



Semantic Interoperability Courses

Course Module 2Core Vocabularies





Disclaimer

This training material was prepared for the ISA programme of the European Commission by PwC EU Services.

The views expressed in this training material 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 study, 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.







Learning Objectives

- → Understand what Core Vocabularies are.
- → Understand how to extend the Core Vocabularies depending on your patterns of information exchange
- → Understand how to use and extend the Core Vocabularies in your own data models.

Outline



1. What? - Introduction

- Definition: Core Vocabularies
- Levels of abstraction: core, domain, information exchange
- Overview of existing Core Vocabularies
- Process and methodology for developing Core Vocabularies

2. Why use the Core Vocabularies?

- Attain a minimum level of cross-domain semantic interoperability
- Patterns of information exchange
- Contexts of use

3. How to use the Core Vocabularies: Linked Data

- Extending the Core Vocabularies to publish Linked Data
- Best practices for publishing Linked Data
- Example: Describe organisations in RDF using standard Vocabularies

4. How to create e-Document formats using the Core Vocabularies

- Guidelines for e-Document engineering using the Core Vocabularies
- Example: Business Activity Registration



Business need

The need for harmonising data models.

The exchange of information in the context of European Public Services is challenging and comes with many **semantic interoperability conflicts**.

Such interoperability conflicts are caused by discrepancies in the interpretation of administrative procedures and legislation, the lack of commonly agreed data models, the absence of universal reference data, etc.



Definition

What is a Core Vocabulary?



Simplified, re-usable, and extensible data models that capture the fundamental characteristics of a data entity in a context-neutral fashion.









European Commission – ISA Programme, 2014 (1) Source: https://joinup.ec.europa.eu/node/43160



Four Core Vocabularies



Fundamental characteristics of a person.



Fundamental characteristics of a legal entity, such as legal identifier, name, company type, activities.



Fundamental characteristics of a location, represented as an address, a geographic name, or a geometry.



Fundamental characteristics of a public service.



Three representation techniques

The same meaning expressed in UML, RDFS, and XSD.



Conceptual model

Reuse existing concepts in CCL, INSPIRE, etc.



RDF schema

Reuses existing RDF vocabularies



XML schema

Reuses Core Components Technical Specification (CCTS) and UBL NDR



ISA Open Metadata Licence v1.1



Registered Organisations Vocabulary Maintained by W3C (Government Linked Data Working Group)



Developed according to an open and inclusive process

Process for developing Core Vocabularies

1. Identify stakeholders

2. Form Working Group

3. Identify chair/cochairs

4. Identify editor(s)

5. Form Review Group

6. Secure **IPR**

7. Establish a working environment and culture

8. Publish drafts

9. Process **comments**

10. Publish the Latest
Call Working Draft
of the vocabulary and
contact the Review
Group, seeking its
feedback

11. Review Last Call Working Draft

12. Gather **evidence of acceptance** by the potential users of the vocabulary

13. The EC to submit documents to the TIE Cluster or ISA Coordination Group

Download the process and methodology for developing Core Vocabularies here:

https://joinup.ec.europa.eu/node/43160



Developed according to an open and inclusive process

Methodology for developing Core Vocabularies (1)

- 1. Identify the Core Vocabularies likely to **meet the needs** of potential users within European institutions
- 2. Working Group to research **existing vocabularies** (provenance, usage, stability)
- 3. Research existing published data and services, avoiding any conflicts with proposed Core Vocabulary
- 4. Articulate the **problem(s)** that the WG is trying to solve in the form of a series of use cases

- 5. Derive a set of **requirements** from the use cases
- Publish the use cases and requirements in a single document
- 7. Create a **concept diagram** (UML)
- 8. Do **not impose** cardinality **rules** or domain/range **restrictions** on vocabulary terms unless necessary

- Use words beginning with an upper or lower case letter or an underscore for all terms in a vocabulary
- 10. Use **simple nouns** for property names
- 11. Use **verbs** for relationship terms
- 12. For each relationship, include a **definition** of its **inverse**

Download the process and methodology for developing Core Vocabularies here:

https://joinup.ec.europa.eu/node/43160



Developed according to an open and inclusive process

Methodology for developing Core Vocabularies (2)

