

# SPARQL - Querying the Web of Data

Seminar WS 2008/2009

## RDF and the Web of Data

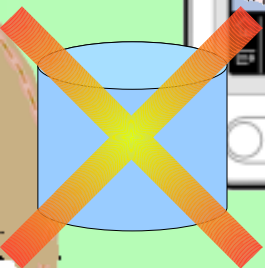
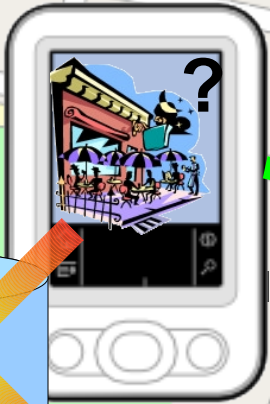
**Olaf Hartig**  
**hartig@informatik.hu-berlin.de**



$52^{\circ}57'N, 13^{\circ}42'O$



$\approx$

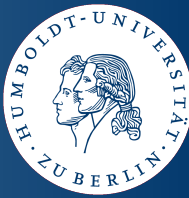


# RDF in General



- Resource Description Framework (**RDF**)
- A **resource** is basically everything
  - E.g. persons, places, Web documents, abstract concepts
- **Descriptions** of resources
  - Attributes and features
  - Relations
- The **framework** contains:
  - A data model, and
  - Languages and syntaxes

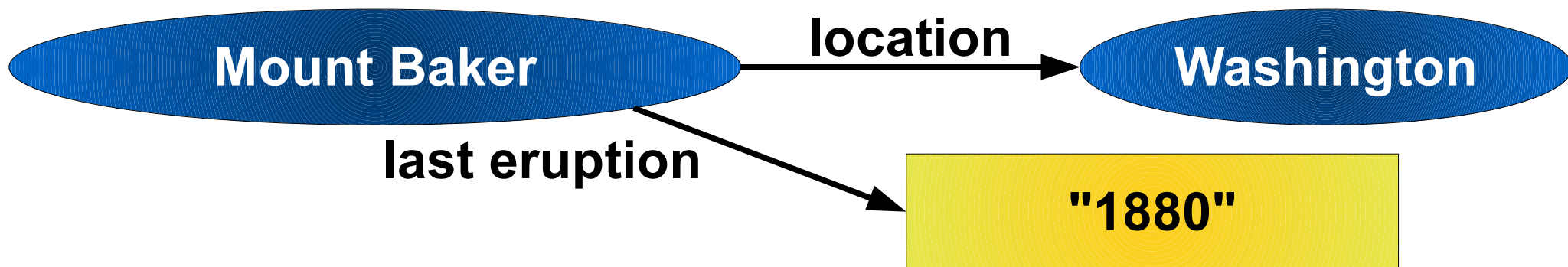
# RDF Data Model



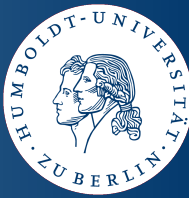
- Atoms of knowledge are **triples** (subject, predicate, object)
- **Subject:** resources
- **Predicate:** properties
- **Object:** resources or literals
- **Examples:**
  - ( Mount Baker , last eruption , "1880" )
  - ( Mount Baker , location , Washington )

# RDF Data Model

- RDF is also a **graph model**
  - Triples as directed edges
  - Subjects and objects as vertices
  - Edges labeled by predicate
- **Example:**
  - ( Mount Baker , last eruption , "1880" )
  - ( Mount Baker , location , Washington )

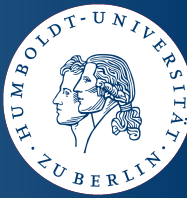


# Uniform Resource Identifier (URI)



- Globally **unique identifier** for resources
- **Syntax:**
  - URI schema (e.g. http, mailto, urn)
  - Colon character (“:”)
  - Scheme-specific part (often hierarchical)
- **Examples:**
  - `http://dbpedia.org/resource/Mount_Baker`
  - `http://www.informatik.hu-berlin.de/~hartig/foaf.rdf#olaf`
  - `urn:isbn:0-486-27557-4`

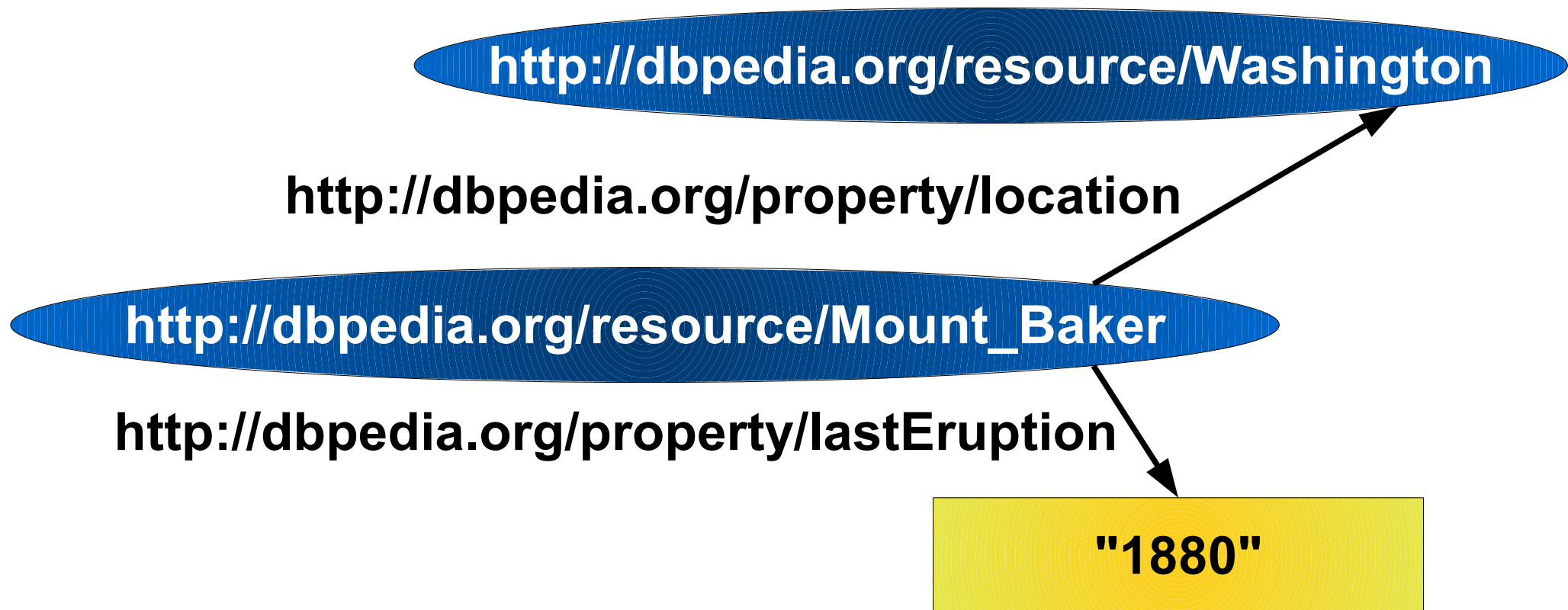
# Uniform Resource Identifier (URI)



