OPEN
DATA
SUPPORT

Training Module 1.3

Introduction to RDF & SPARQL

Presentation metadata

Open Data Support is funded by the European Commission under SMART 2012/0107 'Lot 2: Provision of services for the Publication, Access and Reuse of Open Public Data across the European Union, through existing open data portals'(Contract No. 30-CE-0530965/00-17).

This presentation has been created by PwC

Authors:
Michiel De Keyzer, Nikolaos Loutas and Stijn Goedertier

Disclaimers

1. The views expressed in this presentation are purely those of the authors and may not, in any circumstances, be interpreted as stating an official position of the European Commission.
The European Commission does not guarantee the accuracy of the information included in this presentation, nor does it accept any responsibility for any use thereof.
Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Commission.
All care has been taken by the author to ensure that s/he has obtained, where necessary, permission to use any parts of manuscripts including illustrations, maps, and graphs, on which intellectual property rights already exist from the titular holder(s) of such rights or from her/his or their legal representative.

2. This presentation has been carefully compiled by PwC, but no representation is made or warranty given (either express or implied) as to the completeness or accuracy of the information it contains. PwC is not liable for the information in this presentation or any decision or consequence based on the use of it.. PwC will not be liable for any damages arising from the use of the information contained in this presentation. The information contained in this presentation is of a general nature and is solely for guidance on matters of general interest. This presentation is not a substitute for professional advice on any particular matter. No reader should act on the basis of any matter contained in this publication without considering appropriate professional advice.

Learning objectives

By the end of this training module you should have an understanding of:

- The Resource Description Framework (RDF);
- How to write/read RDF;
- How you can describe your data with RDF;
- What SPARQL is;
- The different types of SPARQL queries;
- How to write a SPARQL query.

Content

This module contains …

- An introduction to the Resource Description Framework (RDF) for describing your data.
  - What is RDF?
  - How is it structured?
  - How to represent your data in RDF.
- An introduction to SPARQL on how you can query and manipulate data in RDF.
- Pointers to further reading, examples and exercises.

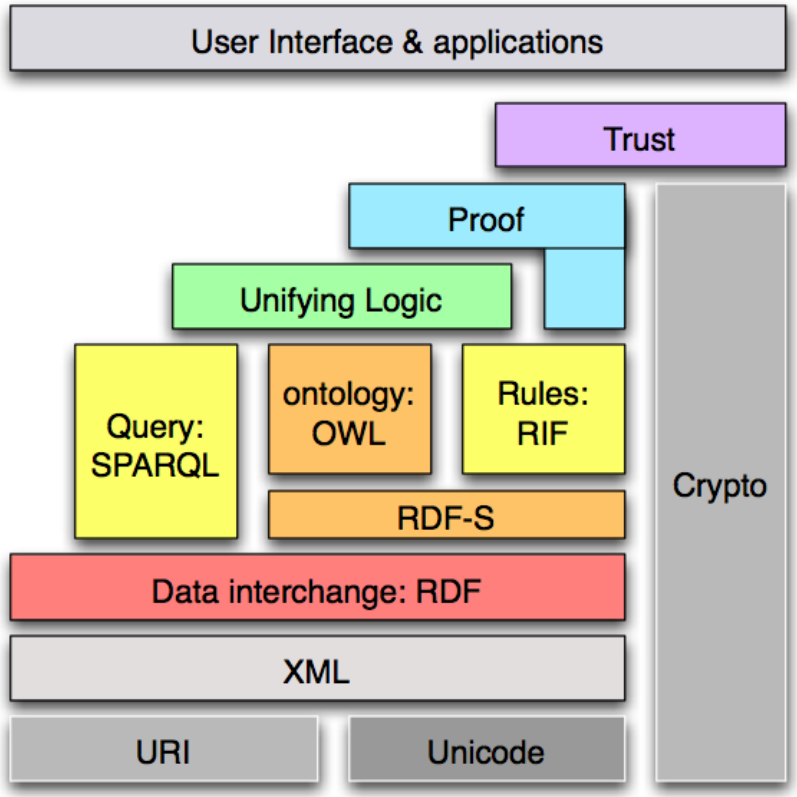
Find more on: ***training.opendatasupport.eu***

Resource Description Framework

An introduction on RDF.

