



**ATHENS Course: Ontologies and the Semantic Web**

# **A Method for Reusing and Re-engineering Non-Ontological Resources for Building Ontologies**

Boris Villazón-Terrazas

[bvillazon@fi.upm.es](mailto:bvillazon@fi.upm.es)

Ontology Engineering Group. Laboratorio de Inteligencia Artificial

Departamento de Inteligencia Artificial

Facultad de Informática

Universidad Politécnica de Madrid

- **Introduction and motivation**
- Non-Ontological Resources (NORs)
- Method for Reusing NORs
- Pattern based Re-engineering Method
- Patterns for Re-engineering Classification Schemes
- Patterns for Re-engineering Thesauri
- Patterns for Re-engineering Lexica
- Technological support

- METHONTOLOGY [Gómez-Pérez et al. 2003], On-To-Knowledge [Staab et al. 2001], and DILIGENT [Pinto et al. 2004] provide guidelines to help researchers to develop ontologies from scratch.
- The development of ontologies (Esperonto<sup>1</sup>, Knowledge Web<sup>2</sup>, SEKT<sup>3</sup>, etc.) has disclosed that a new ontology development paradigm is starting, whose emphasis is on the **reuse and possible subsequent re-engineering of knowledge resources**, as opposed to custom-building new ontologies from scratch.
- In order to support and promote such reuse-based approach, new methods, techniques, and tools are needed.

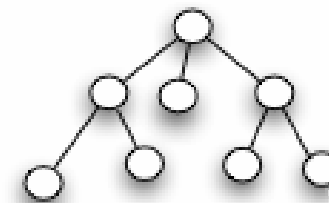
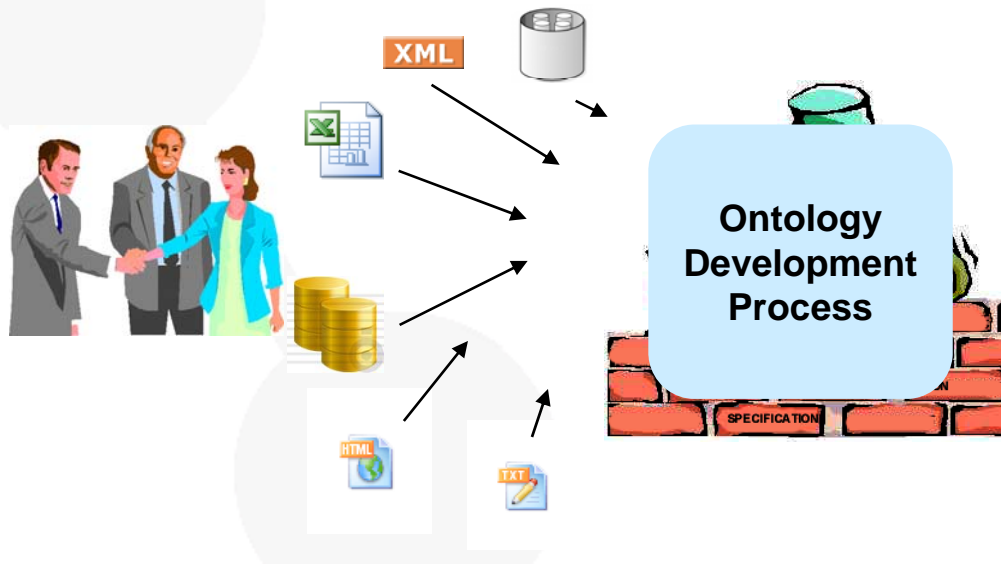
1. [www.esperonto.net](http://www.esperonto.net)
2. <http://knowledgeweb.semanticweb.org/>
3. <http://www.sekt-project.org/>

- In the software engineering community, it is well known that the **reuse of existing resources** helps to reduce costs as well as to disseminate good practices.
- This holds also in ontology engineering, where **reuse of existing knowledge** can be done either by directly reusing resources as they are, or after performing a **re-engineering** process.
- The underlying principle is that reuse allows to save time and money, and promotes the application of good practices.

## From scratch



## Reusing and re-engineering existing resources



- Introduction and motivation
- **Non-Ontological Resources (NORs)**
- Method for Reusing NORs
- Pattern based Re-engineering Method
- Patterns for Re-engineering Classification Schemes
- Patterns for Re-engineering Thesauri
- Patterns for Re-engineering Lexica
- Technological support

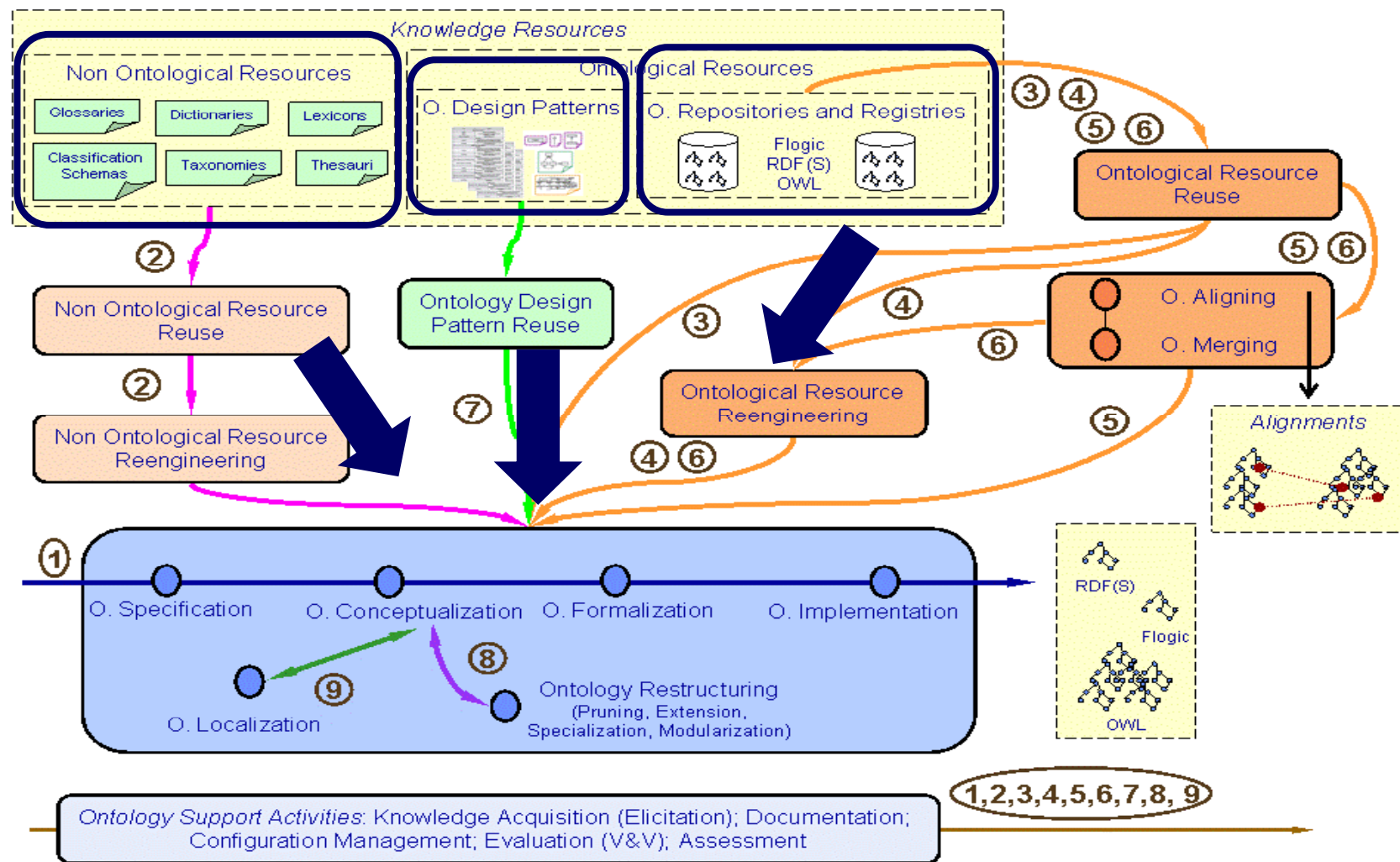
- Knowledge Resource: any information resource that contains or embodies knowledge

## Knowledge Resource

- Ontological Resource is a set of elements, extracted from a set of ontologies, available for solving a need. Elements from this set can be:

- ontologies,
- ontology modules,
- ontology statements
- ontology design patterns.

- Non-Ontological Resource (NOR) is an existing knowledge resource whose semantics has not yet been formalized by means of an ontology.





Knowledge Resources

Ontological Resources

- **Non-Ontological Resource (NOR)** is an existing knowledge-aware resource whose semantics has not yet been formalized by an ontology.
- **Non-Ontological Resource Reuse** refers to the process of choosing the most suitable non-ontological resources for the development of ontologies.
- **Non-Ontological Resource Re-engineering** refers to the process of taking an existing non-ontological resource and transforming it into an ontology.

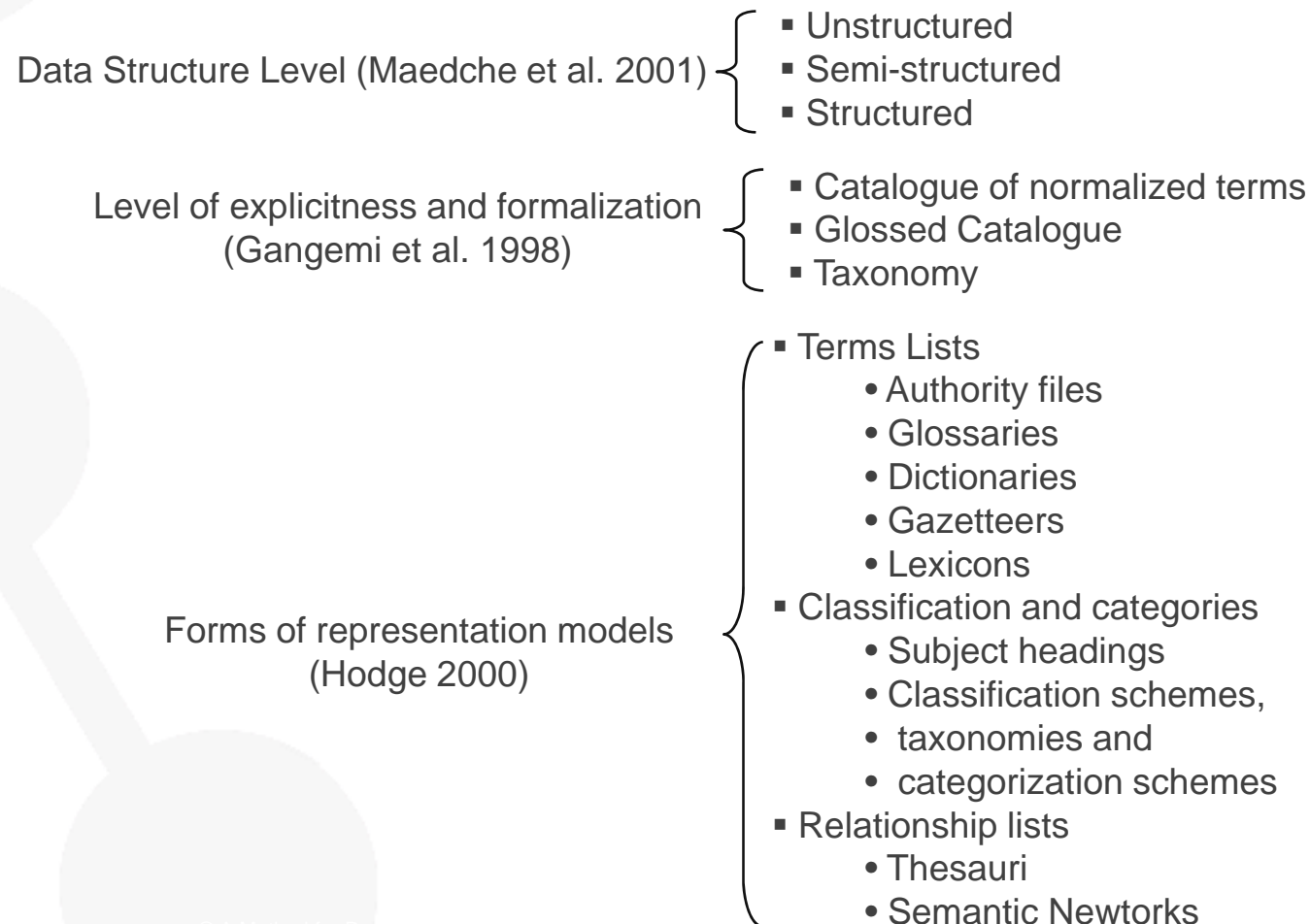


Ontology Support Activities: Knowledge Acquisition (Elicitation); Documentation;  
Configuration Management; Evaluation (V&V); Assessment

1,2,3,4,5,6,7,8, 9

# Non-Ontological Resource

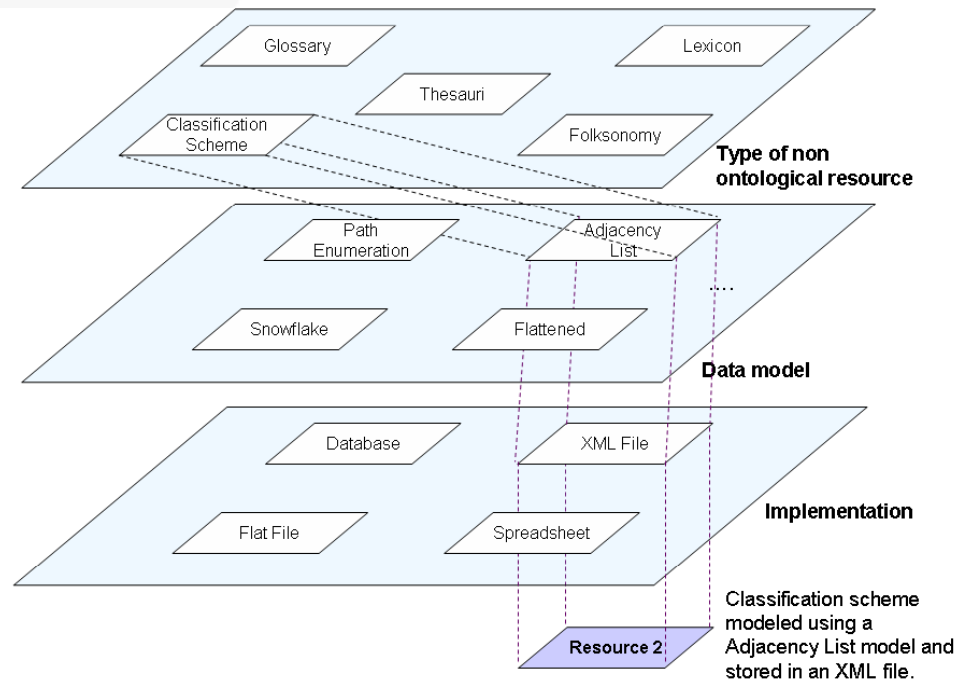
- Non-Ontological Resource (NOR) is an existing knowledge resource whose semantics has not yet been formalized by an ontology.



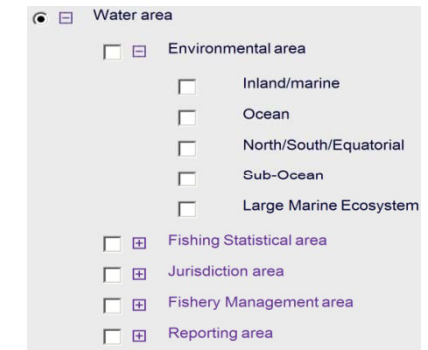
© A Method for Re  
Boris Villazon-Ter

# Patterns for Re-engineering Classification Schemes

## Example – FAO Water Areas Classification Scheme



### Classification Scheme



### Adjacency List

ID	CSI_Name	Parent
20000	Water area	
21000	Environmental area	20000
24020	Jurisdiction area	20000
22000	Fishing Statistical area	20000
21001	Inland/marine	21000
21002	Ocean	21000
21003	North/South/Equatorial	21000
22001	FAO statistical area	22000
22002	Areal grid system	22000

### XML

```

<Classification>
  <Category>
    <NodeId>20000</NodeId>
    <WaterCategory>Water Area</WaterCategory>
    <parentNodeId>1</parentNodeId>
  </Category>
  <Category>
    <NodeId>21000</NodeId>
    <WaterCategory>Environmental area</WaterCategory>
    <parentNodeId>20000</parentNodeId>
  </Category>
  <Category>
    <NodeId>22000</NodeId>
    <WaterCategory>Fishing statistical area</WaterCategory>
    <parentNodeId>20000</parentNodeId>
  </Category>
  <Category>
    <NodeId>24020</NodeId>
    <WaterCategory>Jurisdiction area</WaterCategory>
    <parentNodeId>20000</parentNodeId>
  </Category>
  <Category>
    <NodeId>21001</NodeId>
    <WaterCategory>inland/marine</WaterCategory>
    <parentNodeId>21000</parentNodeId>
  </Category>
  ....
</Classification>
    
```

- Introduction and motivation
- Non-Ontological Resources (NORs)
- **Method for Reusing NORs**
- Pattern based Re-engineering Method
- Patterns for Re-engineering Classification Schemes
- Patterns for Re-engineering Thesauri
- Patterns for Re-engineering Lexica
- Technological support

# Reusing NORs

## Methodological Guidelines (I)

### Non-Ontological Resource Reuse

#### Definition

*Non-Ontological Resource Reuse* refers to the process of choosing the most suitable non-ontological resources for the development of ontologies.

#### Goal

To choose the most suitable non-ontological resources for building ontologies.

#### Input

The ontology requirements specification document (ORSD).

