Development of knowledge-based web services to promote and advance Industrial Symbiosis in Europe (**eSymbiosis**)

LIFE09/ENV/GR/000300

ACTION 2: Service architecture and implementation

D2.1 Waste stream and solution provider knowledge models, ontologies

November 2012

INDEX

INDEX 2

TABLES INDEX 4

FIGURES INDEX 5

1. INTRODUCTION 7

1.1. INDUSTRIAL SYMBIOSIS AND SEMANTICS 7

1.2. THE ROLE OF ONTOLOGIES IN ESYMBIOSIS 8

1.3. INDUSTRIAL SYMBIOSIS PROPERTIES 8

2. ONTOLOGIES 10

2.1. WHAT IS AN ONTOLOGY 10

2.2. ONTOLOGY COMPONENTS 11

2.3. TYPES OF ONTOLOGIES 11

3. ONTOLOGICAL ENGINEERING 13

3.1. DESIGN PRINCIPLES 13

3.2. METHODOLOGIES 13

3.2.1. Cyc method 14

3.2.2. TOVE method 14

3.2.3. METHONTOLOGY 14

3.2.4. SENSUS method 15

3.2.5. On-To-Knowledge method 15

3.3. ONTOLOGY DEVELOPMENT TOOLS 16

3.3.1. Editors 16

3.3.2. Languages 17

4. DEVELOPING ESYMBIOSIS ONTOLOGY 19

4.1. THE PURPOSE OF THE ONTOLOGY 19

4.2. SELECTION OF THE TOOLS 19

4.3. METHODOLOGY 20

4.3.1. Determine the domain and the scope of the ontology 20

4.3.2. Consider reusing existing ontologies 21

4.3.3. Enumerate important terms in the ontology. 21

4.3.4. Define the classes and the class hierarchy 22

4.3.5. Define the properties of classes – slots 28

4.3.6. Define the facets of the slots 31

Owl:onProperty. *31*

Cardinalities *33*

4.3.7. Define the rules *34*

4.3.8. Other issues *34*

5. SEMANTIC SERVICE DESCRIPTION **35**

5.1. SEMANTIC WEB SERVICES AND OWL-S 35

5.2. ESYMBIOSIS SEMANTIC SERVICE DESCRIPTION BASED ON OWL-S 37

6. INPUT – OUTPUT MATCHING **40**

6.1. MATCHING METRICS 41

6.2. MATCHING METHODS BASED ON SEMANTIC WEB SERVICES 41

6.3. TYPES OF MATCHING 42

6.3.1. Direct Matching *42*

6.3.2. Decomposition matching *43*

6.4. THE LEVELS OF MATCHING 43

6.5. THE ALGORITHM: 45

6.5.1. Graph Modelling of the Ontology *45*

6.5.2. The Semantic Distance Measurement (DM) *47*

6.5.3. Property Matching based on Vector Space Modelling (PM) *49*

6.6. INTELLIGENCE EMPOWERED BY THE USE OF SEMANTICS: 50

6.6.1. Alternative Resource Suggestion *51*

6.6.2. Identify technologies/solutions for similar types of resource *51*

6.6.3. Different application for a resource *51*

6.6.4. Partial Matching *51*

6.6.5. Providing general Options *51*

7. TESTING **53**

REFERENCES **58**

TABLES INDEX

Table 1 Properties Required for I/O matching 9
Table 2 Top level concepts of the ontology 25
Table 3 Solutions (Top level) 27
Table 4 Instances of the ontology 28
Table 5 Object Properties 28
Table 6 Data type Properties 30
Table 7 Restrictions on properties 33
Table 8 Service Description Ontology additional data properties in Process.owl 38
Table 9 Service Description Ontology built-in properties in Profile.owl 39
Table 10 Service Description Ontology additional data properties in Service.owl 40
Table 11 Metrics Description 41
Table 12 Properties and restrictions used for modelling the ontology graph 46
Table 13 Distance Measurement semantic similarity between ontology concepts 49
Table 14 Data properties used in vector space matching 50
Table 15 EnviD ltd. solution information 53
Table 16 Hartex ltd resource information 53
Table 17 Profile of registered industries offering potential matches with EnviD ltd 54
Table 18 Complete set of results 56

FIGURES INDEX

Figure 1 Design of the IS ontology 22
Figure 2. Role subsumption. 23
Figure 3 The use of xml:lang tag 31
Figure 4 Restriction example 32
Figure 5 Inferred Axiom 34
Figure 6 OWL-S ontology Model 36
Figure 7 Semantic Web Service Profile Ontology 36
Figure 8 Service Description Ontology 38
Figure 9 Matching Example 42
Figure 10 Direct Matching 42
Figure 11 Resource decomposition matching 43
Figure 12 Levels of Matching 44
Figure 13 An example of restriction defined over a concept 47
Figure 14 eSymbiosis matching 47
Figure 15 Excerpt of the domain ontology used for input-output matching 55
Figure 16 Distance measurement example 55

Revision History

Revision	*Description*	*Date*	*Authors*	*Notes*
Draft v.0.1	Initial Version Document	12/2012	Franjo Cecelja, Aidong Yang, Tara Raafat, Nikolaos Trokanas	
v 1.0	Final Version	05/2013	Franjo Cecelja, Aidong Yang, Tara Raafat, Nikolaos Trokanas	Testing section revised.

1. INTRODUCTION

1.1. INDUSTRIAL SYMBIOSIS AND SEMANTICS

The current practice of IS relies on manual handling of extensive knowledge and information about waste, technologies, logistics and transportation. This information currently exists scattered in various formats such as databases and excel sheets or is simply possessed by the practitioners. As a result, practitioners do not have access to all the required information to make in-depth analysis and discover all possible synergies. The eSymbiosis platform aims to tackle this problem by automating the IS process. This is enabled by defining the knowledge about the IS domain and participating industries in a machine process-able format which is done by the use of semantic technologies specifically ontologies. Ontology in computer science is defined as a mean to model the knowledge within a domain by specifying common concepts and relationship between these concepts in that domain which are then enhanced with further logic to reflect their use and enable the extraction of new knowledge. The IS domain ontology provides a common set of vocabulary for industries to describe their waste or technology [N. Trokanas et al., 2011]. Furthermore, based upon the domain ontology the eSymbiosis platform uses the semantic web service technology to describe every participating industry. Semantic web services (SWS) are the merger between the semantic web and the web services technology and were introduced to add semantic annotation to service elements [S. Battle, 2005]. The annotations could then be processed by machines. This means that the data about every industry registered with the platform would be stored in a machine understandable format hence forming a comprehensive knowledge base which is flexible to update and can be used for automatic matching purposes. SWS also support dynamic systems with changing context. The use of semantics enable the process of input – output matching that allows not only to automate but also to discover partial matches. By partial, we mean possible symbiotic links that do not fulfill all the requirements and would otherwise be ignored. These links might consider different quantities, similar resources or other conditions. In a real IS environment new industries will be constantly joining the system, current participants will be dynamically changing their amount of resource production, their need for resource, its availability or their capability for processing resources. SWS can handle these changes. The eSymbiosis SWS description of participating industries is based on OWL-S [Martin et. Al, 2004] a well-known SWS framework. Within the OWL-S framework, the “Service Profile” provides a way to describe the services offered or required by the users. The eSymbiosis platform utilizes this approach to define both the resource or the technology provided by a certain industry, including all attributes related to them. The OWL_S profile has been modified and expanded to include all these attributes and to fulfil the requirements of the eSymbiosis platform (Figure 1). These attributes are further used as major factors to discover potential synergies.

1.2. The role of ontologies in eSymbiosis

The ontologies are used for the formalisation and the modelling of the knowledge of the technology and waste domains. Ontologies are flexible to implement, share and reuse, especially in conjunction with web technologies and applications. Ontologies are primarily used for knowledge representation in the form of a group of terms that describe a domain of knowledge organised in a hierarchical structure and enhanced with properties/attributes of these concepts and restrictions on these properties. An ontology is associated with instances which represent specific objects of the domain. During the process of registration ontologies navigate the user and indicate requests for data. In more detail, the ontologies provide the path that the user follows during the registration process by parsing the taxonomy of the resources and technologies. During this process, the user is asked to fill in data for properties that are attached to the concepts (data and object properties).

The use of ontologies in the Ontology Web language (OWL) format facilitates the use of synonyms (hence removing the jargon barrier), sharing and reasoning that can automatically generate new knowledge. Ontologies also provide a standardised vocabulary for the given domain and help in eliminating any syntactic issues. Moreover, ontologies allow to informally express the given knowledge which makes them attractive to all audiences no matter the level of coding skills and background. (Trokanas et al. 2012).

1.3. Industrial Symbiosis Properties

The synergy identification and the formation of symbiotic links is reffered to as input – output matching. In order to facilitate this process it is important to identify the properties that describe the resources and the solutions and also affect the potential formation of a symbiotic link. These properties are defined in Table 1.

Table 1 Properties Required for I/O matching

		Property	Description
Input – output matching	Characteristics	Type of input/output	The preferred type of inputs and the associated outputs.
		Quantity/Capacity	The amount of feedstock required by the technology.
		By-Product	The by-products (waste, energy) produced by the technology.
		Availability	The period of time that the solution/resource is available for use.
		Hazardousness	Whether a material is hazardous.
		Pattern of supply/demand	Batch or continuous operation of the solution or batch or continuous production of the resource.
	General Information	Location	The geographic location of the industry.
		Industry Category	The industrial activity.
		Storage Capacity	The available storage.
		Storage Method	Current storage method of the input/output.
		Delivery Method	Current delivery method of the input/output.
	Pre-conditions	Energy Required	The type and amount of energy required.
		Water Required	The amount of water required.

2. ONTOLOGIES

2.1. WHAT IS AN ONTOLOGY

Several definitions exist for ontology. Originating from Philosophy, meaning the explanation (*λόγος - logos*) of being (*ον - on*), today it is used in computer science and knowledge engineering. The most common and most quoted one is the one by Gruber (1993a) who defined ontology as:

"an explicit specification of a conceptualisation"

The term conceptualization indicates that ontology is a simplified view or representation of a given domain of interest. (Sharman et al. 2006)

Guarino and Giaretta (1995) identified several different notions for the term ontology:

1. *Ontology as a philosophical discipline*
2. *Ontology as a an informal conceptual system*
3. *Ontology as a formal semantic account*
4. *Ontology as a specification of a conceptualization*
5. *Ontology as a representation of a conceptual system via a logical theory*
    - *5.1 characterized by specific formal properties*
    - *5.2 characterized only by its specific purposes*
6. *Ontology as the vocabulary used by a logical theory*
7. *Ontology as a (meta-level) specification of a logical theory*

In the same paper the authors suggested another definition for ontology. They described it as:

"a logical theory which gives an explicit, partial account of a conceptualisation"

One of the most explanatory and easy to understand definitions was suggested by Neches and colleagues 1991 (Gómez-Péréz et al. 2004):

"An ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary"

It becomes obvious that the meaning of ontology is highly dependent on the point of view of the author as well as on the forthcoming use of the ontology. Definition, however, is not the most important aspect of the ontology. The respective use, the capabilities, the potentials, the challenges and the benefits related to it are far more important. In few words, an ontology, provides a common vocabulary (terms) for a domain of interest as well as the properties of those terms and the relation among them.

2.2. Ontology components

There are three main components in any ontology regardless the type of it (lightweight or heavyweight). These are the **classes** or concepts which are organised in taxonomies, the **relations** which represent the type of association between the concepts and **attributes** or properties or slots of those concepts. Other components identified in the literature are the **restrictions** on slots (also called facets), the **functions** which are some special cases of relations (Gómez-Péréz et al. 2004), **formal axioms** (Gruber 1993a) used to represent knowledge that is hard or impossible to represent using the other components and **instances** which represent elements or individuals of the ontology.

2.3. Types of ontologies

Similar to other aspect of ontologies the categorisation of them is highly dependent on the point of view and the forthcoming use of the ontology. Several different approaches of ontology classification exist in literature. Mizoguchi and colleagues (1995), proposed the following four types of ontologies according to the respective use of them (Gómez-Péréz et al. 2004).

1. *Content Ontologies used for knowledge reuse. This category includes task, domain and general ontologies.*
2. *Communication Ontologies used for knowledge sharing.*
3. *Indexing Ontologies used for case retrieval.*
4. *Meta-ontologies also known as Knowledge Representation Ontologies.*

Van Heijst and colleagues (1997) categorized ontologies towards two dimensions. The first was the amount and the type of the structure and the second was the issue of conceptualization. According to the first they proposed the following classification (Gómez-Péréz et al. 2004):

1. *Terminological Ontologies which specify the terms used for knowledge representation.*
2. *Information Ontologies which specify storage structure data.*
3. *Knowledge Modeling Ontologies which specify the conceptualization of knowledge.*

According to the second dimension they proposed the following four categories:

1. *Representation ontologies*
2. *Generic ontologies*
3. *Domain ontologies*
4. *Application ontologies*

Another widely used categorisation is that by Guarino (1998) who categorized ontologies according their dependence on particular tasks or points of view. He identified the following four types of ontologies:

1. *Top-level Ontologies which are general ontologies.*
2. *Domain Ontologies which represent knowledge of a specific domain.*
3. *Task Ontologies which are dependent on certain tasks.*
4. *Application Ontologies which are dependent on particular applications.*

Finally, a classification of ontologies focused on the design of them was proposed by Gómez-Péréz et al. (2004). They proposed two types of ontologies:

1. *Lightweight ontologies which include concepts organized in taxonomies, the relationships between those concepts and properties that describe the concepts and*
2. *Heavyweight ontologies which add axioms and constraints to lightweight ontologies.*

3. ONTOLOGICAL ENGINEERING

3.1. DESIGN PRINCIPLES

The main principles of the design of ontologies have been introduced by Gruber (1993b) and accepted by many other researchers (Gómez-Péréz et al. 2004). Gruber identified the following five main principles.

1. ***Clarity****:* The meaning must be effectively communicated and the definitions must be objective, independent of the context and documented in a natural way.
2. ***Coherence****:* Inferences should be consistent to both the formal and informal definitions of an ontology.
3. ***Extendibility****: "An ontology should be designed to anticipate the uses of the shared vocabulary."* (onto-design) A very important aspect for knowledge reuse. A user, besides the designer, should be able to add new terms on the existing vocabulary without the need to process the existing ones.
4. ***Minimal Encoding Bias***: An important aspect for knowledge sharing. The conceptualisation should not be dependent on a particular symbol-level encoding.
5. ***Minimal Ontological Commitment****: "An ontology should require the minimal ontological commitment sufficient to support the intended knowledge sharing activities. An ontology should make as few claims as possible about the world being modelled, allowing the parties committed to the ontology freedom to specialize and instantiate the ontology as needed."*(Gruber 1993b)

3.2. METHODOLOGIES

Several methods and methodologies for the development of ontologies have been suggested. Most of them are based on specific projects that have been developed by researchers. It is important to mention that in a survey conducted in 2007 among ontology developers by University of Madeira (Cardoso, 2007) 60% of the respondents replied that they do not use any methodology when developing an ontology. (Gómez-Péréz et al. 2004) suggest a general framework for any methodology based on the software development process identified by IEEE. They defined three different kinds of activities. The **management activities** such as scheduling, control and quality assurance, **development oriented** activities such as implementation and maintenance and **support activities** such as evaluation, documentation and knowledge acquisition.

