Strategic Industrial Alliances in Paper Industry: XML- vs. Ontology-Based Integration Platforms

Anton Naumenko, Sergiy Nikitin, Vagan Terziyan, Andriy Zharko

Industrial Ontologies Group, Agora Center, University of Jyväskylä, Finland
e-mail: vagan@it.jyu.fi

Abstract

Strategic alliances occur everywhere nowadays as a result of economy globalization. Goals of collaboration differ from alliance to alliance depending on industry, organizational, functional, social, legal, etc. aspects. Generally alliances have characteristics which are not a cumulative sum of characteristics of the parts. Collaborative business intelligence is one of the most desirable characteristics of long-run collaboration. The new opportunities should not be necessarily explicitly specified but should be derived automatically on case-by-case basis. To enable this there is a need to make a platform for integrated consortium (to manage the alliance goals, information and activities) not instead but in addition to existing platforms for consortium partners, which will still autonomously manage their local businesses. In this paper we discuss possible challenges for such integration taking as an example the Paper Industry alliance effort called PaperIXI. We analyze ongoing activities of the alliance (based on available information from open sources), which is promoting an XML-based solution, in contrast to our view, which is based on applying Semantic Web technology. Some motivating examples of industrial implementations based on Semantic Web technology provided by NASA activities are shown, which according to our view can be utilized for the PaperIXI alliance platform.

1 Introduction

Semantic Web [SemanticWeb] is expected to become a next-generation of the Web assuming that besides existing content there will be a conceptual layer of machine-understandable metadata, making the content available for processing by intelligent software. This allows automatic resource integration and provides interoperability between heterogeneous systems. Next generation of intelligent applications will be capable to make use of such metadata to perform resource discovery and integration based on its semantics. Semantic Web, aims at developing a global environment on top of Web with interoperable heterogeneous organizations, applications, agents, web services, data repositories, humans, and so on. On the technology side, Web-oriented languages and technologies are being developed (e.g. RDF [RDF], OWL [OWL], OWL-S [OWL-S], WSMO [WSMO], etc.), and the success of the Semantic Web will depend on a widespread industrial adoption of these technologies. Trend within worldwide activities related to Semantic Web definitely shows that the technology has emerging grows of interest both academic and industry during a relatively small time interval. The growing interest to the Semantic Web, as a research and educational domain, from the academy is evident. New scientific results and interesting challenges in the area appear rapidly. International networks cover topics related to intersections of various former scientific domains with Semantic Web technology and discover new challenging opportunities. Basic standards have been announced and the amount of pilot tools and applications around these standards is exponentially increasing. In spite of growing hype around Semantic Web and appropriate standards, industry developed and is continuously developing own standards for interoperability and integration. There are probably some reasons for refusing wider scale implementation of Semantic Web standards. On the other hand, more and more companies are being involved to various projects related to Semantic Web. Industrial investments to research projects aimed to monitor the status of the technology are also growing. Some companies are extensively involved to the appropriate business. There are at least two categories of such enterprises: those who are producers and providers of Semantic Web based products and services and those who are consumers of these products and services.

In this paper we are trying to make special emphasis on such groups of enterprises, which are going to join their businesses and as a framework for possible integration of their resources and services seriously consider Semantic Web technology.

There are still serious doubts among the Semantic Web community about “killer application” of the technology. Key persons (e.g. Tim Berners-Lee in his keynote speech in WWW2004 in New York) admit that there's no yet killer application discovered to show the public what the framework can do. As the Semantic Web enters what Tim Berners-Lee calls its second phase in 2004, developers must start building applications and make working code available to the public. Moreover, Berners-Lee told his colleagues to forget about looking for a killer application for the Semantic Web: proof of the technology will emerge when new links among information begin to emerge, he said.

DIP consortium [DIP], which leads one of the largest European projects in the area of Semantic Web, believes that a combination of semantic web and web services technology may well deliver the killer application for Semantic Web. They believe the combination can provide an infrastructure that will not only revolutionize information processing, but also the way we access computational resources in general.

According to our view, an integration in general can be considered as a "killer application" of Semantic Web technology, which particularly can be interpreted as heterogeneous data integration, Enterprise Application Integration and Web-service integration among other interpretations. Integration of heterogeneous components can be considered as merging functionality of these components in such a way that the resulting functionality will not be simply a sum of component functionalities, but also something else. Important is that the additional functionality can be automatically derived on the basis of Semantic Web technology. Taking the case with integrating businesses of several companies, the above means that the integrated alliance should have richer business models and opportunities than the same set of companies had, being autonomous; and also that the new opportunities should not be necessarily explicitly specified but can be derived automatically on case-by-case basis. To enable this there is a need to make a platform for integrated consortium (to manage the alliance goals, information and activities) not instead, but in addition to existing platforms for consortium partners, which will still manage their local business. In this paper we further discuss possible challenges for such integration, mainly from a technological point of view, taking as an example the Paper Industry alliance effort called PaperIXI [PaperIXI]. We will try to show ongoing activities of the alliance that promotes an XML-based solution in parallel with some motivating examples of industrial implementations based on Semantic Web technology provided by NASA activities [TQNASA]. Finally, the paper will present our own view, which stands for advantages of applying Semantic Web technology for alliance platforms.

The content of the paper is structured as follows. Section 2 concentrates on the challenges, which partner companies meet in their aiming to a collaborative product data management. Section 3 analyses conceptual differences between ontology-based modeling and document-oriented paradigm. In Section 4 concrete examples, where semantic modeling is more advantageous than a conventional one, are given. Section 5 discusses architectural aspects in building a collaborative business environment. Conclusions are given as a final section for summarization of the gained research results and convictions concerning Strategic Industrial Alliances.

2 Alliance Collaboration around a Supply Chain

2.1 General challenges in organizing a Virtual Enterprise

Every modern alliance faces a challenge of the collaborative product data management in its initiative towards enabling concurrent engineering of a product by partners. Concurrent engineering gives efficient utilization of time and money, but highly increases a communication load. There are several critical issues to support concurrent engineering:

- Building a networked organizations focusing on a target product.
- Facilitating communication among participants of networked organizations.
- Bringing up considerations of later stages such as manufacturing, purchasing, operation, maintenance and recycling to the discussion table.
- Developing a computational infrastructure to support these three issues.

General business process model for management of collaborative product data life-cycle have a form is represented in Figure 1. The left model have been designed within PaperIXI project for pulp&paper industrial Maintenance, Repair and Operation (MRO) life-cycle. The right one was designed by collaborative efforts of Space Shuttle Program Office and NASA's Engineering of Complex Systems (ECS) Group [ECS] in search of a solution, which could improve System Lifecycle Management. For some reason, the model for pulp&paper MRO life-cycle does not form a loop, despite a cyclic nature of modern life-cycles of business processes.

The business process models represent results of a functional modeling using common and formal methods. Particularly for that purpose IDEF notation can used, which allows a functional modeling and represents a business process model as a set of diagrams collected into a hierarchy of decompositions. The process model derives a general picture of a supply chain from activities of corresponding organizations and their units within a collaborative network.

