



A Method for Re-engineering Non-Ontological Resources for Building Ontologies

Boris Villazón-Terrazas, Asunción Gómez-Pérez

{bvillazon,asun}@fi.upm.es

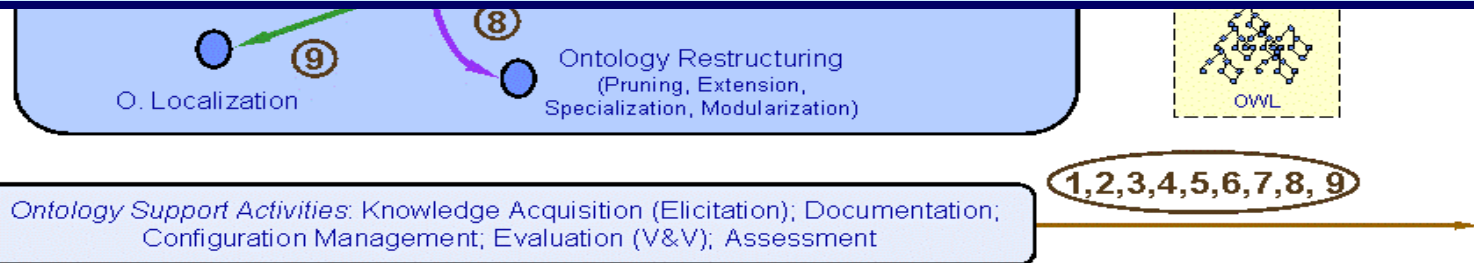
Ontology Engineering Group. Laboratorio de Inteligencia Artificial

Departamento de Inteligencia Artificial

Facultad de Informática

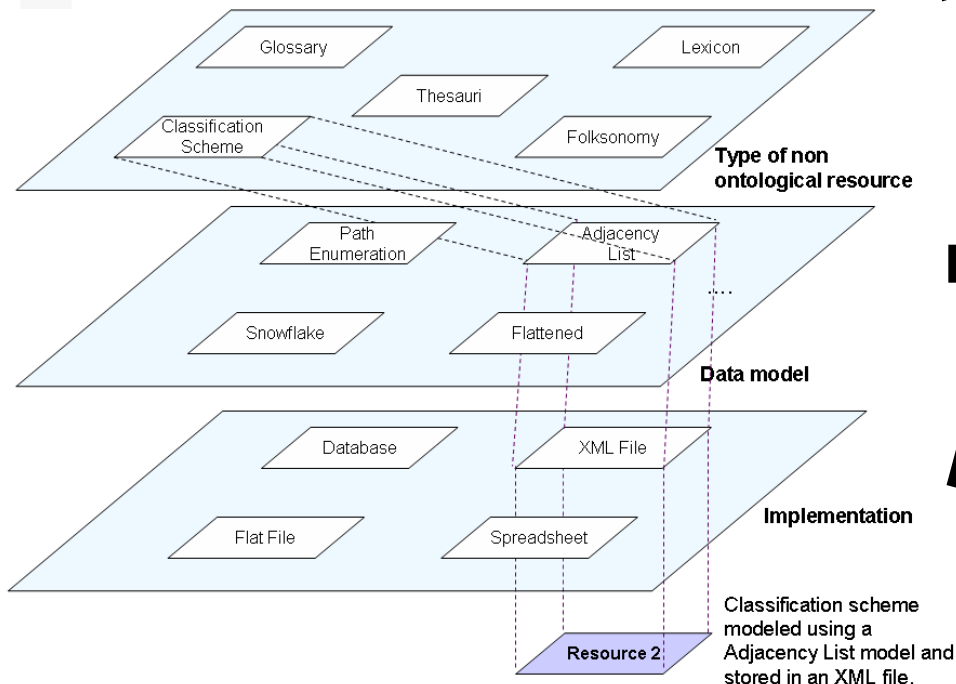
Universidad Politécnica de Madrid

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

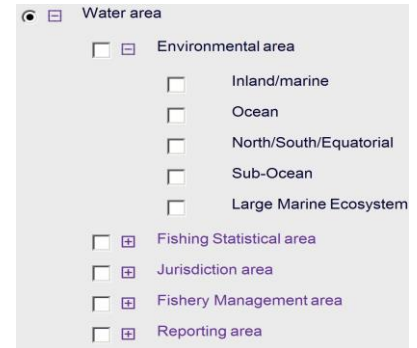


Classification Scheme - Example: FAO Water Areas

- A classification scheme¹ 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 scheme².



