

# Introduction to Knowledge Graphs and Semantic Web Technologies

## RDF and Linked Data

Olaf Hartig

[olaf.hartig@liu.se](mailto:olaf.hartig@liu.se)

# Overview

- Basic concepts of RDF
- Formats for serializing RDF data
- Additional features of RDF
- What is Linked Data?



Image source: <https://www.maxpixel.net/To-Find-Search-Watch-Overview-See-Binoculars-1015267>

# RDF: Basic Concepts

# RDF in General

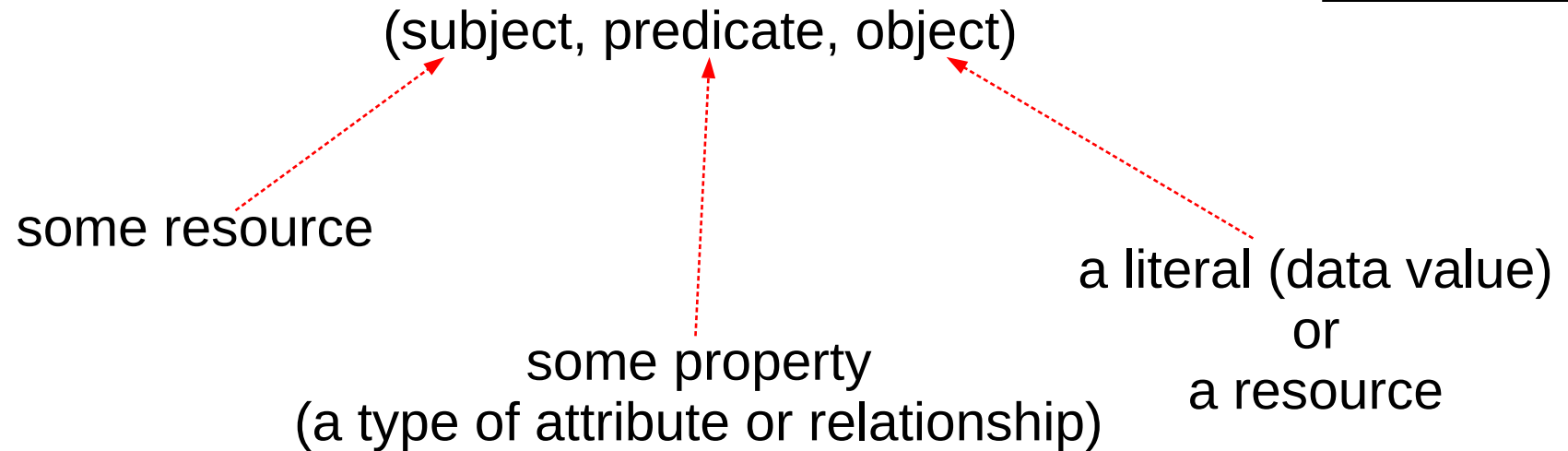


- **R**esource **D**escription **F**ramework
  - family of standards by the W3C
- A **resource** may be everything
  - e.g., persons, places, Web documents, abstract concepts
- **D**escriptions of resources consist of
  - attributes
  - relationships
- The **framework** contains
  - a data model
  - languages and syntaxes

# RDF Data Model



- Data represented in the form of triples



- Example:
  - ( Mount Baker, year of last eruption, 1880 )
  - ( Mount Baker, located in, Washington State )