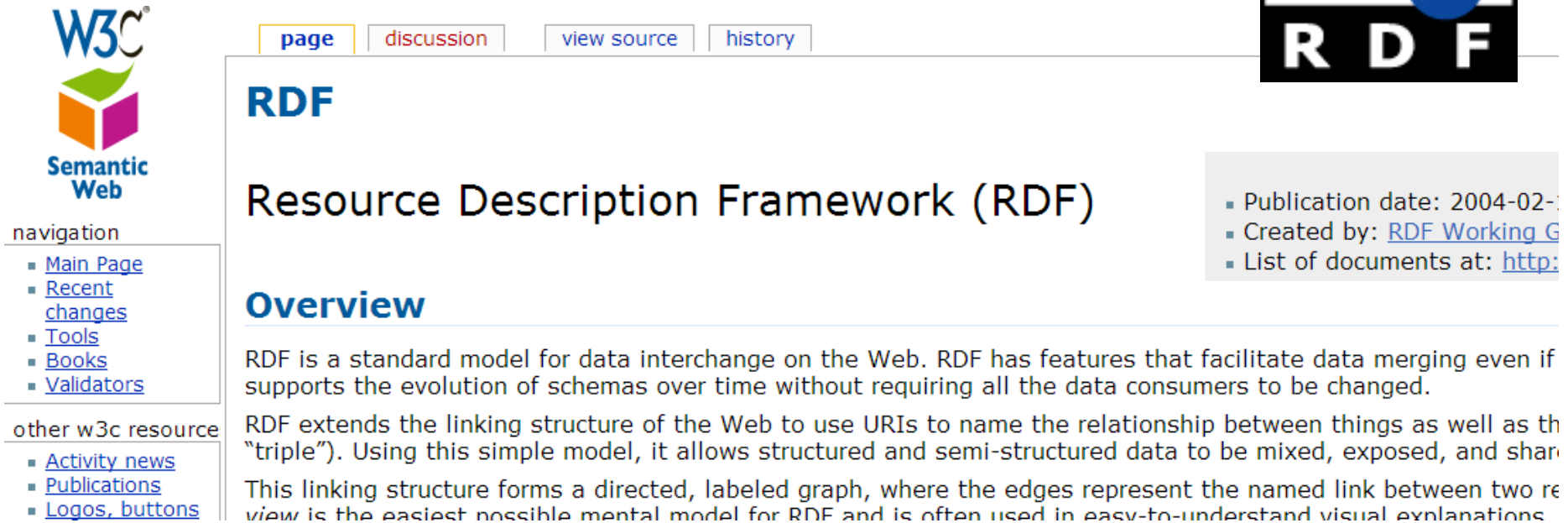
RDF in the stack of Semantic Web technologies

- RDF stands for:
  - **Resource**: Everything that can have a unique identifier (URI), e.g. pages, places, people, dogs, products…
  - **Description**: attributes, features, and relations of the resources
  - **Framework**: model, languages and syntaxes for these descriptions
- RDF was published as a **W3C recommendation** in 1999.
- RDF was originally introduced as a data model for **metadata.**
- RDF was generalised to cover **knowledge of all kinds**.



Where can I find the RDF Specification

http://www.w3.org/RDF/



Example: RDF description of an organisation

Publications Office, 2, rue Mercier, 2985 Luxembourg, LUXEMBOURG

```
<rdf:RDF
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:org="http://www.w3.org/ns/org#"
  xmlns:locn="http://www.w3.org/ns/locn#" >

<org:Organization rdf:about="http://publications.europa.eu/resource/authority/corporate-body/PUBL">
  <rdfs:label> "Publications Office"< /rdfs:label>
  <org:hasSite rdf:resource="http://example.com/site/1234"/>
</org:Organization>

<locn:Address rdf:about="http://example.com/site/1234"/>
  <locn:fullAddress>"2, rue Mercier, 2985 Luxembourg, LUXEMBOURG"</locn:fullAddress>
</locn:Address>

</rdf:RDF>
```

RDF structure

Triples, graphs and syntax.

What is a triple?

RDF is a general syntax for representing data on the Web.

Every piece of information expressed in RDF is represented as a **triple**:

- ***Subject*** – a resource, which may be identified with a URI.
- ***Predicate*** – a URI-identified reused specification of the relationship.
- ***Object*** – a resource or literal to which the subject is related.