#### Output

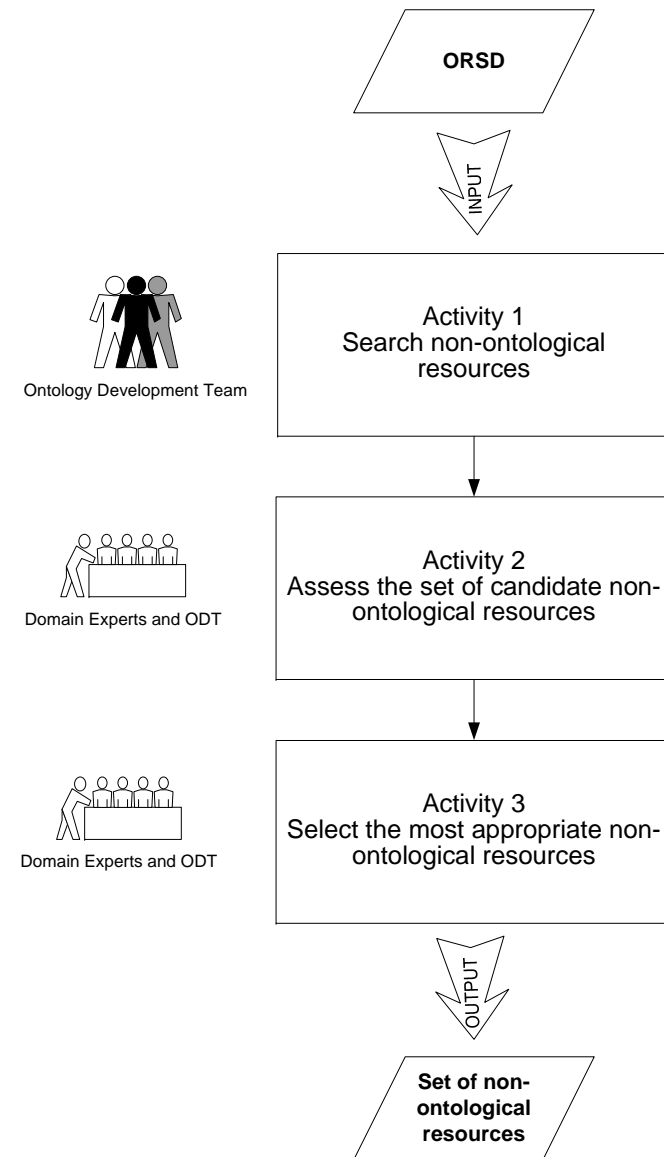
A set of non-ontological resources that to some extent covers the expected domain.

#### Who

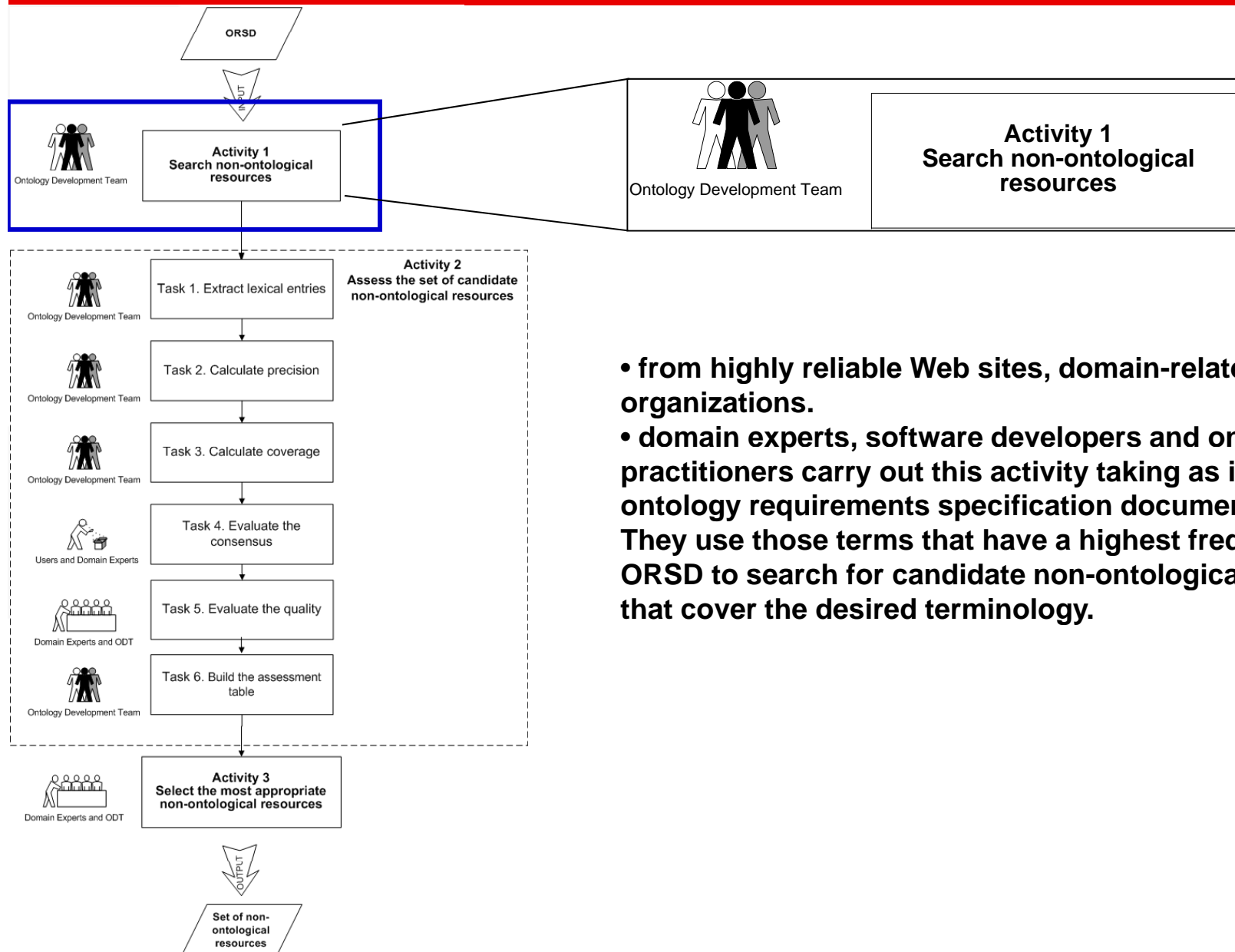
Domain experts, software developers and ontology practitioners.

#### When

After the ontology specification activity and before the non-ontological resource re-engineering process.

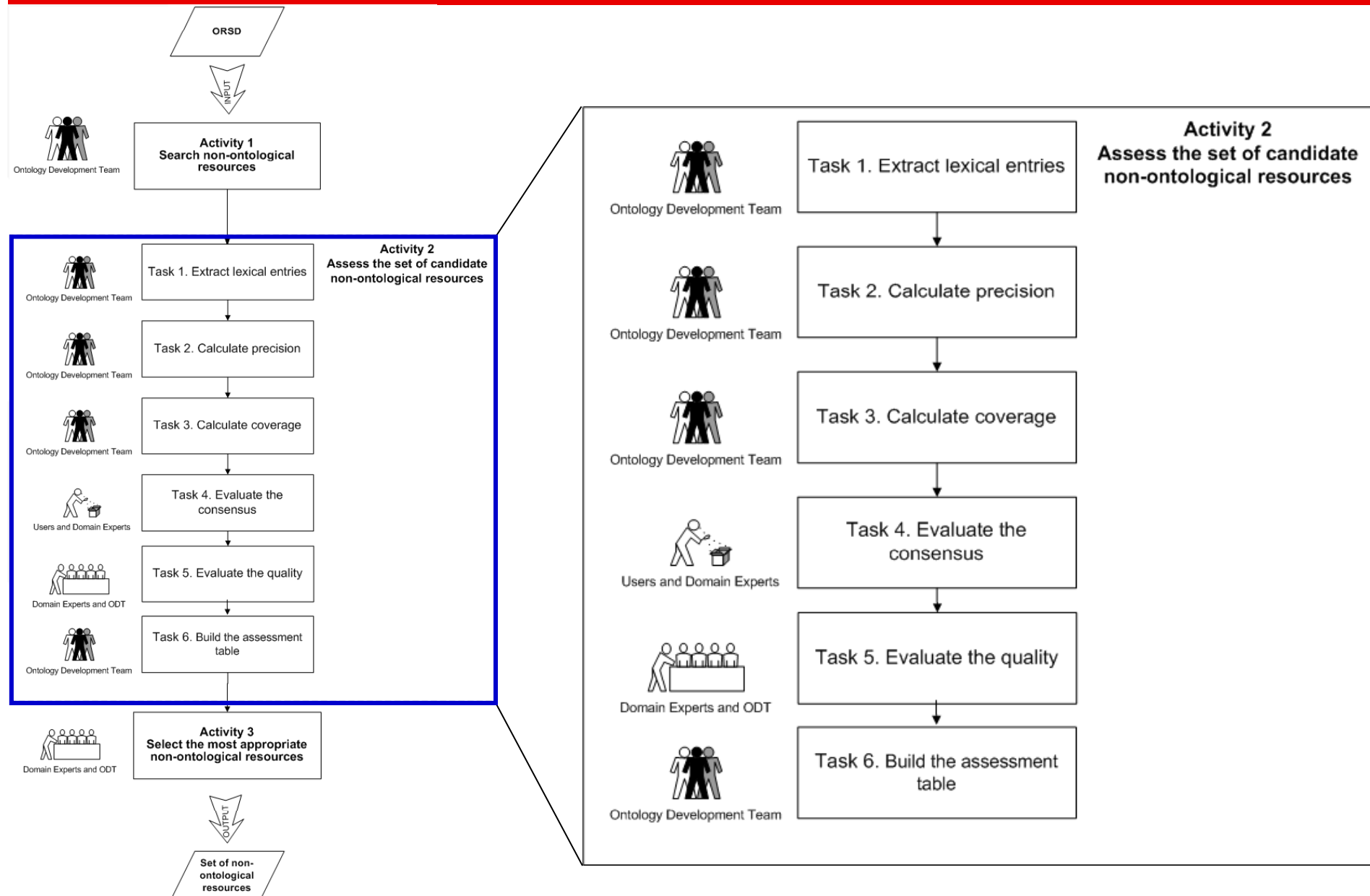


# Methodological Guidelines (II)



- from highly reliable Web sites, domain-related sites and organizations.
- domain experts, software developers and ontology practitioners carry out this activity taking as input the ontology requirements specification document (ORSD). They use those terms that have a highest frequency in the ORSD to search for candidate non-ontological resources that cover the desired terminology.

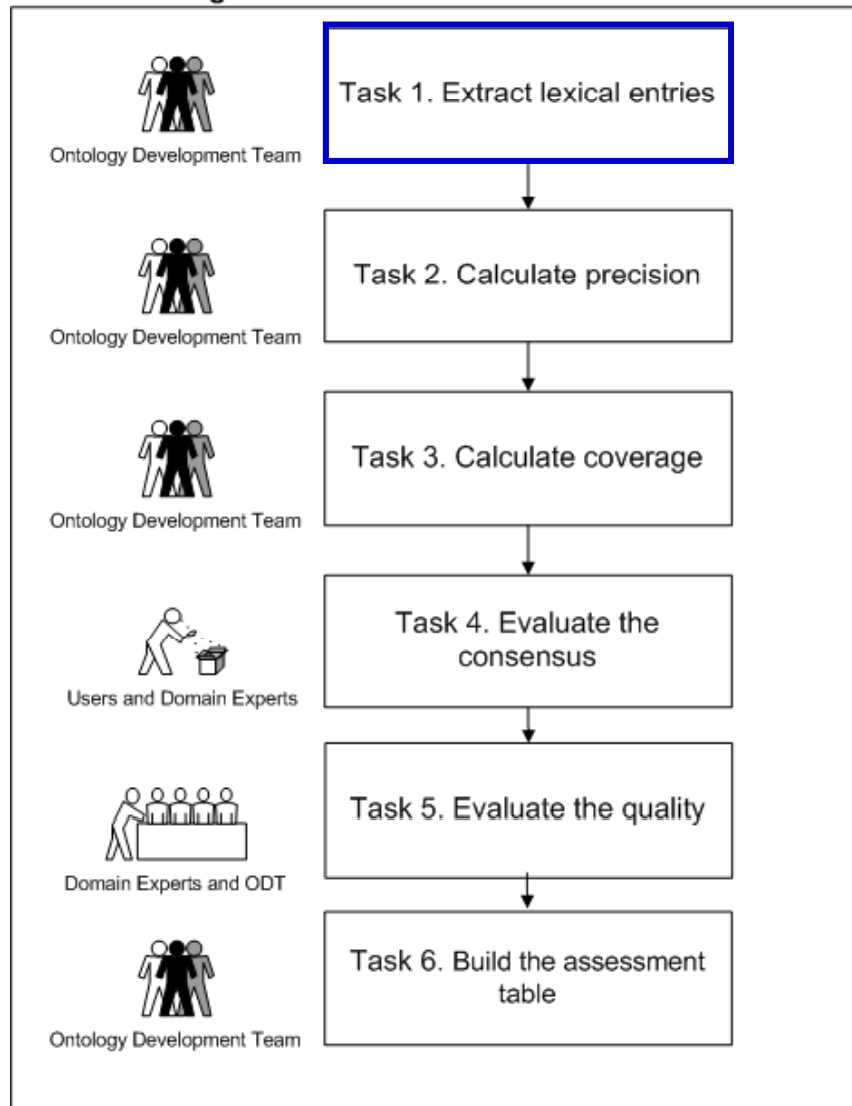
# Methodological Guidelines (III)



# Methodological Guidelines (IV)

## Activity 2

Assess the set of candidate non-ontological resources

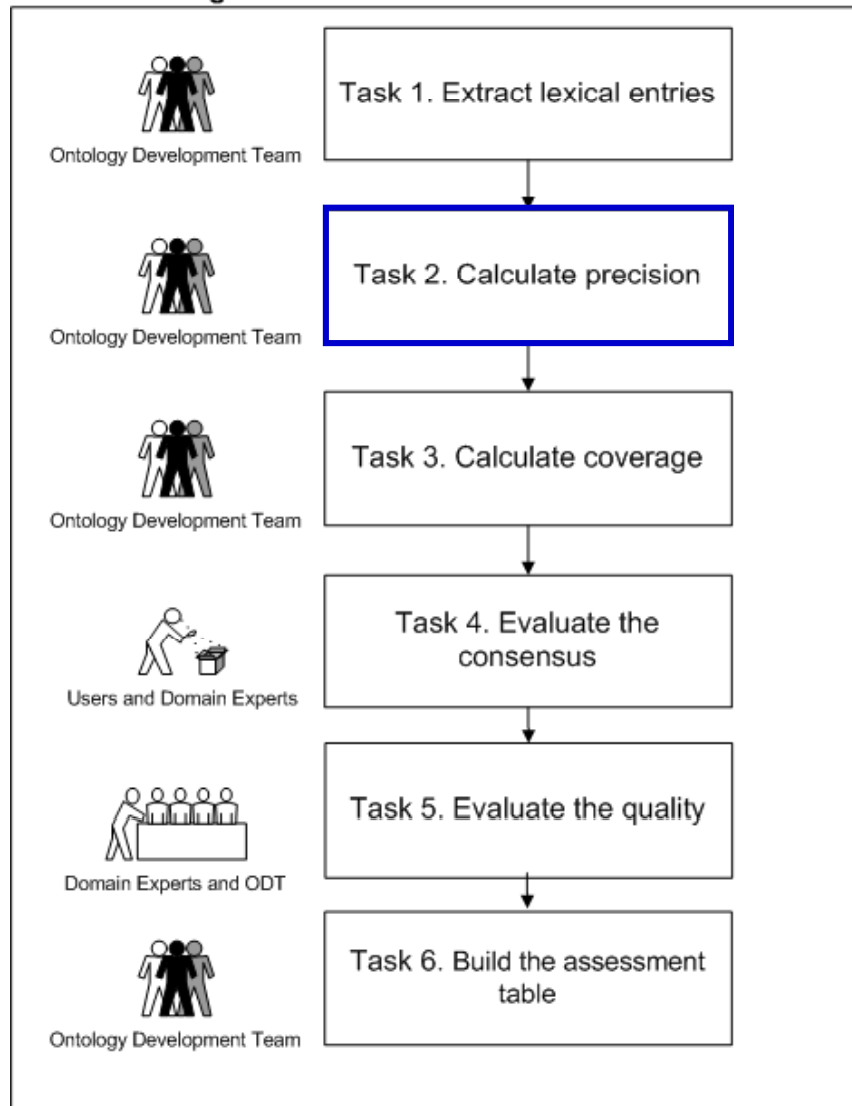


- Extract the lexical entries from the NORs, by using terminology extraction tools



## Methodological Guidelines (V)

## Activity 2

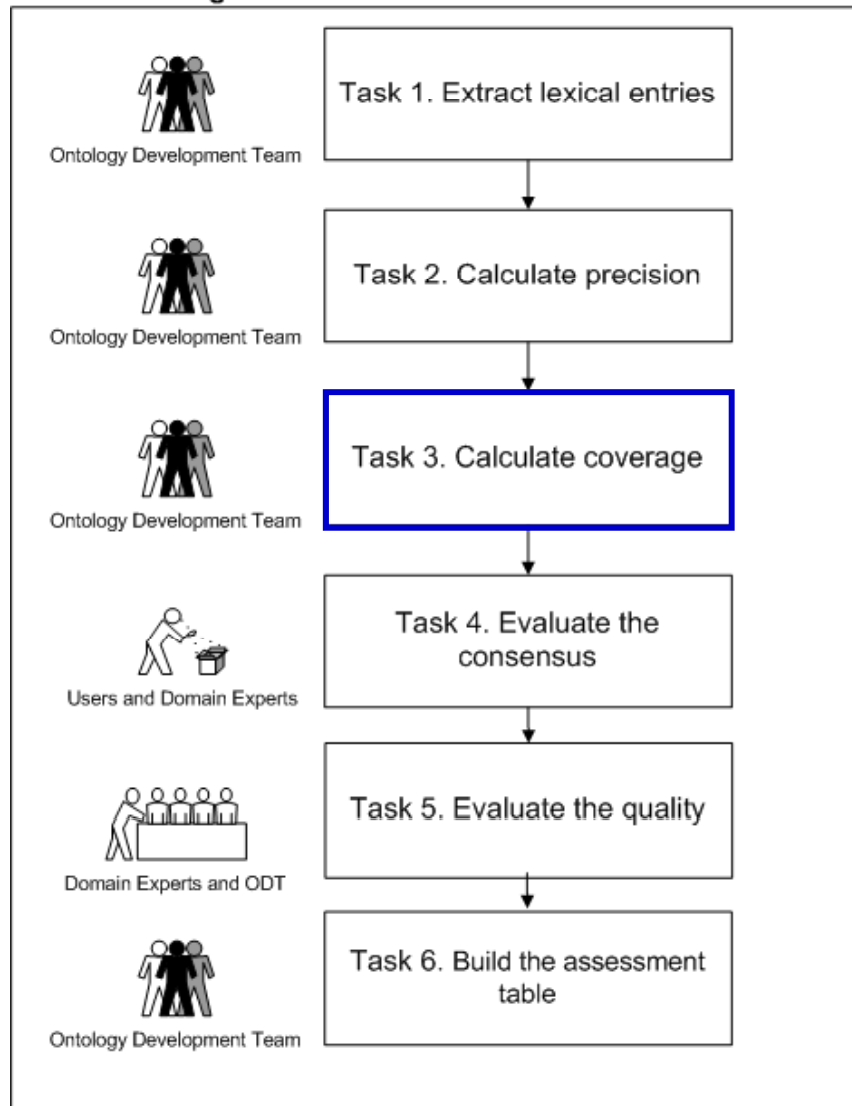
Assess the set of candidate  
non-ontological resources

$$Precision = \frac{|\{NORLexicalEntries\} \cap \{ORSDTerminology\}|}{|\{NORLexicalEntries\}|}$$

- NORLexicalEntries is the set of lexical entries extracted from the NOR.
- ORSDTerminology is the set of identified terms included in the ORSD.

