

Linked Data Exploration in Product Life-Cycle Management

Soumaya El Kadiri, Ana Milicic, and Dimitris Kiritsis

LICP Laboratory, Ecole Polytechnique Fédérale de Lausanne, Switzerland
{soumaya.elkadiri, ana.milicic, dimitris.kiritsis}@epfl.ch

Abstract. Product Life Cycle Management (PLM) strategies have been proven as recommended approach for handling, planning and decision making regarding all aspects of products different life cycle stages. As this domain includes number of actors, system platforms and technologies, gathering information and handling knowledge often emerges as challenging problem. In this paper we propose a methodology to support a PLM approach taking advantage from the emerging semantic web technologies and the underlying paradigm of Linked Data. The main purpose of this methodology is twofold. First, it aims at providing a semantic model for handling different data sources and data types, structuring thus information from all stages of a product life cycle. Second, it extends reasoning mechanisms offered by semantic technologies, to conduct comparative and in-depth data analysis based on data mining and pattern discovery.

Keywords: PLM, Linked data, semantic technologies, ontologies, data analysis.

1 Introduction

In today's modern manufacturing world, products are often highly complex and actors in products life cycle are numerous, diverse and sometimes geographically distant. Yet, to be able to remain concurrent in highly competitive market, quality, ecological impact and pricing have to be optimized over entire life cycle. Different types and formats of data have to be gathered and analyzed on the one hand, and on the other hand, communication between sometimes very diverse software platforms have to be enabled. First step toward this goal is creating unique, all-spanning knowledge base which will map all concepts and relations present in the entire domain. Semantic models are shown to be a good technique for tackling such challenging projects.

In this paper, we first present the challenge in question, by elaborating in details PLM components and complexity. Life cycle stages are defined and key aspects are explained, giving the insight in diversity of processes that need to be taken into consideration. Further on, semantic web technologies are presented and range of available tools is explored. This gives a solid image of possibilities of semantic models, their structure and potential functionalities. Next, concept of "Linked data" is presented, together with the guidelines for defining the domain of relevant product related data. In this context, we propose a methodology based on the application of Linked Data paradigm to the context of PLM. We present the developed generic and specific

ontologies, as an example of reusable semantic models, developed within the European FP7 project LinkedDesign. Finally example of exploitation of linked data to capture additional knowledge is presented, extending thus reasoning mechanisms offered by semantic technologies.

2 Product Lifecycle Management

Product lifecycle management (PLM) is defined as a concept for the integrated management of product related information through the entire product lifecycle [1]. This vision is enabled by recent advances on information and communication technologies and is needed to support current industry needs for faster innovation cycles combined with lower costs. The aim of this integration is to overcome the existing organizational barriers and to streamline the value creation chain, based on the integration of multiple software components [1-3], generally enterprise resource planning (ERP), product data management (PDM) and other related systems, such as computer aided design (CAD) and customer relationship management (CRM) [4]. The term “lifecycle” generally indicates the whole set of phases, which could be recognised as independent stages to be performed by a product, from “its cradle to its grave”. Product lifecycle can be defined by three phases [5-6]:

- Beginning of Life (BOL) including design and manufacturing. In this phase, the product concept is generated and subsequently physically realised. Using many tools, techniques and methodologies, designers, planners and engineers develop the product design and the production process, plan the production facilities and manage manufacture of products with diverse suppliers.

- Middle of Life (MOL) including distribution (external logistic), use and support (in terms of repair and maintenance). In this phase, the product is in the hands of the final customer, i.e. product consumer and/or some service providers, e.g. maintenance actors and logistic providers. The product history related to distribution routes, usage conditions, failures and maintenance can be collected to create an up-to-date report about the status of products.

- End of Life (EOL) where products are retired – actually recollected in the company’s hands (reverse logistic) – in order to be recycled (disassembled, remanufactured, reused, etc.) or disposed. Information from EOL about ‘valuable parts and materials’ and other knowledge that facilitates material reuse should be routed to recyclers and re-users, who can obtain accurate information about product status and product content.

During BOL, the information flow is quite complete as it is supported by several information systems like CAD/CAM/CAE and PDM. However, in contexts where several information systems are used and where other structured and unstructured data sources are spread across different stakeholders, information flow becomes vague or unrecognized [7]. As a consequence, actors involved in each lifecycle phase make decisions based on incomplete and inaccurate information, which leads to operational inefficiencies [6]. The main challenge in this context is to provide a solution ensuring a seamless continuation of knowledge throughout the entire lifecycle and capable of capturing the dependencies between lifecycle stages.