3.2.1. CYC METHOD

Cyc was one of the early approaches of ontology development methodology suggested by Lenat and Guha (1990) based on the **Cyc** project. Most of the enabling technologies for ontologies were not in place at the moment and as a result their method included the following three processes:

1. *Manual coding of common sense knowledge.*
2. *Computer aided extraction of common sense knowledge.*
3. *Computer managed extraction of common sense knowledge.*

3.2.2. TOVE METHOD

Another methodology based on a project related to an enterprise ontology (**Toronto Virtual Enterprise - TOVE**) was suggested by Gruninger and Fox (1996) and it includes the following steps:

1. *Identify motivating scenarios (purpose and respective use).*
2. *Elaborate informal competency questions (questions written in natural language used to specify requirements and evaluate the ontology).*
3. *Specify the terminology of the ontology using first order logic (identify objects and predicates).*
4. *Write competency questions in a formal way using formal terminology defined in the previous step.*
5. *Specify axioms using first order logic.*
6. *Specify completeness theorems.*

3.2.3. METHONTOLOGY

An approach developed by the Ontology group at Universidad Politéchnica de Madrid. METHONTOLOGY (Fernandez-Lopez et al., 1997) focuses on the knowledge level of the ontology. It is based on the software development process, which means that it divides the processes into three categories (Management, Development and Support) and on knowledge engineering methodologies. The **management activities** involve scheduling, control and quality assurance, the **development activities** involve the specification, conceptualisation, formalisation, implementation and maintenance of the ontology and finally, the **support activities** involve knowledge acquisition, integration, evaluation, documentation and configuration management. Gómez-Péréz et al. (2004) divide the above activities in eight tasks:

1. *Building the glossary of the terms (concepts, instances and attributes and their synonyms and acronyms).*

2. *Building concept taxonomies to classify concepts (Top-down, Bottom-up, Middle-out).*

3. *Building ad hoc binary relation diagrams to identify ad hoc relations (e.g. inverse relations) between concepts of the same or different taxonomies.*

4. *Building the concept dictionary (mainly including instances of concepts).*

5. *Describing the binary ad hoc relations in detail.*

6. *Defining instance attributes in detail.*

7. *Defining class attributes in detail.*

8. *Defining constants in detail.*

3.2.4. SENSUS METHOD

Swartout et al. (1997) this method attempts to promote knowledge sharing as it uses a base ontology on which it builds new domain ontologies. SENSUS method includes the following processes:

1. *Identification of the seed terms (key terms of the domain).*
2. *Establish the links between those terms and the base (SENSUS) ontology.*
3. *Add paths to the root.*
4. *Add new domain terms.*
5. *Add complete subtrees.*

3.2.5. ON-TO-KNOWLEDGE METHOD

This project focuses on the intelligent access to large volumes of semi-structured information on the internet (Staab et al. 2001). It suggests a method for ontology learning aiming to reduce the effort of developing an ontology and provides the tools, methods and techniques to achieve this.

1. *Feasibility Study.*
2. *Define the goal and the requirements (includes competency questions).*
3. *Refinement (produce a target ontology).*
4. *Evaluation (evaluate the ontology against requirement and competency questions).*
5. *Maintenance (determine the details about it – who and how).*

As mentioned above 60% of the ontology developers do not follow a specific methodology. That is because most of them are highly dependent on the project they are based on. Moreover, some of them are too technical for user without experience in software development. As a result of that a combination of all the above is to be adopted when developing an ontology. Such an approach has been made by N. F. Noy and D. L. McGuinness (2001). They suggest a combination of all these methods in self explanatory steps.

1. *Determine the domain and the scope of the ontology.*
2. *Consider reusing existing ontologies.*
3. *Enumerate important terms in the ontology.*
4. *Define the classes and the class hierarchy (Top-down. Bottom-up, Combination).*
5. *Define the properties of classes – slots.*
6. *Define the facets of the slots.*
7. *Create instances.*

3.3. ONTOLOGY DEVELOPMENT TOOLS

3.3.1. EDITORS

Most of the existing editors are based on RDF(S) and OWL as ontology development tools. They support, however, other languages as well. A survey conducted by the University of Madeira (Cardojo, 2007) among ontologists regarding the use of ontology editors shows that Protégé is by far the most common used editor with more than 68% of the repondents using it. This comes as no surprise as Protégé was one of the first editors and is expected to remain a leading editor. Other editors used by the respondents were Swoop (with 13.6%) and OntoEdit (12.2%) as well as others with lower portions of acceptance.

Protégé (latest version 4.1) was developed by the Stanford Medical Informatics group of Stanford University. It is an open source and standalone application that is based on Java and supports the use of RDF and OWL files. Ontologies can be exported as RDF(S), OWL and XML Schema. It also offers visualisation of the ontology and supports reasoning tasks like checking consistency. (Hepp et al., 2007; Gómez-Péréz et al. 2004; http://protege.stanford.edu/)

Altova® Semantic Works® (latest version v2010r3 SP1) is a RDF and OWL commercial editor developed by Altova. It is a pure editor and does not support reasoning tasks. Ontologies can be exported in RDL/XML or N-triples format. It also offers an advanced graphical interface. (Hepp et al., 2007, Kashyap et al., 2008, http://www.altova.com/semanticworks.html).

TopBraid Composer (latest version 3.3.2) is part of a wider suite of applications (TopBraid Suite) developed by TopQuadrant. It is based on Eclipse platform. The free version supports RDF and OWL files but does not include support and maintenance. More capabilities like exporting and importing different formats and reasoning tasks are offered by the higher versions (Standard and Maestro). (Kashyap et al., 2008, Gómez-Péréz et al. 2004, **Error! Hyperlink reference not valid.**)

SWOOP is a hypermedia-based ontology editor developed by the MIND lab in University of Maryland. The development process continues by a group of organisations formed by Clark &

Parsia, IBM Watson Research and the University of Manchester. It supports OWL files and provides a browser-like environment. SWOOP also supports reasoning tasks. (http://code.google.com/p/swoop/, http://semanticweb.org/wiki/Swoop)

OntoStudio is a commercial product for creating and maintaining ontologies developed by Ontoprise. OntoStudio was developed to support F-logic but in its latest version it also supports OWL, RDF(S) and RIF. It also supports collaborative ontology development. (http://semanticweb.org/wiki/OntoStudio, http://www.ontoprise.de/en/home/products/ontostudio/)

These may be the most widely used ontology editors there are however many others like **Ontolingua Server** which was the first ontology editor created. Other ontology editors exist, supporting different ontology types. **OntoSaurus** supporting LOOM ontologies, **WebOnto** supporting OCML ontologies, **OilEd** supporting OIL ontologies and many other supporting several different ontology types such as **SemTalk** (RDF(S) and OWL), **COBra** (OWL, GO) as well as others.

3.3.2. Languages

The layer cake of the semantic web suggested by the W3 Consortium, presented in the previous chapter, provides a clear picture of the main languages used in ontology development. XML, RDF and OWL are some of the main but not the only tools. Taye (2010a) in his recent article identifies three types of languages for ontology development. **Vocabularies** for ontologies developed using natural language, **frame-based** languages used for building the structure of an ontology and **logic-based** languages.

eXtensible Markup Language (XML) is a general purpose language (Hepp et al., 2007). The fact that the user can define his own tags classifies it as an extensible language. XML has been the base for development of other later languages such as RDF, DAML-OIL and others.

Resource Description Framework (RDF) is another language recommended by the W3C. Based on XML, it is used for representing resources on the web in a standardised with the use of URIs (Hepp et al., 2007). This way it facilitates information and knowledge exchange. RDF is organised in triples in order to define relationships between concepts (Sharman et al., 2006).

RDF Schema is a specification of the RDF vocabulary description language. It is the second most common language for ontology development according to (Cardojo, 2007) It defines the classes (e.g. rdfs:Class, rdfs:Resource) and properties (e.g. rdfs:Domain, rdfs:subClassOf) that can be used in RDF in order to describe classes and properties in a machine understandable way (W3C; Taye, 2010a).

DARPA Agent Markup Language + Ontology Inference Layer (DAML+OIL) was built on RDF and RDFS in the early 2000. It is a web language for describing web resources. The DAML program was terminated in 2006 as DAML+OIL was succeeded by OWL.

Web Ontology Language (OWL) is built on RDFS and derived from DAML+OIL. It is the newest standard recommended and created by W3C. OWL offers greater interoperability than other languages. According to a survey conducted by the University of Madeira(Cardojo, 2007), OWL is the most common used ontology language (75.9%). It has three versions based on the expressiveness offered. **OWL Lite** is the simplest one that uses the most common features of OWL and is suitable for building taxonomies. **OWL DL** is based on Description Logics. It offers more logical constructs such as negation, disjunction and conjunction as well as inference facilities and it includes the complete OWL vocabulary. Finally, **OWL Full** is the most flexible and most expressive of all three versions and has no restrictions.

4. DEVELOPING ESYMBIOSIS ONTOLOGY

4.1. THE PURPOSE OF THE ONTOLOGY

The eSymbiosis ontology serves a threefold purpose:

(i) Navigate the user through the registration process

During the process of registration, the ontology navigates the user and indicates requests for input. The e-symbiosis system dictates the path that the user follows during the registration process by parsing the taxonomy of the ontology. During this process, the users are prompted to give information about the resource or technology they are registering. For this purpose, labels and other annotation properties are used with an aim to improve the readability of the ontology and hence the user experience. The use of labels not only improves the user friendliness but also allows multilingual modelling for the ontology, an issue to be discussed further in the next section.

(ii) Provide a common vocabulary for IS domain

The ontology provides a standardised vocabulary for the given domain and helps in tackling heterogeneity. It also facilitates the use of synonyms hence removing the jargon barrier and any latent syntactic issues by the use of an industry specific terminology.

(iii) Support the input – output matching for synergy identification

The vocabulary described above is used as a common reference for the match making process. The Resource concept plays a key role in this process as it is used as the reference for calculating similarity between inputs and outputs. The use of semantics allows the identification of partial matches. This feature is further described in the matching section.

4.2. SELECTION OF THE TOOLS

There is a big variety of tools for ontology development as they were analysed in the previous section. The ontology editor that will be used for the development of the specific ontology will be Protégé as it is a standalone and open source application supporting OWL, RDF Schema and it provides reasoning facilities as well as rule encoding through SWRL. Protégé is the most popular editor among ontologists (especially researchers and academics) with 68.2% (Cardozo, 2007). The language used for the ontology development will be OWL which is also the most frequently used language for ontology development and is supported by Protégé.

4.3. METHODOLOGY

The same survey shows that the majority of the ontology developers (60%) (Cardozo, 2007) do not follow a specific methodology. That is because most of the methodologies are based on specific projects and cannot be adopted as they are. As a result of that each project requires either generalisations or specifications. In this project the methodology followed will be close to the one suggested by Natalya F. Noy and Deborah L. McGuinness which involves seven steps (Noy and McGuinness, 2001). The approach followed in step 4 is a combination of top-down and bottom-up approaches.

4.3.1. DETERMINE THE DOMAIN AND THE SCOPE OF THE ONTOLOGY

The ontology aims in representing the knowledge of IS domain. This will include all the knowledge required to register the resources or the technologies available, details of the participants and also compare them and explore the possibility of establishing a symbiotic link. To achieve high standards it is important to create a number of competency questions. These questions play the role of the requirements or specifications and are easy to understand. Moreover, they can be used as a guide for validation of the ontology after it has been developed. Some of the competency questions can be the following:

i. What is the domain of the ontology?
ii. What will be the exact use of the ontology?
iii. Who will use the ontology?
iv. Is it a standalone ontology or part of a semantic application?

These two are general questions that determine the domain of the ontology. The following questions are more specific and their purpose is to define some, if not all, of the requirements of the ontology.

v. What type of participants are there?
vi. What kind of information is necessary for the establishment of a symbiotic synergy?
vii. What types of resources/solutions should the ontology include?
viii. Which part of the information is already known to us and which part should be provided by the participants?

4.3.2. CONSIDER REUSING EXISTING ONTOLOGIES

Several existing ontologies have been considered. The UNSPSC ontology for products and services consists of more than 10,000 concepts divided into 54 categories. This ontology was deemed very detailed for the needs of the project and therefore has not been used.

Ontologies about the industry classification such as

- ✓ NAICS (http://www.srdc.metu.edu.tr/ubl/contextOntology/naics.owl),
- ✓ NACE (http://www.srdc.metu.edu.tr/ubl/contextOntology/nace.owl)
- ✓ ISIC (http://www.srdc.metu.edu.tr/ubl/contextOntology/isic.owl)

have been considered but all of them are based on previous versions of the respective classification. For these reasons, an industry classification based on the latest version of NACE have been developed.

Other relevant ontologies, representing the units of measurement, processes and materials have been considered. None has been re-used as they were either very detailed or very domain specific allowing no flexibility. Although the considered ontologies have not been re-used, it is possible to import them in eSymbiosis ontology if the operation of the platform indicates so even at a later stage. Ontologies that could be reused include:

- ✓ http://sweet.jpl.nasa.gov/2.0/chem.owl
- ✓ http://sweet.jpl.nasa.gov/2.0/chemElement.owl
- ✓ http://sweet.jpl.nasa.gov/2.1/matr.owl
- ✓ http://sweet.jpl.nasa.gov/2.0/chemCompoundOrganic.owl
- ✓ http://sweet.jpl.nasa.gov/2.0/chemCompound.owl
- ✓ http://sweet.jpl.nasa.gov/ontology/substance.owl

4.3.3. ENUMERATE IMPORTANT TERMS IN THE ONTOLOGY.

4.3.3.1. EXISTING CLASSIFICATIONS

For the modelling of the ontologies we have used several existing classifications. The waste classification has been based on the European Waste Catalogue (EWC) and the EWC Stat. These classifications have been enhanced by the addition of links to materials which define the composition of the waste along with the characterisation of the wastes. Existing classifications have been reused for the representation of the industrial sectors (NACE) and also a customised version of CPC product classification. Some material classifications have also been extracted from literature.

4.3.3.2. Developed Classifications

Besides reusing existing classifications we have also developed bespoke classifications that represent the materials, the solutions and other peripheral infomration such as the units of measurement, delivery methods, storage methods and more. Based on the literature review, there was no existing classification for the solutions and more specifically the processing technologies. For that reason, a bespoke classification has been developed. This classification has been in accordance with the Resource classification in an effort to be an intuitive process that can be used by non-experts. Technologies have been classified according to their type, input, industry and their characteristics or parametres.

