Knowledge Sharing Using Ontology Graph-Based: Application in PLM and Bio-Imaging Contexts

Cong Cuong Pham, Alexandre Durupt, Nada Matta, Benoit Eynard

Knowledge sharing using ontology graph-based: application in PLM and Bio-Imaging contexts

Cong Cuong PHAM[1], Alexandre DURUPT[1], Nada MATTA[2], Benoit EYNARD[1]

[1] Sorbonne University, Université de technologie de Compiègne
Department of Mechanical Systems Engineering, UMR CNRS 7337 Roberval,
CS 60319, 60203 Compiegne Cedex, France
phamcong@utc.fr; alexandre.durupt@utc.fr; benoit.eynard@utc.fr

[2]Tech-CICO Lab, Troyes University of Technology, France
nada.matta@utt.fr; guillaume.ducellier@utt.fr

Abstract. Data resources in PLM (Product Lifecycle Management) systems are becoming more and more huge and complex. The heterogeneity of data type and the dependencies among technical information make difficult for users in database exploitation: to query and to share the data. In this paper, we present an ontology-based approach as a promising solution to overcome this issue. An ontology graph-based query interface has been developed with the aim to enhance the knowledge sharing among different types of users (non-technical or coming from diverse expert domains) and then to facilitate the database exploitation. An example in Bio-Imaging domain will be presented as an application field.

Keywords: Product Lifecycle Management (PLM), Bio-Imaging, Knowledge sharing, Ontology graph.

1   Introduction

Product Lifecycle Management (PLM) is a combination of solutions and techniques which enable the efficient management of information through various stages of product lifecycle. These solutions have also tackled the heterogeneity and complexity of data and the challenges in tracking the evolution and the modification of information. Nowadays, with the support of ICT (Information and Communication Technology), PLM databases are becoming more and more complex: the amount of data, the diversity of data types, and especially the dependencies among technical information [1]. Furthermore, new data are always generated and added into database by users of PLM system during their quotidian activities. Normally, these data are related to an individual and created for a concrete purpose. Therefore, they cause the difficulties in data management and knowledge sharing because of their heterogeneity and personality. As consequence, in the context of complex, heterogeneous, and intertwined data resource, a major requirement for an efficient PLM system is to provide users the ability to query data from database and then to share them in community. In recent years, ontology has been widely used in the scientific

community as a promising solution for knowledge sharing. By definition, ontology allows expressing a conceptualization not only in natural language but also in a format that can be interpreted and used by software agents. Therefore, it enables the sharing and the reuse of knowledge. Our aim is to develop an ontology-based knowledge sharing platform, where the understanding of changes and evolutions in the dependencies and linkages among data will be assimilated to all users (non-technical, coming from different areas…). The main objective of this platform is to enhance the data exploitation: in data querying, in data visualization, in technical information sharing and furthermore, in data mining. This article presents our first results on an ontology graph-based query interface which allows performing queries. A case of study will be illustrated in Bio-Imaging domain in which researchers need to have a good understanding of data model as well as the dependencies among data in order to interrogate the database. The rest of paper is organized as follows: In section 2, we present the literature in PLM systems as well as some techniques in the knowledge sharing. Next, in section 3, we propose an approach for ontological model construction. Section 4 describes an ontology graph-based query interface as an application in the context of Bio-Imaging. Finally, section 5 is reserved for discussion and conclusion.

2 Related work

In this section, we firstly present some literatures about PLM systems in the context of heterogeneous and dependent data resources, then some techniques in knowledge sharing. Based on this work, ontology was chosen as the solution in our approach. The methodology and demonstration will be presented in the next sections.

2.1 Product lifecycle management (PLM)

The concept Product Lifecycle Management (PLM) appeared some decades ago. This acronym has been used in different communities such as data management software vendors, academic community, end users… with slightly different interpretations [4]. It is defined as a product centric-lifecycle-oriented business model, supported by Information and Communication Technologies (ICT), in which product data shared among actors, processes and organizations in the different phases of the product and related services [2]. PLM systems manage the increasing of the volume of generated and processed data and information during product lifecycle as well as the traceability and confidentiality issues [3].