- URIs extend the concept of URLs
  - URL of a Web document usually used as its URI
  - Attention: URIs identify not only Web documents
- **Example:**
  - **Me:**  
`http://www.informatik.hu-berlin.de/~hartig/foaf.rdf#olaf`
  - **RDF document about me:**  
`http://www.informatik.hu-berlin.de/~hartig/foaf.rdf`
  - **HTML document about me:**  
`http://www.informatik.hu-berlin.de/~hartig/index.html`

# Example (revisited)

- ([http://dbpedia.org/resource/Mount\\_Baker](http://dbpedia.org/resource/Mount_Baker),  
<http://dbpedia.org/property/lastEruption>, "1880")
- ([http://dbpedia.org/resource/Mount\\_Baker](http://dbpedia.org/resource/Mount_Baker),  
<http://dbpedia.org/property/location>,  
<http://dbpedia.org/resource/Washington>)





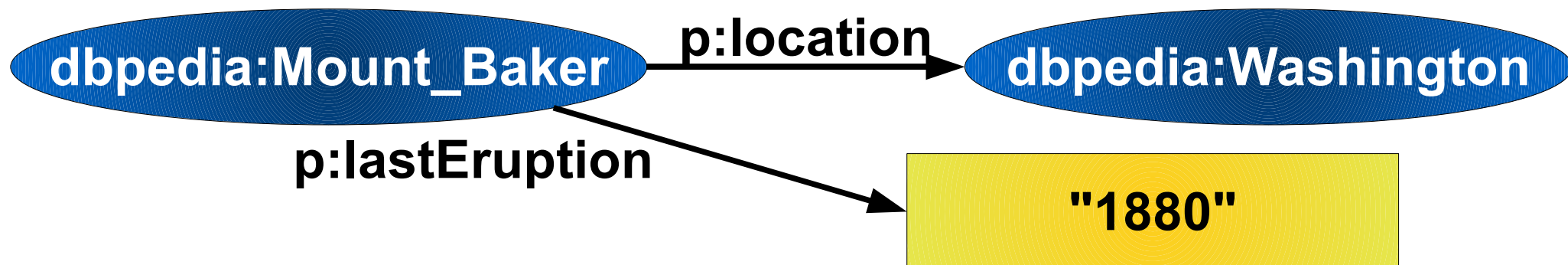
# Compact URIs (CURIE)



- Abbreviated Notation for URIs
- **Syntax:**
  - Prefix name (references the prefix of the URI)
  - Colon character (“:”)
  - Reference part
- URI by **concatenating** the prefix and the reference part
- **Examples:**
  - `dbpedia:Mount_Baker` for  
`http://dbpedia.org/resource/Mount_Baker`
  - `myfoaf:olaf` for  
`http://www.informatik.hu-berlin.de/~hartig/foaf.rdf#olaf`

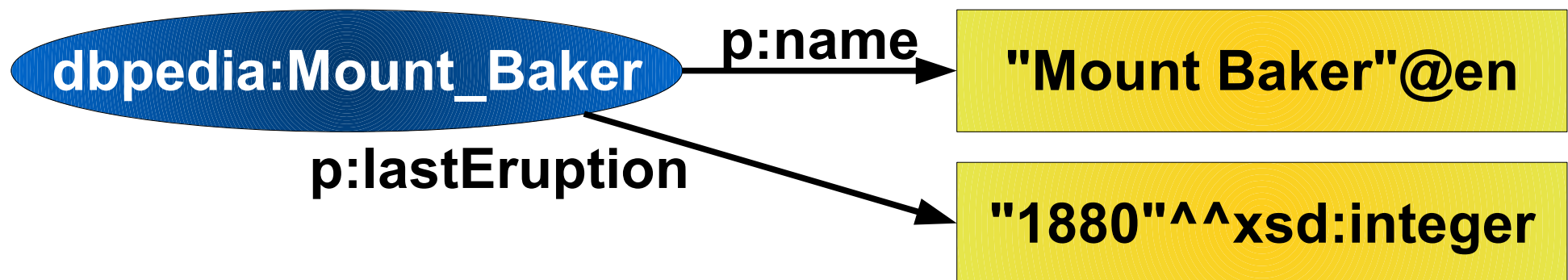
# Example with CURIEs

- Using
  - *dbpedia* for prefix *http://dbpedia.org/resource/*
  - *p* for prefix *http://dbpedia.org/property/*
- we have
  - (dbpedia:Mount\_Baker, p:lastEruption, "1880")
  - (dbpedia:Mount\_Baker, p:location, dbpedia:Washington)

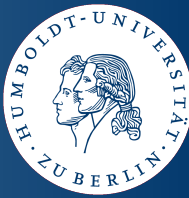


# Literals

- Literals may occur in the **object position** of triples
- Represented by strings
- Literal strings interpreted by **datatypes**
  - Datatype identified by a URI
  - Common to use the XML Schema datatypes
  - No datatype: interpreted as *xsd:string*
- Untyped literals may have **language tags** (e.g. @de)



# N3 – A Readable Syntax for RDF



- Simple notation to list RDF triples:
  - Triples separated by a period (“.”) character
  - Example:

```
<http://dbpedia.org/resource/Mount_Baker>  
  <http://dbpedia.org/property/lastEruption>  
    "1880"^^xsd:integer .  
<http://dbpedia.org/resource/Mount_Baker>  
  <http://dbpedia.org/property/location>  
    <http://dbpedia.org/resource/Washington> .
```

# N3 – A Readable Syntax for RDF



- N3 allows the use of **CURIEs**:
  - **@prefix** directive binds a prefix to a namespace **URI**

```
@prefix dbpedia : <http://dbpedia.org/resource/> .
```

```
@prefix p : <http://dbpedia.org/property/> .
```

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
```

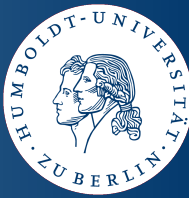
```
dbpedia:Mount_Baker p:lastEruption "1880"^^xsd:integer .
```

