

RDF and Linked Data

Olaf Hartig

Linköping University

TDDD43

Advanced Data Models and Databases

Sep. 25, 2017

Outline

(1) RDF

(2) SPARQL

(3) Linked Data

RDF in General

- **Resource Description Framework**
- A **resource** may basically be everything
 - e.g. persons, places, Web documents, abstract concepts
- **Descriptions of resources**
 - Attributes
 - Relationships
- **The framework contains:**
 - A data model, and
 - Languages and syntaxes



RDF Data Model

- Data comes as a set of **triples** (subject, predicate, object)
- **Subject**: resources
- **Predicate**: properties
- **Object**: literals or resources
- **Examples**:
 - (Mount Baker , last eruption , 1880)
 - (Mount Baker , location , Washington)

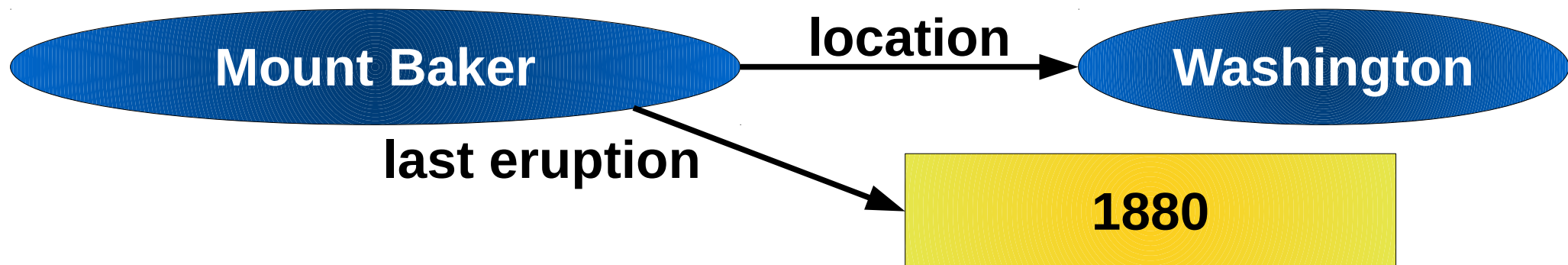
RDF Data Model (cont'd)

- **RDF based data may be understood as a **graph**:**

- Triples as directed edges
- Subjects and objects as vertices
- Edges labeled by predicate

- **Example:**

- (Mount Baker , last eruption , 1880)
- (Mount Baker , location , Washington)