# RDF Data Model (cont'd)

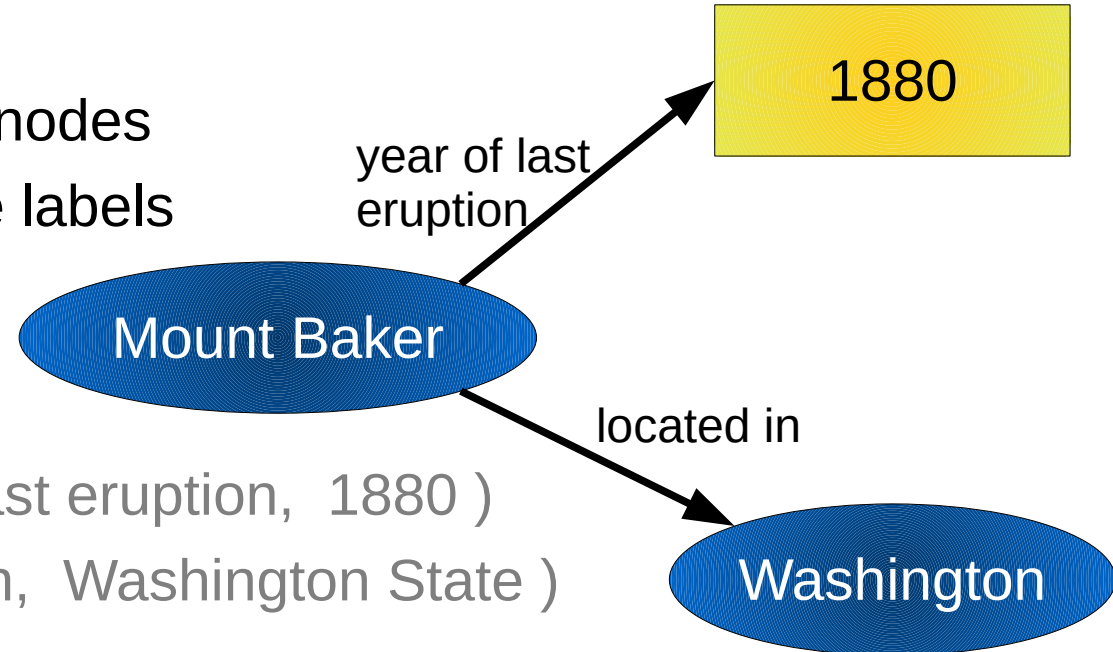


- Data represented in the form of triples  
(subject, predicate, object)

- Set of such triples is called *RDF graph*

- ... and can be visualized as a graph

- triples as directed edges
- subjects and objects as nodes
- predicates used as edge labels

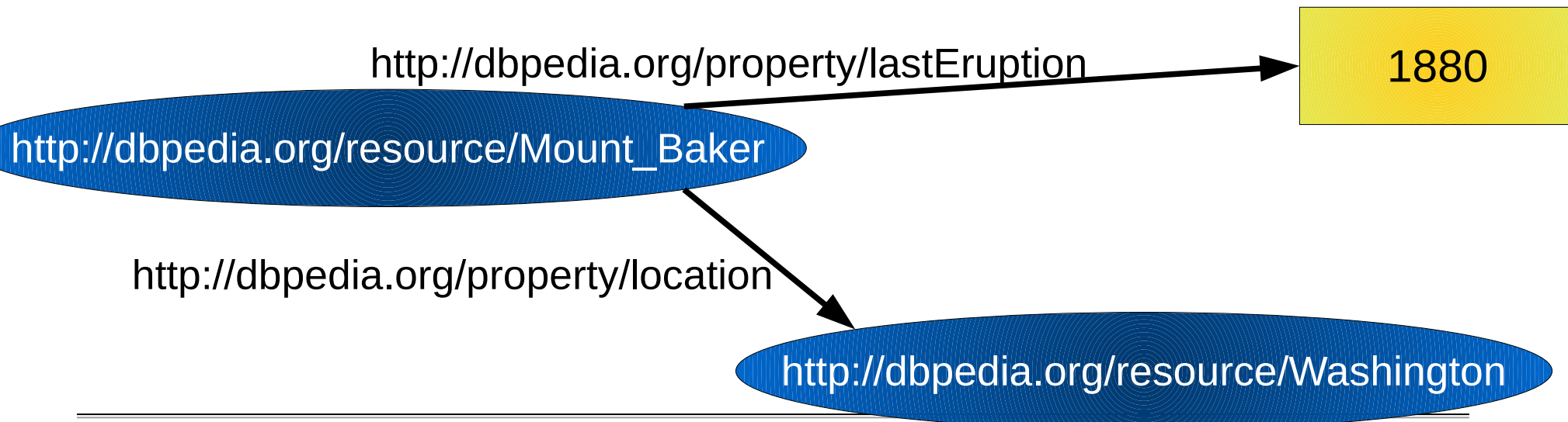


- Example:


- ( Mount Baker, year of last eruption, 1880 )
- ( Mount Baker, located in, Washington State )

# Uniform Resource Identifier (URI)

- Globally unique identifier for resources and for properties
- Typically look like URLs
- Examples:
  - [http://dbpedia.org/resource/Mount\\_Baker](http://dbpedia.org/resource/Mount_Baker)
  - <http://dbpedia.org/property/lastEruption>
  - <http://olafhartig.de/foaf.rdf#olaf>



# Compact URI (CURIE)

- Abbreviated notation for URIs `dbpedia:Mount_Baker`  


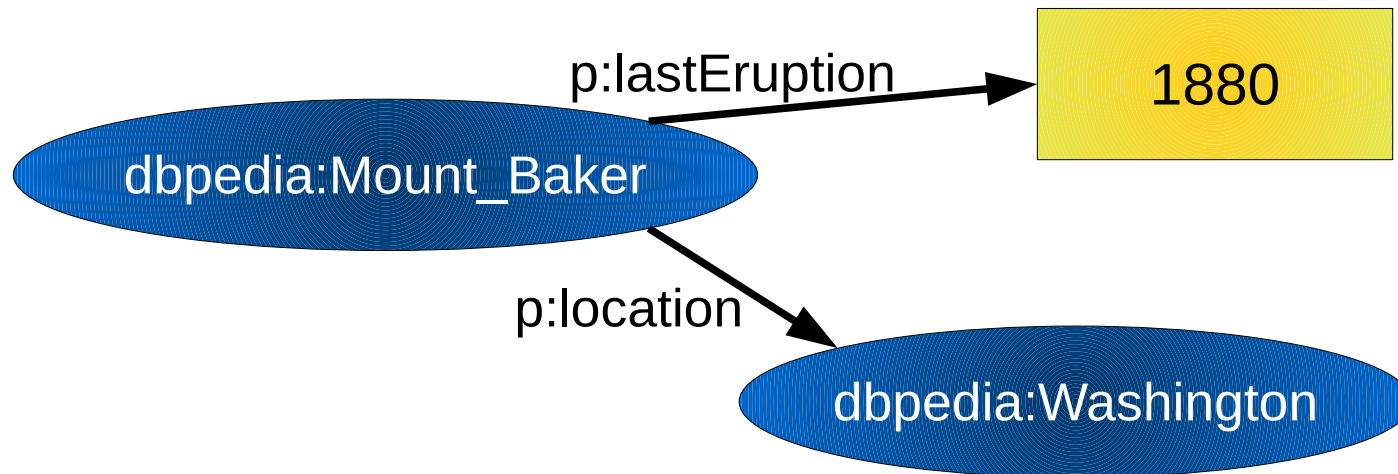
prefix name                      reference part

(references a prefix for building URIs;  
e.g., the prefix <http://dbpedia.org/resource/>)
- Create the full URI by adding the reference part to the prefix referenced by the prefix name
- Further examples:
  - <http://olafhartig.de/foaf.rdf#olaf> may be abbreviated as `od:olaf` by using `od` for <http://olafhartig.de/foaf.rdf#>
  - <http://dbpedia.org/property/lastEruption> may be abbreviated as `p:lastEruption` by using `p` for <http://dbpedia.org/property/>



# Example with Compact URIs

- By using `dbpedia` for `http://dbpedia.org/resource/` and using `p` for `http://dbpedia.org/property/`, our example becomes:




- Represented as triples:  
( `dbpedia:Mount_Baker`, `p:lastEruption`, `1880` )  
( `dbpedia:Mount_Baker`, `p:location`, `dbpedia:Washington` )

# Serialization Formats for RDF Data

# Overview of Serialization Formats for RDF

- RDF/XML

- Turtle 

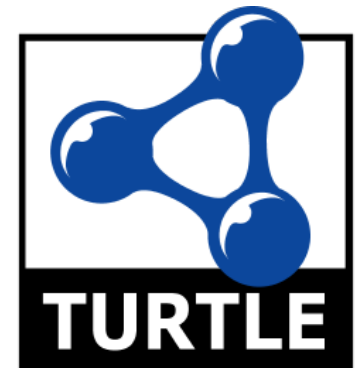
- N-Triples

- RDFa

- JSON-LD

- etc.

Image source: <https://pixabay.com/illustrations/magnifying-glass-looking-for-find-1020141/>



# RDF Turtle

Simple, human-readable notation to list RDF triples:

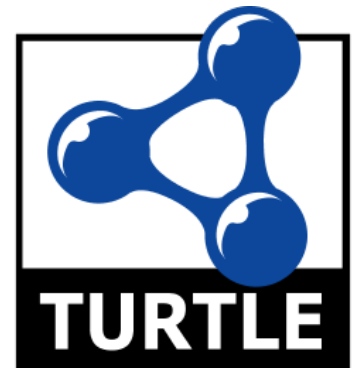
- Triples separated by a period (“.”) character
- @prefix directive to declare prefixes for compact URIs

```
@prefix dbpedia : <http://dbpedia.org/resource/> .
@prefix p : <http://dbpedia.org/property/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

dbpedia:Mount_Baker p:lastEruption 1880 .
dbpedia:Mount_Baker p:location dbpedia:Washington .

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

<http://olafhartig.de/foaf.rdf#olaf> foaf:name "Olaf Hartig" .
```



# RDF Turtle (cont'd)

Shorthand notations (i.e., syntactic sugar):

- Property lists separated by a semicolon (“;”) character
- Object lists separated by a semicolon (“,”) character

```
@prefix dbpedia : <http://dbpedia.org/resource/> .
@prefix p : <http://dbpedia.org/property/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

dbpedia:Mount_Baker p:lastEruption 1880 ;
                    p:location      dbpedia:Washington .

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

<http://olafhartig.de/foaf.rdf#olaf> foaf:name "Olaf Hartig" .
```

# Overview of Serialization Formats for RDF

- RDF/XML
  - an XML representation of RDF data
- Turtle
  - focus on human readability and authoring
- N-Triples
  - for bulk processing, fast import and export; one triple per line
- RDFa
  - RDF inside HTML documents
- JSON-LD
  - JSON format for RDF
- etc.



Image sources: <https://cygri.github.io/rdf-logos/> and <https://json-ld.org/images/>

# (Some) Additional Features of RDF

# Literals

- Data values, can be used in the object position of RDF triples
- Consist of a string representation of the value and a data type
  - data type specifies how to interpret the string representation
  - data type denoted by URI; e.g., <http://www.w3.org/2001/XMLSchema#integer>
  - if no data type given, then it is <http://www.w3.org/2001/XMLSchema#string>
- Literals of type string may additionally have a language tag