13. Use **prepositions** in vocabulary terms only if necessary

14. Use a namespace ending with a **hash character** (#)

15. Keep the namespace as **short** as possible

16. Include a portion that identifies the vocabulary for human readers

17. Do not include any technologyspecific component in the namespace (except HTTP) 18. Do **not restrict** pool of potential users by using a namespace declaring **'ownership'** or **geographical** relevance

19. If necessary, consider meeting *step* 18 using **PURLs**

20. Create and validate the namespace documents in HTML, XML and RDF/XML.

21. Either the WG or the EC must make each one available through {namespace}.ext

22. Either the WG or the EC must set up **content negotiation** to handle requests to the namespace itself 23. When publishing the final version of the Core Vocabulary, **link** the HTML document to an **errata** document

24. Include a Conformance Statement

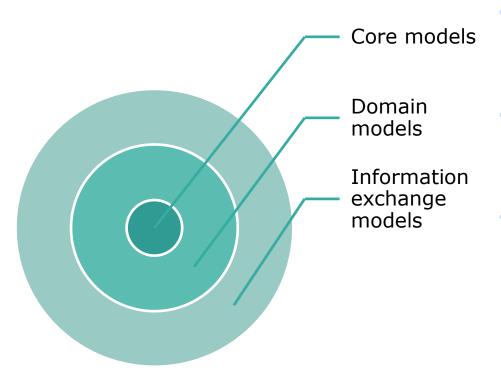
Download the process and methodology for developing Core Vocabularies here:

https://joinup.ec.europa.eu/node/43160



Reuse by extension

Extend the Core Vocabularies into domain models and information exchange models



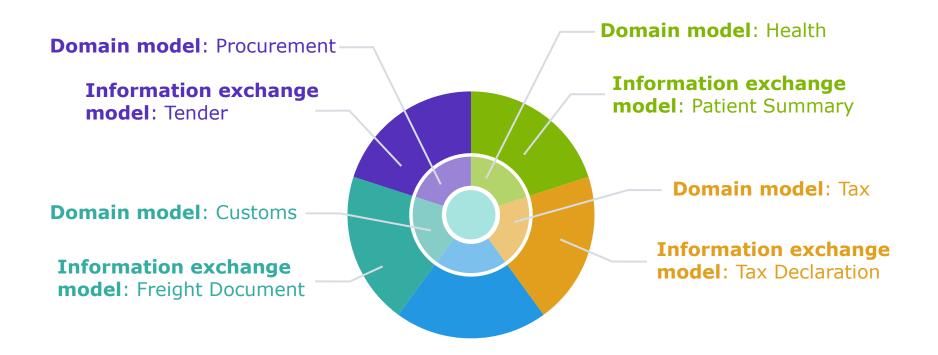
- Core model: a context-neutral data model that captures the fundamental characteristics of an entity.
- Domain model: a conceptual view of a domain that identifies the entities involved and their relationships
- a model that defines and describes the structure and content of a specific information exchange context.

SA 12



Reuse by extension

Extend the Core Vocabularies into domain models and specifications for information exchange

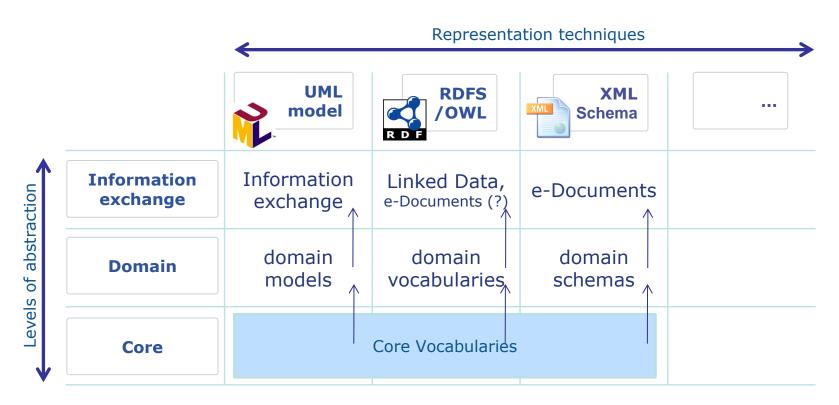


15A 13



Reuse by extension

Extend the Core Vocabularies into domain models and specifications for information exchange