## Methodological Guidelines (VI)

## Activity 2

Assess the set of candidate  
non-ontological resources

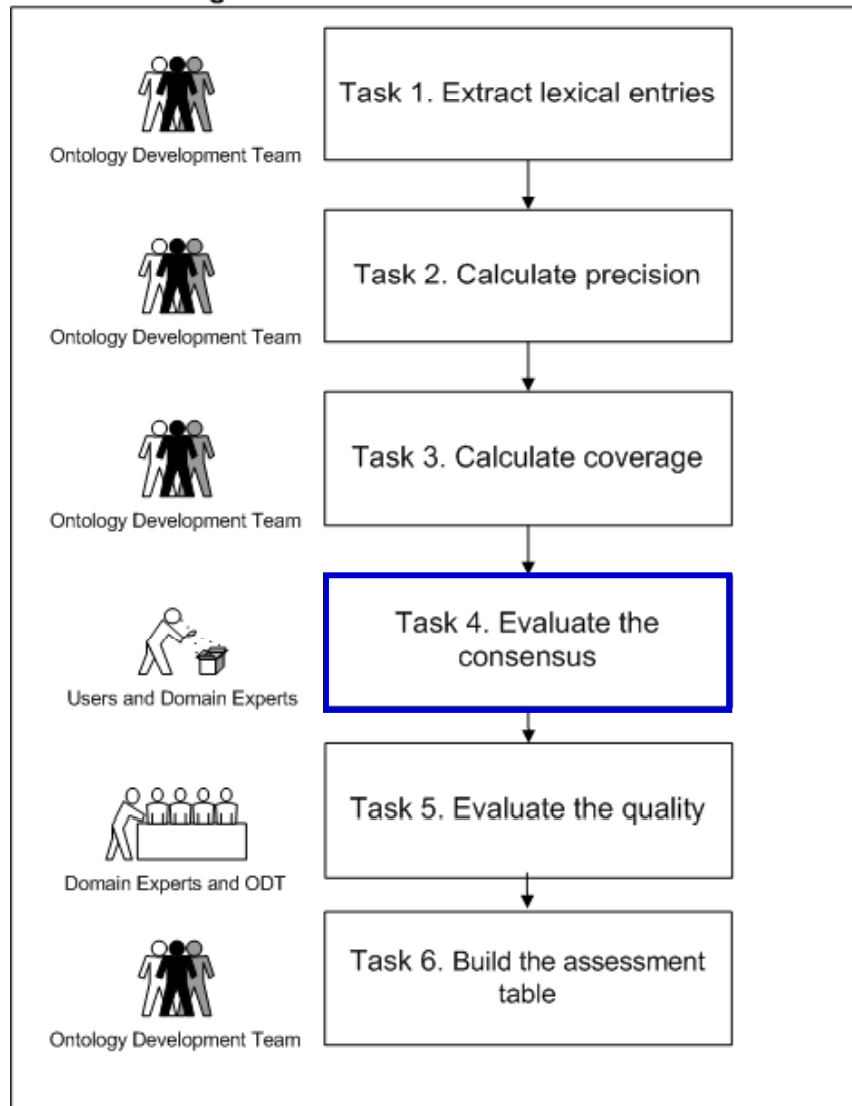
$$Coverage = \frac{|\{NORLexicalEntries\} \cap \{ORSDTerminology\}|}{|\{ORSDTerminology\}|}$$

- NORLexicalEntries is the set of lexical entries extracted from the NOR.
- ORSDTerminology is the set of identified terms included in the ORSD.

# Methodological Guidelines (VII)

## Activity 2

Assess the set of candidate non-ontological resources

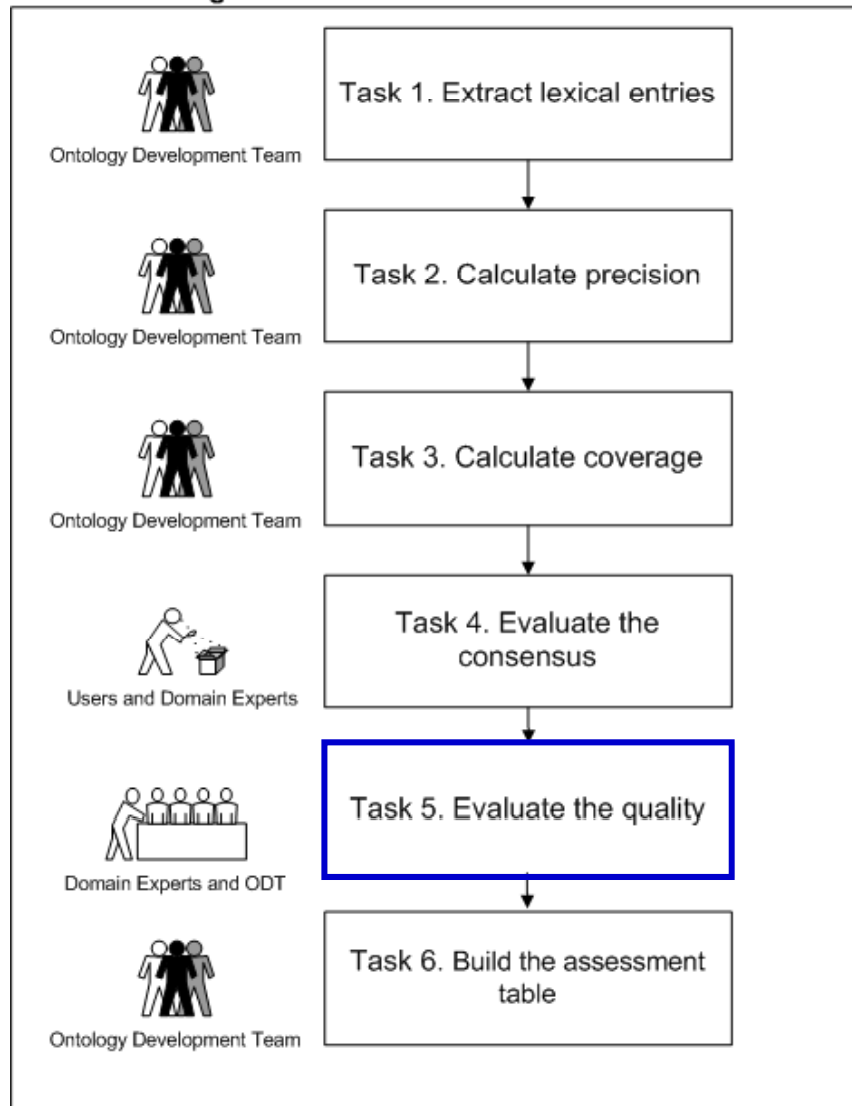


- Consensus is a subjective and not quantifiable criterion.
- Domain experts state whether the NORs have been reach consensus by the community or not.

# Methodological Guidelines (VIII)

## Activity 2

Assess the set of candidate non-ontological resources

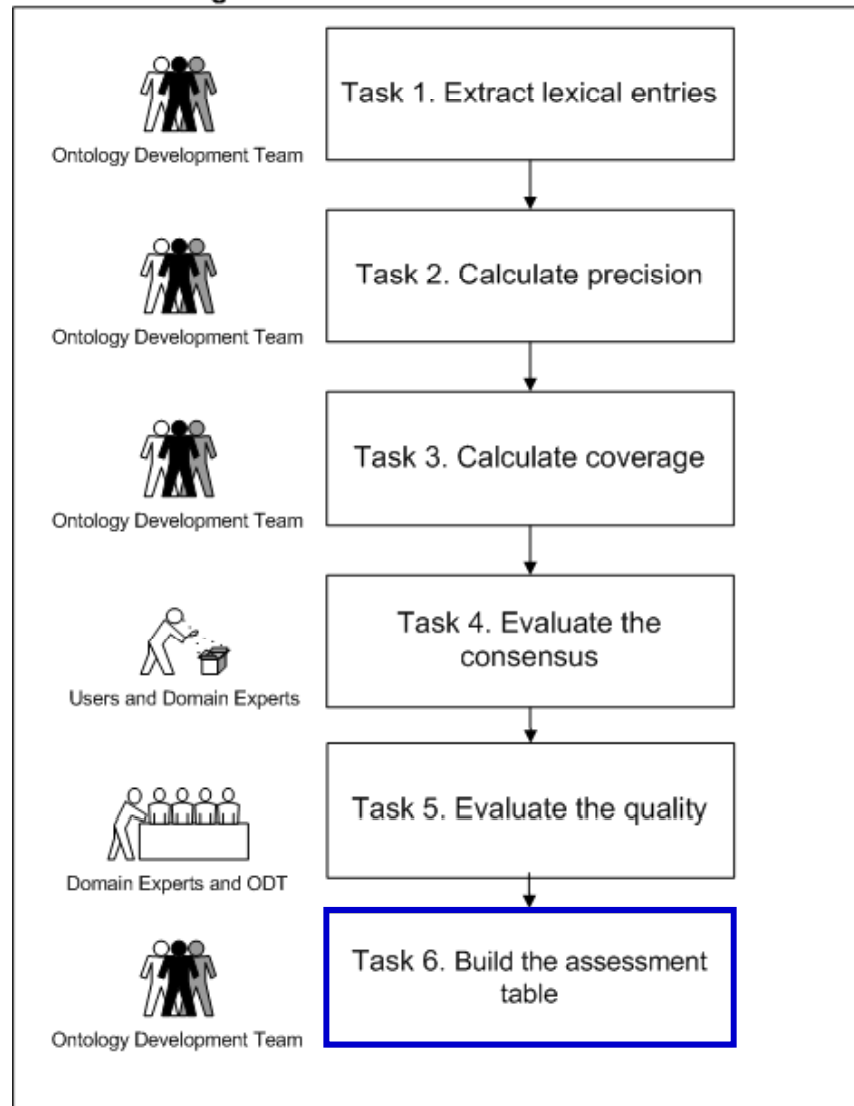


- A deep analysis of the quality of the resource is out of the scope of this thesis. Domain experts, software developers and ontology practitioners state whether the NORs have an acceptable quality. The quality attributes include:
  - well documentation of the resource
  - lack of anomalies of the NOR, such redundancies or inconsistencies
  - reliability of the non-ontological resource, it means analyzing whether we can trust in the resource or not.

# Methodological Guidelines (IX)

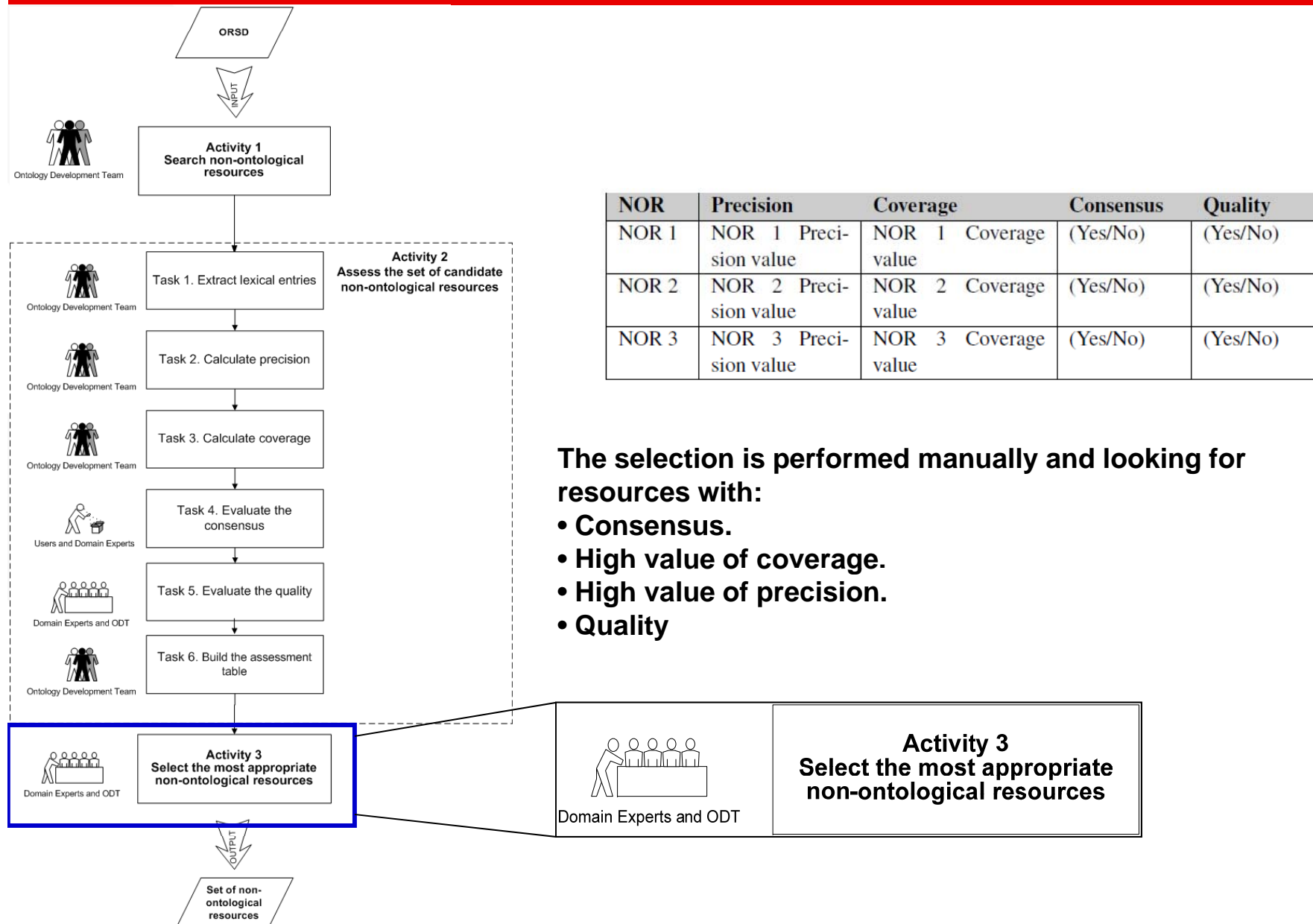
## Activity 2

Assess the set of candidate  
non-ontological resources



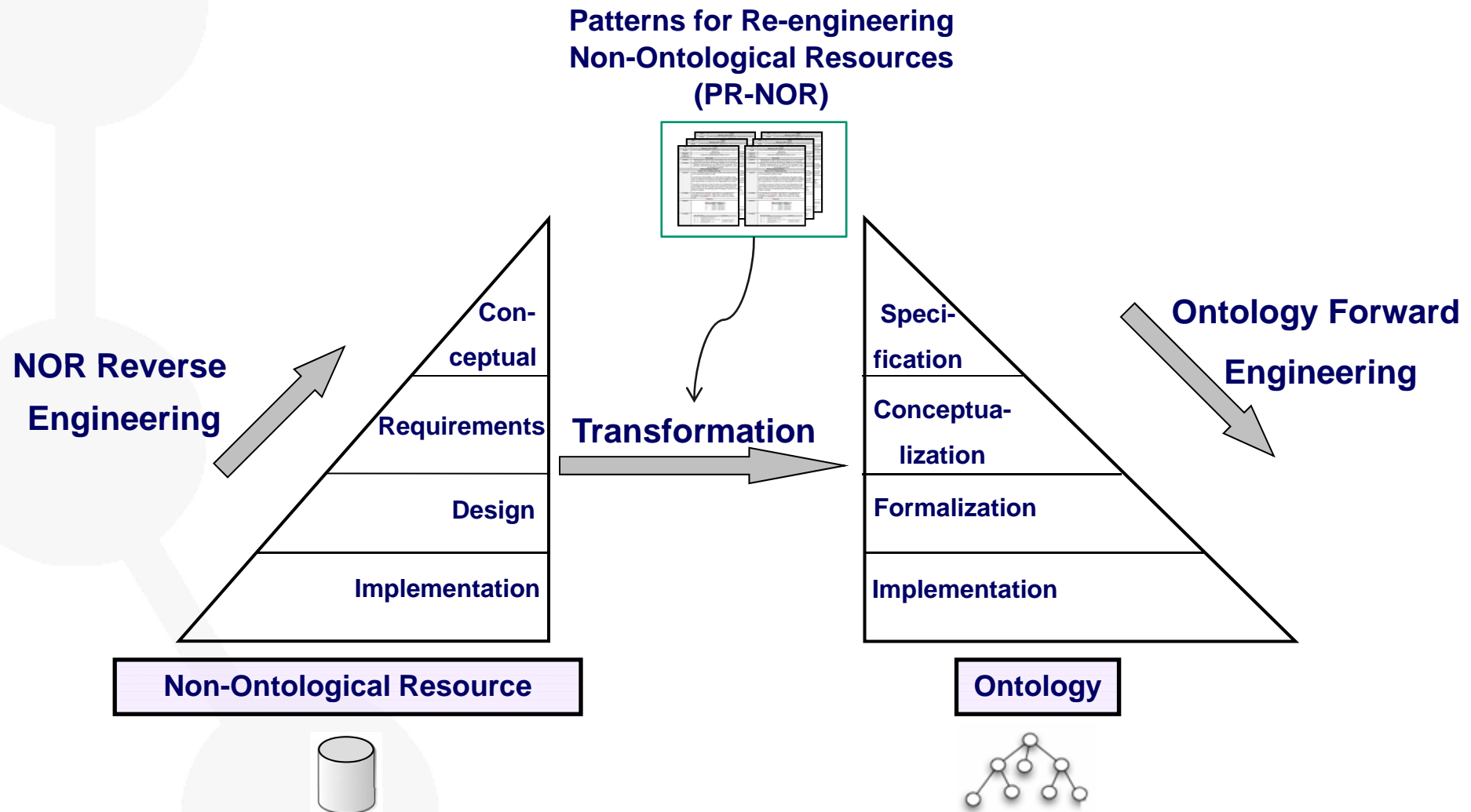
NOR	Precision	Coverage	Consensus	Quality
NOR 1	NOR 1 Precision value	NOR 1 Coverage value	(Yes/No)	(Yes/No)
NOR 2	NOR 2 Precision value	NOR 2 Coverage value	(Yes/No)	(Yes/No)
NOR 3	NOR 3 Precision value	NOR 3 Coverage value	(Yes/No)	(Yes/No)

