Industry Taxonomy Engineering: the Case of the European Software Ecosystem

Ivo Hunink, Rene van Erk
ISVWorld
Utrecht, The Netherlands
{ivo, rene}@isvworld.com

Slinger Jansen, Sjaak Brinkkemper
Utrecht University
Utrecht, The Netherlands
{s.jansen, s.brinkkemper}@cs.uu.nl

ABSTRACT

Presently, no methods exist that support the creation process of an industry taxonomy within a specific domain. Without such a method, taxonomies remain erroneous, making the development of, for instance, a directory of companies for research, extremely hard. This paper presents a method for creating complete and encompassing domain specific taxonomies. With such a method, researchers can create complete, consistent, taxonomies that in turn provide a strong basis for data model development. The method is applied in practice, and the industry taxonomy is evaluated by practitioners.

Categories and Subject Descriptors

D.3.3 [**Programming Languages**]: Language Contructs and Features – *abstract data types, polymorphism, control structures.*

Keywords

Industry Taxonomy Engineering, Software Ecosystem.

1. INTRODUCTION

According to Cheng et al. [1], taxonomies, which represent the vocabularies commonly used by the industry practitioners, are being developed and used increasingly for and by specific industry practitioners, to facilitate information interoperability and retrieval. According to Cheng and others, this interoperability is vital to accessibility, analyzability, and the possibility to combine information, which consequently increases the value of information [1]. As a result, the development of taxonomies is now viewed as a core business of an industry [2]. Without industry specific taxonomies, there is a lack of interoperability, which results in significant economic costs [2], [3], [4].

Nowadays, terms like “classification”, “ontology” and “thesaurus” are used abundantly. The term “taxonomy” is no exception. As mentioned by Rees [5], the distinction between these terms is often blurred. According to Smith et al. [6] the term “taxonomy” is frequently used interchangeably with the words “simple ontology”, and Garshol [7] states that when referring to a taxonomy, it can be just about anything, albeit mostly about an abstract structure for information. Therefore, when using these terms, we need to provide further qualification, in order to remove ambiguity.

In Merriam-Webster, a classification is defined as “systematic arrangement in groups or categories according to established criteria” [8]. A taxonomy is defined as an “orderly classification of plants and animals according to their presumed natural relationships” [9]. An ontology is defined as “a specification of a conceptualization”. Rees [5] redefines these terms to be able to apply them more generally. First of all, Rees states that a classification is used to classify pieces of information in classes, according to external criteria. Then, a taxonomy is used to classify information according to internal criteria and properties. In addition, many authors write that a taxonomy can be compared to a simple ontology [6], [5]. As more properties and relationships are added to such a taxonomy or simple ontology, the term “ontology” becomes more applicable.

To overcome the lack of clarity when it comes to the terms used in this paper, we define each term briefly below. First of all, we define a *classification* as a structure that allows the classification of entities in a class according to external criteria. A *taxonomy* is defined as a structure allowing the classification of entities according to internal criteria, properties and relationships. We intentionally do not call this an ontology or simple ontology because these are used for complex structures that have support for automatic reasoning.

Finally, we need to define the term “industry taxonomy”. Bruno & Richmond [10] define five types of taxonomies that represent information on different levels within an organization. The first type (1) is the functional taxonomy, which organizes the functions performed by an organization. Bruno and Richmond's second taxonomy type (2) is the department taxonomy that is based on and mirrors an organizational chart. The third (3) taxonomy type that is defined is the subject taxonomy which is based on the subjects of information with which an organization might deal. Fourth (4), the product and services taxonomy that is based on the products and services that the organization provides. Finally, the fifth taxonomy type is the location taxonomy, which is based on an organization’s geographical locations. By combining these taxonomy types, a taxonomy is created that represents one organization. Then, by adding elements that represent these types of organizations within an industry, even an industry as a whole can be represented. This could be called an industry taxonomy. Formally, within this paper, we define an industry taxonomy as a structure allowing the classification within one industry of business entity related information , according to internal criteria, properties and relationships, capturing among others the functions performed, its products and its services.

When compiling an industry taxonomy, the most prominent criteria are determined by its intended use. The closer the taxonomy matches its intended use, the more likely it is to be accepted by its end-users. In addition, Guibert et al. [11] states that the more logical this ‘picture’ of the industry is, the better the analysts will be able to see it as a single object of analysis; as a subject with its own autonomy, i.e. with a behavior and with reactions.In addition, a custom developed taxonomy can be specific to an industry, its objectives and cultures. When developing, it must be made sure the taxonomy reflects the understanding and needs of an intended audiences, as well as the range of content to which it will be applied [12]. Therefore, we present a method to engineer Industry Taxonomies, the Industry

Taxonomy Engineering Method (ITAXEM), and apply this method to the European software ecosystem. This enables European software vendors, the scientific world and governmental organizations to get insight in the European software ecosystem, and to identify the gaps between needed and shared information.

This paper continues with a discussion on related work in section 2, which leads, using the research approach in section 3, to the ITAXEM in section 4. In section 5, the ITAXEM is evaluated within a case study focusing on the software ecosystem, which finally leads to the European Software Industry Taxonomy (EUSOIT). Then, the research is concluded and directions for future research are given in sections 6 and 7.