Example: name of a dataset:

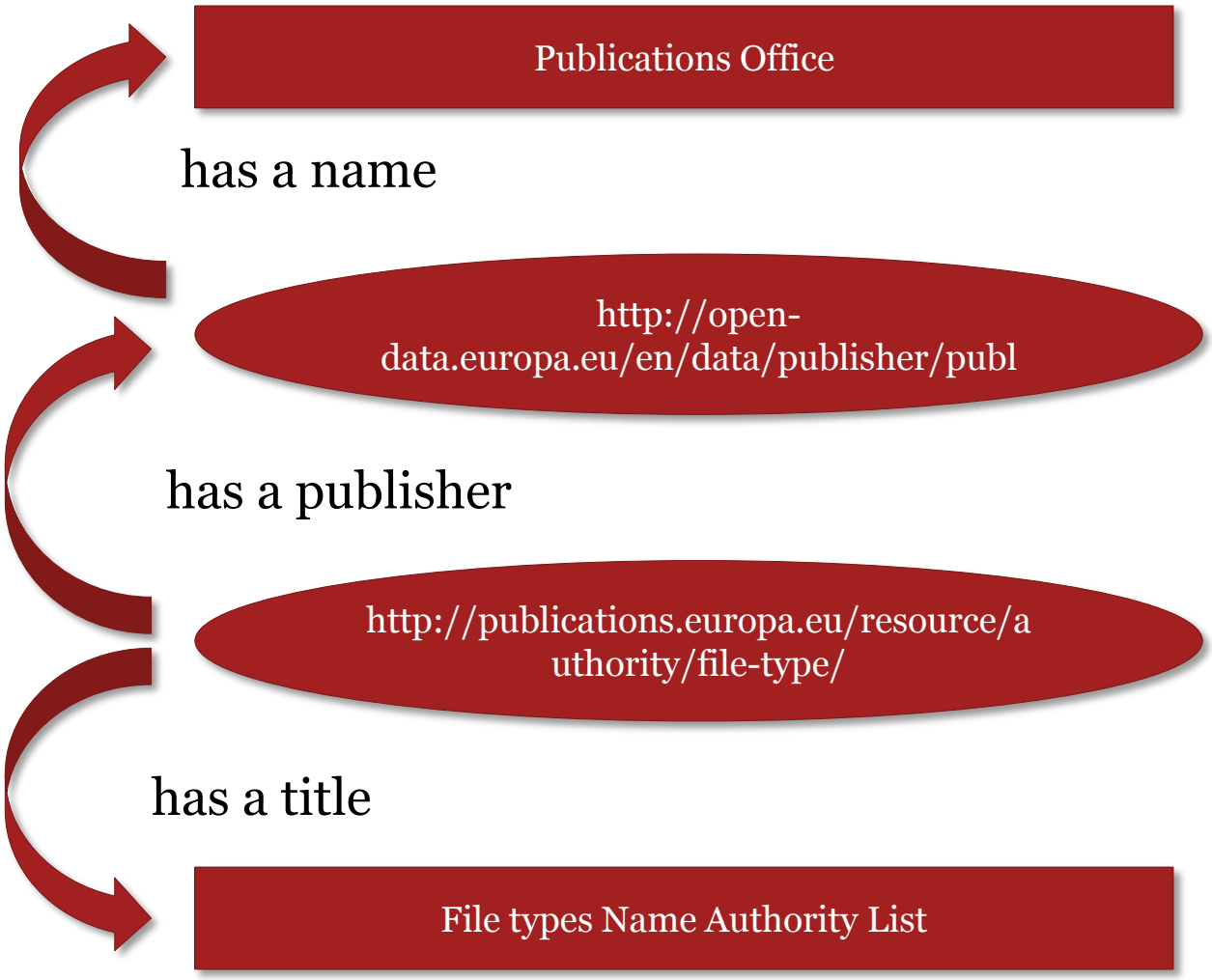
http://publications.europa.eu/resource/authority/file-type/ has a title "File types Name Authority List".

Subject Predicate Object

RDF is graph based

Graph =

A collection of triples



RDF Syntax
RDF/XML

```
<rdf:RDF

  xmlns:dcat="http://www.w3.org/TR/vocab-dcat/"
  xmlns:dct="http://purl.org/dc/terms/"

<dcat:Dataset rdf:about="http://publications.europa.eu/resource/authority/file-type/">
  <dct:title> "File types Name Authority List"< /dct:title>
  <dct:publisher rdf:resource="http://open-data.europa.eu/en/data/publisher/publ"/>
</dcat:Dataset>

<dct:Agent rdf:about="http://open-data.europa.eu/en/data/publisher/publ"/>
  <dct:title>"Publications Office"</dct:title>
</dct:Publisher>

</rdf:RDF>
```

Definition of prefixes

Description of data – triples

Graph

Subject

Predicate

Object

RDF/XML is currently the only syntax that is standardised by W3C.