For these classifications we have used data extracted from the target area of Sterea Ellada and literature. These classifications have been transformed into ontologies as described in section 4.3.4.

4.3.4. Define the classes and the class hierarchy

The eSymbiosis ontology currently consists of around 2000 concepts. These concepts represent more than 50 processing technologies, 20 waste streams, 8 materials streams and 10 products streams.

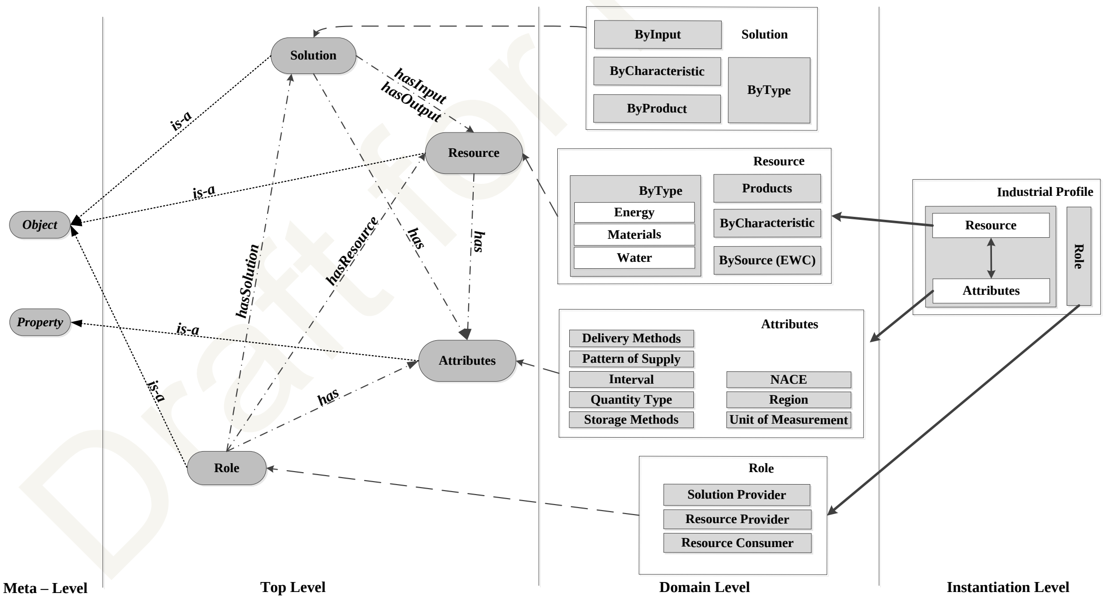

Figure 1 Design of the IS ontology

Our ontology consists of four levels of abstraction (Figure 1): i) meta-level, ii) top-level, iii) domain level and iv) instantiation level, starting from the universal meta-level and leading to the detailed instantiation level.

4.3.4.1. META-LEVEL

This level consists of general concepts that are independent of the domain and can be applied universally. The use of a meta-ontology makes the whole ontology easier to share and reuse but it also provides a better understanding for users outside the domain (Marquardt et al., 2010). The eSymbiosis meta-level consists of the concepts Object and Property. Objects (Trokanas et al. 2012) are all the physical assets associated with the domain. Properties are the attributes that describe the objects defined in the ontology.

4.3.4.2. TOP LEVEL

The top level of the ontology contains abstract concepts of the IS domain which can also be applied in similar domains. These concepts are:

- ✓ **Role** - the different types of participants that can be part of a symbiotic synergy.
- ✓ **Solution** - processing technologies that can process resources (inputs) and produce some others (outputs).
- ✓ **Resource** - materials, waste, energy, products, water and expertise that a user might have to offer or require.
- ✓ **Attributes –** information used to describe the other three top level concepts.

Besides the top level concepts, at this level we have defined some top-level properties that provide the relationships beween the four concepts. These properties are shown in Figure 1.

4.3.4.3. DOMAIN LEVEL

The top level concepts are presented as ontology modules at the domain level. Four ontology modules are described in this section.

***Role*.** Represents the different types of users of the system. It is the “entry point” of the system – where the user navigation begins. Our system supports three different types of users, ResourceProducer, ResourceConsumer and SolutionProvider (Figure 3).

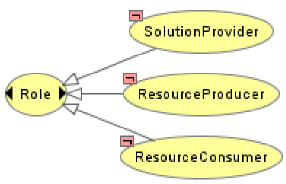

Figure 2. Role subsumption.

Resource. The resource module is based on the integration of multiple sources such as the general knowledge about process industry, existing classifications and expert knowledge. In the eSymbiosis platform, resources are classified in four different ways:

- ✓ **Resource by source**. Based on the European Waste Catalogue (EWC) (EC 2002) which is the classification currently in use by the IS practitioners. EWC is a six digit code based on the source that the resource is derived from. Combined with EWC STAT (EC 2010) (statistical version), this allows the broad classification of almost all types of waste.
- ✓ **Resource by characteristic**. Resources are classified based on key characteristics.
- ✓ **Products classification**. Based on existing product classifications, this aims at allowing users with all levels of expertise to classify their resource. (i.e. a restaurant chain having plastic water bottles, not knowing their composition).
- ✓ **Resource by type**. This sub-module of the ontology is the core sub-module because it has been designed to intrinsically invoke similarity for classes that are close to each other in the subsumption tree – unless otherwise stated. It is used as the common reference for all other sub-modules including the solutionswhich are linked to it by object properties such as hasComposition, hasProduct etc. This classification is used in the process of semantic matching.

Solution. To our knowledge no classification is in place for processing technologies. The importance of expertise in the process industry dictates the need to integrate tacit knowledge in the solution classification. For this reason, extensive input has been extracted from prior experience in the fields of IS and process engineering. Solutions have been classified in 4 different ways, following a principle similar to that for resource classification in an effort to provide an intuitive user registration process.

- ✓ **Solution by input**. Classified based on the resources the solution can process. This is the most straightforward classification that assumes the expert knowledge of the user.
- ✓ **Solution by characteristic**. Classified based on key characteristics.
- ✓ **Solutions by Industry**. Classification based on the industry sector a solution can be found in. This classification requires a lower level of expertise from the person who registers the solution.
- ✓ **Solution by Type**. Classified based on the type of the technology as defined in the literature. Some of the categories are Thermochemical, Mechanical, Biochemical etc.

Attributes. This module contains all the peripheral information that describe and define the main concepts. The use of concepts for the modelling of an attribute has been chosen over the use of properties in cases where it allows more flexibility in the design. In some cases, this approach has been followed in order to enable the use of more than one language. In cases, such as the location, concepts have been favoured in order to follow established approaches such as the ontology for geographical locations.

Some of the most important concepts that form the Attributes module are:

- ✓ *PatternOfSupply*
- ✓ *UnitOfMeasurement*
- ✓ *Location*
- ✓ *QuantityType*

The top level of the concepts is given in the following table along with a short description.

Table 2 Top level concepts of the ontology

Concepts	Description
Role	The participants of IS process.
ResourceProducer	Participants who have a resource available.
SolutionProvider	Participants who have some solution available.
Resource	Resource in IS contains materials, wastes, water, energy etc.
ResourceByType	Resources that can be naturally classified by their type
Materials	The substance or substances out of which a thing is or can be made
Polymers	This concept ranges from synthetic plastics and elastomers to natural biopolymers.
Metals	Solid materials which are typically hard, shiny, malleable, fusible, and ductile, with good electrical and thermal conductivity. Includes ferrous and non-ferrous metals and alloys.
Ceramics	Inorganic, non-metallic materials generally made using clay and other earthen materials through heat and cooling.
Chemicals	Any material with a definite chemical composition.
Minerals	A mineral is a naturally occurring inorganic solid, with a definite chemical composition, and an ordered atomic arrangement.
Composites	Naturally occurring or engineered materials made from two or more constituents.
OrganicMatter	Matter that comes from a once living organism such as plants and animals.
Rocks	The solid mineral materials forming part of the surface of the earth and other planets.
Energy	Usable heat or power.
Electricity	The supply of electric current to a house or other building for heating, lighting, or powering appliances.
Heat	The transfer of energy from one body to another as a result

	of a difference in temperature or a change in phase.
Water	A clear, colorless, odorless, and tasteless liquid, H_2O.
ResourceBySource(EWC)	Based on EWC – waste classification based on the source process.
Products	The totality of goods that can be made available by industries.
EnergyProducts	Goods that can be used for the generation of energy.
Biomass	Organic matter used as a fuel.
Biofuels	Fuels derived directly from living matter.
Coal	A combustible black or dark brown rock consisting mainly of carbonized plant matter, found mainly in underground deposits and widely used as fuel.
NaturalGas	Flammable gas, consisting largely of methane and other hydrocarbons.
Oil	A viscous liquid derived from petroleum.
OilShale	Fine-grained sedimentary rock from which oil can be extracted
Peat	A brown, soil-like material, consisting of partly decomposed vegetable matter.
MaterialProducts	All other goods, not included in class EnergyProducts.
ResourceByCharacteristic	Resources classified based on important physical or chemical properties
BiodegradableResource	Resources that are capable of decaying through the action of living organisms.
Technology	Any technological process that can convert an input to a different output under certain circumstances and with a specific result.
TechnologyByType	Technologies classified by their type.
TechnologyByIndustry	Technologies classified based on the industry they can be applied in
TechnologyByInput	Technologies classified based on their input.
TechnologyByCharacteristic	Technologies classified based on important physical or chemical requirements they have.
Attributes	Information used to describe and define all the concepts of the ontology.
geo:SpatialThing	Imported concept which links to the latitude and longitude information.
Location	Linked to the above concept, the lat and long of the participant.
NACE	Statistical classification of economic activities

	in the European Community.
QuantityType	The physical form of a resource.
PatternOfSupply	The pattern that a resource is produced or required.
Region	Linked to location, embracing the local aspect of IS.
UnitOfMeasurement	Units of measurement.

The high level of the solutions classification is given in Table 3.

Table 3 Solutions (Top level)

By Product	By Type
Energy Products Production	Thermochemical
Biofuels Production	Combustion
Biogases Production	Stoker boilers
Fuel Gas Production	Fluidised bed boiler
Liquid Biofuels Production	Incineration
Biodiesel Production	Rotary Kiln
Bio Oil Production	Direct flame
Material Products Production	Gasification
Metal Production	Moving bed
Aluminium Products Production	Fluidised bed
Primary Aluminium Production	Pyrolysis
Aluminium Foil Production	Fast Pyrolysis
Rubber & Plastic Production	Slow Pyrolysis
Plastics Production	Mechanical
PET Production	Separation
PVC Production	Adsorption
Rubber Production	Distillation
Latex Rubber Production	Crushing
Paper Production	Grinder Machines
Copypaper Production	Mills
Ceramics Production	Biological
Glass Production	Anaerobic Digestion
Glassworks	Wet AD
Glass Wool Production	Mesophilic AD
Cement Production	Composting
Portland Cement Production	Aerobic Composting
Calcium Aluminate Cement Production	Anaerobic Composting
Supersulfated cement Production	Vermicomposting

4.3.4.4. INSTANTIATION LEVEL

This level is populated by the users as they register their resources or solutions. However, instances in our knowledge model do not only represent users. Instances have also been created for the concepts that are members of *Attributes*. To clarify, the instances of the concepts mentioned in the attributes section are listed in the following Table 4.

Table 4 Instances of the ontology

Concept	Instances
PatternOfSupply	{Continuous, Batch}
UnitOfMeasurment	{kg, tonnes, m^3 , litre, acres, m^2}
QuantityType	{Liquid, Solid, Powder, Solution, Emulsion, Gas, Flake, Slurry}

4.3.5. DEFINE THE PROPERTIES OF CLASSES – SLOTS

Object properties (Table 5) have a twofold purpose in our ontology. The most straightforward use is to provide links between concepts and being used as the path of navigation outside the strict limits of the taxonomy. Object properties are meant to facilitate the smooth navigation of the user and enhance the functionality of the information collection facility as navigating exclusively through a taxonomy can scarcely be either an engaging or a user-friendly experience.

Finally, a very important advantage of object properties is that they offer greater flexibility for inference and reasoning. The knowledge modelled through object properties and other elements of the ontology (such as axioms) can be used to infer "new" knowledge and help in minimising the user's effort in providing input in the system.

In the case of eSymbiosis ontology, the knowledge should be there (modelled in the ontology) but only the necessary (in terms of collecting information) knowledge should be addressed to the user. That means that some knowledge can be used only for reasoning and inference purposes.

Table 5 Object Properties

Relationship	Description
geo:location	Term of a basic RDF vocabulary that provides the Semantic Web community with a namespace for representing **lat**(itude), **long**(itude) and other information about spatially-located things, using WGS84 as a reference datum.
belongsToIndustry	Link between participants and the industry sector code (NACE) they belong to.
hasResource	Link between a resource provider and the type of resource

	they have available.
hasTechnology	Link between a solution provider and the type of solution they have available.
hasPatternOfSupply	Links resources to the PatternOfSupply attribute concept. The concept is about the pattern of the demand or availability (Continuous, Batch).
hasApplicationIn	Link between resources and industry sectors for the integration of tacit knowledge about the use of resources in different industries.
canUse	Inverse relation of hasApplicationIn, used for intelligent recommendations.
hasQuantityType	Links resources to the QuantityType attribute concept. The concept is about the physical form of the resource (Solid, Liquid etc.).
hasUnitOfMeasurement	Links resources and solutions to the UnitOfMeasurement attribute concept. The concept is about the unit of measurement of the resource (Kg,Tonnes etc.).
hasComposie	Relation used to provide information about the composition of products and waste types.
isCompositeOf	The inverse of the above.
hasInput	Relation used to link solutions to their inputs.
canProcess	Relation used to link solutions to their main inputs.
needsWater	Relation used to link solutions to their water inputs.
needsEnergy	Relation used to link solutions to their energy inputs.
canBeprocessedBy	Inverse relation of canProcess. Used for tacit knowledge modelling for resource processing.
hasOutput	Relation used to link solutions to their outputs.
hasProduct	Relation used to link solutions to their products.
hasStorageMethod	Relation used to link resources to the storage methods used for their storage.
hasDeliveryMethod	Relation used to link resources to the current method of delivery for resources.
hasInterval	Links resources to the interval related to the amount of resource produced.

For example, in the process of collecting information about the waste of industries, the industry user does not need to know the complex chemical details of the waste type she classifies. In this case, the user should give information about the quantity, the location and the category of waste she produces. The knowledge that lies underneath includes details about the chemical composition and other information that can be used for example in calculating environmental aspects of the waste produced like CO_2 emissions.

A convention that has been used in order to include underlying knowledge, is that a property is addressed to the user only if it has a language label (xml:lang).

Data type properties (Table 6) are also a very important component of the ontology. Such properties are used for the collection of information regarding quantities, names, address details, phone numbers and many others, without which the process would be incomplete. The data type properties are the only actual input of the user besides “clicking” on already modelled knowledge. The effort has been towards minimising the required user input.