2. RELATED WORK

To overcome this lack of interoperability, pre-built taxonomies exist. The term pre-built taxonomy refers to those taxonomies built for a specific reason by an organization. On the one hand, special vendors offer specific taxonomy templates for industries like aerospace, architecture and design, and finance and accounting businesses [12]. On the other hand, industry specific associations offer specific pre-built taxonomies like the Exploration and Production Taxonomy, which has been developed by the Petro-technical Open Software Consortium, the Public Petroleum Data Model Association, and Shell. If a pre-built taxonomy exactly satisfies the objectives of a specific project, it can be used. However, when using a pre-built taxonomy for another purpose than it was built for, using pre-built taxonomies has its disadvantages: they are not specific to an industry's objective and might introduce unfamiliar terminology that makes user training mandatory and more time-consuming [12]. How to develop a taxonomy in a global way is described by many authors [10], [13], [12], [14]. However, these taxonomy development methods stop at the organizational level, and are not formally described, still not allowing to engineer a complete and encompassing domain specific taxonomy.

A method coming close to industry taxonomy engineering is the method of Chosky [13]. He defines an approach for developing taxonomies within and across organizations, rather than for an industry as a whole. His approach consists of eight steps: (1) Select a taxonomy team, (2) Determine Role in Corporate Strategy, (3) Determine business purposes and requirements for taxonomy, (4) Gather and review pre-existing information, (5) Conduct survey and interviews, (6) Create inventories, (7) Rationalize classification, and (8) Finalize Taxonomy [13]. The resulting taxonomy focuses on the purpose of information within and how it is used across the organization. We are not aware of any other methods that engineer industry taxonomies.

3. RESEARCH APPROACH

This research was undertaken using the design research and case study methods. According to Hevner at al. [15], two paradigms characterize the research in the Information Systems discipline. First, behavioral science aims to clarify and/or predict organizational behavior. Second, design-science aims to extend the organizational capabilities by creating new knowledge and artifacts. This research will produce two viable artifacts in the form of the ITAXEM and the EUSOIT. This research, therefore, follows the seven guidelines of design science [15], described briefly below.

In essence, this research bridges the gap between different taxonomy engineering approaches by merging the best of all worlds and setting the standard with the presented ITAXEM (*Problem Relevance*). The utility of the proposed engineering method is tested in the case of the European software ecosystem, which resulted in an effectively described EUSOIT (*Design Evaluation*). The contributions include a novel method to engineering industry taxonomies — the ITAXEM — and the application of this method to the European software ecosystem resulting in the EUSOIT (*Research Contribution*). Hevner [15] states that design-science research must be presented effectively to all audiences. To enable researchers and practitioners to use the ITAXEM to construct industry taxonomies, it is provided in the ITAXEM section of this paper (*Research Communication*).

4. THE ITAXEM CREATION AND EVALUATION

In this section the Industry Taxonomy Engineering Method (ITAXEM) is described, which is later used to engineer the European Software Industry Taxonomy (EUSOIT). The ITAXEM is derived from the Method Association Approach (MAA) as applied in several cases [16], [17], and are used to engineer a method for an organization in a specific situation. Although this research project does not develop a situational method, four similarities are identified between an industry taxonomy and a situational method. First of all, both an industry taxonomy as well as a situational method should suit the needs of a specific situation and/or stakeholders. Second, in both cases, candidate artifacts of different sources are selected, analyzed and stored in a candidate base. Third, in both cases, these candidates are based on UML. Fourth, in both cases, these candidates are used to assemble a new artifact: an industry taxonomy or a situational method.

There are not enough suitable method fragments available in literature and existing methods that aim to develop a taxonomy. Of course, there are methods to develop taxonomies for document management, capture all species in the animal world etc., but not to capture the industry, its organizations and products. Consequently, the justification for not using the Method Association Approach, in the first place, to engineer ITAXEM, lies in the fact that first, there are no suitable method fragments available, and second, the ITAXEM derived from the Method Association Approach (MAA) provides an effective and elegant technique in order to store and analyze candidate taxonomies, and assemble new taxonomies.

As mentioned previously, Choksy [13] defines an approach for developing taxonomies for within and across organizations, rather than for an industry as a whole. The resulting taxonomy focuses on the purpose of information within and how it is used across the organization. Although his approach is not directly aimed on building a taxonomy for an industry as a whole, the techniques and underlying ideas of steps 2 to 7 are used within the method developed here, and more specifically in underlying ideas of conducting interviews. To engineer an Industry Taxonomy, seven steps are performed, shown in figure 1.

Step 1 Identify Requirements. The first step in the ITAXEM is analyzing and identifying the taxonomy's requirements and objective. The objective can be defined using the project's objective, in which the industry taxonomy will be used and/or by interviewing the project's initiators. Second, the primary stakeholders are identified, mainly using common sense and literature. Third, the information needs of those stakeholders are identified. Fourth, the information requirements are inherited from the information needs.

Step 2 Create Candidate Selection Criteria List. In this step, the criteria for selection of candidate sources, in order to fill the taxonomy base, are identified. Further on, candidate taxonomies are selected that meet one or more of these criteria

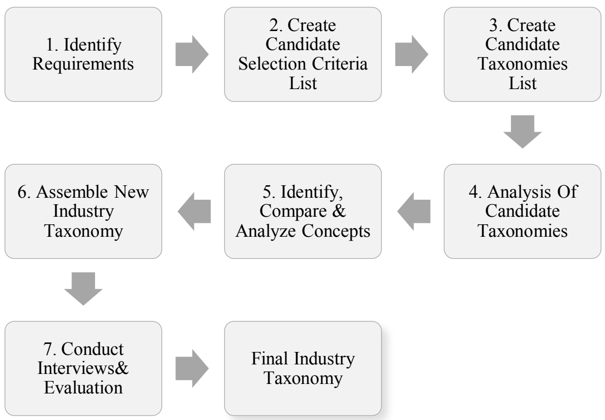

Figure 1. The ITAXEM.