Originating in the car industry, PLM has now been widely applied in various domains including pharmaceutical industry or recently in Bio-Imaging domain [4]. [4] adopted PLM concepts to handle the complexity, heterogeneity and characteristics of Bio-Medical Imaging (BMI) data. However, traditional PLM systems are not flexible as requires by research practices, a requirement of actual works is to enable non-technical users (Bio-Imaging scientists) to query data from the database. In fact, to query database, users need to understand the data model and the dependencies among data in the database. Furthermore, the complexity of the dependencies increases

gradually because of new added data. These new data are usually related to a predefined context and are the work results of a group of individuals, therefore, the others don't understand the nature of these data as well as relationships with existing ones. As consequence, it is required to assimilate the understanding about data dependencies to all users of system for the purpose of database exploitation. Knowledge sharing is therefore studied. We next present some literatures in this aspect.

2.2 Knowledge sharing

Knowledge is defined as information possessed in the mind of individuals which may or may not be new, unique, useful, or accurate related to facts, procedures, concepts, interpretations, ideas, observations, and judgements [6]. There are two forms of knowledge: "tacit" and "explicit". The former exists in the mind and therefore belongs to an individual. The latter exists in the form of words, sentences, documents and other explicit forms. Therefore, explicit knowledge can be better communicated and shared than tacit one.

Knowledge sharing is one of the most important processes in Knowledge Management. It can be defined as activities of transferring or disseminating knowledge from one person, group organization to another. Information Technology (IT) provides techniques to capture knowledge, search, extract content information and present it in a more meaningful form across multiple contexts of use. Some authors [5] [6] have devoted their efforts to construct platforms that enable knowledge sharing by using ITs. [5] used XML Linking Language (XLink) as a method of describing the knowledge and proposed an architecture for sharing this knowledge among users based on peer-to-peer technique. [6] tried to re-define knowledge resources in the network by object-oriented thinking and proposed three-layer knowledge sharing model.

In recent years, the Semantic Web [7] whose ontology is a key component has been used widely as an efficient solution for knowledge sharing systems. Ontology is an explicit, formal specification of a shared conceptualization, it therefore enables the knowledge sharing and knowledge reuse. Next part of section presents some works related to knowledge sharing in PLM based on ontology.

2.3 Ontology-based knowledge sharing in PLM

Ontology has been used in to share knowledge in various domains [8], [9]. In the domain of PLM, many authors have also used it as a solution to tackle the issues in technical information interoperability, knowledge sharing and knowledge reuse. A knowledge layer has been added to commercial PLM systems to solve semantic interoperability problem of heterogeneous data and to fully utilize all available information [10]. In that approach, ontology has been used as a common language across several domains and information sources in manufacturing industries. An ontology model has been built to explicitly define relationships among products, processes and resources, and make this information accessible through a web services.

In the same way, MEMORAe [11] has been integrated in PLM system in order to enable the knowledge sharing [12]. MEMORAe allowed users to construct a shared understanding (tacit and explicit knowledge) through the use of ontologies. According to [12], "under certain conditions, a piece of information shared within a PLM leads to one and only on interpretation, so that under certain conditions, sharing information within PLM systems is sufficient to share explicit knowledge".

[13] introduced an approach based on semantic relationship management to enhance the knowledge management and reuse in collaborative product development. Fig. 2 presents the conceptual model of Relationship Manager in which **Entity** (E) is the key object and it represents any type of product data used in Begin of Life (BOL) phase. *ExpertEntity (EE)* and *RelationshipEntity (RE)* are generated from **Entity**. *EE* represents **Resource**: the metadata, documents stored in CAx application, it is identified by its Uniform Resource Identifier (**URI**). *RE* represents any entity used to link to other Entities.

From this conceptual model, **Entity** is defined as the main class of ontology and it is divided into three subclasses: *RequirementEngEntity*, *MechanicalDesignEntity* and *SimulationEntity* (Fig. 3). The **Entity** class defines two basic semantic relationships: *hasURI* and *hasResource*, respectively to **URI** and **Resource** concepts. According to this, every instance of **Entity** and subclass of **Entity** are characterized by an **URI** scheme and associated with one or more **Resource**(s).
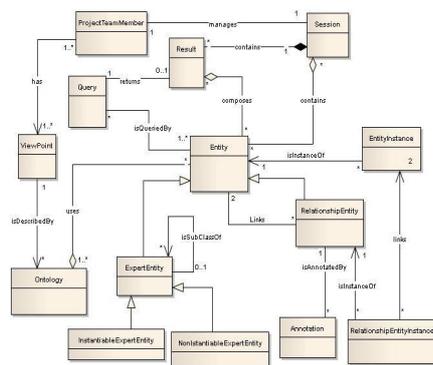
Fig. 2. The extended conceptual model of Relationship Manager
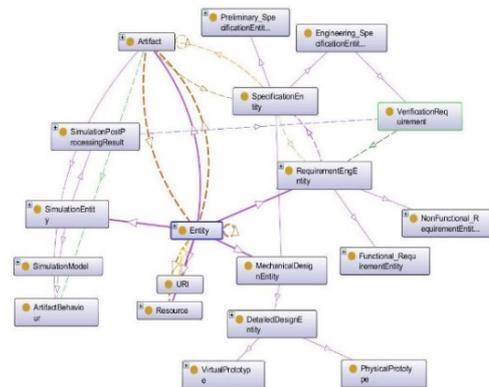
Fig. 3. Main entities of the proposed product BOL ontology