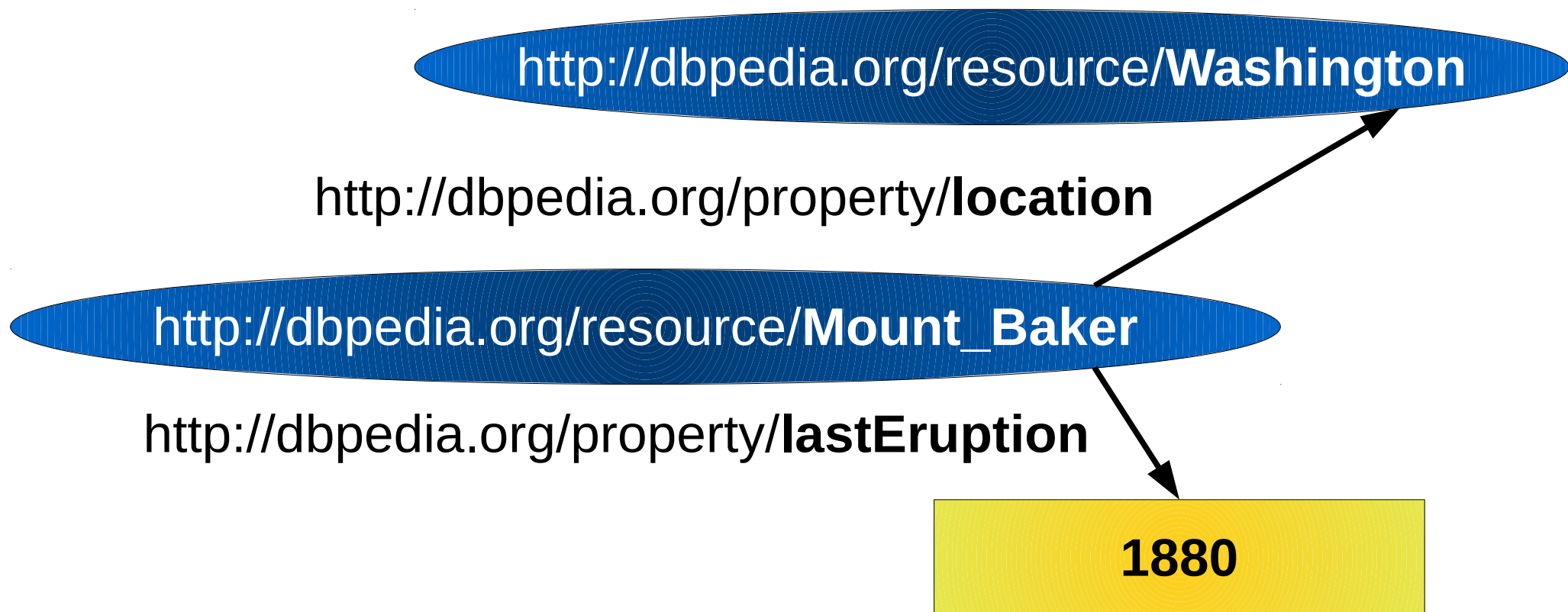


Uniform Resource Identifier (URI)

- **URIs extend the concept of URLs**
 - Globally **unique identifier** for resources
 - URL of a Web document usually used as its URI
 - Attention: URIs identify not only Web documents
- **Example:**
 - Me:
<http://olafhartig.de/foaf.rdf#olaf>
 - RDF document about me:
<http://olafhartig.de/foaf.rdf>
 - HTML document about me:
<http://olafhartig.de/index.html>

Example (revisited)

- (http://dbpedia.org/resource/Mount_Baker,
<http://dbpedia.org/property/lastEruption>, 1880)
- (http://dbpedia.org/resource/Mount_Baker,
<http://dbpedia.org/property/location>,
<http://dbpedia.org/resource/Washington>)



Compact URI (CURIE)

- **Abbreviated Notation for URIs**
- **Syntax:**
 - Prefix name (references a 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://olafhartig.de/foaf.rdf#olaf`

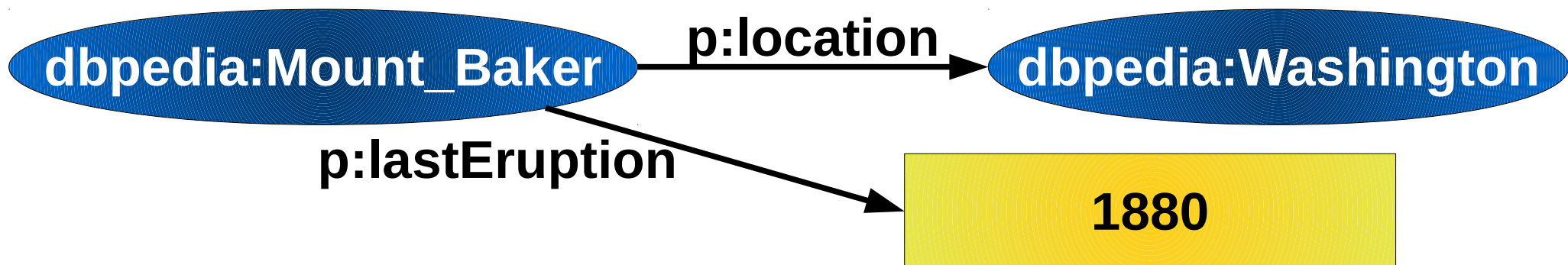
Example with CURIEs

- **Using**

- *dbpedia* for prefix *http://dbpedia.org/resource/*
- *p* for prefix *http://dbpedia.org/property/*

