ICWE 2012 Tutorial

An Introduction to SPARQL and Queries over Linked Data

•••

Chapter 1: Linked Data and RDF

Olaf Hartig
http://olafhartig.de/foaf.rdf#olaf
@olafhartig

Database and Information Systems Research Group
Humboldt-Universität zu Berlin

Outline

Chapter 1:
Linked Data and RDF

Chapter 2:
The SPARQL Query Language

Chapter 3:
Querying Linked Data on the Web

http://olafhartig.de/icwe.html

Chapter 1

- **Motivation**
- **Technical Foundations**
- **The Web of Linked Data**

http://olafhartig.de/icwe.html

The Traditional Web

Traditional Web = Internet + Docs + Links

The Traditional Web

So what is the problem?

The Traditional Web

So what is the problem?

- Web content is only loosely structured
- Difficult for applications to do smart things

The Traditional Web

So what is the problem?

- Web content is only loosely structured
- Difficult for applications to do smart things

Solution:

- Increase the structure of Web content
- Publish data

The Traditional Web

So what is the problem?

- Web content is only loosely structured
- Difficult for applications to do smart things

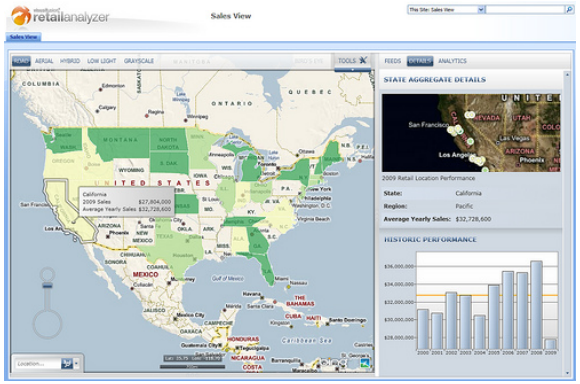
Solution:

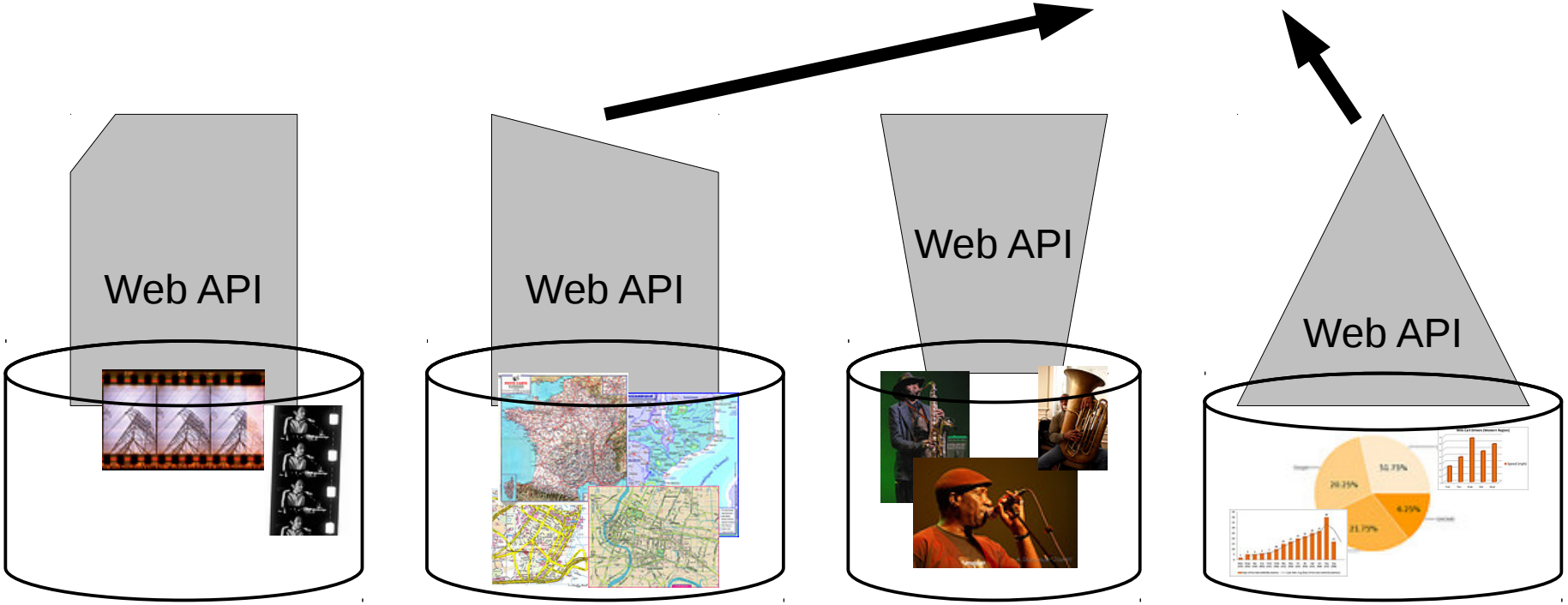
- Increase the structure of Web content
- Publish data

But wait…
don't we do that already?

The Traditional Web

- **Content providers offer access via Web APIs**
- **Mashups combine this data**



