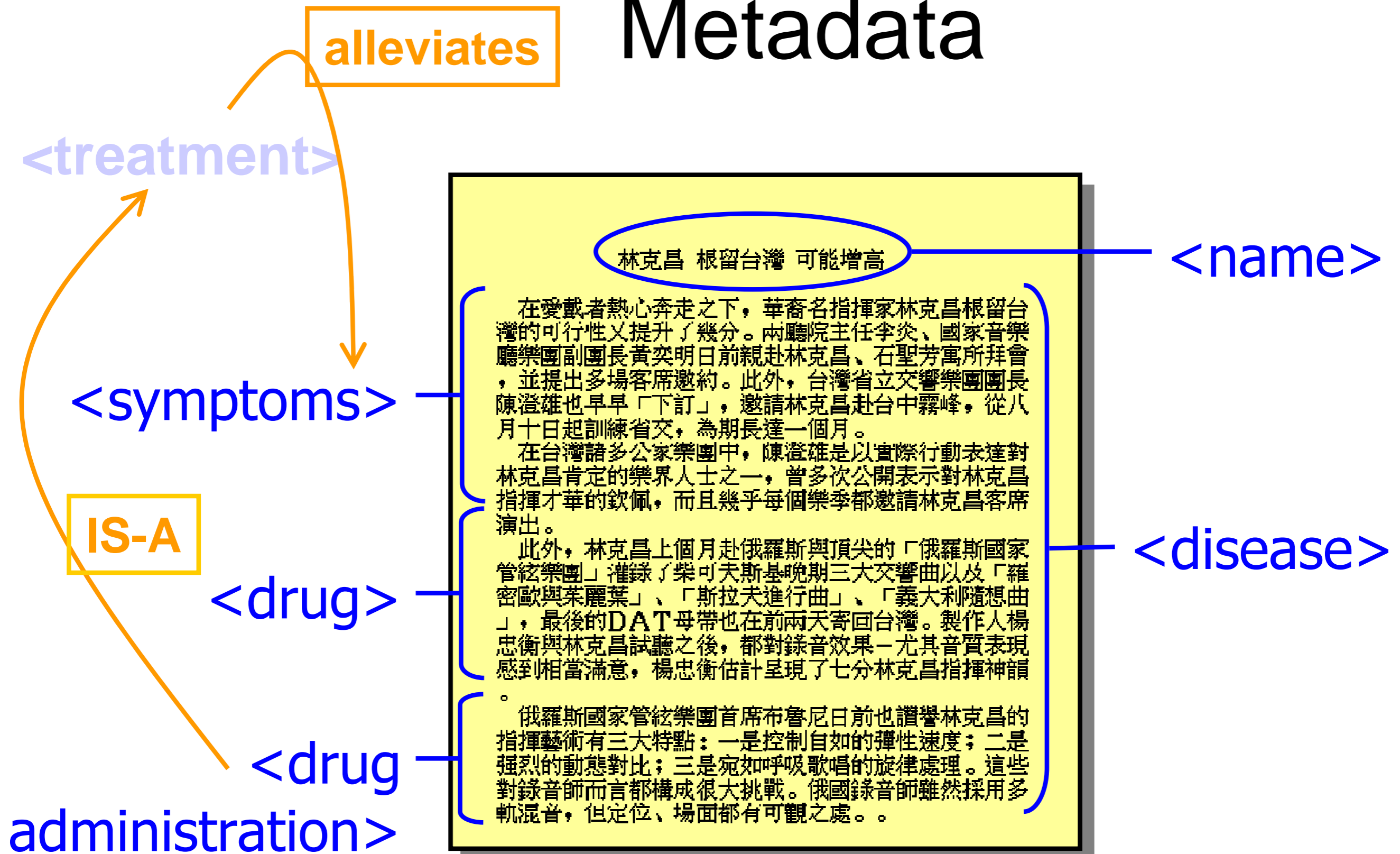
- Attention:
 - Link names are also represented by nodes