```
dbpedia:Mount_Baker p:location dbpedia:Washington .
```

```
dbpedia:Washington p:borderingstates dbpedia:Oregon .
```

```
dbpedia:Washington p:borderingstates dbpedia:Idaho .
```

# N3 – A Readable Syntax for RDF



- N3 provides some syntactic sugar:
  - Property lists separated by a semicolon (“;”) character
  - Object lists separated by a comma (“,”) character

```
@prefix dbpedia : <http://dbpedia.org/resource/> .
```

```
@prefix p : <http://dbpedia.org/property/> .
```

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
```

```
dbpedia:Mount_Baker p:lastEruption "1880"^^xsd:integer ;  
                    p:location      dbpedia:Washington .
```

```
dbpedia:Washington p:borderingstates dbpedia:Oregon ,  
                    dbpedia:Idaho .
```

# N3 – A Readable Syntax for RDF



- More syntactic sugar:
  - Shortcuts for number literals

```
dbpedia:Mount_Baker p:lastEruption "1880"^^xsd:integer ;  
                    geo:lat "48.777222"^^xsd:float ;  
                    geo:long "-121.813332"^^xsd:float .
```

**Equivalent:**

```
dbpedia:Mount_Baker p:lastEruption 1880 ;  
                    geo:lat 48.777222 ;  
                    geo:long -121.813332 .
```

# Classification



- The predefined **property** *rdf:type* enables classifications
  - Object resource represents a category / **class** of things
  - Subject resource is an **instance** of that class

```
@prefix dbpedia: <http://dbpedia.org/resource/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns> .
@prefix umbel-sc: <http://umbel.org/umbel/sc/> .
@prefix yago: <http://dbpedia.org/class/yago/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
dbpedia:Oregon rdf:type yago:StatesOfTheUnitedStates .
dbpedia:Mount_Baker rdf:type umbel-sc:Mountain .
umbel-sc:Mountain skos:definition "Each instance of ▶
Mountain is a topographical feature of significantly ▶
higher elevation ..."@en
```



# Classification



- **Class membership is not exclusive**
  - **I.e. instances may have multiple types**

```
dbpedia:Mount_Baker rdf:type umbel-sc:Mountain ,  
                        umbel-sc:Volcano .
```

- **Classes may be instances of other classes!**

```
dbpedia:Mount_Baker rdf:type umbel-sc:Mountain .  
umbel-sc:Mountain rdf:type umbel-ac:ExistingObjectType .
```

- **Syntactical distinction between classes and instances a priori impossible**

# RDF Schema in General



- **RDF Schema enables specification of schema knowledge**
  - **Definition of the vocabulary used in triples**
  - **Class hierarchies, property hierarchies**
- **RDF Schema semantics enable elementary inferences**

# Predefined Classes

- **RDF Schema defines the following classes**

- ***rdfs:Resource*** – class of all resources

- ***rdfs:Literal*** – class of all literals

- ***rdfs:Class*** – class of all classes

it holds: ( ***rdfs:Class*** , ***rdf:type*** , ***rdfs:Class*** )

- ***rdfs:Datatype*** – class of all datatypes

- ***rdf:Property*** – class of all properties

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns>
```

```
@prefix rdfs : <http://www.w3.org/2000/01/rdf-schema#> .
```

```
@prefix umbel-sc : <http://umbel.org/umbel/sc/> .
```

```
umbel-sc:Mountain rdf:type rdfs:Class .
```

# Class Hierarchies



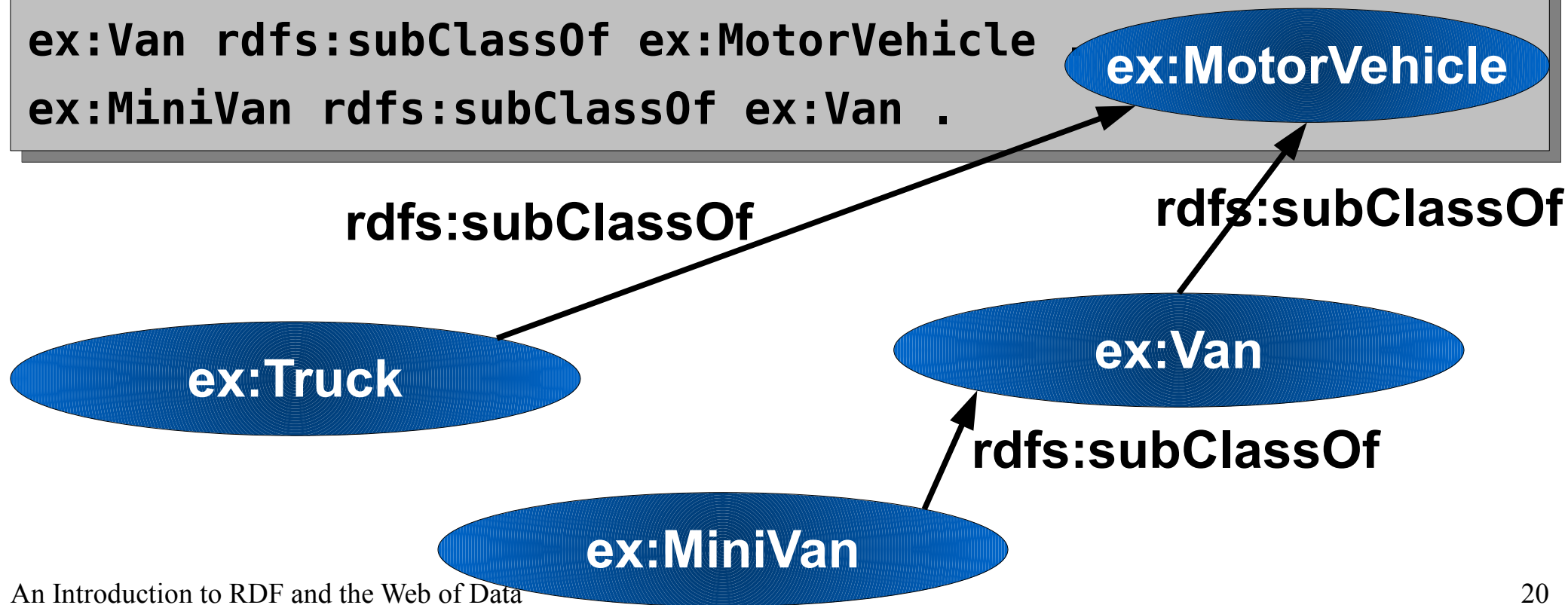
- *rdfs:subClassOf* enables the definition of class hierarchies