RDF Syntax
Turtle

```
@prefix dcat: <http://www.w3.org/TR/vocab-dcat/> .
@prefix dct: <http://purl.org/dc/terms/.

< http://publications.europa.eu/resource/authority/file-type/>
 a <dcat:Dataset> ;
  dct:title "File types Name Authority List" ;
  dct:publisher <http://open-data.europa.eu/en/data/publisher/publ> .

<http://open-data.europa.eu/en/data/publisher/publ>
  a <dct:Agent> ;
  dct:title "Publications Office" .
```

Definition of prefixes

Description of data – triples

Graph

Subject

Predicate

Object

Turtle will be standardised in RDF 1.1.

See also:
http://www.w3.org/2009/12/rdf-ws/papers/ws11

RDF Syntax
RDFa

embedding RDF data in HTML

```
<html>
<head> ... </head>
<body>
...
<div resource="http://publications.europa.eu/resource/authority/file-type/"
typeof= "http://www.w3.org/ns/dcat#Dataset">
<p>
<span property=" http://purl.org/dc/terms/title ">File types Name Authority List<span>
Publisher: <span property="http://purl.org/dc/terms/Agent"> Publications Office</span>
</p></div>
 </body>
```

Subject

Predicate

Object

See also:
http://www.w3.org/TR/2012/NOTE-rdfa-primer-20120607/

How to represent data in RDF

Classes, properties and vocabularies

RDF Vocabulary

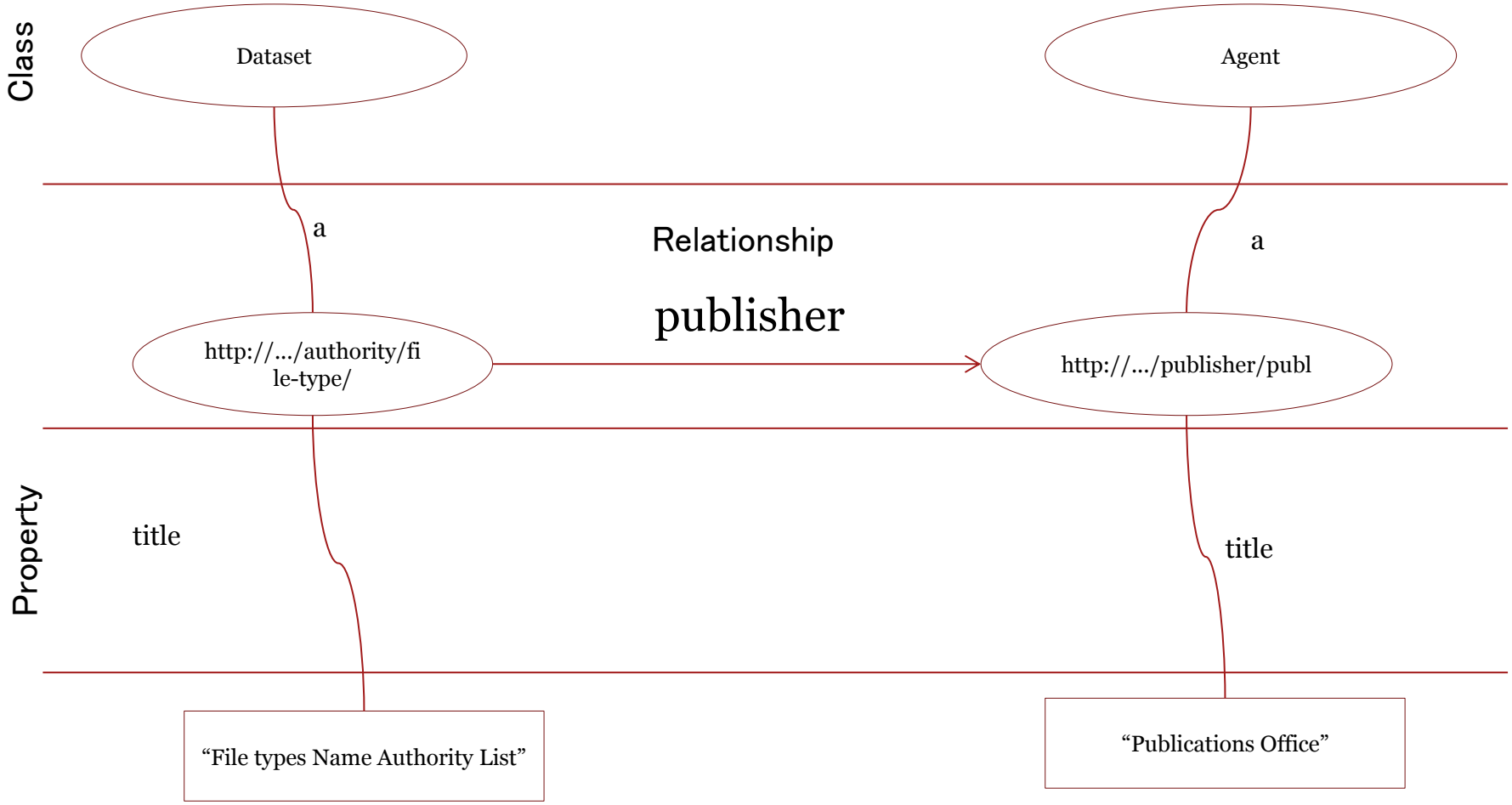
"A vocabulary is a data model comprising classes, properties and relationships which can be used for describing your data and metadata."