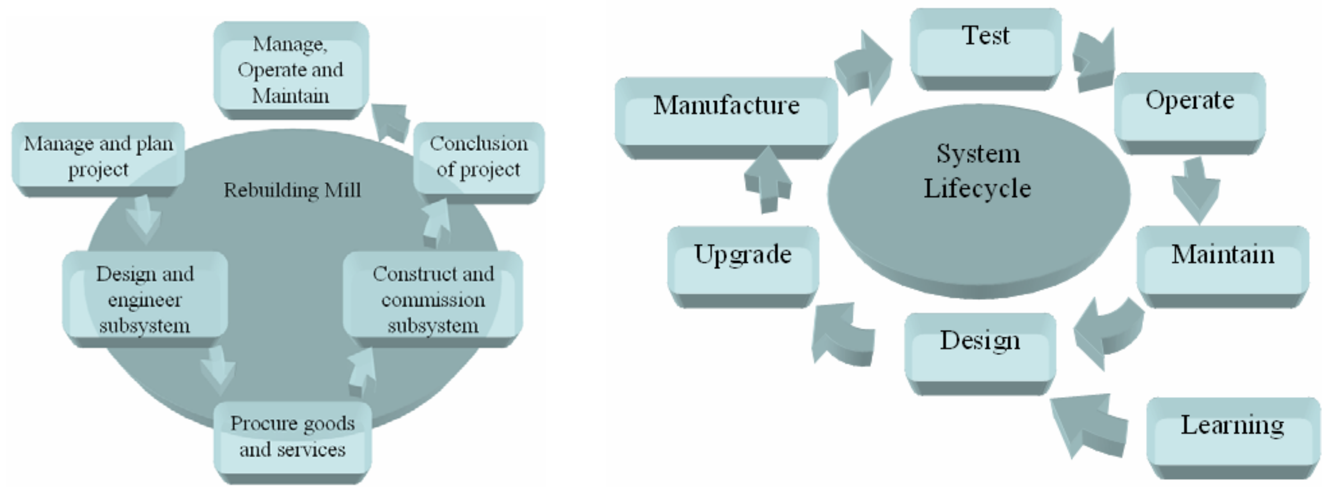

Figure 1. General business process model for management of collaborative product data life-cycle

As it was pointed out above, alliances inevitably encounter challenges concerned with management of a huge amount of product documentation. Several key problems revealed, are [ECS, PaperIXI]:

- Product engineering processes and organizations extend across large enterprise systems with numerous heterogeneous data sources.
- Workgroups are geographically dispersed (efficiency issues, corporate knowledge loss).
- Engineering drawings and Engineering Orders can exceed one million in amount.
- Problem of a poor accessibility leads to a more general problem of data extraction from heterogeneous sources of information usually only partially structured.
- Documentation is difficult to maintain, mainly because of two reasons: (a) frequent change that causes documentation to be outdated soon and (b) lack of collaboration between parties different by interest across and even inside organizations.
- High cost of a delivery of new versions.
- Insufficient information content.
- The processes across the supply chain were mainly limited to companies' own points of view.

Efficient management of documentation is of primary importance across engineering lifecycle. Manufacturing systems sense and alter state of the physical world directly, thus engineering datasets are typically large and have complex interrelationships [TQNASA]. Thereby a network of independent companies that share experience, knowledge and capabilities requires support for information flows among systems across enterprise boundaries. The task of delivering the right information to the right person at the right time constitutes the following requirements: flexibility, efficiency and responsiveness.

The vital needs of emerging "Virtual Enterprise" can be summarized as [TQNASA]:

- Provide a method of incorporating additional lifecycle information (attributes, design notes, assumptions, manufacturing details, test results, process requirements, classification, historical and tracking data, contextual use) through a knowledge capture technique.

- Enhance the usage of integrated engineering models for operations with explicit risk representations and subsequent knowledge capture, characterization and reuse throughout the lifecycle.
- Shared model is needed to communicate, plan and exchange information.
- Provide common view of a system for information search.

2.2 Modeling a problem domain

Continuous product data modeling is a key to allow collaborative product data management in a supply chain. The most significant contribution of the model is a common understanding, interpreting and operating of collaborative organizations. Unified model enables solutions for interoperability between existing tools and elaboration of general tools for industry. In addition to interoperability, a choice of relevant standards for representation and serialization of the model will meet requirements on dynamics of the model through a time. One of such requirements can be effortless modification of tools based on that model. Consolidation of all product data throughout a whole product life-cycle becomes possible thanks to a universal model in the middle.

A business model is built around a universal model of a mill, which can rapidly increase quality and efficiency of the collaboration. The universal mill model concept formalizes meta- and technical data of MRO life-cycle ща a mill. Mill model is divided into a data model and a dictionary data model. Dictionary data model, classification systems, hierarchies and property sets often are out of scope of solution providers [PaperIXI] as Figure 2 shows.

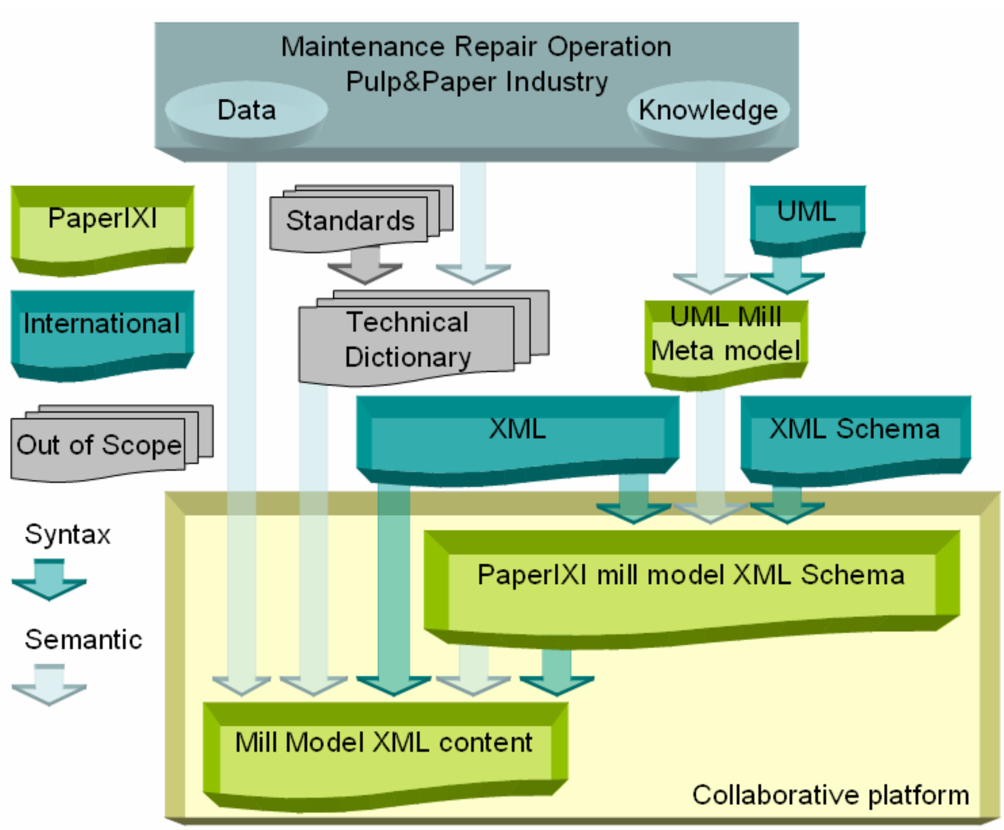

Figure 2. Universal modeling landscape

The main problems and challenges related to a conceptual model common for all alliance members were identified to be the following [PaperIXI]:

- Companies have different conceptualization of a domain and do not have common terminology varying in languages, referred standards, paradigms of modeling.
- Model has to satisfy a wide range of dynamically changed requirements.

- Model has to provide relevant views for different parties based on specific needs.
- Existing standards for product data management are overlapping and incompatible or too generic.
- Generic standards are pending or are too complicated for implementation and they are not used, because they cannot be downloaded free of charge.

Universal meta-model has to follow international standards to ensure its applicability and efficiency on a long-run. It also has to enable transformation of engineering documentation into a form that can represent current state, eliminating ambiguity, uncertainty and misinterpretation [TQNASA]. Additionally, the meta-model and related metadata always have to conform to a reality to be directly usable for current engineering tools and practices.

2.3 Interoperability within an alliance

Interoperability comes from the field of communication. Successful communication between partners in an alliance relies on three principles:

- Common Syntax (the structure of the message).
- Common mechanism of a message delivery.
- Common Semantics (the meaning of a message).