```
@prefix rdfs : <http://www.w3.org/2000/01/rdf-schema#> .  
@prefix ex : <http://example.org/> .
```

```
ex:Truck rdfs:subClassOf ex:MotorVehicle .
```

```
ex:Van rdfs:subClassOf ex:MotorVehicle
```

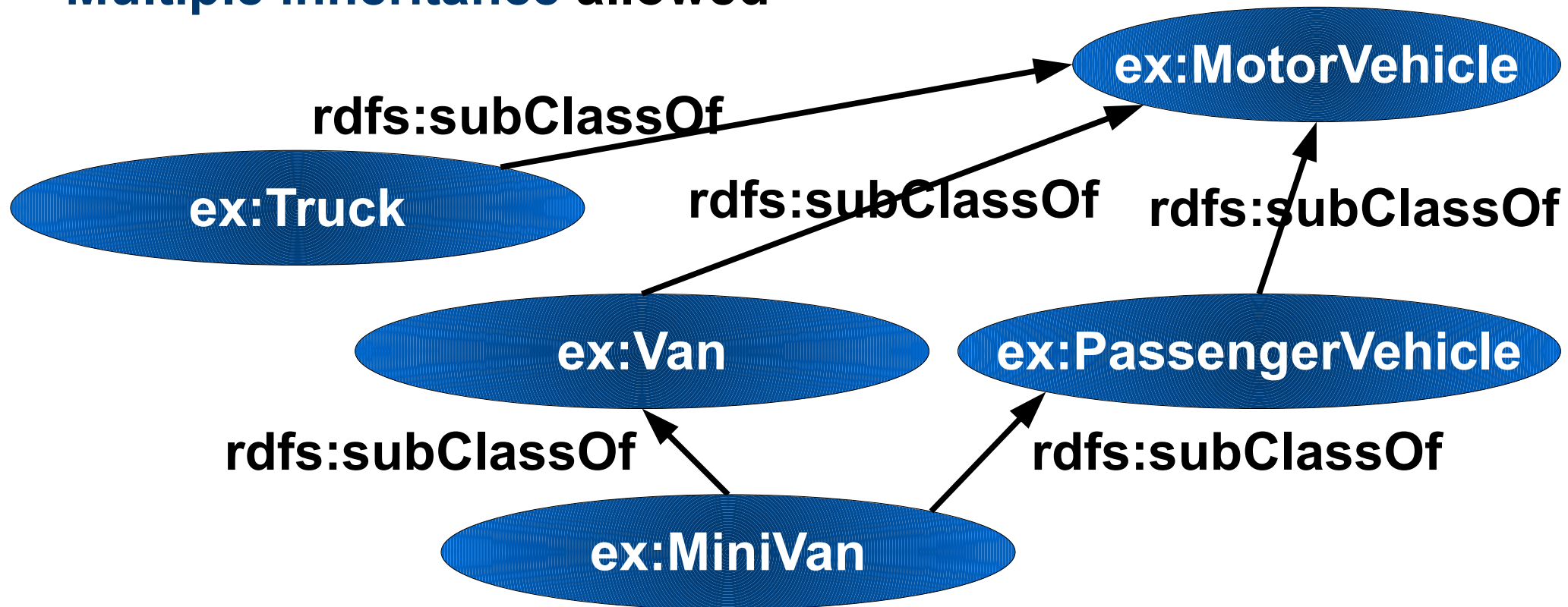
```
ex:MiniVan rdfs:subClassOf ex:Van .
```



# Class Hierarchies



- **Multiple inheritance** allowed



- *rdfs:subClassOf* is **reflexive** – e.g., it holds:

```
ex:Truck rdfs:subClassOf ex:Truck .
```

# Class Hierarchies



- *rdfs:subClassOf* is **transitive**

- E.g., given

```
ex:Van rdfs:subClassOf ex:MotorVehicle .  
ex:MiniVan rdfs:subClassOf ex:Van .
```

- we can infer

```
ex:MiniVan rdfs:subClassOf ex:MotorVehicle .
```

- **Entailment rule:**

( *A* , rdfs:subClassOf , *B* )

( *B* , rdfs:subClassOf , *C* )

---

( *A* , rdfs:subClassOf , *C* )

# Class Hierarchies



- Another entailment rule: 
$$\frac{(a, \text{rdf:type}, A) \quad (A, \text{rdfs:subClassOf}, B)}{(a, \text{rdf:type}, B)}$$

- E.g., from

```
ex:Van rdfs:subClassOf ex:MotorVehicle .
ex:MiniVan rdfs:subClassOf ex:Van .
ex:MyRedVWT3 rdf:type ex:MiniVan .
```

- we may infer

```
ex:MyRedVWT3 rdf:type ex:Van .
```

- and (exploiting transitivity)

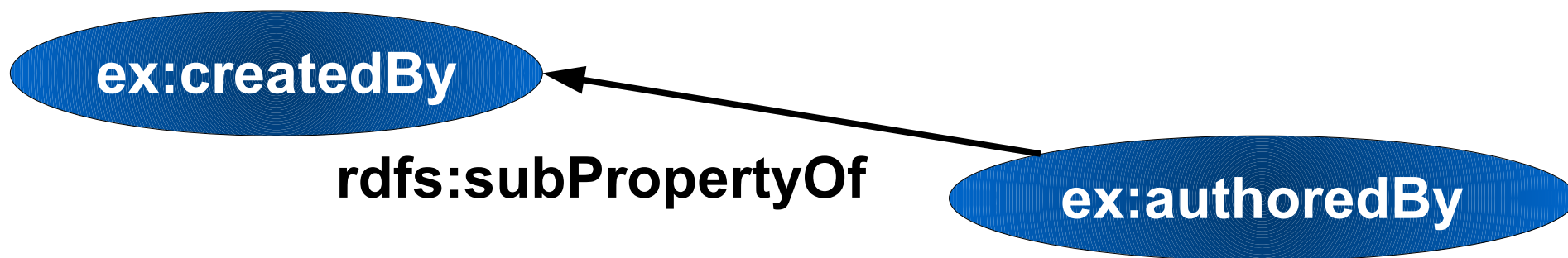
```
ex:MyRedVWT3 rdf:type ex:MotorVehicle .
```

# Property Hierarchies

- Properties usually in predicate position
- Entailment rule:

$$\frac{(a, p, b)}{(p, \text{rdf:type}, \text{rdf:Property})}$$

- Properties are first class citizens (in contrast to OOP)
- Not defined inside classes
- Own hierarchy (specified by *rdfs:subPropertyOf*)





# Property Hierarchies



- Entailment rule:  $( a , p1 , b )$   
 $( p1 , rdfs:subPropertyOf , p2 )$   
 $( a , p2 , b )$

- E.g., from

