ISSS/WS-EC-ECIMF/03/0xx

ISSS/WS-EC/03/xxx

CEN/ISSS Electronic Commerce Workshop

E-Commerce Integration Meta-Framework

CEN Workshop Agreement xxxxx:2003

Final draft
February 2003

ECIMF Project Group

1 Introduction

1.1 Background and Goal Statement

There have been many standardization activities in the area of e-commerce communication. The standard bodies and industry groups in multi-national levels have been promoting several standards. Some of these, with long-standing tradition (like EDI variants), have gained significant acceptance, especially among large industry players. However, these standards are often criticized for their complexity, high implementation cost, multitude of local variants, and extensive demand for expertise knowledge. Other frameworks for electronic commerce, defined more recently in the Internet age, try to avoid those mistakes, and they also have seen some acceptance in selected industry sectors (RosettaNet, OAG, cXML, xCBL, upcoming ebXML …).

However, the proliferation of mutually incompatible standards and models for conducting e-commerce resulted in even more increased demand for interoperability and expert knowledge. So, overall, the isolated efforts of industry groups and standard bodies created quite the adverse effect from what was intended, when it comes to wide acceptance of electronic commerce, especially in the SME market.

These issues slow down the spreading of e-commerce applications, and for this reason the industry is looking for methods to meet the exploding demand in the "new economy" to offer increased QoS, reduction of manual labor and cost, and to meet the requirements of nearly real-time reaction to changing market demands. At the same time the industry is aware that existing e-commerce frameworks require costly adjustments in order to fit their business model to that of specific frameworks, with the perspective that similar costs will follow if the business player wants to participate in other frameworks as well.

1.1.1 E-Commerce Integration Meta-Framework scope

In response to these concerns from the industry, this CEN/ISSS project within Workshop for Electronic Commerce proposes the E-Commerce Integration Meta-Framework (ECIMF):

> ***A meta-framework, which offers a methodology, a modeling language and prototype tools for all e-commerce users to achieve secure interoperability of the service regardless of system platforms and without major adjustments of existing systems.***

The most important characteristic of this project is to present a common approach to enable interoperability without enforcing major changes to the existing infrastructure. This is in contrast with many other widely promoted approaches to interoperability, which require from partners to be strictly conformant to a common standard in order to participate in e-commerce.

There are strong reasons for preferring the "enable" instead of (commonly endorsed) "enforce" approach:

- Business partners may have already made significant investments in building interfaces conforming to some standard(s).
- Commonly used integration methodologies are focused on data translation, which results in complex and inflexible solutions. Changing such integration solutions to accommodate new standards is often infeasible.
- There will always be legacy systems that need to be integrated with the "standard of the year" external interfaces. It is simply not realistic to hope that at some point in time all systems will adopt and fully conform to one common standard for every aspect of business communication.

For these reasons, the interoperability-enabling methodologies, such as the ECIMF approach, will play an increasingly vital role in the e-business communication.

The meta-framework, which the project aims to deliver, is understood as a combination of methodology, modeling notation (meta-models) and guidelines for aligning different aspects of e-commerce – hence the name “meta-framework”, because using these artifacts the users will be able to build concrete integration frameworks.

The main purpose of this meta-framework is to facilitate the interoperability by mapping the concepts and contexts between different existing e-commerce frameworks, across multiple architectural layers. An important premise for this project proposal is the following definition of interoperability:

> ***The interoperability, as seen from the business point of view, takes place when the business effects for the two involved enterprises are the same as if each of them conducted a given business process with a partner using the same e-commerce framework.***

As a consequence of this premise, the project proposes using a top-down approach to the comparative analysis of the e-commerce frameworks, which starts from the business context level. The project also reuses the experiences of other projects in the area of enterprise analysis and modeling.

The approach presented here also addresses integration of internal business processes and applications with external e-commerce interfaces required to conduct business electronically, whichever standard they conform to. This is just a special case of interoperability between differing frameworks. However, this case is crucial for companies in adoption of any e-commerce standard.

1.1.2 Benefits

The development and adoption of the ECIMF standard should benefit especially the following groups:

- SME market:
  The small companies no longer will be forced to restructure at all costs their internal systems in order to conform to whatever framework their bigger partners have. The interoperability bridges that conform to ECIMF will allow them to do it gradually, based on the economic principles, while at the same time allowing them to participate in the e-commerce. This should result in more SME-s joining the e-market, even though their internal economy systems may not yet follow any standard e-commerce framework.

- System integrators:
  The system integrators will be able to use a consistent methodology, and a precise framework for defining the integration bridges. The results of their work can be implemented on various conforming platforms, no longer locking them (and their customers) into a single proprietary tool. The overall cost for the implementing the integration solution, its maintenance and amount of manual labor will be reduced.

- Software vendors:
  The software vendors will be able to offer competitive integration products that conform to the standard framework. This means that their products will be more attractive to the customers, who are more likely to choose a solution that guarantees them certain level of independence. At the same time though, the conformance to ECIMF should allow software vendors to offer clearly understood added values, which are now often misunderstood because of the difficulty in comparing proprietary methodologies.

1.1.3 Relationship to various global e-commerce frameworks

The aim of the ECIMF project is not to propose yet another e-commerce framework. We recognize the efforts of various standardization bodies and industry groups to provide global solutions in this area (e.g. ebXML[1], RosettaNet, xCBL, OAGIS framework, Hewlett-Packard’s e-Speak[2], Microsoft’s BizTalk[3]), as well as other projects offering tailored solutions for specific market or industry sector.

The ECIMF project does not compete with any of these frameworks. We welcome and look forward to cooperate with their representatives in order to enhance the results of this project. The need that the ECIMF wants to address is the interoperability between these frameworks, especially for the transitory periods in

[1] The ebXML project, http://www.ebxml.org/specdrafts/.

[2] The e-Speak framework, Hewlett-Packard, both as a commercial product http://www.e-speak.hp.com, and an OpenSource free Java implementation of the complete framework at http://www.e-speak.net .

[3] The BizTalk framework, Microsoft, http://www.microsoft.com/biztalk/techinfo/BizTalkFramework20.doc , BizTalk repository at http://www.biztalk.org, and the commercial product BizTalk Server http://www.microsoft.com/biztalk, which additionally contains the mapping and orchestration tools.

SME environment (economic and manpower limitations), which are required for adoption of any of the frameworks.

In our opinion at least two factors will continue to adversely affect the wide-spread adoption of e-commerce: one is the fact that quite a few businesses already made commitments to some of the existing frameworks, in terms of internal expertise, investments, partnerships, and adjustments to the technology and models for business interaction imposed by these frameworks. This situation is combined with the current approach to system integration, which very often locks up the companies to specific system integrator and specific proprietary solutions.

The other limiting factor is that extensive knowledge and experience is still required to adequately understand the differences between the frameworks, and even more to implement some level of interoperability – both between the e-commerce frameworks themselves, and between legacy systems and any given framework. Also, though more and more modern frameworks use UML and UMM to describe parts of their models, there is no general meta-framework that would allow implementing interoperability in a structured way, not to mention the fact that many frameworks are defined using imprecise, natural language descriptions.

It's worth noting a fact that is often overlooked: the differences between e-commerce frameworks are much deeper than just differences in their protocols, scenarios and data formats. There is a need for a unified methodology to compare and align also the semantics of central concepts in order to properly understand these differences.

The development of the ECIMF standard builds on the experiences from projects such as:

- ebXML: specifically Business Process Specification Schema (ebBPSS), Collaboration Protocol Profiles and Agreements (ebCCP),
- UN/CEFACT Unified Modeling Methodology (TMWG-N090),
- RosettaNet Implementation Framework v. 2.0 [4] (RNIF2.0),
- BizTalk 2.0 framework [3] (and BizTalk Server commercial tools),
- OAG Integration Specification (OAGIS 7.1),
- OMG's Model Driven Architecture (MDA),
- eCo framework [5]

and others, in order to provide a sufficiently broad and general model for alignment between the frameworks.

Consequently, we see the ECIMF project as a complementary and necessary part of e-commerce adoption, reducing the cost and amount of labor required to adopt any e-commerce framework.

1.2 Project Details

The following list shortly describes the scope for the ECIMF definitions:

[4] RosettaNet, http://www.rosettanet.org.

[5] The eCo Framework, CommerceOne, http://www.commerce.net/eco .

- **Meta-framework modeling methodology** – an approach to model the interactions and transformations required for mapping between different e-commerce frameworks:
    - Top-down analysis, based on the business process integration
    - Multi-layered modeling approach
    - Calibration of concepts within corresponding contexts (semantic translation)

  This part of the project requires close collaboration with the experts in order to reuse as much as possible the experiences collected by groups like ebXML, RosettaNet, OAG, EDI community and others.

  This part of the documentation is contained in section 2 of this document.

- **Meta-framework modeling language** – a precise notation to describe the concepts of e-commerce frameworks, the contexts in which they occur and interact, and the required transformations between them:
    - Business context correspondence (compatibility of economic goals)
    - Semantics of the base building b locks (actors, messages, transactions), data models
    - Scenarios for message exchange (business processes)
    - Access to external resources (URLs, directories, catalogues, databases, etc…)
    - Messaging models
    - Security models and services, as far as they affect the business process and interoperability on the technical level
    - Transport protocols
    - etc.

  For the business process modeling we suggest substantial reuse of the results of ebXML BP work (cf. ebBPSS), with additions of the modeling notation and language to express the transformations between the business processes on different layers.

  This part of the documentation hasn't been developed, since the previous part – methodology, which provides the basis for notation – hasn't been completed.

- **Proof of Concept** – the project will aim to provide a Proof of Concept implementation of the tool-chain needed for realization of the proposed methodology, demonstrating the interoperability between some concrete e-commerce frameworks. The tools developed by the project will be published under Open Source license, freely available for both private and commercial use.

  This part of the documentation is contained in the Appendix ??? of this document. Additionally, the Open Source application module supporting Semantic Translation with labeling is available on the project's dedicated website.

1.3 Original Project Deliverables and Timescales

The timeframe for this project was set up to be 18 months, in the period of June 2001 – December 2002. The manpower allocated on permanent basis to this project was initially planned as follows (expressed in percentage of time involvement times number of people):

- WebGiro: 50% x 1 person
- KTH: 25% x 2 persons

Furthermore, the list below presents prospective manpower that was likely to be involved on a regular basis:

- KTH: 25% x 1 person
- HP: 50% x 1 person
- Microsoft: 50% x 1 person

Unfortunately, in the course of running the project these resources have never been fully realized, which resulted in parts of this CWA being incomplete or missing.

Assuming the above resources, the originally planned deliverables consisted of the following separate documents (which later have been merged into one CWA):

- **General ECIMF methodology (ECIMF-GM):**
  A document (CWA) describing in detail the multi-layered approach, and the specification of the ECIMF methodology. This part should result from the discussions on the general methodology on how to approach the business process integration. The intention is to keep this part vendor- and tool-independent.

  This document, originally intended as a description of formalized methodology, due to the time and resource constraints was put in a form of general guidelines.

- **ECIMF technical specification (ECIMF-TS):**
  A document (CWA) containing the formal technical specification for modeling notation constructs, and the serialized form for the models (i.e. the ECIML and the MANIFEST specifications).

  This specification hasn't been developed, as explained above.

- **The Proof of Concept implementation (ECIMF-POC):**
  It would include the tools to support the methodology – the ECIMF Navigator for conceptual navigation and calibration, integrated with a ManifestFactory implementation in order to produce the MANIFEST recipes based on the model. It would also contain a Proof of Concept mapping of two business processes from different frameworks. This part should include additional examples of mapping, depending on the contributed resources.

  If the timeframe and the resources available are sufficient, a basic ECIML-compliant agent implementation should be created to support the Proof of Concept mapping.

The following milestones were planned for delivering the results:

1.3.1 Initial Proof of Concept (POC) for the approach

Deliverables:

- Reformulate and elaborate on the FAM CWA material in order to show how Conzilla tool can provide structured and contextualized added value to a textual description.
- Provide an initial description of the methodology for comparing the e-commerce frameworks (this will form the draft of ECIMF-GM document).
- Prepare a simple example of mapping the differences between two e-commerce frameworks (e.g. BizTalk and e-Speak), using the proposed approach.

Timescale: 12 June 2001 (Oslo meeting)

Status: delivered, available as a set of PowerPoint slides.

1.3.2 Initial ECIMF specification and basic integration with tools

Deliverables:

- Initial version of the ECIMF-GM and ECIMF-TS documents, and models of a concrete business process in two selected e-commerce frameworks.
- Customization of the Conzilla tool to support the modeling notation introduced in ECIMF-GM.

Timescale: mid-October 2001

Status: partially delivered ECIMF-GM. Initial models in Conzilla.

1.3.3 Refined ECIMF specifications and extended tool-chain

Deliverables:

- Refinement of the ECIMF specifications based on further comparative modeling of the selected frameworks.
- Extended support for the process in the tool-chain: integration of Conzilla, scripting language and the ECIML code generation to form the ECIMF Navigator tool.

Timescale: 1Q2002

Status: partially delivered. Extended ECIMF-GM and Proof-of-Concept documentation. However, Conzilla support lagging behind.

1.3.4 Further refinements to ECIMF specifications, and a reference ECIML-compliant agent implementation

Deliverables:

- More refined ECIMF specifications, and additions to the tool-chain to support the specification.
- Depending on the support from industry partners, a basic reference implementation of the ECIML-compliant server.

Timescale: 4Q2002

Status: partially delivered. ECIMF-POC documentation completed, but the toolkit only supports basic semantic translation support (Conzilla was replaced with Protégé).

1.4 External Liaisons

The project team coordinated its activities with the following projects:

- Other relevant CEN/ISSS/EC-WS projects
- ebTWG,

- RosettaNet,
- Open Applications Group,
- ISO TC/154 Basic Semantic Register

2 General Methodology[6]

2.1 Overview

The ECIMF project deliverables consist of a recommended methodology, presented in this document, and base tools needed to prepare specific comparisons of concrete frameworks (presented in the section 3 of this document, where you can also find the case studies).

The results of following the ECIMF methodology should be clear implementation guidelines for system integrators and software vendors on how to ensure interoperability and semantic alignment between incompatible e-commerce systems. This generic integration rules should be expressed in an implementation-independent language, providing mapping and transformation descriptions/recipes that can be implemented by ECIMF-compliant agents/intermediaries. This ultimately should allow the e-commerce frameworks to interoperate without extensive manual alignment by the framework experts, and will make the integration logic more understandable and maintainable.

2.1.1 Layered approach

The proposed methodology for analysis and modeling of the transformations between the e-commerce frameworks follows a layered approach.

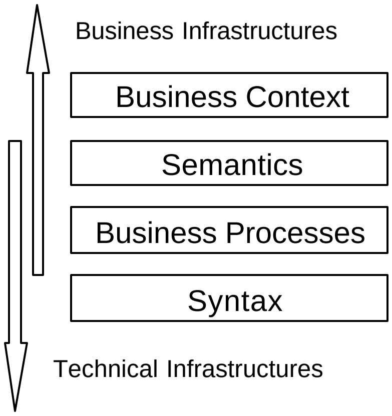


Figure 1 ECIMF layers of integration

This approach means that in order to analyze the problem domain one has to split it into layers of abstraction, applying top-down technique to classify the entities and their mutual relationships:

- First, to establish the scope of the integration task in terms of a **business context** – based on the economic aspects of the partners' interactions,
- Then, to identify the top-level entities and the contexts in which they occur (the data model), and how these contexts affect the **semantic** properties of the concepts,

[6] Editor's note: Originally this section formed a separate document. There may be still some inconsistencies related to this fact.

- Then, to proceed to the next layer in which the interactions (conversation patterns, **business processes**) between the partners are analyzed.
- Then, to go to the lowest, the most detailed level to analyze the messages and data elements **(syntactic level**) in communication between the partners.

Starting from the top-most level, the contexts in which the interactions occur are analyzed and collected, and these contexts affect the semantics of the interactions occurring at the lower layers.

The second dimension of the proposed approach conforms to the Meta-Model Architectures, as described in the MOF standard, introducing the meta-model, model and instance (data) layers. This means that ECIMF will be used to define:

- The modeling notation: a set of modeling concepts with their graphical and XML representation to model the transformations[7],
- The models: concrete transformations between concrete frameworks
- And the model instances of transformations, as realized by an ECIMF-compliant runtime.

Figures 1 and 2 present the ECIMF layers, and how they are applied to define the interoperability model between two incompatible frameworks.

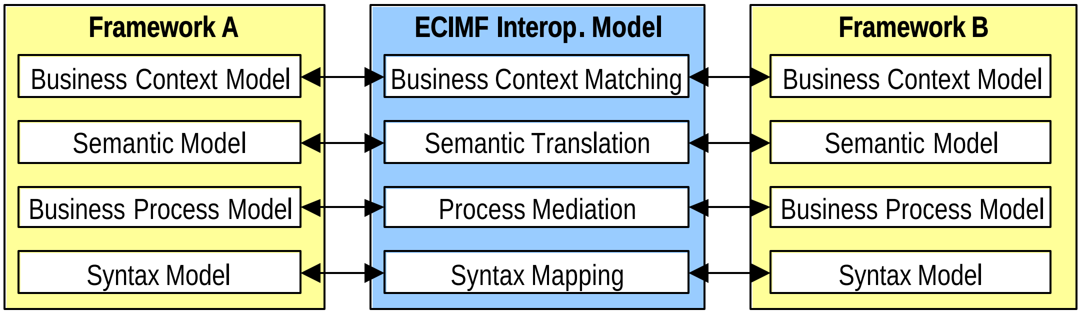


Figure 2 ECIMF methodology – interoperability layers.

Each of these layers is described in detail in the section 2.

2.1.2 Conceptual navigation – ECIMF Navigator

In order to navigate through the framework models and concepts, during the initial stages of the project a prototype tool named Conzilla was introduced, which in later stages of the project was to be augmented with other modules (like data format translating software, automatic generation of interfacing state machines, routing and packaging translators, etc). This extended toolset is called ECIMF Navigator, and its intended use is presented on the Figure 2.