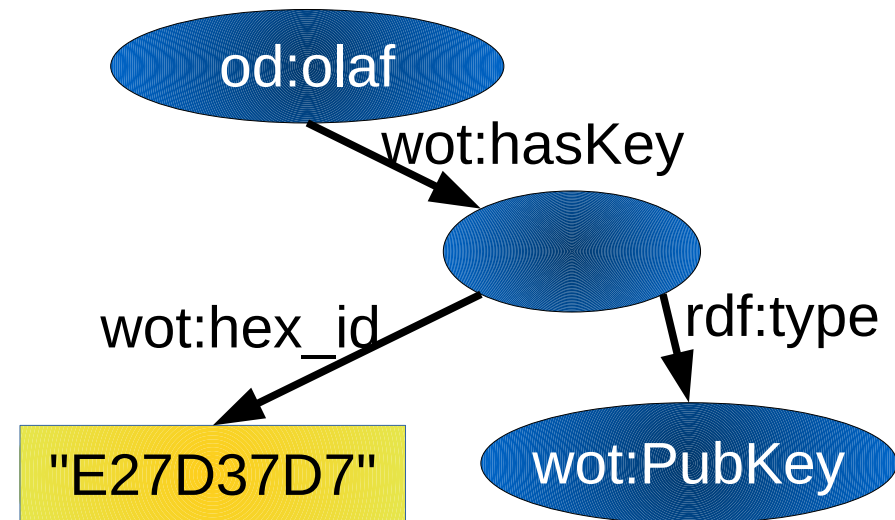
```
@prefix dbpedia : <http://dbpedia.org/resource/> .  
@prefix p : <http://dbpedia.org/property/> .  
@prefix foaf : <http://xmlns.com/foaf/0.1/> .  
@prefix xsd : <http://www.w3.org/2001/XMLSchema#> .
```

```
dbpedia:Mount_Baker p:lastEruption "1880"^^xsd:integer ;  
                    foaf:name "Mount Baker"@en ,  
                               "Mont Baker"@fr .
```



# Blank Nodes

- Unnamed / anonymous resources
  - i.e., no URI
- To distinguish different blank nodes when writing RDF data to a file, the serialization formats use a form of labels for blank nodes
  - scope of blank nodes labels only within each file
  - e.g., in Turtle, blank nodes are represented by `_:label`
    - ... or encapsulated in square brackets:



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



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



# RDF Vocabularies

- Introduce URIs for types/classes of resources
  - use `rdf:type` to associate resources to these classes
- Introduce URIs for properties (i.e., the predicates in triples)
- Define the meaning of the RDF data that uses these URIs
  - more about this when we talk about ontologies

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

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

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

```
<http://olafhartig.de/foaf.rdf#olaf> rdf:type foaf:Person ;  
foaf:name "Olaf Hartig" ;  
wot:hasKey _:x .
```

```
_:x rdf:type wot:PubKey ;
```

```
wot:hex_id "E27D37D7" .
```



# RDF Vocabularies

- Introduce URIs for types/classes of resources
  - use `rdf:type` to associate resources to these classes
- Introduce URIs for properties (i.e., the predicates in triples)
- Define the meaning of the RDF data that uses these URIs
  - more about this when we talk about ontologies

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

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

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

```
<http://olafhartig.de/foaf.rdf#olaf> a foaf:Person ;  
foaf:name "Olaf Hartig" ;  
wot:hasKey _:x .
```

```
_:x a wot:PubKey ;
```

```
wot:hex_id "E27D37D7" .
```



# Named Graphs: Containers for RDF Triples

- Named Graph is an RDF graph that is associated with a URI
  - this URI is typically used to refer to the corresponding RDF graph
- RDF dataset: one default graph + zero or more named graphs
- Use cases:
  - maintaining data from different sources separately
  - graph-level metadata (e.g., provenance information)
  - versioning
  - access control
  - etc
- Serialization formats:
  - JSON-LD: @graph keyword
  - N-Quads: like N-Triples with a fourth element
  - TriG: extends Turtle for RDF datasets



Image source: <http://pngimg.com/download/7778>

# TriG Example

```
dbpedia:Mount_Etna rdfs:seeAlso <http://example.org/d1> .
dbpedia:Mount_Baker rdfs:seeAlso <http://example.org/d2> .

GRAPH <http://example.org/d1>
{  dbpedia:Mount_Etna rdf:type umbel-sc:Volcano ;
   rdfs:label "Etna" . }

GRAPH <http://example.org/d2>
{  dbpedia:Mount_Baker rdf:type umbel-sc:Volcano . }

GRAPH <http://example.org/d3>
{  dbpedia:Beerenberg rdf:type umbel-sc:Volcano ;
   rdfs:label "Beerenberg"@en . }
```

# Linked Data

# Data on the Web

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

Go to CIA.gov

THE WORLD FACTBOOK Countries Maps References About

services: 40.3% (2017 est.)