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>
```

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 Data Models

- 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

Diagram illustrating the Path Enumeration Data Model. The table shows the path from the root to each node. Brackets below the ID column indicate the hierarchy levels: 1 (Water area), 2 (Environmental area, Jurisdiction area, Fishing Statistical area), and 3 (Inland/marine, Ocean, North/South/Equatorial, FAO statistical area, Areal grid system).

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

First Level		
ID	CSI_Name	
20000	Water area	

Second Level		
ID	First Level ID	CSI_Name
21000	20000	Environmental area
24020	20000	Jurisdiction area
22000	20000	Fishing Statistical area

Third Level		
ID	Second Level ID	CSI_Name
21001	21000	Inland/marine
21002	21000	Ocean
21003	21000	North/South/Equatorial
22001	22000	FAO statistical area
22002	22000	Areal grid system

Diagram illustrating the Snowflake Data Model. The hierarchy is represented by three tables: First Level, Second Level, and Third Level. Each table has a column linked to its parent node.

- 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

Diagram illustrating the Adjacency List Data Model. The table shows the hierarchy with a Parent column. Arrows indicate the linking from parent nodes to child nodes.

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

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

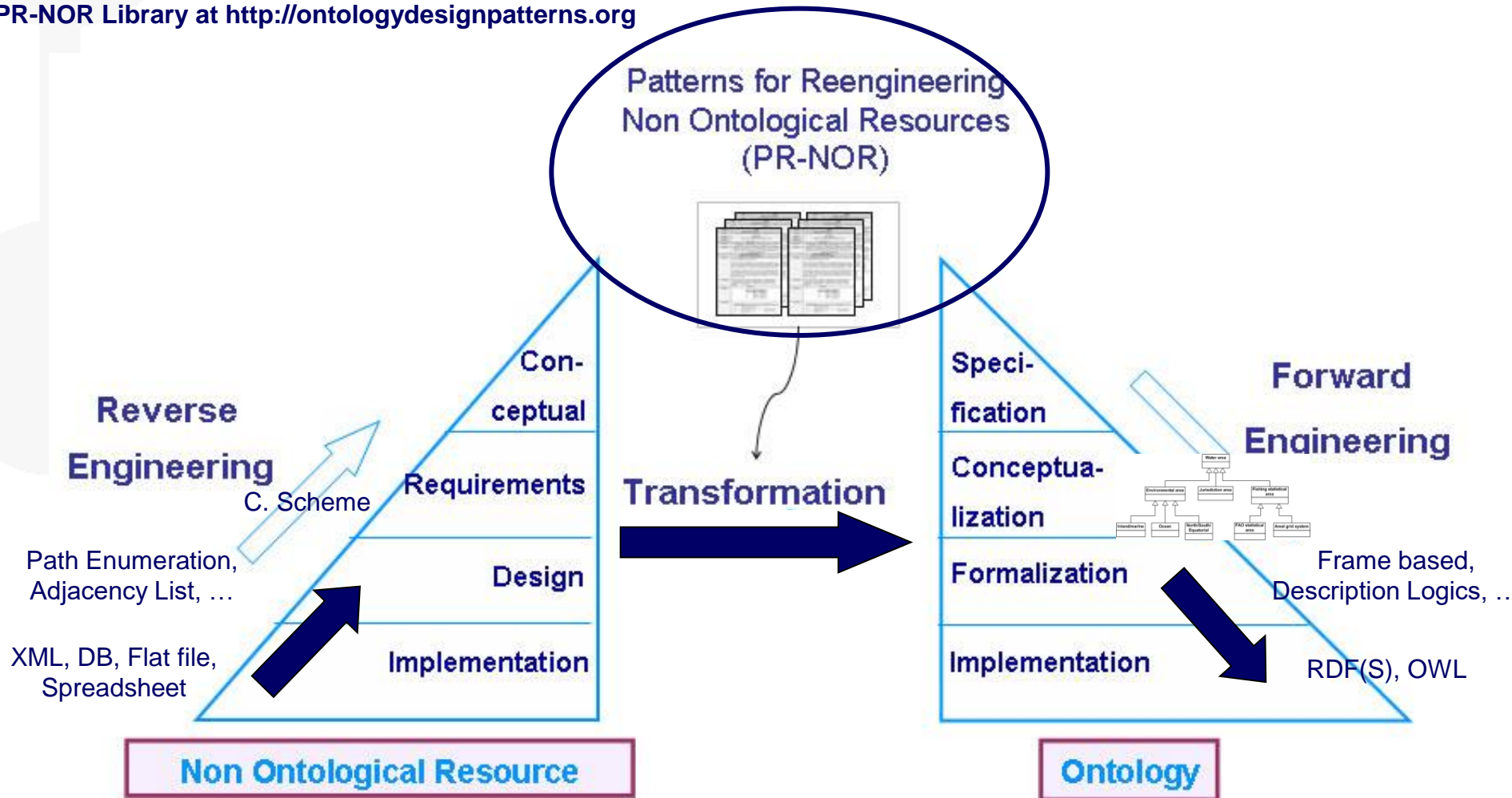
Diagram illustrating the Flattened Data Model. The hierarchy is represented by a single table where each hierarchy level is stored on a different column.

Non-Ontological Resource Re-engineering

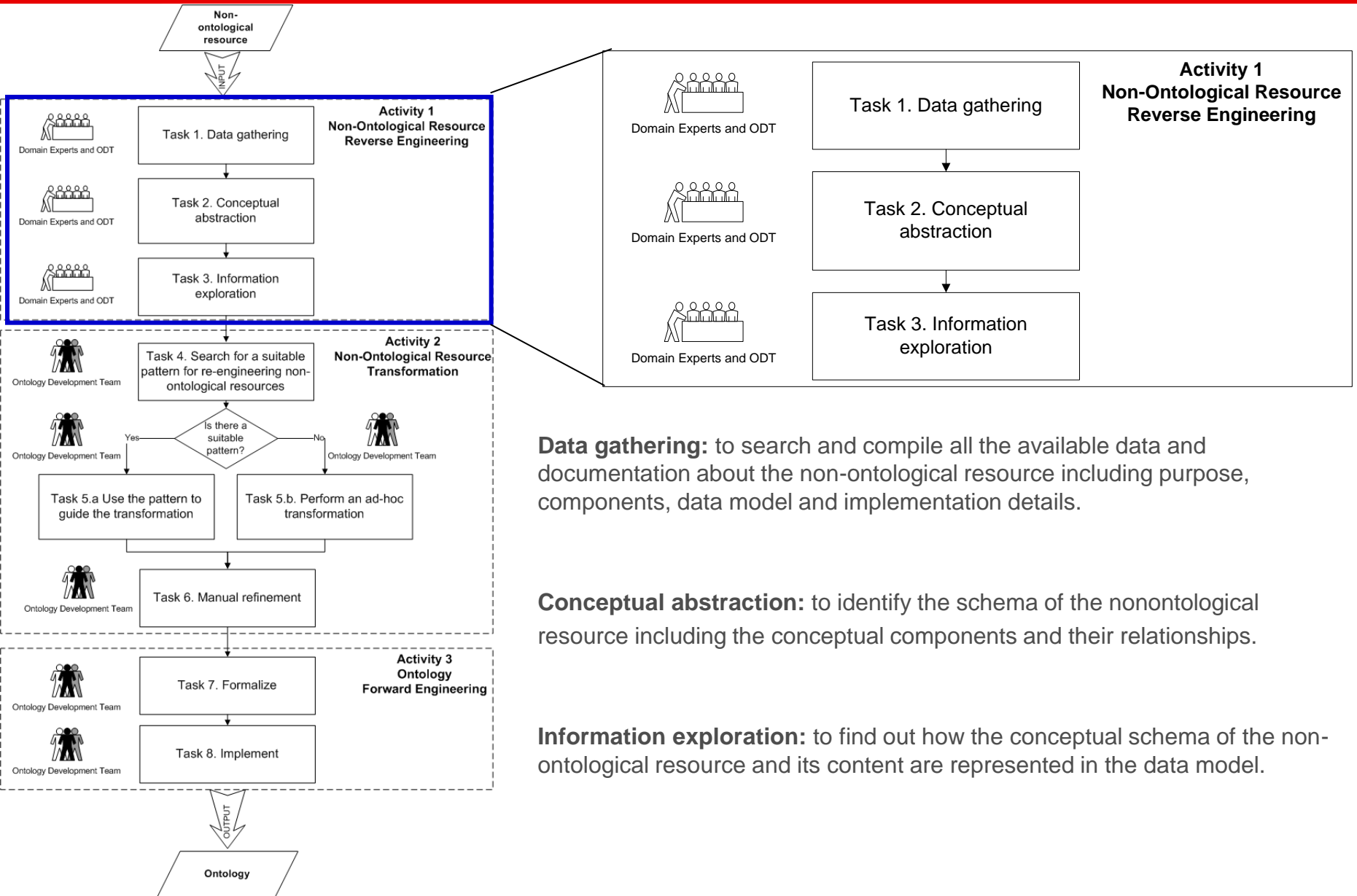
Patterns for re-engineering non-ontological Resources (PR-NOR) define a procedure to transform knowledge-aware resources into ontologies taking into account the resource type (thesaurus, classification scheme, etc.) and their underlying resource data model. For every type of resource and data model, the NeOn method defines a pattern with a well-defined sequence of activities. Our patterns perform the following transformations:

- TBox transformation: for transforming the resource content into an ontology schema.
- ABox transformation: for transforming the resource schema into an ontology schema, and the resource content, into ontology instances.

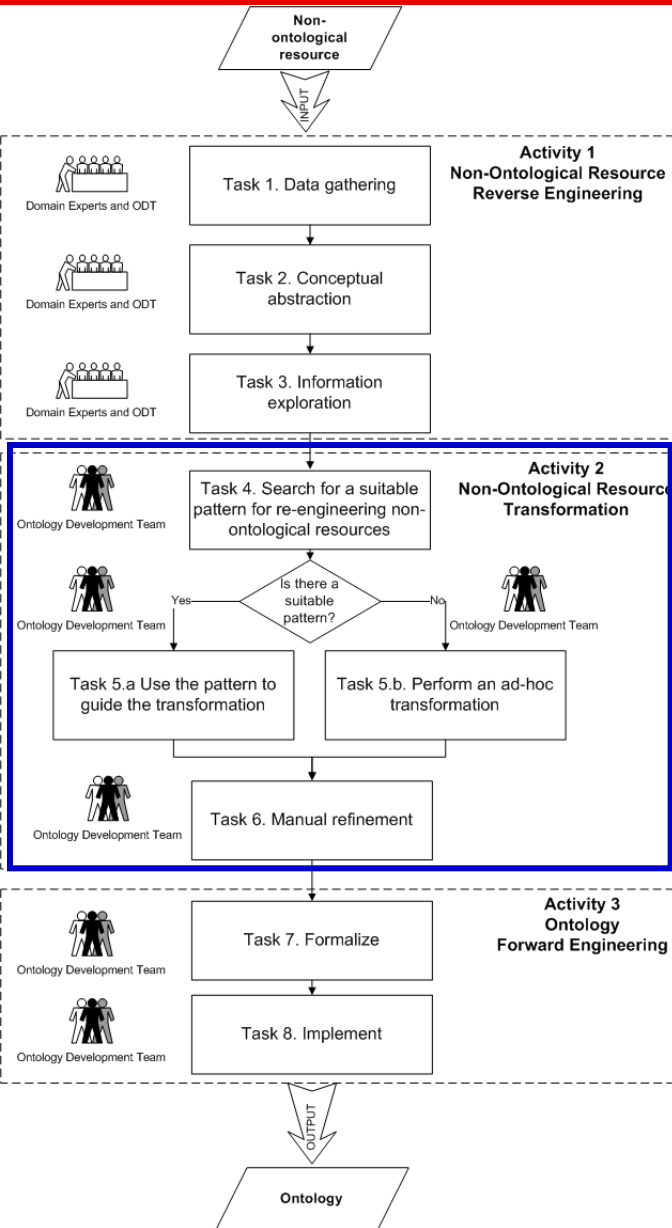
PR-NOR Library at <http://ontologydesignpatterns.org>



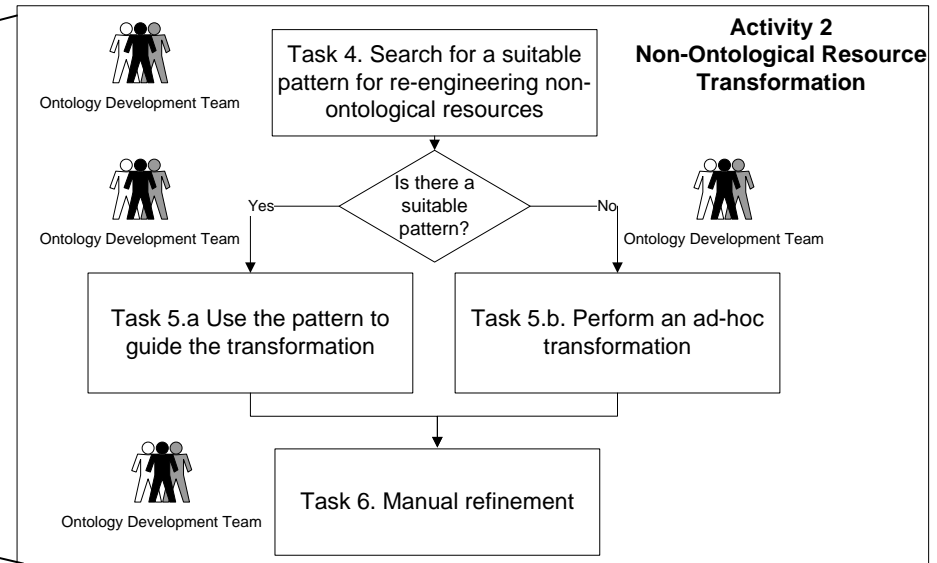
Non-Ontological Resource Re-engineering (II)



Non-Ontological Resource Re-engineering (III)



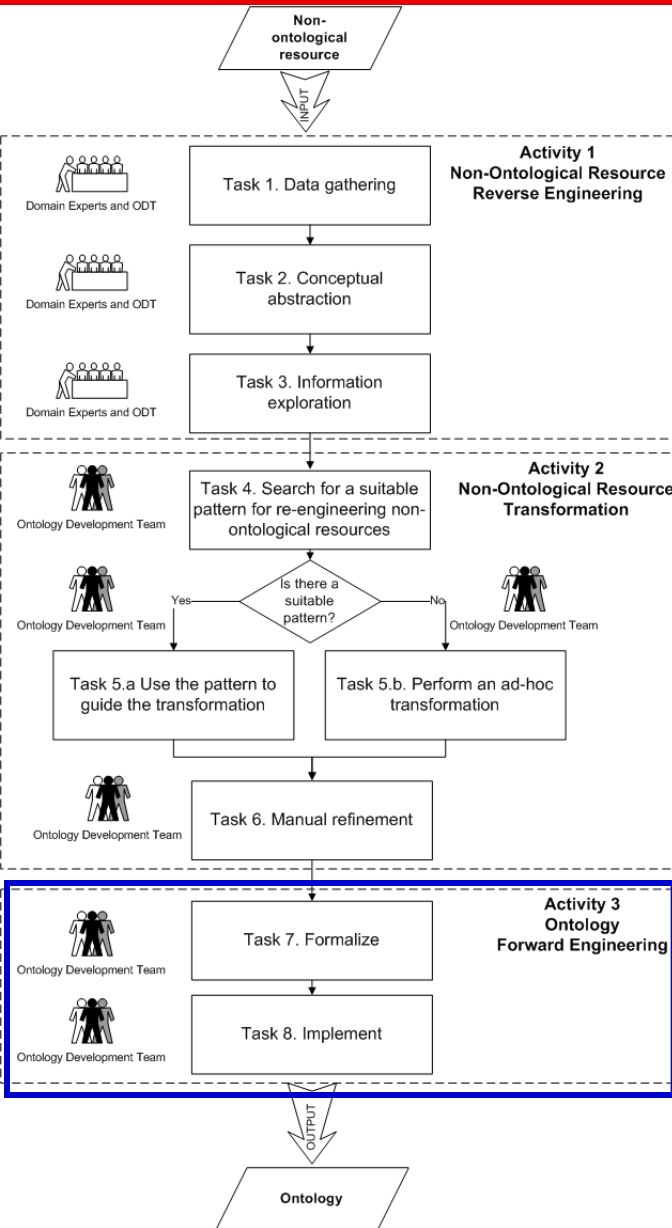
The goal is to generate a conceptual model from the NOR



Criteria for searching for a suitable pattern:

- **NOR Type:** classification scheme, thesauri, lexicon, glossary, or folksonomy