# Methodological Guidelines (X)



- Introduction and motivation
- Non-Ontological Resources (NORs)
- Method for Reusing NORs
- **Pattern based Re-engineering Method**
- Patterns for Re-engineering Classification Schemes
- Patterns for Re-engineering Thesauri
- Patterns for Re-engineering Lexica
- Technological support

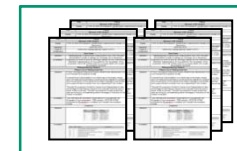
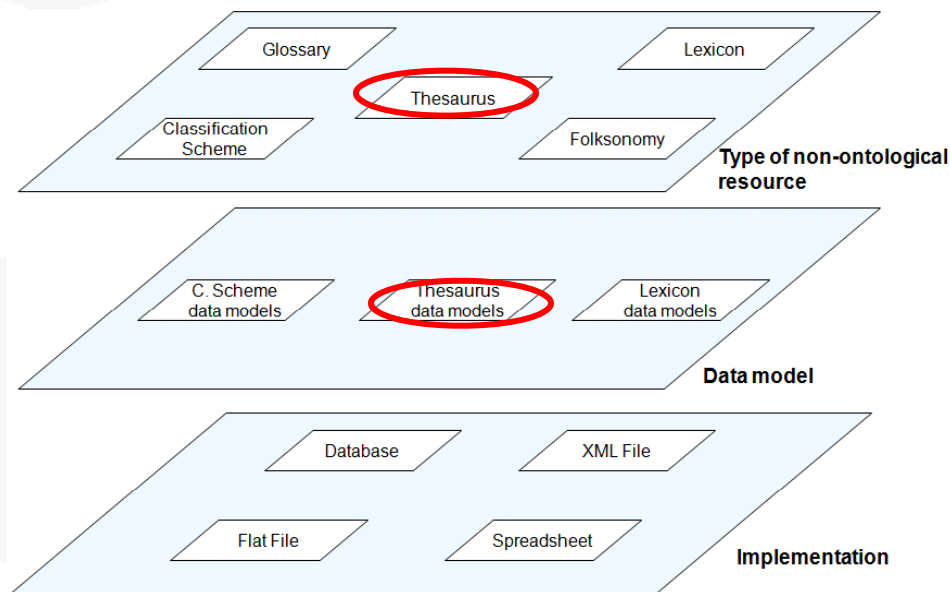
# Re-engineering Model for NORs





# Patterns for Re-engineering Non-Ontological Resources (PR-NORs)

- PR-NORs define a procedure that transforms the NOR components into ontology representational primitives.



**PR-NOR**



- According to our NOR categorization, the data model can be different even for the same type of NOR. For every data model we can define a process with a well-defined sequence of activities to extract the NORs components and then to map these components to a conceptual model of an ontology.
- Each of these processes can be expressed as a pattern for re-engineering NORs.

# Template for the PR-NOR

Slot	Value
<b>General Information</b>	
Name	Name of the pattern
Identifier	An acronym composed of component type + abbreviated name of the component + number
Component Type	Pattern for Re-engineering Non-Ontological Resource (PR-NOR)
<b>Use Case</b>	
General	Description in natural language of the re-engineering problem addressed by the pattern for re-engineering non-ontological resources.
Example	Description in natural language of an example of the re-engineering problem.
<b>Pattern for Re-engineering Non-Ontological Resource</b>	
<b>INPUT: Resource to be Re-engineered</b>	
General	Description in natural language of the non-ontological resource.
Example	Description in natural language of an example of the non-ontological resource.
<b>Graphical Representation</b>	
General	Graphical representation of the non-ontological resource.
Example	Graphical representation of the example of non-ontological resource.
<b>OUTPUT: Designed Ontology</b>	
General	Description in natural language of the ontology created after applying the pattern for re-engineering the non-ontological resource.
<b>Graphical Representation</b>	
(UML) General Solution Ontology	Graphical representation, using the UML profile (Brockmans & Haase, 2006), of the ontology created for the non-ontological resource being re-engineered.
(UML) Example Solution Ontology	Example showing a graphical representation, using the UML profile (Brockmans & Haase, 2006), of the ontology created for the non-ontological resource being used.
<b>PROCESS: How to Re-engineer</b>	
General	Description in natural language of the general re-engineering process, using a sequence of activities.
Example	Description in natural language of the re-engineering process applied to the non-ontological resource example, using the above sequence of activities.
<b>Formal Transformation</b>	
General	Formal description of the transformation by using the formal definitions of the resources.
<b>Relationships (Optional)</b>	
Relations to other modelling components	Description of any relation to other PR-NOR patterns or other ontology design patterns.

INPUT

OUTPUT

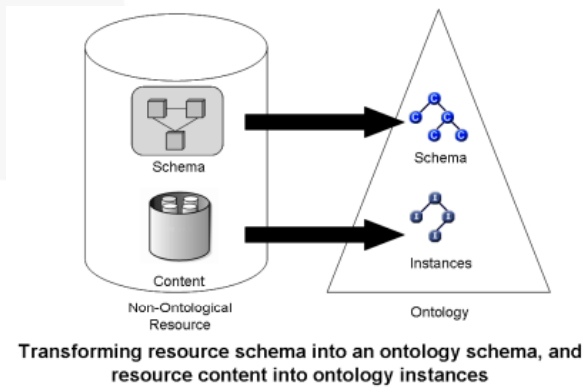
PROCESS

- The re-engineering patterns take advantage of the use of the Ontology Design Patterns<sup>1</sup> for creating the ontology. So, most of the ontologies generated follows the best practices already identified by the community.
- Although we have identified five types of NORs, here we just list patterns for re-engineering classification schemes, thesauri, and lexica.

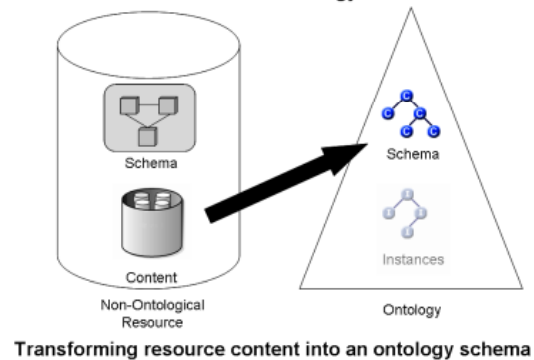


1. <http://ontologydesignpatterns.org>

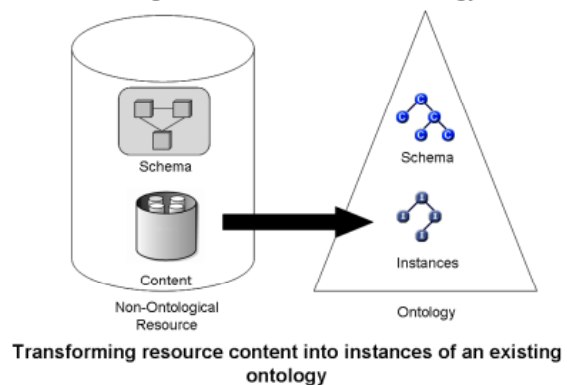
# Transformation approaches



» Abox transformation



» TBox transformation



» Population

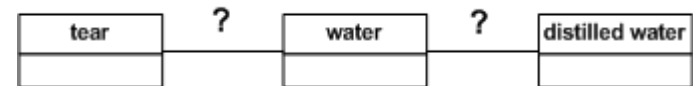
# Semantics of the Relations among the NOR Entities

- TBox transformation: each NOR entity is mapped to a class, and then, the semantics of the relations among those entities is identified. Thus, patterns that follow the TBox transformation approach must disambiguate the semantics of the relations among the NOR entities.

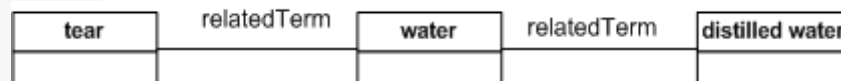
```

<TERM>
<DESCRIPTOR>water</DESCRIPTOR>
<RT>distilled water</RT>
<RT>tear</RT>
</TERM>

```

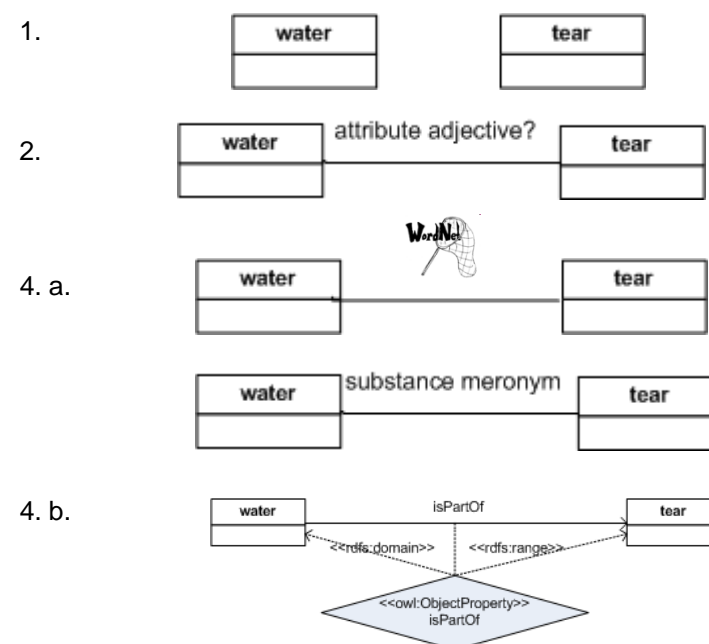
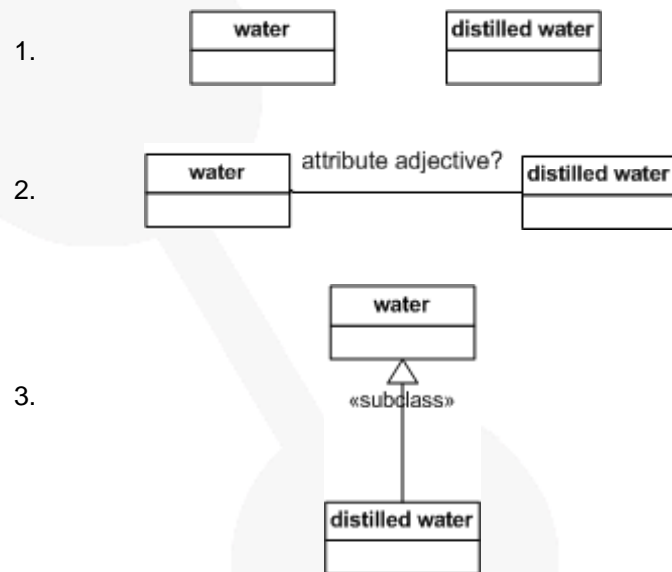


- default relation



# Algorithm for disambiguation

1. Take two related terms from the NOR.
2. Verify whether it is possible to get the *subClassOf* relation by identifying attribute adjectives within the two terms.
3. Relate the two terms by means of the relation *subClassOf*.
4. If it is not possible to get the *subClassOf* relation
  - a. Search in WordNet for a relation between those two terms. Currently the algorithm deal with the following WordNet elements:
    - the hyponym in the relation is interpreted as subClass
    - the hypernym in the relation is interpreted as superClass
    - the meronym in the relation is interpreted as Part
    - the holonym in the relation is interpreted as Whole
  - b. Relate the two terms by means of the relation returned by WordNet.
5. If WordNet gives an empty result, relate the two terms by means of the default relation.



# Re-engineering NORs

## Methodological Guidelines (I)

### Non-Ontological Resource Re-engineering

#### Definition

*Non-Ontological Resource Re-engineering* refers to the process of taking a non-ontological resource and transforming it into an ontology.

#### Goal

Creating an ontology from a non-ontological resource.

#### Input

One or more non-ontological resources selected by the reuse process.

#### Output

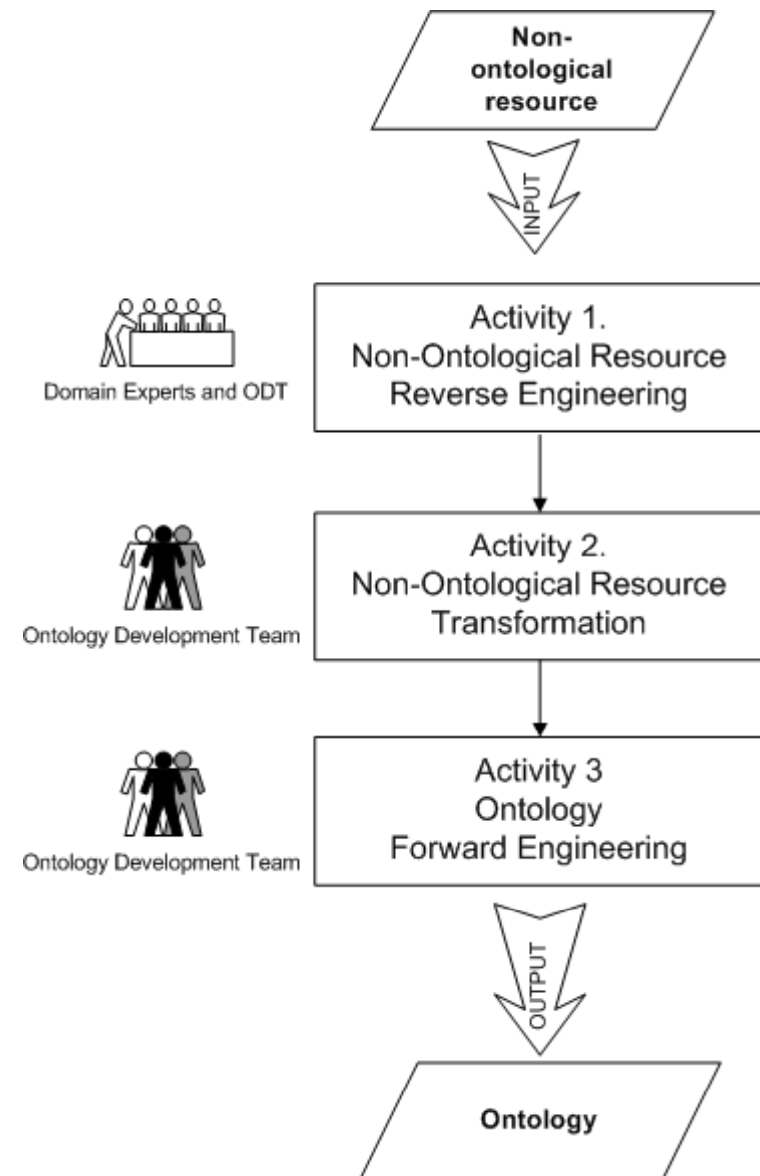
An ontology.

#### Who

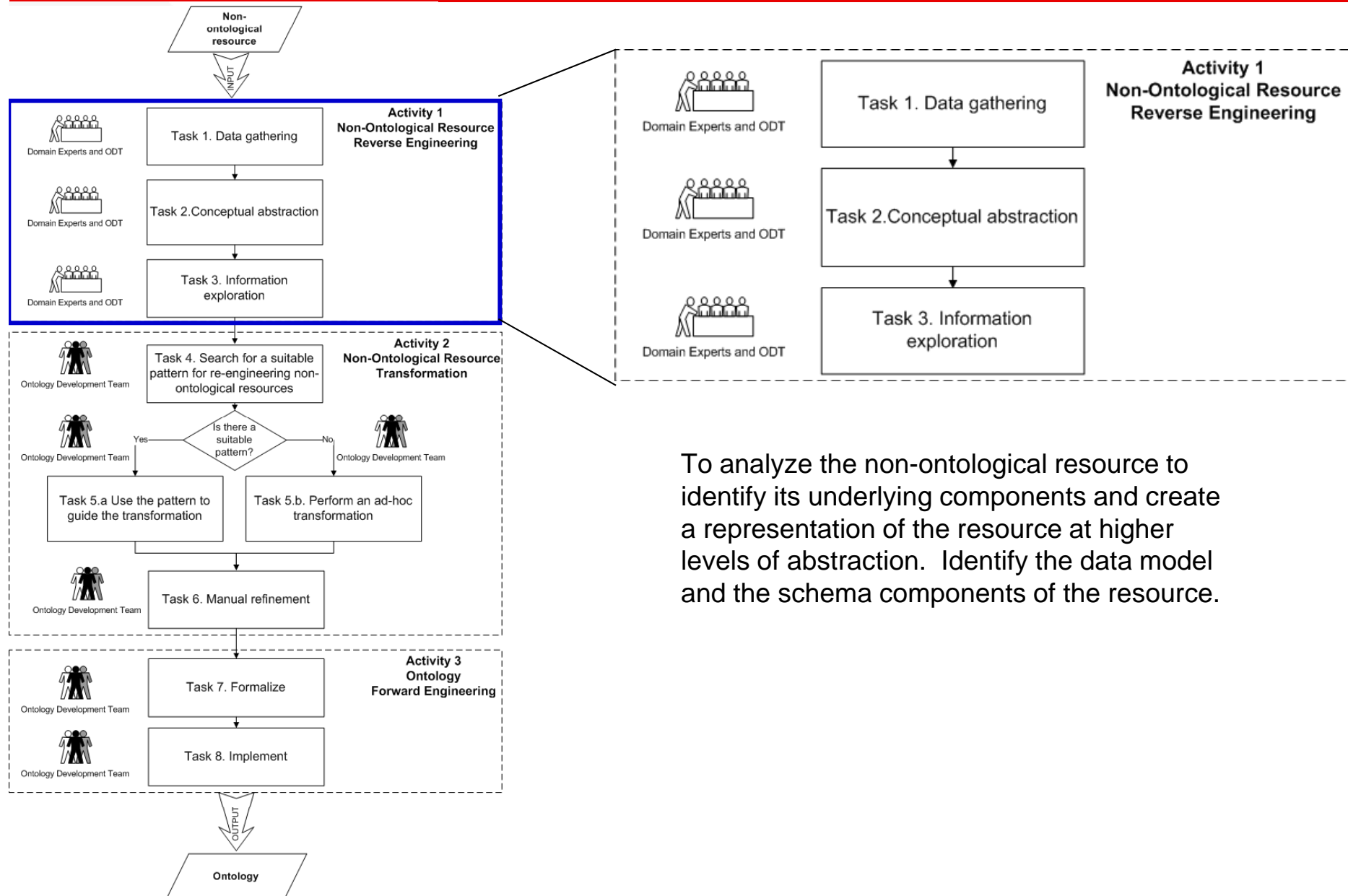
Domain experts, software developers and ontology practitioners.