This ontology enables the capturing and sharing of any product knowledge generated by user. The users can also reuse the available knowledge in order to perform their design tasks.

In the next section, we present our work on the construction of an ontology in the context of Bio-Imaging. This construction process is based on the approach of [13] with a slight difference.

3. Methodology

We initiated by interviewing the scientists at GIN (**N**eurofunctional **I**maging **G**roup) – a laboratory in domain of Bio-Imaging, where the growth and heterogeneity of data resources have been handled by using PLM solutions in Teamcenter (Siemens). During the interview, the scientists have been provided a set of questions related to their quotidian activities. The purpose of these questions were to identify the difficulties of users in manipulating with the information system as well as their need and requirements for the new PLM platform. The interview showed that each researcher has his own individual studies which requires different dataset. Therefore, it is important to enable users to query database themselves.

Furthermore, scientists generate data dynamically by themselves and they want to store them in database. However, this task demands a deep understanding of complex dependencies among data in database, and concepts in the data model. To assimilate this understanding, we decided to use ontology as solution because of its formal expression as well as extensibility and customizability capacity. Based on the approach presented in [13], the construction of ontological model initiated by the data model analysis.

3.1 Data model analysis

Fig. 4 presents the BMI-LM (Bio-Medical Imaging Lifecycle Management) data model used in Teamcenter 9.1 [4]. By adopting PLM solutions in the context of BioImaging, this PLM-oriented data model covered the whole stages of a BMI study from specifications to publications and enabled the flexibility in data management. It contains three types of objects: *Result* objects (Exam, Acquisition, Data Unit, Processing), *Definition* objects (Exam, Acquisition, Data Unit, Processing) and *Reference* objects (Bibliographical, Data). "*Definition*" concepts defined the methods and processes to obtain results, so they have been used for the purpose of data reuse. For example, all the *Processing results* computed by using the same *Acquisition device* and *Processing parameter* can be attached to the same corresponding *Processing definition*.
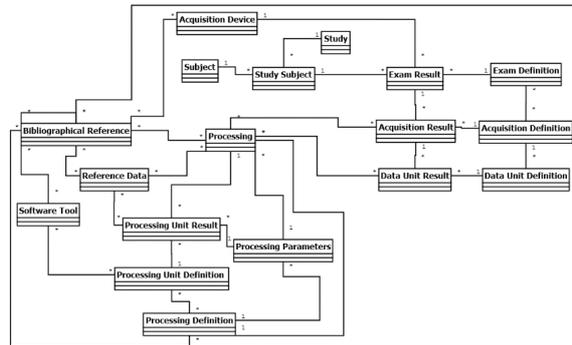

Fig. 4. BMI-LM data model implemented in Teamcenter 9.1[4]

The classification (Fig. 5) has been built based on the data model. BMI data have been classified into branch, classes and subclasses. The classification allows a specific class to be added to a generic item (object in the data model). In comparison with the data model, the classification and its attributes are easier to modify for user than objects attributes, it therefore adapts the flexibility requirement of database.