Outline



1. What? - Introduction

- Definition: Core Vocabularies
- Levels of abstraction: core, domain, information exchange
- Overview of existing Core Vocabularies
- Process and methodology for developing Core Vocabularies

2. Why use the Core Vocabularies?

- Attain a minimum level of cross-domain semantic interoperability
- Patterns of information exchange
- Contexts of use

3. How to use the Core Vocabularies: Linked Data

- Extending the Core Vocabularies to publish Linked Data
- Best practices for publishing Linked Data
- Example: Describe organisations in RDF using standard Vocabularies

4. How to create e-Document formats using the Core Vocabularies

- Guidelines for e-Document engineering using the Core Vocabularies
- Example: Business Activity Registration



Why use the Core Vocabularies?

To attain a minimum level of cross-domain semantic interoperability

- The compliance of characteristics to Core Vocabulary specifications guarantees a minimum of cross-domain interoperability, while providing domain-specific communities
- Core Vocabularies offer freedom and a common starting point for drafting specializations of one's own by adding metadata to the Core
- Increase the possibilities for reuse
- Avoid schema-level conflicts, which are caused by a different logical structure or inconsistencies in metadata

Source: <u>eGovernment Core Vocabularies by Vassilios Peristeras</u>, February 2012

ISA



Why use the Core Vocabularies?

To avoid schema-level conflicts

Examples of schema-level conflicts:

Naming

 Citizen information is verified against the wrong source*

Entity identifier

 Citizens identified by ID card number or national number or none?

Schema isomorphism

 Different attributes on ID cards in different states

Generalization

 Birth certificate of one state can contain all info of birth and family certificates of another state

Aggregation

"full name" or "surname"; "middle name"; "first name"

Schematic discrepancies

 Detailed Information cannot be exchanged due to schematic differences (ex. different xml schemas)

More on semantic conflicts: V. Peristeras - <u>A conceptual analysis of semantic</u> conflicts in pan-European e-government services

^{*} Naming conflicts: evidence placeholders with the same name but different purpose and usage may exist in different Member States, or evidence placeholders with different names may have similar usage and hold similar evidence items.



Contexts in which the Core Vocabularies can be used

- **Development of new systems**: the Core Vocabularies can be used as a default starting point for designing the conceptual and logical data models in newly developed information systems.
- Information exchange between systems: the Core Vocabularies can become the basis of a context specific data model used to exchange data among existing information systems.
- Data integration: the Core Vocabularies can be used to integrate data that comes from disparate data sources and create a data mesh-up.
- **Open data publishing**: the Core Vocabularies can be used as the foundation of a common export format for data in base registries like cadastres, business registers and service portals.

Outline



1. What? - Introduction

- Definition: Core Vocabularies
- Levels of abstraction: core, domain, information exchange
- Overview of existing Core Vocabularies
- Process and methodology for developing Core Vocabularies

2. Why use the Core Vocabularies?

- Attain a minimum level of cross-domain semantic interoperability
- Patterns of information exchange
- Contexts of use

3. How to use the Core Vocabularies: Linked Data

- Extending the Core Vocabularies to publish Linked Data
- Best practices for publishing Linked Data
- Example: Describe organisations in RDF using standard Vocabularies

4. How to create e-Document formats using the Core Vocabularies

- Guidelines for e-Document engineering using the Core Vocabularies
- Example: Business Activity Registration



What is linked data?

Linked data is a set of design principles for sharing machinereadable data on the Web

The **four design principles** of Linked Data (by Tim Berners Lee):

- 1. Use Uniform Resource Identifiers (URIs) as names for things.
- 2. Use HTTP URIs so that people can look up those names.
- 3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
- 4. Include links to other URIs so that they can discover more things.



Linked Data: Best Practices

Use and extend the Core Vocabularies to publish Linked Data

1. Prepare 4. Specify an 2. Select a dataset 3. Model the data stakeholders appropriate licence 5. Good URIs for 6. Use standard 8. Provide machine 7. Convert data Linked Data access to data vocabularies 9. Announce new 10. Recognize the social contract datasets **eGOVERNMENT VOCABULARIES**

Source: http://www.w3.org/TR/ld-bp/



The Core Vocabularies abide by the Linked Data principles

- ISA's Core Person, Core Location and Core Business Vocabularies have been taken as inputs by the Government Linked Data Working Group (GLD WG) of W3C.
- Core Vocabularies:
 - o promote the use of common identifiers for organisations, people and locations in the form of URIs.
 - can be easily combined with other Linked Data vocabularies.
 - can easily be extended with new classes and attributes to fulfil new domain requirements.

