European Commission

OPEN DATA SUPPORT

Training Module 2.4

Designing and developing RDF vocabularies

pwc

PwC firms help organisations and individuals create the value they're looking for. We're a network of firms in 158 countries with close to 180,000 people who are committed to delivering quality in assurance, tax and advisory services. Tell us what matters to you and find out more by visiting us at www.pwc.com.
PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details.

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).

© 2014 European Commission

This presentation has been created by PwC

Authors:
Nikolaos Loutas, Michiel De Keyzer 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:

- What the best practices are for creating an RDF vocabulary for modelling your data
- Where to find RDF vocabularies for reuse.
- How you can create your own RDF vocabulary.
- How to publish your RDF vocabulary.
- The process and methodology for developing semantic agreements developed by the ISA Programme of the European Commission.

CC BY

Content

This module is about…

- The steps for modelling your data.
- How to reuse existing vocabularies to model your data.
- How to create new classes and properties in RDF.
- How and where to publish your RDF vocabulary so that it can be reused by others.

CC BY

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)

CC BY

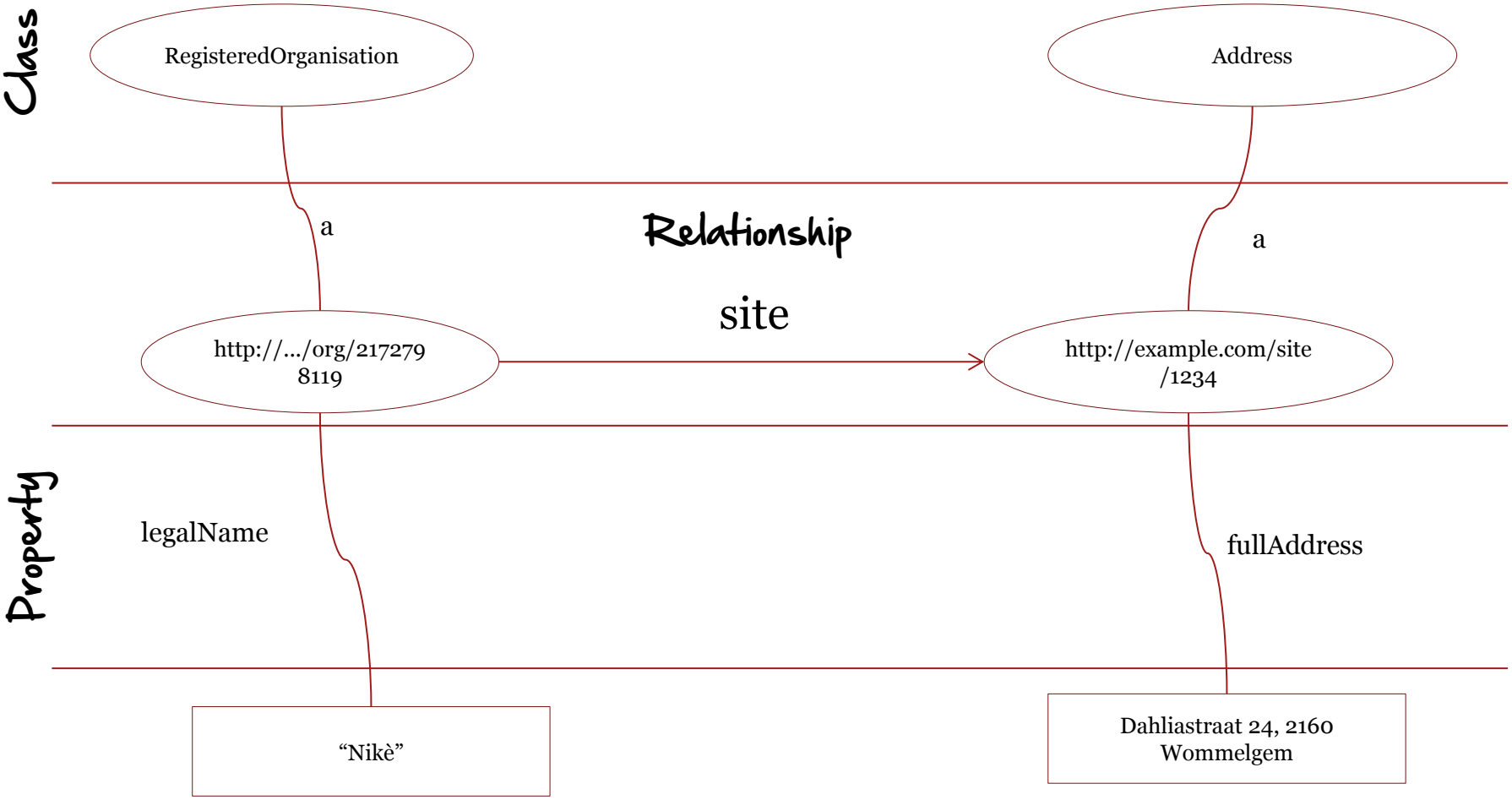
What are classes, relationships and properties?