3 Semantic Web Technologies and Open Web Linked Data

3.1 Semantic Web

The Semantic Web¹ provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is considered as a vision of information that can be readily interpreted by machines, so machines can perform more of the tedious work involved in finding, combining, and acting upon information on the web. The Semantic Web involves three core technologies: Resource Description Framework (RDF), Web Ontology Language (OWL), and Extensible Markup Language (XML). These technologies are combined in order to provide descriptions that supplement or replace the content of Web documents. The machine-readable descriptions enable content managers to add meaning to the content, leading to describe the structure of the knowledge we have about that content. In this way, a machine can process knowledge itself, using processes similar to human deductive reasoning and inference, thereby obtaining more meaningful results and helping computers to perform automated information gathering and research.

3.2 Linked Data Paradigm

The term *Linked Data*² refers to a set of best practices for publishing and connecting structured data on the Web. These best practices have been adopted by an increasing number of data providers over the last three years, leading to the creation of a global data space containing billions of assertions—the Web of Data. Linked data is based on four principles as follows:

All items should be identified using URIs.

- All URIs should be *dereferenceable*— that is, using HTTP, URIs enable anyone (machine or human) to look up an item identified through the URI.
- Looking up a URI leads to more data.
- Links to URIs in other data sets should be included to enable further data discovery.

In contrast to semantic web vision, Linked data is mainly about publishing structured data in RDF using URIs rather than focusing on the ontological level or inference (4).

4 Linked Data in PLM

Having the concept of Linked Data in mind, we propose in this section to define a process of conceptualizing the PLM approach through four main activities: specification, modeling, generation and publication, and exploitation, described in the following sub-sections.

¹ <http://www.w3.org/2001/sw/>

² <http://linkeddata.org/>

4.1 Specification

At this first stage, the need is to define the requirements to design a semantic model covering several aspects of the product lifecycle. In our previous work [8], we have defined a bottom-up approach based on the User Story Mapping method (USM). This method is user-centric and leads to the definition of current and/or expected scenarios and processes. This bottom-up approach, which combined with appropriate tools and methods (such as questionnaires, standards specifications, knowledge based approaches, etc.), resulted in the specification of the knowledge network and domain for sharing and reusing knowledge in collaborative product development.

4.2 Modeling

This second stage aims at modeling the domain and providing a semantic model to cover the requirements of PLM applications. In our previous work, published in [9], we have designed a fundamental ontology which aims to be easily adjusted and adopted for different product engineering systems, thus eliminating the need for repetition of entire design process for every individual company. Being generalized, this ontology needs to be specialized for each specific application. Schema of concepts of this ontology is given in the following figures. This ontology has been specialized to two use-cases: LCC application in product design, published in [10-11], and metrology in manufacturing.

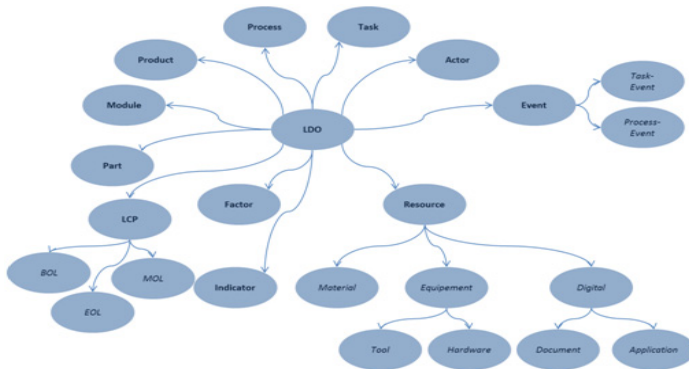


Fig. 1. Generic concepts of LDO

4.3 Generation and Publication of Data Sets

A variety of approaches to mapping data-sources with the ontology and methods for generating RDF statements and publishing them as Linked Data exists in the literature. Virtuoso RDF views [12-14] for instance are production-ready tools for generating RDF representations from relational database content. Some of them even aim at automating partially the generation of suitable mappings from relations to RDF vocabularies.

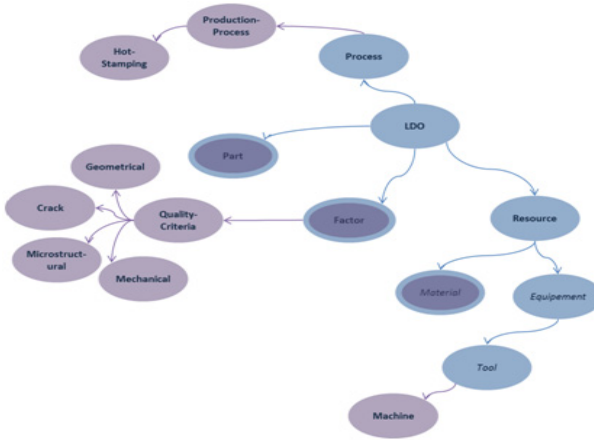


Fig. 2. Specification of ontology for metrology in manufacturing

4.4 Exploitation

This final stage consists of analyzing data taking advantage from structuring as well as reasoning features offered by semantic technologies in order to analyze dependencies between lifecycle stages data from the generated datasets. The next section is dedicated to this phase and provides an industrial use-case for data analysis.