However, the low-level expression of UML schema and the complex relations among classes in the classification bring difficulties for users in querying the database. To overcome this issue, we build an ontology which bases on both of data model and classification. This ontology provides firstly an overview of concepts in the data model and the relationship among them but now represented in a nature language, and therefore it allows end-users to create a query.
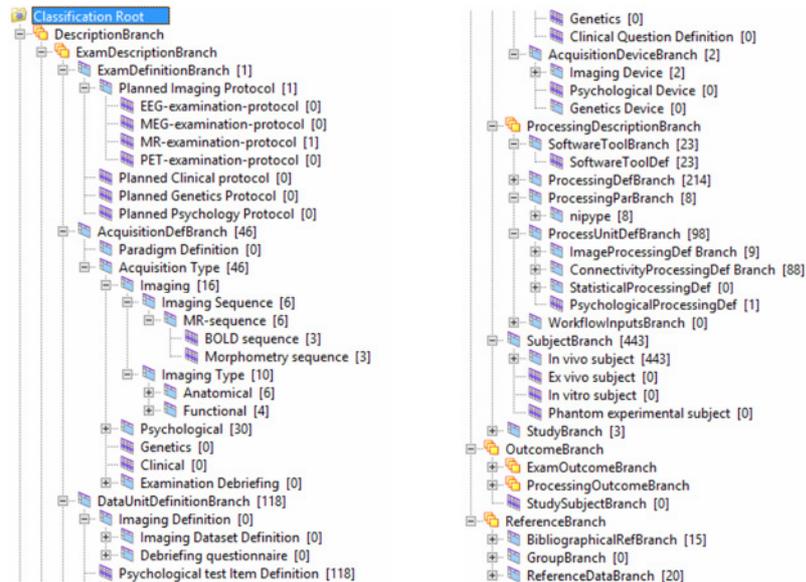

Fig. 5. Classification in TC corresponding with the BMI-LM data model

3.2 Ontological model construction:

The ontology concepts have been identified form data model objects and classes in the classification. We built a tree to represent the hierarchy of these concepts and a graph to illustrate the relationships among them. Ontology concepts are categorized into four major categories: **Tools**, **Data**, **Process** and **Investigator** (Fig. 6) corresponding with acquisition/ processing tools, personal information, acquisition/ processing results and acquisition/ processing definition, bibliographical references. We believe that this hierarchy provides an understandable categorization for users beside of the existing classification and specially when we notify that this ontology is identified from interviews with scientists and it respects their work logic.
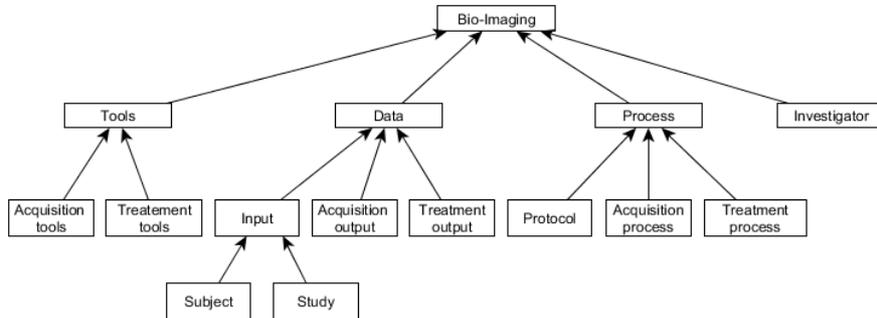
Figure 6: Ontology tree conceptual in Bio-Imaging domain

Then, in the graph of ontology (Fig. 7), we added the relationships among ontology concepts that have been expressed by nature language in order to make them more understandable. For example, An *Acquisition Definition* **generates** some *Acquisition outputs* by **following** some *Protocols*.
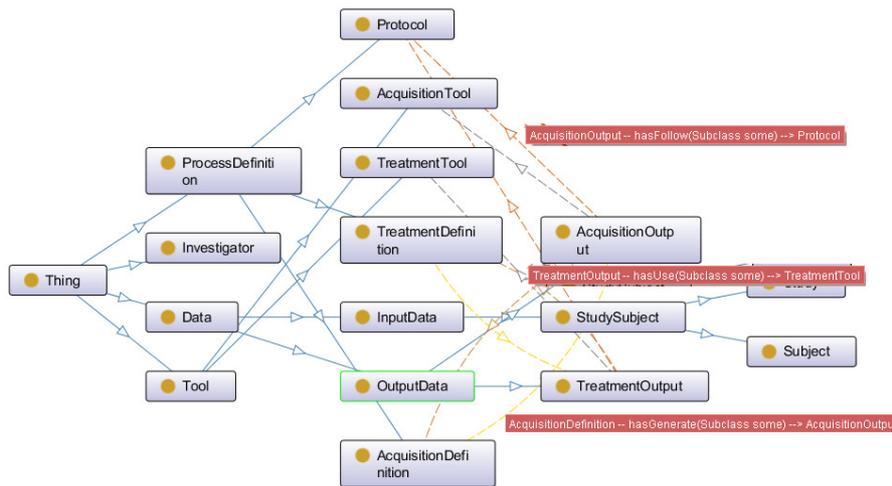