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



Model your data

How to reuse from other vocabularies, define your own terms and publish and promote your vocabularies to describe the data.

CC BY

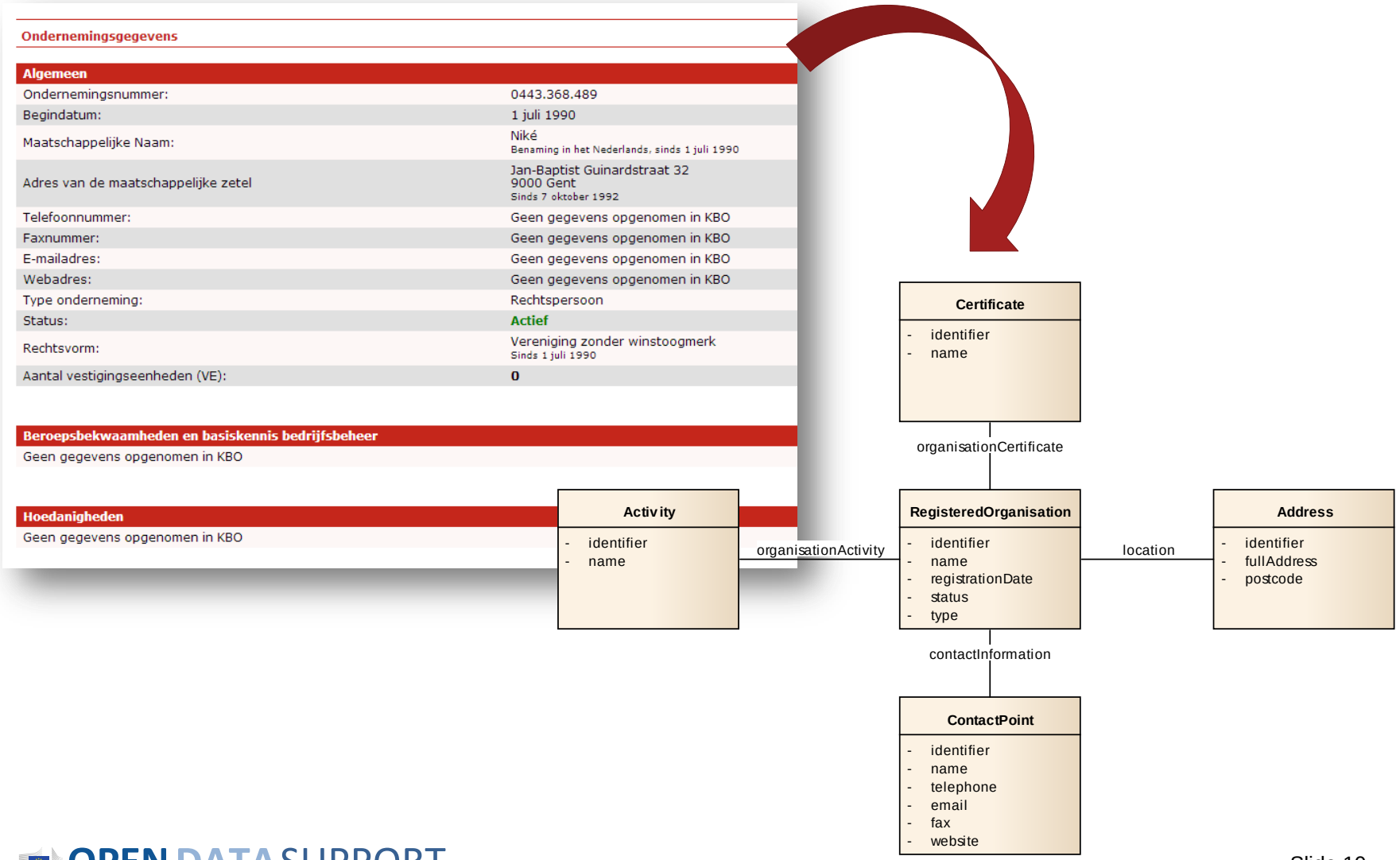
6 steps for modelling your data

1. Start with a **robust Domain Model** developed following a structured process and methodology.
2. Research **existing terms** and their usage and **maximise reuse** of those terms.
3. Where new terms can be seen as specialisations of existing terms, create **sub class** and **sub properties**.
4. Where **new terms** are required, create them following **commonly agreed best practice**.
5. Publish within a **highly stable environment** designed to be **persistent**.
6. **Publicise the RDF schema** by registering it with relevant services.

See also:
https://joinup.ec.europa.eu/community/semic/document/cookbook-translating-data-models-rdf-schemas

CC BY

1 Start with a robust Domain Model



Ondernemingsgegevens

Algemeen	
Ondernemingsnummer:	0443.368.489
Begindatum:	1 juli 1990
Maatschappelijke Naam:	Niké Benaming in het Nederlands, sinds 1 juli 1990
Adres van de maatschappelijke zetel	Jan-Baptist Guinardstraat 32 9000 Gent Sinds 7 oktober 1992
Telefoonnummer:	Geen gegevens opgenomen in KBO
Faxnummer:	Geen gegevens opgenomen in KBO
E-mailadres:	Geen gegevens opgenomen in KBO
Webadres:	Geen gegevens opgenomen in KBO
Type onderneming:	Rechtspersoon
Status:	Actief