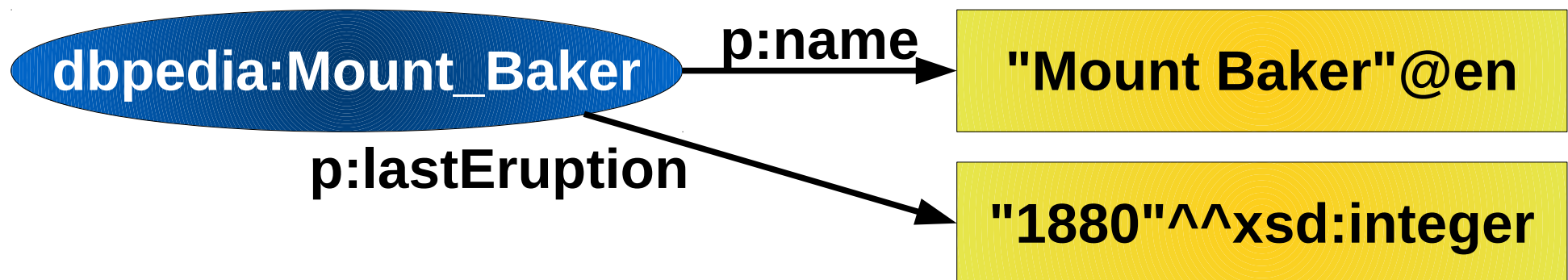
- **we have**

- (dbpedia:Mount_Baker, p:lastEruption, 1880)
- (dbpedia:Mount_Baker, p:location, dbpedia:Washington)



Literals

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



Turtle: A Readable Syntax for RDF

- **Simple, human-readable 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> .
```

Turtle supports 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 .
```

Turtle 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 .
```

Turtle, some 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 .
```