Step 3 Create Candidate Taxonomies List. In this step, the selection process of candidate taxonomies, in order to fill the taxonomy base, is performed. Every collected candidate taxonomy is matched with the previously defined criteria. If a candidate taxonomy is found that matches all of the requirements, the method is completed immediately.

Step 4 Analysis of Candidate Taxonomies. In this step, candidate sources are analyzed, and taxonomies are extracted and documented using the Unified Modeling Language (UML). UML class diagrams, were selected because of the following: First, it is the basis of object-oriented analysis and design and therefore its devised for the purpose of knowledge representation[18]. Second, it is an effective and elegant technique in order to store and analyze candidate taxonomies, and assemble new taxonomies. Third, it allows to design the different kinds of relationships that may appear in taxonomies[19]. Fourth, a taxonomy can be seen as a simple ontology [6], [5]. According to Cranefield & Purvis[20], UML class diagrams can be successfully used to create taxonomies and ontologies. Ignoring constraints that enable automated reasoning is not a problem, because that is not an industry taxonomy requirement. Therefore, using UML class diagrams to show the classes of the taxonomy, the interrelationships between classes, whether or not these interrelationships are inheritance, aggregations, and associations, and to show the operations and attributes of these classes is not a problem.

Step 5 Identify, Compare and Analyze Concepts. In this step, the key concepts of all candidate taxonomies are identified. The purpose of identifying these key concepts is to enable the comparison, selection and analysis of candidate taxonomies in order to assemble a new taxonomy.

Step 6 Assemble New Industry Taxonomy. Taxonomy integration has its roots in the database community, where the process of integration is known as the problem of schema integration or view integration[21]. Inherited from the problem of schema integration is the integration of method fragments, which is called method assembly [22]. Method assembly is divided into two perspectives: (1) the product perspective and (2) the process perspective. The product perspective and candidate taxonomies have many similarities. Most importantly, both exist of classes, attributes and associations. From both the approach of Brinkkemper et al. [22] and the approach of Hakimpour [23], the formal way for candidate taxonomy integration method is inherited. Integration is done per two taxonomies. First, every class-pair within the two candidate taxonomies are compared, and checked witch of the five similarity relationships described below is applicable to them. Common sense can be used to compare two classes and to make sure they refer to the same concept or not. (1) When two classes in two candidate taxonomies are referring to the same concept, having the same name, attributes and associations, then this class is added. (2) When two classes in two candidate taxonomies are referring to the same concept, with a different name, set of attributes and set of associations, then one class is added based on conjunction of both classes. (3) When two classes in candidate taxonomies are within similar specialization, then a specialization relation is added. This specialized class will be defined by the union definition of both classes. (4) When two classes in candidate taxonomies refer to two overlapping or disjoint concepts and have the same specialization concept, they are both added. (5) When two classes refer to two disjoint concepts, they are added separately. For integrating the associations and attributes, three cases can occur. (1) When two associations or attributes in two classes refer to the same relation, then this association or attribute is added. (2) When two associations or attributes in two classes refer to two relations in a specialization similarity, then the relation can be moved to the parent class. (3) When two associations or attributes in two classes refer to two different relations, then they are added separately.

Hakimpour [23] also mentions the case where two associations/attributes in two classes refer to two overlapping relations. To avoid any information loss both attributes, they must both be added. Because attribute types, formats, and precision are not explicitly defined in any of the candidate taxonomies, and the candidate taxonomies are purely about which information is represented, instead of how this information is defined, value conversion between different candidate taxonomies is not required. In more detail, overlapping associations/attributes referring to two overlapping relations, having only different formats, or precision, can be transformed to either preferred format or precision. The main reason is that it's about creating new industry taxonomies, instead of reusing old ones including old data which can normally cause information loss.

Step 7a Conduct Expert Interviews & Evaluation. After finishing the initial industry taxonomy, the candidate taxonomies seem to fit together automatically; yet it should be made sure that the different perspectives of different stakeholders are incorporated in the final taxonomy. Therefore, in this step, semi-structured interviews are conducted to incorporate the perspective of every individual stakeholder group. A logbook is used to keep track of the changes. Moreover, using the expert interviews, one can make sure the industry taxonomy is compliant to all the criteria, requirements and needs mentioned at the start of the execution of the ITAXEM. In step 7b – Evaluation – the initial Industry Taxonomy is evaluated, resulting in the final Industry Taxonomy. In order to evaluate the Industry Taxonomy, a checklist is created, based on the predefined requirements, needs and objective. This way, one can check whether the industry taxonomy is compliant to all the criteria, requirements and needs mentioned at the start of the execution of the ITAXEM.

The result of the ITAXEM is an industry taxonomy matching the needs of the specific situation and industry. The ITAXEM provides an effective and elegant technique in order to store and analyze candidate taxonomies, and assemble new taxonomies.