[7] Since the modeling elements regard multiple layers of the ECIMF approach, hence the name "meta-framework", because they will be used to define interoperability frameworks.

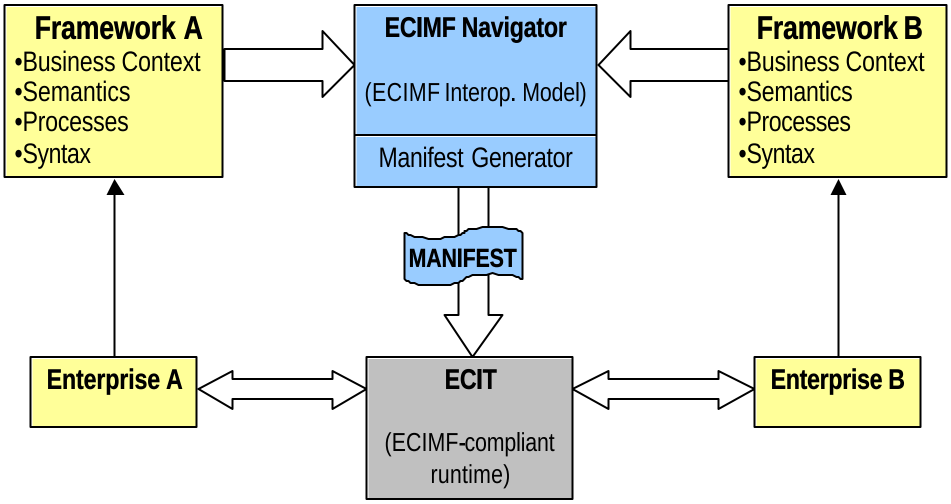


Figure 3 The ECIMF concept of frameworks transformation and alignment.

The ECIMF project used an extension of Conzilla (see http://ww.conzilla.org for more information about the Conzilla project) as a prototype tool for browsing and comparing different e-commerce framework models. One of the goals of the ECIMF project was to extend this tool by necessary backend(s) for producing abstract machine-readable interoperability guides (MANIFEST recipes), expressed in ECIML language.

In later stages, after some limited development and evaluation of future possibilities of the Conzilla platform, the ECIMF project switched to using a well-known knowledge engineering environment Protégé (http://protege.stanford.edu), as it seemed to better match the requirements for extensibility, wider acceptance and sustained maintenance. Concequently, the support for parts of ECIMF methodology has been implemented as Protégé module (so called "tab").

2.1.3 Top-down, iterative process

The ECIMF uses a classic top-down approach for solving the interoperability issues, but combined with an iterative process of refining the higher level models based on the additional information gathered in the process of modeling the lower levels.

This process is described in detail in the Framework Integration Guidelines section.

2.1.4 The modeling notation

The ECIMF project proposes to use an extended UML modeling notation (a UML profile) to express relationships between the semantics and models of the e-commerce frameworks. This E-Commerce Integration Modeling Language ("ECIML"), to be defined as a result of the project, will be a concrete instance of the OMG's MOF meta-meta-model, at the same time re-using as many concepts from standard UML as possible. This puts it in the following relationship to the standard modeling approaches:

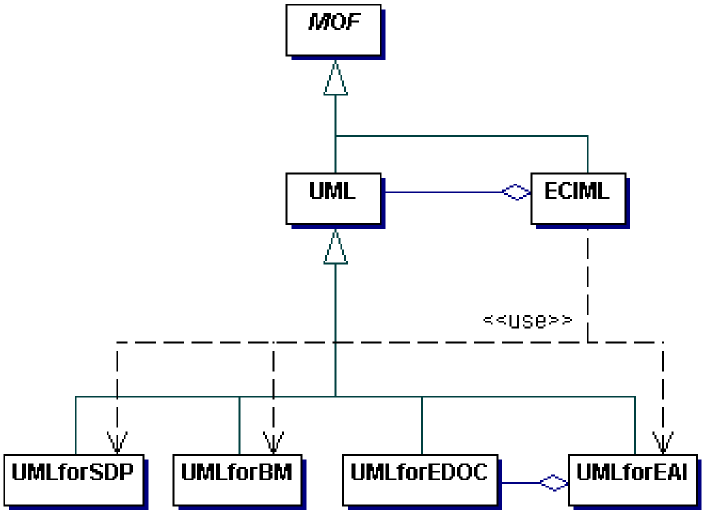


Figure 4 Relationship between the ECIML and other modeling standards.

In other words, the ECIML will be yet another profile of UML 1.4. We will build on the experiences of the projects like pUML (The Precise UML Group), using also the OMG's standards (e.g. CWM, standard UML 1.4 profiles, UML Profile for EAI and UML Profile for EDOC) when appropriate, in order to define a suitable meta-model. We will also reuse as much as possible the specialized concepts developed by the UN/CEFACT Unified Modeling Methodology (UMM), as described in TMWG-N090R10.

One could use the standard UML for modeling the interoperability concepts, but we feel that in its current form it is too generic and lacks necessary precision, and though it's extensible, the way the extensions are specified is often implicit (e.g. stereotyping). In the ECIML meta-model these concepts would be precisely defined. Some of these issues will be addressed in the next major revision of UML standard (2.0), at which point we will evaluate the possibility to use that standard as the sole basis for ECIML.

Consequently, one of the original goals of this project was to define a suitable set of modeling constructs to more adequately address the needs of meta-framework modeling and transformations. However, due to limited resources this part of the project has not been completed.

2.2 Methodology

As mentioned in the overview section, the ECIMF methodology addresses the following four layers of interoperability:

- **Business Context Matching:** this aspect deals with setting up the scope of the integration task – we assume that preparing a complete integration specification for all possible interactions might not be feasible (even if it were possible at all), so the task needs to be limited to the scope needed for solving a concrete business case. This case is identified, the models for each party are prepared, and then it needs to be determined if they match, i.e. if the business partners try to achieve the same business goals.

- **Semantic Translation:** in this step the key concepts and their semantic correspondence is established, so that they can be appropriately transformed whenever they occur in contexts of each of the frameworks (which is also known as "semantic calibration" [CID52]).
- **Business Process Mediation:** in this step the necessary mediation logic is defined, by introducing an intermediary agent that can transform conversation flow from one framework to that of the other, while preserving the business semantics (e.g. the transaction and legal boundaries).
- **Syntax mapping:** in this step the mapping between data elements in messages is defined, based on the already established semantic correspondence and translation rules defined in the first step. Also, the transport protocol and packaging translation is specified.

The following sections describe in detail each of these areas of interoperability.

2.2.1 Business Context Matching

2.2.1.1 Business Context – definition and role

- IT infrastructure exists to support business goals: IT systems don't exist in a void, but they play specific roles in the business.
- Business context is therefore crucial: information is useful only when considered in the right business context. It is the business context that ultimately determines the meaning of data and information exchange.
- Business flow should therefore be considered before technical flow.
- REA modeling framework can be successfully used as the underlying meta-model

Business Context is a collection of:

- ***Agreements / Contracts defining the Commitments***
- ***Collaboration Patterns (using Business Processes) to execute commitments***
- ***Business Objects with their semantics, lifecycle and state, which encapsulate business data and business rules***

2.2.1.2 Resource-Event-Agent modeling framework

REA Enterprise Ontology has been created by William E. McCarthy, mainly for modeling of accounting systems. However, it proved so useful and intuitive for better understanding of business processes that it became one of the major modeling frameworks for both traditional enterprises and e-commerce systems. Recently, it has been extended to provide concepts useful for understanding the processing aspects (processes, recipes) in addition to the economic aspects (economic exchanges). Please see http://www.msu.edu/user/mccarth4/ for more information.

Some of the REA concepts have been used to model the Business Requirements in UN/CEFACT Modeling Methodology ("UMM", formally known as TMWG N090), and the Business Process Analysis Worksheets in ebXML, and it's use is currently a subject of further study in the Business Collaboration Patterns and Monitored Commitments team of the E-Business Transitionary Working Group (eBTWG) - the successor to ebXML.

2.2.1.2.1 Economic exchange as a central concept

- REA ontology focuses on the idea of economic exchange of resources as the basis of business and trading. In REA models, economic agents exchange economic resources in series of events, which fulfill mutual obligations (called Commitments), as specified in an Agreement between the business partners. See also the detailed definitions in the ECIMF-TS document.
- Economic exchange models define collaborations between partners involved in the process, and these collaborations naturally map to business document exchanges (both in paper and in electronic form).

2.2.1.2.2 Value-chain models (REA Enterprise Scripts)

- REA process diagrams show the high-level flows of economic resources in the enterprise, related to the economic events and collaborations between the agents involved in the exchanges. They are sometimes referred to as value-chain diagrams.
- The resource flows between processes in the value-chain diagrams represent the collective unbalanced stock-flows, consumed and produced by the events belonging to given processes.
- Value-chain model (also known as *REA Enterprise Script*) is a series of processes, consisting of exchanges, where collaborations between agents are realized with recipes (groups of ordered tasks).

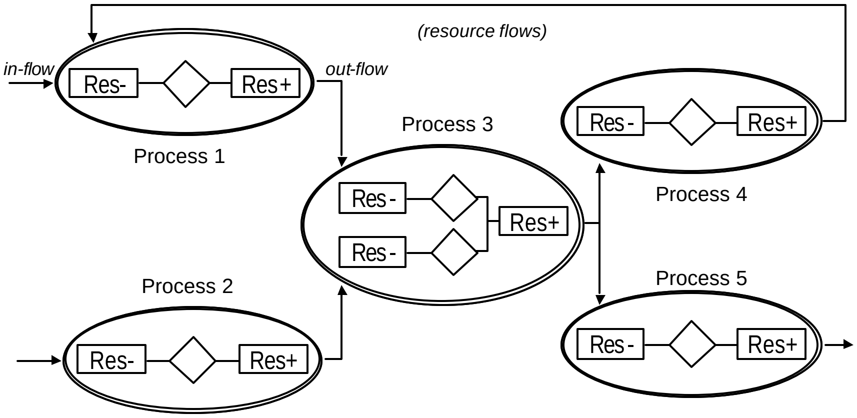


Figure 5 Enterprise value-chain, seen as series of exchanges.

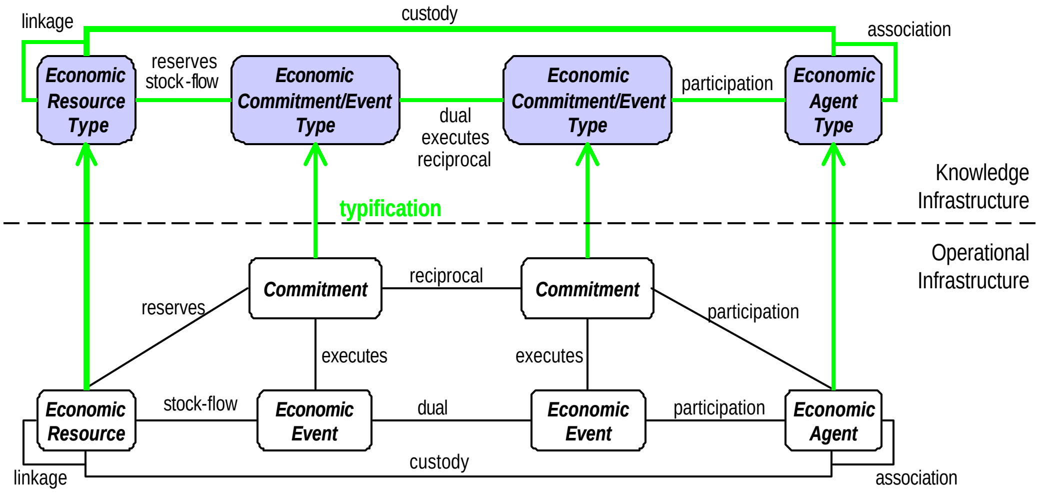


Figure 6 REA meta-model of economic exchanges (simplified).

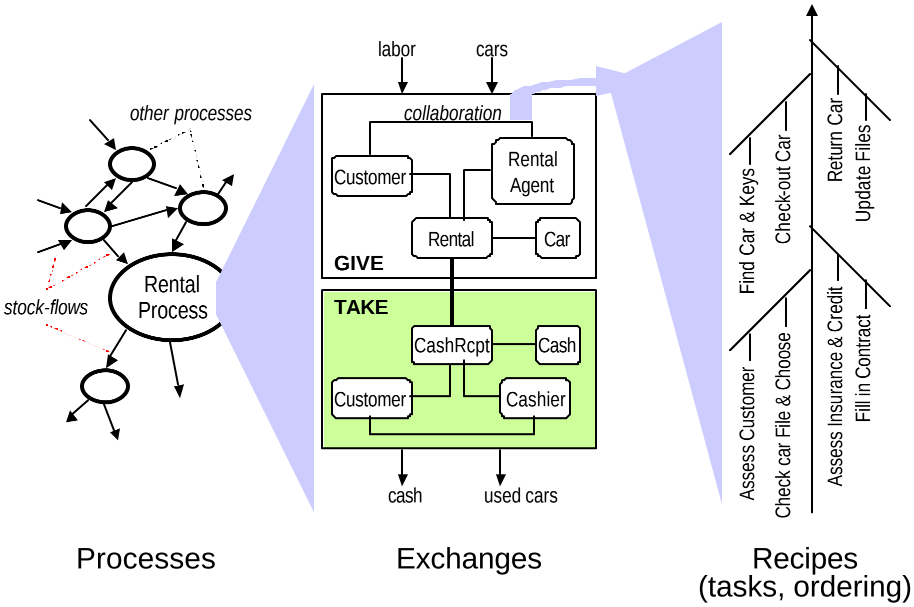


Figure 7 Overview of the processes, exchanges and recipes.

You will find the detailed description of this meta-model in the ECIMF technical specification document (ECIMF -TS).

2.2.1.3 Business Context Matching rules

2.2.1.3.1 Rationale

- Traditional trading partners' agreements: both partners need to agree on:
  - o The type of resources exchanged
  - o The timing (event sequences/dependencies)
  - o The persons/organizations/roles involved

  Also, each of the partners needs to follow the commitments under legal consequences

- Conclusion: in the traditional business, partners achieve common understanding through negotiations, and their results and conditions are then recorded in a formal written contract. In electronic business some standards support creation of electronic TPA's (Trading Partner Agreements). Their formation is a special case of establishing the Business Context Matching described here.

2.2.1.3.2 Matching Rules

Business partners involved in an integration scenario need to consider first whether their business goals and expectations match, before they start solving the technical infrastructure problems. For that purpose, they can create two (or more) business context models, one for each party involved in the integration scenario. The interoperability of the e-commerce scenario, as implemented by two different partners, requires that these models match.

There are several requirements that the models have to meet for them to be considered matching:

2.2.1.3.2.1 #1: Complementary roles

Parties need to play complementary roles (e.g. buyer/seller)

2.2.1.3.2.2 #2: Matching resources

The resources expected in the exchanges need to match to the ones expected by the other partner (e.g. the provided resources could be subtypes of resources requested)

2.2.1.3.2.3 #3: Satisfied timing constraints

The timing constraints on events (commitment specification) need to be mutually satisfiable (e.g. down payment vs. final payment, payment within 24 hours, shipment within 1 week, etc...)

2.2.1.3.2.4 #4: Transaction preservation

The sequence of expected business transactions needs to be the same (even though the individual business activities and resulting conversation patterns may differ). This is especially important for those transactions, which result in legal consequences.

If the above conditions are met, we can declare that the parties follow the same business model to achieve common business goals, and that the differences lie only in the technical infrastructure they use to implement their business model. If any of the above requirements is not met, there is no sufficient business foundation for these parties to cooperate, even in non-electronic form.

A successful completion of this step means that we have established a common business context for both parties. We have also identified the events that need to occur, and the collaborations between agents that support these events. This in turn determines the transactional boundaries for each activity. See example scenario in the Proof-of-Concept section for an illustration of these principles.

(NOTE: this section definitely needs more work…)

This business context model will help us to make decisions in cases when a strict one-to-one mapping on the technical infrastructure level is not possible. It will also help us to decide what kind of compensating actions are needed in case of failures.

2.2.2 Semantic Translation (to be completed)

Figure 8 presents the idea of the semantic translation and the reason why it's a required step in solving the interoperability puzzle. In general, the concepts underlying the foundations on which the IT infrastructures are built, differ between not only the industry sectors, or geographical regions, but even between each company within the same sector. This phenomenon – of different semantics, and different ontologies– causes many complex problems in the area of system integration, and in the area of e-commerce integration specifically.

One of the most common cases that require semantic translation to be performed is when each business party uses a different product catalogue (this situation is sometimes referred to as the "catalog integration", or "catalog merging" problem).

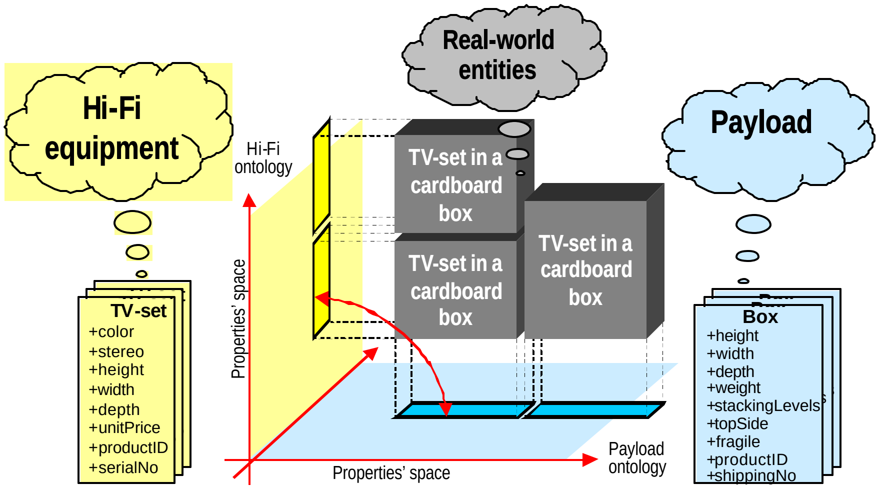


Figure 8 Mapping concepts from different ontologies.