RDF/XML – An XML Syntax for RDF

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

Turtle

```
@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

Outline

(1) RDF ✓

(2) SPARQL

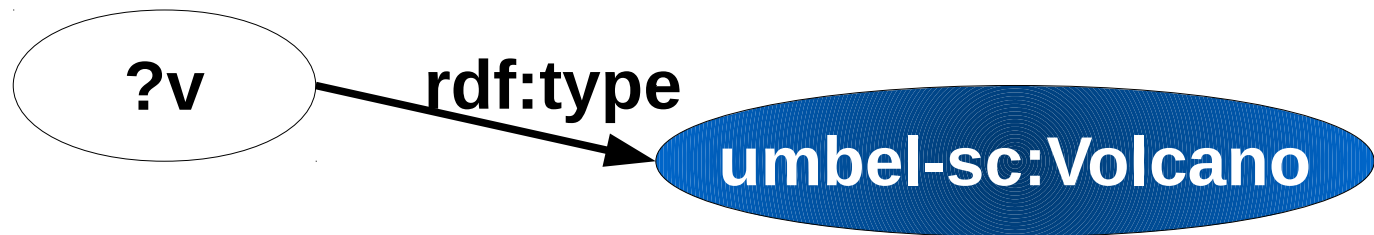
(3) Linked Data

SPARQL in General

- **A family of W3C recommendations**
- **SPARQL Query**
 - Declarative query language for RDF data
 - Our focus today
- **SPARQL Update**
 - Declarative update language for RDF data
- **SPARQL Protocol**
 - Communication between SPARQL processing services (a.k.a. SPARQL endpoints) and clients
- **SPARQL Query Results XML Format**
 - XML format for serializing query results

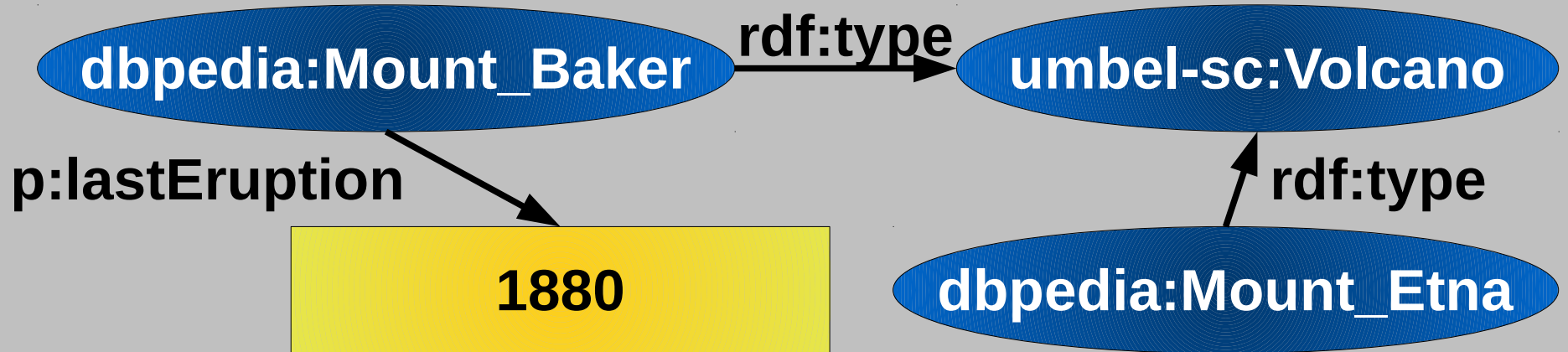
Main Idea of SPARQL Queries

- **Pattern matching:**
 - Describe subgraphs of the queried RDF graph
 - Subgraphs that match the description yield a result
 - Mean: **graph patterns** (essentially, RDF graphs with variables)

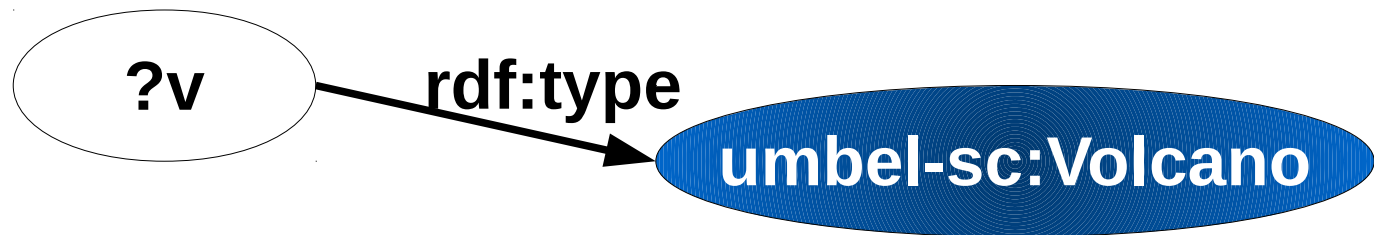


Main Idea of SPARQL Queries

Queried RDF graph:



Result:

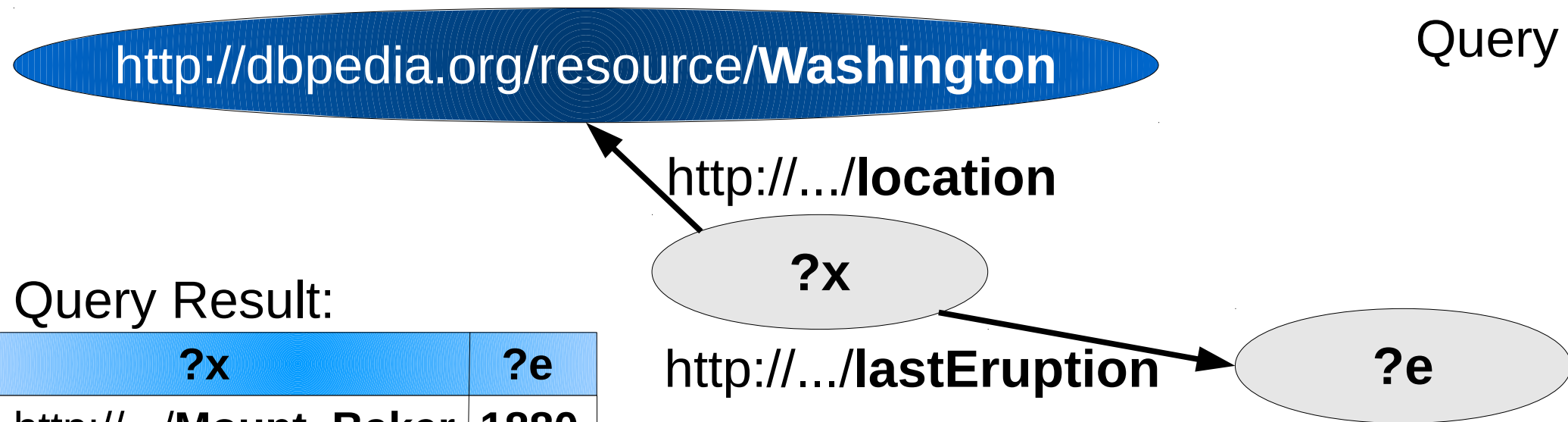


?v
dbpedia:Mount_Baker dbpedia:Mount_Etna

Another Example



Data



Query

Query Result:

<code>?x</code>	<code>?e</code>
<code>http://.../Mount_Baker</code>	<code>1880</code>

Components of a SPARQL Query

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
SELECT ?v
FROM <http://example.org/myGeoData>
WHERE {
    ?v rdf:type umbel-sc:Volcano .
}
ORDER BY ?name
```