```
ex:authoredBy rdfs:subPropertyOf ex:createdBy .  
dbpedia:The_Lord_of_the_Rings ex:authoredBy  
dbpedia:J._R._R._Tolkien .
```

- we may infer

```
dbpedia:The_Lord_of_the_Rings ex:createdBy  
dbpedia:J._R._R._Tolkien .
```

- *rdfs:subPropertyOf* is **reflexive** and **transitive** too

# Property Restrictions



- *rdfs:domain* and *rdfs:range* specify permitted subjects and objects, respectively

```
@prefix rdfs : <http://www.w3.org/2000/01/rdf-schema#> .
```

```
@prefix xsd : <http://www.w3.org/2001/XMLSchema#> .
```

```
@prefix p : <http://dbpedia.org/property/> .
```

```
@prefix ex : <http://example.org/> .
```

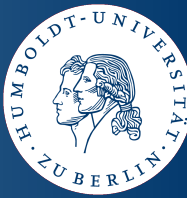
```
@prefix foaf : <http://xmlns.com/foaf/0.1/> .
```

```
ex:authoredBy rdfs:domain ex:Publication .
```

```
ex:createdBy rdfs:range foaf:Person .
```

```
p:lastEruption rdfs:range xsd:integer .
```

# Property Restrictions



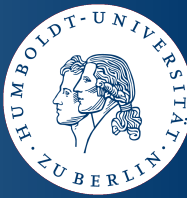
- Entailment rules:

$$\frac{(p, \text{rdfs:domain}, A) \quad (a, p, x)}{(a, \text{rdf:type}, A)}$$

$$\frac{(p, \text{rdfs:range}, A) \quad (x, p, a)}{(a, \text{rdf:type}, A)}$$

- Beware: property restrictions are global and conjunctive
  - Let  $(p, \text{rdfs:domain}, A)$  and  $(p, \text{rdfs:domain}, B)$ ;  
for each  $a$  with  $(a, p, x)$  holds  
 $(a, \text{rdfs:subClassOf}, A)$  and  $(a, \text{rdfs:subClassOf}, B)$
  - Same holds for *rdfs:range*
  - Hence, use the most general class

# Property Restrictions



- Extensional entailment rules:

$$\frac{(p, \text{rdfs:domain}, A) \quad (A, \text{rdfs:subClassOf}, B)}{(p, \text{rdfs:domain}, B)}$$

$$\frac{(p2, \text{rdfs:domain}, A) \quad (p1, \text{rdfs:subPropertyOf}, p2)}{(p1, \text{rdfs:domain}, A)}$$

$$\frac{(p, \text{rdfs:range}, A) \quad (A, \text{rdfs:subClassOf}, B)}{(p, \text{rdfs:range}, B)}$$

$$\frac{(p2, \text{rdfs:range}, A) \quad (p1, \text{rdfs:subPropertyOf}, p2)}{(p1, \text{rdfs:range}, A)}$$

# Further RDF Schema Properties



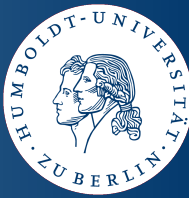
- ***rdfs:label*** – alternative name of a resources
- ***rdfs:comment*** – comment associated to a resource
- ***rdfs:seeAlso*** – reference to a resource with more information about the subject
- ***rdfs:definedBy*** – reference to a resource with a definition of the subject

```
@prefix rdfs : <http://www.w3.org/2000/01/rdf-schema#> .
```

```
@prefix foaf : <http://xmlns.com/foaf/0.1/> .
```

```
foaf:Person rdfs:label "Person"@en , "Person"@de ;  
            rdfs:comment "Class for persons."@en ;  
            rdfs:definedBy <http://xmlns.com/foaf/0.1/> .
```

# RDF Schema Summary



- **RDF Schema (RDFS) provides elementary means to define**
  - **vocabularies and**
  - **a machine-processable meaning of RDF data**
- **RDF data that uses vocabulary described with RDFS can generically be processed by every RDFS-enabled software.**
- **Web Ontology Language (OWL)**
  - **More comprehensive than RDFS**
  - **Property *owl:sameAs* – both URIs refer to the same thing**
- **However, vocabulary-specific processing requires vocabulary-specific rules.**

# Common Vocabularies



- **FOAF (Friend of a Friend)**
  - **Persons and their main properties (e.g. name, email)**
  - ***foaf:knows* relation (enables specification of a network)**
  - **Namespace URI: <http://xmlns.com/foaf/0.1/>**
- **DC (Dublin Core)**
  - **Enables description of created or published resources**
  - **Namespace URI: <http://purl.org/dc/elements/1.1/>**
- **SKOS (Simple Knowledge Organisation Systems)**
  - **Thesauri, classification schemes, taxonomies, ...**
  - **Namespace URI: <http://www.w3.org/2008/05/skos#>**

# Common Vocabularies



- **SIOC (Semantically-Interlinked Online Communities)**
  - **Content and structure of online community sites**
  - **Weblogs, mailing lists, newsgroups,**
  - **Connections between channels and posts**
  - **Namespace URI: <http://rdfs.org/sioc/ns#>**
- **DOAP (Description of a Project)**
  - **(Software) projects**
  - **Maintainer, programming language, source repository, ...**
  - **Namespace URI: <http://usefulinc.com/ns/doap#>**



# Web of Data



- Triples may link different data spaces

```
...  
<http://www.informatik.hu-berlin.de/~hartig/foaf.rdf#olaf>  
foaf:currentProject  
  <http://trdf.sourceforge.net/trdf> .  
...
```

My FOAF file

```
...  
myfoaf:me  
foaf:knows  
  <http://www.informatik.hu-berlin.de/~hartig/foaf.rdf#olaf>  
<http://www.informatik.hu-berlin.de/~hartig/foaf.rdf#olaf>  
rdfs:seeAlso  
  <http://www.informatik.hu-berlin.de/~hartig/foaf.rdf> .  
...
```

My friend's FOAF file

- Triples may link different data spaces

My FOAF file

```
...  
<http://www.informatik.hu-berlin.de/~hartig/foaf.rdf#olaf>  
  foaf:currentProject  
    <http://trdf.sourceforge.net/trdf> .  
...
```

```
<http://trdf.sourceforge.net/trdf>  
  rdfs:seeAlso  
    <http://trdf.sourceforge.net/doap.rdf>  
...
```