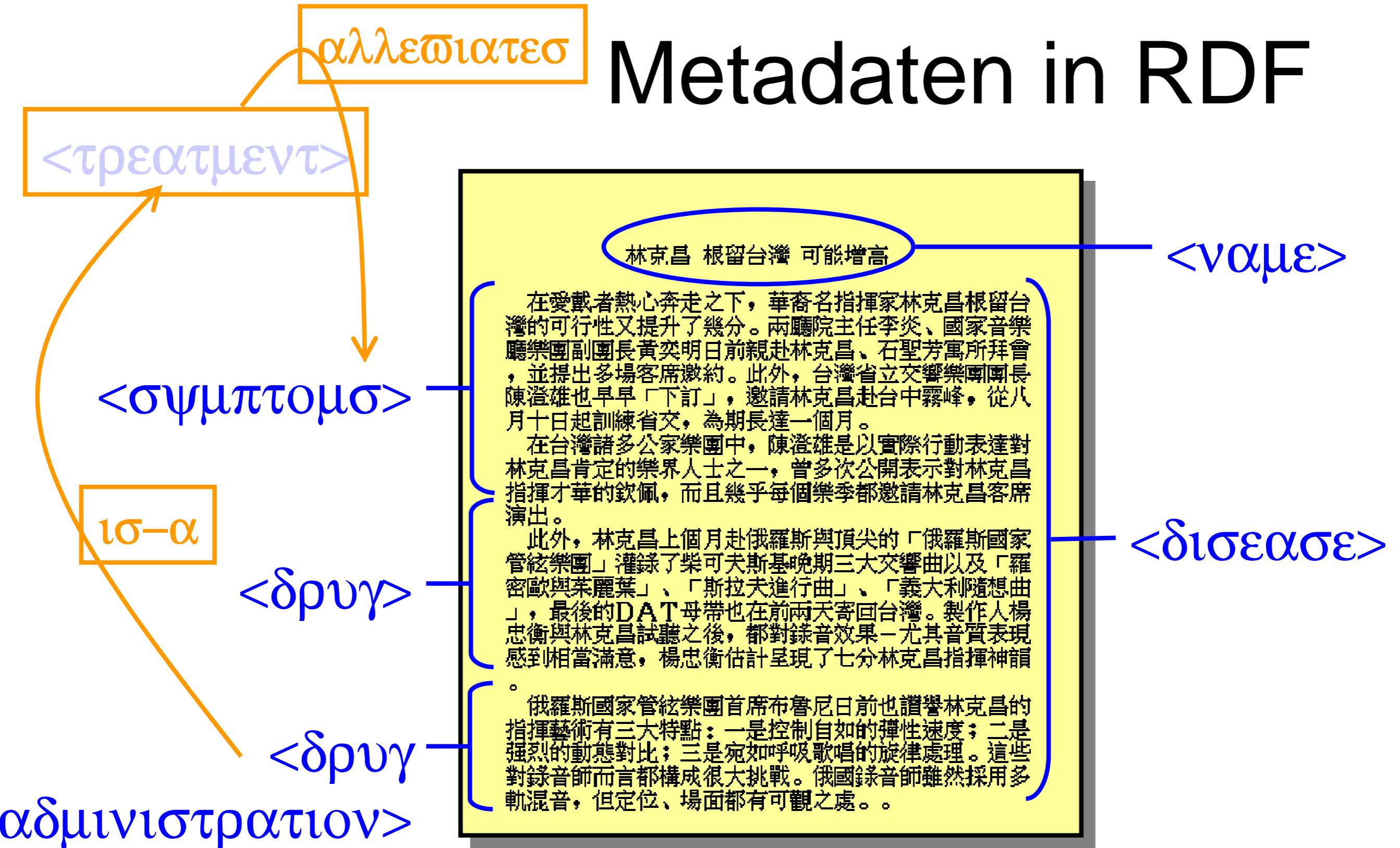
Discussion

- Advantages of RDF:
 - Explicit encoding of a conceptual model consisting of objects, relations and data values
 - Relations between resources can be made explicit, even across models and locations
 - The data model is not sensitive to syntactic variants
 - Some Operations (i.e. Merging) are trivial
- Open Problems
 - No agreement on a specific vocabulary
 - Are country and nation the same after all?
 - Which properties can a country have?