5. CASE STUDY: THE EUROPEAN SOFTWARE ECOSYSTEM

In a traditionally closed market, software vendors are now facing the challenge of opening up their product interfaces, their knowledge bases, and in some cases even their source code[24]. One of the challenges identified by Jansen et al. [24] is the challenge of establishing relationships in the software ecosystem (SECO), and more specifically in the Software Supply Network (SSN). A SECO is defined as a set of businesses functioning as a unit and interacting with a shared market for software and services, together with the relationships among them. A SECO consists of three perspectives: the software vendor level, the software supply network level and the software ecosystem level. A SSN is defined as a network aiming to create competitive advantage for its participants from diverse sources for themselves and for others [24]. According to Jansen et al. [24], methods for relationship identification are required to further assist software vendors in establishing and developing their own SSNs. In addition Sharpe [25] identifies additional challenges faced by the European software ecosystem. First of all, most information available is of low quality. Second, it generally does not contain enough detail. Third, the available information is expensive and not affordable for small ISVs. Fourth, finding potential candidates for partnering with the right skill set, technologies, and mindset etc. is difficult. A way of doing this is by organizing meetings, workshops, and events. A Software Industry Taxonomy can help to overcome these challenges. Unfortunately, there is no such accepted Software Industry Taxonomy that is compliant with the way ISVs, the scientific world, and governmental organizations look at the European software industry. Therefore, to evaluate ITAXEM, it is applied to the case of the European software ecosystem. The reason for focusing on Europe is because the objective of the EUSOIT is to support the development of a new online platform for the European software ecosystem, which has as objective to provide market insight and a collaboration platform for and by European Independent Software Vendors.

EUSOIT requirements. To start, the primary stakeholders, their information needs, the information requirements, and taxonomy's needs are identified. The EUSOIT aims to provide information for three types of stakeholders. First, it should provide Independent Software Vendors with information as input for strategies, opportunities for partnering and enables both national and international benchmarking. Moreover, it allows them to get more in-depth insight in the Software ecosystem, save costs, and consequently improve business strategies. Second, It should allow governmental organizations to have more insight in the past and current status of the software ecosystem and have among others the possibility to benchmark their policies. Third, it should provide the scientific world with a new source of information and thereby many opportunities for future research. Of course, the European software ecosystem has many more types of stakeholders, such as service providers, software policy makers, trade associations, stakeholders in the public sector, chambers of commerce, consumers, etc.. Intentionally, these are left out, or are a sub group of the three groups mentioned above. As second step, the information needs are defined. The primary need of the EUSOIT is to reflect the information contained within the European software ecosystem, and more specifically in Independent Software Vendors defined as an organization that develops software for a market, instead of one-of solutions. The information provided should suit the needs of the three types of stakeholders identified earlier.

The information needs of Independent Software Vendors (ISVs) are identified mainly using the report by Garnett (2009). First of all, ISVs are looking for benchmarks for comparison with their immediate competitive set of companies. This information enables them to understand whether they are ahead or behind their competition. According to (Garnett, 2009) competitor information is the type of information (with 21.7%) that generally is hard to find. Second, ISVs are looking what opportunities are in place to develop new products, in which markets, etc.. According to Garnett, product information with 17.0% is information that either cannot be found or cannot be accessed. Third, ISVs expect to drive business objectives through effective use of technology and intimate understanding of the business needs, customer needs, and market forces and trends within their market. Moreover, market information helps ISVs to define what opportunities and risks are present. According to (Garnett, 2009) market information with 14.1% is information that most often either cannot be found or cannot be accessed. Both governmental organizations and the scientific world are looking for all kinds of information in order to do research on topic of the European Software ecosystem.

Finally, the requirements are derived from the needs identified. Five main information requirements are derived. First of all, the stakeholders are looking for general organizational information like locations, mission, and vision. Then, in order to better understand market forces and trends within the software ecosystem financial information like balance sheets, profit and loss statements and key performance indicators are required. Together, they can solve the gap in market information, which is, with 14.1%, information that most frequently either cannot be found or cannot be accessed. In addition, product Software information including product lines, descriptions, and markets is required. ISVs are looking what opportunities there are to develop new products, in which markets, etc.. Moreover, they can use this information to find complementary software products and/or allow them to find solutions that can be integrated with their own saving money and time. Last but not least, the EUSOIT should give a "complete" picture of the software ecosystem. With a "complete" industry taxonomy we imply that all the interviewed industry experts agree that this industry taxonomy covers all the information of European software ecosystem. These five main requirements are used to make sure the right candidate taxonomies are collection, and serve main input for the evaluation step. **Create Candidate Selection Criteria List.** In this step, five criteria for selection of candidate taxonomy sources are identified. (1) Most importantly, the candidate taxonomy source should be applicable to Independent Software Vendors defined as an organization that develops software in the setting Product Software. (2) The source fulfills one or more aspects of the identified requirements, the identified information needs and identified stakeholders. (3) The source is accepted by the community. This criterion refers to the influence of the source to the software ecosystem. The source should be either sufficiently described in literature or be accepted as commercial source for information. (4) The source should be applicable to one or more countries in Europe as defined previously. (5) The source should be able to describe a rich data set. In other words, it should not only be a phonebook (e.g. name, address and other basic information). Based on the selection criteria, the taxonomies are selected. Selected taxonomies comply with the set of criteria.

