

# RDF and semantic data

---

Adam Carter  
Project Manager, EPCC  
[A.Carter@epcc.ed.ac.uk](mailto:A.Carter@epcc.ed.ac.uk)  
+44 131 650 6009

- Almost synonymous
- As mentioned in the **metadata** lecture, you can use semantic technologies on non-web platforms (e.g. in file metadata)
- They're most often used in an online (web) context
- ...and this is where their power really lies

From: *Linked Data – Design Issues* – Tim Berners-Lee

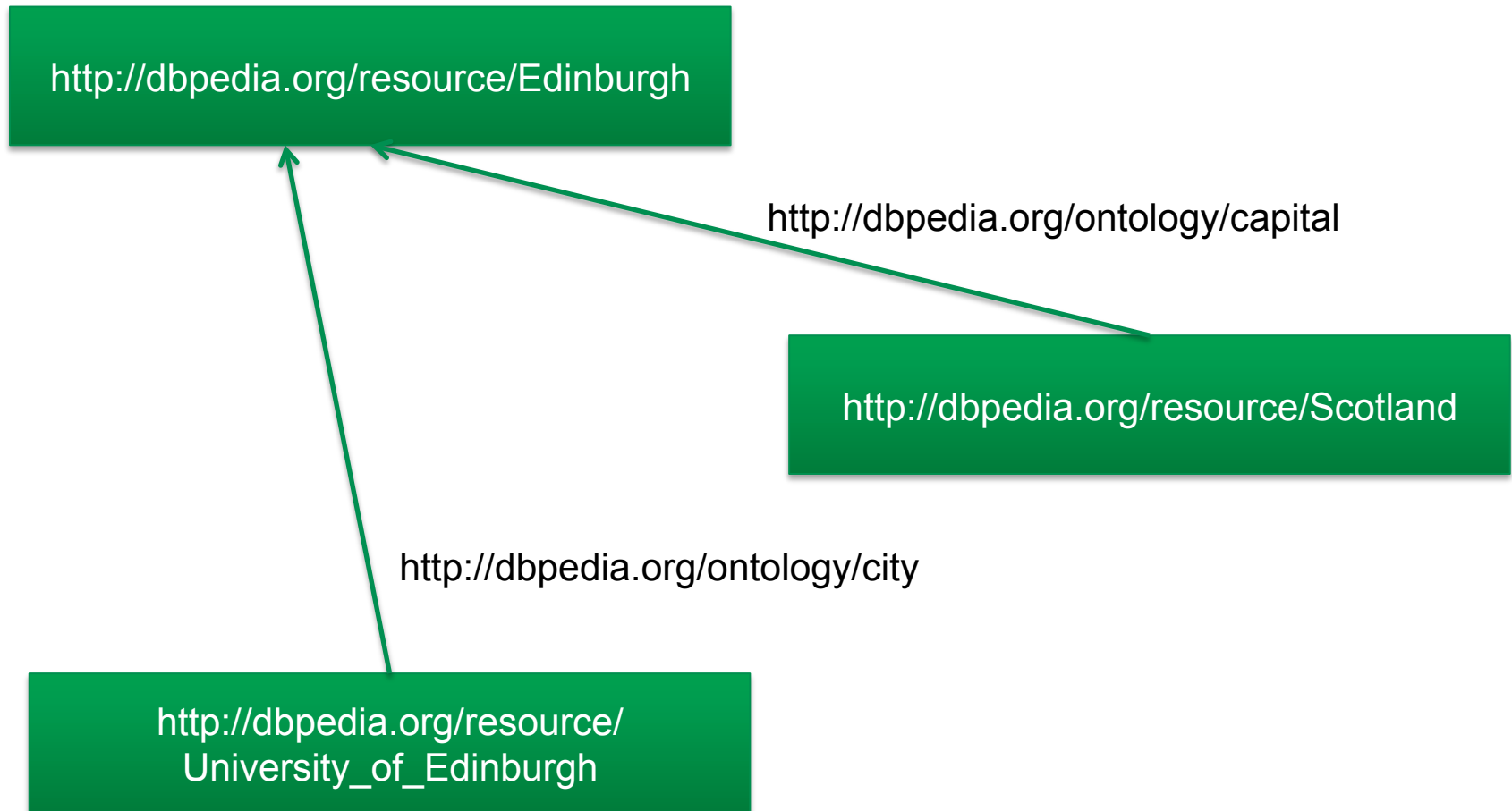
<http://www.w3.org/DesignIssues/LinkedData.html>

- 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, using the standards (RDF, SPARQL)
- Include links to other URIs, so that they can discover more things.