- **Data model:** C.Scheme (path enumeration, adjacency list, snowflake, or flattened), Thesaurus (record-based, relation-based)
- **Transformation approach:** TBox, ABox, or population

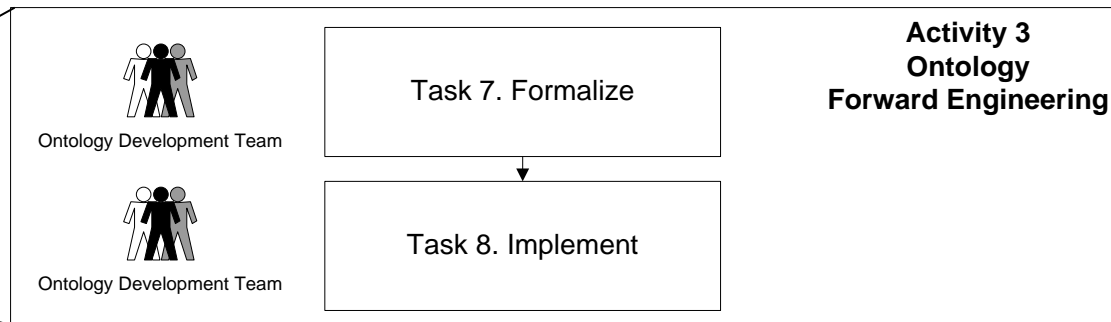
Non-Ontological Resource Re-engineering (IV)



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

Formalize: to transform the conceptual model obtained in task 6 into a formalized model, according to a knowledge representation paradigm as description logics, first order logic, etc.

Implement: to implement in an ontology language.



Ontologies	5
Concepts	1612
Attributes	120
Relations	97
Instances	1674

- NORs pertaining to the domain of human resources were transformed into ontologies.
- We re-engineered classification schemes using the set of patterns.