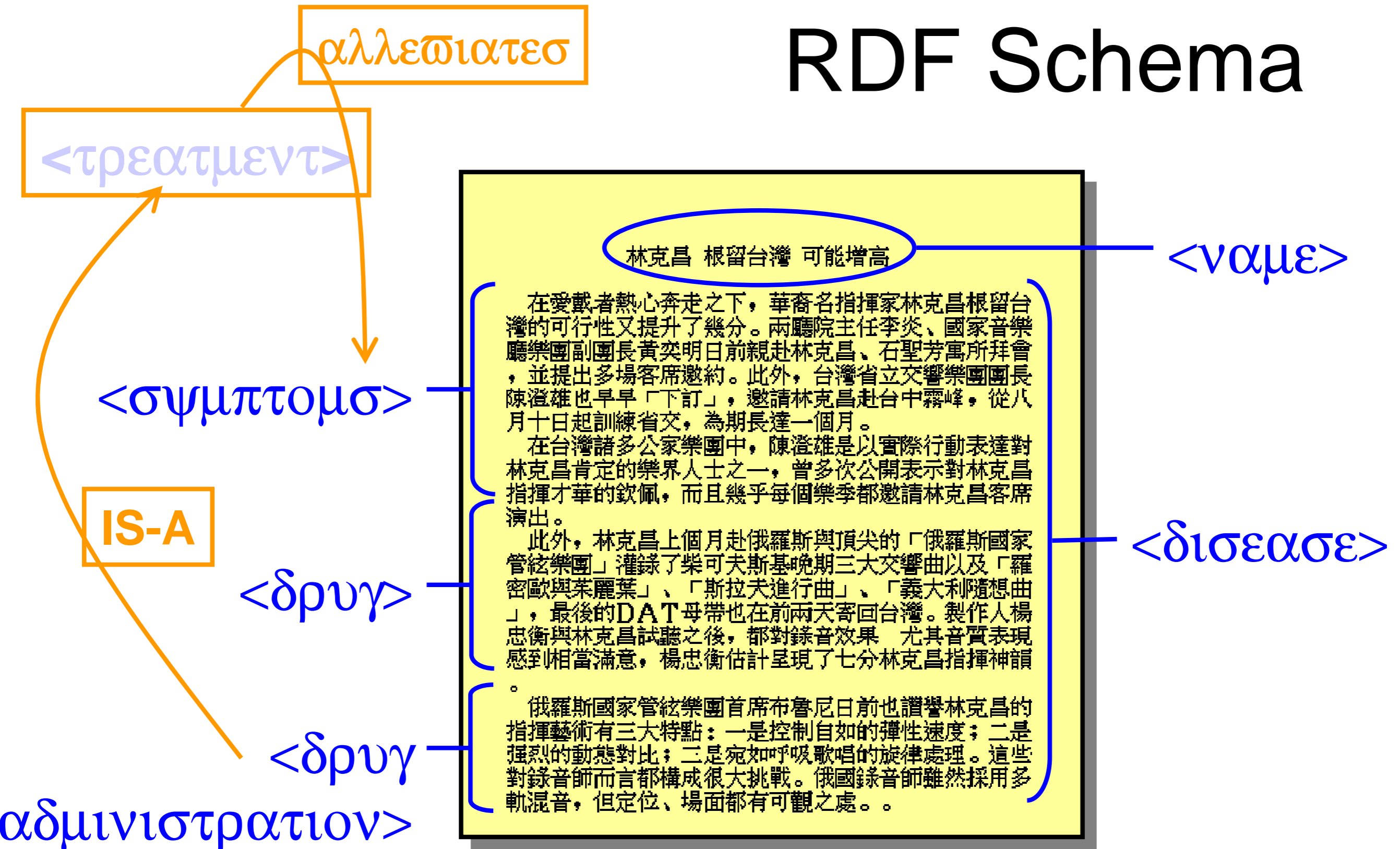
Metadata



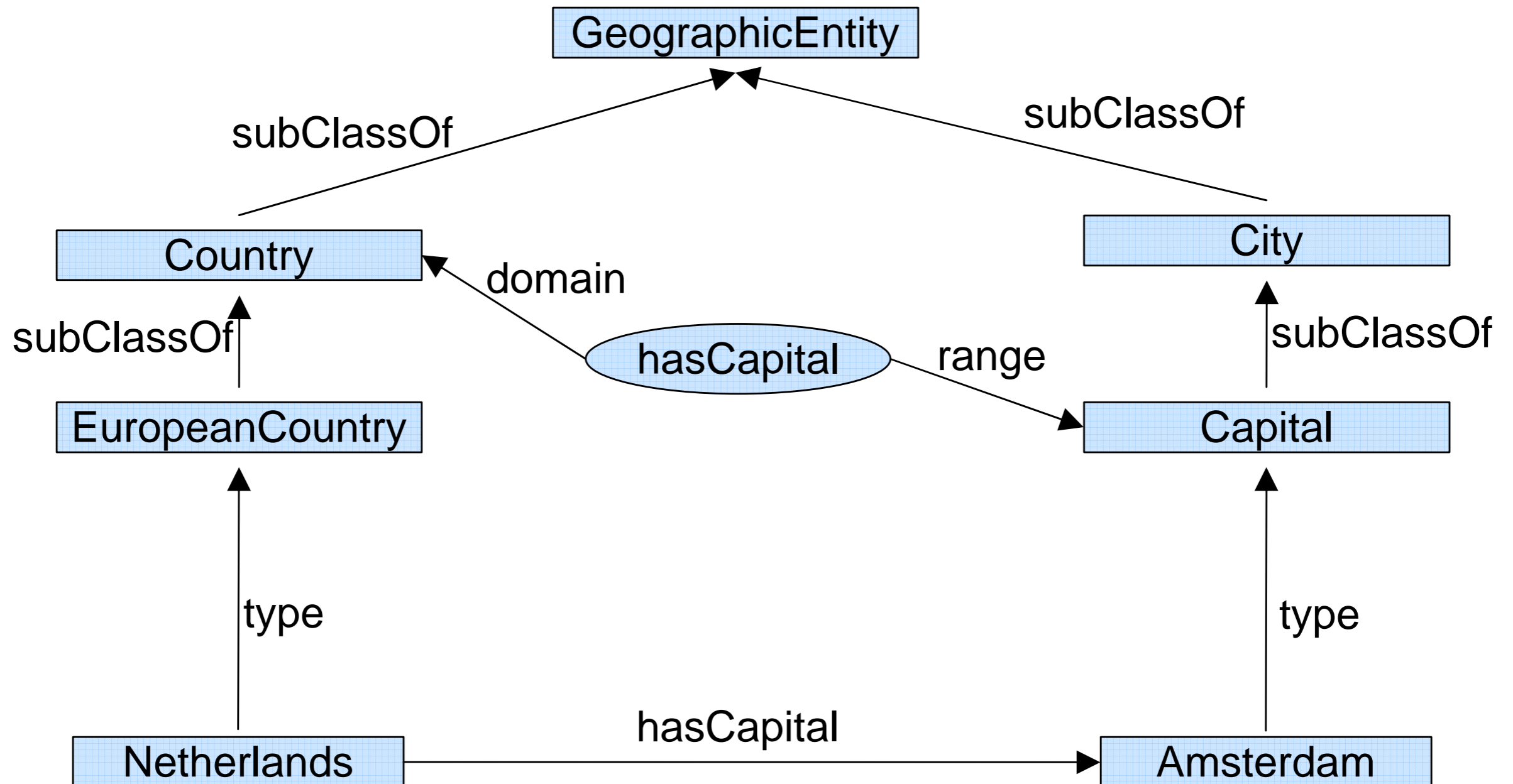
Metadaten in RDF



RDF Schema

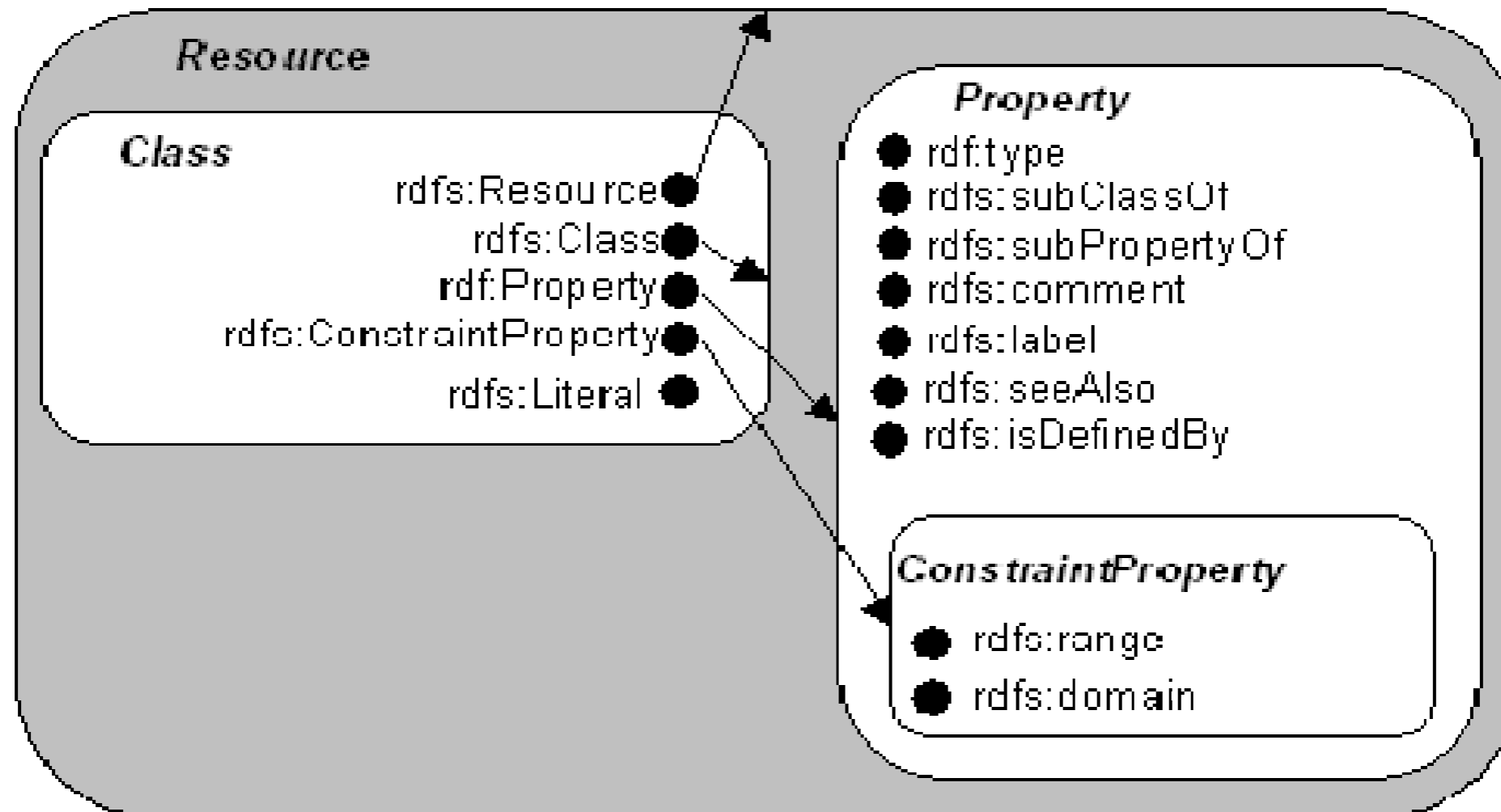


RDF Schema



What is RDF Schema?

- A set of resources and predicates with a predefined meaning:



Observations

- **Relations are first-class citizen!**
 - Relations are defined independent of classes
 - This is different from UML or ER diagrams
 - range/domain have a global meaning
- **There is NO clear separation between an RDF model and its schema, each model can contain some schema elements (=special relations)**
- **There can be multiple metalevels in the Definition**
 - Resources can be classes and instances at the same time
 - This is similar to Model-driven architectures

Derivations in RDF Schema

- Netherlands **Type** EuropeanCountry
EuropeanCountry **subClassOf** Country