- RDF is a data model: the **R**esource **D**escription **F**ramework
- Information is encoded as sets of **triples**:
- **SUBJECT** **PREDICATE** **OBJECT**
- The **cat** sat on **the** **mat**

- **SUBJECT PREDICATE OBJECT**
  - The subject is either
    - a **resource** with a **URI**, or
    - a **blank node**
  - The predicate is
    - a **resource** with a **URI**
  - The object is
    - a **resource** with a **URI**, or
    - a **blank node** or
    - a literal
- RDF triples can be drawn as a graph
  - Each subject or object is a node in the graph
  - Each predicate is an edge in the graph



```
<rdf:Description rdf:about="http://dbpedia.org/  
resource/Scotland">
```

```
    <dbpedia-owl:capital rdf:resource="http://  
dbpedia.org/resource/Edinburgh" />
```

```
</rdf:Description>
```

- Turtle
- N-Triples
- N-Quads
- JSON-LD
- N3
- RDF/XML



`<http://dbpedia.org/resource/Scotland> <http://dbpedia.org/ontology/capital> <http://dbpedia.org/resource/Edinburgh> .`

`<http://dbpedia.org/resource/University\_of\_Edinburgh> <http://dbpedia.org/ontology/city> <http://dbpedia.org/resource/Edinburgh> .`

```
<http://dbpedia.org/resource/Scotland> <http://  
dbpedia.org/ontology/capital> <http://  
dbpedia.org/resource/Edinburgh> .
```

```
<http://dbpedia.org/resource/  
University_of_Edinburgh> <http://dbpedia.org/  
ontology/city> <http://dbpedia.org/resource/  
Edinburgh> .
```

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

```
@prefix dbpOnt <http://dbpedia.org/ontology/> .
```

```
dbpRes:Scotland dbpOnt:capital
```

```
dbpRes:Edinburgh .
```

```
dbpRes:University_of_Edinburgh dbpOnt:city
```


```
dbpRes:Edinburgh .
```

More about turtle: <http://www.w3.org/TR/turtle/>

- rdf
  - <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
- rdfs
  - <http://www.w3.org/2000/01/rdf-schema#>
- owl
  - <http://www.w3.org/2002/07/owl#>
- xsd
  - <http://www.w3.org/2001/XMLSchema#>
- dc
  - <http://purl.org/dc/elements/1.1/>
- foaf
  - <http://xmlns.com/foaf/0.1/>

Find others at <http://prefix.cc>

- A subset of Turtle
- Line-Based: You can't have line-breaks in a triple
- You can't use CURIEs (compact URIs)



```
<html xmlns:wiki="http://en.wikipedia.org/wiki/">
  <head>...</head>
  <body>
    <p>
      Find out more about <a href="[wiki:Biome]">biomes</a>.
    </p>
  </body>
</html>
```



An RDF/XML document is an XML document

```
<?xml version="1.0"?>
```

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dbpedia-owl="http://dbpedia.org/ontology/"
```

```
<rdf:Description rdf:about="http://dbpedia.org/resource/Scotland">
```

```
<dbpedia-owl:capital  
rdf:resource="http://dbpedia.org/  
resource/Edinburgh" />
```

```
</rdf:Description>
```

```
</rdf:RDF>
```

The document contains RDF. Specifically, it's the RDF defined at <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.

Subject

Object

Predicate

Note that the predicate and object appear in the same tag.

```
curl http://www.w3.org/1999/02/22-rdf-syntax-ns#
```

Print the contents of this document to stdout

```
curl -i http://www.w3.org/1999/02/22-rdf-syntax-ns#
```

Print the contents of this document to stdout *with response headers*