Table 6 Data type Properties

Property	Description
confidentialityFlag	Boolean property used to flag confidential information.
hasQuantity	The amount of resource available or required.
hasProcessingPrice	Addressing solution providers – the cost for the resource currently in use.
hasAnnualCost	Addressing solution providers – the annual cost of a resource as feedstock.
isValidFrom	The date the resource/solution becomes available.
isValidUntil	The date the resource/solution stops being available.
hasName	Free text entry for the user to specify the name of the resource/solution.
isBiodegradable	Boolean property used to identify resources that are biodegradable.
isHazardous	Boolean property used to identify resources that are hazardous (contaminated etc.) as defined in the European Waste Catalogue (EWC).
deliveryCapability	Boolean property used to identify whether the user can deliver the resource on offer.
hasStorageCapacity	The amount of resource the user can store when requesting or producing a resource.

Sub-properties can be a great tool for the organisation of all properties. Grouping properties (both object and data type) can be very helpful for sharing and re-using the ontology in question but also for developing applications that can handle effectively such an ontology. Groups could be created based on whether the properties are addressed to the user, the point of the information collection process that they appear or any other use that serve the specific domain and problem.

4.3.5.1. Annotation Properties

Annotation properties are a key element of the eSymbiosis ontology. Their use can span from informative labels, to problem specific custom user defined annotations.

rdfs:label. Labels contain all the information that is addressed to the user. This includes the names of the concepts, properties, questions or comments that aim in helping the user during the registration process. The combined use of labels and xml:lang tags not only allows the multilingual annotation of the ontology but also the "multi-user" annotation. This applies when different type of users might be involved in the registration process. In such cases, some of the properties might be user-specific or some others might have different meanings for different types of users. One relevant example (with both multi-lingual and multi-user annotations) is demonstrated in Figure 3.

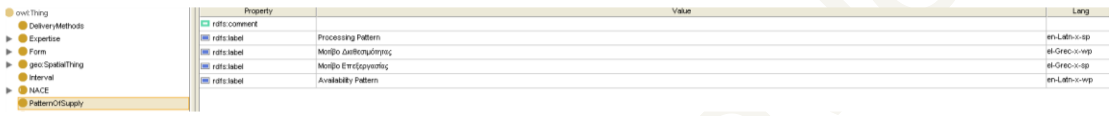

Figure 3 The use of xml:lang tag

User defined annotations. Because the registration process is problem specific, more annotations – other than the defined ones – are required. For that, custom annotation properties have been defined. These annotations can be used for addressing the user (questions), providing guidance to the user (help) or for application specific causes such as numbering the concepts or properties in ascending or descending order, giving weights to object properties or serve any other issue.

4.3.6. Define the Facets of the Slots

Facets or Restrictions (e.g. in Figure 4) or axioms are a core part of ontology engineering. Moving from taxonomy to heavy weight ontology, restrictions are a useful tool in the hands of the ontology engineer. Restrictions serve three main purposes in ontology engineering.

OWL:ONPROPERTY. Limiting which values can be used for the respective property, restrictions or axioms are vital for the information of the ontology to become knowledge. The three types of property restrictions and their role in information collection ontologies are described below.

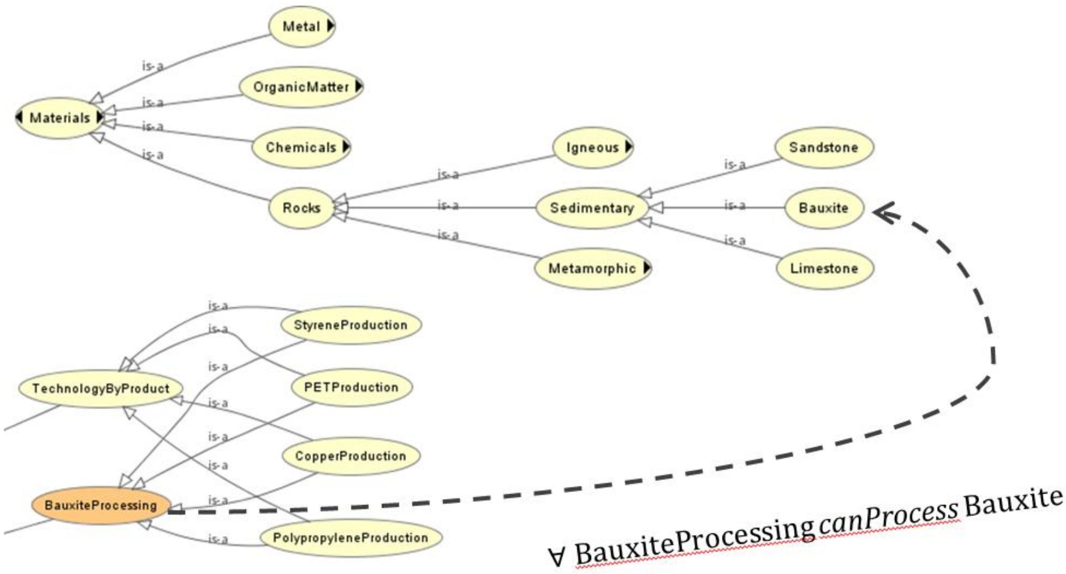

Figure 4 Restriction example

<u>*ALLVALUESFROM*</u>. The allValuesFrom restriction is applied on object properties and links classes. It indicates that "…*the particular class has a local range restriction associated with it*…" W3C owl specification. For the purposes of the information collection process the "only" restriction can help creating a more efficient and personalised experience for the user by controlling the classes that an object property can take values from. This means that the user will be prompted with values that belong to these certain classes. This is a way of validating user input and assuring that the right information is acquired. Moreover, the user will not see unnecessary information, hence reducing the time and effort required and also provide the user with a more consistent view of what is neede form her.

Consider a user entering information about a technology which is capable of processing non-ferrous metals. Through the information collection process instead of providing a list with all possible materials the system should provide a list of the non-ferrous materials following the *Non-FerrousTechnology canProcess allValuesFrom Non-FerrousMaterials*. Although useful the *allValuesFrom* restriction must be carefully used as it can cause inconsistency problems when used as an equivalent class (necessary & sufficient condition).

<u>*SOMEVALUESFROM*</u>. The *someValuesFrom* is a less constricting restriction that can be used as a guideline for the user navigation. It is also useful when multiple choices are allowed. E.g. the *someValuesFrom* restriction could be used if the processing technology could process several materials such as consumables (e.g. oil and water). In this case, the technology should have some values from the non-ferrous metals and also some other materials. This restriction can also be used as a means of input validation. For example, in the case mentioned before, the user must include at

least one non-ferrous metal no matter what her other choices are. Restrictions can be also applied in data type properties which is very helpful for input validation (i.e. certain properties must have a float or integer value).

hasValue. HasValue restriction specifies the value of the property. Their main use in information collection ontologies is for cases where the value of a certain field is fixed. As a result of that, the user will not need to fill in such fields as the information will be automatically inferred from the ontology using the hasValue restriction. Although useful, hasValue restriction can cause problems with the reasoning process as it is not supported by DIG reasoners and also causes problems with direct reasoners when the size of the ontology is large.

CARDINALITIES. Cardinalities allow more flexibility in controlling the number of values that a property is allowed to have, going further than property characteristics like functional.

Min & Max and Exactly. Cardinalities can be used to control the number of possible answers a user could give. Whether they are applied on object or data type properties, they control the number of values each property can link to for each individual not taking into account the range of the property.

Used for Inference. Besides defining the knowledge about concepts by restricting the values that are allowed for each object or data property, restrictions are very important for the inference process.

Some of the most important restrictions are given in Table 7.

Table 7 Restrictions on properties

Domain Class	Property	Type	Value Range
Role	*confidentialityFlag*	Cardinality	=1
Role	*geo:location*	Cardinality	=1
Role	*belongsToIndustry*	Range	NACE
ResourceProducer	*hasResource*	Range & Cardinality	Resource(>1)
SolutionProvider	*hasSolution*	Range & Cardinality	Technology (>1)
Resource	*hasQuantity*	Cardinality	=1
Resource	*hasPatternOfSupply*	Cardinality	=1
Resource	*validFrom*	Cardinality	=1
Resource	*validTo*	Cardinality	=1
Resource	*isBiodegradable*	Cardinality	<1
BiodegradableResource	*isBiodegradable*	Value	true
Products, ResourceBySource	*hasComposite*	Range	Materials
Materials	*isCompositeOf*	Range	Products,

			ResourceBySource
Technology	*canProcess*	Range & Cardinality	Resource (>1)
Technology	*hasProduct*	Range & Cardinality	Resource (>1)
Technology	*needsEnergy*	Range & Cardinality	Energy or Energy Products
Technology	*needsWater*	Range & Cardinality	Water (<1)

4.3.7. DEFINE THE RULES

Rules (and more specifically SWRL rules) can be used to infer new knowledge and also do calculations. The problem with these rules is that no instances can be created by them. Rules could also be used from the reasoner to extract and use the information.

For example, the following rule can be used to calculate the quantity of the product produced by a cement production technology given the conversion rate. In the following example, for a cement production that produces cement (an instance available from user registration) and requires `?z` amount of limestone, then the quantity of cement produced (`?quantity`) is inferred.

```
CementProduction(?x) ∧ hasProduct(?x, ?p) ∧ Cement(?p) ∧ Limestone(?y) ∧
hasQuantity(?y, ?z) ∧  swrlb:multiply(?quantity, ?z, 0.73)

→

hasQuantity(?p, ?quantity)
```

For example, for a cement production technology that can process 100 kg of limestone and an estimated average conversion rate of 73%, the use of this rules leads to the inferred axiom given in Figure 5. This represents the allocation of the quantity of the cement produced.

Figure 5 Inferred Axiom

4.3.8. OTHER ISSUES

In order to allow useful information to be deduced from the names of the concepts and properties we have adopted the following conventions.

- ✓ **Class Names**: No whitespaces CamelCase. (e.g. ResourceProducer)
- ✓ **Property Names**: No whitespaces, starting with a lower case (usually for the prefix) and capitalising every word after that (e.g. hasQuantityType)

5. SEMANTIC SERVICE DESCRIPTION

Benefiting from what semantics have to offer, all participating industries in the esymbiosis platform are described as a "Semantic Web Service". This means that the data about the industry, whether a waste producer, a solution provider, a practitioner or even an intermediary would be stored in a machine understandable way within the system. This will result in the automatic handling of information. Moreover they matching process between industries to create the synergy would not only be automated but would be increased in terms of accuracy and relevance.

5.1. SEMANTIC WEB SERVICES AND OWL-S

Semantic web services were introduced to semantically enrich the description or existing web services.[1] By adding semantic annotation to service elements using a modelling language, the annotations could be processed by machines therefore eliminating human effort[2]. Semantic web services are aimed to provide:

- Greater automation of service selection and invocation
- Automated translation of message content between heterogeneous interoperation services
- Automated or semi- automated approaches to service composition
- Comprehensive approaches to service monitoring and recovery from failure

The mentioned tasks are facilitated with the provision of a richer description of services by the use of ontologies.SWS technology provides specifications for web services to describe their interaction pattern. Description of interaction patterns can be used by the client agents during the discovery as well as the execution time. Intelligence is not only utilized to discover potential matches but could also allow selection and ranking of discovered synergy possibilities. This targets the manual information extraction for creating the link between demander and supplier. This approach also allows building complex services.

The ontological formalism used for Semantic Service Description in the eSymbiosis platform is based on universally recognized OWL-S framework [Martin et al. 2004]. OWL-S is a well known semantic

[1] Silva, P. d. A. and C. M. F. A. Ribeiro , "On the use of Ontology Reconciliation Techniques in SOA", Third International Conference on Next Generation

[2] Verma, K. and A. Sheth ,"Semantically Annotating a Web Service." IEEE Computer Society, 2007

web service framework which is developed using the OWL semantic web description language. OWL-S is itself an ontology and is used for semantically describing web services. Every OWL-S description is an instance of the ontology. Figure 6 shows different components of the OWL-S ontology model.

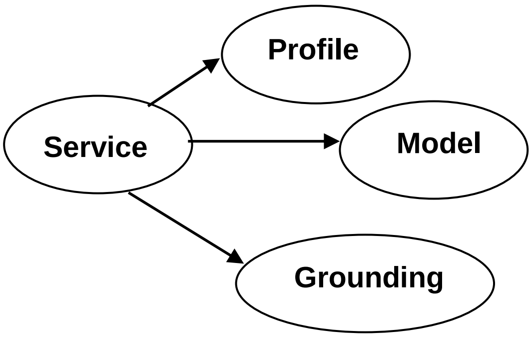

Figure 6 OWL-S ontology Model

Within the OWL-S framework, the Service Profile provides a way to describe the services offered by the providers, and the services needed by the requesters. For the eysmbiosis process the service "*profile"* is used to describe each site registered by the user. However only information required for the matching process will be stored in the profile as the user's profile will be loaded in the matchmaker in the process of finding the industry synergy. Each user registered as a semantic service can have various profiles. In other words the link between service and profile is $1-n$. Since the profile is used for finding the potential matches it is best to describe different resources owned by the user.

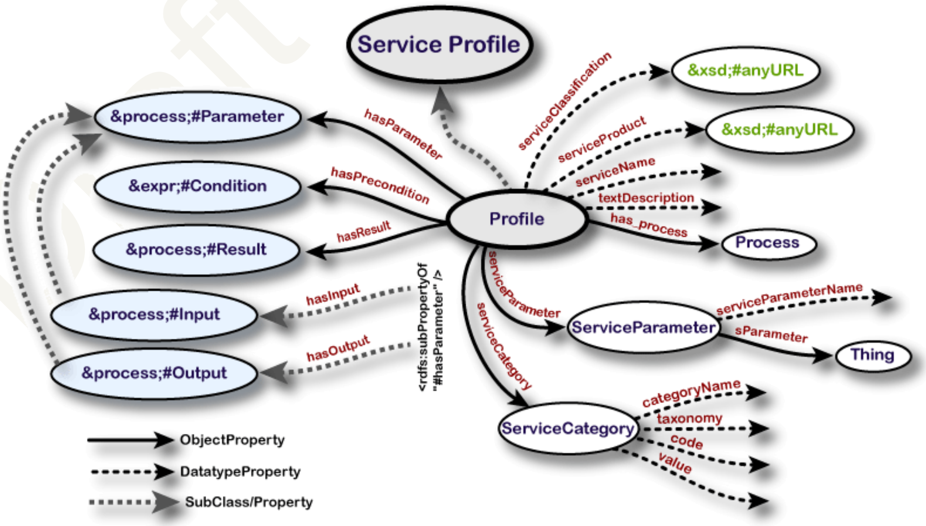

Figure 7 Semantic Web Service Profile Ontology

5.2. eSymbiosis Semantic Service Description based on OWL-S

The Semantic profile of each industry is formed when the user has submitted all the data through the dynamic web portal. Every user will be an instance of the OWL-S ontology which has been modified to incorporate properties related to the IS resources, specifically the ones serving as essential metrics for potential synergy discovery. Every Industry has a set of inputs and outputs referring to a resource type. The information about each resource is presented in the format of data properties attached to the input/output concept. The value assigned to these properties are either numerical or point to a concept within the domain ontology (Figure 8).