Rechtsvorm:	Vereniging zonder winstoogmerk Sinds 1 juli 1990
Aantal vestigingseenheden (VE):	0

Beroepsbekwaamheden en basiskennis bedrijfsbeheer

Geen gegevens opgenomen in KBO

Hoedanigheden

Geen gegevens opgenomen in KBO

CC BY

2 *Reuse existing terms and vocabularies*

- General purpose vocabularies: DCMI, RDFS
- To name things: rdfs:label, foaf:name, skos:prefLabel
- To describe people: FOAF, vCard, Core Person Vocabulary
- To describe projects: DOAP, ADMS.SW
- To describe interoperability assets: ADMS
- To describe registered organisations: Registered Organisation Vocabulary
- To describe addresses: vCard, Core Location Vocabulary
- To describe public services: Core Public Service Vocabulary
- To describe datasets: DCAT, DCAT Application Profile, VoID

CC BY

Creating application profiles

- **Different domains** have **different requirements** for domain-specific semantics, e.g. classification concepts.
- Generic RDF vocabularies usually provides the **core base classes** needed to allow extensions to add specific sub-class structures or classification schemes as required.
- In such cases, reusers are encouraged to define **application profiles particular to an application domain** by specifying (if required) **sub-classes**, **sub-properties** and **controlled vocabularies**.
- For example,
  - DCAT Application profile for data portals in Europe
  - Registered Organization vocabulary as an application profile of the Organization ontology.