```
mbpac2:~ acarter$ curl -i http://www.w3.org/1999/02/22-rdf-syntax-ns#
HTTP/1.1 200 OK
Date: Mon, 06 Oct 2014 15:02:13 GMT
Server: Apache/2
Content-Location: 22-rdf-syntax-ns.ttl
Vary: negotiate,accept,accept-charset
TCN: choice
Last-Modified: Tue, 25 Feb 2014 02:36:35 GMT
ETag: "1210-4f331f4ed92c0;4f3320aa1f34b"
Accept-Ranges: bytes
Content-Length: 4624
Cache-Control: max-age=21600
Expires: Mon, 06 Oct 2014 21:02:13 GMT
P3P: policyref="http://www.w3.org/2014/08/p3p.xml"
Access-Control-Allow-Origin: *
Content-Type: text/turtle; charset=utf-8
```

```
curl -H "Accept: application/rdf+xml"
```

Use a specific  
request header

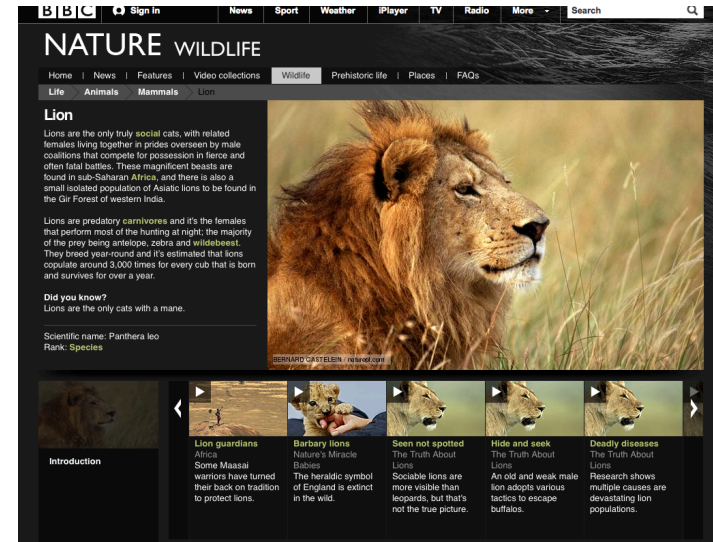
```
http://www.example.com/resource
```

```
[{
  "@context": {
    "capital": "http://dbpedia.org/ontology/capital",
  },
  "@id": "http://dbpedia.org/resource/Scotland",
  "capital": "http://dbpedia.org/resource/Edinburgh"
},{
  "@context": {
    "city": "http://dbpedia.org/ontology/city",
  },
  "@id": "http://dbpedia.org/resource/University_of_Edinburgh",
  "city": "http://dbpedia.org/resource/Edinburgh"
}]
```

- DBpedia
- MusicBrainz
- BBC
- Freebase
- FOAF
- US Census data
- CiteSeer
- GeoNames
- YAGO
- CIA Factbook



<http://www.bbc.co.uk/nature/life/Lion>



`curl -H "Accept: application/rdf+xml"`

<http://www.bbc.co.uk/nature/life/Lion>

see also [www.bbc.co.uk/ontologies](http://www.bbc.co.uk/ontologies)

```

$ curl -H "Accept: application/rdf+xml" http://www.bbc.co.uk/nature/life/Lion
<?xml version="1.0" encoding="utf-8"?><rdf:RDF
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:dc="http://purl.org/dc/terms/"
  xmlns:dctypes="http://purl.org/dc/dctypes/"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:sg="http://www.w3.org/2003/06/sg-schema#"
  xmlns:po="http://purl.org/ontology/po/"
  xmlns:vo="http://purl.org/ontology/vo/"
  <rdf:Description rdf:about="nature/species/Lion">
    <foaf:primaryTopic rdf:resource="nature/species/Lion#species"/>
    <rdfs:seeAlso rdf:resource="nature/species"/>
  </rdf:Description>

  <vo:Species rdf:about="nature/life/Lion#species">
    <rdfs:label>Lion</rdfs:label>

    <vo:name rdf:resource="http://www.bbc.co.uk/nature/species/Lion#name"/>

    <foaf:depiction rdf:resource="http://ichef.bbci.co.uk/naturelibrary/images/ic/640x360/1/1/lion/lion_1.jpg"/>