- RDF Vocabularies are ***sets of terms*** used to describe things.
- A term is either a ***class or a property***.
  - Object type properties (relationships)
  - Data type properties (attributes)

What are classes, relationships and properties?

- **Class.** A construct that represents things in the real and/or information world, e.g. a person, an organisation, a concepts such as “health” or “freedom”.
- **Relationship.** A link between two classes; for the link between a document and the organisation that published it (i.e. organisation *publishes* document), or the link between a map and the geographic region it depicts (i.e. map *depicts* geographic region). In RDF relationships are encoded as object type properties.
- **Property.** A characteristic of a class in a particular dimension such as the legal name of an organisation or the date and time that an observation was made.

Examples of classes, relationships and properties



Reusing RDF vocabularies

- Reuse greatly **aids interoperability** of your data.

  Use of dcterms:created, for example, the value for which should be a data typed date such as 2013-02-21^^xsd:date, is immediately processable by many machines. If your schema encourages data publishers to use a different term and date format, such as ex:date "21 February 2013" – data published using your schema will require further processing to make it the same as everyone else's.

- Reuse **adds credibility** to your schema.

  It shows it has been published with care and professionalism, again, this promotes its reuse.

- Reuse is **easier** and **cheaper**.

  Reusing classes and properties from well defined and properly hosted vocabularies avoids your having to replicate that effort.

See also:
https://joinup.ec.europa.eu/community/semic/document/cookbook-translating-data-models-rdf-schemas
http://www.slideshare.net/OpenDataSupport/model-your-data-metadata

Where can I find existing vocabularies?

http://joinup.ec.europa.eu/

See also:
http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/CommonVocabularies

Well-known vocabularies

DCAT-AP	**Vocabulary for describing datasets in Europe**
Core Person Vocabulary	Vocabulary to describe the fundamental characteristics of a person, e.g. the name, the gender, the date of birth…
DOAP	Vocabulary for describing projects
ADMS	**Vocabulary for describing interoperability assets.**
Dublin Core	**Defines general metadata attributes**
Registered Organisation Vocabulary	Vocabulary for describing organizations, typically in a national or regional register
Organization Ontology	for describing the structure of organizations
Core Location Vocabulary	Vocabulary capturing the fundamental characteristics of a location.
Core Public Service Vocabulary	Vocabulary capturing the fundamental characteristics of a service offered by public administration
schema.org	Agreed vocabularies for publishing structured data on the Web elaborated by Google, Yahoo and Microsoft

Model your own vocabulary as an RDF Schema

If there is no suitable authoritative reusable vocabulary for describing your data, use conventions for describing your own vocabulary:

- RDF Schema (RDFS)
- Web Ontology Language (OWL)

Example: definition of a class :

cpsv:PublicService a rdfs:Class, owl:Class;
rdfs:label "Public Service"*@en;*
rdfs:comment "This class represents the service itself. As noted in the scope, a public service is the capacity to carry out a procedure and exists whether it is used or not. It is a set of deeds and acts performed by or on behalf of a public agency for the benefit of a citizen, a business or another public agency."*@en.*

See also:
http://www.slideshare.net/OpenDataSupport/model-your-data-metadata

Introduction to SPARQL

The RDF Query Language

About SPARQL

SPARQL is the standard language to query graph data represented as RDF triples.