XML is becoming a common standard for syntax. Historically, TCP has become a common mechanism for communication between systems, supporting higher level communication mechanisms such as HTTP over TCP/IP. However, having a common syntax and common mechanism of a delivery is not enough. Interoperability requires that the systems have a common definition of what is to be shared or communicated. We need an infrastructure to support semantic alignment [IAC] and recently XML format is considered to be not enough for this purpose.

Product data sharing and exchange in a collaborative environment of an alliance raise several issues and requirements to modeling principles and format for data exchange. The following categories of problems of documentation exchange and sharing in a life-cycle of *maintenance, repair and operation* of a mill have been identified [PaperIXI]:

- Language problems spotlight multilingual audience of technical documentation across the globe.
- Versioning problems are caused by a need of early adoption of preliminary versions of documents to meet high requirements of concurrent engineering in collaborative environment.
- There are different types of media associated to products that vary through almost all possible types of digital media. Efficient exchange and sharing in this context lead to a more generic problem of information extraction and integration from heterogeneous sources.
- A need in formal specification of metadata about cooperation such as goals, process, states, conditions, context, competences, responsibilities, partners' roles, etc.
- Huge amount of technical information for complex products requires scalable solution for data specification, advanced performance of a search for precise information and put high requirements on issues of user interface usability.
- Ability to understand client's information needs is one of the most important enabling factors for information sharing and exchange.
- Nomenclature and identification problems constitute a present situation, caused by misunderstanding in terminology, classifications, identification systems, coding systems, etc.

The process of Product Data integration also poses challenges for alliance partners. Some of them are related to data models and corresponding data structures. The sticking point of this aspect is in differences of modeling principles for the domain between companies-partners. This fact leads to irrelevancy in structures that are used by maintenance units and component or system suppliers [PaperIXI, ECS] (Figure 3).

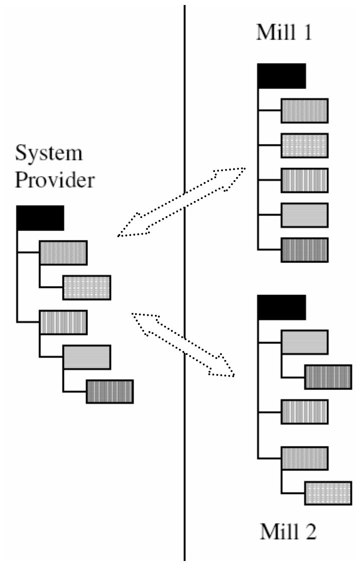

Figure 3. Heterogeneity in product structuring between partners of an alliance

On the other hand, structuring principles being applied to operational documents are also heterogeneous among the companies. As a result, updating partially changed documents becomes difficult, because a document represents an atomic element of information in this context. Furthermore, system or component suppliers make documents that do not describe only the equipment delivered, but the same document describes often several other equipments that are irrelevant for the document reader [PaperIXI].

As a next challenge in the Product Data integration, an absence of common classification and identification comes up. Each Mill maintenance system utilizes its own identifiers for the Product Data items that are not relevant for the system suppliers. The same kind of heterogeneity dominates in the coding systems used for identifying [PaperIXI, ECS]:

- Products, individuals or functional property (consider PSK5941 [PSK] as a possible candidate for it).
- Units of measurement (an example – UNECE Codes [UNECE] for Units of Measure Used in International Trade).
- Countries (consider ISO3166 [ISO-3166]).
- Companies (sample: D-U-N-S number [DUNS] or IANA Private Enterprise Number [IANA]).
- Document metadata (e.g., MIME format [MIME], language [ISO-639]).

Additionally, some coding systems in computer unreadable form are used within some organizations.

If to consider the challenges related to data transformation needs that occur within the Paper Mill Lifecycle, some difficulties exist there, too. They are partially concerned with removing obsolete data after upgrading instrumentation or management process, with checking validity of data that are distributed among the participants of the supply-chain [PaperIXI]. Data are handled sequentially through the supply-chain (Mill -> Engineering consultant -> System supplier -> Component supplier) and in any time, when someone handles the data, errors may happen. Same data is copied in several places and changes in the requirements do not reach all the users.

The fact that Mill maintenance systems do not store all the manufacturing details represented in bills of materials prevents the automation of Knowledge Management as well. Consequently, a particular data supplier cannot produce all the data needed in a certain Mill maintenance system. At the same time, partners receive redundant data and its removing requires manual efforts. Currently, as the life-cycles of a supply-chain tend to form a loop, one-way transformation is not enough. Transformations during a communication of a feedback in Mill maintenance create problems of interoperability once again [PaperIXI].

3 Building a Collaborative Environment for a Supply Chain Alliance

3.1 Conventional Meta-modeling

As it was stated in Section 2, a common domain model is a starting point in building an efficient collaborative environment for a supply chain alliance. An approach selected for a design of the model will predetermine the overall efficiency of the implementation of the environment.

UML has been widely used for designing the domain models. If to take a meta-model for a mill supply chain, it will represent mill and product structures and spatial relation between components of a mill taking into consideration a need to describe technical documentation. Component, document and relation will be main concepts of the mill model. A candidate domain model may define function, product and individual as its three main components, while giving a possibility to extend them by companies' specific components (Figure 4). Attachments refer to external technical documents. Objects can also have relations between each other (such as functional or product structures), that are represented using the Relation concept. This aspect of a modeling gives a possibility to overcome the existing problem of insufficient explicit interlinking between pieces of information in a collaborative environment of a supply chain [TQNASA]. This approach is anticipated to provide a solution for transferring references to necessary documents from minds of people to their explicit formal definitions.

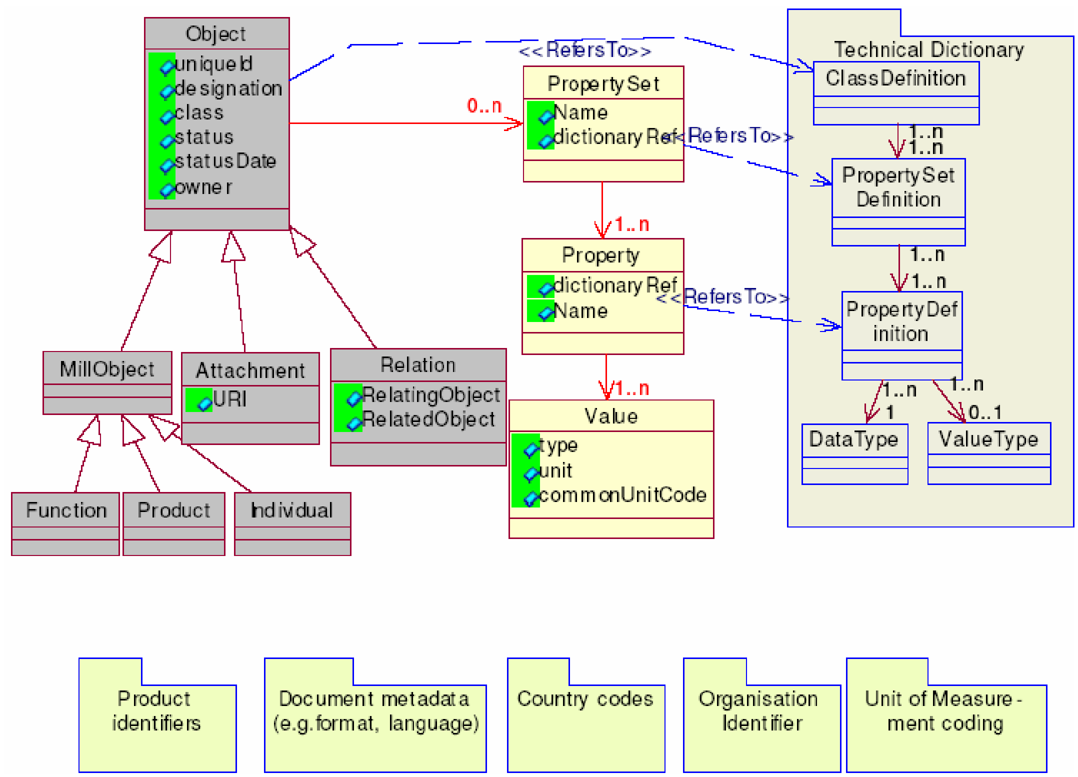

Figure 4. Candidate data model, adopted from [PaperIXI]

- Functional mill object defines function of the mill process, equipment or component.
- Product mill object defines generic product in the mill, not containing information of the individual physical objects.
- Individual mill object defines physical object in the mill, containing information such as serial number, maintenance history etc.
- Attachment defines a link to the external document.
- Relation contains link to the objects that are related to each other.