**Unemployment rate**

5.83% (2019 est.)  
6.32% (2018 est.)

**note:** these official rates may not include those working at near-subsistence farming

country comparison to the world: 96

**Unemployment, youth ages 15-24**

**total:** 27%  
**male:** 27.8%  
**female:** 25.9% (2019 est.)

country comparison to the world: 42

**Population below poverty line**

14.3% (2012 est.)

Civil Liberties

Government

Introduction

Geography

People and Society

Environment

Government

Economy

Energy

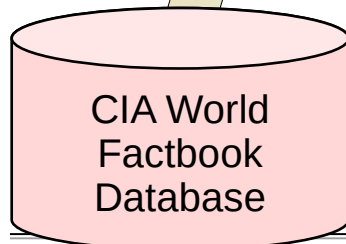
Communications

Transportation

Military and Security

Terrorism

Transnational Issues



IMDb All Search IMDb Sign In EN

War Child

TV Movie · 1999 · 25m

YOUR RATING ☆ Rate

Cast & crew · IMDbPro All topics

Documentary War

+ Add to Watchlist

Director Michael Davie

**Storyline** Edit

kosovo albania

Add full plot · Add synopsis

Taglines Young people affected by the war in Kosovo

Genres Documentary · War

**Details** Edit

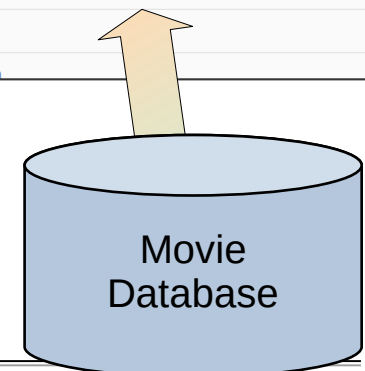
Release date July 12, 1999 (United States)