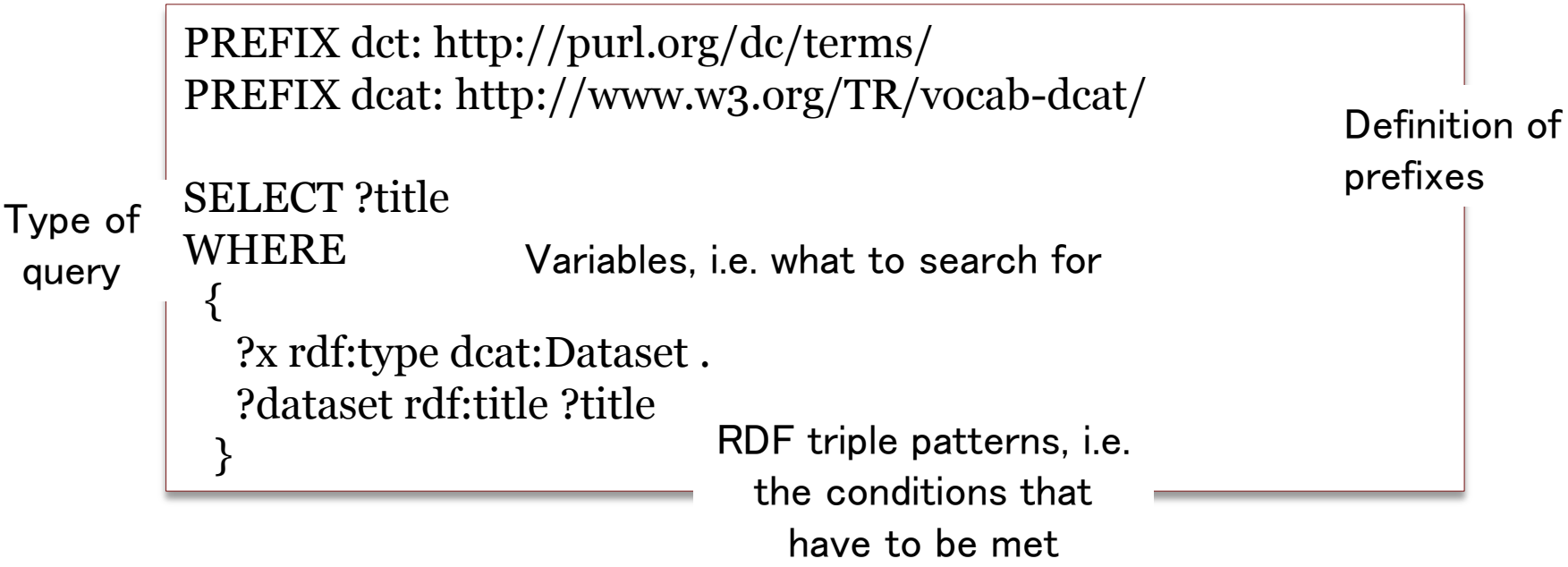
- **S**PARQL **P**rotocol and **R**DF **Q**uery **L**anguage
- One of the three core standards of the Semantic Web, along with RDF and OWL.
- Became a W3C standard January 2008.
- SPARQL 1.1 is a W3C Recommendation since March 2013.

Types of SPARQL queries

- **SELECT**
  Return a table of all X, Y, etc. satisfying the following conditions …
- **CONSTRUCT**
  Find all X, Y, etc. satisfying the following conditions … and substitute them into the following template in order to generate (possibly new) RDF statements, creating a new graph.
- **DESCRIBE**
  Find all statements in the dataset that provide information about the following resource(s) … (identified by name or description)
- **ASK**
  Are there any X, Y, etc. satisfying the following conditions …

See also:
http://www.euclid-project.eu/modules/chapter2

Structure of a SPARQL Query



SELECT – return the name of a dataset with particular URI

Sample data

```
<http://.../authority/file-type/>   rdf:type  dcat:Dataset.
<http://.../authority/file-type/>   dct:title  “File types Name Authority List“ .
<http://.../authority/file-type/>    dct:publisher  < http://open-data.europa.eu/en/data/publisher/publ>.

< http://.../publisher/publ>  rdf:type    dct:Agent .
< http://.../publisher/publ>  dct:title    “Publications Office” .
```