There is direct mapping between the data properties and object properties defined for the resource concept (refer to Nikos's section of Resource properties) and the properties of the input and output of each industries' semantic profile.

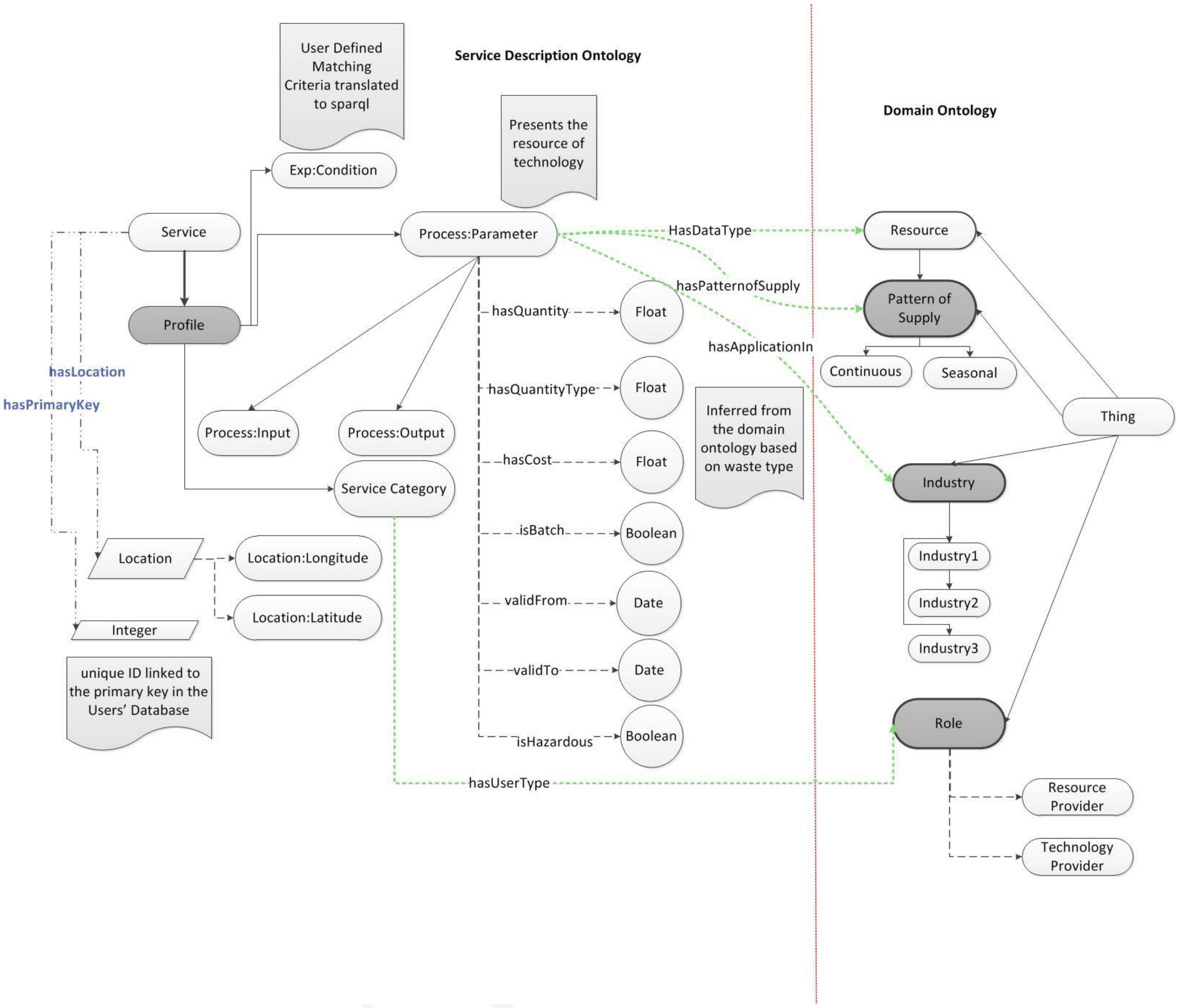

Figure 8 Service Description Ontology

Table 8 Service Description Ontology additional data properties in Process.owl

Property Name	Property type	Description	Value or Domain ontology concept
ParameterType	*anyURI*	The resource type of industry as classified by in the domain ontology	*Chosen by industry from a **resource** type concept within the domain ontology*
hasProcessingPrice	*Float*	For the solution provider	*Maps to the Processing prices data property in the ontology*
hasQuantityType	*anyURI*	The Physical form of the resource	*Maps to the **QuantityType** concept in the domain ontology*
hasQuantity[3]	*Float*	Datatype property of the	*Maps to the Quantity data*

[3] *Quantity to be converted to the smallest unit within the quantity type category*

		quantity of the available resource	*property in the ontology*
hasPatternOfSupply	*anyURI*	Is the availability of this resource **batch** or **Continuous**	*Maps to the Pattern of supply property in the ontology*
hasAnnualCost	*Float*	What is the annual cost for the supply of the resource currently used?	*Maps to the Processing prices in the ontology*
delilveryCapability	*Boolean*	Whether the industry has delivery capabilities	*Maps to the delivery Capability data property in the domain ontology*
hasApplicationIn	*anyURI*	Refers to the application of the resource within certain industries	*Maps to the **NACE** concept within the ontology*
hasMinDelivery	*Float*	Refers to the minimum amount of resource delivered	*Maps to the minimum delivery property within the ontology*
hasEWCCode	*Integer*	Refers to the EWC code registered for the resource	*Maps to **ResourcebySource** concept in the domain ontology*
isHazardous	*Boolean*	Shows whether the resource is hazardous or not	*Maps to the isHazardous property within the domain ontology*
validFrom	*Date*	This identifies the date from which the resource is available	*Maps to the valid From property within the domain ontology*
validTo	*Date*	Identifies the date until which the resource is available	*Maps to the domain ontology*

OWL-S framework has built in properties and concepts which could be used for defining additional attributes of the system users

Table 9 Service Description Ontology built-in properties in Profile.owl

Property Name	Property type	Description	Value
ServiceClassification	anyURI	refers to the NACE code	Maps to concept ***NACE*** in the domain ontology
ServiceName	Literal	Industry's site name	Industry's site name
TextDescription	Literal	Industry's business	Describes industry's business in general terms
Exp:Condition	Sparql	Used for pre-conditions	Sparql query

		when matching processing technologies	
Service Category	anyURI	Used to categorize the role of industry	Maps to ***Role*** concept in the domain ontology
ServiceParameter	anyURI	Any extra parameters required for the service description which links to a concept within the domain ontology could be defined using this parameter	Location and Industry type are two parameters defined at this level of the ontology mapping to ***IndustryType*** concept and Location property in the domain ontology

Table 10 Service Description Ontology additional data properties in Service.owl

Property Name	Property type	Description	Value
hasPrimaryKey	Integer	A unique ID linking each site to its original industry	*connect SWS profile of each technology to a record in a database that stores additional proprietary and non-semantic information*
hasStorageCapacity	*float*	Refers to the storage capacity of the Industry	*Maps to the storage capacity property within the ontology*
hasLocation hasLongitude hasLatitude	Float/float	Refers to the location of the site	*Based on longitude and latitude values of the industry's location*

6. INPUT – OUTPUT MATCHING

Semantic relevance between members of the eSymbiosis platform is established by matching respective description ontologies using the purposely designed inference engine and ontology matchmaker engine.The input and output of the industries searching for a possible synergy will be compared based on firstly and more importantly the semantic relationship between the type of resources and secondly on the semantic similarity between qualitative attributes enriching the resource objects.

6.1. Matching Metrics

The metrics used for the matching are all modelled in the service description ontology. Table 11 provides an overview of metrics identified by domain experts as essential for the practice of IS.

Table 11 Metrics Description

Matching Metrics	Explanation
Type of Resource	Refers to the knowledge classification in the domain ontology
Quantity of Resource	The matching industry should be able to process not only the type but also the amount. This might result in multiple matches which can be combined to match the requested amount.
Pattern of supply	Whether it is supplied in batches or continuously
Availability	Presents the availability period of a resource
Location	Presents the location of each site belonging to industries, based on longitude and latitude

6.2. Matching methods based on semantic web services

By using semantic technology in the platform, the platform enables more than just automation. It builds intelligence within the system that allows types of matching which are otherwise not possible. More precisely, by using the implicit and tacit knowledge modelled and inferred from the ontology through relationships between material and the knowledge about their application and use, the platform is capable to offer alternative resources. In situations where an exact solution is not identifiable, it is possible to identify solutions which handle similar type of resources and present them to the user. The level of semantic similarity is indicated to the user. The knowledge about application of resources in certain industries, which is modelled through the *hasApplication* relation in the domain ontology, empowers the creation of suggestive matches where an alternative(s) application area for the resource exists. Another aspect empowered by semantic matching is the notion of "partial matching." By current practice and in conditions where registered industries can only partially satisfy a request, the "exact" matching procedure returns no results, completely discarding possibilities, where the requester might have been able to negotiate a symbiosis or find a combination of matches to fulfil its needs. In contrast, the semantic matching engine provides the requesting industry with a list of these matches as potential choices and opportunities for processing, re-using and replacing their resources while measuring and ranking their degree of relevance. This allows industries to analyse all possibilities and therefore make informed decisions in establishing a synergy.

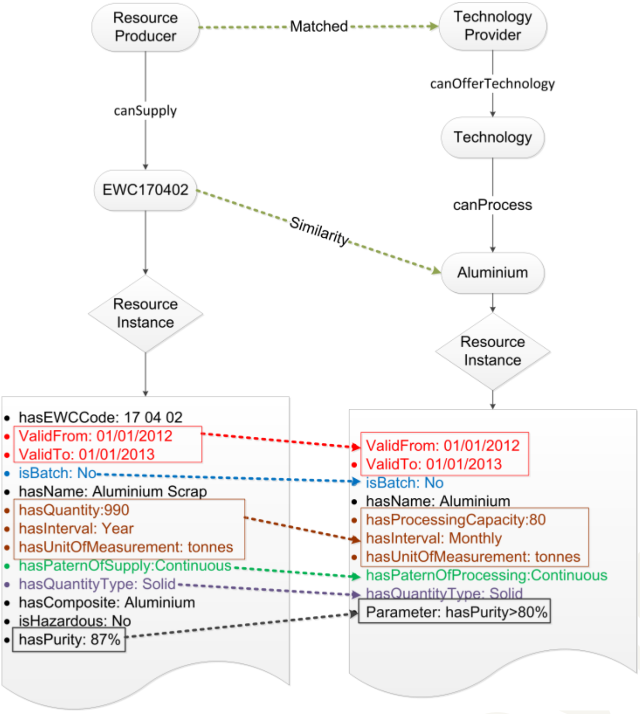

Figure 9 Matching Example

6.3. Types of Matching

6.3.1. Direct Matching

Direct matching occurs when the input and output of the two industries are matched taking into account any semantic similarity that exists between the resource types. In consequence, the system is not only bound to identifying exact matches but can also semantically interpret the relationship that different concepts have with each other to provide intelligent solutions to the requester, as shown in Figure 10**Error! Reference source not found.**.

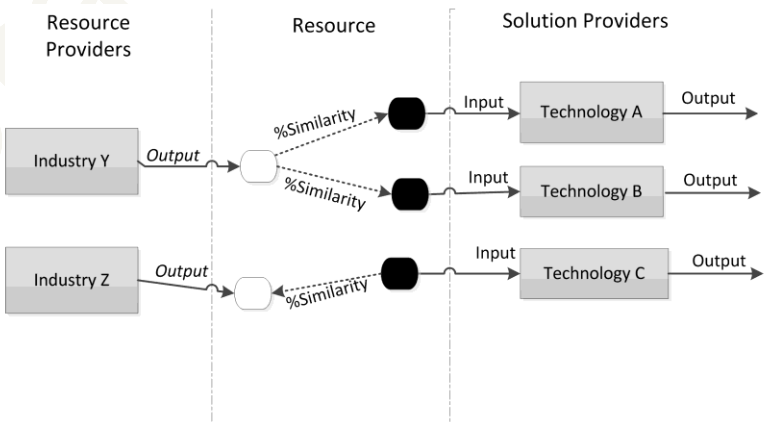

Figure 10 Direct Matching

6.3.2. *DECOMPOSITION MATCHING*

Resource decomposition is performed when a certain type of resource could be decomposed before being processed or used by any of available solutions. The knowledge regarding the composition of the resource is mainly extracted by the tacit knowledge modelled in the domain ontology using the property *hasComposite* and also by the information each user provides regarding the composition of its resource, during the registration process. This type of matching increases chances of discovering a beneficial symbiotic match, as there might be technologies which are capable of processing the composites of the resource rather than all the resource itself. This requires a one–to-many match, between the waste/resource produces and technology providers which are capable of processing any composition. The resource decomposition type of matching is illustrated in Figure 11. The composition could also happen regarding the availability of the resource.

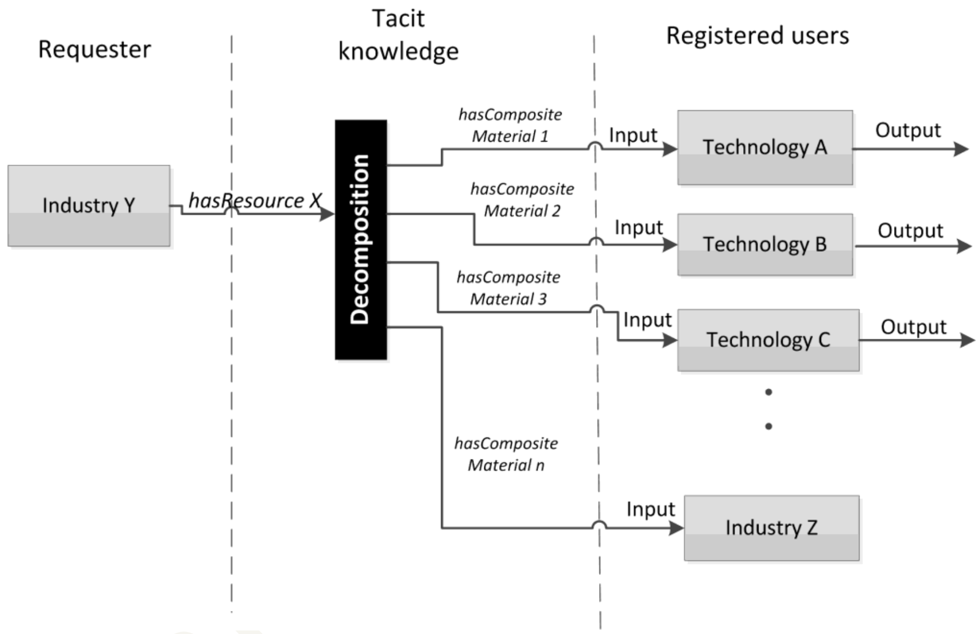

Figure 11 Resource decomposition matching

The process of matching is performed in three stages: i) elimination, ii) semantic matching and ranking, and iii) performance ranking, as shown inFigure 12.