In the example presented on Figure 8, a real-world entity - TV-set in a cardboard box - is represented very differently in two domain ontologies - the ontology of Hi-Fi equipment, and the transportation ontology. Although two representations may refer to the same real entity, in order to communicate that fact to the users of the other ontology we need to perform a semantic enrichment, in order to determine the proper classification of the concept in the other ontology.

What's even worse, we may discover (as is often the case) that the concepts overlap only partially, and the conditions under which they match the concepts from the other ontologies are defined by complex formulas, dependent potentially on several factors such as values from external resources , time, geographical region etc. In this case, the physical dimensions of the TV-set concept are confusingly homonymous to the dimension properties of the Box concept, but in the first case they refer to the TV-set chassis, and in the second case they refer to the cardboard box dimensions. Furthermore, the Box dimensions might be allowed to take only certain discrete values (e.g. according to a normalized cardboard container types), so in order to determine their values based on the information available in the TV-set concept, it is necessary to access some external resource (a cardboard box catalogue).

2.2.2.1 Describing semantic mapping

2.2.2.1.1 Semantic Translation meta-model

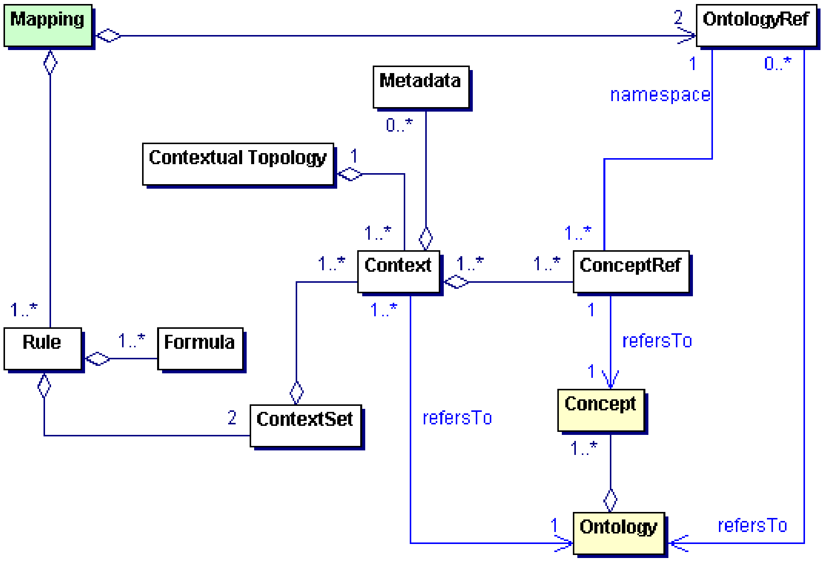


Figure 9 Semantic Translation meta-model

Figure above presents the meta-model for capturing the rules of semantic correspondence between concepts belonging to two different ontologies. This meta-model has been developed based on the principles of contextual navigation, which means that the proper understanding of a concept requires considering the context in which it occurs.

Furthermore, the translation rules (mappings) only refer to the original ontologies and concepts, which means that the original definitions, constraints, relationships and axioms are not recorded in the translation rules, but are only represented by unique identifiers (references). The reason for this is that especially in the e-commerce scenarios these source ontologies are usually completely separate, and maintained by separate organizations. These two concepts (Ontology and Concept) are accordingly marked as “external” in the list below.

- *Ontology*: the original full domain ontology (external)
- *Concept*: concepts defined in the original Ontology (external)
- *Mapping*: a top-level container for the semantic mapping rules, applicable to a pair of ontologies, as specified by the *OntologyRef*-s. (The *Mapping* is marked green in the diagram as the starting point for reading the whole meta-model.)
- *OntologyRef*: a URN uniquely identifying the referred ontology (possibly allowing to access it remotely).

- *ConceptRef*: a namespaced reference to individual *Concept*-s defined in the original *Ontology*. A URN, which possibly allows to access remotely the concept definition in the original ontology.
- *Context*: built on the basis of the original *Ontology* (refersTo), consists of related concepts represented by *ConceptRef*-s, which are considered relevant to the given transformation rule (the exact and full relationship of the *Concept*-s is defined in the original ontology - *Context* captures just the fact that they are related for the purpose of mapping).
- *ContextSet*: a group of one or more *Context*-s referring to the same *Ontology*.
- *Rule*: a rule that defines how to translate between the concepts in a *ContextSet* from one ontology, to the corresponding concepts in a *ContextSet* from the other ontology. A *Rule* consists of exactly two *ContextSet*-s, each one referring to respectively one of the ontobgies, and a set of *Formula*-s, which define the valid transformations on these *ContextSet*-s.
- *Formula*: a formal expression defining how translation is performed between concepts from the source *ContextSet* to those in the target *ContextSet*.

The reason for defining the *ContextSet*, in addition to *Context*, is that probably we would like to use concepts from several contexts belonging to a single *Ontology*, and map them to several contexts in the other. But at the same time there is a requirement to state explicitly that we always map between exactly two different ontologies.

2.2.2.1.2 Algorithms for discovering the semantic correspondence

(Many exist, none ideal or fully automatic. There is a need to use several in parallel, plus heuristics…)

2.2.2.1.3 The Formula language

(Needs to be more complex than first-order logic. Probably a full-fledged programming language, e.g. XSLT, JavaScript, XQuery, etc.)
It is yet to be defined what kind of language will be used to describe the transformations between the models. The following is a short list of the requirements that need to be satisfied:

- Preferably Open Source implementations available
- Highly portable
- Well-known: this is needed in order to ease the adoption
- Strongly typed: the transformations need to be precisely defined, and it's preferred that most logical errors would be discovered during the parsing/compilation, not at the runtime.
- High level (additional tools for manipulation of complex programmatic structures, database and directory access, etc…)

The candidates that we consider at this stage are Java, JavaScript, XSLT, XQuery and Python.

2.2.2.2 Example model

Below is an example of (part of) the model built with the Semantic Translation meta-model.

(NOTE: for now the Formula language is unspecified, and in this example a JavaScript-alike language was used).

```
Rule:rule1
| | |
| | +-- ContextSet:set1 {Ontology 1}
| |       \Context:Party
| |       \Context:Address
| |       \Context:PartyIdentification
| |       \Context:Name
| +--ContextSet:set2 {Ontology 2}
|         \Context:Agent
|         \Context:Location
|         \Context:Name
|         \...
\Formula:formula1
|   \body: "set2.Name = set1.Name"
\Formula:formula2
|   \body: "set2.Location.Address.Street1 =
                   set1.Address.Street;
            set2.Location.Address.Street2 =
                   concat(set1.Address.Zip, set1.Address.City);"
\Formula3:Formula ...
        .....
```

(NOTE2: There is also a working hypothesis that one could use a rule of thumb to treat the ebXML aggregate core components as Contexts, and most primitive core components as concepts - but this needs further research, and discussions with the eBTWG community.)

2.2.3 Business Process Mediation (to be completed)

2.2.3.1 Business Process Models

The elements of Business Process models describe the major steps in the interaction scenario that need to be performed in order to successfully execute the mutual commitments. In this step we identify the business transaction boundaries, and the activities that need to be performed in order to fulfill them, or what kind of activities are needed to rollback (or compensate) for failed transactions.

A *business process* (according to [REA],[ebXML],[UMM]) consists of a sequence of *business activities* performed by one business partner alone, and *business interface activities* performed by two or more business partners. In the ECIMF methodology we will be interested primarily in aligning the *business interface activities*, although in most cases understanding both types of activities is needed in order to understand the business process constraints. These activities realize the collaborations between the involved business Agents, and they also support the economic exchanges identified in the Business Context models. Further, we will use the term BusinessActivity to mean the business interface activity.

In this model, each collaboration task is further decomposed into *business activities*, which may involve one or more *business transactions*, which in turn are executed with help of *business documents* and *business signals*.

2.2.3.1.1 Business Process Meta-model

Here are more detailed descriptions of each of the modeling elements:

- *BusinessProcess*: contains one or more economic exchanges, which in turn contain two or more BusinessCollaborationTasks each.
- *BusinessCollaborationTask*: a logically related group of BusinessActivities, which realizes the collaboration between two Agents in a given Event.
- *BusinessActivity*: a business communication (initiated by a requesting or responding business Agent). BusinessActivities may lead to changes in state of one or both parties.
- *BusinessTransaction*: a set of BusinessDocuments and BusinessSignals exchanges between two parties that must occur in an agreed format, sequence and time period. If any of the agreements are violated then the transaction is terminated and all business information and business signal exchanges must be discarded (possibly some additional compensating actions need to be taken as well).
- *BusinessDocument*: a message sent between partners as a part of information exchange, which contains business data (payload).
- *BusinessSignal*: a message that is transmitted asynchronously back to the partner that initiated the transfer of business process execution control (by sending a BusinessDocument), which doesn't contain any business data, but instead just signifies acknowledgement or error condition.

(NOTE: probably this meta-model needs to be harmonized with UMM or eBTWG, but there is also a need to provide a ***simplified*** *version…)*

2.2.3.1.2 Business Process Models

Business processes are most often modeled using UML activity diagrams (or similar notation), where each diagram represents one of the collaborations. This view relates to the Business Context view in the following way:

- The collaboration links between Agents correspond 1:1 to BusinessCollaborationTasks. This means that for the typical economic exchanges there will always be two BusinessCollaborationTasks– one for the "give" part, and one for the "take" part of the exchange.

In addition to that, the BusinessProcess view enhances the understanding of the Business Context, because it allows us to correlate various Events that are dependent on each other even if they don't belong to the same economic exchange (e.g. consumption of resources, replenishment and sales tasks are dependent on each other, but they are not likely all to be part of the same BusinessCollaborationTask between two specific partners).

2.2.3.1.3 Business Collaboration Tasks and Business Transactions

- The BusinessCollaborationTasks support the execution of the BusinessEvents identified in the previous step. There should be as many Business Tasks as many collaboration links were in the Business Context models.
- BusinessEvents are realized by one or more BusinessTransactions. Consequently, BusinessCollaborationTasks consist of one or more BusinessTransactions
- BusinessCollaborationTasks are represented as UML activity diagrams, showing the activities of both collaborating agents. These diagrams usually contain two parts (swimlanes): one for the requesting (initiating) party, the other for the responding party. The diagrams should also contain the messages passed between the parties.

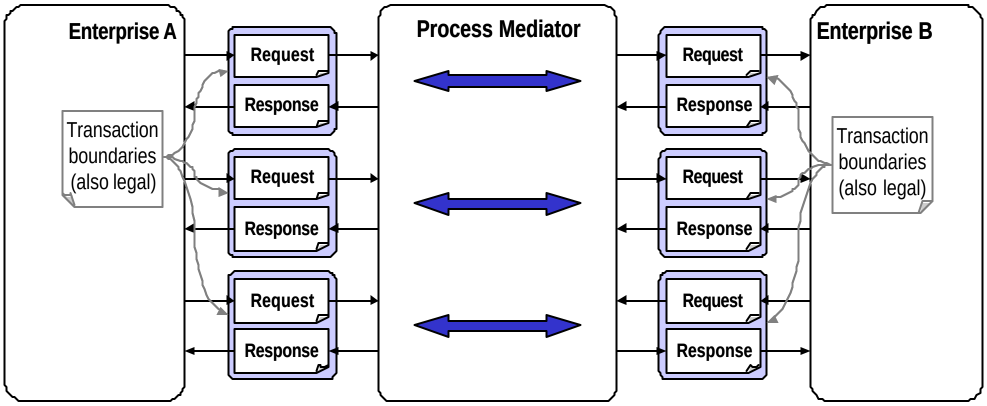


Figure 10 Example scenario that requires Process Mediator.

2.2.3.2 Business Process Mediation Model

The mediation between two different conversation patterns (which may involve different low-level technical transactions) needs to be designed and managed in a Business Process Mediation model.

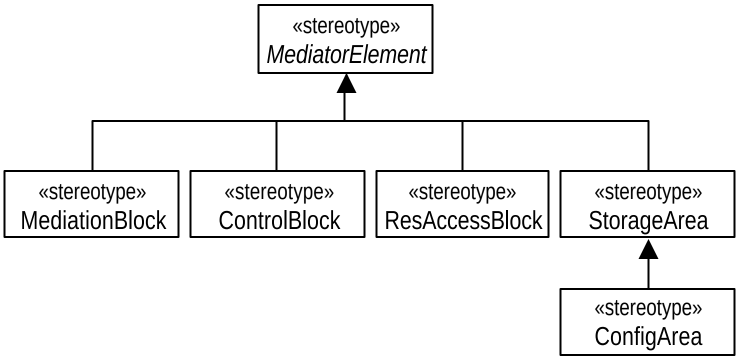
2.2.3.2.1 Business Process Mediation Meta-model

(NOTE: the working hypothesis is that the model elements will be responsible for reconciliation of concrete aspects of conversations. The current idea of the internal structure of the model is as follows:

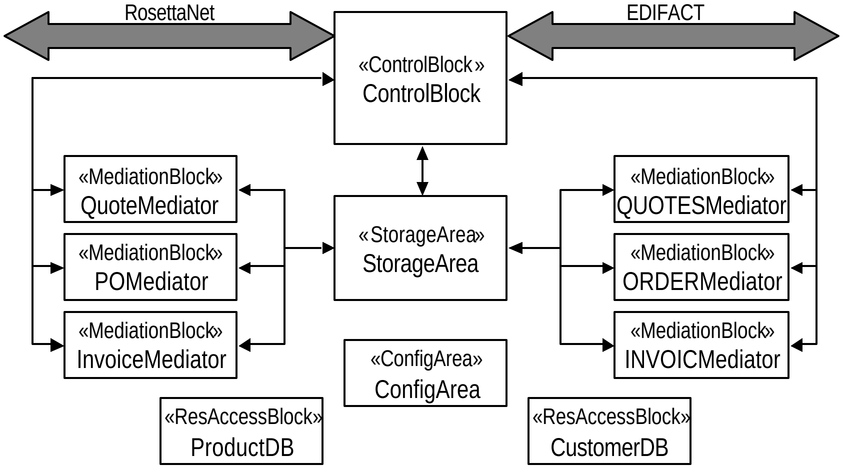
- *there will be mediation blocks handling the flow of each business transaction – totally the number of distinct business transactions on one side plus the number of distinct business transactions on the other side. These mediation blocks will be responsible for handling the details of conversations according to a given framework, within the boundaries of one specific transaction.*
- *there will be resource wrapper blocks, allowing for uniform access to external resources*
- *there will be one controlling block, responsible for managing the overall flow of transactions.*

- *there will be a common storage area, which any mediation block or the controlling block can access in order to store intermediate data – such as previous messages*
- *similar to that, there will be a configuration area accessible to all blocks, containing the configuration parameters.*

To summarize, the following diagram presents the meta-model:



And the diagram below presents a mediation model example:



Again, this is just a working hypothesis – any comments are much welcome!)

2.2.3.2.2 <u>Checking the task alignment</u>

(to be completed…)

2.2.3.2.3 <u>Creating the Mediation elements</u>

(to be completed…)

The process of building this part of the integration model is very closely related to the Semantic Translation, because very often a semantic correspondence needs to be established between the concepts, transactions, messages and information elements.

A Process Mediator is responsible for monitoring the conversation flows between each partner and itself, and according to the mapping rules it should generate appropriate stimuli (in form of message flows) in order to achieve desired state changes in each partner's Business Objects, while preserving the transaction boundaries.

Readers are referred to the Proof-of-Concept section, which illustrates these principles on a real example.

2.2.4 Syntax Mapping (to be completed)

2.2.4.1 Data element mapping

(using the semantic mapping rules. Syntax mapping is often preformed with XSLT, plus optionally the straightforward wrappers for non-XML formats)

2.2.4.2 Message format mapping

(see above. Additionally, it needs to ensure the well-fomedness and validity of messages according to the format specifications.)

2.2.4.3 Message packaging mapping

(ebXML CPP/CPA ?)

2.2.4.4 Transport protocol mapping

(ebXML CPP/CPA ?)

2.2.5 MANIFEST recipes

The meta-framework definitions/recipes for interoperability are named "MANIFEST". The language to be used in these definitions will be called E-Commerce Integration Modeling Language ("ECIML"), and will be based on XML representation of ECIMF models, rules and definitions.

A MANIFEST document consists of a set of interoperability recipes, based on the transformation model prepared using ECIML notation and then expressed in a serialized (XML) format. The MANIFEST-s will be identified by a unique ID, and stored in the repository from which an ECIML-compliant agent can retrieve it. The agent, based on the transformations specified in the MANIFEST recipe, will create necessary processing structures to align the message handling and interactions between the agents belonging to different frameworks. It should also be possible for ECIML-compliant modeling tools to re-use already existing MANIFEST recipes to adjust the interoperability model to specific needs. It is expected that some publicly available repository will store the commonly used templates for inter-framework alignment, so that less experienced or knowledgeable users can leverage the accumulated expertise of framework experts, and by making relatively minor adjustments re-use the templates as their own MANIFEST recipes.

The specifics of the repository need to be further discussed. Initially we suggest possibility of using either ebXML or UDDI to store the MANIFEST recipes.

2.3 The ECIMF-compliant runtime toolkit

The project aimed to provide a simple implementation of the E-Commerce Integration Toolkit ("ECIT"), consisting of the ECIMF Navigator (extending existing toolkits, like

Conzilla or Protégé) and a basic implementation of ECIML-compliant agent, and make these available on an Open Source basis. However, in order to fully leverage the ECIMF approach, we expect the software vendors to follow our initiative and provide complete implementations as proprietary products – still, compatible with the open standard.