#### When

After the non-ontological resource reuse process and before the conceptualization activity.



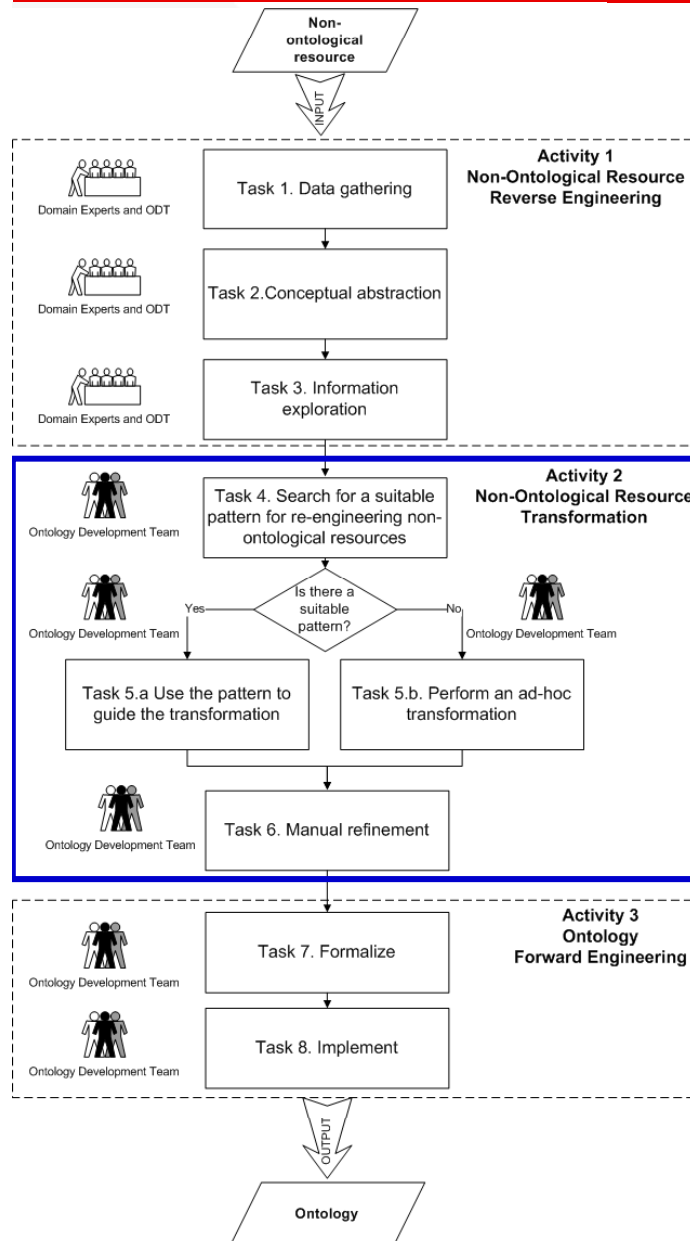
# Methodological Guidelines (II)



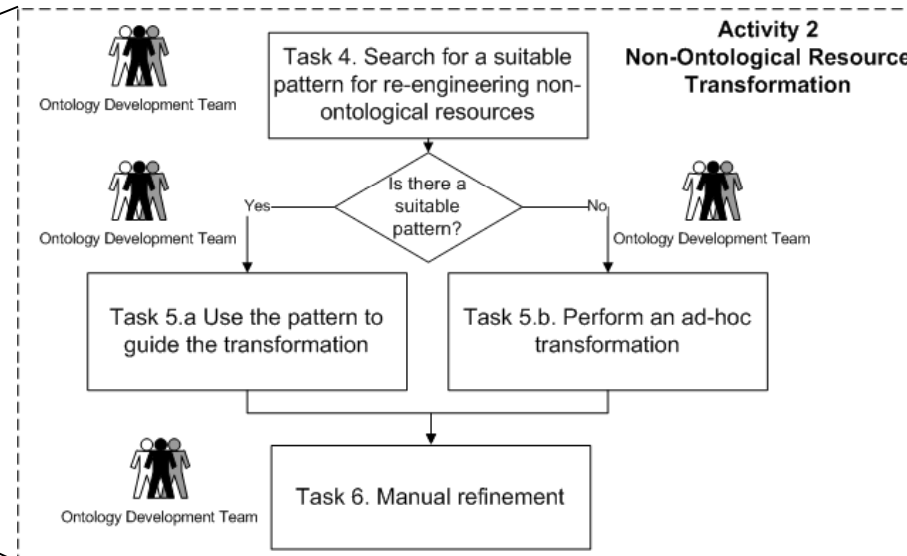
To analyze the non-ontological resource to identify its underlying components and create a representation of the resource at higher levels of abstraction. Identify the data model and the schema components of the resource.



# Methodological Guidelines (III)



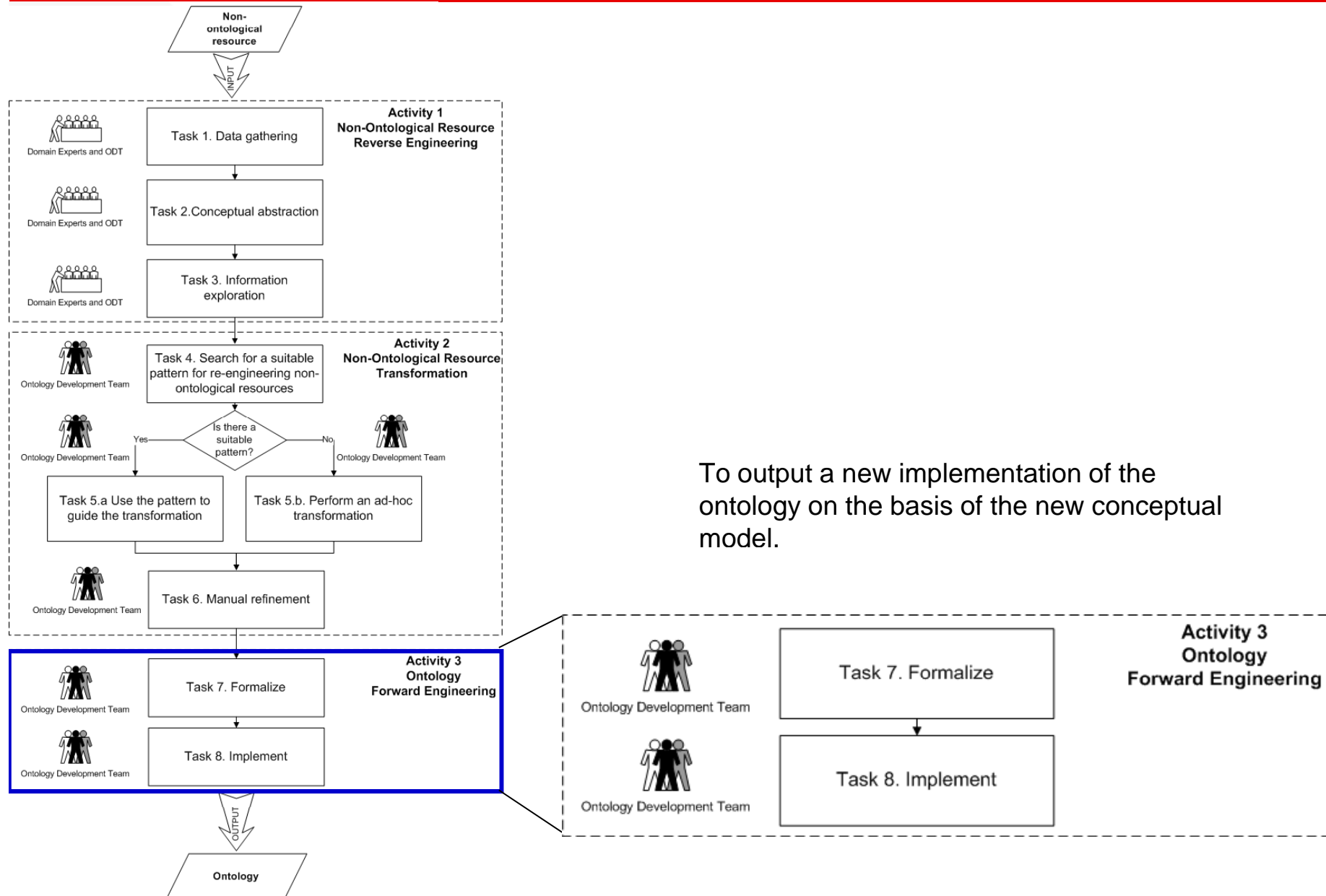
The goal is to generate a conceptual model from the NOR



Criteria for searching for a suitable pattern:

- **NOR Type:** classification scheme, thesauri, or lexicon
- **Data model:** C.Scheme data model, thesaurus data model, or lexicon data model
- **Transformation approach:** TBox, ABox

## Methodological Guidelines (IV)



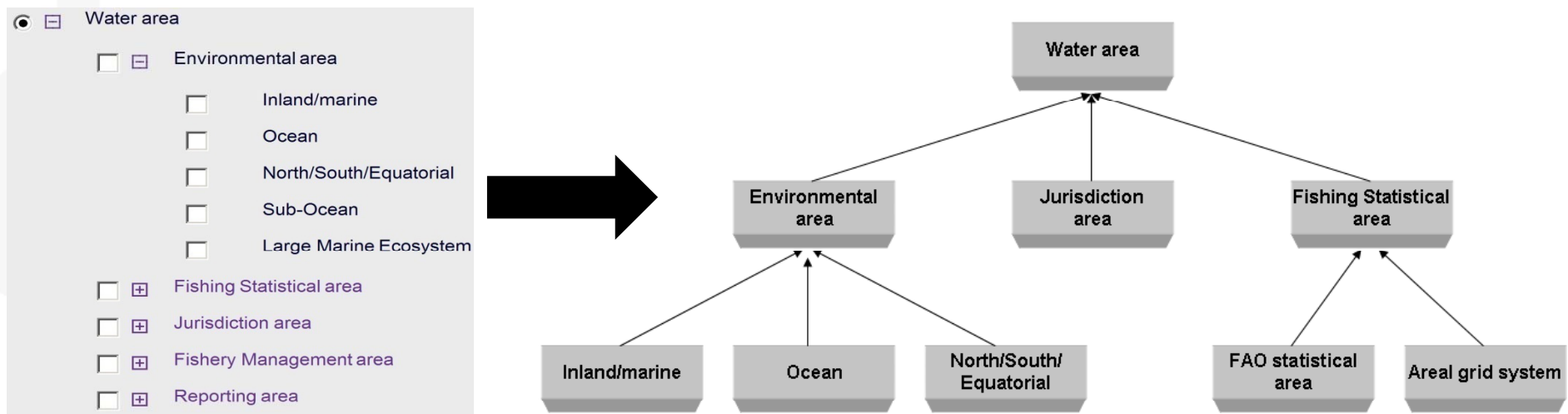
To output a new implementation of the ontology on the basis of the new conceptual model.

- Introduction and motivation
- Non-Ontological Resources (NORs)
- Method for Reusing NORs
- Pattern based Re-engineering Method
- **Patterns for Re-engineering Classification Schemes**
- Patterns for Re-engineering Thesauri
- Patterns for Re-engineering Lexica
- Technological support

- Classification schemes
  - Components
  - Data Models
  - Example
- Patterns for Re-engineering Classification Schemes into Ontologies

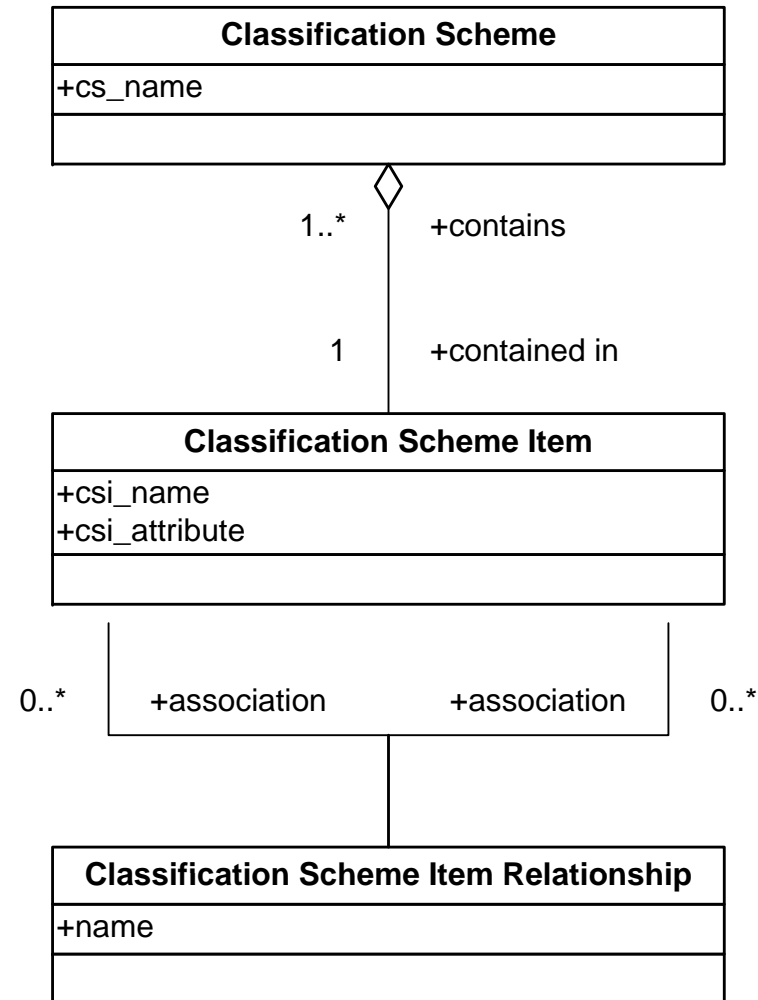
## Classification schemes

- A classification scheme<sup>1</sup> is the descriptive information for an arrangement or division of objects into groups based on characteristics, which the objects have in common. E.g. water area classification scheme<sup>2</sup>.



1. International Standard Organization (ISO). Information technology - Metadata registries – Part 1: Framework, 2004. Report ISO/IEC FDIS 11179-1.
2. <http://www.fao.org/figis/servlet/RefServlet>

## Classification Scheme Components



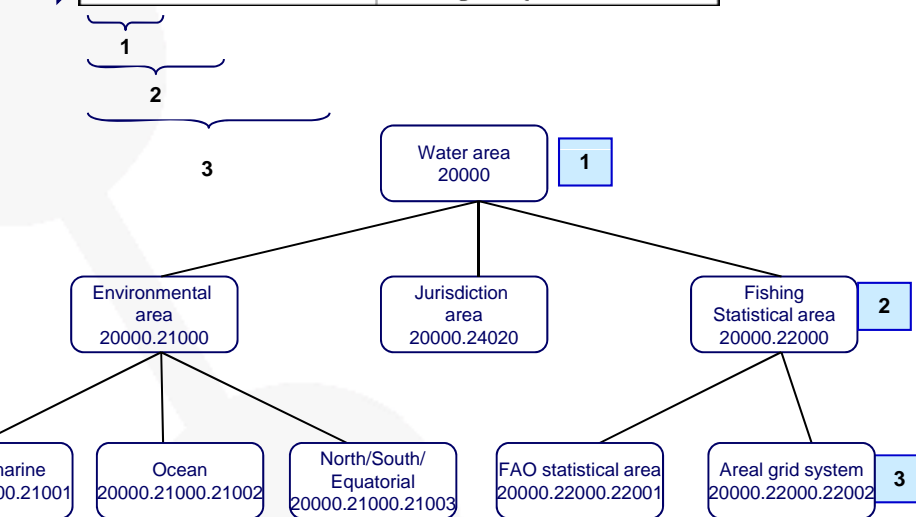
1. International Standard Organization (ISO). Information technology - Metadata registries – Part 1: Framework, 2004. Report ISO/IEC FDIS 11179-1.

# Patterns for Re-engineering Classification Schemes

## Classification Scheme Data Models (I)

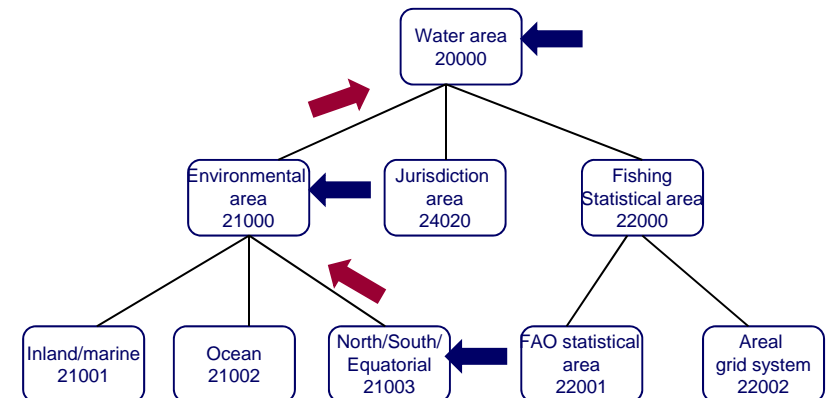
- Path Enumeration Data Model** is defined as a model that stores for each node the path (as a string) from the root to the node.