Query

```
PREFIX dcat: <http://www.w3.org/TR/vocab-dcat/>
PREFIX dct: <http://purl.org/dc/terms/>

SELECT ?dataset

WHERE
 {
    <http://.../authority/file-type/>  org:hasRegisteredSite  ?dataset .
 }
```

Result

name
“File types Name Authority List”

SELECT - return the name and publisher of a dataset

Sample data

```
<http://.../authority/file-type/>   rdf:type  dcat:Dataset.
<http://.../authority/file-type/>   dct:title  “File types Name Authority List“ .
<http://.../authority/file-type/>    dct:publisher  < http://open-data.europa.eu/en/data/publisher/publ>.

< http://.../publisher/publ>  rdf:type   dct:Agent .
< http://.../publisher/publ>  dct:title   “Publications Office” .
```

Query

```
PREFIX dcat: <http://www.w3.org/TR/vocab-dcat/>
PREFIX dct: <http://purl.org/dc/terms/>

SELECT ?dataset ?publisher

WHERE
 {http://.../authority/file-type/  dct:publisher ?publisherURI.
   http://.../authority/file-type/ dct:title  ?dataset.
   ?publisherURI  dct:title  ?publisher  . }
```

Result

dataset	pubisher
“File types Name Authority List”	“Dahliastraat 24, 2160 Wommelgem”

SPARQL Example – EU ODP (1)

SPARQL

You can search for the metadata stored in the EU Open Data Portal triple store by using the SPARQL endpoint query editor below.

Namespaces

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX odp:  <http://open-data.europa.eu/ontologies/ec-odp#>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
```

SPARQL Query

```
SELECT  ?Graph ?DatasetTitle  ?Publisher ?DatasetURI WHERE { graph
?Graph  {?DatasetURI a <http://www.w3.org/ns/dcat#Dataset>;
dc:publisher ?Publisher; dc:title ?DatasetTitle.
FILTER(regex(?DatasetTitle, "Gas prices for household consumers"
,"i")) }  } LIMIT 10
```

Format

HTML

Limit results

10

Run query *Get query URL*

Sample queries

'Retrieve dataset with specific title (eg.'Gas prices for household consumers')

'Retrieve number of datasets per publisher'

'Retrieve all publishers'

Retrieve all datasets that have been modified after a certain date

Retrieve all the resources from a dataset with a title that contains specific words (eg.'Gas prices for household consumers')

Further reading

'Introducing Linked Data and the Semantic Web'
A step-by-step introduction

'Linked Open Data – The essentials'
A high-level overview for policy makers

'Bathing water quality'
A case study from data.gov.uk

Guides and tutorials to more Linked Data

Graph	DatasetTitle	Publisher	DatasetURI
http://open-data.europa.eu/data/dataset/ChAwt8m5GkLf8YOPtai7OA	"Gas prices for household consumers"@en	http://publications.europa.eu/resource/authority/corporate-body/ESTAT	http://ec.europa.eu/eurostat/product?code=ten00113

SPARQL Example – EU ODP (2)

SPARQL

You can search for the metadata stored in the EU Open Data Portal triple store by using the SPARQL endpoint query editor below.

Namespaces

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX odp:  <http://open-data.europa.eu/ontologies/ec-odp#>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
```

SPARQL Query

```
SELECT  distinct (?Publisher) WHERE { graph ?g {?DatasetURI a
dcat:Dataset; dc:publisher ?Publisher}  } ORDER BY (?Publisher) LIMIT
10
```

Format

HTML

Limit results

10

Run query *Get query URL*

Sample queries

'Retrieve dataset with specific title (eg.'Gas prices for household consumers')

'Retrieve number of datasets per publisher'

'Retrieve all publishers'

Retrieve all datasets that have been modified after a certain date

Retrieve all the resources from a dataset with a title that contains specific words (eg.'Gas prices for household consumers')

Further reading

'Introducing Linked Data and the Semantic Web'
A step-by-step introduction

'Linked Open Data – The essentials'
A high-level overview for policy makers

'Bathing water quality'
A case study from data.gov.uk

Guides and tutorials to more Linked Data guides

SPARQL Example – EU ODP (2)