```
...  
<http://trdf.sourceforge.net/trdf> doap:name "tRDF" ;  
                                     doap:created "2008-03-03" .  
...
```

DOAP file

- **Linked data principles**<sup>1)</sup> (set of best practices for publishing and deploying data on the Web using RDF):
  1. Use URIs as names for things.
  2. Use HTTP URIs so that people can look up those names.
  3. When someone looks up a URI, provide useful RDF data.
  4. Include RDF statements that link to other URIs, so that they can discover related things.
- These principles allow a true **Web of data**
- **RDF links have a machine-processable semantic (in contrast to links between Web documents)**

<sup>1)</sup><http://www.w3.org/DesignIssues/LinkedData.html>

```
...  
<http://www.informatik.hu-berlin.de/~hartig/foaf.rdf#olaf>  
foaf:currentProject  
  <http://trdf.sourceforge.net/trdf> .  
...
```

My FOAF file

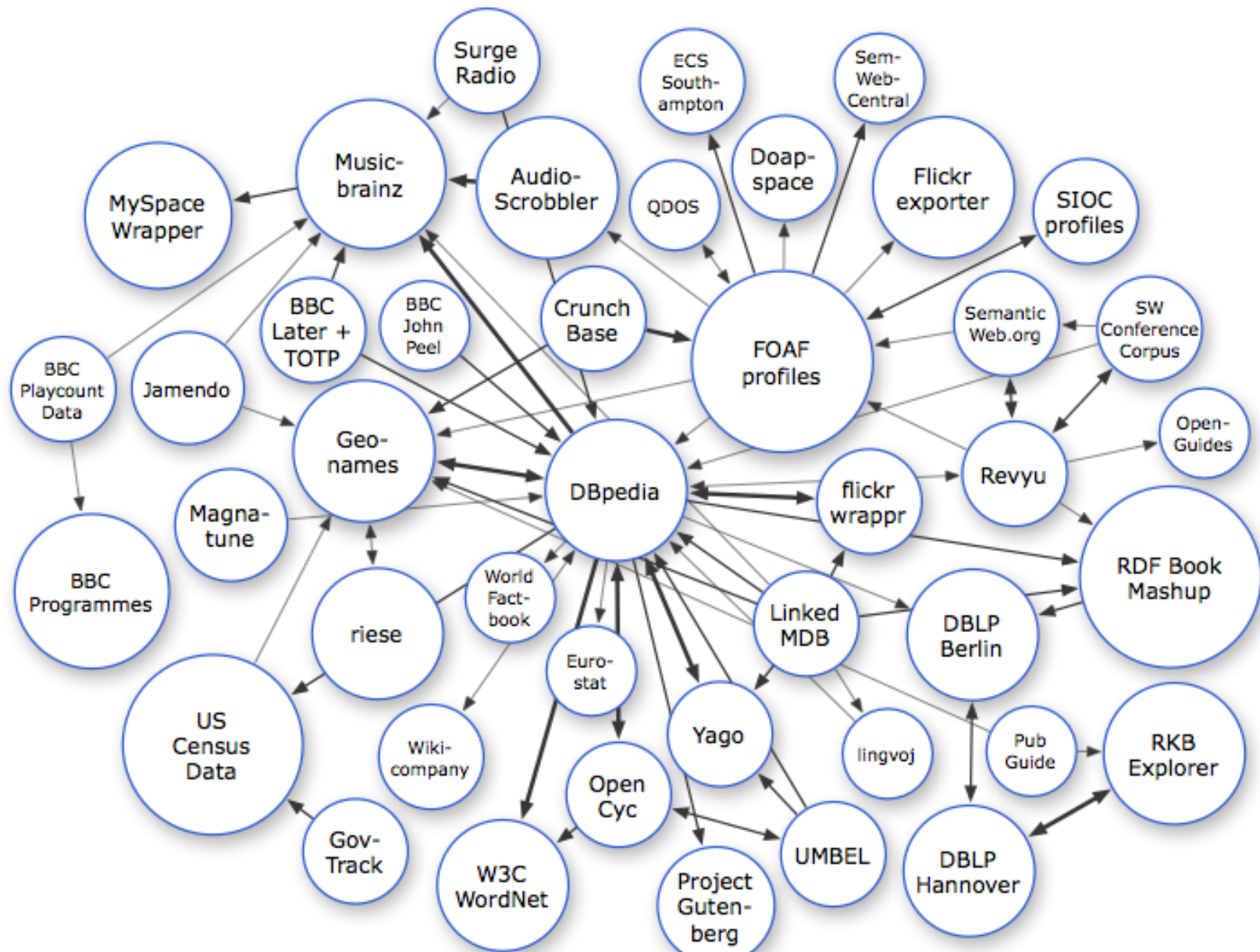
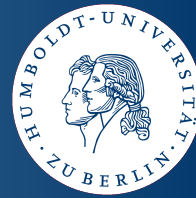
- HTTP Request: GET <http://trdf.sourceforge.net/trdf>
- Server response:  
303 See Other – <http://trdf.sourceforge.net/doap.rdf>
- 2nd Request: GET <http://trdf.sourceforge.net/doap.rdf>

```
...  
<http://trdf.sourceforge.net/trdf> doap:name "tRDF" ;  
                                   doap:created "2008-03-03" .  
...
```

Redirection

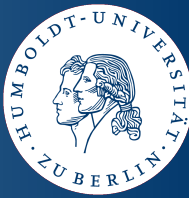
Project

# Web of Data



By courtesy of the Linking Open Data community project As of September 2008  
License: CC-BY-SA  
<http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

# Web of Data



- **Selected open RDF datasets:**

<b>Dataset</b>	<b>Description</b>	<b>Triples</b>
<b>Dbpedia</b>	<b>Structured information extracted from Wikipedia</b>	<b>ca. 117M</b>
<b>U.S.Census</b>	<b>2000 U.S. Census data</b>	<b>ca. 1B</b>
<b>GovTrack</b>	<b>U.S. Congress data</b>	<b>ca. 13M</b>
<b>riese</b>	<b>EuroStat data</b>	<b>ca. 5M (3B)</b>
<b>DBLP</b>	<b>Bibliographic information on major computer science journals and conference proceedings</b>	<b>ca. 28M</b>
<b>MusicBrainz</b>	<b>Data about artists, records, songs etc.</b>	<b>ca. 36M</b>
<b>BBC</b>	<b>Data about BBC Programmes</b>	<b>ca. 2M</b>
<b>CrunchBase</b>	<b>Directory of technology companies, people, and investors</b>	<b>ca. 700K</b>

# Open World Assumption

- The absence of a triple is **not** relevant
- The knowledge:

```
@prefix dbpedia : <http://dbpedia.org/resource/> .
```

```
@prefix p : <http://dbpedia.org/property/> .
```

```
dbpedia:Washington p:borderingstates dbpedia:Oregon .
```

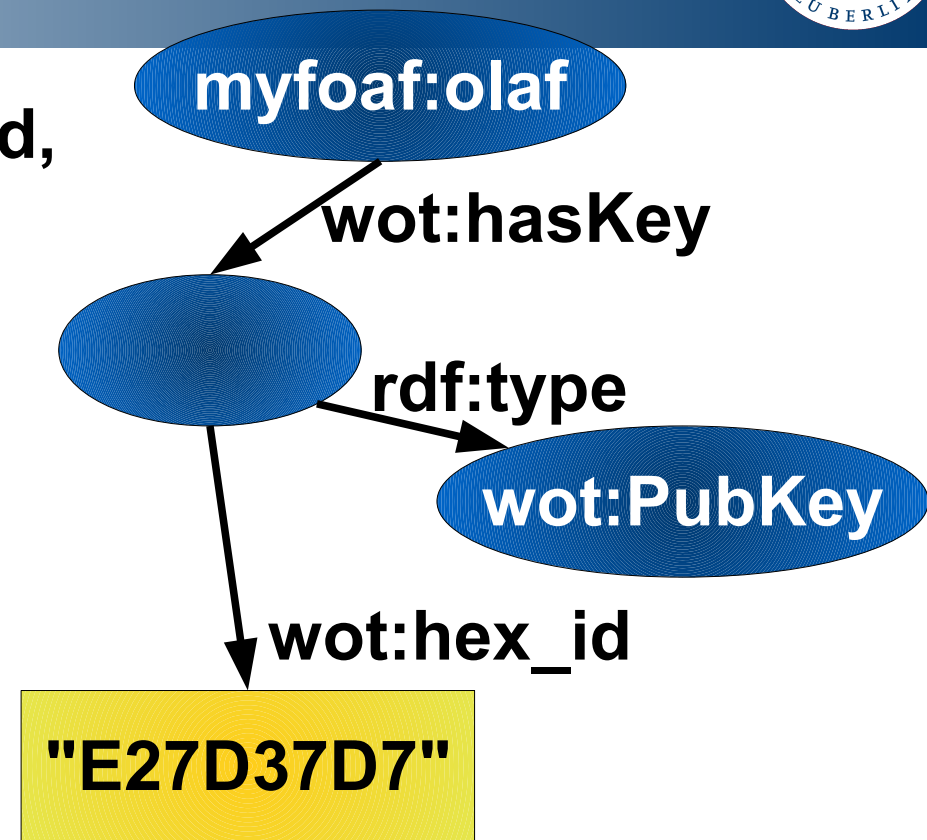
**does not mean Washington has one bordering state!**

- If **you** have no other triples doesn't mean they are not true.

# Blank Nodes



- Blank nodes represent unnamed, **anonymous resources**
- Not identified by a URI
- Blank node identifiers
  - Identification of blank nodes in triple serializations
  - Form: `_:xyz`
  - Significant only within a single RDF graph



```
myfoaf:olaf wot:hasKey _:x .
_:x rdf:type wot:PubKey ;
wot:hex_id "E27D37D7" .
```

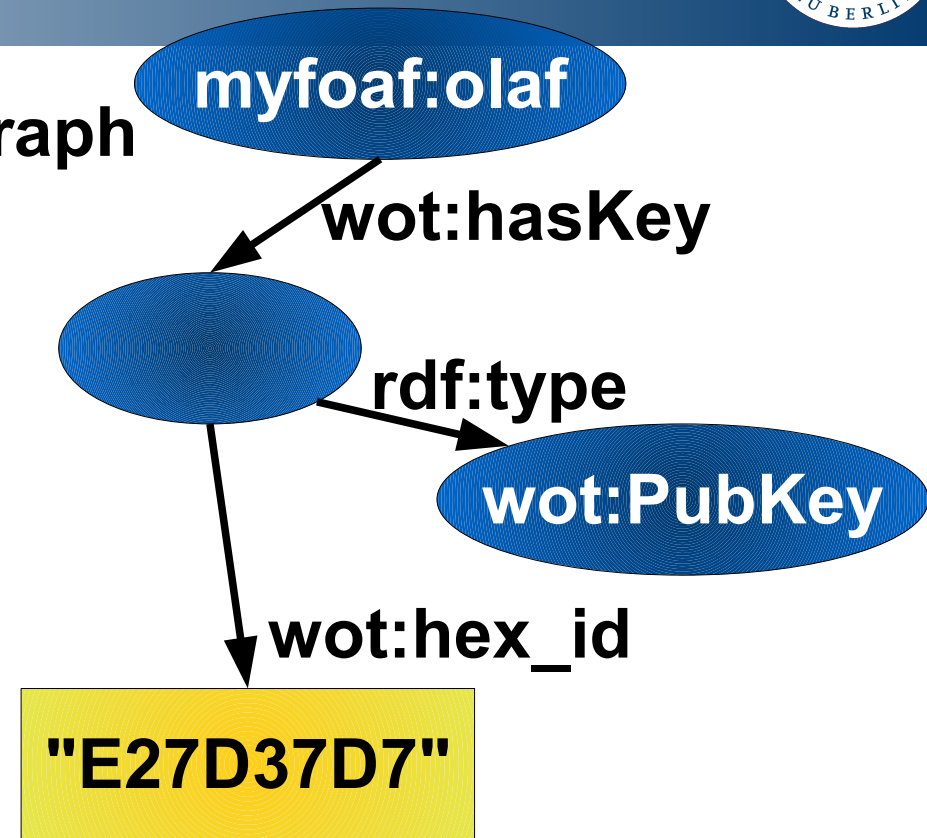


# Blank Nodes



- Blank nodes break the global graph

- Cannot be referenced
- **Not reusable**
- Name your resources (linked data principle)



- Abbreviated syntax in N3:

```
myfoaf:olaf wot:hasKey [ rdf:type    wot:PubKey ;  
                        wot:hex_id  "E27D37D7" ]
```

# Groups of Things



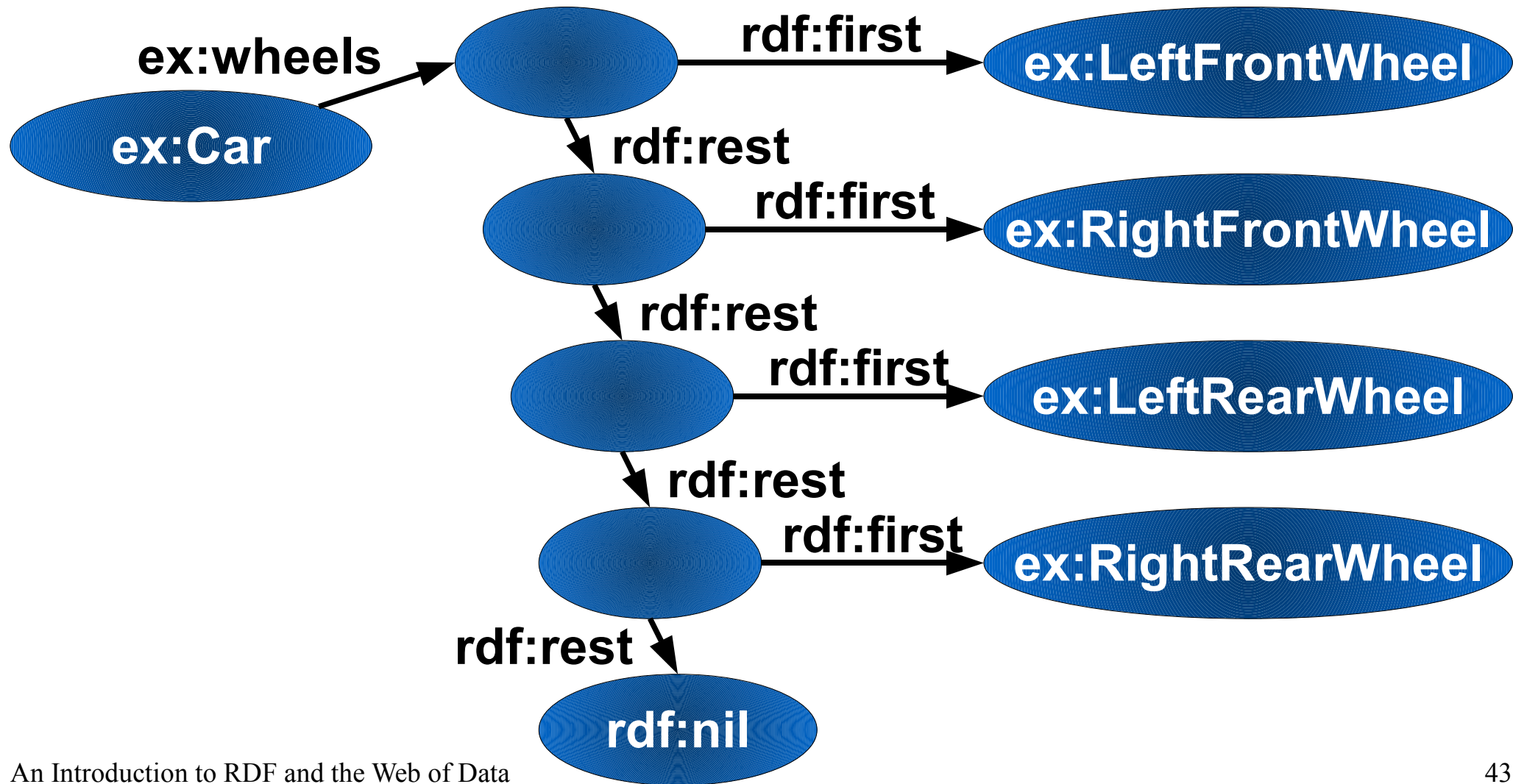
- Containers are an open group
  - Contain resources or literals, possibly duplicates
  - *rdf:Seq* – ordered list
  - *rdf:Bag* – set (unordered)
  - *rdf:Alt* – for alternatives

```
dbpedia:Mount_Etna ex:eruptions [ rdf:type rdf:Bag ;  
                                   rdf:_1 "1669" ;  
                                   rdf:_2 "1949" ;  
                                   rdf:_3 "1971" ;  
                                   rdf:_4 "2001" ] .
```

# Groups of Things



- Collections
  - Closed list of resources or literals, possibly duplicates



# Groups of Things



- **Collections in N3**

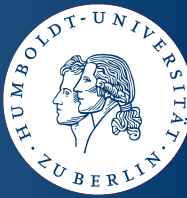
```
ex:Car ex:wheels _:a ;
_:a rdf:first ex:LeftFrontWheel ; rdf:rest _:b .
_:b rdf:first ex:RightFrontWheel ; rdf:rest _:c .
_:c rdf:first ex:LeftRearWheel ; rdf:rest _:d .
_:d rdf:first ex:RightRearWheel ; rdf:rest rdf:nil .
```

- **Shortcut**

```
ex:Car ex:wheels ( ex:LeftFrontWheel ex:RightFrontWheel
                  ex:LeftRearWheel ex:RightRearWheel ) .
```

- **Generic access with SPARQL impossible**

# Reification



- Reification allows statements about statements

represented by `_:s`

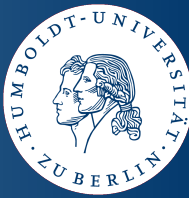
```
ex:LinkedDataPrinciples rat:rating "great" .
```

```
_:s rdf:type rdf:Statement ;  
    rdf:subject ex:LinkedDataPrinciples ;  
    rdf:predicate rat:rating ;  
    rdf:object "great" .
```

```
_:s dc:creator myfoaf:olaf .
```

about

# RDF/XML – An XML syntax for RDF



```
@prefix dbpedia : <http://dbpedia.org/resource/> .
```

N3

```
@prefix p : <http://dbpedia.org/property/> .
```

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
```

```
dbpedia:Mount_Baker p:lastEruption "1880"^^xsd:integer .
```

```
dbpedia:Mount_Baker p:location dbpedia:Washington .
```

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:p="http://dbpedia.org/property/">
  <rdf:Description rdf:about="http://dbpedia.org/resource/Mount_Baker">
    <p:lastEruption ▶
      rdf:datatype="http://www.w3.org/2001/XMLSchema#integer" ▶
      >1880</p:lastEruption>
    <p:location rdf:resource="http://dbpedia.org/resource/Washington"/>
  </rdf:Description>
</rdf:RDF>
```

RDF/XML

# Further Reading

- **W3C RDF Specifications – <http://www.w3.org/RDF/>**
  - **RDF Primer**
  - **RDF: Concepts and Abstract Syntax**
  - **RDF Vocabulary Description Language 1.0: RDF Schema**
  - **RDF Semantics**
  - **RDF/XML Syntax Specification (Revised)**
  - **RDF Test Cases**
- **Information about Linked Data – <http://linkeddata.org/>**