ID	CSI_Name
20000	Water area
20000.21000	Environmental area
20000.24020	Jurisdiction area
20000.22000	Fishing Statistical area
20000.21000.21001	Inland/marine
20000.21000.21002	Ocean
20000.21000.21003	North/South/Equatorial
20000.22000.22001	FAO statistical area
20000.22000.22002	Areal grid system



- Adjacency List** is a recursive structure for hierarchy representations that comprises a list of nodes with a linking column to their parent nodes.

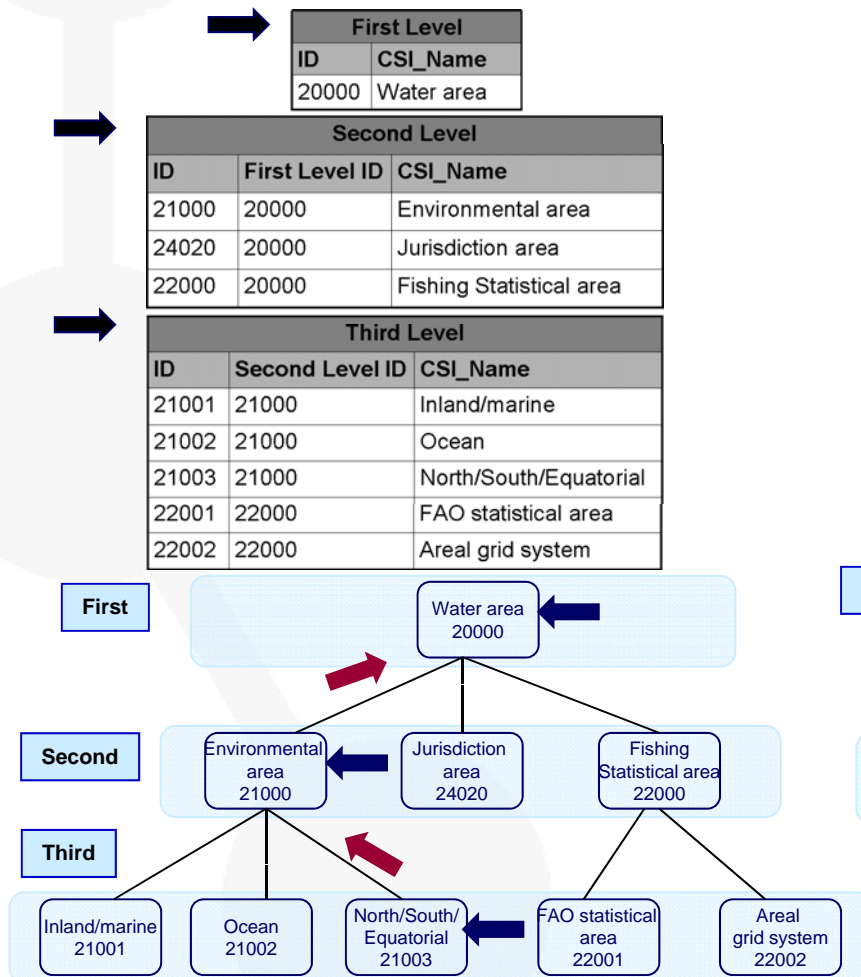
ID	CSI_Name	Parent
20000	Water area	
21000	Environmental area	20000
24020	Jurisdiction area	20000
22000	Fishing Statistical area	20000
21001	Inland/marine	21000
21002	Ocean	21000
21003	North/South/Equatorial	21000
22001	FAO statistical area	22000
22002	Areal grid system	22000



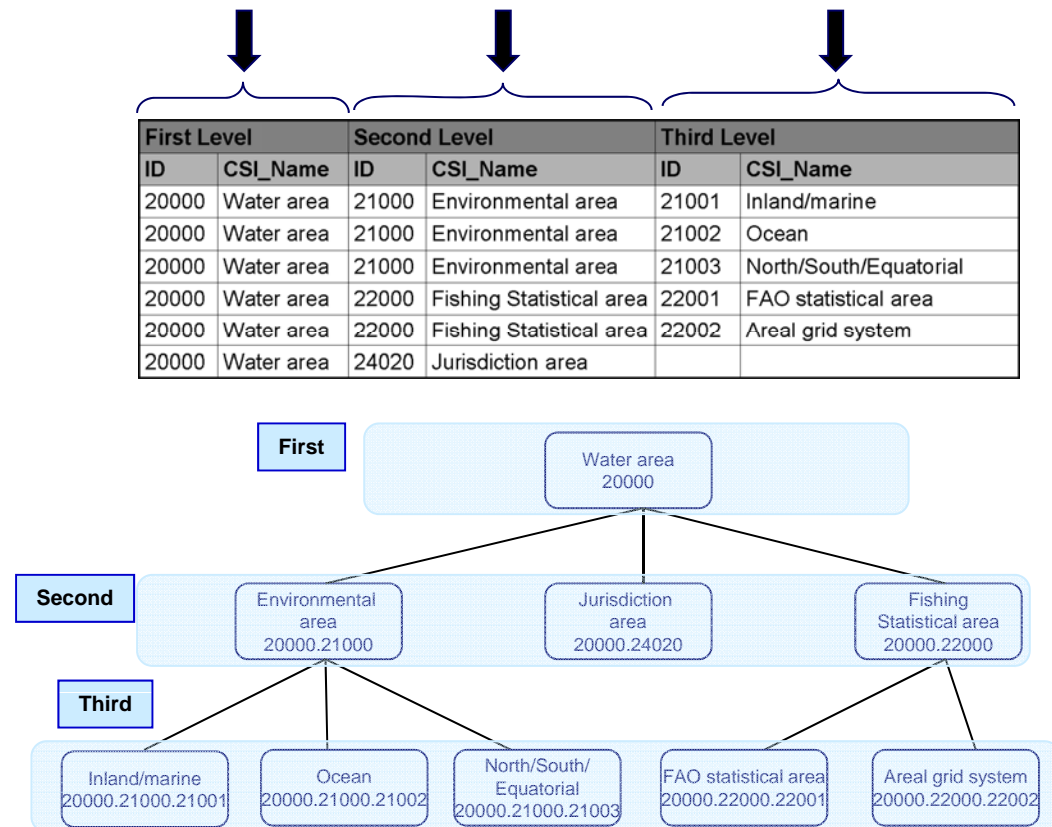
# Patterns for Re-engineering Classification Schemes

## Classification Scheme Data Models (II)

- Snowflake Data Model** is a normalized structure for hierarchy representations. For each hierarchy level a entity is created. In this model each hierarchy node has a column linked to its parent node.



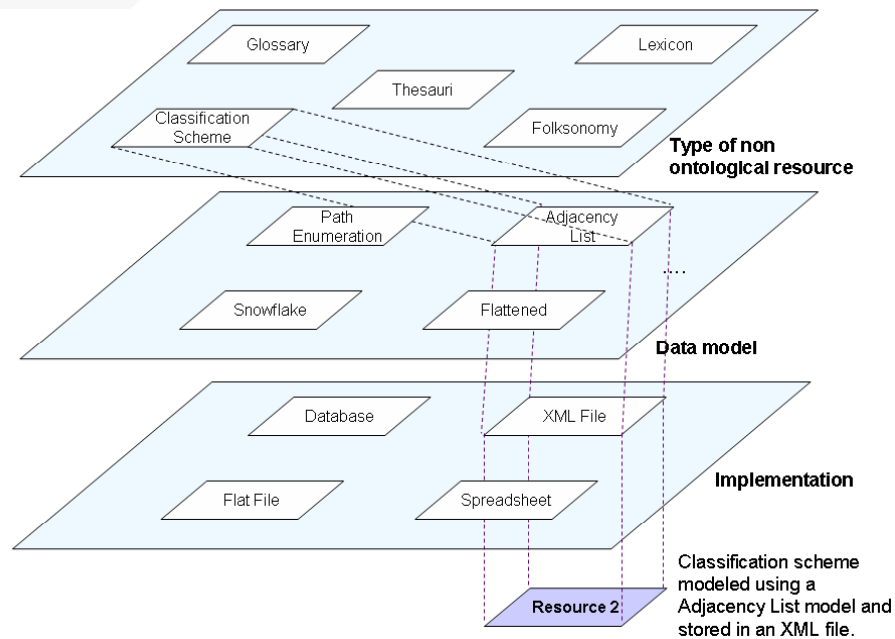
- Flattened Data Model**, is a denormalized structure. The hierarchy is represented with an entity where each hierarchy level is stored on a different column.



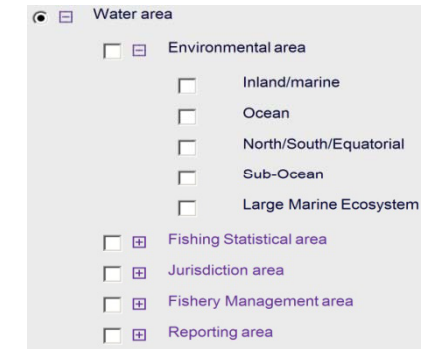


# Patterns for Re-engineering Classification Schemes

## Example – FAO Water Areas Classification Scheme



### Classification Scheme



### Adjacency List

ID	CSI_Name	Parent
20000	Water area	
21000	Environmental area	20000
24020	Jurisdiction area	20000
22000	Fishing Statistical area	20000
21001	Inland/marine	21000
21002	Ocean	21000
21003	North/South/Equatorial	21000
22001	FAO statistical area	22000
22002	Areal grid system	22000

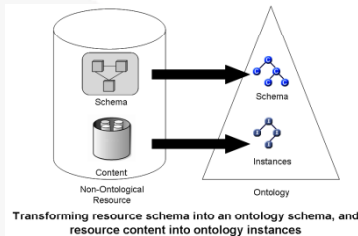
### XML

```

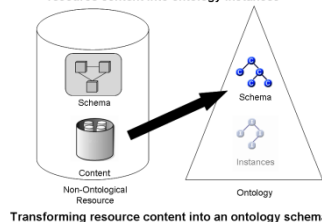
<Classification>
  <Category>
    <NodeId>20000</NodeId>
    <WaterCategory>Water Area</WaterCategory>
    <parentNodeId>1</parentNodeId>
  </Category>
  <Category>
    <NodeId>21000</NodeId>
    <WaterCategory>Environmental area</WaterCategory>
    <parentNodeId>20000</parentNodeId>
  </Category>
  <Category>
    <NodeId>22000</NodeId>
    <WaterCategory>Fishing statistical area</WaterCategory>
    <parentNodeId>20000</parentNodeId>
  </Category>
  <Category>
    <NodeId>24020</NodeId>
    <WaterCategory>Jurisdiction area</WaterCategory>
    <parentNodeId>20000</parentNodeId>
  </Category>
  <Category>
    <NodeId>21001</NodeId>
    <WaterCategory>inland/marine</WaterCategory>
    <parentNodeId>21000</parentNodeId>
  </Category>
  ...
</Classification>
    
```

## Patterns for Re-engineering Classification Schemes into Ontologies

N	Identifier	Type of NOR	NOR Data Model	Transformation approach	Target
1	PR-NOR-CLTX-01	Classification Scheme	Path Enumeration	TBox	Ontology Schema
2	PR-NOR-CLTX-02	Classification Scheme	Adjacency List	TBox	Ontology Schema
3	PR-NOR-CLTX-03	Classification Scheme	Snowflake	TBox	Ontology Schema
4	PR-NOR-CLTX-04	Classification Scheme	Flattened	TBox	Ontology Schema
5	PR-NOR-CLLO-10	Classification Scheme	Path Enumeration	ABox	Ontology
6	PR-NOR-CLLO-11	Classification Scheme	Adjacency List	ABox	Ontology
7	PR-NOR-CLLO-12	Classification Scheme	Snowflake	ABox	Ontology
8	PR-NOR-CLLO-13	Classification Scheme	Flattened	ABox	Ontology



ABox transformation

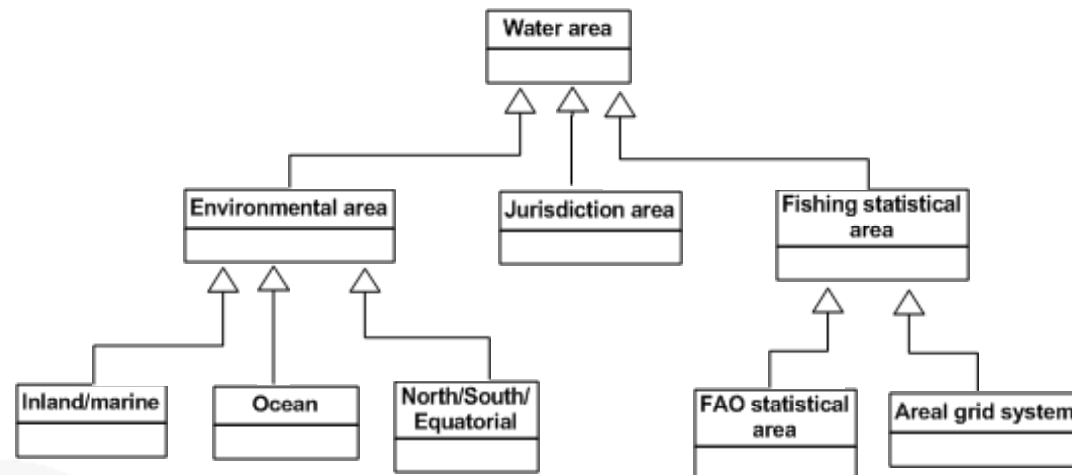


TBox transformation

# Patterns for Re-engineering Classification Schemes

Pattern for re-engineering a classification scheme, which follows the adjacency list data model, into an ontology schema

ID	CSI_Name	Parent
20000	Water area	
21000	Environmental area	20000
24020	Jurisdiction area	20000
22000	Fishing Statistical area	20000
21001	Inland/marine	21000
21002	Ocean	21000
21003	North/South/Equatorial	21000
22001	FAO statistical area	22000
22002	Areal grid system	22000



- Introduction and motivation
- Non-Ontological Resources (NORs)
- Method for Reusing NORs
- Pattern based Re-engineering Method
- Patterns for Re-engineering Classification Schemes
- **Patterns for Re-engineering Thesauri**
- Patterns for Re-engineering Lexica
- Technological support

- Thesauri
  - Components
  - Types of thesaurus
  - Data Models
  - Example
- Patterns for Re-engineering Thesauri into Ontologies

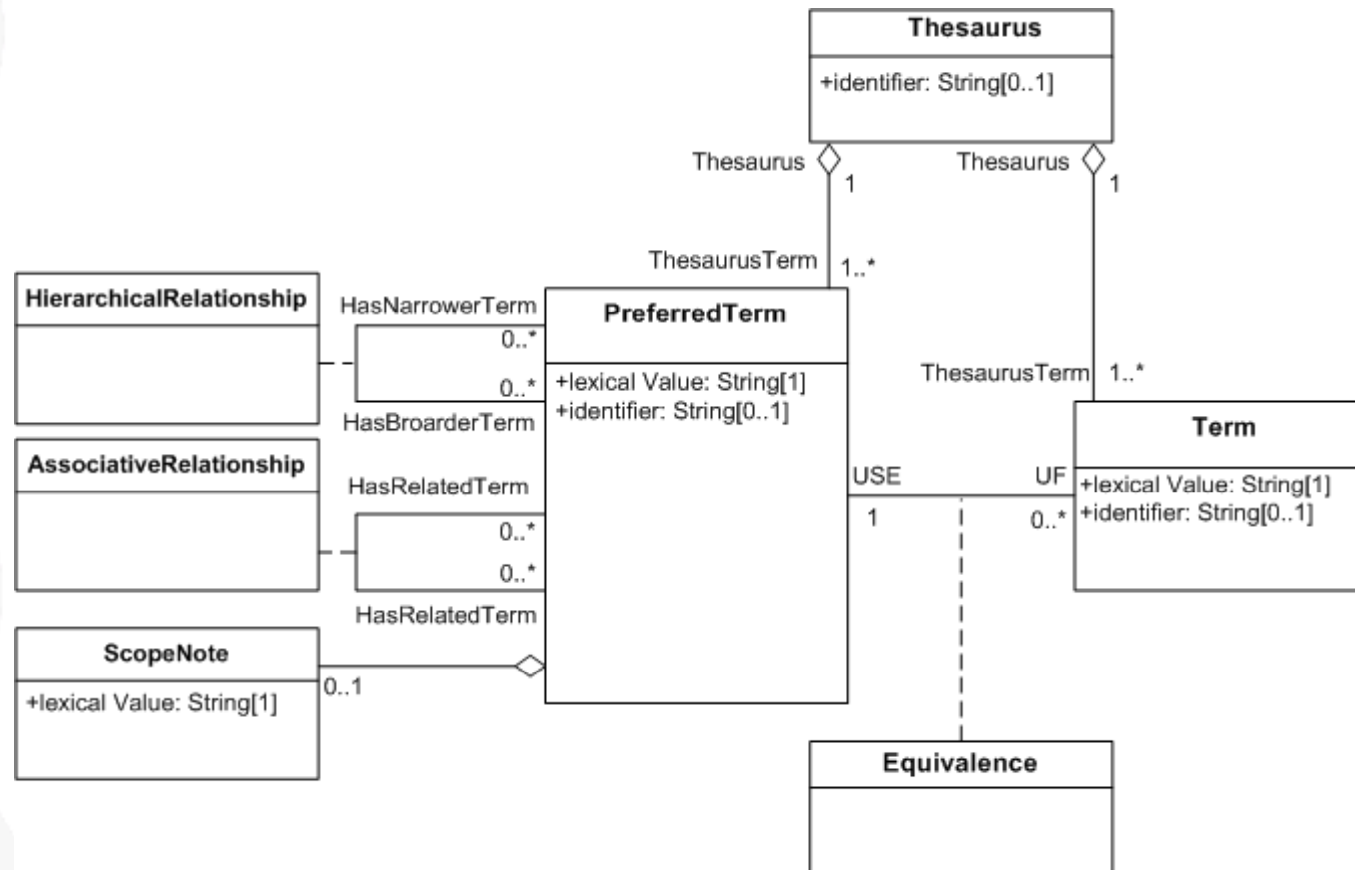
- Thesauri<sup>1</sup> are controlled vocabularies of terms in a particular domain with hierarchical, associative and equivalence relations between terms. Thesauri are mainly used for indexing and retrieving of articles in large databases. An example of thesaurus is the AGROVOC<sup>2</sup> thesaurus.