Figure 7: Conceptual graph of ontology

Ontology graph can be developed in more detail by expanding each concept (node) into it sub-concepts (sub-nodes). Sub-nodes inherit all attributes and have the same relationships with their parent. For example: **Tools** class is divided into *Acquisition tools* and *Treatment tools* (as illustrated in ontology tree), therefore, they have the same relationship "isUsedBy" with **Data** and their parent class **Tools**.

In the next section, we deal with an application in Bio-Imaging domain where this ontology has been used to help users in making a query to the database.

4. Application

We developed a query interface based on the data model, the classification and the ontology. Our aim is to provide to users a multiple layer view, from conceptual level (ontology graph) to low-data-level (data model, classification) in order to help them to query database.

4.1 Ontology-based graph query interface

The ontology graph and the data model are represented in a graph while the classification is presented in a tree (Fig. 8). Here we take an example of query frequently used by scientists at GIN:

*"Querying all subjects (*StudySub*) who **have** certain characteristics (*sex = "man"*, age <= "45") and have **passed** an Acquisition (*AcquisitionRes*) (name = "AcqName", date <= "01.12.2014") which **suffers** a Treatment (*ProcessRes*) (name = "ProcessName", description contains "Description")".*

In this example, we want to query all *Subjects* related to some *Acquisitions* and *Treatments*. Firstly, user chooses three concepts from the ontology graph. These concepts are linked to the corresponding objects in the graph of data model (a representation of data model UML schema but in a graph form): StudySub, AcquisitionRes, ProcessRes. User then defines the value of each selected object's attribute by using the tree of classification. Fig. 9 illustrates the process of objects selection from the query object StudySub to ProcessRes. Finally, a query is generated from conditions defined by user.

4.2. Query making and query execution process

The query generated from the query interface will then be sent to and executed in server. By now, we use TCXquery [14] as a Query Processor that makes PLM content in the database as usable as XML document. Therefore, the query defined by users is transformed into XQuery language. An extract of query in xQuery language is cited as bellow. The xPath (through all objects) is generated automatically by using a graph pathfinding algorithm.

```
return
Teamcenter:Query($teamcenter,$query)[./F_IMAN_master_fo
rm/GIN4_StudySubMaster/@gin4_gender = "M" and ./F_IMAN_
master_form/GIN4_StudySubMaster/@gin4_datebirth >=
"1970"]/F_GIN4_rel_ExamResIMA/GIN4_ExamRes/F_GIN4_rel_A
cquisition/GIN4_Acquisition[./@name = "AcqName" and
./@examDate <= 01.12.2014]/F_GIN4_rel_ProcessingRes/
GIN4_ProcessRes[./@name = "ProcessName" and constain(.
/@description,"Description")]
```

The results of query will be sent back and visualized in the form of graph (Fig. 9). Users can edit, do some analysis or save these results for further researches.
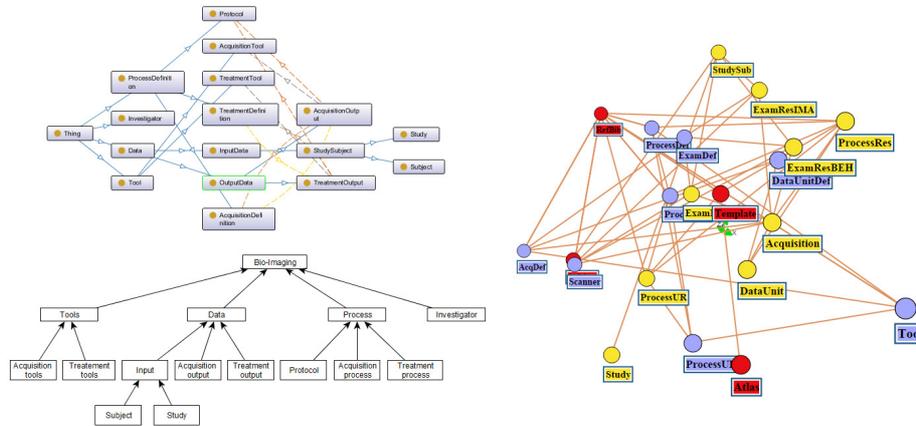
Fig. 8. Multiple layer view of query interface