The alpha-stage version of this toolkit has been implemented based on the Protégé framework, and is distributed under Mozilla Public License (a non-restrictive, business-friendly open source license).

2.4 Frameworks Integration Guideline

The main objective of the ECIMF project is to provide clear guidelines and methodologies for building interoperability bridges between different incompatible e-commerce standards.

This section presents a general guideline to solving this issue in case of two incompatible e-commerce frameworks F1 and F2. Annex 1 gives additional supporting information.

The guideline has been divided into several steps, to be performed sequentially and iteratively, as needed. The steps follow the methodology described in the previous section – the layers on the top are addressed first, since they give the broadest context necessary for understanding of the lower-level data transformations. The successful completion of all steps will result in a set of interoperability rules, enforced by a framework mediating agent, which will allow parties using different frameworks to cooperate towards common business goals.

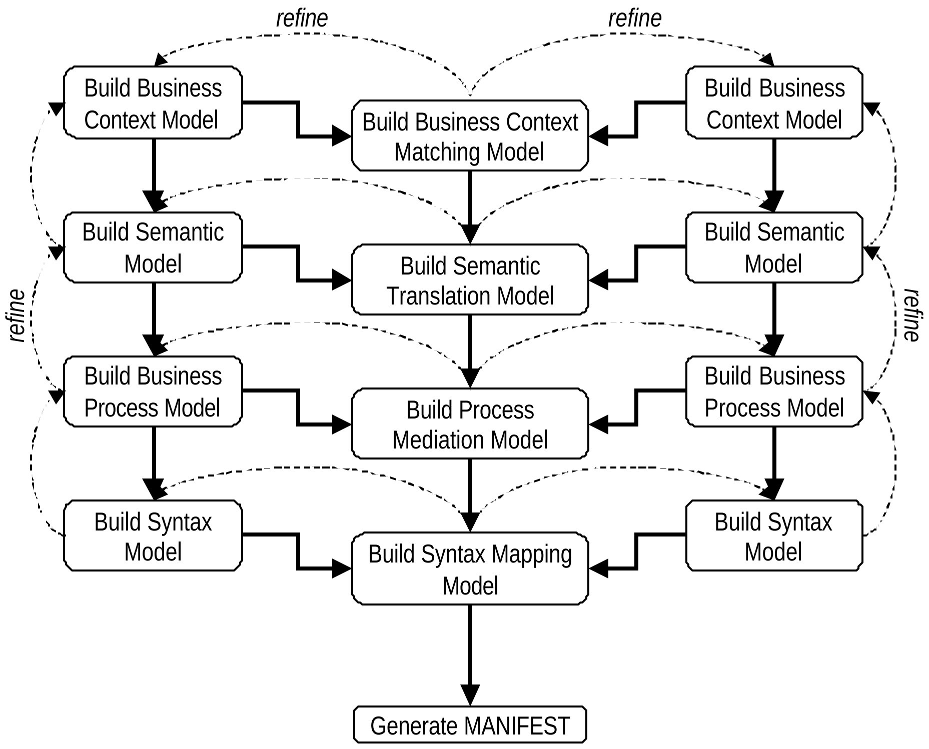

refine
refine
Build Business Context Model
Build Business Context Matching Model
Build Business Context Model
Build Semantic Model
Build Semantic Translation Model
Build Semantic Model
refine
refine
Build Business Process Model
Build Process Mediation Model
Build Business Process Model
Build Syntax Model
Build Syntax Mapping Model
Build Syntax Model
Generate MANIFEST

Figure 11 The process of modeling the integration recipes between two e-commerce frameworks.

The guideline has a modular structure, reflected in the fact that in each step several so-called *alternative procedures* have been defined. Each *alternative procedure* refers to a well-defined unit of work that needs to be done (a part of integration step), and allows you to replace or extend the approach suggested for that step with other methods of your choice, as long as they provide you with similar results (artifacts) as the input to the next step. The boundaries of each alternative procedure are clearly marked, and the input/output deliverab les are specified.

You can also find a common meta-model defined in each of the steps, which serves as a common vocabulary (shared ontology) for understanding the incompatible frameworks.

One important thing to note here is that the integration modeling between two frameworks is asymmetric, i.e. the integration model will usually contain two elements that refer to the same individual model elements, but defined differently depending on the direction in which the data is traveling.

The subsections below present the details of the guideline.

2.4.1 Analysis of the Business Context Matching

2.4.1.1 Creating Business Context Models

A **business context model** shows a concrete business scenario expressed with the use of economic modeling elements, e.g. those found in the REA meta-model. We suggest using the following standard UML diagrams for that purpose:

- Class diagrams to show the specific types of entities involved.
- Collaboration diagrams to show a specific scenario populated with specific instances of participating entities.
- Value-chain diagrams (REA process diagrams), to clearly define the flows of resources, and how they depend on the collaboration between partners.

For examples of such business context models, please see the ECIMF-POC document.

2.4.1.2 Checking the Business Context Matching Rules

Each of the context matching rules needs to be checked, and any additional requirements or assumptions made need to be recorded, so that they can be used to understand the interactions in the lower layers of the ECIMF model.

Below is a table that summarizes this step of the guideline:

Business Context Matching	
Input	Traditional business knowledge, legal agreements between partners, industry specific rules, legal constraints, specific business goals, common business practices and codes of conduct
Output	Tw o Business Context Models for the integration scenario, defined in a set of UML diagrams (class, collaboration, activity), and an analysis of their matching (and any additional requirements on which the matching depends).

2.4.2 Creating the Business Process Mediation Model

2.4.2.1 Creating the Business Process models

A **business process model** shows concrete business collaboration, expressed as series of business activities and transactions between the partners. We suggest using the standard UML activity diagrams for that purpose, one diagram for each collaboration.

2.4.2.1.1 Identify the Business Collaboration Tasks

For each collaboration link in the Business Context diagram, a Business Collaboration Task is created.

2.4.2.1.2 Identify the Business Transactions

For each collaboration, and for each Agent, the business transactions are discovered and described. Since the Agents possibly use different frameworks, there might be different transactions expected even for the same collaborations.

For examples of such business process models, please see the ECIMF-POC document.

2.4.2.2 Creating the Mediation model

(NOTE: describe how the process mediation model can be created, using concepts from the Mediation meta-model.)

(NOTE2: the relationship to eBTWG BOT's [Business Object Types] need to be analyzed. BOT's define not only the class (+properties), but also the behavior, state and methods. As such, they are the best candidates to provide the intermediate internal model, and the problem of process mediation could be reduced to the problem of reconciling the state diagrams of the key BOT's. Please see the analysis in PowerPoint slides at http://www.ecimf.org/events/Brussels-20020220/ECIMF-eBTWG.ppt).

Below is a table that summarizes this step of the guideline:

Business Process Mediation	
Input	Business Context models, other information on business processes supporting the business context, semantics of the business processes (obtained in the next step), etc.
Output	Business Process Models, Business Process Mediation Model for the integration scenario, defined in a set of diagrams (activity/business process, ECIMF process mediation diagram)

2.4.3 Creating the Semantic Translation Model

2.4.3.1 Acquiring the source ontologies

(NOTE: describe the process of discovering the ontologies from e-commerce standards, best practices, business rules etc…)

2.4.3.2 Selection of the key concepts

(NOTE: describe how the business context and business process models help to determine the key concepts …)

2.4.3.3 Creating the mapping rules

(NOTE: describe how the mapping rules can be created, based on one of the alternative procedures …)

Below is a table that summarizes this step of the guideline:

Semantic Translation	
Input	Tw o source ontologies, obtained from formal specifications, UML models, textual descriptions, knowledge of domain experts etc.
Output	Semantic Translation Model, containing rules for equivalence of the key concepts.

2.4.4 Creating Syntax Mapping Model

2.4.4.1 Data element mapping

(NOTE: describe how the external formats can be mapped to internal representation …)

2.4.4.2 Message format mapping

(NOTE: describe how the message well-formedness rules can be satisfied. This may involve proactive "asking" for more information in order to satisfy the demands of a given message format…)

2.4.4.3 Message packaging mapping

(NOTE: describe how the message packaging [encoding, charset, MIME, etc] can be aligned)

2.4.4.4 Transport protocol mapping

(NOTE: describe how the transport protocol parameters need to be defined.)

Below is a table that summarizes this step of the guideline:

Syntax Mapping	
Input	Semantic Translation Model, simple mapping of primitive data types, external resources to be used.
Output	Syntax Mapping Model, containing the exact mapping of data elements, message formats, packaging and transport protocols.

For additional details, and more information on alternative procedures available for each of these steps, please refer to the Annex.

2.5 References

[UMM]: *Unified Modeling Methodology*; UN/CEFACT TMWG N090R9.1; available from: UN/CEFACT TMWG. A copy of the draft can be also found at: http://www.ecimf.org/doc/other/TMWG_N090R9.1.zip

[ebCDDA]: *Core Components Discovery and Analysis*; ebXML, May 2001; available from: http://www.ebxml.org/specs/ebCDDA.PDF

[ccDRIV]: *Catalog of Context Drivers*; ebXML, May 2001; available from: http://www.ebxml.org/specs/ccDRIV.PDF

[CID52]: *Conceptual Navigation and Multiple Scale Narration in a Knowledge Manifold*; Ambjörn Naeve; KTH, 1999; available from: http://cid.nada.kth.se/sv/pdf/cid_52.pdf

[OB00]: *Ontology-Based Integration of Information — A Survey of Existing Approaches*; H. Wache, T. Vögele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann and S. Hübner; University of Bremen, 2000; available from: http://www.tzi.de/buster/papers/SURVEY.pdf

[SAGV00]: *Semantic Translation Based on Approximate Re-Classification*, Heiner Stuckenschmidt, Ubbo Visser; University of Bremen, 2000; available from: http:// www.tzi.de/buster/papers/sagv-00.pdf

[SW00]: *A Layered Approach to Information Modeling and Interoperability on the Web,* Sergey Melnik, Stefan Decker; Stanford University, 2000; available from: http://www-db.stanford.edu/~melnik/pub/sw00/sw00.pdf

3 Proof-of-concept – scenario analysis

3.1 Editor's note

Originally this section formed a separate document. There may be still some inconsistencies related to this fact.

3.2 Purpose and scope

This section presents a step-by-step example of how the ECIMF can be used to prepare a set of recipes for interoperability between two e-commerce partners.

In this scenario, one partner, referred to as a Customer, produces Hi-Fi equipment of various sorts, and needs to ship them to the merchants. The other partner, referred to as Shipping Agency, offers services of shipping goods.

The Customer uses RosettaNet Implementation Framework 2.0 (RNIF) as his e-commerce interface, whereas the Shipping Agency uses EDI (EDIFACT D99.A).

This example follows the steps outlined in the Frameworks Integration Guidelines (in General Methodology section).

3.3 Business Context Matching

In this step, two Business Context models are built and compared, in order to check whether they can match the expectations of the other business partner.

3.3.1 Creating the Business Context Models

The diagrams below have been built using REA modeling elements, here expressed as UML stereotypes.

(NOTE: they present only a subset of the full diagram! E.g. there should be a Resource:Payload and Resource:Labor which is transformed or used by the Events…)

Figure 1 presents the business context diagram for the shipping agency. Here are the key elements of that diagram:

- The agency expects the payment first, and only then delivers the service
- The roles of ShippingAgent and Cashier are split into two different entities (persons, divisions …)
- ShippingAgent and Cashier collaborate with each other in order to satisfy the business rules (payment needs to be fulfilled first, and only then the shipment takes place)
- Both ShippingAgent and Cashier collaborate with the Customer.

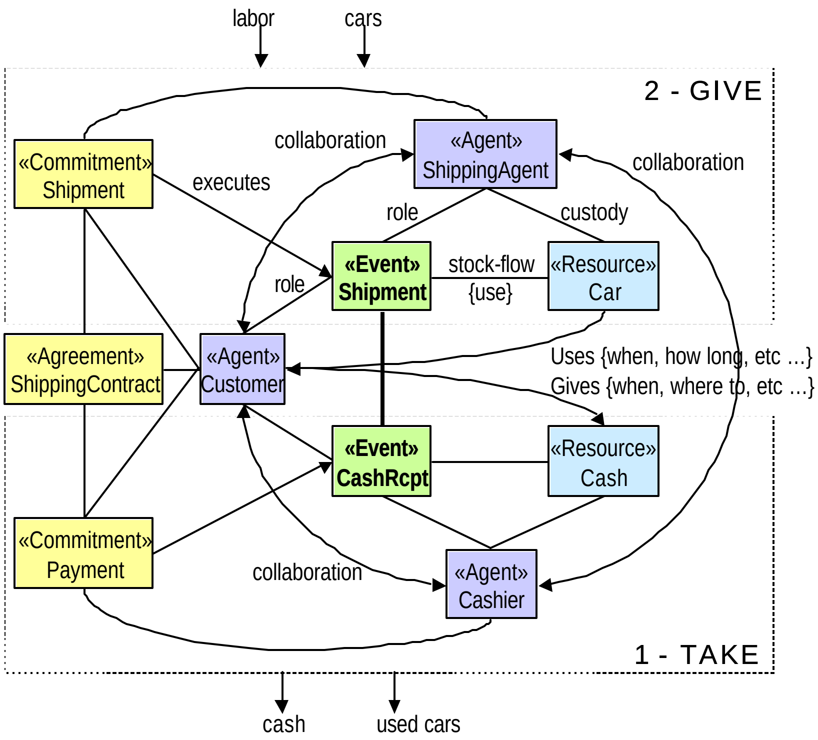


Figure 12 Business Context model as seen by the shipping agency.

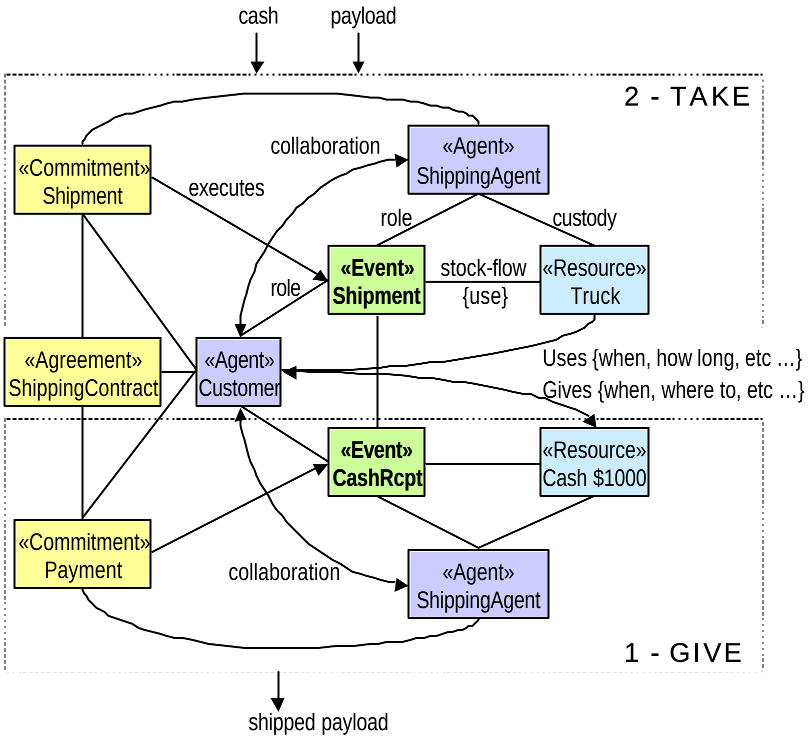


Figure 13 Business Context model as seen by the customer.

Now, for the customer the business context can be represented as shown on the next Figure. The key elements are:

- Customer expects first to give cash, then receive a service
- Customer wants to deal with the same entity for both events
- Customer has some specific demands on the kind of car, and the amount of cash.

3.3.2 Checking the Business Context Matching

From the diagrams above it is clear that in order for these two partners to be able to collaborate – in the traditional or in the electronic way – the following criteria have to be met (which ECIMF calls "business context matching rules"):

- *#1: Partners need to play complementary roles*: which is here the case. Note: although the Customer has a limited view of the Shipping Agency organizational structure (he wants to deal with just the ShippingAgent), it still has to be determined if he is able to deal with two separate persons/entities, which is required by the Shipping Agency (ShippingAgent and Cashier).
- *#2: Expected resources need to be equivalent*: in this case, parties need to agree on the exact kind of transportation used, and the exact amounts of money to be paid. They need to also agree on several additional properties of using the transportation (when, how long, from where, etc …) and providing the payment (when, where to, what currency etc…).
- *#3: Timing constraints need to be mutually satisfiable*: in this case, the Customer is able to satisfy the requirement of the Shipping Agency that he needs first to pay. Further timing constraints may show up when analyzing the collaboration patterns between the parties.
- *#4: Transaction boundaries need to be preserved*: in this case, there are two transactions: payment and shipment, possibly consisting of several lower-level technical transactions. All supporting communication between the partners needs to be aligned in such a way that it preserves these boundaries for each of them.

After additional negotiations, we can state that these two Business Contexts match. These additional requirements identified in this step need to be recorded.

(NOTE: how?)

For the sake of this example, we assume that both parties agreed to follow the model presented on the first Figure.

3.4 Process Mediation

3.4.1 Create Business Process models

Based on the Business Context models, we determined that the collaborations we are interested in are the following:

- **Payment collaboration task**: involving Customer and Cashier
- **Shipment collaboration task**: involving Customer and ShippingAgent.

Based on that, we should be able to identify concrete business processes existing within each organization, which support these collaborations. Also, it should be possible to identify the business transactions, which involve the electronic communication between the partners, and sending of electronic business documents.

3.4.1.1 Identify the Business Transactions

For all collaboration tasks we need to describe two sets of transactions, each according to the framework used by the Agent. As an example, we will analyze in detail the Payment Collaboration Task.