The mill model has layered structure that spreads from a mill structure to product structures. The solution represented above does not assume intervention into or modification of existing product data structures being utilized by every partner in the supply-chain. It is just advisable to apply hierarchical approach in the modeling and to communicate the resulting structures to every related partner [PaperIXI]. Issues of product data modeling are considered as belonging to internal processes of companies and have been left out of scope of the analysis. Higher degree of granularity for the structures in question is also recommended, because it makes them more flexible, but then increasing the difficulty of their management.

To overcome the heterogeneity in the item classification and identification, it is recommended to gradually switch to common international standards. More smooth way in the transition can be creation of a consortium-wide specific standard as an aligned union of the existing partner-specific, e.g. Forest Industry Grouping for Materials and Services or Swedish SSG product database [SSG]. However, even for the cluster-wide standard referencing to more common international standards are recommended. Additionally, to cope with a hitch between suppliers of instrumentation and parties, which exploit it, it is advisable to use in maintenance data exchange the same classification system that is selected for the business transactions of the cluster domain [PaperIXI].

Much higher level of automation (combining, validation and mappings) can be achieved in Knowledge Management within the consortium, if to follow a computer-readable format in data definitions. The proof-of-concept data representation format is based on PaperIXI Mill model XML schema combined with RosettaNet Technical Dictionary (RNTD) format [RNTD].

Common XML schema that was designed within the PaperIXI project based on the presented model is dedicated to solving the challenges related to document structures. Compliance of the documents with the PaperIXI manuals XML schema provides better modularity for them and therefore better flexibility in automated document updates, integration and delivery of relevant content. The strict following to the defined XML schema in the document formatting opens broad possibilities for automated transformations of the content as required initially. Using XML-based tools, it is simple to create filters for removing unwanted object from the structures, assuming that the objects to be filtered are identified, e.g. by defining the product classes that are not transferred.

PaperIXI limits consideration of possible candidates for data exchange standards to business transaction standards and STEP family standards. The own product data exchange standard was introduced to occupy place between mentioned above. PaperIXI mill model XML data format intends to efficient exchange of mill model objects. XML schema of the data format reflects mill model concepts on upper level while giving possibility to specify information about any sub-concepts. Format is defined to be as simple as possible. Result of this is that it cannot reflect complex standards like STEP while keeping ability to deliver sufficient data of MRO life-cycle according to mill model definition. Generally it allows exchanging of mill model objects, mill and product hierarchies, objects' properties and references to documentation.

3.2 Semantic Meta-modeling

As another approach to a formal modeling of a problem domain, which meets adequately the requirements described in Section 2, is an ontological one [RDF, OWL, WSMO]. This modeling approach using graph model represents information content as a set of interrelated semantic concepts and their properties with references to external specifications, which formally define and justify the semantics. Ontological approach also assumes close co-existence of data and metadata definitions, while making semantics easily readable and identifiable by software.

The semantics of data models often constitute an informal agreement between the developers and the users of the data model [Meersman] and which finds its way only in application that use the data model. E.g., in many cases, the data model is updated on the fly as particular new functional requirements pop up. In the context of open environments (as Semantic Web), ontologies represent knowledge that formally specify agreed logical theories for an application domain [Guarino98]. Ontological theories, i.e. a set of formulas intended to be always true according to a certain conceptualization [Ushold], consist of domain rules that specify – or more precisely, approximate – the intended meaning of a conceptualization.

Ontologies and data models, both being partial accounts (although in a varying degree) of conceptualizations [Guarino95], must consider the structure and the rules of the domain that one needs to model. However, unlike task-specific and implementation-oriented data models, ontologies, in principle and by definition should be as much generic and task-independent as possible. The more ontology approximates the ideal of being a formal, agreed and shared resource, the more shareable and reusable it becomes. As Ushold remarks, reusabilility and reliability are system engineering benefits that can be derived from the use of ontologies [Ushold].

Typical domains, where the application of the Semantic Web's approach becomes crucial, are those, which utilize in their life-cycles large amounts of documentation with heterogeneous underlying structures [OWL Req]. One of the most evident examples is a paper industry with its great variety of engineering documentation that relates to design, manufacturing, testing, etc. The documents have underlying hierarchical structures, which are heterogeneous through different document sets. Due to close relationships between the content of the documents, they have frequent cross-references. Using the semantic modeling approach, Technical Dictionary, meta-model of the mill and further conceptualization will be represented by ontology that gives possibility for application to identify and interpret data. A concrete example for the Paper Mill model can be related to design documentation, where typical users are:

- Maintenance engineer looking for all information relating to a particular part (e.g., "control valve").
- Design engineer looking at constraints on re-use of a particular sub-assembly.

To be applicable to the mentioned cases of usage, the relevant ontology must define appropriate constraints. These constraints can be utilized for advanced search or checking consistency. Here is one possible example of a constraint:

controlValve(X) => CardinalityOf(out-port(X)) = 2

femaleMember(X) AND controlValve(Y) AND isComponentOf(X,Y) => length(X) < length(Y)

According to the analysis in [Spyns et al], data models, such as database or XML-schemas, typically specify the structure and integrity of data sets. Thus, building data models for an enterprise usually depends on the specific needs and tasks that have to be performed within this enterprise. Further analysis attempts to determine, in which aspects the semantic modeling approach can be more advantageous than conventional methodologies.

Generally, the approach of semantic modeling can be considered as an attempt of switching from document-centric to knowledge-centric approach. Knowledge-centric approach allows product data annotation so that it becomes unambiguous and classified. OWL is envisioned to be an efficient solution supporting the complex assets over their long life-cycle [Price].

As one of the earliest solutions for the challenges being in focus of this paper, a project led by TopQuadrant [TQ] company ordered by NASA Ontologies, can be mentioned. The project results have proven that Ontology-based solution provides a new value to data sharing by setting a semantic bridge between heterogeneous components by explicit linking relevant pieces of data. NASA Enterprise Architecture Process Ontology uses different types of associations to interlink involved objects. As far as ontology defines classification schemes, relations and rules to describe domain, data being exchanged stays consistent with all components because it is bound to ontology.

Figure 5 depicts interrelations of domain, ontology, data and semantic standards.

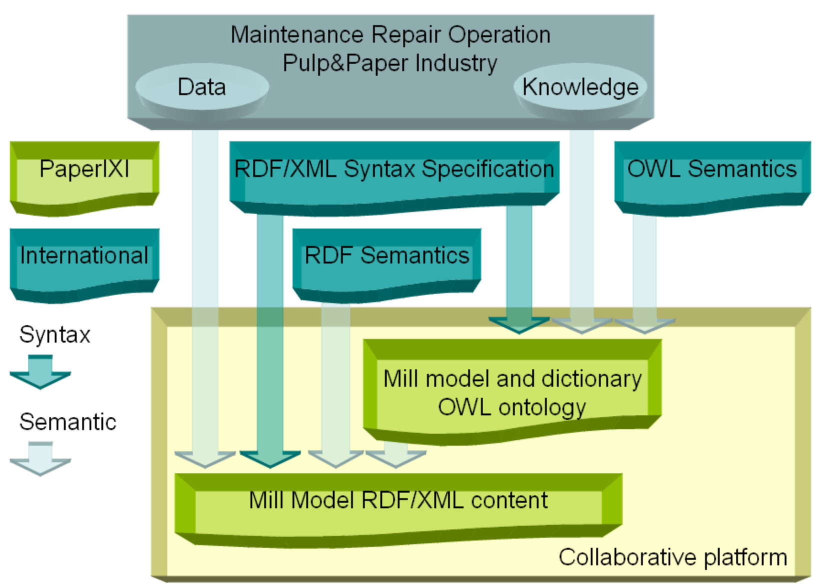

Figure 5. Modeling landscape of the semantic vision

Knowledge and data of the paper domain are spread over levels according to division on data and metadata. Table 1 summarizes the levels and solutions of PaperIXI as a representative of a conventional modeling approach and Semantic Web based vision.