Create Candidate Taxonomies List. In this step, the candidate taxonomies are collected and selected. Based on the selection criteria, the selected taxonomies are: "Amadeus" [26], "ITEuropa" [27], "Business Accounting & Finance"[28], "Partner Pathway to Business Performance" [29], "Truffle 100 2008" [30], "Softsearch"[1], "Capterra"[1], "The Software Network" [1]. All of these taxonomies comply with the set of criteria and besides that, cover parts of the identified information requirements. The selected candidate taxonomies are briefly described in the next section, including its background, its content, references and why it meets the criteria.

Analysis of Candidate Taxonomies. In this step, the candidate sources are analyzed, and taxonomies are extracted, the selected taxonomies are described. More in detail, the analyze process, key concepts, trade-offs, and analyses problems and corresponding solutions per candidate taxonomy are described. Only the "Amadeus taxonomy" is described here in full detail.

The Amadeus database offers standardized organizational profiles based on consolidated and unconsolidated information, including financial profiles, activities and ownership on approximately 11 million companies throughout. The main focus of Amadeus is on the financial profile, containing 24 balance sheet items, 25 profit and loss account items. Moreover, they defined 26 standard ratios as most important to measure organizational performance. In addition, they define the status of an organization using eleven indicators, and the legal form of an organization using three legal form indicators. Finally, they also define the board members and auditors.

Identify, Compare and Analyze Concepts. In this step, the key concepts of all candidate taxonomies are identified. The key concepts identified are shown in table 1. A cell containing a "●" indicates that the concept is present in candidate taxonomy. An empty cell indicates that the concept is not present in the candidate taxonomy.

Assemble EUSOIT. The candidate taxonomy integration activity describes a simple and elegant process in order to integrate several candidate taxonomies. Integration is done per two candidate taxonomies. First of all, before starting, one has to make sure to apply each of the five rules, as described in the previous section, during every moment of the integration process. Then, every class-pair within the two candidate taxonomies is analyzed and compared, and checked whether any of the five similarity relationships.

Describing the complete process of integrating each candidate taxonomy pair would be too overwhelming. Therefore, one example is chosen in order to clarify and show the process in practice. In this example, we integrate the candidate taxonomy of the Truffle Top 100, and the candidate taxonomy of ITEuropa. Both the Truffle Top 100 and the ITEuropa candidate taxonomy do contain an object representing the concept of a sales

● = Concept is present in candidate taxonomy Empty = Concept is not present in candidate taxonomy	Amadeus	ITEuropa	Business Accounting & Fin.	Partner Pathway	Truffle 100 2008	Softsearch	Capterra	The Software Network
Financial Profile	●		●					
Assets	●		●					
Liabilities	●		●					
Balance Sheet	●		●					
Financial KPIs & Ratios	●			●	●			
Profit & Loss Statement	●		●					
Legal Form, Ticker Etc.	●							
Addresses, Locations, Etc.	●							
Contact Details		●						
Divisions	●							
Board Members	●							
Employees	●							
Auditors	●							
Partnerships				●				
General Product Inf.		●						
Technical Product Inf.		●						
Hardware Platforms		●						
Software Platforms		●						
Product Lines		●						
Vertical Industries		●			●	●	●	●
Horizontal Applications		●			●	●	●	●
Org. Profile		●			●			
Staff Breakdowns		●			●			
Sales Breakdowns		●			●			

Table. 1. The identified key concepts per candidate

breakdown. It is called the "revenue software activities breakdown" in truffle top 100, and "sales from company activity breakdown in %". The truffle top 100 calls this concept the "revenue software activities breakdown", which consist of 7 attributes: license, maintenances, related services, open source support, software as a service and asp, hardware and software resale, and other revenues. On the other hand we have ITEuropa that calls this concept the "service sales breakdown in %", which consist of the 6 attributes consulting, implementation, support and helpdesk, outsourcing and managed services, internet services training, and other activities.

According to the process described in step 6 of the ITAXEM, we first have to compare the class-pair. Although we have to check other class-pairs as well, common sense says that we should not compare for example employees and hardware platforms or what so ever. Next, we check witch of the five similarity relationships is applicable to them. The two classes in the two candidate taxonomies are referring to the same concept, but are not having the same name, attributes and associations. Therefore, we cannot just add one of the two classes (rule 1). According to rule two, they have a different name, set of attributes and set of associations, and so then one class is added based on conjunction of both classes (rule 2). It has nothing to with a specialization relations (rule 3 and 4), and they aren't disjoint either (rule 5). Concluding from this first phase, we have to add one class based on conjunction of both classes.

To cover the semantic heterogeneity, we have to apply the three rules as defined in step 6. We have to do this for each attribute pair. Some are easy to mix and match. In example the attribute

[1] Softsearch, Capterra and the Software Network are databases (http://www.softsearch.com; http://www.capterra.com; http://www.thesoftwarenetwork.com)

"internet services" and "Software as a Service and ASP" are both referring to the same relation, but have a different name. Therefore, this relation is added under the name "Software as a Service and ASP". This can be done for each attribute. Finally, it results in Software Activity Revenue Breakdown, consisting of 7 attributes: Licenses, Maintenance, Professional Services, Support, Software as a Service and ASP, Hardware and Software Resale, Other Revenues.