Components of a SPARQL Query

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
SELECT ?v
FROM <http://example.org/myGeoData>
WHERE {
    ?v rdf:type umbel-sc:Volcano .
}
ORDER BY ?name
```

- **Prologue:**
 - [Prefix definitions](#) for using compact URIs (CURIEs)

Components of a SPARQL Query

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
SELECT ?v
FROM <http://example.org/myGeoData>
WHERE {
    ?v rdf:type umbel-sc:Volcano .
}
ORDER BY ?name
```

- **Result form specification:**
 - SELECT for projection
(similar to projection in relational algebra)
 - Other forms: DESCRIBE, CONSTRUCT, and ASK

Components of a SPARQL Query

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
SELECT ?v
FROM <http://example.org/myGeoData>
WHERE {
    ?v rdf:type umbel-sc:Volcano .
}
ORDER BY ?name
```

- **Dataset specification:**
 - Specify the RDF dataset to be queried (use URIs that identify particular RDF graphs in your RDF database)

Components of a SPARQL Query

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
SELECT ?v
FROM <http://example.org/myGeoData>
WHERE {
    ?v rdf:type umbel-sc:Volcano .
}
ORDER BY ?name
```

- **Query Pattern:**
 - WHERE clause specifies the graph pattern to be matched

Basic Graph Pattern

- Set of triple patterns (i.e., RDF triples with variables)
- Variable names prefixed with “?” (or “\$”)
- Turtle syntax
 - Including syntactic sugar (e.g., property and object lists)

```
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/>
SELECT ?name
WHERE {
    ?v rdf:type umbel-sc:Volcano ;
        rdfs:label ?name .
}
```

Basic Graph Pattern (Example)

```
dbpedia:Mount_Etna rdf:type umbel-sc:Volcano ;           Data*
                      rdfs:label "Etna" .
dbpedia:Mount_Baker rdf:type umbel-sc:Volcano.
dbpedia:Beerenberg  rdf:type umbel-sc:Volcano,
                      umbel-sc:NaturalElevation ;
                      rdfs:label "Beerenberg"@en ;
                      rdfs:label "Беренберг"@ru .
```

- **Question:** What are the names of all (known) volcanos?

```
SELECT ?name WHERE {                                     Query*
    ?v rdf:type umbel-sc:Volcano ;
    rdfs:label ?name . }
```

*Prefix declarations omitted

Result:

?name
"Etna"
"Беренберг"@ru
"Beerenberg"@en

Optional Graph Pattern

```
dbpedia:Mount_Etna rdf:type umbel-sc:Volcano ;  
                    rdfs:label "Etna" .  
dbpedia:Mount_Baker rdf:type umbel-sc:Volcano .  
dbpedia:Beerenberg rdf:type umbel-sc:Volcano ;  
                    rdfs:label "Beerenberg"@en .
```

Data

- **Question:** What are *all* (known) volcanos and their names?

```
SELECT ?v ?name WHERE {  
  ?v rdf:type umbel-sc:Volcano ;  
  rdfs:label ?name . }
```

Query

- **Problem:** Mount Baker *missing* (b/c no name in the data)

?v	?name
dbpedia:Mount_Etna	"Etna"
dbpedia:Beerenberg	"Beerenberg"@en

Optional Graph Pattern

- Keyword **OPTIONAL** indicates optional patterns

```
SELECT ?v ?name WHERE {  
    ?v rdf:type umbel-sc:Volcano .  
    OPTIONAL { ?v rdfs:label ?name }  
}
```

Query

?v	?name
dbpedia:Mount_Etna	"Etna"
dbpedia:Mount_Baker	
dbpedia:Beerenberg	"Beerenberg"@en