The following table contains the example list of business transactions, together with their business documents, identified for the Customer:

Party	Customer		
Collaboration Task	Payment Collaboration		
Framework	RNIF 2.0		
Transaction name	**Initiator / Responder**	**Request document**	**Response document**
PIP3A1: Request for quote	Initiator	QuoteRequest	QuoteConfirm
PIP3A4: Request Purchase Order	Initiator	PORequest	POConfirm
PIP3C3: Notify of Invoice	Responder	Invoice	
PIP3C6: Notify of remittance advice	*Initiator*	*RemittanceAdvice*	
Message delivery control	*Any*	*Secure Flow*	

In a similar manner, we identify the transactions for the Shipping Agency:

Party	ShippingAgency		
Collaboration Task	Payment Collaboration		
Framework	EDIFACT		
Transaction name	**Initiator / Responder**	**Request document**	**Response document**
Request for quote	Responder	REQUOTE	QUOTES
PIP3A4: Request Purchase Order	Responder	ORDERS	ORDRSP
Notify of Invoice	Initiator	INVOIC	
Notify of remittance advice	Responder	REMADV	
Message delivery control	*Any*	*APERAK, CONTRL*	

However, at this point we discover that the Customer's system doesn't implement the PIP3C6 – in the RosettaNet framework this is optional. We also discover that RosettaNet uses so called SecureFlows for communication control, whereas EDIFACT uses two messages: APERAK and CONTRL. Furthermore, we see that in EDIFACT framework use of these two messages is also sometimes optional. We need to further study their semantics – see the section on Semantic Translation.

It is useful also to picture these collaborations in a common diagram. This is presented on the Figure below. The business transactions are shown here also, as rounded boxes containing the business documents. These transactions change the states of each partner's Business Objects. Areas of potential problems are marked with red color.

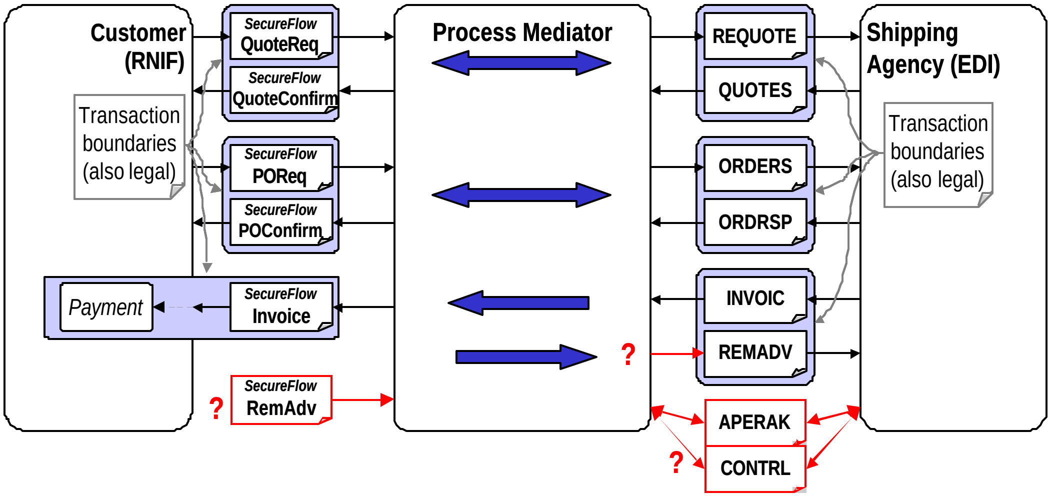


Figure 14 Process Mediation for the Payment Collaboration Task.

3.5 Semantic Translation

This step of integration helps to discover the underlying data model and the differences in meaning of the concepts used by each e-commerce framework. As it will be demonstrated, these differences will affect the design of both the process mediation and the syntax mapping.

For the sake of this example, let's assume that the customer wants to ship TV-sets from the factory to the shops.

This step will make use of the individual ontologies, a shared vocabulary and external resources in order to map between the key concepts in each of the frameworks.

Please note that generally the mappings are not symmetric, i.e. different rules and possibly different external resources need to be used when translating concepts from Customer to Shipping Agency than the other way around. For this reason, two sets of rules will always be present for each concept.

3.5.1 Acquire the source ontologies

For the purpose of this example, we acquired necessary concepts from each of the e-commerce frameworks – RNIF and EDIFACT respectively. We also made quite a few assumptions, which in the real case would have to be obtained from the particular IT system implementation, message implementation guidelines, product catalogues, company's procedures etc.

3.5.2 Select the key concepts

Let's start from the mapping of the two representations of a real-world entity (TV set), which is the subject of the shipment. These representations differ in each framework, because of their different scope.

This entity is represented in the ontology of the Customer as a TV-set – a kind of Hi-Fi equipment, while in the ontology of the Shipping Agency it is represented as a Box – a kind of Payload.

3.5.3 Create the mapping rules

The table below presents the semantics of the two corresponding concepts – TV-set in the Customer ontology, and Box in the Shipping Agency ontology – and the mappings required between the two representations, whenever they occur in the business documents.

Customer: TV-set **Properties**	**Semantic Translation** **Mapping Rules**	**Shipping Agency: Box** **Properties**

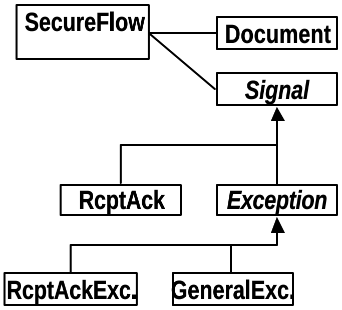
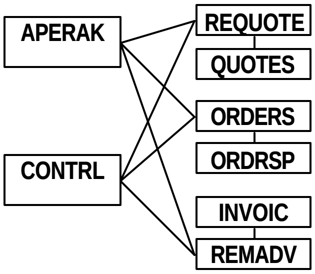
`Height`
`Width`
`Depth`
Represent the physical dimensions of the TV set chassis.
	Tv_set ? Box: dimension values will always be higher, but discrete. Need to be obtained from a cardboard box catalogue (external resource)
	Box ? Tv_set: dimension values will always be lower. Need to be obtained from a TV products catalogue (external resource) using `productID`
		`Height`
		`Width`
		`Depth`
		Represent the physical dimensions of the cardboard box used to ship the electronic equipment of any kind. The values are discrete, because only certain box sizes are available.

Not available (N/A)	Tv_set ? Box: needs to be obtained from a product catalogue (external resource) Box ? Tv_set: not needed	`Weight` Represents the weight of the box with the contents.
N/A	Tv_set ? Box: needs to be obtained from a product catalogue (external resource) Box ? Tv_set: not needed	`StackingLevels` Represents the number of levels the boxes can be stacked, one on top of the other.
N/A	Tv_set ? Box: always set to True. Box ? Tv_set: not needed	`Fragile` Marks the payload as fragile (requiring special care during transportation)
`Color`	Tv_set ? Box: not needed Box ? Tv_set: needs to be obtained from a product catalogue (external resource)	N/A
`Stereo`	Tv_set ? Box: not needed Box ? Tv_set: needs to be obtained from a product catalogue (external resource)	N/A
`UnitPrice`	Tv_set ? Box: not needed Box ? Tv_set: needs to be obtained from a product catalogue (external resource)	N/A
`ProductID` Product identification (type)	Tv_set ? Box: concatenate with the `serialNo` Box ? Tv_set: split into `ProductID` and `serialNo`, based on a required `serialNo` length.	**`ProductID`** **Product identification (type), including serial number. Primary identification data**
`SerialNo` **Serial number. Primary identification data**	Tv_set ? Box: see rule above Box ? Tv_set: see rule above	N/A

There are several interesting observations that can be made based on this example:

- Several external resources need to be consulted in order to prepare the mapping. It is possible to record the fixed values in the translation rules, but it would be more flexible to be able to query these resources dynamically, during run time.
- However, some of the values can be specified explicitly in the rules, and have fixed value (e.g. the `fragile Box` property).
- The translation rules are definitely not symmetric– e.g. different external resources may need to be consulted in order to supply missing data.
- There is a property, which uniquely identifies the corresponding physical entity (`Tv_set.serialNo` and `Box.productID`), although it is defined differently and requires processing.
- The properties related to physical dimensions are confusingly homonymous, although in reality their relationship is governed by a complex formula (and requires use of external resources).

Before proceeding to the last step (syntax mapping), let's analyze the message delivery control mechanisms, as these were identified as problematic during the process mediation step.

Customer (RNIF)	Semantic Translation	Shipping Agency (EDI)
 SecureFlow Document Signal RcptAck Exception RcptAckExc. GeneralExc. `SecureFlow` consists of a business document (containing business data), and a responding business signal (acknowledgement).	The RNIF business documents map 1:1 to EDI business messages, e.g.: QuoteRequest ? REQUOTE QuoteConfirm ? QUOTES PORequest ? ORDERS POConfirm ? ORDRSP etc … However, individual data elements can be missing, and will have to be collected from the previous messages, or supplied explicitly in the rules, or obtained from external resources.	 APERAK CONTRL REQUOTE QUOTES ORDERS ORDRSP INVOIC REMADV In this particular case, the EDI system uses `APERAK` and `CONTRL` messages only to signal exceptions. Acknowledgements are implicit, in the form of response business documents.
`ReceiptAck` This signal means that the document business data has been accepted for further processing (which implies also well-formedness)	RNIF ? EDI: not needed – don't forward. EDI ? RNIF: needs to be synthesized from the response document. Possible problems with timing constraints… (ack. too late)	N/A – implementation choice (positive acknowledgements are implicit).
`ReceiptAckException` This signal means the document was not well-formed (parsing errors). Business data was not considered at all.	The semantics of both messages is identical, which means a 1:1 mapping can be applied, both ways.	`CONTRL` This message is sent when parsing errors occur. Business data was not considered at all.
`GeneralException` This signal means that there were errors in the business data processing (though it means	RNIF ? EDI: always map to `APERAK`	`APERAK` In this implementation, this message is sent only when an error occurs when processing

Customer (RNIF)	Semantic Translation	Shipping Agency (EDI)
processing (though it means implicitly the document was well-formed).	EDI ? RNIF: map only if the APERAK message carries an error status.	error occurs when processing business data (though it means implicitly the document was well-formed).

Again, this analysis brings a couple of interesting observations:

- The differences in the semantics of message flow control mechanisms will affect the implementation of the process mediator, because some messages need to be created, removed, or sent at different times than the originating messages. Conclusion: there is no simple 1:1 mapping between messages, and the process mediator is really needed.
- The business documents map 1:1 in this example. However, as shown on the Figure 3, the RNIF side doesn't produce the RemittanceAdvice message, which the EDI side needs for completion of the low-level transaction. This message needs to be either synthesized by the process mediator (by accessing an external resource, such as the payee's bank), or the RNIF side needs to implement it.
- The timing constraints for ReceiptAck (times defined in RNIF, which define how long the sender has to wait for an acknowledgement before concluding a failure) may be impossible to satisfy in this scenario. The EDI side doesn't produce required ReceiptAck signals, and they need to be created based on the response EDI messages – which may be sent too late to satisfy the timing limits defined in RNIF.

After completing this step, we are very well prepared to define the low-level syntax mapping – transformation of the data elements in individual messages.

3.6 Syntax mapping

According to the layered ECIMF model, the syntax mapping – i.e. the translation between the individual data elements – is the lowest layer of interoperability, and it is affected by the rules defined in all the higher layers.

Let's take for example a fragment of mapping between the PurchaseOrderRequest and ORDERS. Figure 5 shows the fragments of each message and the mapping links between the data elements.

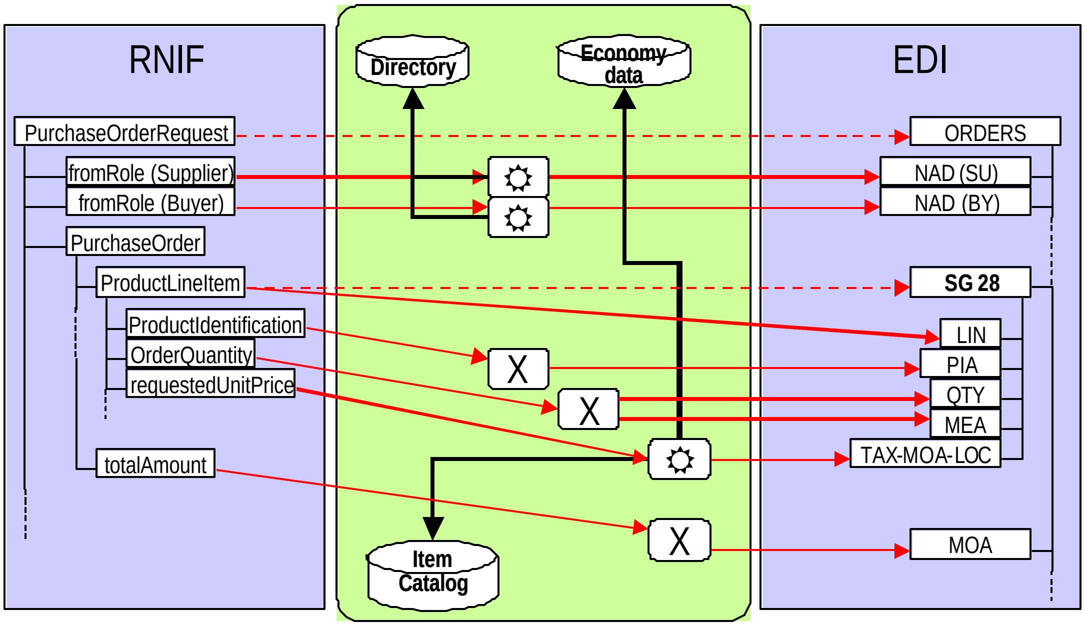


Figure 15 Message syntax mapping.

Again, a few observations can be made based on this example:

- This is a one-way mapping, as the arrows on the red links indicate. This means, that this mapping is valid for translation of PurchaseOrderRequest messages into ORDERS messages, and not necessarily vice versa (in fact, in our example different external resources will be needed to perform the translation in the other direction).
- The dashed lines represent the instance links, i.e. for each instance on one side a corresponding instance on the other side is created. In this case, for one `PurchaseOrderRequest` document one `ORDERS` message is created, and similarly for one `ProductLineItem` one Segment Group 28 (`SG28`) is created. Note, however, that additional limitations need to be considered here, which come from the limitations on the allowed number of the given data elements in a message. In this case, there can be no more than 200000 (according to EDIFACT D99.A) occurrences of `SG28` in a single `ORDERS` message. If there are more `ProductLineItems` than that, they probably need to be divided into two `ORDERS` messages – however, this changes significantly the flow of the low-level transactions, as presented on the Figure 3.
- The boxes with a toothed wheel represent complex processing, with the use of external resources. This is needed e.g. if the identification schemas for parties are different, or in the above-mentioned example of different product classifications.
- The boxes with an "X" represent simple data transformation, like numeric or string operations. E.g. as identified in the Semantic Translation step, the product ID used in EDI (`PIA` element) needs to be a concatenation of the sub-elements of the `ProductIdentification` element in RNIF.

In this step also the differences in the transport protocols and packaging are considered. Some differences (like use of FTP vs. SOAP) will require providing additional protocol parameters, e.g. FTP username and password, SOAP service name, a WSDL file, details of the MIME packaging etc. Some of these parameters can be expressed using ebXML CPP/CPA.

3.7 Generation of MANIFEST

As the final step, based on the models and transformation rules prepared in the steps above, a MANIFEST needs to be generated - an abstract recipe for interoperability between RNIF and EDI, within the given scope.

The example syntax of the MANIFEST document could look like the sample below:

```
<?xml version=’1.0’?>
<Manifest>
   <BusinessContextMatching name=’Shipment’>
      <BusinessContext id=’WidgetsLtd’> ... </BusinessContext>
      <BusinessContext id=’JoeShipping’> ... </BusinessContext>
   </BusinessContextMatching>
   <ProcessMediation>
      <Framework id=’RNIF’ name=’WidgetsLtd’>
         <BusinessProcessDefinition location=’uddi:...PIP3A4...’/>
         <BusinessProcessDefinition location=’uddi:...’/>
         <BusinessProcessDefinition location=’uddi:...’/>
      </Framework>
      <Framework id=’EDI’ name=’JoeShipping’>
         <BusinessProcessDefinition>
            ... (here it follows, defined using ebXML BPSS)...
         </BusinessProcessDefinition>
      </Framework>
      <MediationRules>
         ...
      </MediationRules>
   </ProcessMediation>
   <SemanticTranslation>
      <OntologyRef id=’RNIF’>urn:ont1 ...</OntologyRef>
      <OntologyRef id=’EDI’>urn:ont1 ...</OntologyRef>
      <Rule id=’rule1’>
         <SourceCtxSet id=’set1’/>
         <TargetCtxSet id=’set2’/>
         <formula id=’formula1’/>
         <formula id=’formula2’/>
      </Rule>
      <ContextSet id=’set1’><context id=’ctx1’/></ContextSet>
      <ContextSet id=’set2’><context id=’ctx2’/></ContextSet>
      <Context id=’ctx1’>
         <ConceptRef id=’tv_set’>urn:...TV-set</ConceptRef>
      </Context>
      <Context id=’ctx2’>
         <ConceptRef id=’box’>urn:...Box</ConceptRef>
      </Context>
      <Formula id=’formula1’>
         <body>
set2.ctx2.box.productID := set1.ctx1.tv_set.productID +
   ″ ″ + set1.ctx1.tv_set.serialNo;
         </body>
      </Formula>
      <Formula id=’formula1’>
         <body>
set2.ctx2.box.fragile := true;
```

```
         </body>
      </Formula>
   </SemanticTranslation>
   <SyntaxMapping>
         <Mapping>
            <SourceMessage>PurchaseOrderRequest</SourceMessage>
            <TargetMessage>ORDERS</TargetMessage>
            <Rules>
            </Rules>
         </Mapping>
         ...
   </SyntaxMapping>
</Manifest>
```

(This example uses the Semantic Translation ontology, developed for the purpose of this project – see http://www.ecimf.org/contrib/onto/ST/index.html for more details).