6.4. THE LEVELS OF MATCHING

the process of input – output matching is performed in two stages: i) the stage of elimination, and ii) the stage of calculating the similarity measures between request instance and instances which were not eliminated.

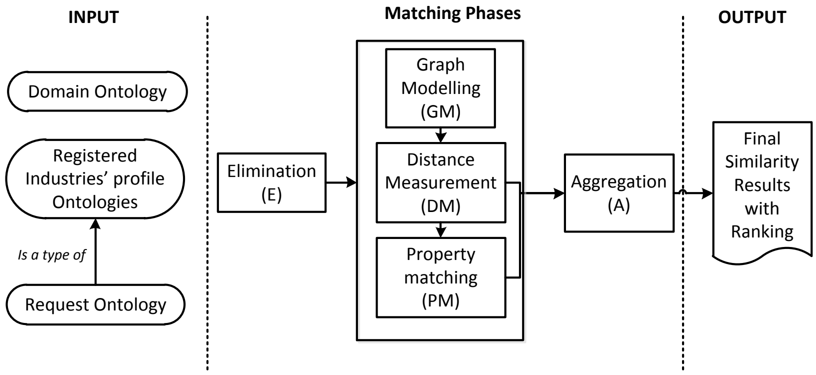

Figure 12 Levels of Matching

The process of eliminations is introduced to minimise redundant matching and hence to computationally speed up the process. Three categories of elimination are introduced: i) elimination based on requestor role, ii) elimination based on the nature of the resource in terms of hazardousness, and iii) elimination based on availability of resources. Additionally, Availability, which is characterised by properties *isValidFrom* and *isValidTo*, is measured by the time overlap between the requestor and matched instances.

Role: Users of the IS platform are either offering a resource (waste, material, product, technology) or looking for a resource. Therefore if the industry is a resource producer it will only be matched against resource consumers and enablers (including technologies and experts) and vice versa. The user type has been introduced as a concept within the domain ontology and the subclasses presenting possible roles (technology provider, resource consumer, etc.) have been introduced as disjoints. If a user can provide both technology and resource they will have two different profiles separately registered under the relevant user type. Each instant of service description has a property known as the "service Category" which points to the "*usertype"* concept within the domain ontology and is used to perform elimination.

Hazardousness: If the requestor is providing or asking for a hazardous material, all instances which are not categorised as hazardous are eliminated or vice versa. The hazardousness of the resource is implied by a Boolean data property within the service description ontology.

Availability: In the practice of IS, availability is one of the most important factors. If a resource is requested by a user, they have need for it through a clearly specific period. Therefore, if other registered users are not able to provide this resource or process it during the required period they should not be considered a potential match and should be eliminated at the very first stage. However, the eliminations will be done only if there is 0 overlap between the availability of registered resources and the requested resource. If there is an overlap (however small) a partial match could exist which is a major focus of semantic matching and those users should not be eliminated from the potential list of matches .Their overlap period will be used for calculating the final degree of match in the aggregation phase. If resource A is available during time$T(A) = (\boldsymbol{A}start, \boldsymbol{A}end)$ and Resource B is available during $T(B) = (Bstart, Bend)$ then their overlap is calculated as below:

$$Overlap(A,B) = T(A)\ \Pi\ T(B) = Min(Aend, Bend) - Max(Astart, Bstart)$$

$$If \quad \begin{cases} Overlap(A,B) \leq 0 & eliminate\ service \\ Ovelrap(A,B) > 0 & use\ overlap\ degree\ for\ similarity\ measurement \end{cases}$$

6.5. THE ALGORITHM:

The semantic relevance of the industries is primarily based on the similarity of the resource type (Distance Measurement in **Error! Reference source not found.**) which is further enhanced by atching of resource characteristics and general industry information (Attribute Matching in **Error! Reference source not found.**), which are aggregated together into a similarity measure.

6.5.1. GRAPH MODELLING OF THE ONTOLOGY

An ontology O is defined by its set of concepts C organized in corresponding subsumption H_C, established relation R_C between concepts in the form of Object Properties also organized entities I with and axioms A used to infer knowledge from existing ones and data properties :

$$O = \{C, H_c, R, H_r, I, R_I\ A\}$$

where

$$R = \{R_I, R_E\}$$

$$R_I = \{R_o, R_d, R_a\}$$

$$R_E = Built\ in\ owl\ properties$$

The object properties themselves can then be further defines

R_C : The object properties between concepts C.

R_D : The data type properties between concepts and values.

R_A : The annotation properties of all entities in the ontology O.

In order to calculate the semantic relevance between concepts of the ontology (*C*) using distance measurement techniques, the ontology first needs to be modelled into a graph format. There are numerous works done on modelling an ontology into a graph for the purpose of ontology matching or alignment. Each work takes a different approach and elements of the ontology to present the graph. Some modelling are based on bipartite graphs taking into account only the hierarchical structure of the ontology and translating superclass and subclass relationships(H_r) into a graph. (Similarity flooding) Some other works also take into consideration the domain and range of the internal properties (GMO)(vector space model). However no graph modeling process to our knowledge models axioms (*A*). As explained, the axioms or better known as restrictions in OWL (reference), play an important role in presenting semantic relevance in the domain of IS. We have therefore modelled the IS domain ontology as a directed bipartite graph based on both external and internal properties as well as the restrictions (axioms) defined over internal properties.

$$Graph\ G = \{V, E\} \begin{cases} V = \{C\}\ represnts\ the\ set\ of\ Vertices(nodes)\ which\ are\ ontology\ concpets \\ E = \{R_E, R_o, A\}\ represnts\ the\ set\ of\ \ Edges \end{cases}$$

The identified elements have all been translated as links between graph nodes (ontology concepts). The focus of the semantic matching is the type of resource and therefore only relevant properties have been used to form the graph in order to avoid any unnecessary calculation and reduce time of measurement process. While the properties and axioms demonstrate the links between the concepts there is still a need to identify the strength of each link. Each element is therefore presented with a different weight. Due to the nature of the distance measurement algorithm (shortest path) explained in section 3.3, stronger links have been given a smaller weight. The relationship between a class and itself will be maximum presented as 0 as well as *the owl:equivalentclass*. All properties and their inverse have the same weight. Axioms take the same weight as the property they are defined over. Table 12 presents the elements of the ontology used as links in the graph model along with the weights assigned to them. The graph is formed over the inferred ontology since through reasoning new relationships will be inferred which directly affect the links between various concepts. This way we assure that no semantic relevance between two concepts has been omitted.

Table 12 Properties and restrictions used for modelling the ontology graph

Edge	Weight
Subsumption(is-a)	0.5
equivalency	0
canProcess	0.9
canBeProcessedBy	0.9
hasApplicationIn	0.7
canUse	0.7
hasComposite	*0.6*
isCompositeOf	*0.6*

In OWL there are four different type of property restrictions including *Cardinality, SomeValuesFrom, AllValuesfrom* and *HasValue*. Apart from cardinality the other type of axioms defined on any of the mentioned internal properties are considered as a link between the class over which the restriction is defined and the value of the restriction. Consider the following example where the concept "*EWC120101"* has a restriction **hasComposite some Aluminium** defined over it. This means that We can translate this restriction to a "*hasComposite*" link between EWC120103 and Aluminium .

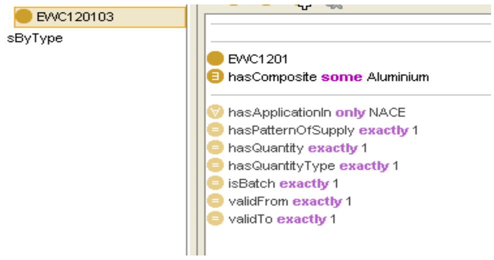

Figure 13 An example of restriction defined over a concept

Below is a snapshot of a small section of the domain ontology modelled as a graph.

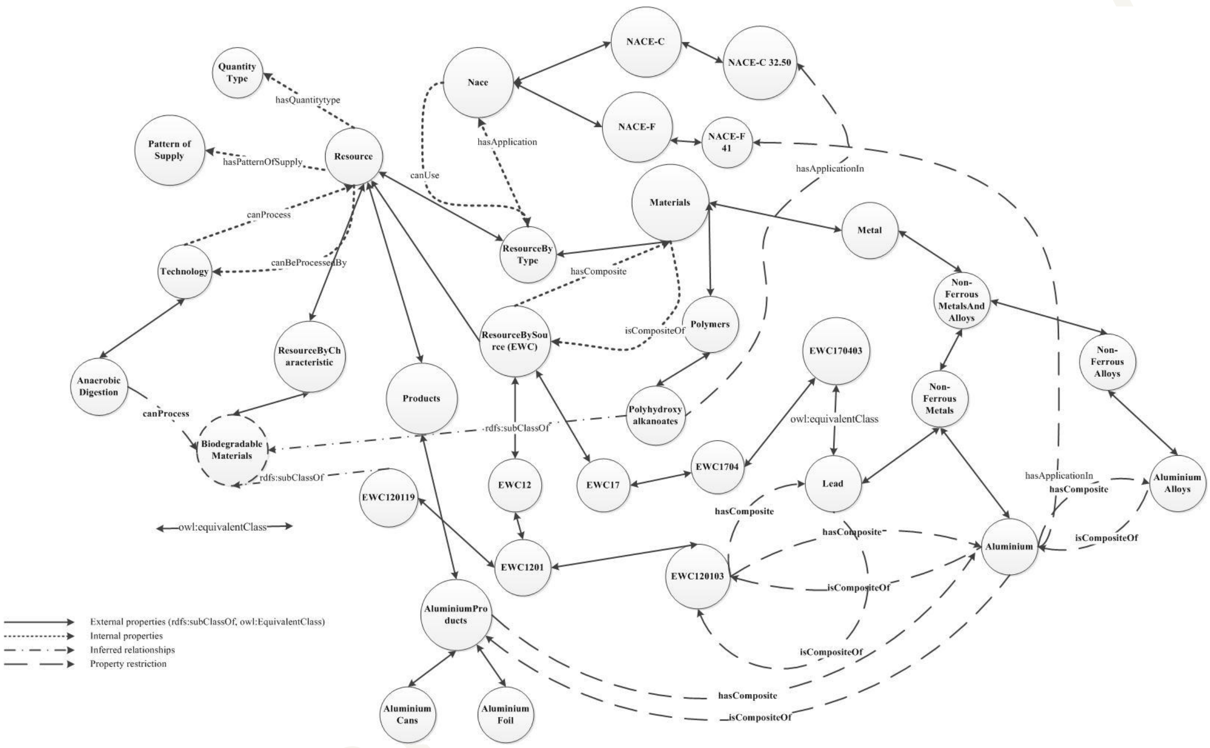

Figure 14 eSymbiosis matching

6.5.2. THE SEMANTIC DISTANCE MEASUREMENT (DM)

The elimination phase passes a list of users which might be potential matches with the request to next matching phase which is Distance Measurement (DM). This is where the semantic similarity between the type of resource specified in the request is measured against the type of resources registered by the other users on the list.The type of resource is specified in the semantic description ontology of the industry through the " *hasType*" property. (service description figure)

The similarity measurement uses a modified version of the *Djkstra*'s shortest path algorithm (reference) upon the graph model of the domain ontology . The modification allows for consideration of disjoint classes through the process of finding the shortest path between concepts. This means that any path that goes through disjoints is disregarded as disjoint concepts do not contribute to the distance measurement similarity in the domain of discourse. As an input, the algorithm takes an nxn matrix where *n* presents the number of nodes in the graph and therefore number of concepts in the ontology. The elements of the matrix are filled based on the weight of the links connecting the nodes. Where there is a direct link between two nodes the weight value is allocated to that element and where there is no direct link between the value "inf" is assigned. Matrix A is an example of the input matrix created for 5 nodes (*Material, EWC120103,Aluminium,metaldust, Biodegradable Material*) of the graph in figure 4.The nodes appear in the mentioned order as indexes of both the rows and columns of the matrix.

$$A = \begin{pmatrix} 0 & inf & inf & inf & 0.5 \\ inf & 0 & 0.6 & 0 & inf \\ inf & 0.6 & 0 & inf & inf \\ inf & 0 & inf & 0 & inf \\ 0.5 & inf & inf & inf & 0 \end{pmatrix}$$

Having the matrix as the input to the algorithm the paths (distances) are then measured for all pairs of nodes. The path lengths determine the dissimilarity between the two nodes which needs to be normalized and converted to similarity. Normalization is done based upon the diameter of the graph or the ***longest distance in the corpus*** which in other words represents the longest, shortest path between two nodes in the graph. The normalization result is between [0,1]; 0 showing maximum similarity and 1 showing no similarity. The similarity between concepts is calculated by formula ?. Table 13 shows a snapshot of similarity measurement between 10 concepts taken from our example graph.

$$\boldsymbol{similiarty} = \left(\mathbf{1} - \underbrace{(\boldsymbol{dissimilarity}/\boldsymbol{max\ dissimilarity})}_{\textbf{Normalized dissimilarity}}\right) * \mathbf{100}$$

Draft for Review

Table 13 Distance Measurement semantic similarity between ontology concepts

	Resource	Material	MaterialsByType	MaterialsBySource_EWC	Plastic	Aluminium	metal_dust	EWC12	EWC120103	AnaerobicDigestion	NACE_A	BiodegradableMaterials	Polyhydroxyalkanoates
Resource	100%	87%	74%	74%	62%	36%	36%	62%	36%	0%	0%	74%	62%
Material	87%	100%	87%	87%	74%	49%	49%	74%	49%	0%	69%	87%	74%
MaterialsByType	74%	87%	100%	0%	87%	62%	62%	0%	0%	0%	56%	74%	74%
MaterialsBySource_EWC	74%	87%	0%	100%	0%	46%	62%	87%	62%	0%	56%	74%	62%
Plastic	62%	74%	87%	0%	100%	0%	0%	0%	0%	0%	44%	74%	87%
Aluminium	36%	49%	62%	0%	0%	100%	0%	0%	0%	0%	18%	36%	0%
metal_dust	36%	49%	62%	62%	0%	0%	100%	74%	0%	0%	18%	36%	0%
EWC12	62%	74%	0%	87%	0%	59%	74%	100%	74%	0%	44%	62%	49%
EWC120103	36%	49%	46%	62%	0%	85%	0%	74%	100%	0%	18%	36%	23%
AnaerobicDigestion	0%	64%	51%	51%	51%	13%	13%	38%	13%	100%	33%	77%	64%
NACE_A	0%	69%	56%	56%	44%	18%	18%	44%	18%	0%	100%	56%	44%
BiodegradableMaterials	74%	87%	74%	74%	74%	36%	36%	62%	36%	0%	56%	100%	87%
Polyhydroxyalkanoates	62%	74%	74%	62%	87%	0%	0%	49%	23%	0%	44%	87%	100%