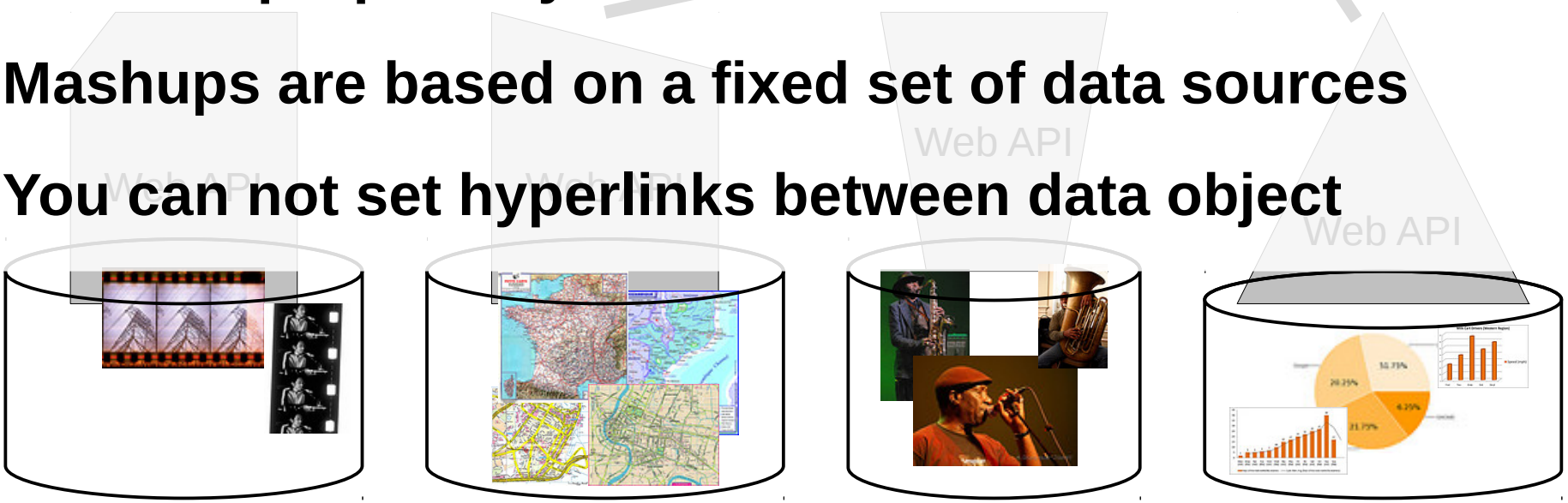


The Traditional Web

- **Content providers offer access via Web APIs**
- **Mashups combine this data**

Shortcomings:

- **APIs are proprietary**
- **Mashups are based on a fixed set of data sources**
- **You can not set hyperlinks between data object**



Chapter 1

- Motivation
- **Technical Foundations**
- The Web of Linked Data

http://olafhartig.de/icwe.html

Technical Foundations

There is no magic – Linked Data is based on well-established (Semantic) Web technologies.

- **RDF as shared data model**
- **HTTP to access data on the Web**
- **URIs**
  - Globally unique identifiers for entities
  - Pointers to data
- **Hyperlinks**

RDF in General

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

RDF Data Model

- **Data comes as a set of triples (subject, predicate, object)**

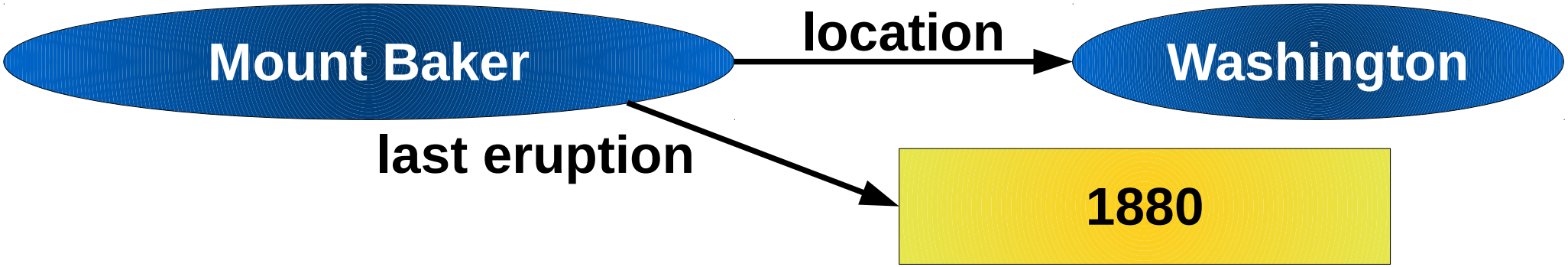
- **Subject: resources**
- **Predicate: properties**
- **Object: literals or resources**

- **Examples:**
  - (Mount Baker , last eruption , 1880)
  - (Mount Baker , location , Washington)

RDF Data Model

- **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)

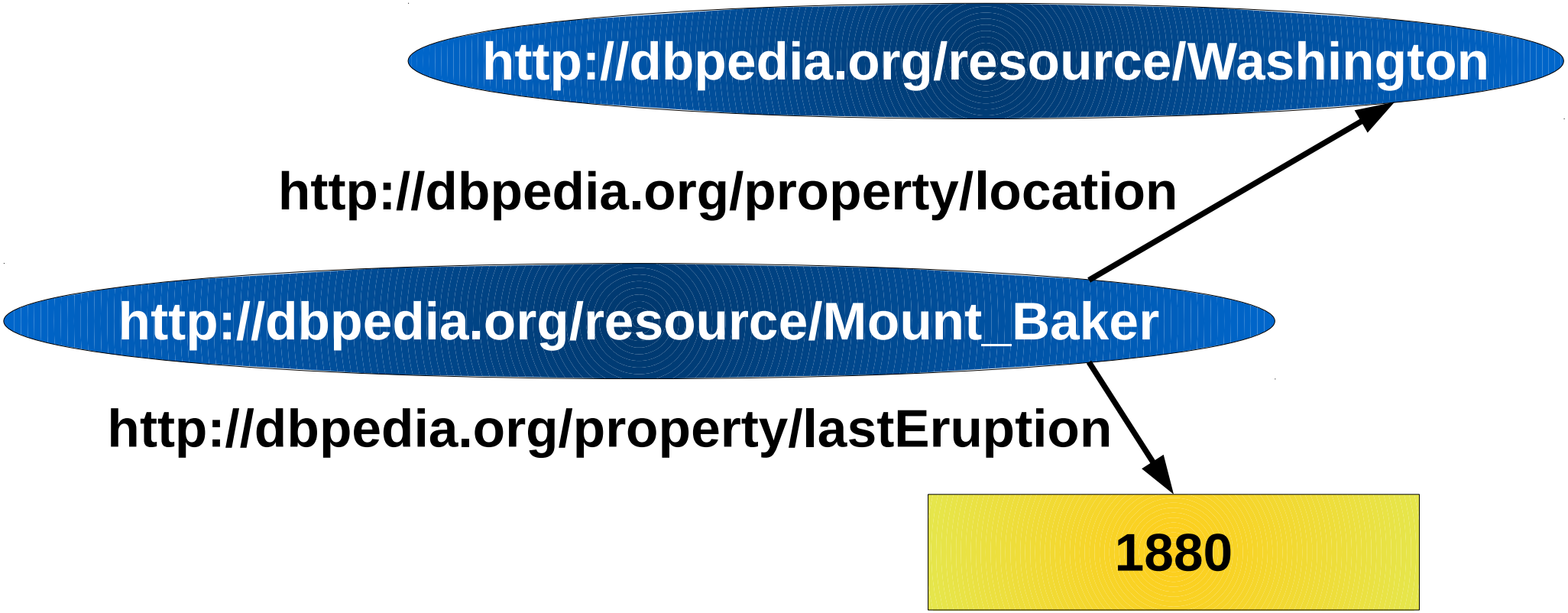


Uniform Resource Identifier (URI)