Note that for the purpose of configuring the ECIMF-compliant runtime, only the process mediation and syntax translation rules are needed. However, the models of the two other layers are included as well in order to facilitate exchange of the ECIMF models between the modeling tools, and to preserve the knowledge collected during the process of mapping.

In the next step, as presented previously in the Figure 5, the ECIMF-compliant agent receives the MANIFEST and instantiates the necessary adapters. This may involve setting up processing pipelines for messages, creating state machines to keep track of complex interactions, creating translation maps for message elements, reading parameters provided by the communicating parties, etc. This reference environment for execution of the MANIFEST recipe can be provided as a commercial product.

Finally, at this stage it is possible for the parties to successfully establish business interaction, even though they use different e-commerce frameworks to express their activities.

3.8 Implementation: ECIML-compliant agent

(This section is incomplete. Please see Annex 2 for some initial materials)

4 ECIMF Toolkit – description

4.1 Introduction

This software module has been created to illustrate and investigate various methodologies for concept mapping and alignment between different e-commerce standards. These standard e-commerce frameworks are represented as ontologies - shared conceptualizations of a given problem domain as seen by their respective user communities.

In our project we decided to follow a semantic translation approach, which uses an upper-level shared ontology that provides concepts-labels to identify similar concepts in each respective ontology. This approach means that instead of building a full-mesh N*(N-1) collection of translations for each pair of existing e-commerce frameworks, it is sufficient to prepare N translations from that framework by attaching conceptual labels taken from this shared ontology.

Under this approach, the following steps need to be performed:

- Attach labels taken from shared ontology to your concepts
- Find corresponding labels in the foreign ontology
- Apply more steps to refine the relationships:
    - Local context
    - Automated, formal reasoning and inference
    - External context – semantic enrichment
    - Heuristics (best practice and rule of thumb ?)
- Define the translation rules in a formal way

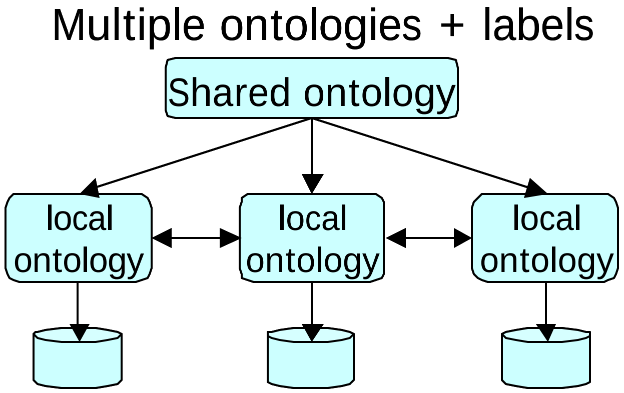


Figure 16 Shared ontology approach to semantic translation.

This software is provided under the terms of Mozilla Public License (see Mozilla site for details).

4.2 Limitations

Originally, this tool was intended as a more or less complete implementation of various modules of ECIMF-compliant agent, as described in the section on project goals. However, due to unexpected shortage of human resources, only the alpha-quality version of semantic translation module has been implemented. So, currently the tool is very limited in its scope and functionality:

- the ontologies that represent e-commerce standards need to be supplied in Protege .pprj format. This usually means that you have to convert them first from some other format, using either Protege built-in import modules or enter the concepts manually... There is a simple EDIFACT import module under development, as well as DTD/XSD import modules.
- the tool currently supports mapping through labeling. It is theoretically possible to use it for other mapping methodologies by using the scripting capabilities, but it would be inconvenient.
- the custom search script function is not supported yet, although the amount of work needed to complete it is small.
- there are numerous layout problems.
- other limitations exist, to be sure...

4.3 Simple usage scenario

First of all, users are strongly advised to first read the introductory material in the presentation http://www.ecimf.org/events/Paris-20020610/Interoperability.ppt - it helps to understand the principles behind the methodology implemented in this tool.

Let's step through a scenario, in which you will map the concepts in the included sample projects:

1. **Download, install, and start the tool.** The exact steps will depend on your platform - under Windows, when the installation process is completed, there will be a new item in your Start/Programs list called ECIMF-ST.
   When the tool is started, a console window will also appear, where you can find all sorts of useful debug information.
2. **Labeling:** in this step, you will attach labels to the concepts in each of the individual ontologies
   - Press LOAD button to load SOURCE project. A file selection dialog will open. Go to projects/ subdirectory and select source.pprj project.
   - In a similar way, load the LABELS project from labels.pprj.
   - Highlight one of the concepts in the SOURCE project. The bottom-right panel will show you the details of the concept, including a list of labels attached to it.
     (NOTE: you may want to resize the main window and/or individual panels by dragging the dividers between the panels)
   - Modify the list of labels by creating ("C" button) a label from scratch, attaching ("+" button) a label from the LABELS ontology, editing ("E" button) or removing ("-" button) a label. You will be mostly interested in using the "+" and "-" buttons.

Note that individual concepts can have multiple labels, each of them possibly characterising the concept in a different way. The labeling process helps to fix the SOURCE concept in the conceptual topology that can be described by the LABELS concepts. In other words, the more specific labels are attached to the SOURCE concept, the less ambiguous its definition is according to the LABELS ontology.

- Follow similar steps, but with the TARGET project.

3. **Mapping:** in this step, you will create a formula that describes a mapping between one of the SOURCE concepts and the TARGET concepts:

   - Select the "Mapping" tab. Note that the layout here is different - on the left the SOURCE project is presented, on the right there is the TARGET project, and in the middle you can see the panel where the mapping hints will be presented. You can access also the full MAP project if you want to browse the formulas.
   - Select one of the concepts in the SOURCE project.
   - Press the "Find in TARGET" button to show the hints.
     Note: the "Conf" button shows the various possible algorithms for finding the corresponding labels. Currently, the scripting is not implemented here, but please take a look at various possibilities of searching and matching...
   - If some hints are found, you can select them for use in a formula by checking the "Use?" checkboxes.
   - Create a new formula by clicking on "Create" button. You can also change the name of the formula.
   - A pop-up dialog will appear that lets you edit the formula in your favorite scripting language.
     This panel also shows you what kind of data sources are available to you in this context. A special name called "SOURCE" refers always to the input concept that you selected for mapping. Also, the target concepts that were found are available under their names.
     The properties of each concept are available directly as instance variables, so you can e.g. use a notation "SOURCE.name" to refer to the input concept name.
   - Press OK to save the formula to the MAP project.

   You can review the m apping formulas by clicking on a "Map" tab in the middle panel. Then select the "Formula" concept, and in the bottom-left panel select one of the instances of formulas.

   (Note: currently a layout problem prevents you from viewing the formula body).

4.4 Additional information

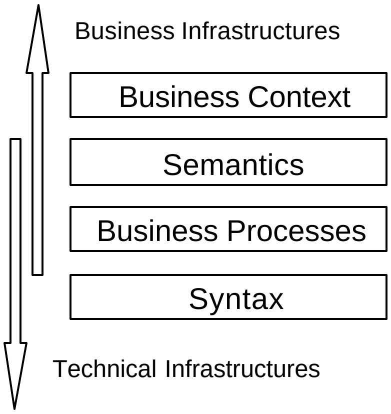
Additional information about the tool can be obtained from the author (Andrzej Bialecki ab@getopt.org). You may also check the PowerPoint slides here: http://www.ecimf.org/events/Paris-20020610/ECIMFToolkit.ppt).

The source code for the tool is included in the installation package downloadable from http://www.ecimf.org/software .

5 Summary and conclusions

The ECIMF project made an attempt to address the interoperability problems by providing a single general and holistic view of all major aspects involved in solving concrete integration scenarios between e-commerce partners.

The most important outcome of the project seems to be the 4-aspect model of interoperability:



We have investigated various existing approaches that address each of these areas, and tried to indicate which of them need further research. Based on this, we present the following conclusions and recommendations for future work.

5.1 Interoperability of Business Contexts

REA models (retro-fitted to virtual organizations) help to understand interoperability issues on the value-chain level. This is because they provide a formal framework to describe contractual commitments and their relationship to partners' collaborations, transactions and processes. They also help to identify differences in local business context.

Recently, REA Enterprise Modeling Framework has been adopted as a central part of business models in ebXML.

The conclusion of ECIMF project is therefore that the application of this or similar framework is required for proper understanding of business-related constraints of integration scenarios. We recommend that further work be expended to formalize this aspect of interoperability, and especially how it influences the interoperability of technical infrastructures.

5.2 Semantic Interoperability

Today there are islands of well-defined semantics for use in e-commerce, such as universal classification schemas (EAN/UCC, UNSPSC …) and standard e-commerce frameworks (RosettaNet, OAGIS, ebXML, xCBL …).

But there is no generally available, overall and unified business semantics across existing standards. Similar business concepts are being expressed differently, using different semantic depth, which results in ambiguous and overlapping concepts when considered in an integration scenario. This in turn leads to drastic increase in complexity and cost of integration. This also prevents ad-hoc collaboration scenarios between partners using different e-commerce frameworks. Well-established older standards will linger, so that this aspect of integration will not go away any time soon.

The ECIMF project group has identified the need for better and more effective methods for semantic mapping. Some of the most promising methods use upper-level shared ontologies – however, there is no such common unified ontology available at the moment. Readers are encouraged to review Annex 3, where this problem is discussed in depth.

Some of the existing projects are working intensively in this area, specifically:

- ISO TC/154 Basic Semantic Register: provides a cross-linked reference to key concepts across several existing e-commerce standards.
- ECIMF Semantic Mapping Tool: provides a prototype tool to facilitate semantic translation process, with use of shared ontology.
- OntoWeb projects: several projects, e.g. on ontology-based integration of content standards (SIG1), and industrial applications of ontologies (SIG4)

and other similar projects. However, there is still much to be done before the average e-commerce user begins to benefit from this work.

The ECIMF project clearly identifies this issue as a fundamental integration problem, and recommends both further basic research into efficient methods of semantic mapping, and a development of upper-level shared e-commerce ontology for the purpose of such mapping.

5.3 Interoperability of Business Processes

The ECIMF project has identified the need to reconcile incompatible definitions of business processes, as specified by different e-commerce frameworks.

Although good and comprehensive models for business process modeling exist (e.g. the one developed by UN/CEFACT ebXML project), there is little or no work being done on process mediation *across* standards. This is a very complex and non-obvious issue, which involves elements like transaction preservation, observing the timing constraints, compensation for failed transactions, legal consequences of failed transactions, partial fulfillment and others.

The ECIMF project recommends further research in this area. We also suggest that a separate, well-defined module (here referred to as Process Mediator) should be responsible for addressing these issues. Initial requirements and suggestions for possible architectures have been presented in this document.

Currently, the project members are aware of just one research project, which tries to address this integration aspect in a systematic way – the Process Broker project at Swedish Royal Institute of Technology (http://www.dsv.su.se/~pajo/processbroker/index.html), led by prof. Paul Johannesson.

5.4 Syntactic interoperability

The issue of syntax mapping is the most common aspect of interoperability being addressed today by software vendors. There are many existing software suites which concentrate mainly on this aspect, while offering only very limited functionality in all other integration aspects, as identified above.

Unfortunately, as the ECIMF project concludes, interoperability of message formats and transport protocols is also the last issue to be addressed when implementing integration solutions, and probably the most straightforward – that is, as soon as all other constraints (semantic and dynamic) are well understood. This low-level mapping quickly becomes very complex and difficult to maintain, if it is not driven by underlying higher-level models.

Therefore we recommend that vendors of integration software suites should concentrate on development of model-driven tools for system integration, taking into account the high-level e-commerce models being developed by recognized standard bodies and industry forums (such as UN/CEFACT ebXML, RosettaNet, OMG, OAG, UBL and others).

5.5 Software tools

ECIMF Project has delivered a prototype tool that illustrates some of the principles developed by the project. This tool is available under liberal Open Source license (so called Mozilla License), and can be downloaded, together with Java source code, from http://www.ecimf.org .

In the course of working on the tool, one of the obstacles was the lack of machine-readable models of e-commerce frameworks. In some cases, like EDIFACT directories, even though such sources exist they require substantial development effort (or investment) to process the data due to their historically complex formats.

Therefore ECIMF project members recommend that efforts should be spent to prepare (or convert) machine-readable models of existing e-commerce frameworks in popular formats, such as XML, and XML applications like XMI and RDF.

If these existing sources of e-commerce concepts and models become easily available for processing and analysis in contemporary well-documented formats, for which parsers and development tools are freely available, then we should expect both a significant increase in reuse of this rich heritage, and a decrease in cost of software solutions.

6 Acknowledgments

The ECIMF project group gratefully acknowledges valuable input and support received from the following individuals and organizations:

- CEN/ISSS WS-EC members, especially:
    - Man-Sze Li, IC-Focus
    - Tony Fletcher, Choreology Ltd.
    - Prem Couture, Cyscom
    - Martin Bryan, SGML Centre
    - Andrzej Bialecki, Ampiro AB, Sweden
- Bob Haugen, Logistical Software LLC
- William McCarthy, Michigan State University
- Paul Johannesson from IT University, Department of Computer and Systems Sciences, Sweden
- Ambjörn Naeve, Mikael Nilsson and Matthias Palmér, Royal Institute of Technology (KTH), Sweden
- Heiner Stuckenschmidt, University of Bremen
- WebGiro AB, Sweden
- And other contributors

7 Annex 1 – Additional supporting materials for the Frameworks Integration Guideline

(Non-normative)

(NOTE: the parts in Times New Roman require still significant amount of work – both editing and conceptual. The parts in Arial seem to be mostly OK… The notes in italics mark the areas requiring additions and discussions.)

1. Business Context Matching

Business Context Matching	
Input	Traditional business knowledge, legal agreements between partners, industry specific rules, legal constraints, specific business goals, common business practices and codes of conduct
Output	Two Business Context Models for the integration scenario, defined in a set of UML diagrams (class, collaboration, activity), and an analysis of their matching (and any additional requirements on which the matching depends).
Alternative Procedures	
REA	REA ontology [REA], [REAont]
UMM	Business Requirements View in Chapter 9.2 of [UMM] (can be considered a specialized subset of REA)
EbXML	Business Process Analysis Worksheets and Guidelines [bpWS] (which are also based on REA principles)
SimpleREA	Described below.

1.1. Creating Business Context Models

Simple REA

Here we describe a simplified procedure useful for modeling of simple business cases (based on subset of REA, with relationships to UMM BRV and BTV; it should also be compatible with ebXML). As a result of the pragmatic process described below, you will create an economic exchange diagram, which provides a high-level overview of the parties involved in the business activities; and a value-chain diagram which puts this exchange in a context of the whole enterprise.

1. **Economic Exchange Diagram**

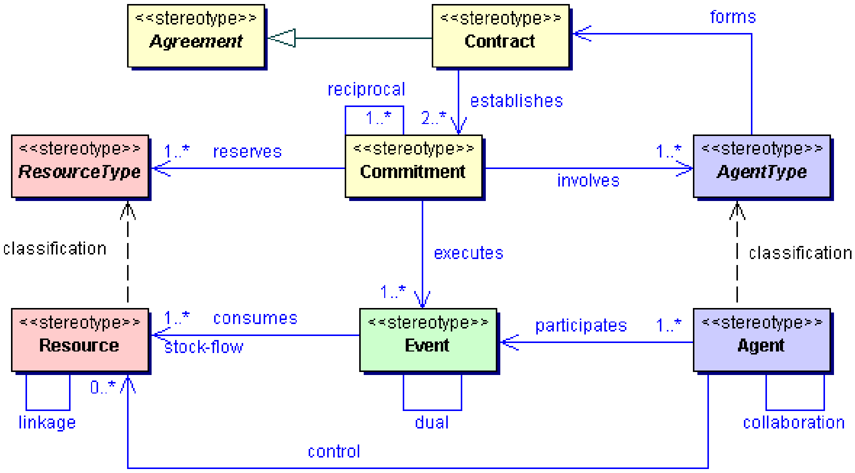
   1.1. Meta-model

   Describe the entities involved in the business case at hand, using the following terms (represented as UML stereotypes):

   - ***AgentType***: the role that a business partner plays in the scenario (e.g. buyer, seller, payer etc…). This is an abstract classification of the concrete Agents involved.
   - ***Agent***: if needed, specifies a concrete representative of a business party, which fulfills a given partner type (e.g. a sales clerk [= seller], a customer [= buyer]).
   - ***Agreement***: an agreement is an arrangement between two partner types that specifies in advance the conditions under which they will trade (terms of shipment, terms of payment, collaboration scenarios, etc.) A special kind of agreement (contract) commits partners to execute specific events, in which economic resources are exchanged.
   - ***Commitment***: an obligation to perform an economic event (i.e. transfer ownership of a specified quantity of a specified economic resource type) at some future point in time.
   - ***EventType***: an abstract classification or definition of an economic event. E.g. rental, service order, direct sales, production (of goods from raw materials), etc …
   - ***Event***: an economic event is the transfer of control of an economic resource from one partner type to another partner type. Examples would include the concrete sales, cash-payments, shipments, leases, deliveries etc. Economic Events usually cause changes in the state of each partner type (so called business events). Therefore they are directly related to (and determine) the transaction boundaries.
   - ***ResourceType***: an economic resource type is the abstract classification or definition of an economic resource. For example, in an ERP system, ItemMaster or ProductMaster would represent the Economic Resource Type that abstractly defines an Inventory item or product. Forms of payment are also defined by economic resource types, e.g. currency.
   - ***Resource***: if needed, specifies a quantity of something of value that is under the control of an enterprise, which is transferred from one partner type to another in economic events. Examples are cash, inventory, labor service and machine service. Contracts deal with resource types (abstract definitions), whereas events deal with resources (real entities). You may use this distinction if needed.