6.5.3. Property Matching based on Vector Space Modelling (PM)

The service description of users(industries) which have a resource type matching degree of over 50% in the distance measurement phase will be send to this stage. This is the phase where the data properties characterizing the resource concept and also the service itself are matched. As resources need to be compared taking into account numerous properties, a suitable approach is the adaption of vector space modelling .This way the two services will be compared taking all properties into consideration. Each Service description is modelled into a vector as follow

$$\boldsymbol{Sj} = \{\boldsymbol{V}(\boldsymbol{d1}).\boldsymbol{w1}, \boldsymbol{V}(\boldsymbol{d2}).\boldsymbol{w2}, \ldots .. \boldsymbol{V}(\boldsymbol{dn}).\boldsymbol{wn}\}$$

$Where\ \boldsymbol{S}\ is\ the\ Service\ Description\ Vector, \boldsymbol{di}\ is\ the\ data\ property\ attached\ to\ the\ resource$

$where\ \ \boldsymbol{0} < i < n\ ;$

$\boldsymbol{n}\ = Max\ Number\ of\ properties\ attached\ to\ the\ concept;$

$\boldsymbol{V}(\boldsymbol{di}) is\ the\ value\ of\ that\ property\ and$

$$\boldsymbol{wi}\ \ is\ the\ weight\ of\ that\ property$$

It is important for the accurate results to make sure that the service descriptions being compared are of the same dimensions. Therefore if one property value is missing from a description we will replace it with 0. Also vector models are presented in discrete numeric format, therefore continuous data formats such as availability period are converted to a percentage of overlap in reference to the request value. Non numeric values are such as pattern of supply are presented in a pre-defined numeric format. The location of industries is calculated as the distance to the requesting industry's location. Table 3 shows properties used for matching in with respect to the practice of IS

Table 14 Data properties used in vector space matching

Data Property	Value Type	Description
Quantity	Float	The exact value itself will be put in the vector
Location (presented as distance)	Longitude and Latitude	Based on the longitude and latitude the distance to the requester is measured and the value of the distance is used to model the vector
Availability (Valid From- Valid to)	Date	Is measured as the percentage of overlap with the availability specified in the request
Pattern of Supply	Predefined text	Continuous: 1 Batch: 2

Based on properties mentioned in table 1 the service description vector for each user will be modelled as follow:

$$\underline{\boldsymbol{S} = \left(\boldsymbol{v}(\boldsymbol{Quantity}), \boldsymbol{v}(\boldsymbol{PatternofSupply}), \boldsymbol{v}(\boldsymbol{Availability}), \boldsymbol{v}(\boldsymbol{distance})\right)}$$

When the vectors are formed we will use the cosine similarity algorithm to calculate the similarity between the Request vector and potential matching services vectors which were extracted in the previous phase of matching. The similarity has to be calculated between the request and all potential matches one by one.

The formula used for calculating the cosine similarity is as follow:

$$similarity = \cos(\theta) = \frac{A.B}{||A||||B||} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

$Where\ \boldsymbol{A} = \ Service\ Request\ vector$

$\boldsymbol{B} = \ Service\ description\ vector\ of\ each\ potential\ match.$

After input – output matching the users are provided with the set of match ranking based on the approximation of possible outcomes based on performance metrics. This gives the user an idea of how each synergy establishment is likely to benefit the industry from various aspects. The factors included are environmental, financial and social aspects. This enables users to consider potential matches in the context of a wider range of benefits and make informed choices based on their particular business priorities.

6.6. Intelligence empowered by the use of Semantics:

The use of semantics for modelling the IS domain and the industries themselves takes the process of IS further than only automation. It builds intelligence within the system that allows types of matching which are otherwise not possible.

6.6.1. ALTERNATIVE RESOURCE SUGGESTION

Using the implicit and tacit knowledge modelled and inferred from the ontology through relationships between material and the knowledge about their application and use, the system is capable of offering the user with alternative resources and consequently, substitute synergy possibilities.

6.6.2. IDENTIFY TECHNOLOGIES/SOLUTIONS FOR SIMILAR TYPES OF RESOURCE

In situations where an exact solution is not identifiable the system is capable of identifying solutions which handle similar type of resources and present them to the user.

6.6.3. DIFFERENT APPLICATION FOR A RESOURCE

The knowledge about application of resources in certain industries, which is modelled through the "has Application" relation in the domain ontology; empowers the creation of suggestive matches where an alternative(s) application area for the resource exists.

6.6.4. PARTIAL MATCHING

Another aspect empowered by semantic matching is the notion of "partial matching." In conditions where registered industries can only partially satisfy a request, the "exact" matching procedure would return no results, completely discarding possibilities, where the requester might have been able to negotiate a symbiosis or find a combination of matches to fulfil its needs. The semantic matching engine would provide the requesting industry with a list of these matches as potential choices and opportunities for processing, re-using and replacing their resources while measuring and ranking their degree of relevance. This allows industries to analyse all possibilities and therefore make informed decisions in establishing a synergy.

6.6.5. PROVIDING GENERAL OPTIONS

In addition to discovering potential synergy opportunities with members of the eSymbiosis, the users can exploit the platform to gain knowledge about various options and solutions that exists for a type of resource. These options are provided regardless of physical existence of an eSymbiosis member to provide the solution. The suggestions merely serve as general information. As an example, a user looking to identify possible uses of "aluminium" will be provided with the following options.

Potential uses for "Aluminium":

- Aluminium Foil Production
- Bearing Steel Production
- Extruded Aluminium Profiles Production
- Iron Powder Production
- Primary Aluminium Production

- Remelting of Aluminium Scrap
- Rolled Aluminium Sheet Production
- Scrap-based Aluminium Production
- Virgin Aluminium Production

7. TESTING

EnviD ltd is a solution provider, an enterprise that produces chemicals for a wider market. It has an anaerobic digestion process facility which is planned to be free in foreseeable future. The company wishes to fill in free capacities and to explore the opportunities for new partnerships. Through the registration process EnviD ltd registers as a solution provider and provides other information essential for the matching process, as shown in Table 15.

Table 15 EnviD ltd. solution information

Process	***Required resources***	***Quantity (t/w)***	***Pattern of supply***	***Availability***		***Geographical location***	
				From	***To***	***Latitude***	***Longitude***
Anaerobic digestion	*Lignocelluloses*	*150*	*continuous*	*09/08/2012*	*08/12/2015*	*38.339*	*23.61278*

Note that the properties which have been filled to describe the resource and the company itself are those marked as mandatory properties. These properties serve as minimum requirement for identifying potential matches and are implemented as a must be field property in the ontology.

Hartex ltd is a cardboard production company producing various heavy machinery packaging products through pulping pine chips. The company has a production plan for the coming four year period and expects to generate a large quantity of waste. Hartex ltd has engaged a waste manager who identified waste as EWC type EWC30308 – the waste produced from paper and cardboard production. The company registers as a resource provider with details given Table 16.

Table 16 Hartex ltd resource information

Resource	***Quantity (t/w)***	***Pattern of supply***	***Availability***		***Geographical location***	
			From	***To***	***Latitude***	***Longitude***
EWC30308	*90*	*continuous*	*04/07/2012*	*03/06/2016*	*38.34567*	*23.63116*

A high number of other companies have already been registered and hence instantiated the ontology either as solution providers or as resource providers. Note that each solution provider also provides a product which can be treated as a resource for other solution providers, the chaining criterion in IS sense.

After the matching request by EnviD ltd has been placed, the matching process starts in stages. The elimination stage eliminates all the instances i) which are instantiated in the domain ontology as a solution provider role, ii) which have different type of hazardousness and iii) which have no overlap of time availability. The remaining companies that could potentially provide matches with similarity in addition to Hartex ltd are listed in Table 17. SP stands for Pattern of Supply and 1 shows a continuous pattern while 2 presents a batch production. Q stands for Quantity.

Table 17 Profile of registered industries offering potential matches with EnviD ltd

ID	Company	Resource type(output)	Q	SP[§]	Availability		Location	
					Valid From	Valid to	Lat	Long
27	Timberium	Wood	230	2	14/11/2014	14/12/2015	38.325	23.600
22	WoodSol	MDF	50	2	07/08/2013	17/10/2015	38.326	23.581
187	GrePack	Cardboard	450	2	07/08/2013	17/12/2014	38.329	23.612
144	Farmex	EWC020103	80	1	07/08/2012	14/12/2015	38.325	23.631
44	The Fishery	EWC020705	70	2	09/08/2012	09/09/2016	38.342	23.581
1	Municipality	EWC030301	90	1	06/11/2012	05/07/2016	38.345	23.611
19	Furnumil	Lignocellulosic	200	1	04/07/2012	17/10/2015	38.378	23.631

The distance measurement similarity h_k^C (Section 6.5.2) between requesting and other instances in the domain ontology is used as the measure of input/output resource type match. The measurement starts by presenting the resource cluster of the ontology which is enriched with the qualitative attributes of the semantic objects and numerical data, as a directed graph.

The excerpt of the domain ontology illustrating the input–output matching between EnviD ltd and Hartex ltd. is shown in Figure 15. This part of the ontology also has associated two restrictions along the relationship *hasComposite*: $f_D = \exists\, hasComposite\ \mathrm{Paper}$ on the concept EWC030308 with semantic that EWC030308 type of waste can be composed of paper and $f_D = \exists\, hasComposite\ \mathrm{Ligno-cellulosesProducts}$ also on the concept EWC030308 with semantic that EWC030308 could also be composed of lignocelluloses.

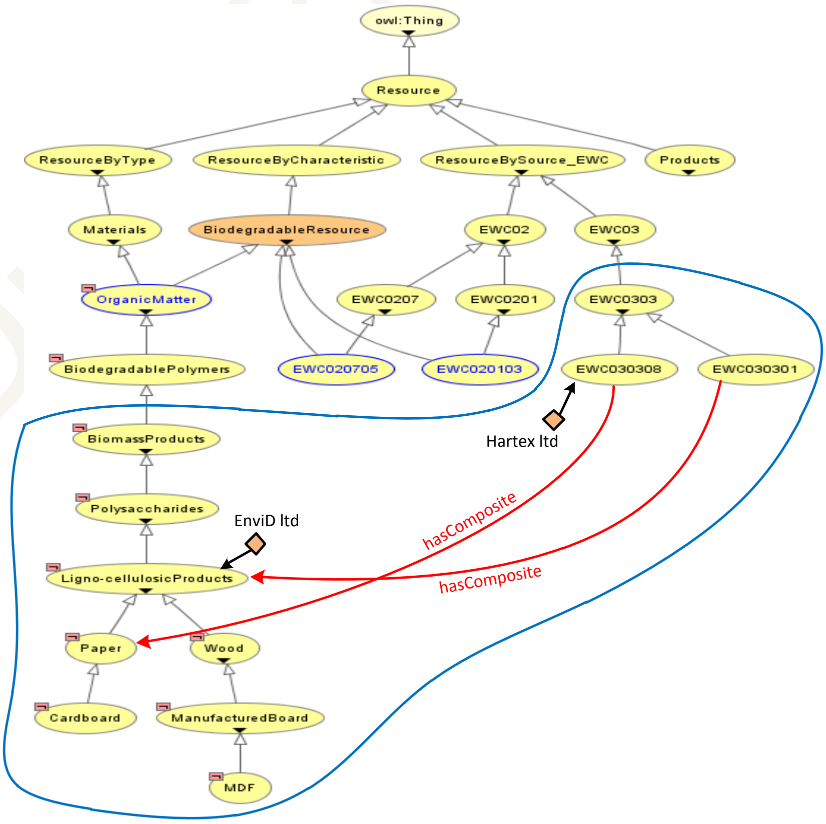

Figure 15 Excerpt of the domain ontology used for input-output matching

The respective ontology graph matrix, which is the input to the distance matching algorithm; including concepts in the order of *"EWC0303", "EWC030301", "EWC030308", "BiomassProducts", "Polysaccharides", "Lingo-cellulosicProducts", "Paper", "Wood", "Cardboard"," ManufactureBoard"* has the form

$$M = \begin{pmatrix} 0 & 0.4 & 0.4 & inf & inf & inf & inf & inf & inf & inf \\ 0.4 & 0 & inf & inf & inf & 0.5 & inf & inf & inf & inf \\ 0.4 & inf & 0 & inf & inf & inf & 0.5 & inf & inf & inf \\ inf & inf & inf & 0 & 0.4 & inf & inf & inf & inf & inf \\ inf & inf & inf & 0.4 & 0 & 0.4 & inf & inf & inf & inf \\ inf & 0.5 & inf & inf & 0.4 & 0 & 0.4 & 0.4 & inf & inf \\ inf & inf & 0.5 & inf & inf & 0.4 & 0 & inf & 0.4 & inf \\ inf & inf & inf & inf & inf & 0.4 & inf & 0 & inf & 0.4 \\ inf & inf & inf & inf & inf & inf & 0.4 & inf & 0 & inf \\ inf & inf & inf & inf & inf & inf & inf & 0.4 & inf & 0 \end{pmatrix} \quad (7.1)$$

As explained in the matching algorithm (Section 6.5) , the elements in matrix M present values of a direct link between two nodes, whereas nodes which are not linked are assigned with value infinity. Edges along class-subclass relationship are weighted 0.5 as states in Table 12, whereas edges along restriction on the relationship "*hasComposite*" have the weight 0.6 (Table 12) to reflect the semantic reality associated with IS.

The shortest distance between the two concepts *"EWC030308"* and *"Lingo-cellulosicProducts"* are recovered by two edges through the concept *Paper* with the total value 0.9, as shown in Figure 16, which normalised to the longest path between them in the ontology gives $\delta' = (distance)/D_c = 0.9/4.5 = 0.2$ and then the similarity is

$$h_k^C = (1 - 0.2) = 0.8 \; (= 80\%) \quad (31)$$

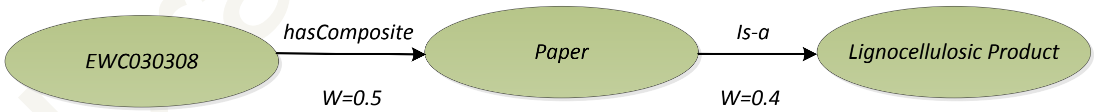

Figure 16 Distance measurement example

The property similarity is calculated between the two companies using properties $P_i^4 = \{v(\,Availability), v(\,Quantity), v(PatternOfSupply), v(\,Location)\}$. Since the requestor has not provided any information for the other matching metrics properties and regarding that their importance in the matching is non-critical, they have been eliminated as a matching dimension. The *Availability* is calculated as the overlap period $T_o = 1216$ days which results in 100% overlap between the two companies. The property *Location* is calculated as the distance between the two companies using Haversine formula which in normalised form gives $L_{distance} = 1.7$. The quantity is also calculated in relevance to the required quantity indicating a 60% coverage of the required quantity by EnviD. Hence, P_i^4 is modelled as 4-dimensional vectors which for EnviD ltd and Hartex ltd are $\mathbf{p}_1 = (100, 100, 100, 0)$ and $\mathbf{p}_2 = (60, 100, 100, 37.36)$, respectively.