- **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/~hartig/foaf.rdf#olaf`
  - RDF document about me:
    `http://olafhartig.de/~hartig/foaf.rdf`
  - HTML document about me:
    `http://olafhartig.de/~hartig/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)

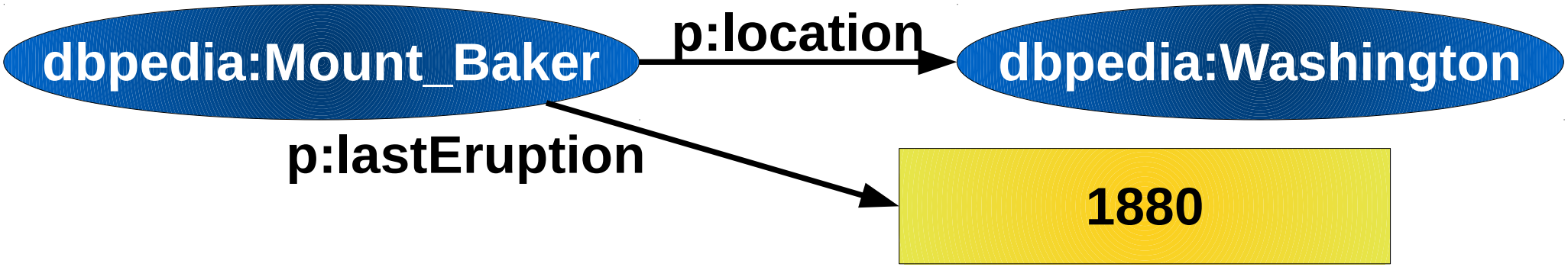


Compact URIs (CURIE)

- **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://olafhartig.de/~hartig/foaf.rdf#olaf`

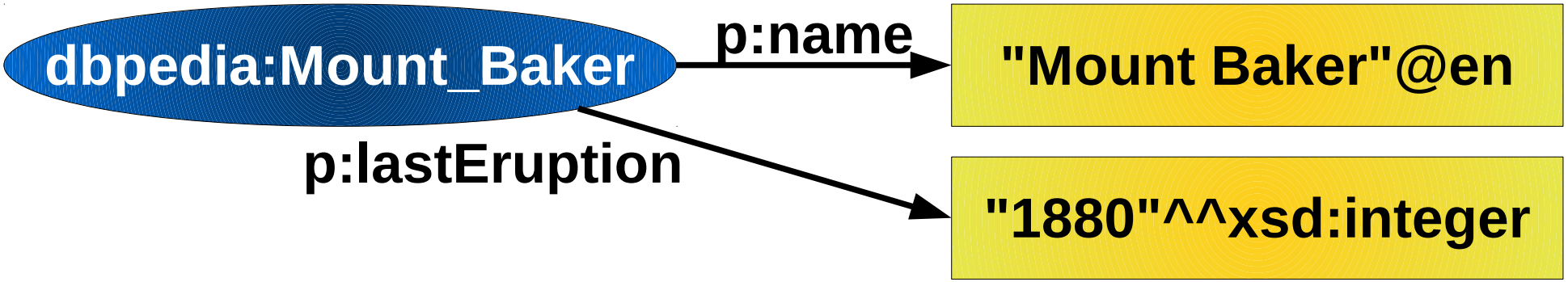
Example with CURIEs

- **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)**



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 – A Readable Syntax for RDF

- **Turtle supports 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 .
```

Turtle – A Readable Syntax for RDF

- **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 – 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 .
```

RDF/XML – An XML syntax for RDF

Turtle

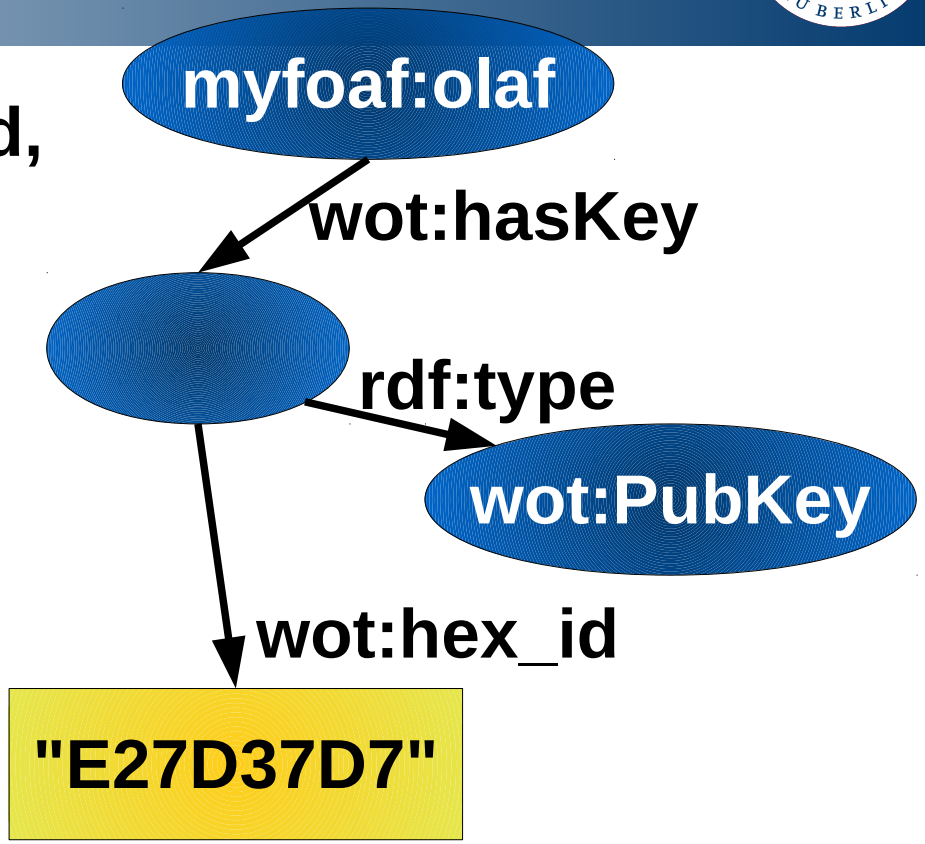
```
@prefix dbpedia : <http://dbpedia.org/resource/> .
@prefix p : <http://dbpedia.org/property/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

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

RDF/XML