ISCO-88 (COM)
International Standard Classification
of Occupations
(for European Union purposes)



FOET

Classification of fields
of education and
training



NACE

Statistical Classification of
Economic Activities in the
European Community



ISO 3166

English country names
and code elements



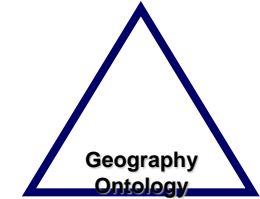
*Pattern for re-engineering a
classification scheme modelled
with a Path Enumeration Data Model*



*Pattern for re-engineering a
classification scheme modelled
with a Snowflake Data Model*

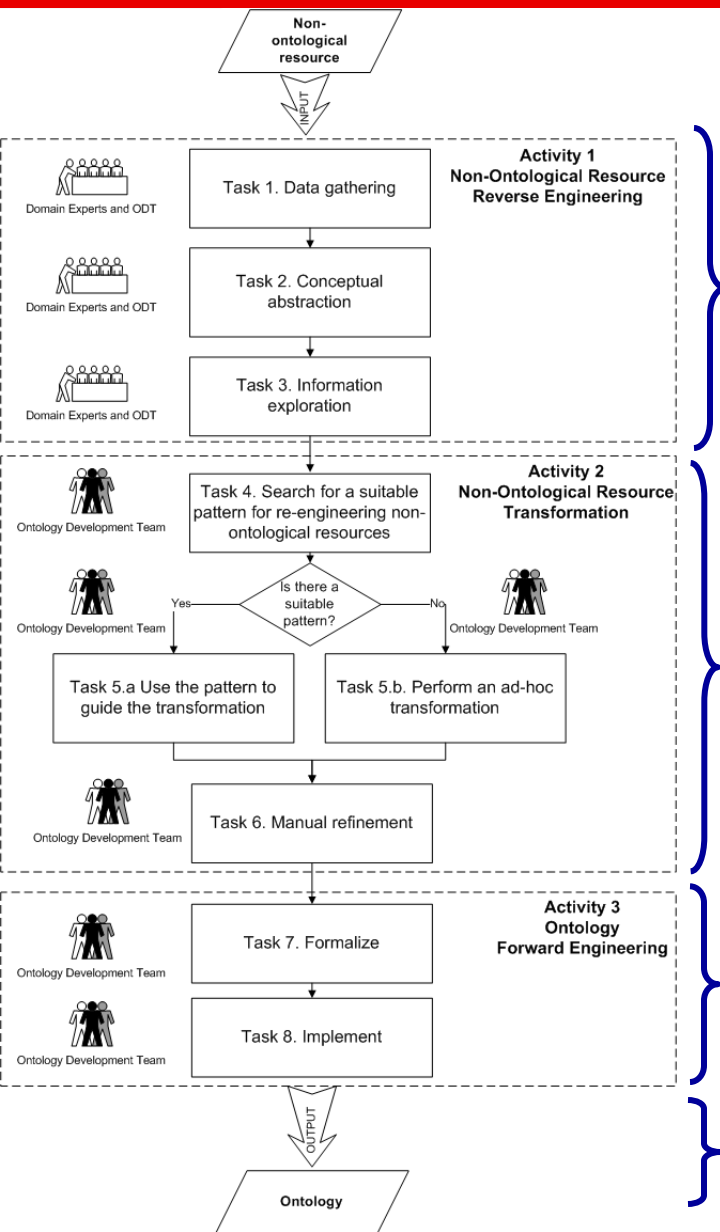


*Pattern for re-engineering a
classification scheme modelled
with an Adjacency List Data Model*



ISTAT
Italian Geography
Standard





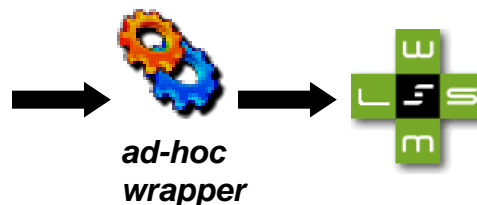
Adjacency list

Criteria for searching a suitable pattern:

- NOR Type: **classification scheme**
- Data model: **adjacency list**
- Transformation Approach: **TBox**



Pattern for re-engineering a classification scheme modelled with an Adjacency List Data Model PR-NOR-CLTX-02





A Method for Re-engineering Non-Ontological Resources for Building Ontologies

Boris Villazón-Terrazas, Asunción Gómez-Pérez

{bvillazon,asun}@fi.upm.es

Ontology Engineering Group. Laboratorio de Inteligencia Artificial

Departamento de Inteligencia Artificial

Facultad de Informática

Universidad Politécnica de Madrid