EN : Oryza	BT ( subclassOf ) : Poaceae
	NT ( hasSubclass ) : Oryza sativa
	NT ( hasSubclass ) : Oryza perennis
	NT ( hasSubclass ) : Oryza rufipogon
	NT ( hasSubclass ) : Oryza longistaminata
	NT ( hasSubclass ) : Wetland rice
	NT ( hasSubclass ) : Oryza glaberrima
	NT ( hasSubclass ) : Upland rice
	NT ( hasSubclass ) : Oryza punctata
	RT : Rice fields
	RT : Cereal crops
	RT : Rice
EN : Rice	BT ( subclassOf ) : Cereals
	NT ( hasSubclass ) : Broken rice
	NT ( hasSubclass ) : Basmati rice
	RT : Rice straw
	RT : Oryza
	RT : Rice flour
	UF : Paddy

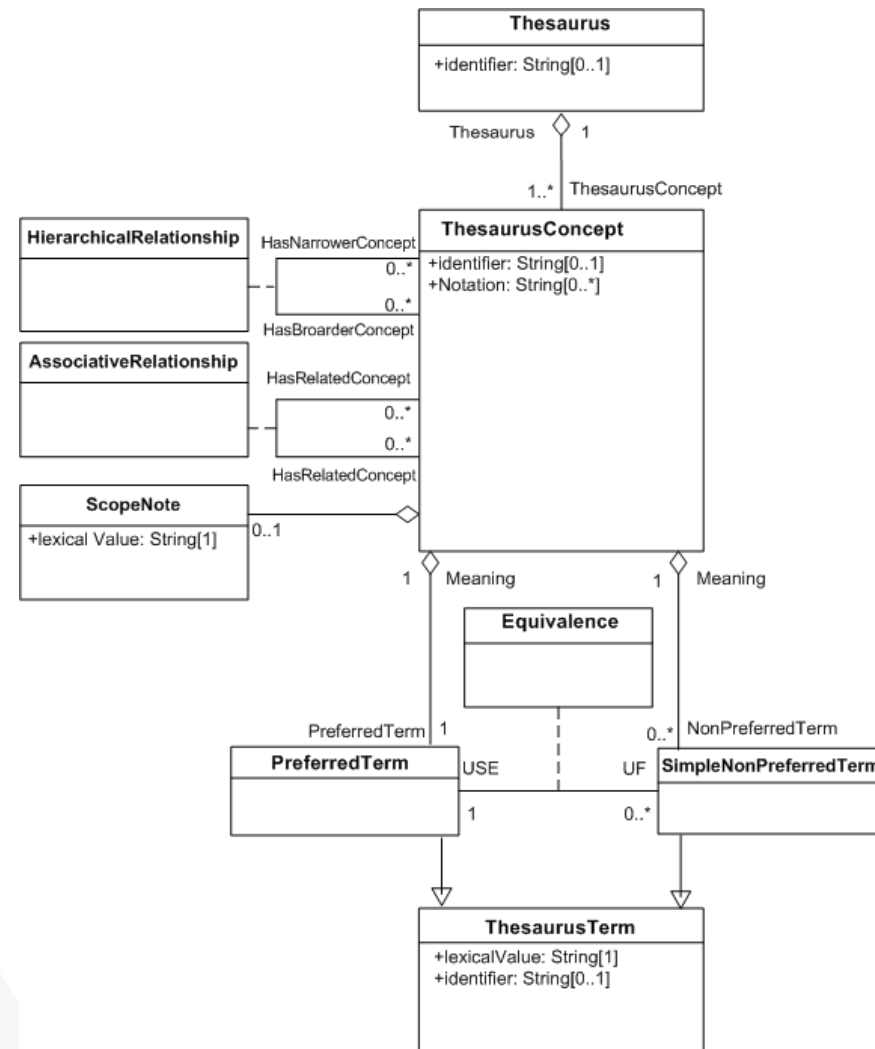


1. International Standard Organization (ISO). Documentation – Guidelines for the establishment and development of monolingual thesaurus, 1986. Report ISO 2788.

2. <http://www.fao.org/agrovoc/>



1. International Standard Organization (ISO). Documentation – Guidelines for the establishment and development of monolingual thesaurus, 1986. Report ISO 2788.



1. British Standards Institution, BSI. Documentation – Structured vocabularies for information retrieval - Guide - Part 5: Exchange formats and protocols for interoperability., 2005. Report BS 8723-5.



## Record-based data model for a term-based thesaurus

EN : Oryza	BT ( subclassOf ) : Poaceae
	NT ( hasSubclass ) : Oryza sativa
	NT ( hasSubclass ) : Oryza perennis
	NT ( hasSubclass ) : Oryza rufipogon
	NT ( hasSubclass ) : Oryza longistaminata
	NT ( hasSubclass ) : Wetland rice
	NT ( hasSubclass ) : Oryza glaberrima
	NT ( hasSubclass ) : Upland rice
	NT ( hasSubclass ) : Oryza punctata
	RT : Rice fields
	RT : Cereal crops
	RT : Rice
EN : Rice	BT ( subclassOf ) : Cereals
	NT ( hasSubclass ) : Broken rice
	NT ( hasSubclass ) : Basmati rice
	RT : Rice straw
	RT : Oryza
	RT : Rice flour
	UF : Paddy

Term	BT	NT	RT	UF
Rice	Cereals	Broken rice Basmati rice	Rice straw Oryza	Paddy
Oryza	Poaceae	Oryza sativa Oryza perennis Oryza rufipogon Oryza longistaminata Wetland rice Oryza glaberrima Upland rice Oryza punctata	Rice fields Cereal crops Rice	

## Relation-based data model for a term-based thesaurus

EN : Oryza	BT ( subclassOf ) : Poaceae
	NT ( hasSubclass ) : Oryza sativa
	NT ( hasSubclass ) : Oryza perennis
	NT ( hasSubclass ) : Oryza rufipogon
	NT ( hasSubclass ) : Oryza longistaminata
	NT ( hasSubclass ) : Wetland rice
	NT ( hasSubclass ) : Oryza glaberrima
	NT ( hasSubclass ) : Upland rice
	NT ( hasSubclass ) : Oryza punctata
	RT : Rice fields
	RT : Cereal crops
	RT : Rice
EN : Rice	BT ( subclassOf ) : Cereals
	NT ( hasSubclass ) : Broken rice
	NT ( hasSubclass ) : Basmati rice
	RT : Rice straw
	RT : Oryza
	RT : Rice flour
	UF : Paddy

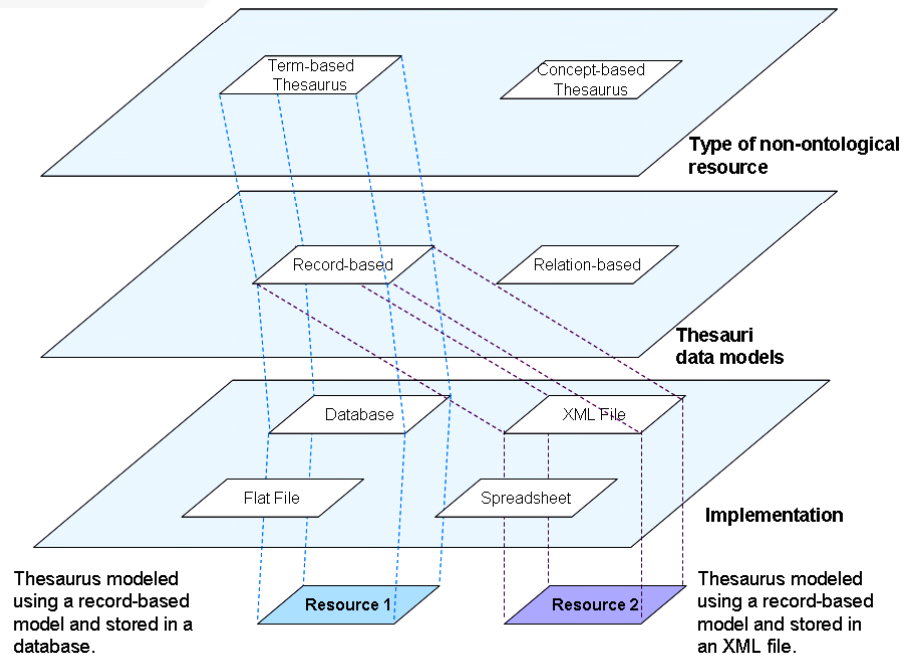
(1) Term Entity	
TermCode	Term
1001	Term1
1002	Term2
1003	Term3
1004	Term4
1005	Term5

(2) Term-Term Relationship Entity		
TermCode1	TermCode2	RelID
1001	1003	10
1003	1004	20
1002	1005	10
1003	1005	30

(3) Relationship Entity		
RelID	RelDesc	RelAbr
10	Broader Term	BT
30	Related Term	RT
20	Used For	UF

# Patterns for Re-engineering Thesauri

## Example – FAO Agrovoc thesaurus



### Term-based Thesaurus

EN : Oryza	BT ( subclassOf ) : Poaceae
	NT ( hasSubclass ) : Oryza sativa
	NT ( hasSubclass ) : Oryza perennis
	NT ( hasSubclass ) : Oryza rufipogon
	NT ( hasSubclass ) : Oryza longistaminata
	NT ( hasSubclass ) : Wetland rice
	NT ( hasSubclass ) : Oryza glaberrima
	NT ( hasSubclass ) : Upland rice
	NT ( hasSubclass ) : Oryza punctata
	RT : Rice fields
	RT : Cereal crops
	RT : Rice
EN : Rice	BT ( subclassOf ) : Cereals
	NT ( hasSubclass ) : Broken rice
	NT ( hasSubclass ) : Basmati rice
	RT : Rice straw
	RT : Oryza
	RT : Rice flour
	UF : Paddy

### Record-based model

Term	BT	NT	RT	UF
Rice	Cereals	Broken rice Basmati rice	Rice straw Oryza	Paddy
Oryza	Poaceae	Oryza sativa Oryza perennis Oryza rufipogon Oryza longistaminata Wetland rice Oryza glaberrima Upland rice Oryza punctata	Rice fields Cereal crops Rice	

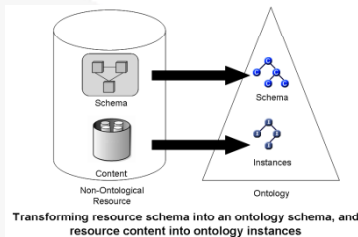
### XML

```

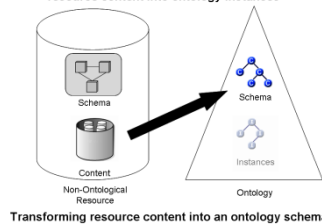
<thesaurus>
  <terms>
    <term name="Rice">
      <broaderTerm>Cereals</broaderTerm>
      <narrowerTerm>Broken rice</narrowerTerm>
      <narrowerTerm>Basmati rice</narrowerTerm>
      <relatedTerm>Rice straw</relatedTerm>
      <relatedTerm>Oryza</relatedTerm>
      <usedFor>Paddy</usedFor>
    </term>
    <term name="Oryza">
      <broaderTerm>Poaceae</broaderTerm>
      ...
    </term>
    ...
  </terms>
</thesaurus>
  
```

# Patterns for Re-engineering Thesauri into Ontologies

N	Identifier	Type of NOR	NOR Data Model	Transformation approach	Target
1	PR-NOR-TSLO-01	Term-based thesaurus	Record-based	TBox	Ontology Schema
2	PR-NOR-TSLO-02	Term-based thesaurus	Relation-based	TBox	Ontology Schema
3	PR-NOR-TSLO-11	Term-based thesaurus	Record-based	ABox	Ontology
4	PR-NOR-TSLO-12	Term-based thesaurus	Relation-based	ABox	Ontology



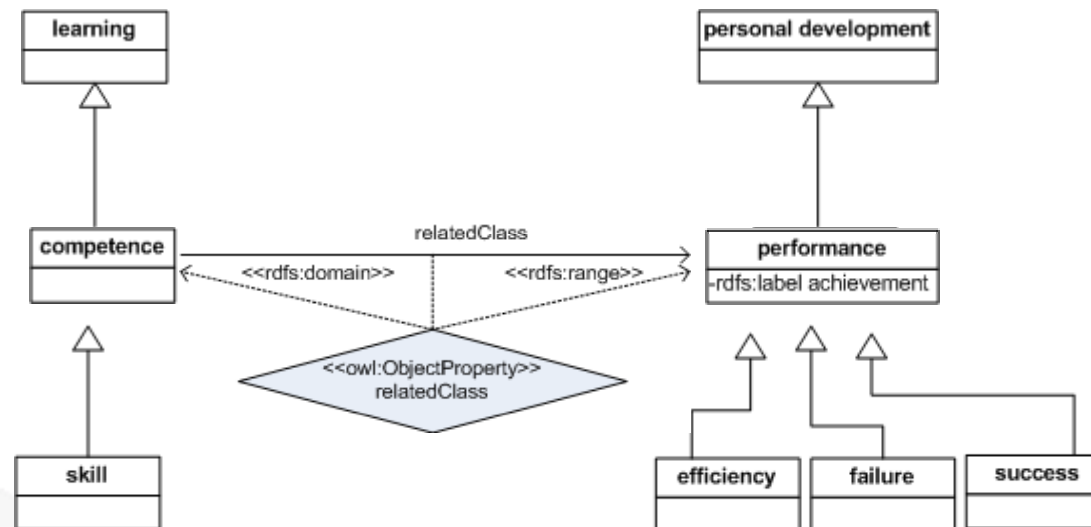
ABox transformation



TBox transformation

Pattern for re-engineering a term-based thesaurus, which follows the record-based model, to an ontology schema

Term	BT	NT	RT	UF
competence	learning	skill	aptitude know how knowledge performance	
performance	personal development	efficiency failure success	competence productivity	achievement



- Introduction and motivation
- Non-Ontological Resources (NORs)
- Method for Reusing NORs
- Pattern based Re-engineering Method
- Patterns for Re-engineering Classification Schemes
- Patterns for Re-engineering Thesauri
- **Patterns for Re-engineering Lexica**
- Technological support

- Lexica
  - Components
  - Data Models
  - Example
- Patterns for Re-engineering Lexica into Ontologies

- A lexicon<sup>1</sup> is a list of words in a language (a vocabulary) along with some knowledge of how to use each word. A lexicon may be general or domain-specific; we might have, for example, a lexicon of several thousand common words of English or German, or a lexicon of the technical terms of dentistry in some language. An example is WordNet<sup>2</sup>

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

### Noun

- {09411430} [S: \(n\) river](#) (a large natural stream of water (larger than a creek)) *"the river was navigable for 50 miles"*
  - [part meronym](#)
    - {09274500} [S: \(n\) estuary](#) (the wide part of a river where it nears the sea; fresh and salt water mix)
    - {09405396} [S: \(n\) rapid](#) (a part of a river where the current is very fast)
    - {09475292} [S: \(n\) waterfall, falls](#) (a steep descent of the water of a river)
  - [domain term category](#)
  - [has instance](#)
  - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
    - {09448361} [S: \(n\) stream, watercourse](#) (a natural body of running water flowing on or under the earth)
  - [part holonym](#)
    - {09476011} [S: \(n\) water system](#) (a river and all of its tributaries)

[WordNet home page](#)

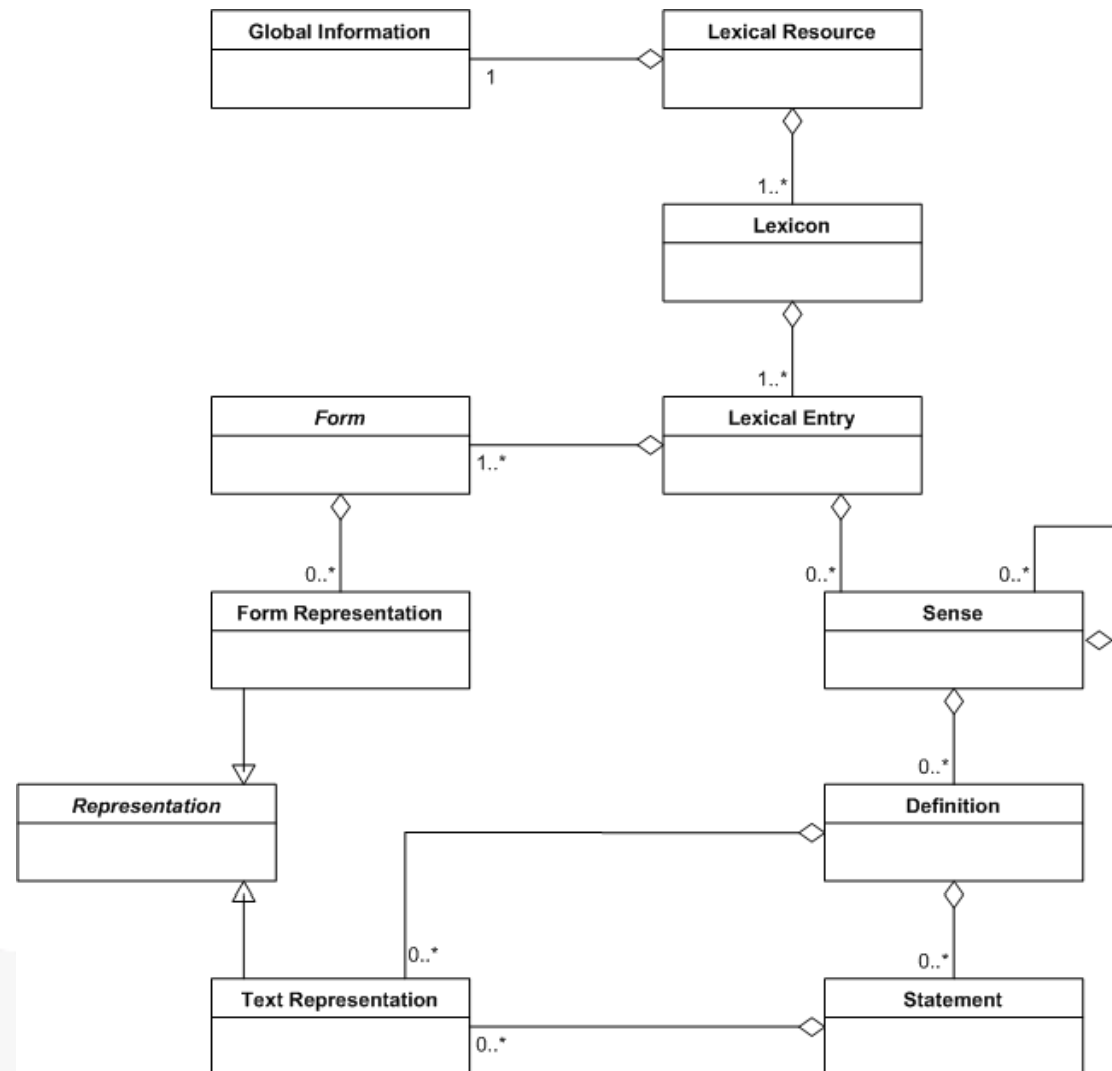


1. G. Hirst. Ontology and the lexicon. In Handbook on Ontologies in Information Systems, pages 209–230. Springer, 2004.
2. <http://www.wordnet-online.com>