```
<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>
```

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*
  - Scope: a single RDF graph




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

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

Vocabularies and Ontologies

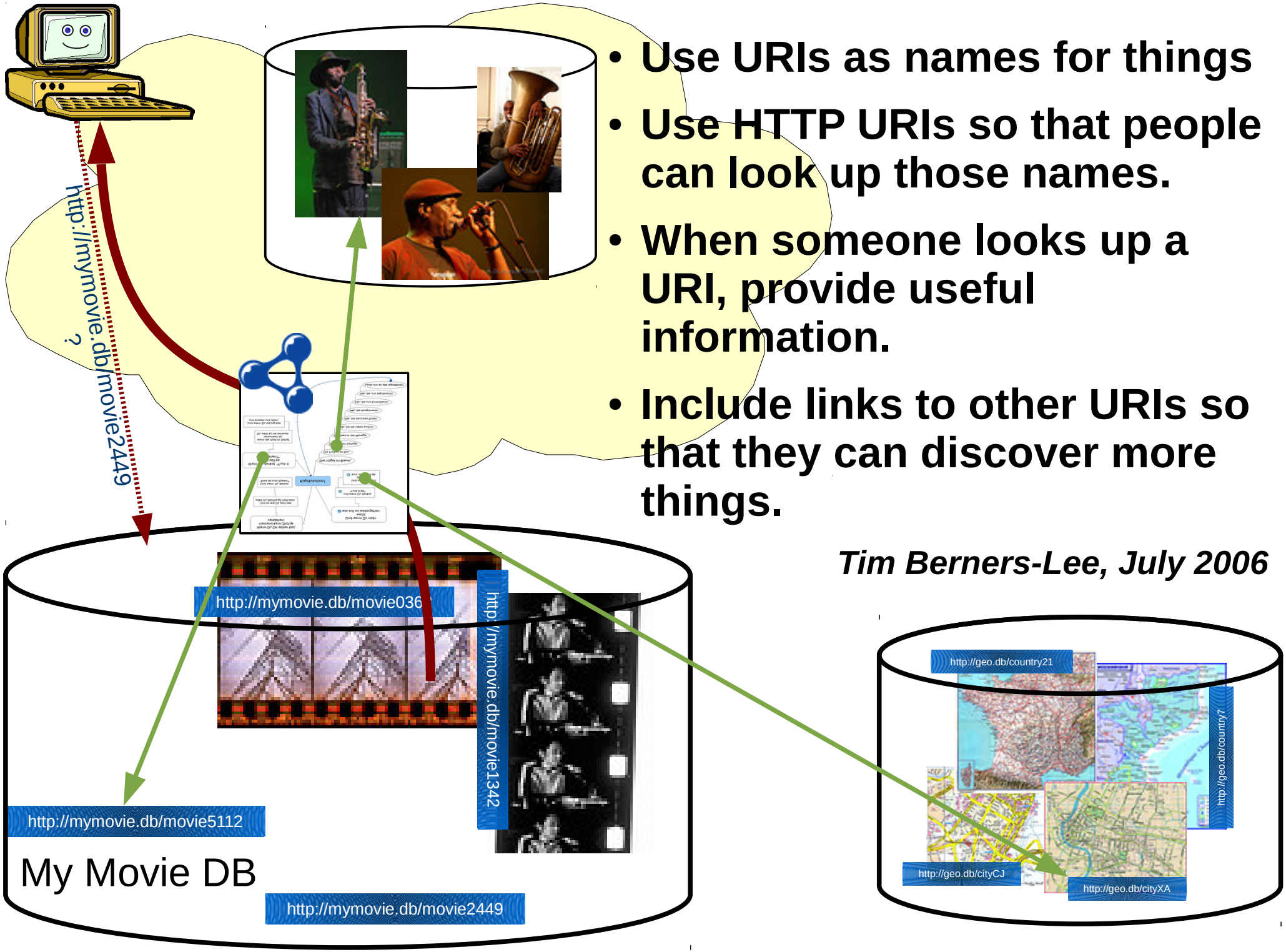
- **Defined using RDFS or OWL**
- **A plenty of vocabularies exist:**
  - People
  - Social media