A 22



Example: Describe organisations in RDF using standard Vocabularies

- Organisations can be described in RDF using a combination of the Registered Organization Vocabulary and the more general Organization Ontology.
 - Registered Organization Vocabulary: simplified, reusable and extensible data model; describes organisations that have gained legal entity status through a formal registration process (Registered Organizations - RegORG)
 - Organization Ontology (ORG): describes the several parts of an organisation.
- Case example: PricewaterhouseCoopers Enterprise Advisory is a registered legal entity in the Belgian company register.

Source: https://joinup.ec.europa.eu/node/52998/



Example: Describe organisations in RDF using standard Vocabularies

Registered Organization Vocabulary

- Describe essential elements of a registered organisation
- Data fields in official extracts of business registers
 - the legal name of the organisation
 - the registered number of the organisation
 - the legal address of the organisation
 - o the **activities** for which the organisation is registered for
 - the type of organisation
- Each organisation is identified by a unique URI



Source: https://joinup.ec.europa.eu/node/52998/



Example: Describe organisations in RDF using standard Vocabularies

Registered Organization Vocabulary: PricewaterhouseCoopers Enterprise Advisory

Legal name

<rov:Registeredorganisation
rdf:about="http://kbopub.economie.fgov.be/kbopub/
toonondernemingps.html?ondernemin
gs
nummer=415622333">
<rov:legalName>PricewaterhouseCoopers Enterprise
Advisory</rov:legalName>
</rov:Registeredorganisation>

Registered number

<rov:registration> <adms:Identifier
rdf:about="http://example.com/Reg415622333">
<skos:notation>0415.622.333</skos:notation>
<adms:schemeAgency>Belgian Base Register for
Companies</adms:schemeAgency> </adms:Identifier>
</rov:registration>

Type

<rov:companyType> <skos:Concept
rdf:about="http://example.com/Cooperatievevennoot
schap"> <rdfs:label>Cooperatieve
vennootshap</rdfs:label> </skos:Concept>
</rov:companyType>

Legal address

<rov:registeredAddress> <locn:Address
rdf:about="http://example.com/ra415622333">
<locn:postCode>1932 Zaventem</locn:postCode>
<locn:fullAddress>Belgium, Woluwedal
18</locn:fullAddress> </locn:Address>
</rov:registeredAddress>

Activities for which the company is registered for

```
<skos:Concept rdf:about="http://example.com/ca7022">
<rdfs:label>Business and other management
consultancy activities</rdfs:label> </skos:Concept>
<skos:Concept
rdf:about="http://example.com/ca74142">
<rdfs:label>Other business and management
consultancy activities</rdfs:label> </skos:Concept>
```

Source: https://joinup.ec.europa.eu/node/52998/

Outline



1. What? - Introduction

- Definition: Core Vocabularies
- Levels of abstraction: core, domain, information exchange
- Overview of existing Core Vocabularies
- Process and methodology for developing Core Vocabularies

2. Why use the Core Vocabularies?

- Attain a minimum level of cross-domain semantic interoperability
- Patterns of information exchange
- Contexts of use

3. How to use the Core Vocabularies: Linked Data

- Extending the Core Vocabularies to publish Linked Data
- Best practices for publishing Linked Data
- Example: Describe organisations in RDF using standard Vocabularies

4. **How** to create e-Document formats using the Core Vocabularies

- Guidelines for e-Document engineering using the Core Vocabularies
- Example: Business Activity Registration



e-Document formats

Extending the Core Vocabularies

Definitions:

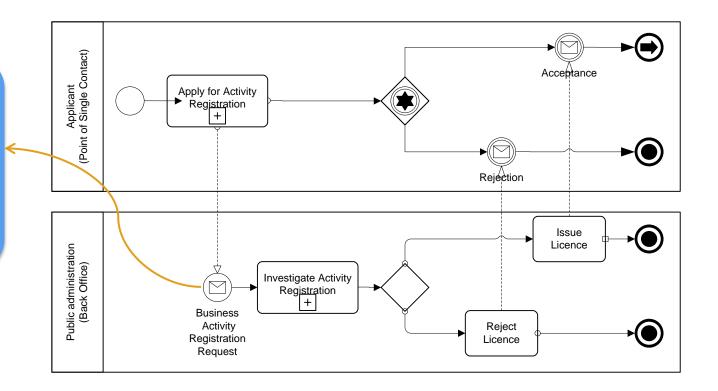
- e-Document: any document in electronic format containing structured data (and possibly also unstructured data) used in the context of an administrative process.
- e-Document format: is a specification that lays down the syntax (structure) and semantics of a particular type of e-Document.

→ The Core Vocabularies can be used as a starting point to define e-Document formats



Example: Business activity registration

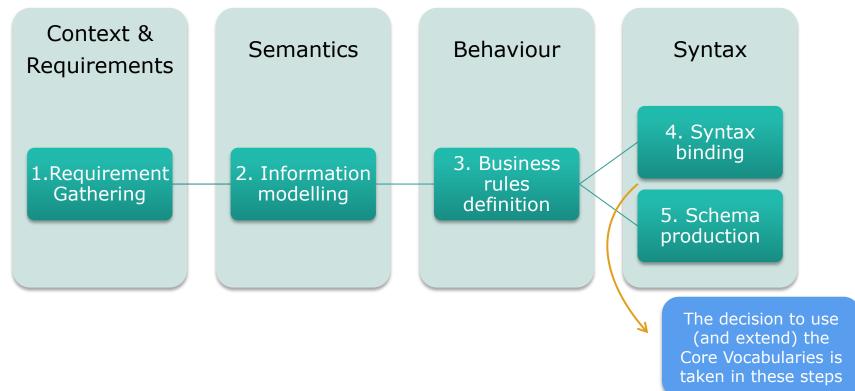
Use the Core
Vocabularies
(i.e. Registered
Organization
Vocabulary) as a
starting point for
the e-Document
'Business Activity
Registration
Request'



SA 28

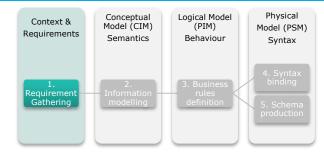


Methodology for e-Document engineering



D1.2 - Guidelines on e-Document engineering for public administrations



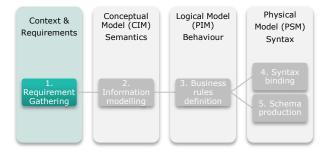


The first step is to precisely define the objective of the business process.

- Goals: describe specific goals to be achieved with the exchange of e-Documents
- Scope: describe the scope derived from the goals
- Key examples: describe key examples as real-life scenarios to depict the business process flow
- Specific requirements: gather specific requirements that e-Documents must fulfil linked to the goals



Example: business activity registration

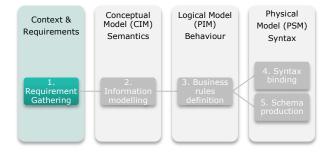


Goals

Goal ID	Goal Name	Goal Description
G1	Improve Business Process Performance	To simplify the business activity registration procedure both for the businesses and competent authorities
G2	Improve Management Efficacy	To harmonize the business activity registration both at European level and at national level.
G3	Decrease Costs and save time	To enable competent authorities to check for validity and suitability of the information and supporting documents submitted by the businesses.
G4	Improve Security	To increase the security and reliability of the business activity registration transactions



Example: business activity registration



Scope

Scope statement

A user accesses a website to get information on the documents that have to be presented in a destination country (being a foreign country or their own) in order to register a business activity. The website system provides the user with information on the documents he has to upload in order to be able to submit the business activity registration request to the destination country. It is outside of the scope the process by which the website system describes the documents to be submitted.

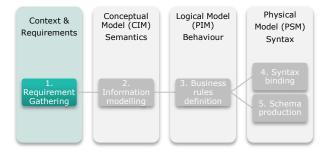
The website collects the electronic unstructured documents and metadata from the business.

The website creates the e-Document with the metadata about the user, the business, the activity and the documents uploaded by the user.