Conduct Expert Interviews & Evaluation. Empirical data is collected through a total of 9 semi-structured interviews resulting in qualitative data and finally in an in-depth perspective on the software ecosystem. The outlines for the interview consists of four main areas: (1) What is the interviewees information strategy, and do they need, (2) How do they finance their information strategy, (3) what information are they currently using, (4) what information would be at hands in the ideal world.

By conducting the interviews as described, it is made sure that the different perspectives of all stakeholders are incorporated in the final taxonomy. In the beginning, an audio recording and a recording on paper were kept to make sure everything is well documented (see appendix H). In addition, a logbook was used to keep track of the changes made to the industry taxonomy.

In general, all interviewees approved that the initial EUSOIT already gives a complete view of the European software ecosystem. Nevertheless, most of the interviewees suggested modifications to one or more parts of the industry taxonomy, in order to improve it. Not all of the modifications are described in this section. In order to give a brief view of how the process of merging the interview results worked, the results of interviewee 4 described briefly below.

The first and most important modification the interviewee noted is that the EUSOIT as a whole was completely focused on the ISV side, while disregarding the market side. Therefore, he introduced a concept called Market Profile, containing four attributes: minimal target company size, a maximum target company size and a target market description. The main reason why the interviewee insisted on this to be added is that it is important for a company to know what the product's target market is. In the end, this modification made sense because the whole EUSOIT was completely focused on the ISV side, not on the market side, and were added.

Evaluation of EUSOIT. In this final activity the draft EUSOIT is evaluated. After evaluation, it is upgraded to final version of the industry taxonomy. The reason for evaluation is to make sure the industry taxonomy is complete. With a "complete" industry taxonomy we imply that all the interviewed industry experts agree that this industry taxonomy covers the whole European software ecosystem. In order to evaluate the EUSOIT, a checklist is created, matching all the criteria, requirements and needs mentioned at the start of the execution of the ITAXEM. The checklist consists of the following five checks; (1) Is the EUSOIT clear and understandable? (2) Does it give a complete view of the European Software ecosystem? With a "complete" industry taxonomy we imply that the taxonomy covers the whole industry. (3) Does it offer the information needed by the three major stakeholders; the ISVs, the governmental organizations, and the scientific world? (4) Does it offer a good foundation for the new online platform for ISVs? (5) Does it provide all the information to give in-depth market insight into the European software ecosystem?

All interviewees approved that the EUSOIT gives a complete view of the European software ecosystem. The information needed by the three major stakeholders; the ISVs, the governmental organizations, and the scientific world, defined in the beginning of the execution, and gathered during the interviews is also incorporated. Therefore the EUSOIT offers all the information needed by the three stakeholders. Last but not least, it offers a good foundation for the new online platform, and matches the objective of this platform, because it showed already its usefulness in practice. Finally we can state that each item on the checklist is checked, implying that the EUSOIT is compliant to all the criteria, requirements and needs mentioned at the start of the execution of the ITAXEM.

6. DISCUSSION

The research presented makes a contribution in the area of industry taxonomy engineering, with the formalization of a method for engineering industry taxonomies. In addition, this research applies this method by engineering the European Software Industry Taxonomy. The output of an industry taxonomy engineering project, is a structure that enables the classification of business entity related information, within one industry, according to internal criteria, properties and relationships, capturing among others, the functions performed, and the products and services supplied. At the stakeholder's level, the industry taxonomy accommodates the viewpoints of multiple stakeholders, which include the industry's employees, customers, and governmental organizations. In addition, the more consistent with reality this 'picture' of the industry is, the easier the end-users are able to see it as a single object of analysis; as a subject with its own autonomy, i.e. with a behavior and with reactions. If the end-users see it this way, and it is accepted, it will also largely influence the way information is displayed to them, in for example web initiatives.

In the past, no formal approaches to industry taxonomy engineering existed. Therefore, every approach, in this area, is more than welcome. Due to the early stage of this research, the ITAXEM has its weaknesses and limitations. Nevertheless, it can be used, to engineer an industry taxonomy. It allows every industry professional to jump on the industry taxonomy engineering bandwagon, and in turn allows them to create an industry taxonomy. Building an Industry-wide taxonomy can reap the additional benefit of clarifying the industry's organization, both internally, and how organizations fit within their supply chain, how they connect to other organizations, and how one defines the relationships in between. The challenges, weaknesses, and limitations as noticed by the researches during the execution of the project, are described briefly below.

Quality of Sources. During the step in which candidate taxonomies are collected and selected, a limited number of candidate taxonomy were collected. We were sure that there were many candidate taxonomy available. However, it was hard to define the quality of these candidate taxonomies. Moreover, it was hard to find candidate taxonomies besides the ones already mentioned by the project's initiators. Finally we reduced the number of candidate taxonomies into a small set of relevant candidates, and meeting high enough quality standards, by selecting those candidates that meet the criteria.

Stakeholder Identification. One aspect to keep in mind is to understand the goals and target audiences of an industry taxonomy. Every industry and industry taxonomy initiative has different types of organization, relationships, and stakeholders in different user communities with different priorities, not to mention the way of funding which largely influence the initiative itself. The ITAXEM does not explicitly state the way stakeholders need to be identified. Moreover, there is a high chance that the initiative itself has its own reasons to create the taxonomy, and thereby itself already defined its stakeholders. In the case of the EUSOIT, the reasons were to create an industry taxonomy only for three groups of stakeholders: Independent Software Vendors, Governmental Organizations and the Scientific World. However, this does not imply that other stakeholders, like system integrators and analysts, are not addressed. These stakeholders, and others, fit in one of the three groups, which allows us to make sure their needs are addressed as well. The reason for not explicitly state and interview these groups, is due to time constraint as well as they fit in one of the groups mentioned above. The project team must consider the organizational goals, audiences, intended usage, and desired level of access when determining the stakeholders, their information requirements and needs. One must keep in mind, that when using an industry taxonomy, it is created primary for the intended stakeholders, and not necessary accommodates viewpoints of others.