- Optional patterns may result in unbound variables

Constraints on Solutions

- **Syntax: Keyword FILTER followed by filter expression**

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
PREFIX p: <http://dbpedia.org/property/>

SELECT ?v
WHERE {
    ?v rdf:type umbel-sc:Volcano ;
        p:lastEruption ?le .
    FILTER ( ?le > 1900 )
}
```

- **Filter expressions contain operators and functions**
- **Operators and functions operate on RDF terms**

Unary Operators in Constraints

Operator	Type(A)	Result type
! A	xsd:boolean	xsd:boolean
+ A	numeric	numeric
- A	numeric	numeric
BOUND(A)	variable	xsd:boolean
isURI(A)	RDF term	xsd:boolean
isBLANK(A)	RDF term	xsd:boolean
isLITERAL(A)	RDF term	xsd:boolean
STR(A)	literal / URI	simple literal
LANG(A)	literal	simple literal
DATATYPE(A)	literal	simple literal

Binary and other Operators

- **Logical connectives `&&` and `||`**
 - for `xsd:boolean`
- **Comparison operators `=`, `!=`, `<`, `>`, `<=`, and `>=`**
 - for numeric datatypes, `xsd:dateTime`, `xsd:string`, `xsd:boolean`
- **Comparison operators `=` and `!=`**
 - for other datatypes
- **Arithmetic operators `+`, `-`, `*`, and `/`**
 - for numeric datatypes
- **Furthermore:**
 - `REGEX(String,Pattern)` or `REGEX(String,Pattern,Flags)`
 - `langMATCHES(A,B)`
 - etc.

Constraints (Example)

```
dbpedia:Mount_Etna rdf:type umbel-sc:Volcano ;  
                    rdfs:label "Etna" .  
dbpedia:Beerenberg rdf:type umbel-sc:Volcano,  
                          umbel-sc:NaturalElevation ;  
                    rdfs:label "Beerenberg"@en ;  
                    rdfs:label "Бере́нберг"@ru .
```

Data

- **Question:** What volcanos have an “e” in their name?

```
SELECT ?v WHERE {  
  ?v rdf:type umbel-sc:Volcano ;  
    rdfs:label ?name .  
  FILTER( REGEX(STR(?name), "e") )  
}
```

Query

?v
dbpedia:Beerenberg
dbpedia:Beerenberg

Constraints (Example, cont'd)

```
dbpedia:Mount_Etna rdf:type umbel-sc:Volcano ;  
                    rdfs:label "Etna" .  
dbpedia:Beerenberg rdf:type umbel-sc:Volcano,  
                           umbel-sc:NaturalElevation ;  
                    rdfs:label "Beerenberg"@en ;  
                    rdfs:label "Беренберг"@ru .
```

Data

- **Question:** What volcanos have an “e” in their name?

```
SELECT ?v WHERE {  
  ?v rdf:type umbel-sc:Volcano ;  
    rdfs:label ?name .  
  FILTER( REGEX(STR(?name), "e", "i") )  
}
```

Query

?v
dbpedia:Mount_Etna
dbpedia:Beerenberg
dbpedia:Beerenberg

Components of a SPARQL Query

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
SELECT ?v
FROM <http://example.org/myGeoData>
WHERE {
    ?v rdf:type umbel-sc:Volcano .
}
ORDER BY ?name
```

- **Solution modifiers:**

- Only for SELECT queries
- Modify the **result set** as a whole (not single solutions)
- Keywords: DISTINCT, ORDER BY, LIMIT, and OFFSET

SPARQL 1.1

- **New features of SPARQL 1.1 Query:**
 - Aggregate functions (e.g., COUNT, SUM, AVG)
 - Sub-queries
 - Negation (EXISTS, NOT EXISTS, MINUS)
 - Assignments (e.g., BIND, SELECT expressions)
 - Property paths (navigation à la XPath)
 - Basic query federation (SERVICE, BINDINGS)
- **SPARQL 1.1 Update:**
 - Graph update (INSERT DATA, DELETE DATA, INSERT, DELETE, DELETE WHERE, LOAD, CLEAR)
 - Graph management (CREATE, DROP, COPY, MOVE, ADD)