Fig. 8. Illustration of the objects selection process and the visualization of return results

5. Conclusion

In this paper, we presented an approach using ontology as a solution to overcome the issues of database exploitation in the context of heterogeneous and distributed data. We then implemented this ontology in a semantic query system, and as the first results, the scientists at GIN can query database themselves without know previously the data model. As future work, we will focus on the test of query interface proposed with various sets of queries. For scientists at GIN, the return results need to be captured, represented, saved, enhanced, shared and reused by other users and in a

different context. Furthermore, it will be necessary to link the data (instance) in the database with concepts in the ontology model. The aim of this work is to enable the using of semantic query language (SPARQL for example) to query the database. This implementation will enhance the search performance of system.

References

1. Rahmani, T., Bougain, S., Gerhard, D.: Ontology-Based Integration of Heterogeneous Material Data Resources in Product Lifecycle Management. In: 2013 IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 4589–4594 (2013).
2. Terzi, S., Bouras, A., Dutta, D., Garetti, M., Kiritsis, D.: Product lifecycle management-from its history to its new role. Int. J. Prod. Lifecycle Manag. 4, 360–389 (2010).
3. Belkadi, F., Bosch-Mauchand, M., Kibamba, Y., Duigou, J.L., Eynard, B.: Functional Architecture and Specifications for Tolerancing Data and Knowledge Management. In: Rivest, L., Bouras, A., and Louhichi, B. (eds.) Product Lifecycle Management. Towards Knowledge-Rich Enterprises. pp. 35–45. Springer Berlin Heidelberg (2012).
4. Allanic, M., Durupt, A., Joliot, M., Eynard, B., Boutinaud, P.: Application of PLM for Bio-Medical Imaging in Neuroscience. In: Bernard, A., Rivest, L., and Dutta, D. (eds.) Product Lifecycle Management for Society. pp. 520–529. Springer Berlin Heidelberg (2013).
5. Sato, H., Otomo, K., Masuo, T.: A knowledge sharing system using XML Linking Language and peer-to-peer technology. In: 2002 Symposium on Applications and the Internet, 2002. (SAINT 2002). Proceedings. pp. 26–27 (2002).
6. Zhang, J., Liu, Y., Xiao, Y.: Internet Knowledge-Sharing System Based on Object-Oriented. In: Second International Symposium on Intelligent Information Technology Application, 2008. IITA '08. pp. 239–243 (2008).
7. Domingue, J., Fensel, D., Hendler, J.: Introduction to the Semantic Web Technologies. In: Domingue, J., Fensel, D., and Hendler, J. (eds.) Handbook of Semantic Web Technologies. pp. 1–41. Springer Berlin Heidelberg (2011).
8. Yoo, D., No, S.: Ontology-based economics knowledge sharing system. Expert Syst. Appl. 41, 1331–1341 (2014).
9. Davies, J., Duke, A., Stonkus, A.: OntoShare: Using Ontologies for Knowledge Sharing. In: Semantic Web Workshop (2002).
10. Raza, M.B., Harrison, R.: Ontological knowledge based system for product, process and resource relationships in automotive industry. CEUR Workshop Proc. 748, 23–36 (2011).
11. Abel, M.-H.: Competencies management and learning organizational memory. J. Knowl. Manag. 12, 15–30 (2008).
12. Arduin, P.-E., Le Duigou, J., Penciuc, D., Abel, M.-H., Eynard, B.: Knowledge Sharing Within Extended Enterprises: Case of Product Lifecycle Management systems. In: European Conference on Knowledge Management. pp. 63–71. , portugal, Portugal (2014).
13. Assouroko, I., Ducellier, G., Boutinaud, P., Eynard, B.: Knowledge management and reuse in collaborative product development - A semantic relationship management-based approach. Int. J. Prod. Lifecycle Manag. 7, 54–74 (2014).
14. Sriti, M.-F., Boutinaud, P.: PLMXQuery: Towards a Standard PLM Querying Approach. In: Rivest, L., Bouras, A., and Louhichi, B. (eds.) Product Lifecycle Management. Towards Knowledge-Rich Enterprises. pp. 379–388. Springer Berlin Heidelberg (2012).