The Project Team. The ITAXEM does not explicitly describe who should execute the ITAXEM. Foremost, the team responsible for engineering the industry taxonomy should include industry experts, having knowledge of the technical domain, as well as knowledge of business processes. Corcoran [31] describes four important steps to ensure a successful team for taxonomy engineering. She states that, before one begins, preparations should be made. Before even real taxonomy engineering is started, first, (1) the objectives need to be defined, (2) a (financial) sponsor needs to be found, (3) a taxonomy project leader should be assigned, (4) a team needs to be recruited, that matches the objectives, and has enough knowledge of the technical domain, as well as knowledge of business processes, to fulfill the objectives.

Industry Taxonomy Governance. As Corcoran [31] writes, a taxonomy is never finished, and an industry taxonomy should never be perceived as complete. It is a dynamic artifect that evolves over time. She states as well that building the perfect, exhaustive taxonomy is not only futile but counterproductive to the business. This implies the need for Industry Taxonomy Governance. Engineering an industry taxonomy does not stop at deployment. One should manage the whole lifecycle of the industry taxonomy." Is the taxonomy achieving its objectives?", "Does it provide enough detail for the end-users?", "Can it be adjusted for the situation next years?" - All questions which should be able to be answered by the taxonomy owner. An industry taxonomy-engineering project should be focused on what end-users care about most on that specific moment. It just should be the start of a new life of the industry taxonomy, because an objective of building perfect industry taxonomy contains all concepts available, and meeting all end-user's expectations, will never be achieved. Coming back to the ITAXEM, one should not expect an exhaustive industry taxonomy as result. Rather, one should expect to kick-start the lifecycle of such an artifact, and given the right tools, allowing the end-users to really make the industry taxonomy as complete as possible.

Braun, Schmidt & Walter [14] wrote about a comparable phenomenon, what they called the ontology maturing process. They state that most of the current methodologies for building ontologies rely on specialized knowledge engineers, which counts as well for the ITAXEM, because an industry taxonomy can be seen as a simple ontology. One of its limitations is that it is executed by a team of industry experts, which only participate just in the beginning of its life-cycle. Although it tries to accommodate the viewpoints of all stakeholders, which include industries employees, customers, and governments, the ITAXEM does not describe on how to act on the changing viewpoints of future stakeholders, and how to improve and maintain the industry taxonomy during its life-cycle.

In the real-world setting, the needs for maintenance of domain specific taxonomies emerge in the daily work of end-users. Therefore, Braun, Schmidt & Walter [14] introduces the ontology maturing processes which triggers end-users to engage in ontology engineering process, and merge it within their everyday work processes. They also argue that these kinds of artifacts cannot be formalized from scratch, but rather continuously evolve in a maturing process from an in- formal cloud of ideas to a formal taxonomy. Therefore, the ontology maturing process was presented by Braun, Schmidt & Walter [14], which consists of four steps. First of all, the emergence of ideas should allow end-users to introduce not well-defined concepts to the taxonomy. Then, in the consolidation in communities phase, more formal concepts can be created from these tags. Third, in the formalization step, the newly developed concepts are organized into the relationships. Finally, the axiomatization phase takes place, which allows the end-users and experts to add logical formalism, to capture the more in-depth semantics. The last step for ITAXEM is not needed, in the first place, because we only need a basic knowledge representation. The other three steps could be a useful add-on to the Industry Taxonomy Governance process. Such an improvement process overcomes two main problems: First, it eliminates the problem of the time lag between emergence of ideas and their inclusion in an industry taxonomy by experts, and second, it allows better communication between end-users and the owners of the taxonomy. By applying Industry Taxonomy Governance, an industry taxonomy is never perceived as complete. On the contrary, it evolves over time, allowing to meet the viewpoints and needs of future stakeholders.

Industry Classifications. Another part not yet covered by the ITAXEM is a formal sub-method that enables the development of industry classifications, that are part of an industry taxonomy. In the case of the EUSOIT, two examples of industry classification are the horizontal applications consisting among others of Accounting and Taxes, Business Intelligence / Data Warehousing, Business Planning / Continuity, Calendar/Scheduling etc. and vertical industries consisting among others of Agriculture, Apparel & Fashion, Automotive, and Banking & Financial etc.. Industry classifications like these two select the essential characteristics of applications, technologies, industries and markets and divides those characteristics into a smaller number of salient, preferable mutually boxes. We suggest adding one method into the ITAXEM, allowing the project team to develop industry classifications in a structured way.