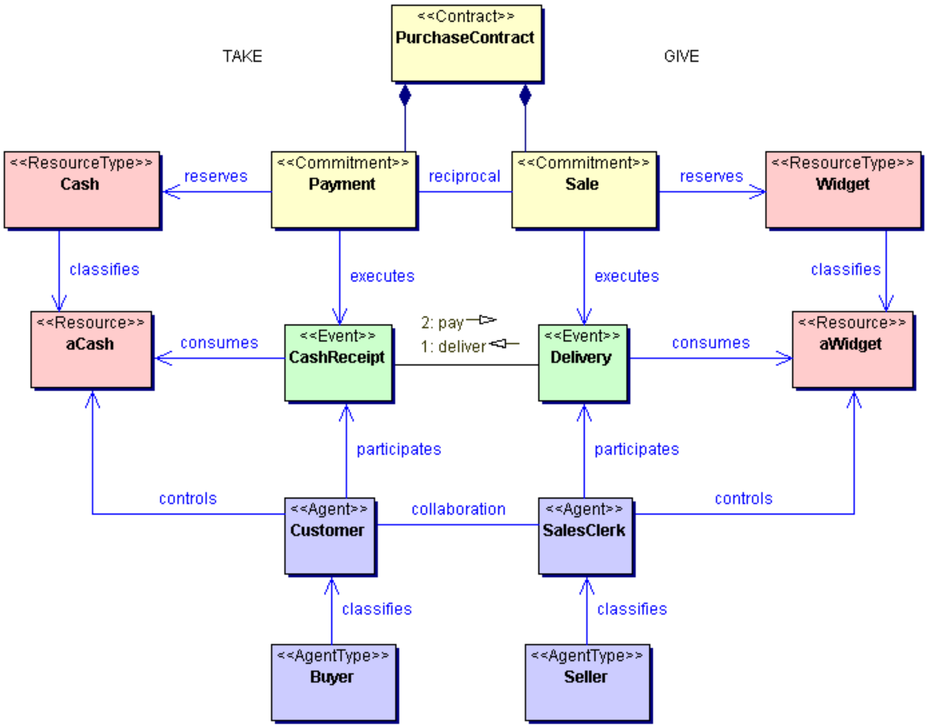
1.2. Meta-model and constraints

The meta-model for building the economic exchange diagrams is presented on the figure below:



The entities have been color-coded. The collaboration between Agents is realized with the BusinessTasks (collaboration protocol), which may be represented as UML activity diagrams.

1.3. Model example



The coloring schema on this diagram corresponds to that on the meta-model diagram.

Note: this diagram shows instances (concrete entities) of types specified above in the meta-model diagram. This is indicated by the UML stereotypes (labels in guillemots). Notice the two messages exchanged in this model – the first is to deliver, the second to pay (but it may be the other way around – an advance payment). This diagram helps us to identify the business transactions (in this case: {deliver, pay}), and also shows us the timing constraints (in this case: first deliver, then pay).

(NOTE: any useful real-life scenario would be more complicated. It could e.g. contain a catalog lookup, negotiation, shipment, blanket agreement, etc… This diagram serves therefore only as an illustration of the approach).

1.2. Checking the Business Context Rules

7.1.1.1.1 #1 Complementary roles

Parties need to play complementary roles (e.g. buyer/seller)

7.1.1.1.2 #2 Expected resources

The resources expected in the exchanges need to be equivalent to the ones expected by the other partner (e.g. cash for goods)

7.1.1.1.3 #3 Timing constraints

The timing constraints on events (commitment specification) need to be mutually satisfiable (e.g. down payment vs. final payment)

7.1.1.1.4 #4 Transaction boundaries

The sequence of expected business transactions needs to be the same (even though the individual business actions may differ)

2. Business Process Mediation

Business Process Mediation	
Input	Business Context models, other information on business processes supporting the business context.
Output	Business Process Models, Business Process Mediator Model for the integration scenario, defined in a set of diagrams (activity/business process, ECIMF process mediation diagram)
Alternative Procedures	
UMM + ECIMF-PM	UMM-BOV, and the ECIMF Process Mediation Model
UML-EDOC + ECIMF-PM	UML-EDOC, and the ECIMF Process Mediation Model
EbXML + ECIMF-PM	Business Process Specification Schema, and the ECIMF Process Mediation Model

2.1. Creating the Business Process Models

(to be completed...)

2.2. Creating the Business Process Mediation model

2.2.1. Check the Business Tasks alignment

- Identify **request and response messages**.
  (NOTE: this step will benefit from information collected in BOV and FSV models, if available (cf. [UMM]))
- Determine **legal obligations** boundaries: which interactions and messages bring what legal and economical consequences. This can be established based on the relationship to the business context diagram.
  (NOTE: needs more substance…)
- Determine the **business transaction boundaries**, rollback (compensation) activities and messages for failed transactions. The transaction boundaries can be better identified with the help of the business context diagram.
  (NOTE: needs more substance…)
- **Identify the differences** in message flow, by comparing message flows between requesting/responding parties for each business task.

2.2.2. Create the Mediation Elements between Business Tasks

- **Missing messages/elements**: are those that are present in e.g. Framework 1 business task B_x (we use the notation $F_1(B_x)$ for that), but don't occur in the corresponding $F_2(B_y, B_z, \ldots)$. This is also true about the individual data elements, which may become available only after certain steps in the conversations, different for each framework. These messages and data elements will have to be created by the mediator, based on already available data from various sources, such as:
  - previous messages
  - configuration parameters
  - external resources

  and sent according to the expected conversation pattern.
- **Superfluous or misplaced messages/elements**: are those that don't correspond directly to any of the required/expected messages as specified in the other framework. Also, they may be required to arrive in different order. The mediator should collect them (for possible use of information elements they contain at some later stage) without sending them to the other party, or change the order in which they are sent. The business context diagram will help determine what kind of re-ordering is possible without breaking the transaction boundaries (it should be possible to change the order within the transaction boundaries without breaking their semantics, but not across them).
- **Different constraints** (time, transactional, legal…): this issue is similar in complexity to resolving the semantic conflicts (see below), and a similar approach could be taken.

  (NOTE: namely???)

3. Semantic translation *(to be completed)*
(NOTE: needs to be harmonized with the methodology section!!!)

- **Identify the key concepts** in use for message exchanges conducted according to each framework, within the context of the selected corresponding business tasks:
  - **For each message** in B_i identify the key indispensable information elements that decide about the success of the information exchange from the business point of view in each of the frameworks:

    $M_i(E_1, E_2, \ldots, E_n)$
  - **For each message** M_i in B_i, based on the framework model, identify the key concepts that these information elements represent. In terms of OO and UML modeling, use the information collected in the previous step to build an object diagram, where instances of classes represent the key concepts (perhaps already identified in the formal framework description) and properties take the values from the message elements:

    $M_i(C_1(E_1, E_2, \ldots), C_2(E_m, E_n, \ldots), \ldots, C_n(E_x, E_y, \ldots))$

    This notation means that each message M_i contains a set of key concepts (classes) – information elements, which decide the meaning of the message.
  - **Collect the key concepts** in a unique set:

    $F_1(C_1, C_2, \ldots, C_n, \ldots, C_x, \ldots, C_z)$

  (NOTE: this is a bottom-up approach. Needs to be re-worked to better reflect the overall top-down approach).
  (NOTE 2: this step corresponds to the process of building conceptual topology of frameworks F1 and F2, which are sets of conceptual neighborhoods [CID52]).
- **Collect more semantic data** about each concept, as expressed by each framework's specifications, in a form of properties and constraints:

  $C_i(p_1, p_2, \ldots, p_m, c_1, c_2, \ldots, c_x)$

  We introduce the notation P_i to denote a property with its accompanying constraints. Therefore we may express the above as follows:

  $C_i(P_1, P_2, \ldots, P_m, c_n, \ldots, c_x)$

  These additional semantic data will probably point to some obvious generalizations, which in turn may lead to reduction of the set of unique concepts.

(NOTE 1: The steps detailed above lead to creation of framework ontologies – or, in the language of [UMM], Lexicons with core components. Similarly, the process described below corresponds to finding a translation between ontologies [OB00] – although, since the ontologies are built from scratch here, the approach to use shared vocabulary may provide useful reduction in complexity (cf. [OB00]). The latter approach is similar to the process described in [ebCDDA] for discovery of domain components and context drivers).
(NOTE 2: the Business Operational View [UMM] model of the frameworks, if available, is a very appropriate source for this kind of information)
(NOTE 3: two concepts $F_1(C_x)$ and $F_2(C_y)$ may in fact represent one real entity – however, due to the different contexts in which they are described they may appear to be non-equal. Such cases will be resolved in the following steps)

- **Generate hypotheses about corresponding concepts** in the other framework:
  - Concepts are likely to correspond if they:
    - have similar properties
    - are similarly classified
    - play similar roles (similar relationships with other concepts, occur in similar contexts)
- **Test each hypothesis:**

Semantic Translation	
Input	Ontologies for each framework, containing the key concepts
Output	Semantic Translation rules, defining the correspondence between the key concepts
Alternative Procedures	
BUSTER	Approximate re-classification (described below)
Subsumption	Check the constraints on the properties, describe the differences in property specifications (such as scale, allowed values, code lists, classification) and check the correctness of classification based on the following criteria: • The **necessary conditions** for concept $F_i(C_x)$ is set of values/ranges of some of its properties that are true for all instances of that concept. Therefore, if a concept C_y doesn't display them, it cannot be classified as C_x. Necessary conditions help to rule out false correspondence hypotheses. • The **sufficient conditions** for concept $F_i(C_x)$ is a set of properties and constraints, when met automatically determine the concept classification. Sufficient conditions help us to identify the concepts that surely correspond because they show all sufficient conditions. Example: "TV-set" meets sufficient conditions for being a "house appliance". However, it fails to meet the necessary conditions for a "cleaning house appliance".
Anchor-PROMPT	
Cupid	
MOMIS	
Ontomorph	
Upper-level ontology labeling	*(using terms from upper-level ontology to label the concepts, and then prepare translation formulas based on the formal subsumption algorithms)*

Approximate re-classification

If the above steps result in well-defined rules of correspondence for most cases of the observed concept occurrence, the hypothesis can be considered basically true. It is probably not feasible to strive for exact solution in 100% cases – we may allow certain exceptions. There are several ways to determine the level of proximity:

- **Rough classification:** the concept definition can be treated as having its upper and lower bounds. The upper bound (the most precise) is necessary conditions, and the lower bound (the most general) is the sufficient conditions. We may declare that $F_1(C_x)$? $F_2(C_y)$ even when necessary conditions are not met, but sufficient ones are.
- **Probabilistic classification:** we can determine (based on e.g. available pre-classified data sets) the significance of each property on the result of classification, and so calculate the probability of entity belonging to a specific class.
- **Fuzzy classification:** for each property we define a fuzzy rule, which describes the level of similarity of the tested property. Then, the best match is defined when maximum number of rules gives positive results.

- **Other hypotheses**: if the hypothesis cannot be proven with a sufficient degree of certainty, other hypotheses need to be formulated and tested.
- **Possible difficulties** that may arise:
  - There is **no corresponding** concept: may be there are too many unknown properties to determine the corresponding concept in F_2, because in the context of F_1 they were irrelevant. In this case, the information required to find $F_2(M_x(C_y))$ needs to be supplied from elsewhere, based on properties of the real entities that

$F_1(M_i(C_j))$ and $F_2(M_x(C_y))$ refer to - we need to provide more semantics about the concepts than what is found in the framework specifications (usually from a human expert).

- There are **many corresponding** concepts, depending on which property we choose: we could arbitrarily choose the one that plays the most vital role from the business point of view – and choose which properties decide that an instance of a concept in F1 could be classified as an instance of corresponding concept in F2:

  $F_1(C_x(P_i))$? $F_2(C_y(P_j))$

  See also the section above on probabilistic classification.
- The **conflicts in property** constraints cannot be easily resolved. This case calls for help from the domain expert.
- **Describe the rules and exceptions** (if any), and in what contexts they occur.

(NOTE: there are three ways to address this problem, according to [OB00]:

- *Create a single global ontology, which will include concepts from both frameworks. Not feasible for even moderately complex cases.*
- *Create mappings between concepts in ontologies (this is the approach suggested above, although [OB00] warns again that it leads to very complex mappings)*
- *Using shared vocabulary, re-build the ontologies from scratch – the result will be somewhat automatically aligned. Then, prepare the translation rules, which should be now much simpler.)*

4. Syntax translation (to be completed)

- **Data element mapping**

  (NOTE: describe how the external formats can be mapped to internal representation …)

- **Message format mapping**

  (NOTE: describe how the message well-formedness rules can be satisfied. This may involve proactive "asking" for more information in order to satisfy the demands of a given message format…)

- **Message packaging mapping**

  (NOTE: describe how the message packaging [encoding, charset, MIME, etc] can be aligned)

- **Transport protocol mapping**
  - **Align packaging and transport** protocols, based on the specifications in each framework.
- *(to be continued…)*

8 Annex 2 – Example Architecture of ECIMF-compliant Toolkit

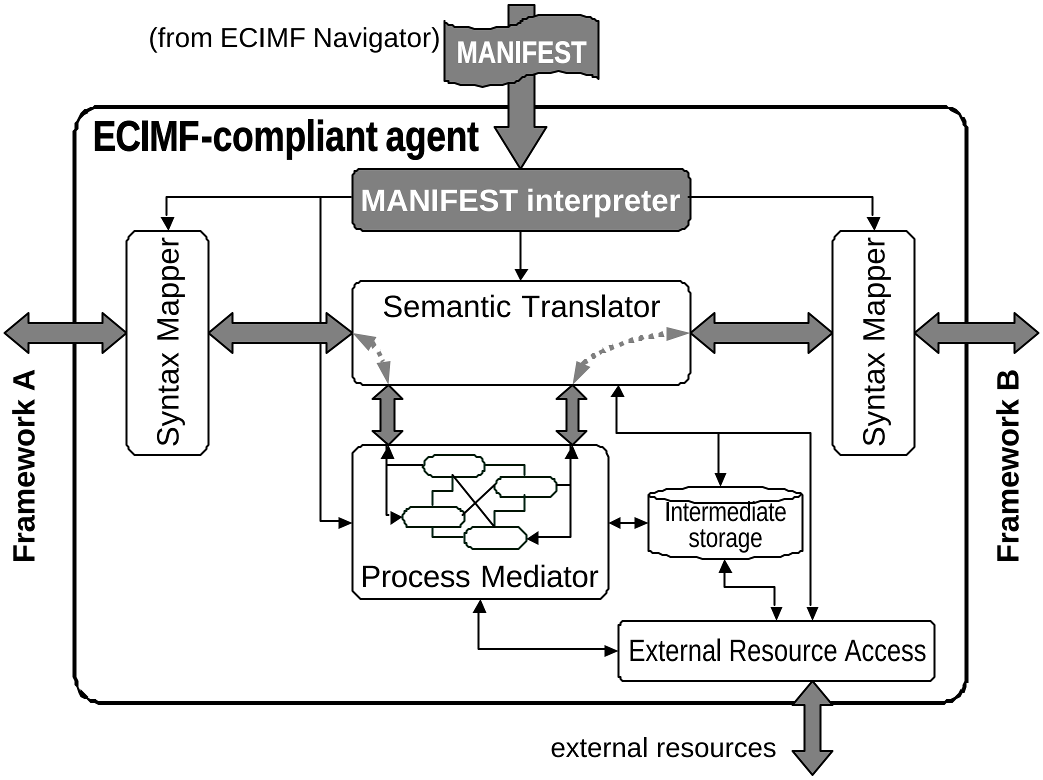


Figure 17 Example of ECIT (ECIMF-compliant agent) facilitating message exchange.

Figure above presents a block diagram of an ECIMF-compliant integration agent. The data flow (represented by thick gray arrows) goes first through the low-level data format adapters (named "Syntax Mappers"), then proceeds to the "Semantic Translators" module, and finally is controlled by the "Process Mediator". The "MANIFEST Interpreter", which uses the information provided in the "MANIFEST" specification prepared in the ECIMF Navigator, configures all these building blocks.

It is important to note that in this model, the ECIMF-compliant agent operates not only on the currently arrived data in the current message, but also uses some historical data stored in the intermediate storage, as well as the data available from external resources.

8.1.1 Syntax Mapper

The syntax Mapper is responsible for translating the message format and transport protocol to/from the internal model representation, which is then used by other modules. This could involve e.g. translating from EDI to XML, and then building an XML Document Object Model (DOM) tree as a representation of the incoming message. Further processing in the Semantic Translation module processes that internal model representation.

8.1.2 *Semantic Translator*

This module is responsible for changing the information model according to the translation rules, so that the information contained in the original message is understandable for the other party according to its (different) data model and meaning. This module operates only on the internal representation of the data.

8.1.3 *Process Mediator*

This module aligns the conversational patterns of each of the frameworks. It should be noted that this might require working not only with the currently received data in the message, but also with some historical data in the context of the same conversation. Also, there may be a need for using a given piece of information later during the same conversation, as specified by the differing message formats. For these reasons, the process mediator needs to use an intermediate storage, in which the data related to the context of current conversation may be kept.

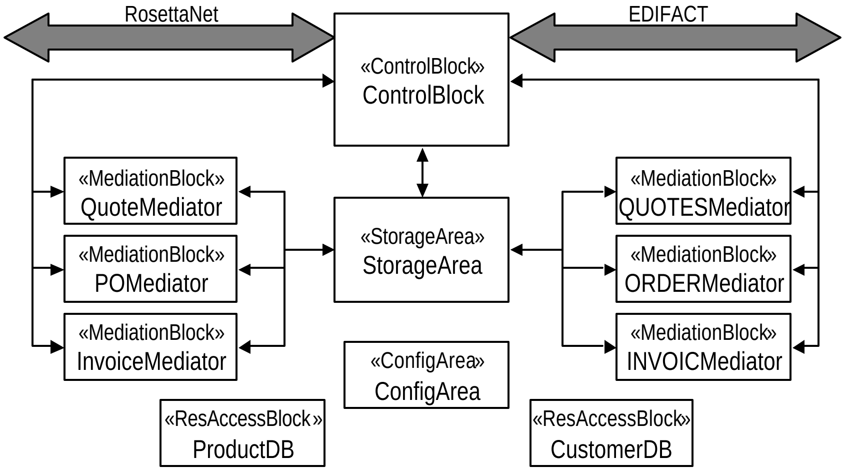


Figure 18 Process Mediator model.