→ Netherlands **Type** Country
- aspirin alleviates headache
alleviates **range** symptom
→ headache **Type** symptom

Observations

- RDF Schema semantics allows to automatically add implicit information to a model
- This can be done by recursively applying very simple derivation rules
- An RDF model can NEVER be inconsistent on the formal level („freedom of speech“)
 - Schema Elements do not define Constraints
 - They tell you what to derive

Putting it all together

- An RDF file with schema and data elements

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  xmlns:geo="http://www.geography.org/schema.rdf#">

  <rdfs:Class rdf:about="#Country">
    <rdfs:subClassOf rdfs:resource="#GeographicEntity"/>
  </rdfs:Class>

  <geo:Country rdf:about="#Netherlands">
    <geo:hasCapital rdfs:resource="#Amsterdam"/>
  </geo:Country>
  <geo:Capital rdf:about="#Amsterdam"/>

  <rdf:Property rdf:about="#hasCapital">
    <rdfs:domain rdfs:resource="#Country"/>
    <rdfs:range rdfs:resource="#Capital"/>
  </rdf:Property>
</rdf:RDF>
```

- Parts of this model could be located anywhere!

Summary

- RDF is a simple, graph-based data model for metadata on the web
- RDF has an XML syntax for:
 - Exchanging RDF Models
 - Embedding RDF Models into web pages
- Advantages over XML
 - Data model is agnostic to syntactic variations
 - Information from different models and locations can easily be linked
 - Some important operations are trivial (i.e. merging two models)
- RDF Schema defines special resources and predicates for defining vocabularies
 - Vokabular: Class, SubClassOf, domain, range
- Implicit information can be derived using simple derivation rules
- There is no clear separation between model and schema, schema elements can be part of an RDF model