Outline

(1) RDF ✓

(2) SPARQL ✓

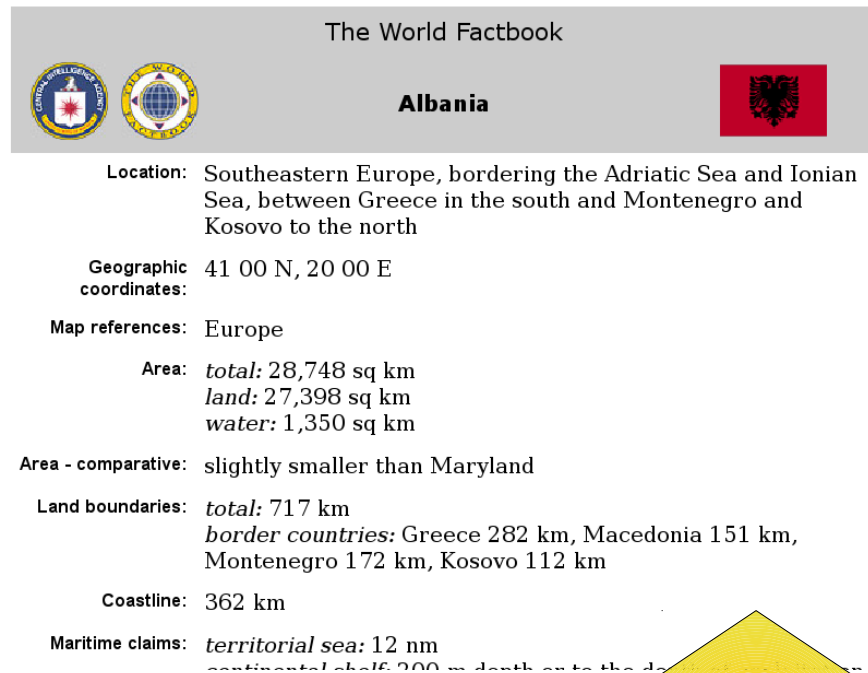
(3) Linked Data

Data on the Web



The screenshot shows the IMDb page for the TV movie "War Child" (1999). The page includes a navigation bar with links like "NOW PLAYING", "MOVIE/TV NEWS", "MY MOVIES", "DVD & BLU-RAY", "IMDb TV", "MESSAGE BOARDS", "SHOWTIMES & TICKETS", and "IMDbPro". The main content area displays the movie title, overview, user rating (awaiting 5 votes), director (Michael Davie), release date (12 July 1999), genre (Documentary), tagline ("Young people affected by the war in Kosovo"), and plot keywords ("Spoiler alert! R"). Additional details include runtime (USA: 25 min), country (USA), and language (English). A yellow arrow points from the "MovieDB" database icon to the IMDb page.

[Country List](#) | [World Factbook Home](#)



The screenshot shows the CIA World Factbook page for Albania. The page includes the title "The World Factbook", the country name "Albania", and the Albanian flag. The main content area displays various facts about Albania, including location, geographic coordinates, map references, area, land boundaries, coastline, and maritime claims. A yellow arrow points from the "CIA World Factbook" database icon to the page.

Location: Southeastern Europe, bordering the Adriatic Sea and Ionian Sea, between Greece in the south and Montenegro and Kosovo to the north

Geographic coordinates: 41 00 N, 20 00 E

Map references: Europe

Area: total: 28,748 sq km
land: 27,398 sq km
water: 1,350 sq km

Area - comparative: slightly smaller than Maryland

Land boundaries: total: 717 km
border countries: Greece 282 km, Macedonia 151 km, Montenegro 172 km, Kosovo 112 km

Coastline: 362 km

Maritime claims: territorial sea: 12 nm
continental shelf: 200 m depth or to the depth of exploitation