  - Commerce
  - Events
  - Radio and TV programmes
  - Music

  etc.

Chapter 1

- Motivation
- Technical Foundations
- **The Web of Linked Data**

http://olafhartig.de/icwe.html



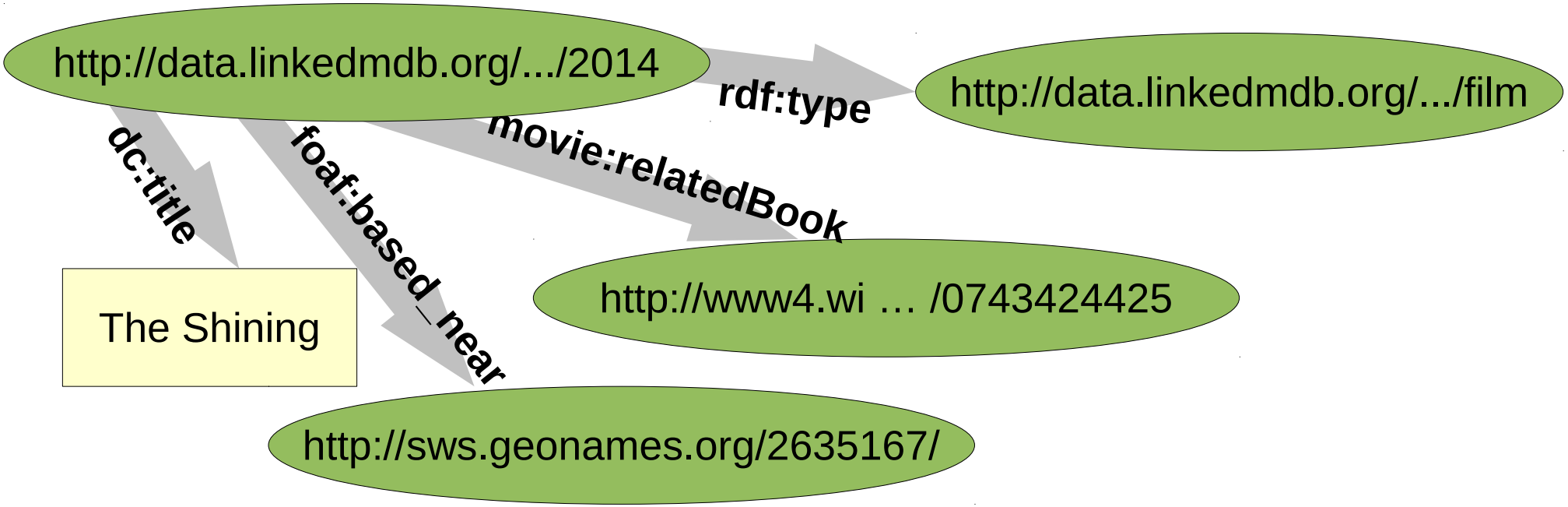
- **Use URIs as names for things**
- **Use HTTP URIs so that people can look up those names.**
- **When someone looks up a URI, provide useful information.**