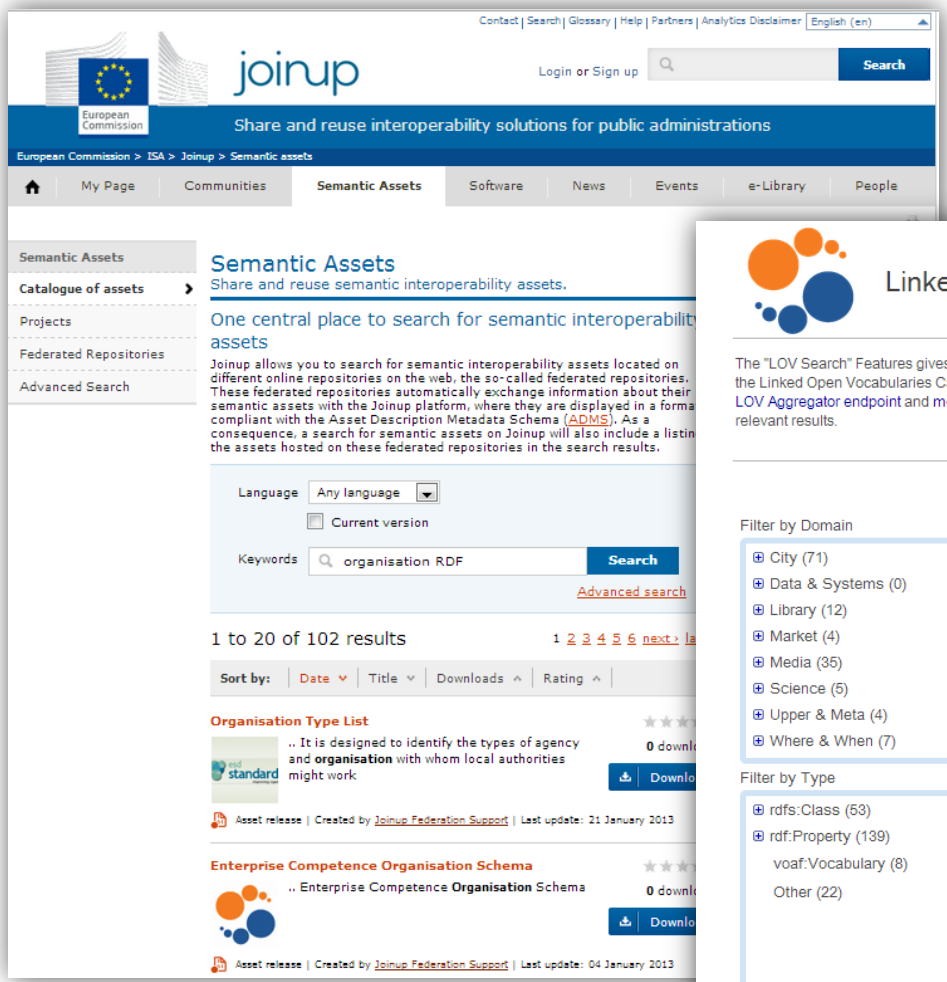
See also:
joinup.ec.europa.eu/asset/dcat_application_profile/home

CC BY

Advantages of reuse

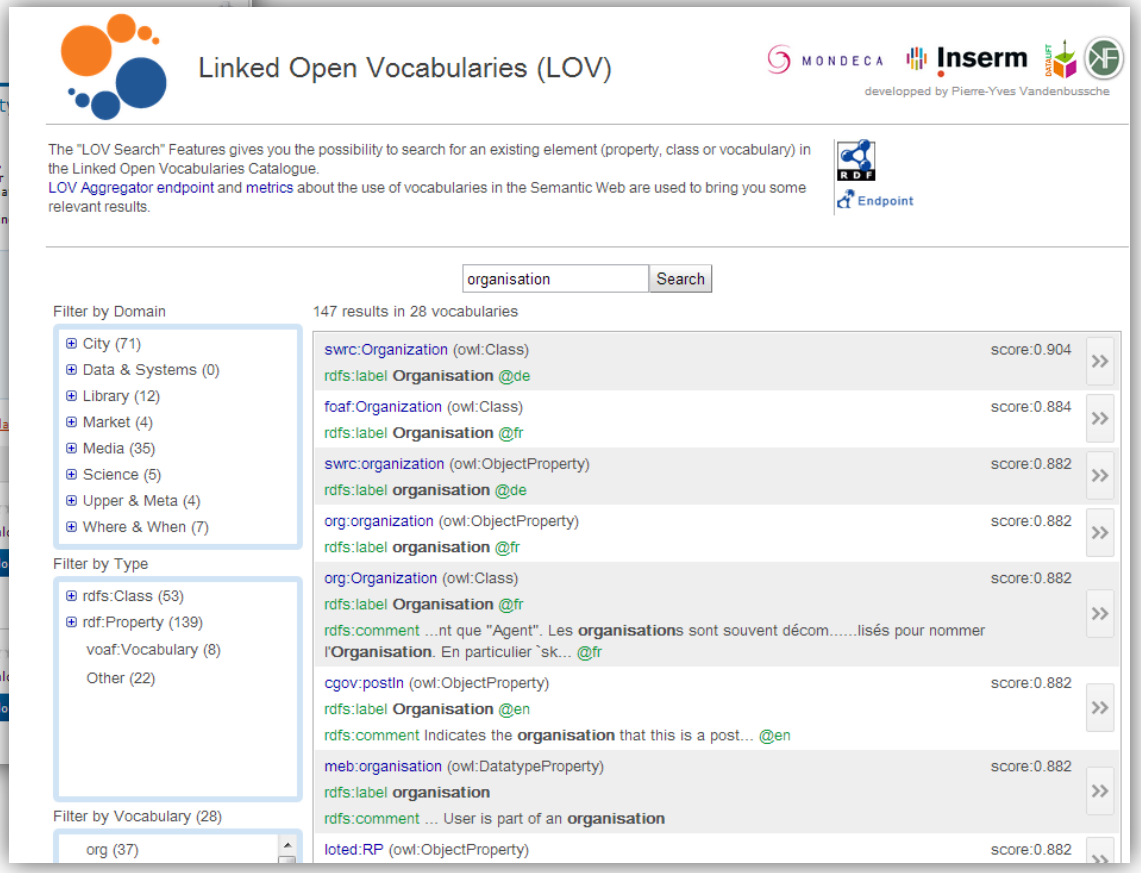
- Reuse greatly **aids interoperability** of your data
  - Use of dcterms:created, for example, the value for which should be a data typed date such as 2013-02-21^^xsd:date, is immediately processable by many machines. If your schema encourages data publishers to use a different term and date format, 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.