Table 1. Levels and components of data abstraction.

<table>
<tr><th>Levels</th><th>Description</th><th>PaperIXI</th><th>Semantic vision</th></tr>
<tr><td>Meta model</td><td>A formal language used to represent modeling concepts</td><td>UML, EXPRESS</td><td rowspan="2">OWL Web Ontology Language RDF/XML Presentation Syntax.
Ontology of MRO in paper industry domain</td></tr>
<tr><td>Technical Dictionary</td><td>Dictionary and classification of concepts within domain</td><td>Out of scope</td></tr>
<tr><td>Implementation schema/model</td><td>A formal representation of concepts within the domain (a view onto real-world objects and ideas).</td><td>Domain specific XML schema</td><td>RDF/XML Syntax Specification</td></tr>
<tr><td>Data transfer file</td><td>An instance of a implementation schema. Defines specific information within the structure and meaning imposed by the chosen schema</td><td>XML data</td><td>RDF/XML data</td></tr>
<tr><td>Database implementation</td><td>The concepts implemented in the database storage</td><td>SAP, Oracle, …</td><td>Joseki, Sesame RDQL, Protégé, Jena …</td></tr>
</table>

4 Advantages of OWL over XML

Let us consider several examples of how XML-based and Semantic Web based solutions cope with changes in domain models. Such dynamics is considered to be an inevitable feature of collaborative environments used by modern supply chain alliances. Structure of XML-documents is tree-based, while RDF-document is a graph. This fundamental discrepancy results in different properties of documents and collections of documents.

Example 1:

If technical dictionary of Mill model is extended with new property and this property becomes obligatory for certain Class Definition. Then XML-documents developed according to PaperIXI-schema and Technical Dictionary should be modified in order to match new class structure. In case of ontology-based model if certain class changes we need to modify corresponding documents also, but here comes significant difference. In order to provide these changes to XML-document we need to change its tree-structure. This could cause consequences demanding to change XSLT-stylesheets, transforming this document to other formats. And what about updating all XML-documents in order to be consistent with our new class description? What RDF gives us is ability to append new node to graph without changing previously created structure. So we can simply extend our model with new statement about certain object. It is not even necessary to put this extension to the same file where object is. We can create separate file stating that following list of objects have changed and gained new property.

Example 2:

Another use case is more interesting. What if it becomes reasonable to split class definition “A” into two disjoint subclasses “B” and “C” and reassign all the instances of class A to classes B or to class C respectively? Figure 6 illustrates class inheritance before and after splitting.

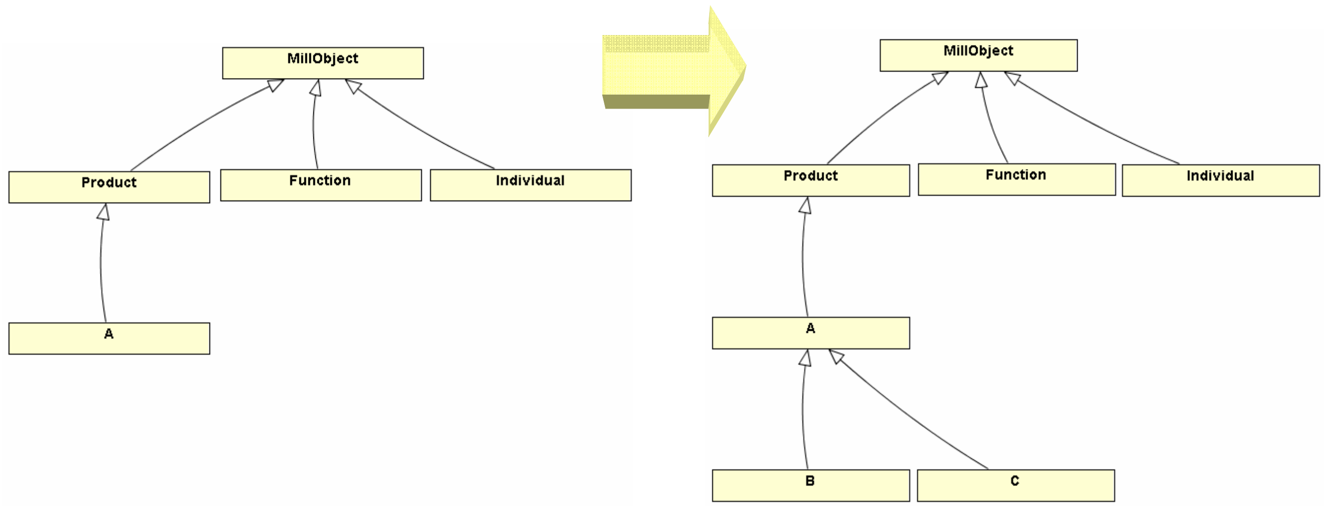

Figure 6. New subclasses in dictionary

In XML-document we have to change all the values of “class” attribute of corresponding objects.

```
<Product designation="MOT:QWE1236" class="A"
   uniqueId="AF08AF09-6400-42d7-A603-7FD352524C96"
   status="Final" statusDate="2004-01-04" owner="Motor Manufacturer Inc.">
```

, becomes

```
<Product designation="MOT:QWE1236" class="B"
   uniqueId="AF08AF09-6400-42d7-A603-7FD352524C96"
   status="Final" statusDate="2004-01-04" owner="Motor Manufacturer Inc.">
```

What can be the consequences? Possible loss of interlinking and possible changes to XSLT again. If XSLT rules rely on class names and implement logic over it, they become useless. What happens to RDF? Again we just add another statement, not necessary to the same file where all description resides. The benefit is obvious – we save previously assigned names and add new ones without loosing the semantics. So, old engines continue working with the same class names and new engines get needed names which are not in contradiction with saved old ones because direct instance of any class is instance of its all upper classes. This is the value brought by ontological modeling and RDF-based serializations.

The significant advantage of RDF lies in dynamics. By providing relation definitions between objects via statements, RDF allows building complex structures, where semantics of relationships is not reflected in a structure of a document. In XML we put subnodes to node and assume some relation. For example:

```
 <PropertySet dictionaryRef="PXIS1001">
        <Name lang="en">General information</Name>
        <Property dictionaryRef="PXIP1001">
           <Name lang="en">Name</Name>
           <Value type="en">Winder</Value>
        </Property>
 </PropertySet>
```

It is implied that tag <Name> belongs to tag <PropertySet>. We can say that this particular PropertySet has Name “General information”, so tag <Name> plays role of “hasName” property or we can say that <Name> belongs to <PropertySet>. RDF uses following construction:

```
<rdf:Statement rdf:about="StatementID">
        <rdf:subject> PropertySet Unique Identifier</ rdf:subject>
        <rdf:predicate>Name</ rdf:predicate>
        <rdf:object>Winder</ rdf:object>
</rdf:Statement>
```

This Statement is independent node and other nodes can refer to it by its StatementID. So if we need to add new property, we just add a new statement. This way of representing relationships allows reaching a new level of dynamics in document structure changes. Relations between objects do not influence document structure and this brings new value to document management features. For example if one XML-document has its copies distributed in several places, then any changes to its structure require changes to all copies in order to be consistent. So, new versions should be substituted for old ones. RDF allows constructing a model from distributed sources, so new changes can be stored either remotely or locally and appended if needed.