- **Include links to other URIs so that they can discover more things.**

Tim Berners-Lee, July 2006

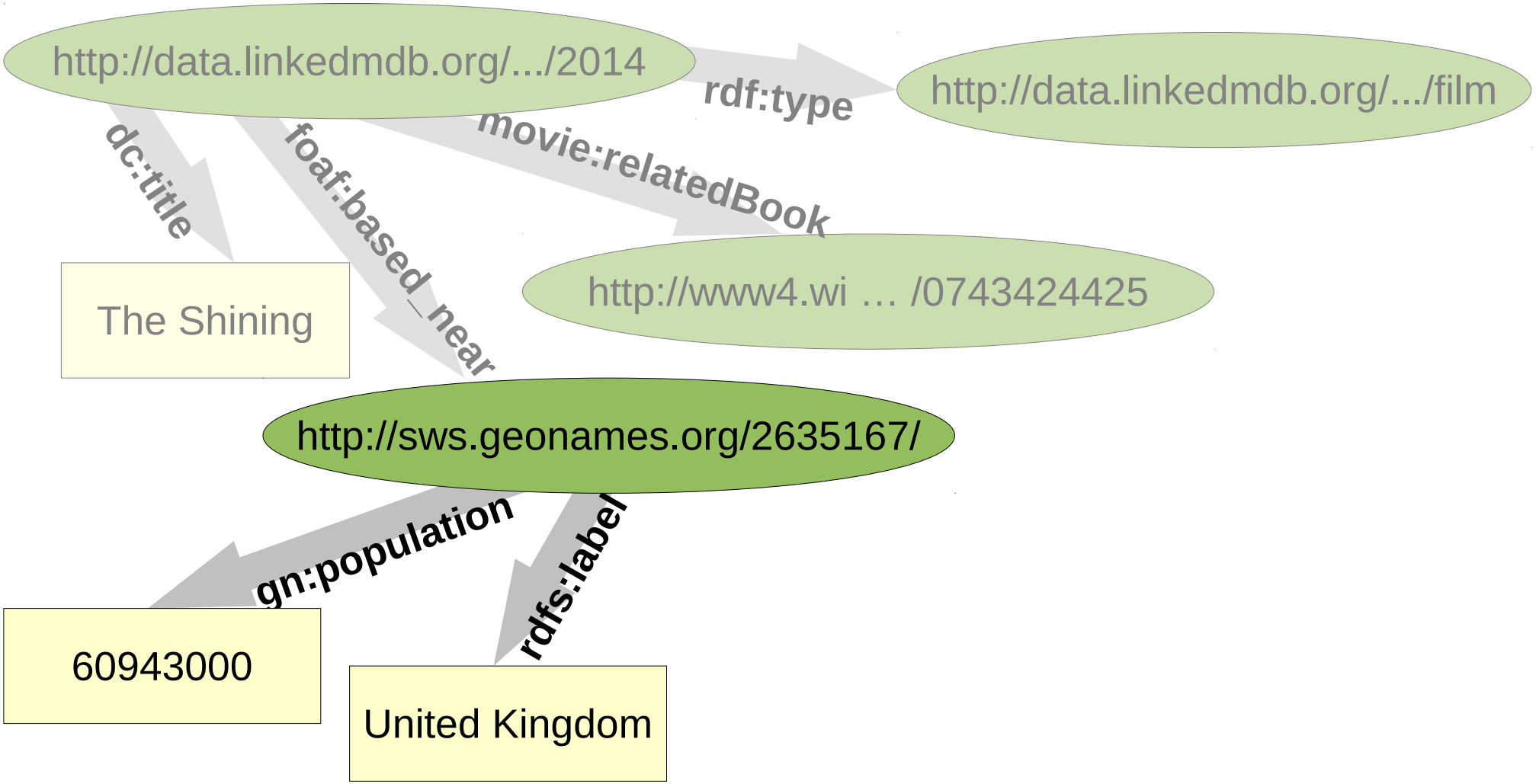
Linked Data – An Example

http://data.linkedmdb.org/.../2014

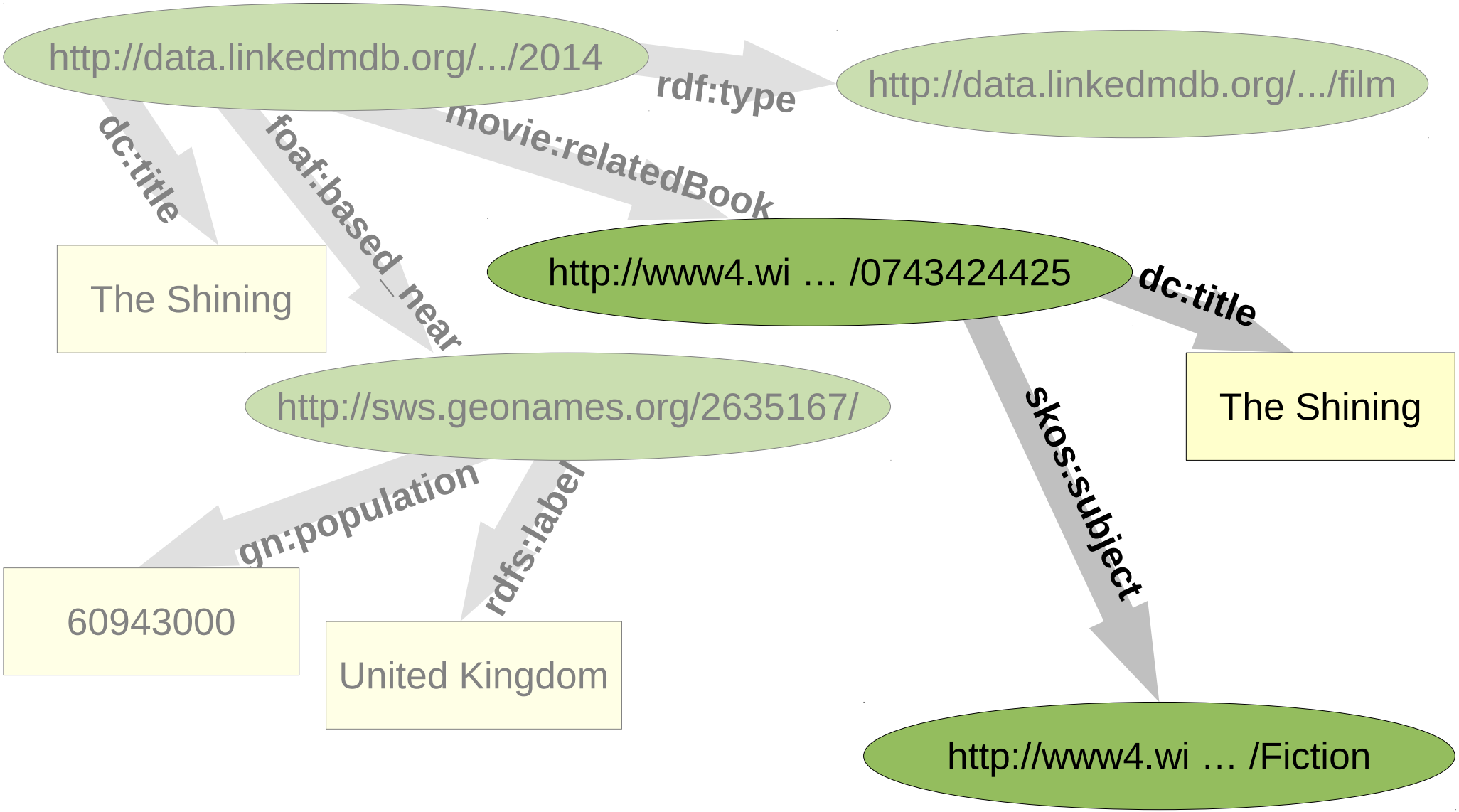
Linked Data – An Example



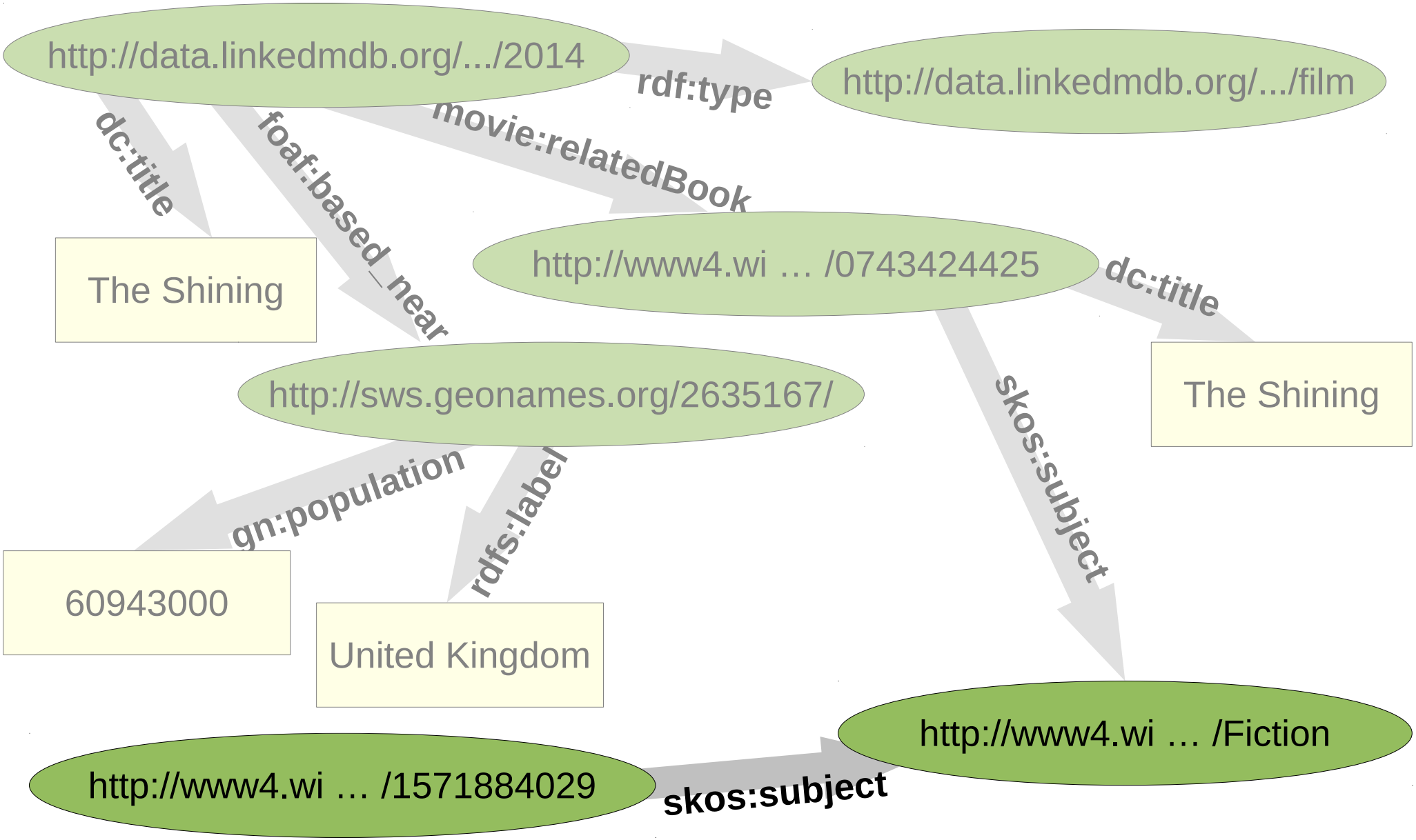
Linked Data – An Example



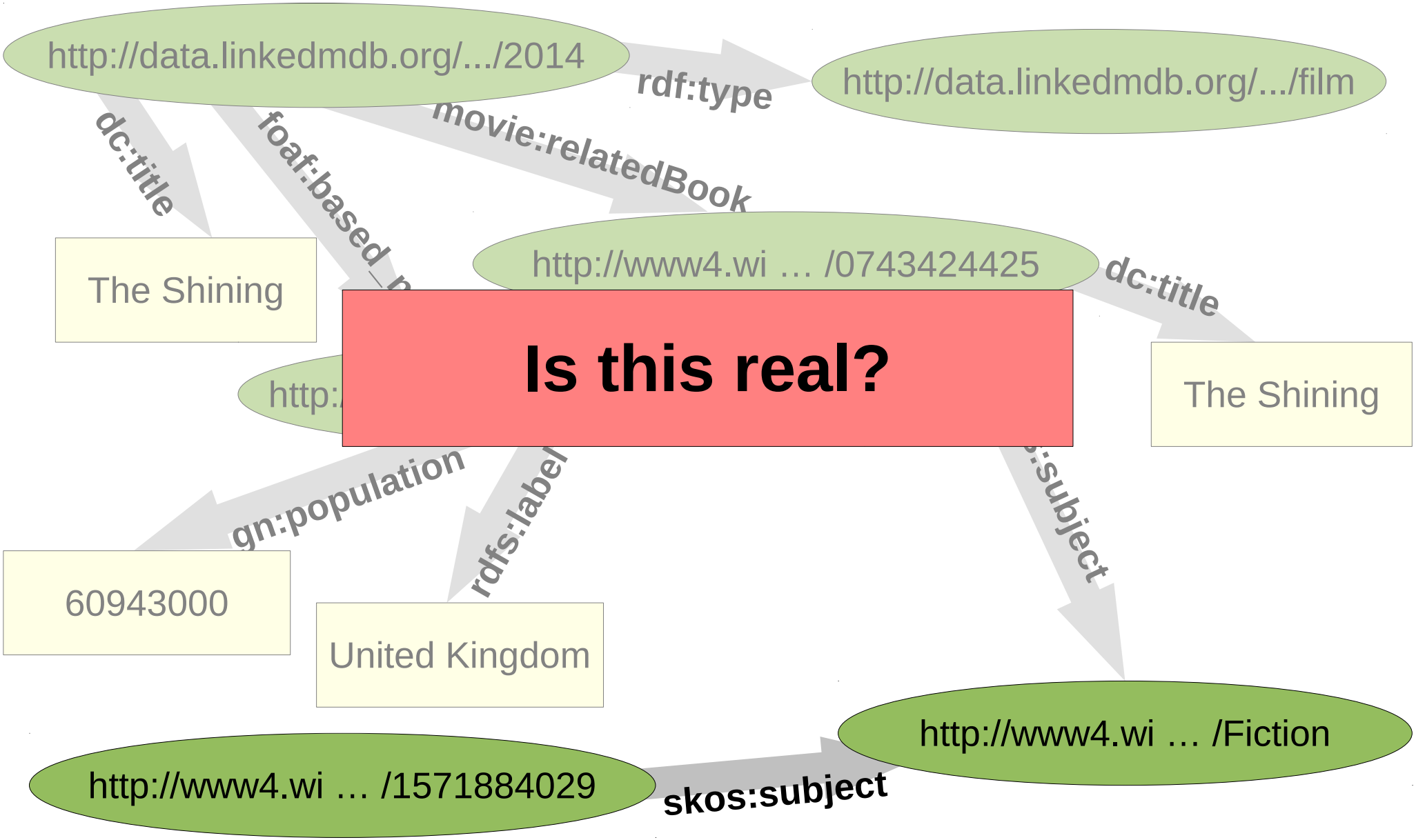
Linked Data – An Example



Linked Data – An Example