    <dc:description>Lions are the only truly social cats, with related females living together in prides overseen by male coalitions that compete for possession in fierce and often fatal battles. These magnificent beasts are found in sub-Saharan Africa, and there is also a small isolated population of Asiatic lions to be found in the Gir Forest of western India.
  </dc:description>
  </rdf:Description>
</rdf:RDF>

```



“An ***RDF vocabulary*** is a collection of IRIs intended for use in RDF graphs. For example, the IRIs documented in [RDF11-SCHEMA] are the RDF Schema vocabulary. RDF Schema can itself be used to define and document additional RDF vocabularies.”

-- *RDF 1.1 Concepts and Abstract Syntax*

<http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>

- Defines terminology that can be used in RDF documents (e.g. classes, properties)
- Can also be used to specify taxonomies for both classes and properties
- OWL documents are themselves RDF documents

- In files using a serialisation
  - RDF/XML, Turtle, etc.
- In a triplestore
  - A kind of database
- Note that much semantic web data is distributed
  - A single RDF statement can relate entities described in three different places

- Bulk RDF download (in the RDF/XML, Turtle, etc.)
- Web query
  - HTTP GET Accept:application/xml+rdf
- SPARQL
  - A query language for trip

- On the Semantic Web, resources are represented by URIs.
- Most resources are *not* web pages, e.g., “<http://dbpedia.org/resource/Edinburgh>” is the City of Edinburgh. You can’t get this over http!
- When you type this address into a web browser, you are redirected to another URL “<http://dbpedia.org/page/Edinburgh>”
- Using **curl** shows you the actual response, namely “303 See Other”

```
mbpac2:~ acarter$ curl -i http://dbpedia.org/resource/Edinburgh
HTTP/1.1 303 See Other
Date: Mon, 06 Oct 2014 15:25:24 GMT
Content-Type: text/html; charset=UTF-8
Content-Length: 0
Connection: keep-alive
Server: Virtuoso/07.10.3211 (Linux) x86_64-redhat-linux-gnu VDB
Location: http://dbpedia.org/page/Edinburgh
```



- The other way to differentiate between the document returned by HTTP and the object itself is by use of a fragment identifier, e.g.
  - <http://www.w3.org/2000/01/rdf-schema#Literal>
- More details at <http://www.w3.org/TR/cooluris/>

- SPARQL is to RDF what SQL is to data in relational databases
- Types of SPARQL queries:
  - SELECT
    - Returns a table of results
  - CONSTRUCT
    - Used to construct a new RDF graph
  - DESCRIBE
    - Find all the statements that mention a resource
  - ASK
    - A query that just returns true/false

Typically expressed as XML or JSON, e.g., with  
*SPARQL Query Results XML Format*  
<http://www.w3.org/TR/rdf-sparql-XMLres/>

```
:t1 foaf:name "Adam Carter" .  
:t1 foaf:based_near :Edinburgh .  
:t2 foaf:name "Barack Obama" .  
:t2 foaf:based_near :WashingtonDC .  
  
SELECT ?name  
WHERE {  
  ?x foaf:name ?name  
}
```

Returns: Adam Carter, Barack Obama

Note the common shorthand: a resource that begins with ":" means that the full name is prepended by the URI of the containing document.

The implication here is that entries like ":Edinburgh" are defined in full elsewhere in the document.

# Format of a query

```
PREFIX ns1: <...>
PREFIX ns2: <...>
...
SELECT ...
FROM <...>
FROM NAMED <...>
WHERE {
...
}
GROUP BY ...
HAVING ...
ORDER BY ...
LIMIT ...
OFFSET ...
VALUES ...
```



Define the dataset (optional)

- SPARQL is also a protocol
  - Full name: SPARQL over HTTP

“SparqlQuery is the protocol's only interface. It contains one operation, query, which is used to convey a SPARQL query string and, optionally, an RDF dataset description.”

-- from <http://www.w3.org/TR/rdf-sparql-protocol/>

```
GET /sparql/?query=EncodedQuery&default-graph-uri=http://www.other.example/books HTTP/1.1
Host: www.other.example
User-agent: my-sparql-client/0.1
```

Here *EncodedQuery* uses standard URL encoding



- Once you have machine readable facts with well defined meanings, you can reason, e.g.
  - Adam Carter lives in Edinburgh
  - Edinburgh is in Scotland
  - Therefore, Adam Carter lives in Scotland

- RDF is a data model for representing semantic data
- It relies on giving every resource (every “thing”) a unique label (a URI) and describing relationships between resources with **triples**
- RDF can be serialised in multiple ways
- Using content negotiation and redirection allows correct differentiation between objects and their description in HTML
- Descriptions of the meanings of sets of terms are combined into **ontologies** which themselves can be expressed as RDF using **OWL**