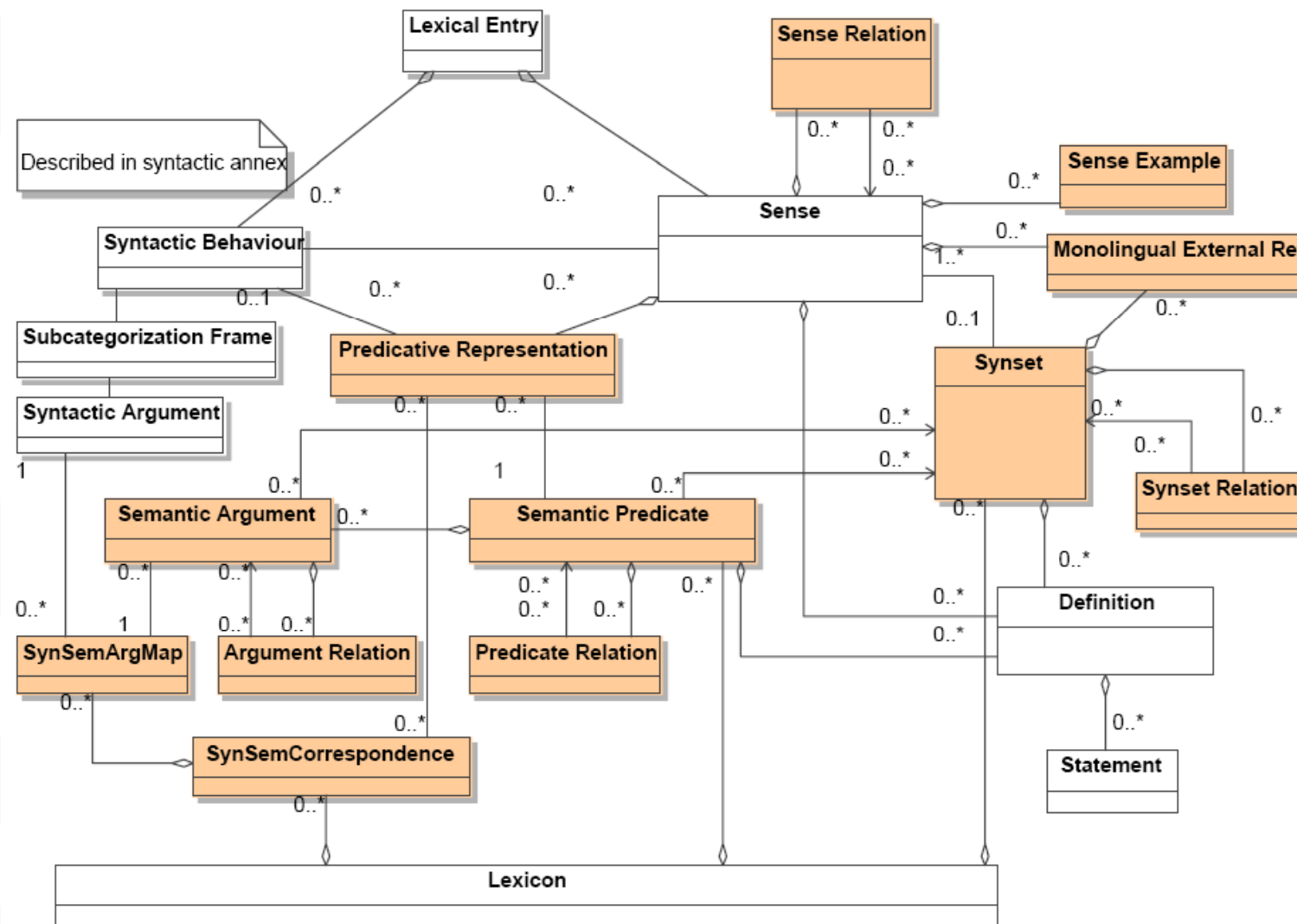
# Patterns for Re-engineering Lexica

## Lexical Markup Framework (LMF)



1G. Francopoulo, M. George, N. Calzolari, M. Monachini, N. Bel, M. Pet, and Claudia Soria. Lexical markup framework (lmf). In Proceedings of the fifth international conference on Language Resources and Evaluation, LREC 2006, Genoa, Italy, 2006.

# WordNet-LMF – Semantic Extension Package



1. Soria, C., Monachini, M., and Vossen, P. 2009. Wordnet-LMF: fleshing out a standardized format for wordnet interoperability. In Proceeding of the 2009 international Workshop on intercultural Collaboration (Palo Alto, California, USA, February 20 - 21, 2009). IWIC '09. ACM, New York, NY, 139-146

# Lexicon data models

- Record-based data model

Word	Gloss	POS	Part Meronym	Part Holonym	Hypernym	Hyponym	...
river	a large natural stream of water (larger than a creek); "the river was navigable for 50 miles"	N	estuary rapid waterfall	water system	stream		
...	...						

- Relation-based data model

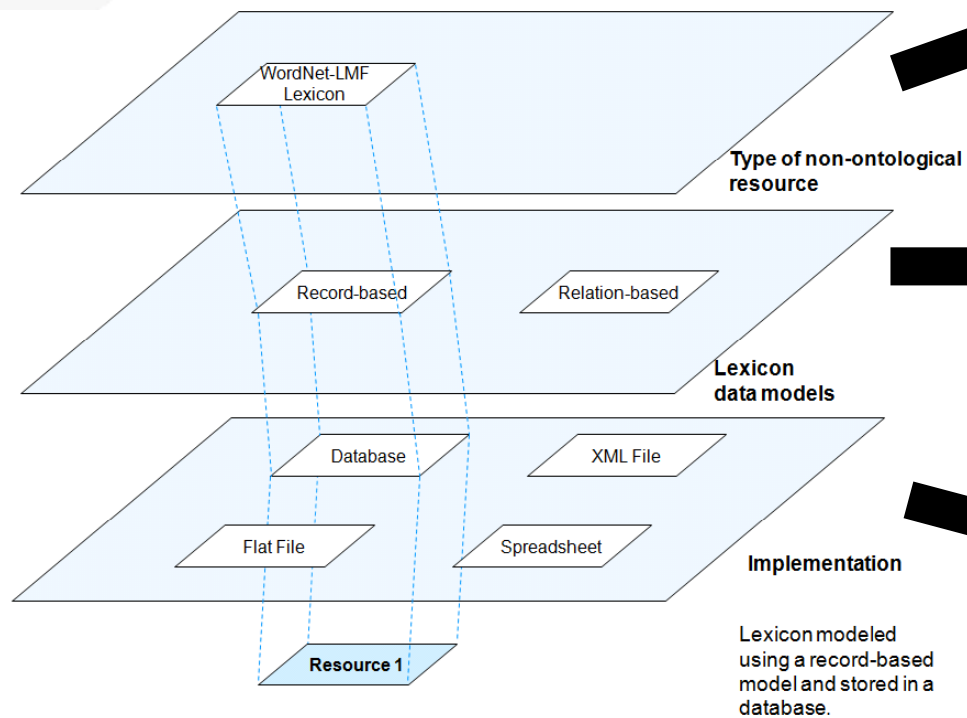
Synsetid	Word	POS	Gloss	...
108614198	river	n	a large natural stream of water (larger than a creek); "the river was navigable for 50 miles"	...
108814882	rapid	n	a part of a river where the current is very fast	...
108696219	stuary	n	the wide part of a river where it meets the sea; fresh and salt water mix	...
108854154	stream	n	a natural body of running water flowing on or under the earth	...
...	...	...	...	...

Synset1id	Synset2id	Linkid
108614198	108696219	11
108614198	108854154	1
...	...	...

Linkid	Link
1	hypernym
11	part holonym
12	part meronym
...	...

# Patterns for Re-engineering Lexica

## Lexicon Example



### WordNet-LMF Lexicon

Key: "S;" = Show Synset (semantic) relations, "W;" = Show Word (lexical) relations

#### Noun

- (09411430) [S; \(n\) river](#) (a large natural stream of water (larger than a creek)) *"the river was navigable for 50 miles"*
  - [part meronym](#)
    - (09274500) [S; \(n\) estuary](#) (the wide part of a river where it nears the sea, fresh and salt water mix)
    - (09405396) [S; \(n\) rapid](#) (a part of a river where the current is very fast)
    - (09475292) [S; \(n\) waterfall, falls](#) (a steep descent of the water of a river)
  - [domain term category](#)
  - [has instance](#)
  - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
    - (09448361) [S; \(n\) stream, watercourse](#) (a natural body of running water flowing on or under the earth)
  - [part holonym](#)
    - (09476011) [S; \(n\) water system](#) (a river and all of its tributaries)

[WordNet home page](#)

### Record-based model

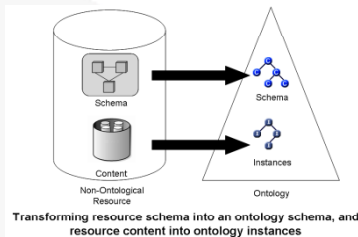
Word	Gloss	POS	Part Meronym	Part Holonym	Hypernym	Hyponym	...
river	a large natural stream of water (larger than a creek); "the river was navigable for 50 miles"	N	estuary rapid waterfall	water system	stream		
...							

### Database

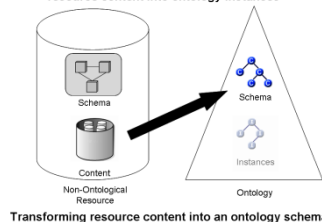
linkid	link	recurses
92	domain member category	0
1	hypernym	1
4	instance hyponym	1
80	pertainym	0
50	also	0
93	domain region	0
30	antonym	0
40	similar	0
70	verb group	0
21	entail	1
15	substance holonym	1
14	member meronym	1
81	derivation	0
11	part holonym	1
96	domain usage	0
12	part meronym	1
98	member	0
13	member holonym	1
94	domain member region	0
16	substance meronym	1
97	domain	0
2	hyponym	1
3	instance hypernym	1
96	domain member usage	0
23	cause	1
60	attribute	0
71	participle	0
91	domain category	0

# Patterns for Re-engineering Lexica into Ontologies

N	Identifier	Type of NOR	NOR Data Model	Transformation approach	Target
1	PR-NOR-LXLO-01	WordNet-LMF Lexicon	Record-based	TBox	Ontology Schema
2	PR-NOR-LXLO-02	WordNet-LMF Lexicon	Relation-based	TBox	Ontology Schema
3	PR-NOR-LXLO-10	WordNet-LMF Lexicon	Record-based	ABox	Ontology
4	PR-NOR-LXLO-11	WordNet-LMF Lexicon	Relation-based	ABox	Ontology

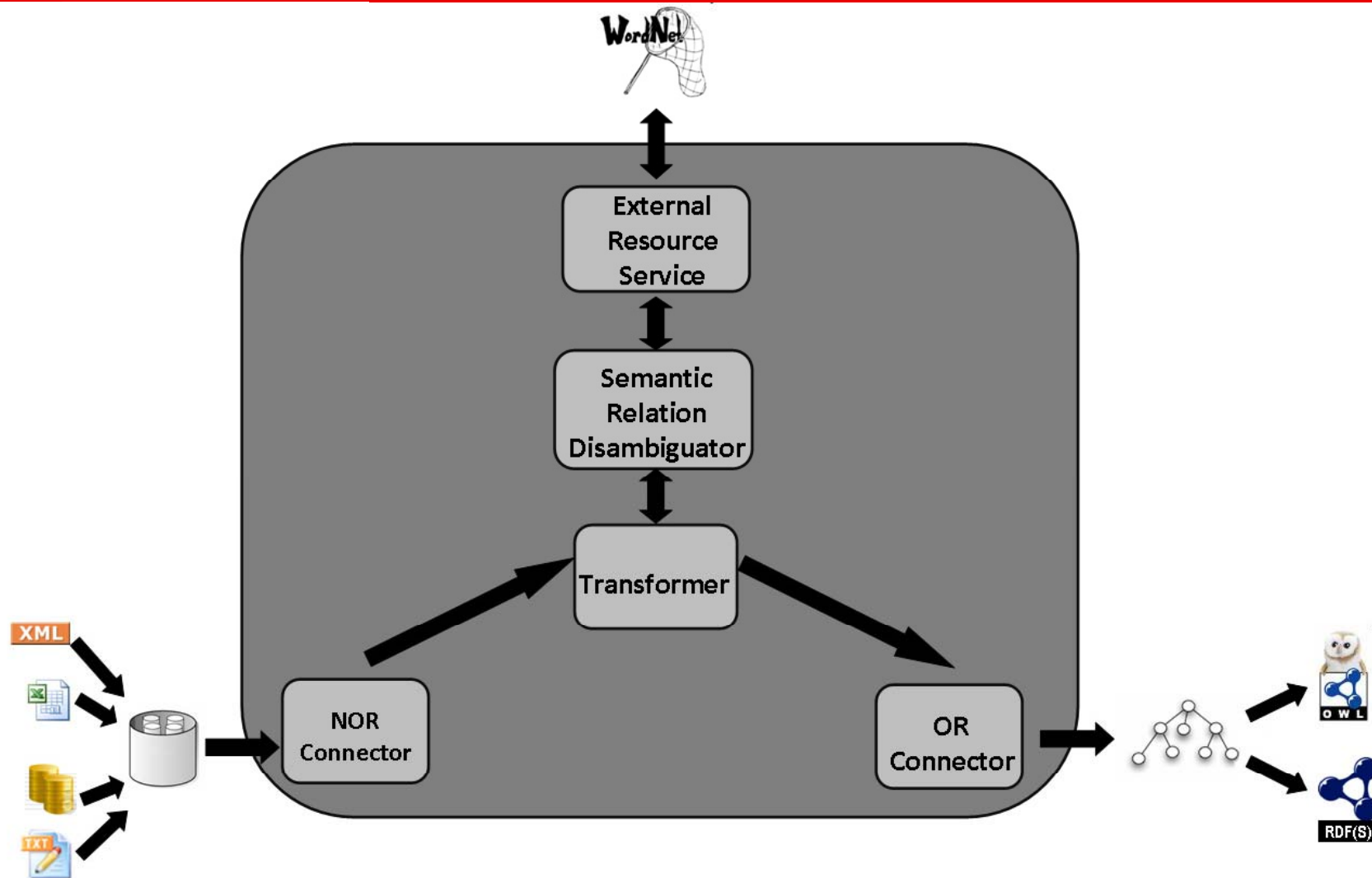


ABox transformation



TBox transformation


- Introduction and motivation
- Non-Ontological Resources (NORs)
- Method for Reusing NORs
- Pattern based Re-engineering Method
- Patterns for Re-engineering Classification Schemes
- Patterns for Re-engineering Thesauri
- Patterns for Re-engineering Lexica
- **Technological support**



<http://mccarthy.dia.fi.upm.es/nor2o/>

# Technological support

## PR-NOR library at the ODP Portal



navigation

- Main page
- List patterns
- Pattern types
- Modeling Issues
- Domains
- Training
- Events

contribute

- Submit a pattern
- Submit an exemplary ontology
- Post a modeling issue
- Review a pattern
- Feedback about the portal
- Request an ODP account

help

- About ODP
- What is a pattern?
- What is an exemplary ontology?
- How to post a pattern
- Training

catalogues

submissions:reengineeringodps [discussion](#) [view source](#) [history](#)


### Submissions:ReengineeringODPs

Below you find the currently proposed Re-engineering OPs (RPs).  
New proposals of RPs are very welcome. Please [post a new proposal](#) if you want to contribute.

#### Proposed Re-engineering ODPs

submissions:pattern for re-engineering a term-based thesaurus, which follows the recordbased data model, into an ontology schema [discussion](#) [view form](#) [history](#)

### Submissions:Pattern for re-engineering a term-based thesaurus, which follows the recordbased data model, into an ontology schema



If you are a member of [quality committee](#) please visit the [evaluation section](#)

If you are author of this proposal or you want to contribute to this pattern's review, you can:

- ask for a review [post your open review](#)
- specify if this revision takes in account any of the review(s) [add a new scenario](#) for Pattern for re-engineering a term-based thesaurus, which follows the recordbased data model, into an ontology schema

In general, it could be useful to visit the [evaluation section](#) to have informations about the evaluation process of this proposal

Current revision ID: **8956**

#### General information

<b>Name</b>	Pattern for re-engineering a term-based thesaurus, which follows the recordbased data model, into an ontology schema
<b>Problem</b>	Re-engineering a term-based thesaurus which follows the record-based model to design an ontology schema.

#### Non-Ontological Resource

<b>Description</b>	<p>A non-ontological resource holds a term-based thesaurus which follows the record-based model.</p> <p>A thesaurus represents the knowledge of a domain with a collection of terms and a limited set of relations between them. The record-based data model is a denormalized structure, uses a record for every term with the information about the term, such as synonyms, broader, narrower and</p>
--------------------	---

<http://ontologydesignpatterns.org/wiki/Submissions:ReengineeringODPs>







**ATHENS Course: Ontologies and the Semantic Web**

# **A Method for Reusing and Re-engineering Non-Ontological Resources for Building Ontologies**

Boris Villazón-Terrazas

[bvillazon@fi.upm.es](mailto:bvillazon@fi.upm.es)

Ontology Engineering Group. Laboratorio de Inteligencia Artificial

Departamento de Inteligencia Artificial

Facultad de Informática

Universidad Politécnica de Madrid