Example 3:

Another use case possible within a long-run life-cycle of a supply chain is changes to a model. What if a conventional model will require changes in its structure some day? Let's consider following changes to the model (see Figure 7).

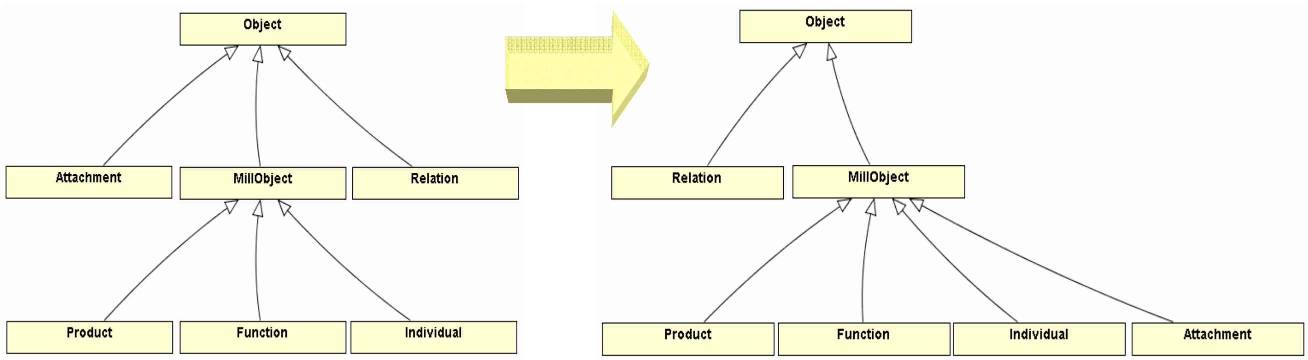

Figure 7. Model change

When "Attachment" object moves to another place in model hierarchy it might obtain new attributes from parent classes and what than happens to XML documents? Again their structure should be changed as in Example 1.

The paradigm of RDF is knowledge-oriented, while XML is document-oriented. Deeper analysis of this difference shows that RDF lifts a model description from the level of a document and structural limitations to more generic level of ontology model, where a document can also be presented as a concept.

4 Architectural Solutions Enabling Collaborative Business Environment

One of the architectural solutions that enable collaborative business environment was developed within a project led by TopQuadrant [TQ] company as set of Ontology-Based Collaborative Environments with supporting tools.

Semantic Collaborative Environment provides facilities to manage highly structured data and conforms to the needs of simple web-based UI's, version control, administration, agreed namespaces and dependencies, changes propagation to all dependent ontologies, provenance management and brings together different content sources from different contexts (integration). The environment architecture is Semantic Collaborative Environment Architecture (see Figure 8).

The Semantic Collaborative Environment Architecture is used to put together Ontology Lifecycle Management Environments, System Engineers Workbench, Mission Planners Workbench, Community of Practice Environments, Enterprise Architects Workbench and ITGovernance Workbench [TQ NASA].

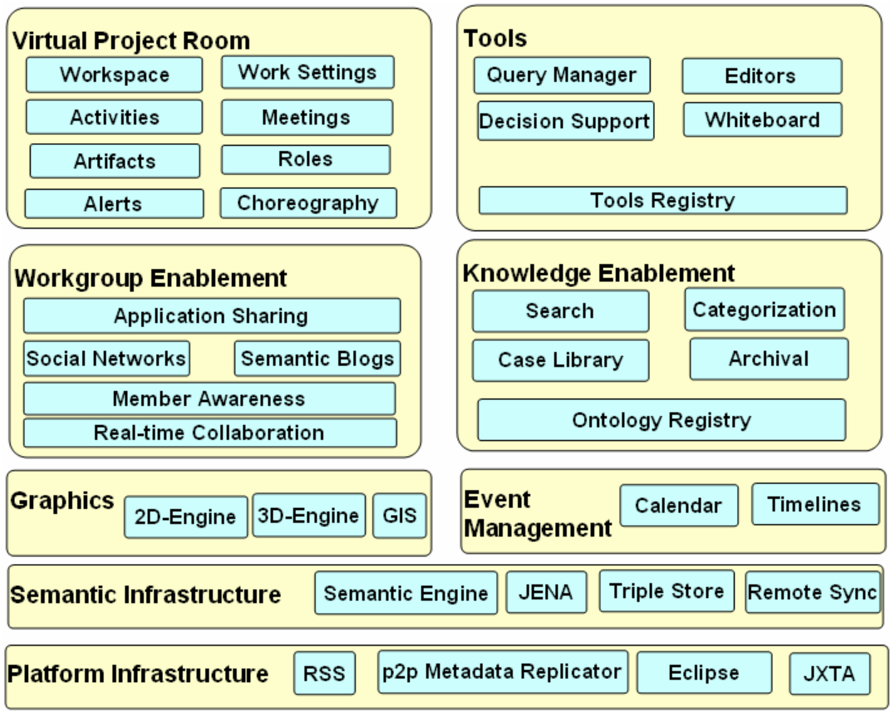

Figure 8. Semantic Collaborative Environment Architecture (adopted from [TQ])

As an architectural solution for the mentioned challenges on Product Data integration, a centralized approach of Mill model server was proposed by PaperIXI. The central server will store documents structured according to the recommended XML schemata and related to whole Paper Mill life-cycle. This solution will allow a single-point way of applying the recommended requirements and will help to perform all automated activities in an organized manner. The activities include efficient communication of changes to all parties who are responsible for delivering data to the changed part of the Mill model, efficient removal of obsolete data and replacing it by the up-to-date data. The use of the XML format with its explicit definitions of data types will allow automated checking of the data validity by existing XML validators. However, a technical validity does not guarantee a real correctness of those data. Hence, a possibility of human inspection upon the data must be also provided in the system. The availability of the server additionally eliminates a need in intermediary parties, which only copy the requirements to the sub-contractors. In this technical solution a repository will collect all engineering data according to the mill model. State of the repository will represent in this case ongoing modernization project, while old outdated data of mill and products is deleted together with technical documentation.

As a relevant solution which targets the needs of alliances in modeling and automation of Business Process execution, Proactivity Stage of the SmartResource project performed by Industrial Ontologies Group[1], can be mentioned. These efforts aim at design and implementation of the OWL-based framework for semantic modeling of automated proactive behavior for software agents that realize business logic for alliance participants. The collaborative environment for an alliance is planned to be based on distributed architecture (Peer-to-Peer) to preserve private preferences of each participant and to support high scalability of the environment. The flexible way that has been chosen for modeling dynamic business logic using OWL assumes control of the agent behavior with help of OWL constraints that are applied to OWL classes, which reflect corresponding members of an alliance.

A clarifying example is shown in Figures 9 and 10. As a first step of semantic modeling of an alliance behavior, representatives of companies, which are potential members of an alliance, must be annotated. Figure 9 shows very simple example of such annotation using Protégé-2000 tool. Such representatives (instances) of ICT companies as Hewlett Packard, Siemens and TietoEnator with corresponding customers have been defined among other companies (Paper_Making and Automation). Samples of restrictions that the alliance applies to the properties of its members, such as allowed partners for

[1] http://www.cs.jyu.fi/ai/OntoGroup: official website of Industrial Ontologies Group.

every member, are shown on Figure 10. The restrictions serialized from Protégé-2000 into OWL/XML are added to a database of the reasoning agents that care about maintenance of policies of an alliance. Due to a flexibility of the OWL-based modeling, dynamics of the alliance (joining/leaving of partners) and complexity of management of alliance policies the proposed approach is anticipated to be very efficient. Please, notice that the simplified examples on the screenshots have nothing to do with real companies and their alliances but they are given just for clarifying the concept of collaborative business ontology.