The similarity measurement provides a cosine similarity (Section 6.5.3) $h_k^{V,c} = 0.949$ and Euclidean similarity $h_k^{V,E} = 0.65$. When combined together as the final result of property similarity (PM) phase of the matching of t between the two companies calculated as $h_k^V = \left(h_k^{V,c} + h_k^{V,E}\right)/2 = 0.64$ they present a 64% similarity.

Having measured the two phases, the aggregated similarity between the two companies EnviD ltd and Hartex ltd from eq. (4.19) is $h_k = \left(\alpha h_k^C + \beta h_k^V\right)/(\alpha + \beta) = 0.74\ (= 74\%)$ where $\alpha = 0.6$ and $\beta = 0.4$ reflecting that the type of resource of the companies has a greater effect on the possibility of the establishment of a synergy.

Following the same procedure, the matches with other companies are determined and the final results are summarised in Table 18.

Table 18 Complete set of results

Company	Semantic distance Similarity	Average property similarity	Aggregated results	Similarity percentage
Furnumil	1	0.487325	0.79493016	79%
Timberium	0.89	0.221338	0.622535374	62%
Municipality	0.87	0.909041	0.885616538	89%
GrePack	0.78	0.239821	0.56392828	56%
Hartex	0.80	0.64	0.743529759	78%
WoodSol	0.67	0.283113	0.51524522	52%
Farmex	0.44	0.764722	0.569888781	57%
The Fishery	0.44	0.398214	0.423285785	42%

8. CONCLUSIONS & SUMMARY

Activity 2.1. involved the design and implementation of waste streams and processing technologies knowledge models that support the semantic description of users and their resources or solutions. The ontology (eSYMBIOSIS ontology) is the backbone of the platform. The eSYMBIOSIS ontology supports the three key functionalities of the eSYMBIOSIS platform:

(i) Navigation of the user through the registration process. The path that the user follows during the registration process is dictated by the ontology. During this process the concepts and properties of the ontology are used for information collection.

(ii) Provide a common vocabulary for IS domain. The ontology provides a standardised vocabulary for the given domain and helps in tackling heterogeneity.

(iii) Support the input – output matching for synergy identification. The ontology is used as a common reference for the matchmaking process. The Material stream is the key in this process as it is used as the reference for calculating similarity between inputs and outputs. The use of semantics allows for partial matching.

The use of ontologies and semantics facilitates a new approach for Industrial Symbiosis, based on semantic description of processing technologies and resources. The participating industries are described using Semantic Web Service (SWS) description ontology adapted for IS, along with information about their technologies and/or resources.

The ontology, as at present, has four (4) levels of abstraction and consists of some 2000 concepts. All ontology concepts have been labelled (in both Greek and English) to enable sharing and reusing.

SWS ontologies have established relationships with the IS domain ontology which enables acquisition of explicit knowledge from the users, their purposeful discovery and partial matching with other technologies based on the input output type, technological, economic and environmental properties. The level of suitability to participate in IS is assessed by their contribution. Input – output matching, the key to form IS networks, is performed by a matchmaker built for IS network integration as a direct matching solution. The algorithm behind the matchmaker uses both tacit and explicit knowledge in order to calculate the similarity score.

9. ΣΥΜΠΕΡΑΣΜΑΤΑ & ΠΕΡΙΛΗΨΗ

Η Δράση 2.1. περιλαμβάνει το σχεδιασμό και την υλοποίηση των οντολογιών (μοντέλα γνώσης) που περιγράφουν τα απόβλητα και τις τεχνολογίες επεξεργασίας και υποστηρίζουν τη σημασιολογική (semantic) περιγραφή των χρηστών και των πόρων τους ή τεχνολογιών τους. Η οντολογία (οντολογία eSYMBIOSIS) είναι ο κορμός της πλατφόρμας. Η οντολογία eSYMBIOSIS υποστηρίζει τρεις βασικές λειτουργίες της πλατφόρμας eSYMBIOSIS:

(i) Πλοήγηση του χρήστη μέσω της διαδικασίας εγγραφής. Η πορεία που ακολουθεί ο χρήστης κατά τη διάρκεια της διαδικασίας εγγραφής υπαγορεύεται από την οντολογία. Κατά τη διάρκεια αυτής της διαδικασίας συλλέγονται πληροφορίες με τη χρήση των οντοτήτων (concepts) και των ιδιοτήτων (properties) της οντολογίας.

(ii) Παρέχει ένα κοινό και ενιαίο λεξιλόγιο για τον τομέα της Βιομηχανικής Συμβίωσης. Η οντολογία παρέχει ένα τυποποιημένο λεξιλόγιο για το συγκεκριμένο τομέα και βοηθά στην αντιμετώπιση της ανομοιογένειας μεταξύ ορολογιών.

(iii) Υποστήριξη της ταυτοποίησης εισροών - εκροών για τη δημιουργία συμβιωτικών συνεργειών. Η οντολογία χρησιμοποιείται ως κοινό σημείο αναφοράς για τη διαδικασία ταυτοποίησης. Η οντολογία που περιγράφει τα υλικά είναι το κλειδί σε αυτήν τη διαδικασία, καθώς χρησιμοποιείται ως αναφορά για τον υπολογισμό της ομοιότητας (similarity) μεταξύ των εισροών και εκροών. Η χρήση της σημασιολογικών τεχνολογιών επιτρέπει την ταυτοποίηση χρηστών που δεν καλύπτουν το σύνολο των προδιαγραφών.

Η χρήση των οντολογιών δίνει τη δυνατότητα για μια νέα προσέγγιση στη βιομηχανική συμβίωση, με βάση τη σημασιολογική περιγραφή των τεχνολογιών επεξεργασίας και των πόρων. Οι συμμετέχουσες βιομηχανίες περιγράφονται χρησιμοποιώντας την οντολογία Υπηρεσιών του Σημασιολογικού Ιστού (Semantic Web Service), προσαρμοσμένη στα πλαίσια της Βιομηχανικής Συμβίωσης, μαζί με πληροφορίες σχετικά με τις τεχνολογίες τους και / ή τους πόρους.
Η οντολογία, στη σημερινή της μορφή, έχει τέσσερα (4) επίπεδα και αποτελείται από περίπου 2000 έννοιες (concepts). Όλες οι έννοιες οντολογίας έχουν επισημανθεί (στα ελληνικά και τα αγγλικά) για να διευκολυνθεί η επαναχρησιμοποίησή της.

Οι οντολογίες Υπηρεσιών του Σημασιολογικού Ιστού (Semantic Web Service) συνδέονται με τη οντολογια eSymbiosis, επιτρέποντας την εξαγωγή ρητών πληροφοριών από τους χρήστες και την εκμετάλλευσή τους κατά τη διαδικασία ταυτοποίησης εισροών – εκροών. Η ταυτοποίηση γίνεται από έναν αλγόριθμο που λαμβάνει υπ’ όψη τη ρητή και την άρρητη γνώση που περιλαμβάνουν οι οντολογίες κατά τον υπολογισμό του βαθμού ομοιότητας (similarity score).

References

Alberta, C. C. and H. G. Kevin (2008). Use of Industrial By-Products in Urban Transportation Infrastructure: argument for increased industrial symbiosis. IEEE, IEEE.

Bain, A., M. Shenoy, W. Ashton and M. Chertow (2010). Industrial symbiosis and waste recovery in an Indian industrial area Resources, Conservation and Recycling **54**(12): pp 1278-1287.

Behera, S., J. Kim, S. Lee, S. Suh and H. Park (2012). Evolution of 'designed' industrial symbiosis networks in the Ulsan Eco-industrial Park: 'research and development into business' as the enabling framework. Journal of Cleaner Production **29-30**: pp 103 - 112.

Bridgwater, A. (2003). Renewable fuels and chemicals by thermal processing of biomass. Chemical Engineering Journal **91**: pp 87 - 102.

Chertow, M. (2000). INDUSTRIAL SYMBIOSIS: Literature and Taxonomy. Annu. Rev. Energy Environ. **25**: pp 313-337.

Chertow, M. (2004). Industrial Symbiosis. Encyclopedia of Energy: pp 407-415.

Chertow, M. (2007). Uncovering Industrial Symbiosis. Journal of Industrial Ecology **11**(1): pp 11-29.

Chertow, M. and J. Ehrenfeld (2012). Organizing Self-Organizing Systems - Toward a Theory of Industrial Symbiosis. Journal of Industrial Ecology **16**(1): pp 13-27.

Cherubini, C., G. Jungmeier, M. Wellusch, T. Willke, I. Skiadas, R. Van Ree and E. de Jong (2009). Toward a common classifi cation approach for biorefi nery systems. Biofuels, Bioproducts & Biorefining **3**: pp 534 - 546.

EC (2002) "European Waste Catalogue."

EC (2010). COMMISSION REGULATION (EU) No 849/2010: GENERATION OF WASTE. Official Journal of the European Union **28.9.2010**: pp.

Eggelston, S. (2006). METAL INDUSTRY EMISSIONS. 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Institute for Global Environmental Strategies (IGES) for the IPCC**:** pp.

Grant, G., T. Seager, G. Massard and L. Nies (2010). Information and Communication Technology for Industrial Symbiosis. Journal of Industrial Ecology **14**(5): pp 740-753.

Gruber, T. (1995). Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies **43**(5-6): pp 907-928.

Hu, W., N. Jian, Y. Qu and Y. Wang (2005). A Graph Matching for Ontologies. K-Cap 2005 Workshop on Integrating Ontologies**:** pp 43-50.

Jacobsen, N. (2006). Industrial Symbiosis in Kalundborg, Denmark: A Quantitative Assessment of Economic and Environmental Aspects. Journal of Industrial Ecology **10**(1-2): pp 239-255.

Jae-Yeon, W., K. Jung-Hoon, L. Sang-yoon and P. Hung-Suck (2006). Industrial Symbiosis as an Integrated Business/Environment Management Process: The Case of Ulsan Industrial Complex. IEEE, IEEE.

Jensen, P., L. Basson, E. Hellawell, M. Bailey and M. Leach (2011). Quantifying 'geographic proximity': Experiences from the United Kingdom's National Industrial Symbiosis Programme. Resources, Conservation and Recycling **55**: pp 703-712.

Junli, W., D. Zhijun and J. Changjun (2006). GAOM: Genetic Algorithm based Ontology Matching. IEEE Asia-Pacific Conference on Services Computing (APSCC'06), IEEE Asia-Pacific Conference on Services Computing (APSCC'06).

Karadimas, N., V. Loumos and O. Mavrantza (2004). Quality of Service Ensuring in Urban Solid Waste Management. SECOND IEEE INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS.

Karlsson, M. and A. Wolf (2008). Using an optimization model to evaluate the economic benefits of industrial symbiosis in the forest industry. Journal of Cleaner Production **16**: pp 1536e1544.

Khalid, A., M. Arshad, M. Anjum, T. NMahmood and L. Dawson (2011). The anaerobic digestion of solid organic waste. Waste Management **31**: pp 1737 - 1744.

Korhonen, J. (2004). Industrial ecology in the strategic sustainable development model: strategic applications of industrial ecology Journal of Cleaner Production **12**(8-10): pp 809-823.

Lehtoranta, S., A. Nissinen, T. Mattila and M. Melanen (2011). Industrial symbiosis and the policy instruments of sustainable consumption and production. Journal of Cleaner Production **19**: pp 1865e1875.

Lombardi, D. and P. Laybourn (2012). Redefining Industrial Symbiosis: Crossing Academic–Practitioner Boundaries. Journal of Industrial Ecology **16**: pp 28 - 37.

Martin, D., M. Burstein, O. Hobbs, D. MsDwrmott, S. McIlaith, S. Narayanan, M. Paolucci, B. Parsia, T. Payne, E. Sirin, N. Srinivasan and K. Sycara. (2004). "OWL-S: Semantic Markup for Web Services." Retrieved 20/09/2012, 2012, from http://www.w3.org/Submission/OWL-S/.

Martin, M. and M. REklund (2011). Improving the environmental performance of biofuels with industrial symbiosis. Biomass and Bioenergy **35**: pp 1747 - 1755.

McCarthy, I. (1995). Manufacturing classification: Lessons from organizational systematics and biological taxonomy. Integrated Manufacturing Systems **6**: pp 37 - 48.

Melnik, S., H. Garcia-Molina and E. Rahm (2002). Similarity flooding - a versatile graph matching algorithm. Proc 18th Int Conf Data Eng.

Meneghetti, A. and G. Nardin (2012). Enabling industrial symbiosis by a facilities management optimization approach. Journal of Cleaner Production **35**: pp 263 - 273.

Mirata, M. and T. Emtairah (2005). Industrial symbiosis networks and the contribution to environmental innovation: The case of the Landskrona industrial symbiosis programme Journal of Cleaner Production **13**(10-11): pp 993-1002.

NACE (2008). NACE Rev. 2: Statistical classification of economic activities in the European Community, eurostat - European Commission**:** pp.

Raafat, T., F. Cecelja, A. Yang and N. Trokanas (2012). Semantic Support for the Industrial Symbiosis Process. Computer-Aided Chemical Engineering **30**: pp 452 - 456.

Raafat, T., N. Trokanas, F. Cecelja and M. Cirovic (2013). Semantic Algorithm for Industrial Symbiosis Service. To be published in European Journal of Operational Research: pp.

Ratafia-Brown, J., L. Manfredo, J. Hoffmann and M. Ramezan (2002). MAJOR ENVIRONMENTAL ASPECTS OF GASIFICATION-BASED POWER GENERATION TECHNOLOGIES, NETL, U.S. Department of Energy, Office of Fossil Energy, National Energy Technology Laboratory**:** pp.

Schmoch, U. (2008). Concept of a Technology Classification for Country Comparisons, Final Report to the World Intellectual Property Organisation (WIPO)**:** pp.

The-Carbon-Trust (2011a). International Carbon Flows: Aluminium**:** pp.

The-Carbon-Trust (2011b). International Carbon Flows: Steel**:** pp.

Tous, R. and J. Delgado (2006). A vector space model for semantic similarity calculation and OWL ontology alignment. Proceedings of DEXA**:** pp 307-316.

Trokanas, N., T. Raafat, F. Cecelja, A. Kokossis and A. Yang (2012). Semantic Formalism for Waste and Processing Technology Classifications Using Ontology Models. Computer-Aided Chemical Engineering **30**: pp 167 - 171.

Turban, E. and J. Aronson (2001). Decision Support Systems and Intelligent Systems. New Jersey, Prentince-Hall, Inc.

Veiga, L. and A. Magrini (2009). Eco-industrial park development in Rio de Janeiro, Brazil: a tool for sustainable development Journal of Cleaner Production **17**(7): pp 653-661.