You can find reusable RDF vocabularies on…



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

http://lov.okfn.org/



OPEN DATA SUPPORT

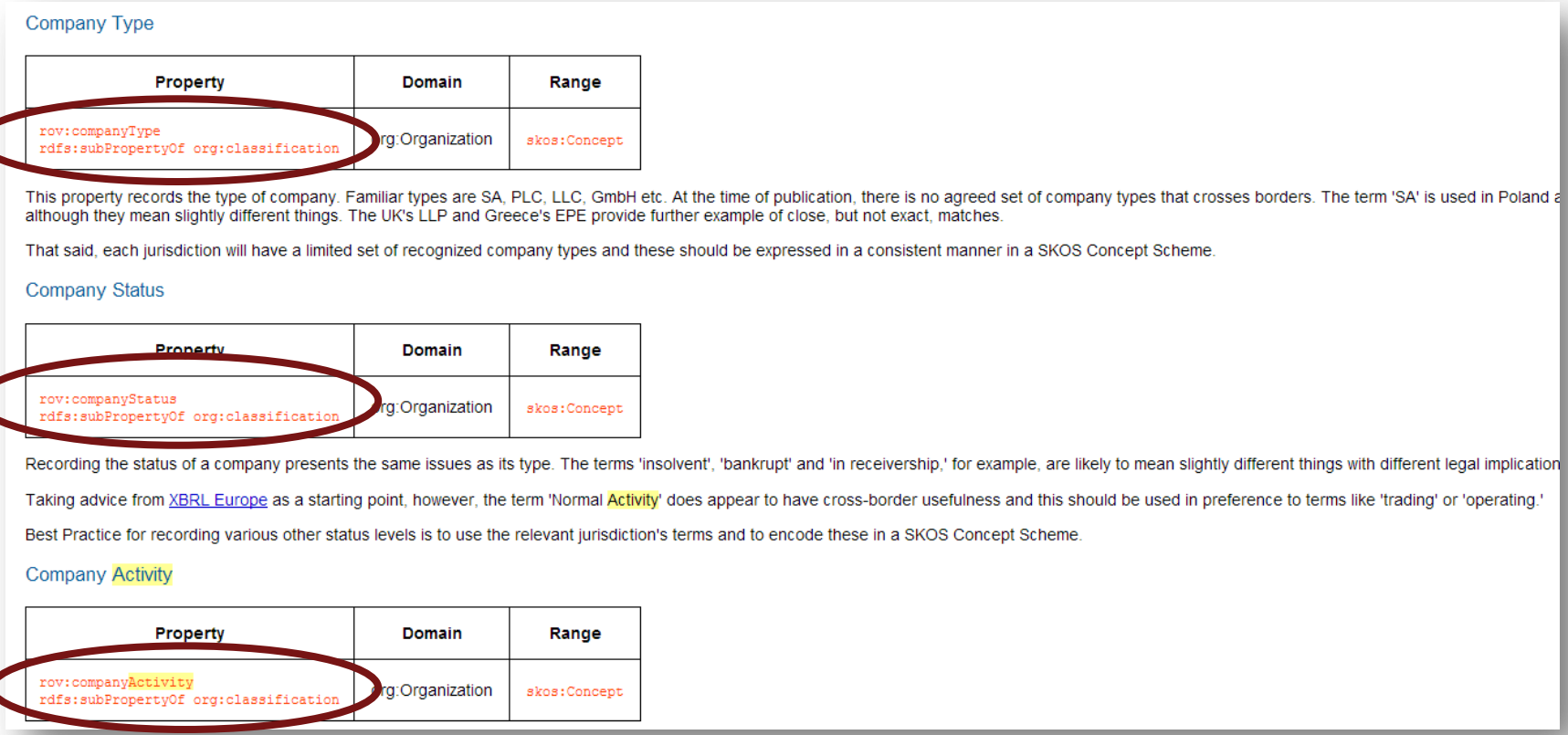
CC BY

3 *Creation of sub-classes and sub-properties*

- RDF schemas and vocabularies often include **terms that are very generic**.
- By creating **sub-class** and **sub-property** relationships, systems that understand the super property or super class may be able to interpret the data even if the more specific terms are unknown.
- Do **not create sub-classes and sub-properties simply to allow you to use your own term** for something that already exists.

Creation of sub-properties - example

The Registered Organization vocabulary defines three sub-properties of org:classification: companyType, companyStatus and companyActivity.



Company Type

Property	Domain	Range
rov:companyType rdfs:subPropertyOf org:classification	org:Organization	skos:Concept

This property records the type of company. Familiar types are SA, PLC, LLC, GmbH etc. At the time of publication, there is no agreed set of company types that crosses borders. The term 'SA' is used in Poland a
although they mean slightly different things. The UK's LLP and Greece's EPE provide further example of close, but not exact, matches.

That said, each jurisdiction will have a limited set of recognized company types and these should be expressed in a consistent manner in a SKOS Concept Scheme.

Company Status

Property	Domain	Range
rov:companyStatus rdfs:subPropertyOf org:classification	org:Organization	skos:Concept

Recording the status of a company presents the same issues as its type. The terms 'insolvent', 'bankrupt' and 'in receivership,' for example, are likely to mean slightly different things with different legal implication