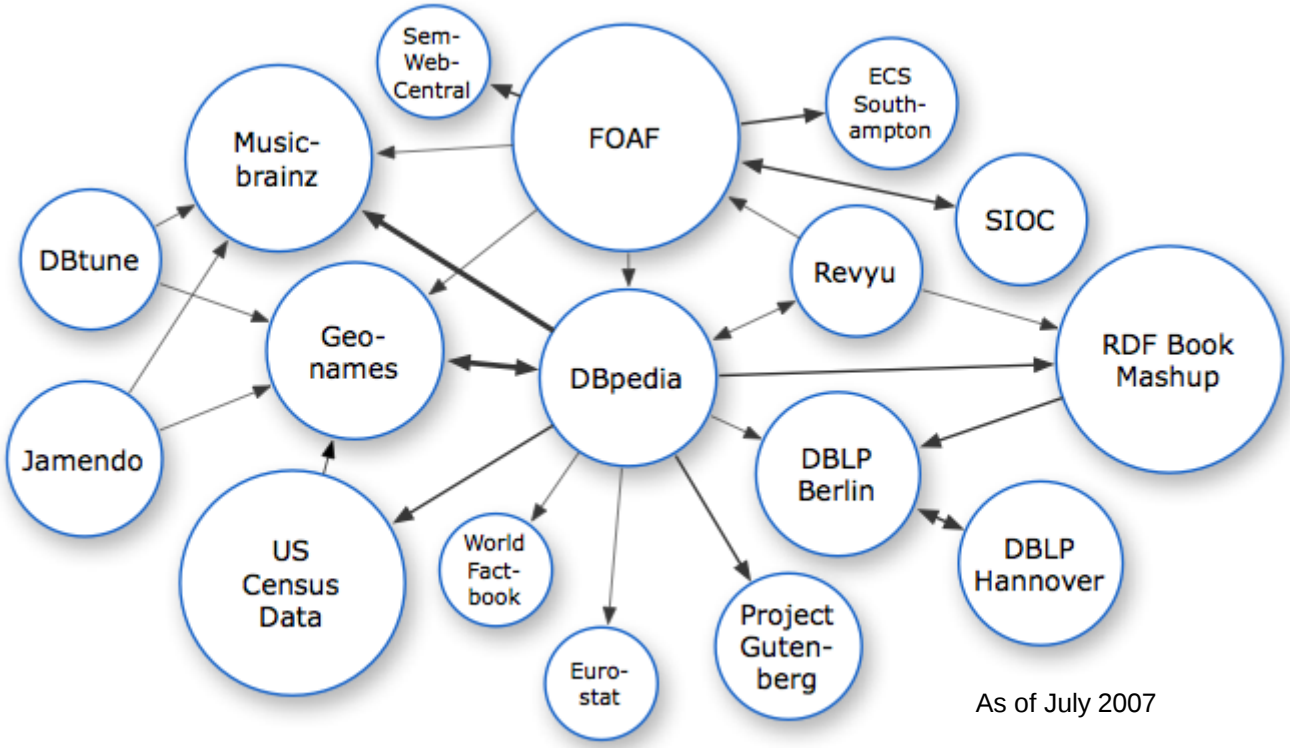
Linked Data – An Example



W3C Linking Open Data Project

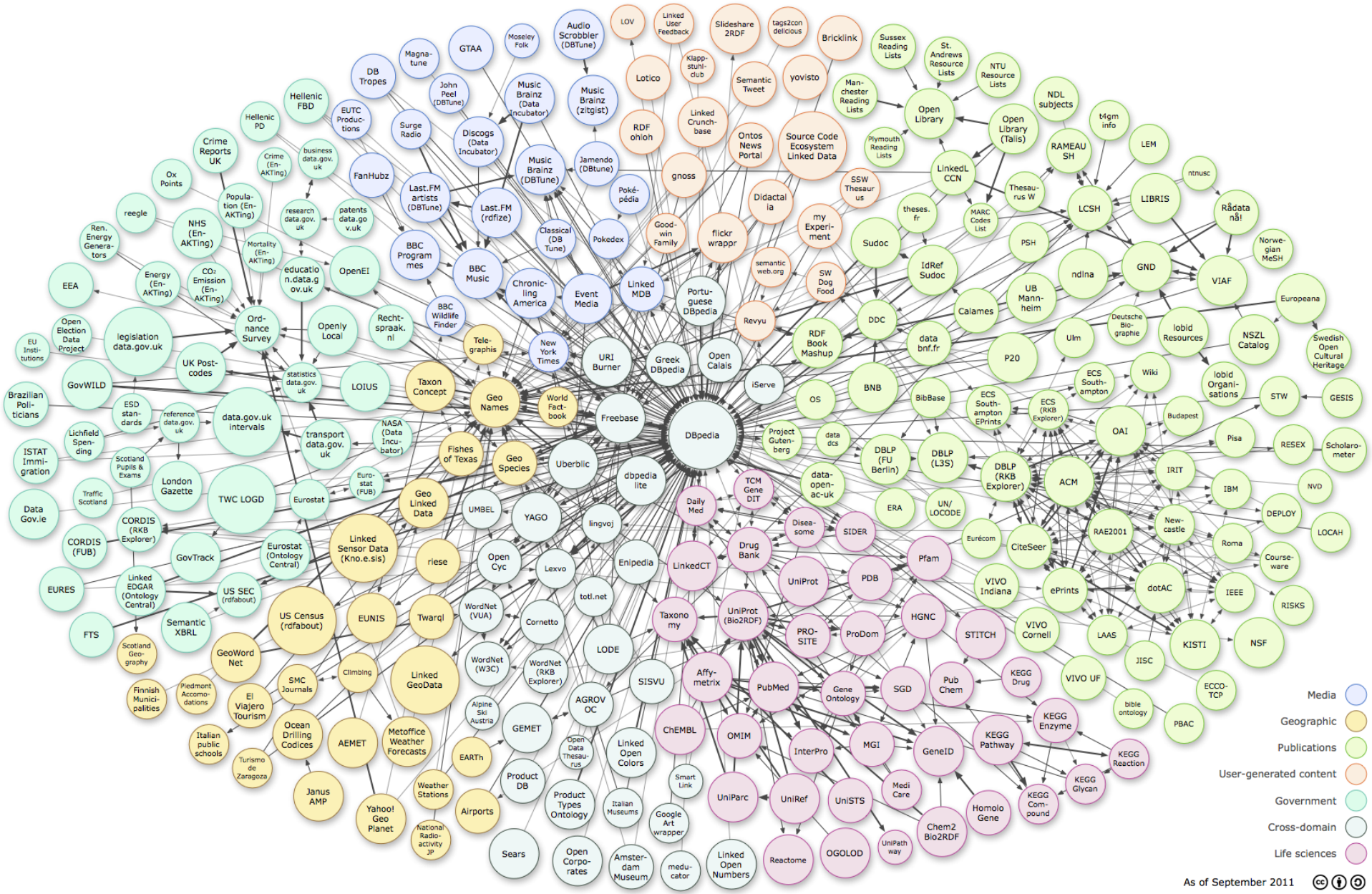
- **Grassroots community effort**
- **Publish existing, open license datasets as Linked Data**
- **Interlink things between different data sources**

W3C Linking Open Data Project



As of July 2007

W3C Linking Open Data Project



Linked Data Publishers

- UK government
- US government
- Thomson Reuters (Open Calais)
- MetaWeb (Freebase)
- BBC
- NY Times
- Best Buy
- CNET

etc.

Outline

Chapter 1:
Linked Data and RDF

Chapter 2:
The SPARQL Query Language

Chapter 3:
Querying Linked Data on the Web

http://olafhartig.de/icwe.html

These slides have been created by
Olaf Hartig

http://olafhartig.de

Some of the slides in the motivations section are inspired by similar slides from Chris Bizer and from Juan Sequeda.