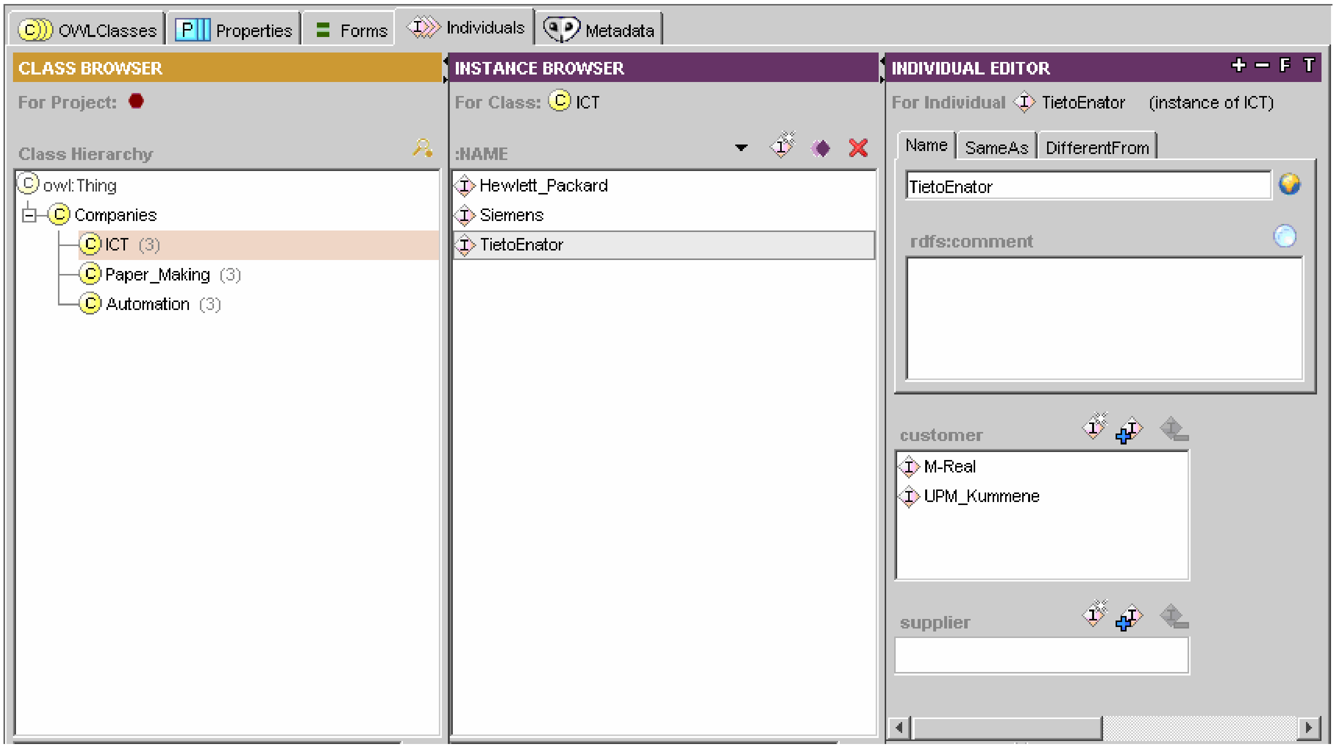

Figure 9. Declaring representatives of ICT companies

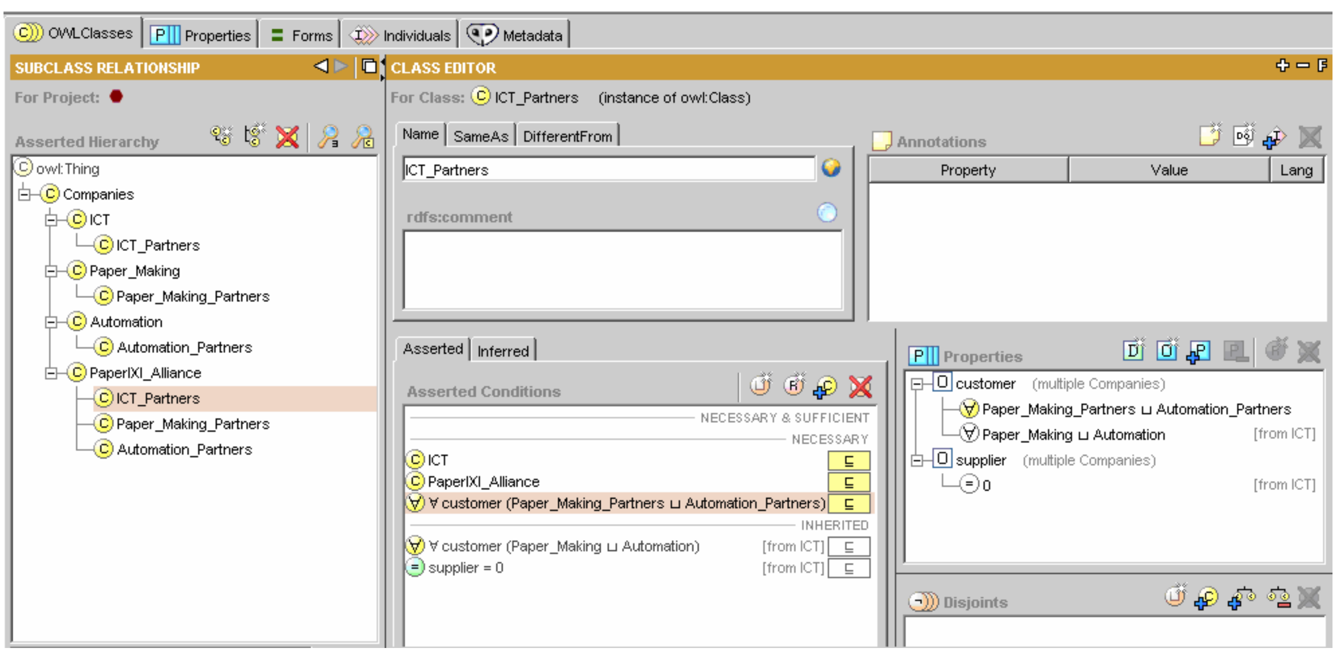

Figure 10. Describing restrictions on ICT companies – alliance partners

Conclusions

Strategic Alliances occur everywhere nowadays as a result of economy globalization. Goals of collaboration differ from alliance to alliance depending on industry, organizational, functional, social and legal aspects. What is the major reason? Generally the reason is that an alliance has characteristics which are not a cumulative sum of characteristics of parts. There are several characteristics of an alliance that are not dependent on grounding for collaboration, but they can be enabled by correct use of existing technologies. Collaborative business intelligence is one of the most desirable characteristics of long-run collaboration.

Whatever goals and grounds of collaboration are, the most demanding requirement is quality of supporting technologies and level of their utilization. Both referred projects in the paper such as PaperIXI and NASA provide results of analysis of integration of supply chain of complex product data life-cycle management. Both approaches enable community knowledge management, collaboration intelligence and learning characteristics. Projects differ in sets of technologies. PaperIXI solutions rely on XML technologies and NASA project relies on Semantic Web technologies.

Another sample of attempts to build global platforms for business process integration, which is related to the Industrial Maintenance domain, can be PROTEUS [Hausladen&Bechheim, Proteus, Thron et. al.] and SmartResource [Kaikova et. al., SR, Terziyan 2005a, Terziyan 2005b] European initiatives. The first one utilizes XML standard as a basis for integration, the second one relies on ontological modeling. The choice of Semantic Technology in the SmartResource project allowed reaching comparable results with considerably smaller resources.