5 Data Exploitation and Knowledge Discovery

5.1 Preparation of Data

Striving towards sharing of data and knowledge exchange between different life-cycle phases has led to immense improvements in manufacturing world. Data collected during exploitation and maintenance of a product are used by designers to make better decisions during design of future products. Same data are used by engineers to improve manufacturing line designs. Data from manufacturer and maintenance are used for improving recycle and reuse decisions. Yet, selection of data to be exchanged is always very case specific, and it's based on some already assumed cause-effect relations. Design of unified data bases, opens a possibility of having overview of all available information, but in these cases it becomes challenging to manage these high volume data, especially in a case of more complex products, whose manufacturing and maintenance require number of individual processes performed by different actors.

By modeling entire product life-cycle domain in one semantic model, we are able to maintain the same advantage of having overview of all available information and also exploit on structure of RDF metadata. In our work done so far on the domain of product life-cycle, we have used ontology to model this RDF statements graph. This choice was made due to number of additional advantages of ontology such as, reasoning over data, consistency check-up and variety of visualization options.

Having concepts from all phases of life cycle in one model, allows us to define assumed cause-effect relations, thereby completing the model. The real added value is created once we employ reasoning over data, and in this case, we benefit from rule inference specifically. Starting from base set of rules and relations, ontology will automatically expand this network, thus creating all possible rules and relations that can be inferred. This new relations give a very good insight in where unknown cause-effect processes might be found.

5.2 Application of Data Mining

Important part of exploitation of data merged from entire product life-cycle is data mining and pattern discovery, since it results in potentially yet unknown dependencies present in data. Using entire collection of data is unsuitable for processing due to high diversity of data and too large number of topics merged. Appropriate data sets have to envelope potentially correlated concepts and still not span over too wide part of the domain. Rules and relations within ontology, defined by human experts as well as those inferred, are natural indicator on where dependencies exist and thus, they can be used for extracting useful data sets. By merging attributes of concepts between which there is a relation, we get data sets for which data-mining exploration could give new knowledge and understanding of product life-cycle. Since all the attribute values will be of correct type and within predefined ranges, detecting missing data or outliers becomes obsolete. Having data structured within ontology, thus simplify the time consuming data preprocessing step.

Our use-case is driven from metrology applications in automotive industry, and is limited to design and manufacturing processes. Following the proposed methodology described in the previous sections, and using the ontology network implemented in our previous works, we have applied, at this stage of developments, a manual approach to generate related datasets. One example of benefits coming from gathering data from different life stages of a product is given in a Figure 2 and Figure 3.

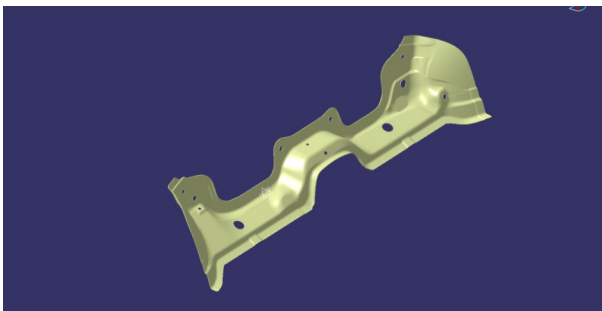


Fig. 3. CAD model of a part

Figure 2 displays CAD model of a part, as it was designed. On the other hand Figure 3 displays 3-D scans of physical parts manufactured according to design. On a left-hand side image, the manufacturing process was performed successfully, while on the right-hand side shows crack. Having these data sets from design stage and

manufacturing stage in same data model, enables development of a tool for automatic crack discovery. This is only one example of prospective exploitation of semantic data model spanning entire product life cycle.