Taking advice from XBRL Europe as a starting point, however, the term 'Normal Activity' does appear to have cross-border usefulness and this should be used in preference to terms like 'trading' or 'operating.'

Best Practice for recording various other status levels is to use the relevant jurisdiction's terms and to encode these in a SKOS Concept Scheme.

Company Activity

Property	Domain	Range
rov:companyActivity rdfs:subPropertyOf org:classification	org:Organization	skos:Concept

CC BY

Defining a sub-property in RDF

```
<rdf:Property rdf:about="rov:companyType">
        <rdfs:label xml:lang="en">company type</rdfs:label>
        <rdfs:comment xml:lang="en" rdf:parseType="Literal">
          This property records the type of company. Familiar types are SA, PLC, LLC, GmbH etc. Each jurisdiction will have a limited set of recognised company types and these should be used in a consistent manner  using a skos:Concept as described in the <a href="#skos:Concept">Code</a> Class.
        </rdfs:comment>
        <rdfs:isDefinedBy rdf:resource="http://www.w3.org/ns/regorg#"/>
        <rdfs:range rdf:resource="skos:Concept"/>
        <rdfs:subPropertyOf rdf:resource="org:classification" />
        <dcterms:identifier>legal:companyType</dcterms:identifier>
</rdf:Property>
```


CC BY

4 *Where new terms are required, create them following commonly agreed best practices*

- ✓ Classes begin with a capital letter and are always singular, e.g. skos:Concept.
- ✓ Properties begin with a lower case letter, e.g. rdfs:label.
- ✓ Object properties should be verbs, e.g. org:hasSite.
- ✓ Data type properties should be nouns, e.g. dcterms:description.
- ✓ Use camel case if a term has more than one word, e.g. foaf:isPrimaryTopicOf.