Publisher
http://publications.europa.eu/resource/authority/corporate-body/BUDG
http://publications.europa.eu/resource/authority/corporate-body/CFCA
http://publications.europa.eu/resource/authority/corporate-body/CNECT
http://publications.europa.eu/resource/authority/corporate-body/COMP
http://publications.europa.eu/resource/authority/corporate-body/COR
http://publications.europa.eu/resource/authority/corporate-body/DEVCO
http://publications.europa.eu/resource/authority/corporate-body/DGT
http://publications.europa.eu/resource/authority/corporate-body/EAC
http://publications.europa.eu/resource/authority/corporate-body/EASO
http://publications.europa.eu/resource/authority/corporate-body/EBA
http://publications.europa.eu/resource/authority/corporate-body/ECFIN
http://publications.europa.eu/resource/authority/corporate-body/ECHA
http://publications.europa.eu/resource/authority/corporate-body/EEA
http://publications.europa.eu/resource/authority/corporate-body/EIB
http://publications.europa.eu/resource/authority/corporate-body/EIGE
http://publications.europa.eu/resource/authority/corporate-body/EIT
http://publications.europa.eu/resource/authority/corporate-body/ELARG
http://publications.europa.eu/resource/authority/corporate-body/EMPL

SPARQL Update

Can be used for…

- Adding data (INSERT)
- Deleting data (DELETE)
- Loading RDF Graph (LOAD / LOAD .. INTO)
- Clearing an RDF Graph (CLEAR GRAPH)
- Creating RDF Graphs (CREATE GRAPH)
- Removing RDF Graphs (DROP GRAPH)
- Copying RDF Graphs (COPY GRAPH … TO GRAPH)
- Moving RDF Graphs (MOVE GRAPH … TO GRAPH)
- Adding RDF Graphs (ADD GRAPH TO GRAPH)

See also:
http://www.euclid-project.eu/modules/chapter2
http://www.w3.org/TR/sparql11-update/

Summary

- RDF is a general way to express **data** intended for publishing on the **Web**.
- RDF data is expressed in **triples**: subject, predicate, object.
- **Different syntaxes** exist for expressing data in RDF.
- SPARQL is a standardised language to **query** graph data expressed as RDF.
- SPARQL can be used to **query** and **update** RDF data.

Group questions

Go to the SPARQL end-point of the EU Open Data Portal and

(1) Query for all the publishers

(2) Query for the number of datasets per publisher

(3) Query for all the datasets of DG SANCO

https://open-data.europa.eu/en/linked-data

Thank you!

...and now YOUR questions?

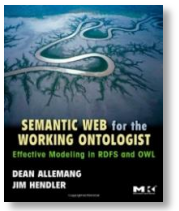
References

- Semantic Web Stack. W3C. http://www.w3.org/DesignIssues/diagrams/sweb-stack/2006a.png
- Resource Description Framework. W3C. http://www.w3.org/RDF/
- Linked Data Cookbook. W3C. http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook
- Cookbook for translating data models to RDF schemas. ISA Programme. https://joinup.ec.europa.eu/community/semic/document/cookbook-translating-data-models-rdf-schemas
- Common Vocabularies / Ontologies / Micromodels. W3C. http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/CommonVocabularies
- SPARQL Query Language for RDF. W3C. http://www.w3.org/TR/rdf-sparql-query/
- Module 2: Querying Linked Data. EUCLID. http://www.euclid-project.eu/modules/course2
- Module 2: Querying Linked Data. EUCLID. http://www.euclid-project.eu/modules/course2
- SPARQL 1.1 Update. W3C.. http://www.w3.org/TR/sparql11-update/

Further reading

Learning SPARQL. Bob DuCharme.

http://www.learningsparql.com/



Semantic Web for the working ontologist. Dean Allemang, Jim Hendler.

http://workingontologist.org/

EUCLID
EdUcational Curriculum for the usage of LInked Data

EUCLID - Course 2: Querying Linked Data

http://www.euclid-project.eu/modules/course2

Related projects and initiatives

Joinup, https://joinup.ec.europa.eu/

Linked Open Vocabularies, http://okfn.org/

W3C GLD WG, http://www.w3.org/2011/gld/wiki/Main_Page
W3C Schools – Learn RDF
http://www.w3schools.com/rdf/default.asp

EUCLID, http://euclid-project.eu/

TopBraid Composer

Protégé Ontology Editor , http://protege.stanford.edu/

XML Summer School http://xmlsummerschool.com/

Be part of our team…

Find us on

Open Data Support
http://www.slideshare.net/OpenDataSupport

Open Data Support
http://goo.gl/y9ZZI

Join us on

http://www.opendatasupport.eu

Follow us

@OpenDataSupport

Contact us

contact@opendatasupport.eu