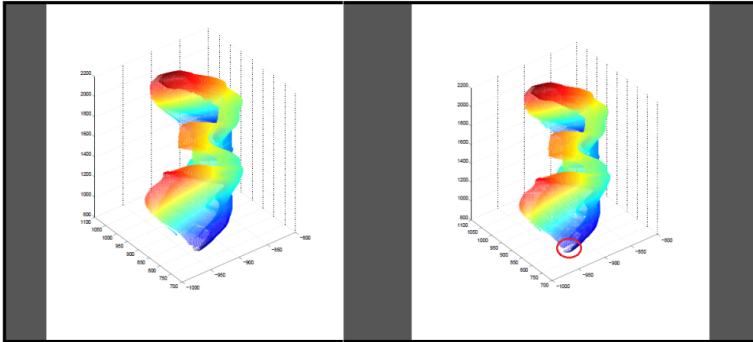


Fig. 4. Scan of a manufactured part, without crack (left) and with crack (right)

6 Conclusion

The application of Linked Data paradigm in the context of PLM seems to be a promising approach to deal with numerous and scattered system data sources involved in the lifecycle. Ontology as a key enabler of Linked Data are interoperable systems that can be merged and aligned, thus eliminating the need for any future translators or communication bridges between different systems. The larger the domain covered with ontology model is, the more powerful exploitation mechanisms become. Design and implementation of such unified data and knowledge model, opens endless possibilities for application of tools for automated data analysis, knowledge extraction, reasoning over data, patterns discovery and domain visualization. The proposed approach can be considered a step toward more efficient knowledge reuse and new querying techniques. Finally, numerous technics in the field of learning mechanisms and artificial intelligence can be applied in this context in order to extend reasoning capabilities beyond inference-based technics offered by semantic technologies. By applying these technics and taking advantage from the Linked Data paradigm and related enabling technologies, we will be able to provide anticipation capabilities supporting decision making in product lifecycle management.

References

1. Saaksvuori, A.: Immonen: Product Lifecycle Management. Springer, Berlin (2004)
2. Abramovici, M., Schulte, S.: Product Lifecycle Management – Logische Fortsetzung der PDM-Ansätze oder Neuauflage des CIM-Debakel. I2P-Integrierte Informationsverarbeitung in der Produktentstehung – (k)ein Gegensatz zwischen Innovation und Kostensenkung, VDI (2004)

3. Arnold, V., Dettmering, H., Engel, T., Karcher, A.: *Product Lifecycle Management beherrschen: ein Anwenderhandbuch für den Mittelstand*. Springer, Berlin (2005)
4. Boos, W., Zancul, E.: PPS-Systeme als Bestandteil des Product Lifecycle Management. In: Schuh, G. (ed.) *Produktionsplanung und -steuerung, Grundlagen, Gestaltung und Konzepte*. Springer, Berlin (2006)
5. Kiritsis, D., Bufardi, A., Xirouchakis, P.: Research issues on product lifecycle management and information tracking using smart embedded systems. *Advanced Engineering Informatics* 17 (2003)
6. IMTI Inc. : *Modeling and Simulation for Product Life-Cycle Integration and Management*, White Paper (2002)
7. Jun, H., Dimitris, K., Paul, X.: Research issues on closed-loop PLM. *Computers in Industry* 58(8-9) (December 2007)
8. Milicic, A., Perdikakis, A., El Kadiri, S., Kiritsis, D., Ivanov, P.: Towards the definition of domain concepts and knowledge through the application of the user story mapping method. In: Rivest, L., Bouras, A., Louhichi, B. (eds.) *PLM 2012. IFIP AICT*, vol. 388, pp. 58–69. Springer, Heidelberg (2012)
9. Perdikakis, A., Milicic, A., El Kadiri, S., El Kadiri, S., Kiritsis, D., Pardalis, Alexandrou: *Ontology Usage in the Context of Next Generation Product Development*. In: *APMS 2012, Rhodes Island-Greece*, September 24-26 (2012)
10. Milicic, Perdikakis, El Kadiri, Kiritsis, Terzi, Fiordi, Sadocco: *Specialization of a Fundamental Ontology for Manufacturing Product Lifecycle Applications - A case study for Lifecycle Cost Assessment*. In: *EI2N 2012, Rome, Italy*, September 12-13 (2012)
11. Milicic, Perdikakis, El Kadiri, Kiritsis: *PLM Ontology Exploitation through Inference and Statistical Analysis - A Case Study for LCC, St Petersburg, Russia* (June 2013)
12. Blakeley, C.: *Rdf views of sql data declarative sql schema to rdf mapping* (2007)
13. Erling, Mikhailov: *RDF support in the Virtuoso DBMS*. In: *Proceedings of the 1st Conference on Social Semantic Web, GI edn. LNI*, vol. P-113. Bonner Kollen Verlag (September 2007) ISSN 1617-5468
14. Bizer, C.: *Seaborne.: D2RQ - treating non-RDF databases as virtual RDF graphs*. In: *ISWC 2004 (posters)* (November 2004)