Defining a new class - Organisation

```
<rdf:RDF
        xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
        xmlns:org="htpp://example.org/organisation-schema">
<rdf:Class rdf:about="org:Organisation">
        <rdfs:label xml:lang="en">Organisation</rdfs:label>
        <rdfs:comment xml:lang:"en">
        Legal entity that is registered in an official national or regional register.
        </rdfs:comment>
</rdf:Class>
```


CC BY

Defining a new property - registrationNumber

<rdf:RDF

xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"

xmlns:org="htpp://example.org/organisation-schema">

<rdf:Property rdf:about="org:registrationNumber">

<rdfs:label xml:lang="en">registrationNumber</rdfs:label>

<rdfs:comment xml:lang:"en">

The number that a registered organisation receives upon registration in the official register.

</rdfs:comment>

</rdf:Class>

CC BY

Defining domain and range restrictions

```
<rdf:RDF
        xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
        xmlns:org="htpp://example.org/organisation-schema"
        xmlns:locn="http://www.w3.org/ns/locn#">
<rdf:Property rdf:about="org:isLocated">
        <rdfs:label xml:lang="en">isLocated</rdfs:label>
        <rdfs:comment xml:lang:"en">
        The official address of the registered organisation's headquarters.
        </rdfs:comment>
        <rdfs:domain rdf:resource="org:Organisation"/>
        <rdfs:range rdf:resource="locn:Address">
</rdf:Class>
```

http://example.org/org/1234	org:isLocated	http://dbpedia.org/page/Brussels

CC BY

5 *Publish within a highly stable environment designed to be persistent*

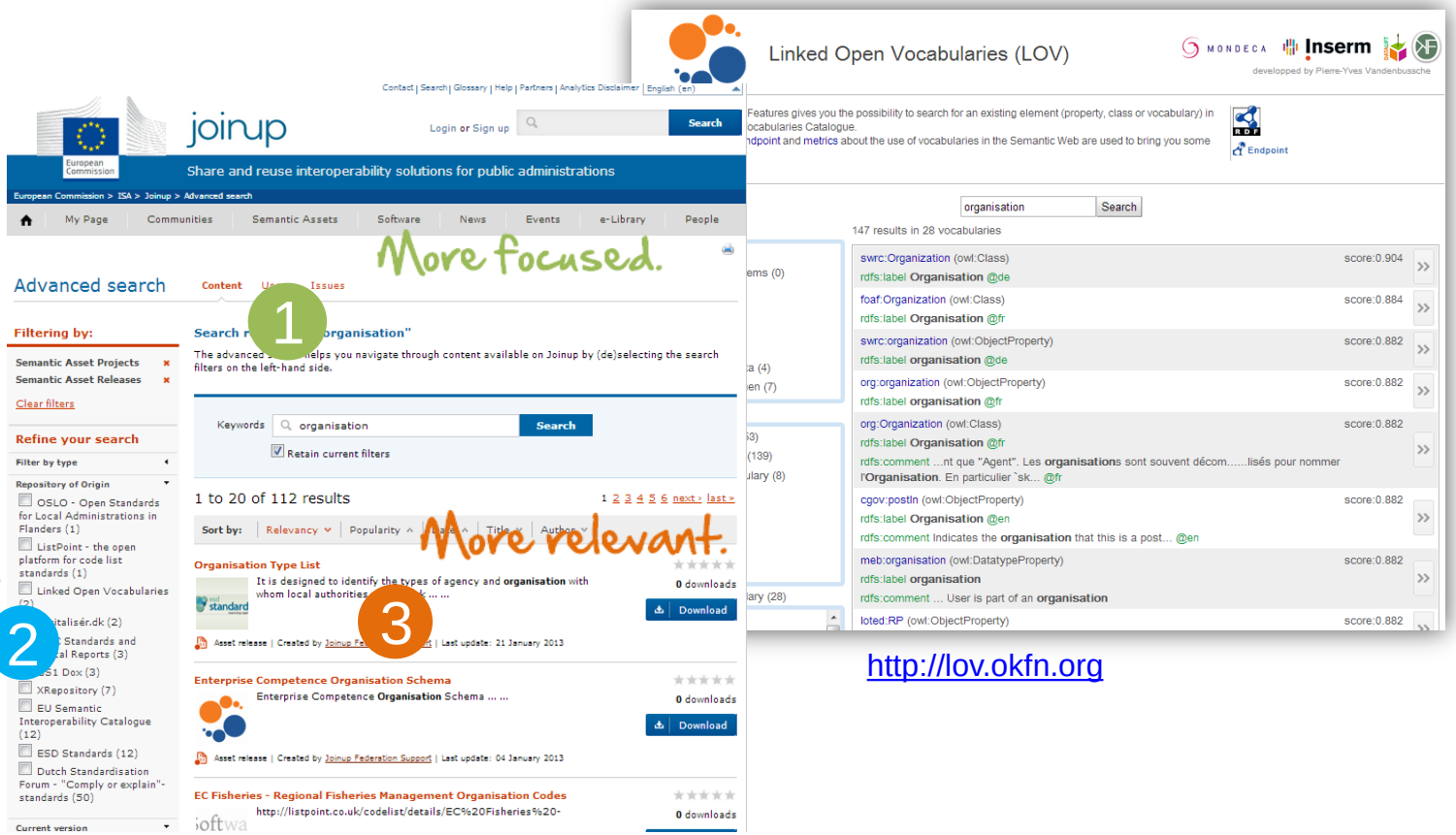
- Choose a stable namespace for your RDF schema (e.g. W3C, Purl…)
- Use good practices on the publication of persistent Uniform Resource Identifiers (URI) sets, both in terms of format and of their design rules and management.
- Examples:
  - http://www.w3.org/ns/adms
  - http://purl.org/dc/elements/1.1