7. CONCLUSION

Presently, no methods exist that help engineering an industry taxonomy within a specific domain. Without such a method, taxonomies remain erroneous, making the development of, for instance, a directory of companies for research, extremely hard. This paper presents a method for creating complete and encompassing domain specific taxonomies. First of all, the method presented offers a structured manner in order to engineer an industry taxonomy within a specific domain. The method is efficient because the project team borrows what already exists, and builds what does not yet exist. Moreover, it is effective because one makes sure the industry taxonomy accommodates the viewpoints of its stakeholders. The method is evaluated in practice, and the engineered industry taxonomy is evaluated by expert interviews. Consequently, the ITAXEM is applicable to any industry, from the Agriculture industry and Apparel & Fashion industry to the Travel industry.

8. REFERENCES

1 Cheng, Chin, Lau, Gloria, Law, Kincho, Pan, Jiayi, and Jones, Albert. Regulation retrieval using industry specific taxonomies. *Artificial Intelligence and Law* (2008), 277-303.

2 Trippe, Bill. Can XML Drive Taxonomies and Categorization?, p30-34.

3 Brunnermeier SB, Martin SA. Interoperability costs in the US automotive supply chain (2002).

4 Gallaher MP, O'Connor AC, Dettbarn JL et al. Cost analysis of inadequate inoperability in the capital facilities industry (2004).

5 Rees, R van. CIB73 2003. In *Clarity in the usage of the terms ontology, taxonomy and classification* (2003), 1-8.

6 Smith, M.K., Welty, C., and McGuinness, D.L. *OWL Web Ontology Language Guide*. W3C, 2004.

7 Garshol, Lars Marius. Metadata? Thesauri? Taxonomies? Topic Maps! *InterChange* (2004), 17-30.

8 *Merriam-Webster Dictionary - Classification*. Merriam-Webster, Incorporated, 2010.

9 *Merriam-Webster Dictionary - Taxonomy*. Merriam-Webster, Incorporated, 2010.

10 Bruno, D and Richmond, H. The Truth About Taxonomies. *Information Management Journal*, 37, 2 (2003), 44.

11 Guibert, Bernard, Laganier, Jean, and Volle, Michel. An essay on industrial classifications. *Economie et statistique* (February 1971).

12 Cisco, S L and Jackson, W K. Creating order out of chaos with taxonomies. *The Information Management Journal* (2005), 45-50.

13 Choksy, C E B. 8 Steps to Develop a Taxonomy. *The Information Management Journal* (2006), 31-41.

14 Braun, Simone, Schmidt, Andreas, and Walter, Andreas. Ontology Maturing: a Collaborative Web 2.0 Approach to Ontology Engineering. In *Proceedings of the Workshop on Collaborative Construction of Structured Knowledge (CKC) at the 16th International World Wide Web Conference (WWW 07)* (Banff, Canada 2007), WWW2007.

15 Hevner, A. R., March, S. T., and Park, J. Design Science in Information Systems Research. *MIS Quarterly* (2004), 79-99.

16 Weerd, Inge Van De and Brinkkemper, Sjaak. Meta-Modeling for Situational Analysis and Design Methods. In *Handbook of Research on Modern Systems Analysis and Design Technologies and Applications*. Minnesota State University, Mankato, 2008.

17 Luinenburg, Lutzen. *Conceptual design modeling of CMS-based web applications*. Utrecht, 2007.

18 Kiko, Kilian and Atkinson, Colin. A Detailed Comparison of UML and OWL. *Technical Report / Department for Mathematics and Computer Science* (2005), 2-58.

19 Bedford, Denise. Presentation to Canadian Metadata Forum. In *Canadian Metadata Forum* (2005), Library and Archives of Canada.

20 Cranefield, Stephen and Purvis, Martin. UML as an Ontology Modelling Language. In *In Proceedings of the Workshop on Intelligent Information Integration, 16th International Joint Conference on Artificial Intelligence*. IJCAI, Stockholm, 1999.

21 Batini, C, Lenzerini, M, and Navathe, S. B. A comparative analysis of methodologies for database schema integration. *ACM Computing Surveys (CSUR)* (December 1986), 323-364.

22 Brinkkemper, Sjaak, Motoshi, Saeki, and Harmsen, Frank. Meta-Modelling Based Assambly Techniques for Situational Method Engineering. *Information Systems* (1999), 209-228.

23 Hakimpour, F. and Geppert, A. Resolving Semantic Heterogeneity in Schema Integration: an Ontology Based Approach. *FOIS '01* (2001), 297-308.

24 Jansen, S., Finkelstein, A., Brinkkemper, S. A Sense of Community: A Research Agenda for Software Ecosystems. In *International Conference on Software Engineering (ICSE 2009)* (Vancouver, Canada 2009).

25 Sharpe, Mike. *Software 2.0: Rebooting Europe's Software Industry*. 2009.

26 BUREAU VAN DIJK ELECTRONIC PUBLISHING. *Amadeus*. 2009.

27 *ITEuropa*. 2008.

28 Davies, Tony and Boczko, Tony. *Business Accounting and Finance*. McGraw-Hill, New York, 2005.

29 Lawton, Matthew and Graham, Stephen. Microsoft Competencies: Partner Pathway to Business Performance. *IDC* (2006), 1-7.

30 TRUFFLE 100. *Truffle 100 2008*. Paris, 2008.

31 Corcoran, Mary. Industry insights. *Online*, 26, 5 (2002), 76.

32 Yin, Robert K. Criteria for judging the quality of research designs. In *Case Study Research - Design and Methods*. Sage Publications Inc., London, 2003.

33 *Merriam-Webster Dictionary - Ontology*. Merriam-Webster, Incorporated, 2010.