Country of origin United States

Language English

Filming locations Albania

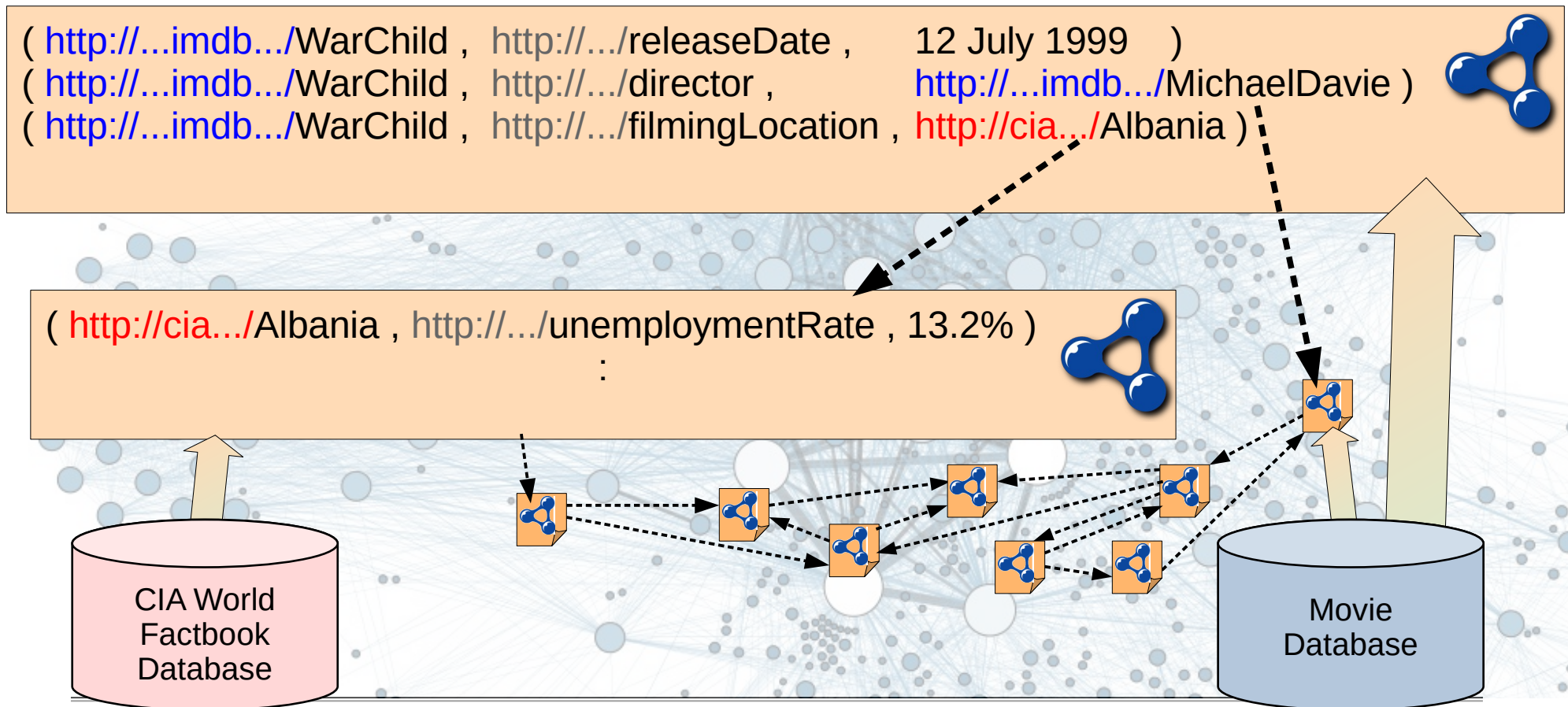
Production company National Geographic Television



# Linked Data Publishing Principles

Goal: publishing data on the WWW in a structured form that is standardized and machine-readable

- data model: RDF
- global identifiers: URIs
- access protocol: HTTP





# Adoption for Open Data

- Started as a grassroots community effort in 2007
  - publish existing, open-license datasets as Linked Data
  - including data links between different datasets
- 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 (Feb–Jun 2014): ca. 4.1B triples in ca. 44M docs from ca. 48K sites
  - lodstats (as of Sep 24, 2017): ca. 149B triples in ca. 3K datasets