XML based technologies are proven to be successful in efficient data exchange and integration solutions by wide spread and utilization. Obviously XML is good candidate, but not the best. The paper shows that XML is good while if to consider a domain to be static, but cases of complex assets with long life-cycle dynamic nature of knowledge change appropriateness of the XML format.

We state that semantic web standards and technologies are good option to enable formal management of domain knowledge in long run for efficient collaboration. These technologies support business intelligence solution in a long-run scope, because of their purpose to provide a framework for knowledge management in ontological way. They allow lifting of document centric product data management to knowledge centric. Underlying graph model of semantic web standards is more generic and can describe both tree structures of XML and relational schemas of databases [Mazzocchi]. Definitely OWL is the number one candidate for technical dictionary and classification systems, but having RDF/XML as a format for data gives consolidated solution for creation of a collaborative platform and tools. OWL can easily incorporate new learnt knowledge without changes to semantic based tools, whereas changes in a model cause unwanted technical consequences in a case with using the XML schema.

We tried to show in the paper that Semantic Web standards solve challenges of product data management more elegantly providing more opportunities for business intelligence tools. But XML is a much more mature technology than Semantic Web in terms of size of user community, availability of support tools, and viability of business models relying on the technology. Therefore, Semantic Web standards can be adopted in situations where the capability to represent semantics is important enough to overcome XML's maturity advantages. Nature of collaboration in paper industry around a paper mill with a life-cycle of nearly fifty years definitely requires a solution that has been initially designed to support knowledge management in changing domains.

Acknowledgements

This research has been performed as a part of the SmartResource ("Proactive Self-Maintained Resources in Semantic Web") project in Agora Center (University of Jyväskylä, Finland) and funded by TEKES and industrial consortium of following companies: Metso Automation, TeliaSonera, TietoEnator and Science Park. The authors are also grateful to Mr. Jouni Pyötsiä and Mr. Arto Marttinen from Metso Automation for useful consulting, PaperIXI project group and TopQuadrant for publishing useful materials and analytical information on the Web.

References

[DIP] Data, Information, and Process Integration with Semantic Web Services, http://dip.semanticweb.org/

[DUNS] Data Universal Numbering System (DUNS), http://www.rurdev.usda.gov/rbs/farmbill/duns.htm.

[ECS] Engineering of Complex Systems Group, NASA, http://ecs.arc.nasa.gov/

[Guarino95] Guarino N. & Giaretta P., (1995), Ontologies and Knowledge Bases: Towards a Terminological Clarification, in Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing, N. Mars (ed.), IOS Press, Amsterdam, pp 25 – 32.

[Guarino98] Guarino N., (1998), Formal Ontologies and Information Systems, in Guarino N. (ed.), Proc. of FOIS98, IOS Press, pp. 3 – 15.

[Hausladen&Bechheim] Hausladen I., Bechheim C., *E-Maintenance Platform as a Basis for Business Process Integration*, INDIN04, 2nd IEEE International Conference on Industrial Informatics, Berlin, Germany, pp 46-51, 24-26/06/2004.

[IAC] Industry Advisory Council, Interoperability Strategy: Concepts, Challenges, and Recommendations, Available in: http://xml.coverpages.org/bcm.html

[IANA] Internet Assigned Numbers Authority (IANA) for Private Enterprise Number, http://www.iana.org/cgi-bin/enterprise.pl.

[ISO-3166] ISO3166 Standard, Country Codes, http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html.

[ISO-639] ISO639 Standard, Language Codes, http://www.w3.org/WAI/ER/IG/ert/iso639.htm.

[Kaikova et. al.] Kaikova H., Khriyenko O., Kononenko O., Terziyan V., Zharko A., Proactive Self-Maintained Resources in Semantic Web, *Eastern-European Journal of Enterprise Technologies,* Vol. 2, No. 1, 2004, ISSN: 1729-3774, pp. 4-16.

[Mazzocchi] Mazzocchi S., It's All About Graphs, White Paper, February 11, 2004, available at: http://www.betaversion.org/~stefano/linotype/news/46/, last accessed 17th April 2005.

[Meersman] Meersman R., (1999), The Use of Lexicons and Other Computer-Linguistic Tools in Zhang Y., Rusinkiewicz M, & Kambayashi Y. (eds.), Semantics, Design and Cooperation of Database Systems, in The International Symposium on Cooperative Database Systems for Advanced Applications (CODAS 99), Springer Verlag, Heidelberg, pp. 1 – 14.

[MIME] Multipurpose Internet Mail Extensions (MIME), http://www.ietf.org/rfc/rfc2045.txt.

[OWL] Web Ontology Language, http://www.w3.org/2004/OWL/

[OWLS] Ontology Web Language for Web Services (OWL-S) 1.0, http://www.daml.org/services/owl-s/1.0/.

[OWL Req] OWL Web Ontology Language Use Cases and Requirements, http://www.w3.org/TR/webont-req/

[PaperIXI] PaperIXI project , http://pim.vtt.fi/paperixi/.

[Price] D. Price, R. Bodington, Applying Semantic Web Technology to the Life Cycle Support of Complex Engineering Assets, In: S.A. McIlraith et al. (Eds.), Proceedings of the ISWC 2004, LNCS, vol. 3298, Springer, 2004, pp. 812-822.

[Proteus] PROTEUS – European ITEA project on a generic e-Maintenance platform, http://www.proteus-iteaproject.com/.

[PSK] PSK5941 Standard, Purchase of machinery and equipment for industry. Interchange of electronic documentation. Machine and equipment data. http://www.psk-standardisointi.fi/.

[RDF] Resource Description Framework, http://www.w3.org/RDF/

[RNTD] RosettaNet Technical Dictionary (RNTD) format, http://www.rosettanet.org/.

[SemanticWeb] Semantic Web, http://www.w3.org/2001/sw/.

[SR] SmartResource – Tekes project on Proactive Self-Maintained Resources in Semantic Web, http://www.cs.jyu.fi/ai/OntoGroup/SmartResource_details.htm.

[Spyns et al] Spyns P., Meersman R. & Jarrar M., Data modelling versus Ontology engineering. SIGMOD Record: Special Issue on Semantic Web and Data Management, 31(4): 12 - 17, December 2002.

[SSG] SSG Product Database (SKC), http://www.ssg.se/Eng/Foretagsinfo/skc-databasen_ny.asp.

[Terziyan, 2005a] Terziyan V., Semantic Web Services for Smart Devices Based on Mobile Agents, In: *International Journal of Intelligent Information Technologies*, Vol. 1, No. 2, April-June 2005, Idea Group, ISSN 1548-3657, pp. 43-55.

[Terziyan, 2005b] Terziyan V., SmartResource: Utilizing Semantic Web Services to Monitor Industrial Devices, In: *Proceedings of the XML Finland 2005: XML - the Enabling Technology for Integrating Business Processes*, March 8-9, 2005, Pori, Finland.

[Thron et. al.] Thron M., Bangemann T., Schmidt D., *Middleware for a Global Maintenance Platform,* 2004 IEEE International Workshop on Factory Communication Systems, Sept. 22-24, 2004, WFCS 2004, Proceedings pp. 381-384.

[TQ] TopQuadrant: Ontology Development and Solution Envisioning for Semantic Web Applications, http://www.topquadrant.com/

[TQNASA] "Semantic Technology and Model-Based Life-Cycle Support for NASA Space Vehicle Engineering", Ralph Hodgson, March 2005, http://www.topquadrant.com/documents/talks/TQ Semantic Technology for Model-Based Life-Cycle Support (v4).pps

[UNECE] United Nations Economic Commission for Europe (UNECE), Units of Measure Used in International Trade. http://www.unece.org/.

[Ushold] Ushold M. & King M., (1995), Towards a Methodology for Building Ontologies, in Proc. of the Workshop on Basic Ontological Issues in Knowledge Sharing (IJCAI95 workshop).

[WSMO] Web Services Modeling Ontology (WSMO), http://www.wsmo.org/.