The website submits the e-Document instance to the destination country Point of Single Contact. The Point of Single Contact in the destination country acknowledges the business activity registration request and forwards it to the proper authority for license issuance.



Example: business activity registration

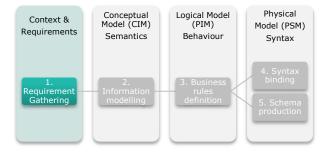


Key example

Key Example Identifier	Key Example Description
	A French business person browses the French PSC looking for registration his business activity in Germany.
	The PSC website offers a page with the possible countries and activities per country he can register.
	The French business person picks up on obtaining a license for opening a store in Germany.
	The PSC website provides detailed information on the documents needed to obtain this license through a form.
KE1	The French business person uses the form to upload the requested documents.
	The PSC website requests the French business person to log in or register in order to get information about his business.
	The PSC website packs all documents and submits that to the German PSC
	The PSC website sends back to the French business person the acknowledgement of reception from the German PSC



Example: business activity registration

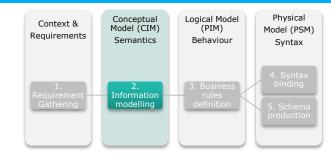


High level requirements

Requirement identifier	Requirement name	Requirement statement	Rationale	Reference to goals
R1	Business information	The business requesting the registration of the activity has to be identified	The receiving PSC needs to know the business requesting the business registration activity to be able to understand which are the documents he has to receive.	G1, G4
R2	Requestor	The person requesting the service on behalf of the business has to be identified	The receiving PSC has to ensure the requestor is capable of requesting the service on behalf of the business.	G4
R3	Business activity	The business activity to be registered has to be identified	The receiving PSC has to know for which business activity the requester is registering for.	G1, G2
R4	The provided documents have Documents to be identified and their purpose has to be described		The receiving PSC has to be able to identify unstructured documents to automate the	G1, G2, G3
R5	Identification	The business request has to be identified	The business request has to be uniquely identifiable, with information about its issuance.	G1, G2, G3



2. Information modelling



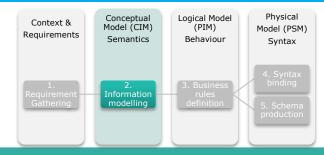
This phase identifies and describes the information to be exchanged in e-Documents according to the requirements specified in the first step.

- Capture business terms in an information model describing the explicit semantics of every data element: attributes and cardinalities
- Describe the relationships between information components and requirements
- Depict information model requirements using a conceptual modelling language (ISO11179 MDR)
- Identify and reuse semantics and concepts from standard vocabularies where possible



2. Information modelling

Example: business activity registration

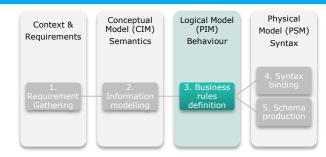


Reference to								
Information Requirement Identifier	Business Term Name	Usage	Business Requirement Identifier	Business Rule Identifier	Cardinality	Concept location	Standard Concept Name	Concept Description
IR4	Business activity	Activity performed by the legal entity, which is requested for registration	R3		11	Registered Organization Vocabulary	Organisation Activity	The activity of an organisation should be recorded using a controlled vocabulary. Several vocabularies exist, many of which map to the UN's ISIC codes. The preferred choice for European interoperability is NACE.
IR5	Business name	Name of the legal entity that is requesting the business activity registration	R1		11	Registered organisation Vocabulary	Legal Name	The legal name of the business. A business might have more than one legal name, particularly in countries with more than one official language.
IR6	Business legal form	Type of the legal entity that is requesting the business activity registration	R1	ıc	11	Registered Organization Vocabulary	Organisation Type	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. Each jurisdiction needs a limited set of recognized company types and these should be expressed in a consistent manner.

A 36



3. Business rules definition



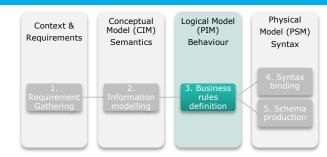
In the previous step (information modelling) the business terms and facts were described. However, there are still action **assertions**, **constraints** and **derivations** concerning some aspects of the e-Document. These business rules are described according to the goals and requirements of the first step.