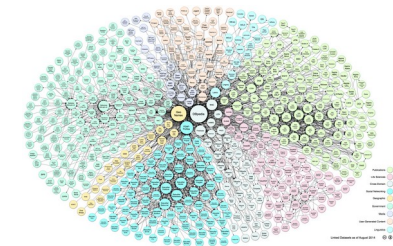
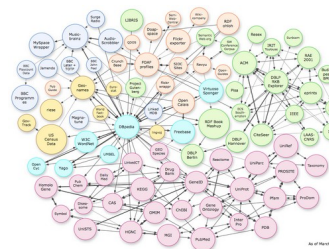
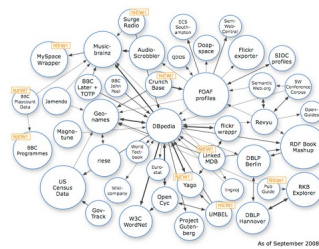
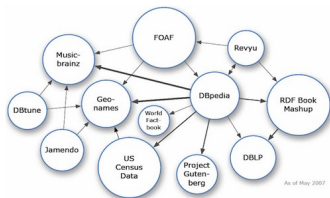
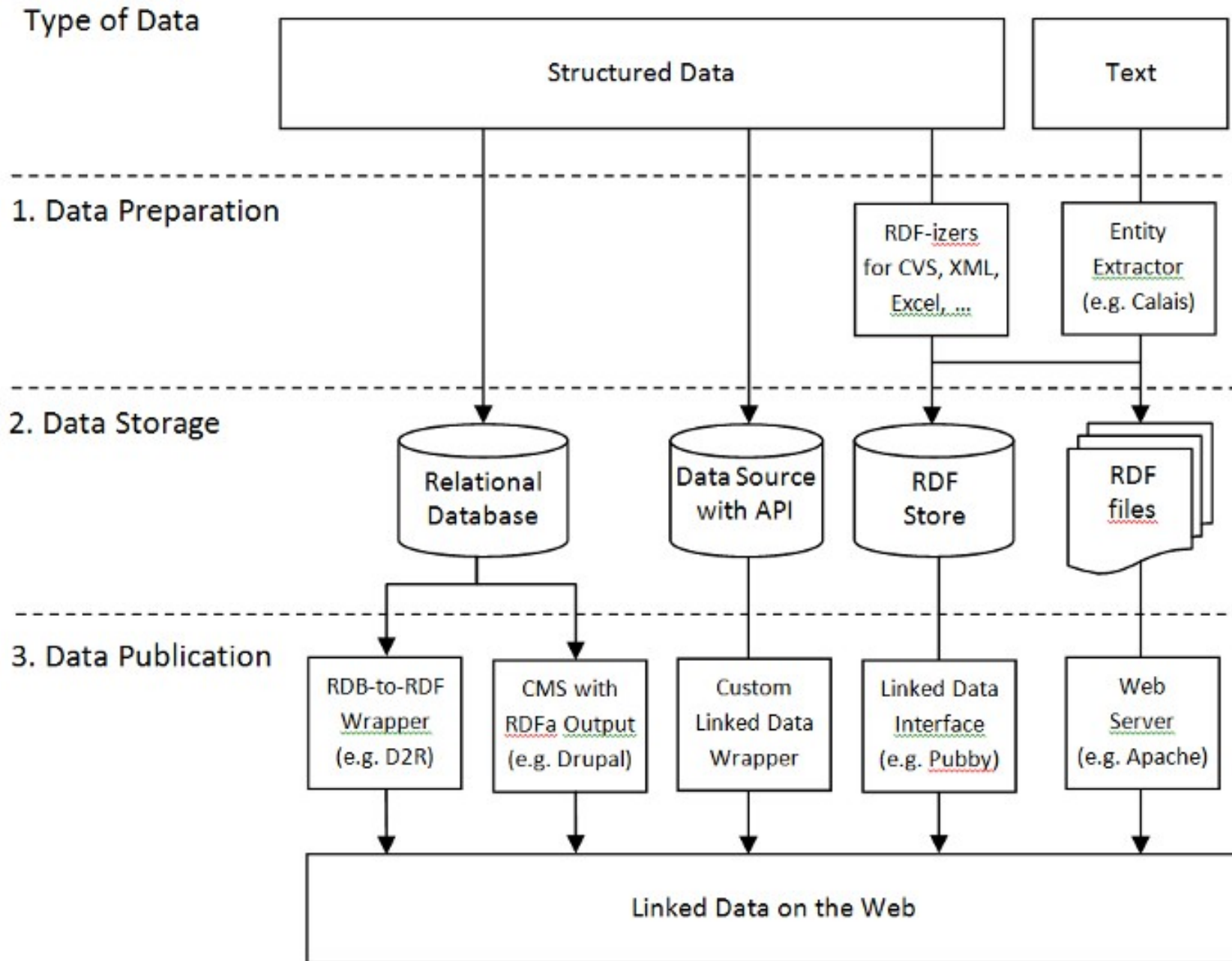


Image source: <https://lod-cloud.net/>

# Providing a Linked Data Interface



[www.liu.se](http://www.liu.se)