See also:
https://joinup.ec.europa.eu/community/semic/document/cookbook-translating-data-models-rdf-schemas
http://www.slideshare.net/OpenDataSupport/design-and-manage-persitent-uris

CC BY

6 *Publicise the RDF schema by registering it with relevant services*

Once your RDF schema is published you will want people to know about it. To reach a wider audience register it Joinup and Linked Open Vocabularies.



Conclusions

Analyse

- Start with a robust Domain Model developed following a structured process and methodology.
- Research existing terms and their usage and maximise reuse of those terms.

Model

- Where new terms can be seen as specialisations of existing terms, create sub class and sub properties as appropriate.
- Where new terms are required, create them following commonly agreed best practice in terms of naming conventions etc

Publish

- Publish within a highly stable environment designed to be persistent.
- Publicise the RDF schema by registering it with relevant services.

CC BY

Group exercise

http://www.visualpharm.com

In groups of 2, create the RDF description of a vocabulary for representing a citizen.

http://www.visualpharm.com

According to you, what are the main barriers to the reuse of existing RDF vocabularies?

Take also the online test here!

CC BY

Thank you!

...and now YOUR questions?

CC BY

References

Slides 9:

- Linked Data Cookbook. W3C. http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

Slide 10-23:

- ISA Programme. Cookbook for translating Data Models to RDF Schemas. https://joinup.ec.europa.eu/community/semic/document/cookbook-translating-data-models-rdf-schemas

Slide 16, 18,-21:

- W3C. An organization ontology. http://www.w3.org/TR/vocab-org/

Slide 23:

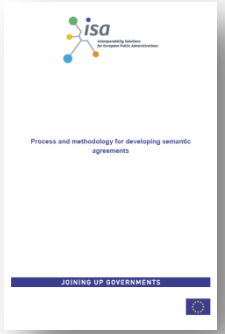
- ADMS Brochure. ISA Programme. https://joinup.ec.europa.eu/elibrary/document/adms-brochure

CC BY

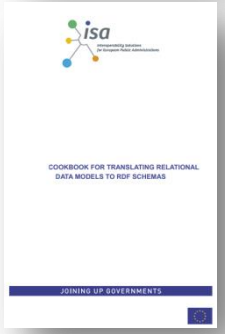
Further reading

Linked Data Cookbook, W3C Government Linked Data Working Group,

http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook



EC, ISA Process and methodology for developing semantic agreements,
https://joinup.ec.europa.eu/community/core_vocabularies/document/process-and-methodology-developing-semantic-agreements



EC ISA, Cookbook for translating Data Models to RDF Schemas
https://joinup.ec.europa.eu/community/semic/document/cookbook-translating-data-models-rdf-schemas

Related projects and initiatives

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

Linked Open Vocabularies (LOV), http://lov.okfn.org/

EC ISA, e-Government Core Vocabularies,
https://joinup.ec.europa.eu/community/core_vocabularies/home

W3C Schools – Learn RDF
http://www.w3schools.com/rdf/default.asp

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

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

CC BY