- Identify integrity constraints on the information model and describe them as business rules
- Define inferences and mathematical calculations that the e-Document elements must fulfil
- Define conditional business rules and co-occurrence constraints that e-Document elements must fulfil
- Define sets of allowed values for coded data elements



3. Business rules definition

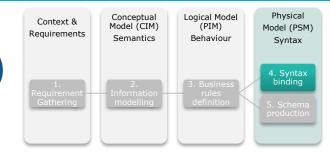
Example: business activity registration



Business Rule ID	Rule	Refer to Information Requirements	Refer to High Level Requirements	Error level
BR1	The business activity must refer to a NACE activity	IR4	R3	Fatal
BR2	The legal form of the business must be recognized by the business' country of origin	IR7	R1	Fatal



4. Syntax binding (reuse)

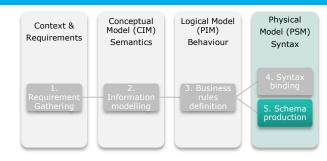


Syntax binding is one of the options to produce physical artefacts in order to help developers implement the e-Documents according to the e-Document format rules. With syntax binding, the information requirement model is mapped to an **existing** syntax model and the usage guidelines are specified.

- Map the information model to a standard syntax when this syntax fulfils most of the goals and requirements of the project
- Create a usage guideline on the syntax for implementers
- Create validation artefacts for business rules and code lists
- List minor gaps and/or requirements that cannot be fulfilled using the selected syntax



5. Schema production (partial reuse)



The second option is to **produce** a **new** e-Document format. This option should be followed when there are no recognized international standards for the industry and business process the project is targeting.

- Map common information model components to available Common Vocabulary schemas (e.g. ISA Core Vocabularies, UBL common library, UN/CEFACT Core Components Library)
- Create new e-Document formats using a standard NDR to automate the schema production
- Create validation artefacts for business rules and code lists



5. Schema production (partial reuse)

Example: business activity registration

XSD Schema: BusinessActivityRegistrationRequest.xsd



Context & Conceptual Model (CIM) Semantics Logical Model (PIM) Behaviour Physical Model (PSM) Syntax

1. Requirement Gathering Information modelling definition F. Schema production



e-Document engineering

Tools

Tool	Description
SemanticMDR	Information Modelling
	Metadata Registry
Metadata Workbench	Schema creation
	UN/CEFACT NDR
Xgenerator	Schema creation
https://joinup.ec.europa.eu/software/xgenerator/description	XÖV NDR / any (configurable)
Crane Software GC-to- UBL NDR script	Schema creation
http://www.cranesoftwrights.com/resources/ubl/index.htm#	OASIS UBL NDR
gc2ublndr	
eDoCreator	Schema creation
http://srdc.com.tr/dist/#/edocreator	OASIS UBL NDR
GEFEG.FX	Information Modelling +Schema creation
http://www.gefeg.com/en/index.htm	CEFACT NDR, OASIS UBL NDR,
Enterprise Architect	Information Modelling +Schema creation
+ ShapeChange	GML NDR







Illustration: Business Activity Request XML Schema

CoreVocabularyBasicComponents.xsd

(namespace prefix: cvc)

CoreVocabularyAggregateComponents.xsd

(namespace prefix: cva)

BusinessActivityRegistrationRequest.xsd

The global elements cvc:LegalName and cva:Cvbusiness can be reused in any schema



References

<u>A conceptual analysis of semantic conflicts in pan-European e-government services</u>, Peristeras, V., Loutas, N., Goudos, S. K., & Tarabanis, K., 2008, Journal of Information Science, 34(6), 877-891.

<u>Enterprise Integration Patterns</u>, Hohpe & Woolf, 2003, Addison-Wesley Professional

Best Practices for Publishing Linked Data, W3C, 2014

<u>Describe organizations in RDF with Core Business Vocabulary and ORG Ontology</u>, November 2012

<u>Process and methodology for Core Vocabularies</u>, ISA, November 2011 <u>eGovernment Core Vocabularies by Vassilios Peristeras</u>, February 2012

A 44





Project Officers Vassilios.Peristeras@ec.europa.eu Suzanne.Wigard@ec.europa.eu Athanasios.Karalopoulos@ec.europa.eu

Visit our initiatives

















Get involved



Follow @SEMICeu on Twitter



Join the **SEMIC** group on LinkedIn



Join the SEMIC community on Joinup