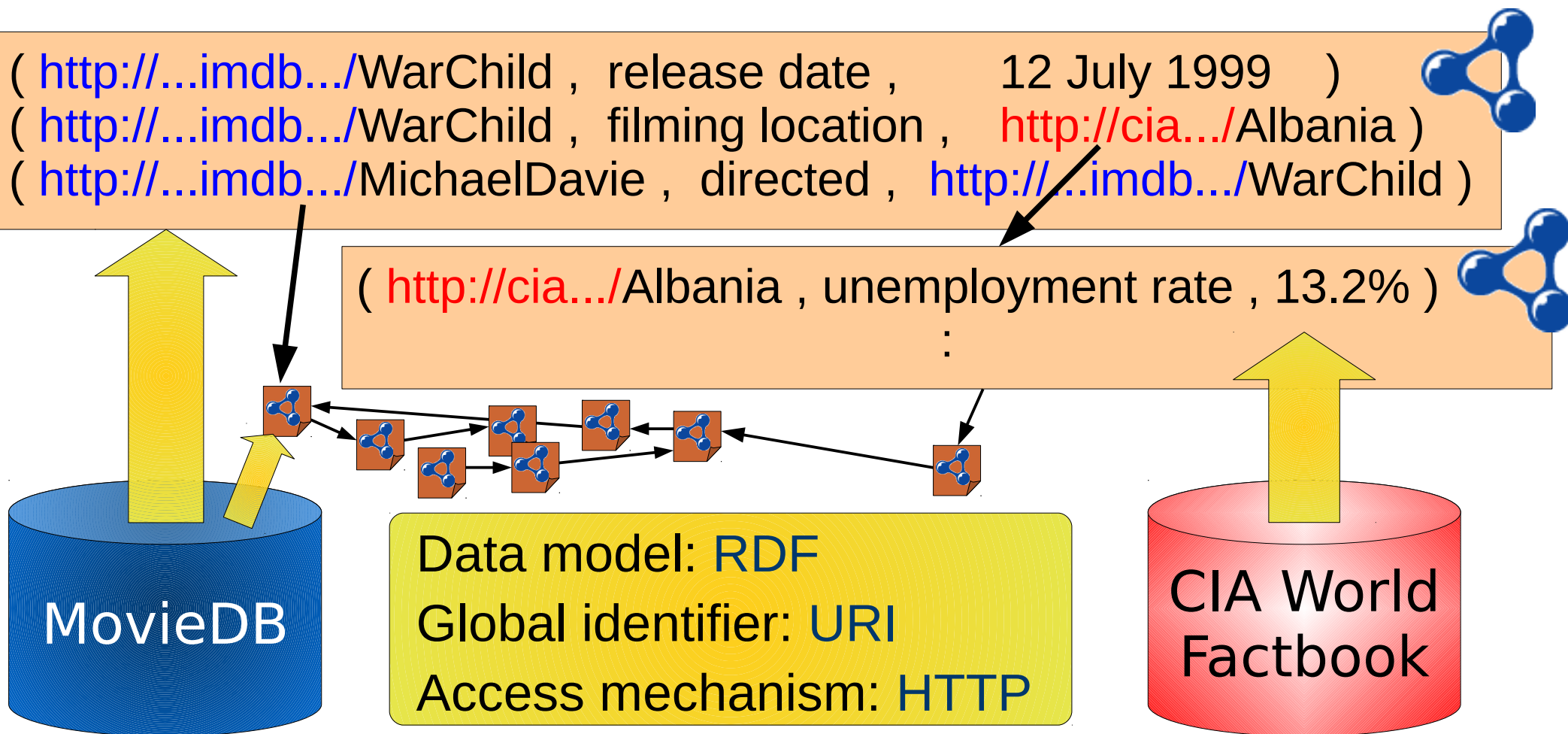
MovieDB

Traditionally, published in HTML documents that are designed for human consumption

CIA World Factbook

Linked Data Publishing Principles

- **Goal: publishing structured data on the WWW in a standardized, machine-readable manner**



Adoption

- **Started as a grassroots community effort in 2007**
- Publish existing, open license datasets as Linked Data
- Interlink things between different data sources
- **Prominent publishers joined the effort**
 - e.g., BBC, NY Times, Library of Congress, Thomson Reuters, Springer, Nature, Best Buy, Sears, Renault, UK Government
- **Numbers**
 - BTC 2014 Crawl (February – June 2014):
ca. 4.1B triples in ca. 44M docs from ca. 48K sites
 - lodstats (as of September 24, 2017):
ca. 149B triples in ca. 3.0K datasets



The Web of Linked Data

...a globally distributed network of data

...which we may understand as a huge distributed database

How do we enable applications to **query this data**?

- Active area of research!
- Looking for a thesis topic in this area?
Contact me!