Process Mediator needs to collect all information available from the input messages, complement them with information from other resources (e.g. external directories, configuration parameters), and generate appropriate output messages, which contain necessary information in order to complete given transactions, according to the target framework specification and within its timing constraints.

9 Annex 3 – MULECO: Multilingual Upper-Level Electronic Commerce Ontology

9.1 Editor’s note

The following material included in this Annex has been created as a draft proposal for a separate CEN/ISSS project. This material is far from being complete, and – since the project hasn’t been started due to the lack of resources – cannot be completed at this stage in this forum. However, in the opinion of EC Workshop members it provides a good starting point for anyone wanting to continue this work, and because it discusses the issues of semantic interoperability and the use of ontologies in e-commerce, it has been decided that it should be incorporated verbatim into the final ECIMF CWA as an informal Annex.

ECIMF Project Group gratefully acknowledges the efforts of Mr. Martin Bryan as the primary author of this initiative and editor of the following material.

9.2 What the project hopes to achieve

This CEN/ISSS Electronic Commerce Workshop project will research the most efficient means of developing a multilingual upper-level ontology for describing and identifying the relationships between electronic commerce applications and the ontologies used to describe them. In particular it will investigate how information related to business processes can be integrated with existing techniques for classifying businesses, their products and services.

There are many existing and proposed "electronic commerce ontologies". The vast majority have been defined monolingually, or in at most three or four languages, often from the same language group. The problem is that different trading partners tend to use different ontologies, and tend to prefer ontologies developed in their native language or in a "neutral" language, which is often English. It is, therefore, difficult to identify points of overlap between ontologies, and it is also difficult for people to find relevant terms in ontologies using their native language.

Figure 19 The relationship of MULECO to eCommerce Applications

The aim of MULECO is to develop a mechanism that will allow existing ontologies to identify their inter-relationships by identifying the relationships between themselves and a set of terms defined in a multilingual ontology that has been designed specifically to allow people to find terms using their native language. We realise that it is not possible, or desirable, to create and maintain a multilingual ontology that covers all terms used in all business applications in all European languages. What is needed is a way of classifying entries at the upper-most levels of existing ontologies in a form that takes account of the sort of terms used by people when they are trying to locate the term(s) they wish to use. To do this we need to extend existing business classification schemes to take account of things like business processes, variant names within different user communities, exclusion properties (e.g. no peanuts), etc. Such extensions need to be based on a well documented model that is based on properly researched linguistic characteristics, such as that provided by the Expert Advisory Group in Language Engineering Standards in The EAGLES Guidelines for Lexical Semantic Standards provided in Chapter 6 of *EAGLES LE3-4244: Preliminary Recommendations on Lexical Semantic Encoding -- Final Report* (http://www.ilc.pi.cnr.it/EAGLES96/EAGLESLE.PDF).

The MULECO project will develop an upper-level ontology, expressed as an extended network of industry descriptors, commercial terms and business roles, that will be recorded in a way that allows each entry to be addressed from other ontologies and applications by means of a Uniform Resource Identifier or an XML Path/Query.

The upper-level ontology will take as its start point existing standardized industry and process classification schemes, such as the International Standard for Industrial Classification (ISIC) used as the basis for the NACE classification of European business. The project will take note of the work being done by the IST CLAMOUR project to formally define such classification schemes. In particular it will extend currently used techniques for data classification, based on hierarchical classification of terms into broader and narrower meanings, by allowing for more complex relationships, in particular those relating to the relationships of wholes and parts which are vital to the mapping of the relationships between business processes. By defining a set of business relevant relationships between terms the project will allow classification hierarchies to become a controlled network of related words that forms an ontology rather than a classification scheme.

The ontology will be expressed in a language that provides the following functionality not currently found in electronic commerce ontologies based on languages such as RDF, OIL, KIF, etc, which the members of the CEN/ISSS Electronic Commerce Workshop feel are required to model different kinds of relationships between multilingual electronic commerce ontologies:

1. The ability to uniquely identify the domain (e.g. industry sector) in which each term is employed
2. The ability to formally record the meaning of the term within a particular domain
3. The ability to identify other domains in which the same meaning applies
4. The ability to record alternative terms that have the same meaning within the original domain
5. The ability to identify alternative terms used for the same meaning in other domains
6. The ability to identify an exactly equivalent term used in a different language
7. The ability to identify a nearly equivalent term used in a different language
8. The ability to identify terms that form a part of an object defined by a term
9. The ability to identify wholes that a term forms a part of
10. The ability to identify an opposite term or property (e.g. water-resistant/water-soluble)
11. The ability to record relationships between terms or properties
12. The ability to identify opposite relationships (e.g. isMother/isChild)
13. The ability to declare properties that record measurements
14. The ability to declare properties that record times
15. The ability to associate terms with specific points in process chains

Monolingual ontologies that are linked to the multilingual ontology will be able to make use of equivalences expressed in the multilingual ontology to extend their search potential. This will allow companies that have developed electronic commerce applications for a single country/language to extend their applications to other European countries and beyond without having to change their underlying data dictionaries. With the forthcoming extension of the European Single Market into Eastern Europe and the Mediterranean there will be an increasing need for tools that allow the creation and maintenance of complex multilingual business ontologies of the type to be developed by this project. The project will evaluate the problems associate d with developing multilingual ontologies, methodologies and techniques for overcoming them and the advantages to be gained from their use.

This project will incorporate and build on the concepts currently being developed to introduce monolingual ontologies into the Semantic Web. It will introduce such concepts into electronic commerce applications that are aimed at improving the flow of information between businesses within different language communities. At present most of the development work on the Semantic Web is postulated on the basis of using English language terms to identify the relationship between web resources and ontologies. Existing tools for applying the World Wide Web Consortium (W3C)'s Resource Description Framework (RDF) to the identification of related resources are generally postulated on the manual indexation of resources. Business applications require that this work be automated so that resource relationships can be identified automatically in a timely manner as part of business processes, without any human intervention. To be able to do this in a multilingual environment requires the use of a new generation of methodologies and tools. The project will seek to develop methodologies and tools for the creation and maintenance of multilingual ontologies, and for the querying of such ontologies.

The project hopes to:

1. Develop a methodology for expressing a general-purpose ontology for describing the full gamut of electronic commerce applications in multiple languages
2. Develop an open source tool to support the development and maintenance of a multilingual upper-level ontology
3. Populate a multilingual ontology with Internet-addressable terms for describing electronic commerce applications and services, and the relationships between them
4. Identify a set of existing electronic commerce ontologies and associate them with relevant terms in the multilingual upper-level ontology.
5. Input draft specifications into the European and international standardization process.

The results of the project will be reviewed by members of the CEN/ISSS Electronic Commerce Workshop and other relevant standardization organizations.

9.3 Existing Techniques

The following techniques have been studied as possible bases for MULECO:

- The EAGLES Guidelines
- Techniques for the Definition of Ontologies
  - IEEE Standard Upper-level Ontology (SUO)
  - DAML+OIL
- A Thesaurus Interchange Format in RDF
- XML Representation of ISO 13250 Topic Maps
- Unified Modeling Language (UML)

- The Object-Role Modeling (ORM)
- The Common Warehouse Metamodel (CWM) Business Nomenclature Package
- ISO 11179: Specification and Standardization of Data Elements
- ISO DIS 16642 :Terminological Markup Framework (TMF)
- ISO 704: Principles and methods of terminology
- The International Standard for Industrial Classification (ISIC)

9.4 The EAGLES Guidelines

The EAGLES project was concerned with Natural Language Processing (NLP). As such it had a very wide theme, and needed to cater for the large number of circumstances in which text is used. Many of its features were concerned with word d isambiguation in different contexts that are not directly applicable to the more limited applications for which business semantics are required. This section only discusses those features of the EAGLES Guidelines that are directly relevant for the description of business semantics.

The EAGLES Guidelines for Lexical Semantic Standards provided in Chapter 6 of *EAGLES LE3 - 4244: Preliminary Recommendations on Lexical Semantic Encoding -- Final Report* (http://www.ilc.pi.cnr.it/EAGLES96/EAGLESLE.PDF) points out that:

> "Hierarchical networks [describing hyperonym/hyponym (broader/narrower term) relationships] are very powerful structures because classifications at the top can be inherited to large numbers of word meanings that are directly or indirectly related to these top levels."

and

> "to achieve consistency in encoding hyponymy relations, the best approach is to build the hierarchy top down starting from a limited set of tops or unique beginners … Having an overview of the classes, even at a very high level, makes it possible to more systematically check the possible classes. Furthermore, a systematized top level makes it easier to compare and merge different ontologies."

Business semantics will need someone to develop a top level hierarchy suitable for business uses if they are to be able to interoperate.

As is pointed out in the EAGLES Guidelines, many thesauri cluster words that are related in an unstructured way. For example, the standardized medical thesaurus MESH contains the following entries related to transportation:

```
Transportation
... Aviation
... ... Aircraft
... ... ... Air Ambulances
... ... Space Flight
... ... ... Extravehicular Activity
... ... ... Spacecraft
```

The terms `Space Flight` and `Extravehicular Activity` do not represent subclasses of transportation vehicles but are, rather, types of activities related to certain vehicles. Because of this, MESH can only be used to globally extract words that are related; it cannot be used to make inferences such as: all the things that can be used to transport people, goods, etc.

Ontologies of business semantics need to be structured in such a way that they can be used for making inferences.

Words can have different meanings in different contexts. A term that has more than one meaning is said to exhibit polysemy. Words that share the same meaning within a particular context are synonyms. Synonyms should be able to replace each other in stated contexts. If their replacement is not always possible they are referred to as near-synonyms. Near-synonyms have meanings that partially overlap each other. Terms that share the same parent hyperonym but do not overlap in meaning are known as co-homonyms. It is important that ontologies of business semantics be able to make these distinctions within the relationships they record.

Word-sense disambiguation is an important subtask for Information Retrieval, Information Extraction or Machine Translation. One of the key factors in disambiguation is the identification of the domain with which the relevant text is concerned. If you have identified the domains in which each meaning of a term applies you can disambiguate meanings by utilizing information relating to the domains of discourse within a resource.

While hyperonym/homonym relationships work for nouns they are not so useful for other parts of speech, which are generally harder to disambiguate. For most business related classification schemes, however, verbs and other parts of speech are of relatively low importance in identifying meaning. (Verbs identify relationships or actions: they can be useful to identify the role played by particular agents on particular objects. Roles can be classified to create thematic roles. Adjectives are used to describe properties of nouns, e.g. brown gloves. Adverbs, prepositions, conjunctions, etc, are not widely used in electronic business messages. Of key importance to business, however, are terms used for the quantification of measurements and for defining time.)

Many lexicons permit multiple hyperonyms (broader terms) to be associated with a homonym (narrower term). Three types of hyperonym have been identified within the EAGLES project: exclusive, conjunctive and non-exclusive. For exclusive hyperonyms one of a choice of meanings must be determined by context. Conjunctive hyperonyms allow more than one meaning to be associated with a given context. If either multiple meanings or a single meaning can apply in a given context the hyperonym is deemed to be non-exclusive.

The EAGLES-based EuroWordNet distinguishes between Entities, Concepts, Events and States. Each of these is further divided, with up to 5 levels of subdivision. A typical EuroWordNet entry has the form:

```
[ -ORTHOGRAPHY : horse
  -WORD-SENSE-ID : horse_1
  -BASE-TYPE-INFO : [ BASE-TYPE: ANIMAL
                      LX-RELATION: LX-HYPONYM]
                    [ BASE-TYPE: OBJECT
                      LX-RELATION: LX-HYPONYM]
  SYNONYMS : Equus_caballus_1
  HYPERONYMS : [HYP-TYPE: conjunctive
                 HYP-ID: animal_1]
                [HYP-TYPE: conjunctive
                 HYP-ID: equid_1]
                [HYP-TYPE: non-exclusive
                 HYP-ID: pet_1]
                [HYP-TYPE: non-exclusive
                 HYP-ID: draught_animal_1]
  HYPONYMS : [HYP-TYPE: disjunctive
              HYP-ID: mare_1]
```

```
            [HYP-TYPE: disjunctive
              HYP-ID: stallion_1]]
```

Meronymy is defined as a lexical part-whole relationship between elements. A good example is provided by human body parts. "Finger" is a meronym of "hand" which is a meronym of "arm" which is a meronym of "body". The "inverse relation" is called holonymy. “Body" is the holonym of "arm" which is the holonym of "hand" which is the holonym of "finger". The co-meronymy relationship is one between lexical items defining sister parts (arm, leg, head are co-meronyms of body). Meronymy is different from taxonomy because it does not classify elements by class. That is to say, the hierarchical structuring of meronymy does not originate in a hierarchy of classes (toes, fingers, heads, legs, etc, are not hierarchically related).

Not all meronyms are related to a single holonym. For example, "nail" is more general than its holonym "toes" as it can also be part of a finger as well. Cruse[8] introduced the notions of super-meronym ("nail" is a super-meronym of "toes") and hypo-holonym ("toes" is a hypo-holonym of "nail") to allow for this.

The EAGLES paper recommends that "any lexical semantic standard should record a simple binary relation of antonymy where possible between [opposite] word senses". For example, "north" is the antonym of "south", and vice versa.

All of the above techniques can usefully be applied to multilingual business ontologies.

The on-going work, within the ISLE project for the development of International Standards for Language Engineering (http://www.ilc.pi.cnr.it/EAGLES96/isle/), on a Multilingual ISLE Lexical Entry (MILE) will extend the EAGLES Guidelines to cover the relationships between entries in different languages.

9.5 Techniques for the Definition of Ontologies

An ontology is a particular system of categories that provides a certain vision of the world. In the simplest case, an ontology describes a hierarchy of concepts related by subsumption relationships (e.g. lower-level terms meet the criteria set for higher-level terms). An ontology is the general framework within which catalogues, taxonomies, terminologies, etc, may be organized.

The key ingredients that make up an ontology are a vocabulary of terms and a precise specification of what those terms mean. But ontologies also analyse the fundamental categories of *objects, their current state,* and whether they form a *part* or the *whole* of something else, as well as the relations between parts and the whole and their laws of dependence.

Recently ontologists have started to explore the potential of process- and task-based ontologies. Rather than trying to describe all the characteristics of a particular universe or business domain, these more limited, time-aware, ontologies seek to provide information that is relevant for the management of a particular process or the completion of a specified task. One advantage of taking this approach is that it is much easier to develop intelligent agents that can make inferences based on such specialized domain knowledge than it is to create general-purpose inference engines of the type typically employed in AI systems. At this stage the question of how to identify the relationship between ontologies developed for specific processes and tasks has not been explored. Because it includes facilities for identifying the processes used by a particular business domain, however, MULECO should make it easier for intelligent agents, as well as human users, to identify ontologies that are relevant to their particular processes and tasks.

[8] Cruse, D. A. 1986. *Lexical Semantics*. Cambridge University Press

A formal ontology is the result of combining the intuitive, informal method of classical ontology analysis with the formal, mathematical method of modern symbolic logic. Over the years a wide range of formal ontologies have been proposed. To make it possible for ontologies to exchange data a number of "knowledge representation languages" have been developed, including KIF, Ontolingua, SNePS, HOL and Conceptual Graphs. Of these the most influential seems to have been the Knowledge Interchange Format (KIF). The basis for the semantics of KIF is a conceptualization of the world in terms of objects and relations among those objects. There are nine types of terms in KIF -- individual variables, constants, character references, character strings, character blocks, functional terms, list terms, quotations, and logical terms.

9.5.1 IEEE Standard Upper-level Ontology (SUO)

KIF, which is in the process of being published as a US standard by ANSI (see http://logic.stanford.edu/kif/dpans.html), has been chosen by IEEE as the basis for a Standard Upper-level Ontology (SUO). This upper ontology is limited to concepts that are meta, generic, abstract and philosophical, and therefore are general enough to address (at a high level) a broad range of domain areas. As well as very high level constructs such Independent Entity and Relative Entity SUO will cover such things as Agents, Persons and Organizations, using KIF definitions of the form:

```
(subclass-of Agent Object)
(subclass-of Person Agent)
(subclass-of Organization Agent)
(subclass-of Publisher Organization)
(subclass-of University Organization)
(disjoint Person Organization)
(subclass-of LegalObligation InstitutionalObligation)
```

and constructs for basic business functions, such as:

```
(subclass-of Quantity SpatialForm)
(subclass-of Weight Quantity)
(subclass-of Arrangement Schema)
(subclass-of Number Arrangement)
(subclass-of Set Arrangement)
```

SUO will also define instances of particular relationships, using formulations such as:

```
(instance-of hasAnnotation BinaryRelation)
(nth-domain hasAnnotation 1 Object)
(nth-domain hasAnnotation 2 TextObject)
```

and

```
(instance-of subProcess BinaryRelation)
(nth-domain subProcess 1 Process)
(nth-domain subProcess 2 Process)
```

Definitions can be assigned to SUO concepts using documentation statement of the form:

```
(documentation Agent "An active animate entity that voluntarily initiates an
action.")

(documentation Arrangement "Mathematical structures that do not have spatial
dimensions: numbers, sets, lists, algebras, grammars, and the data structure of
computer science. Arrangement includes the subclasses whose names are derived
from _taxis_, the Greek word for "arrangement", including taxonomies and syntax.
All the syntactic forms in natural languages, programming languages, and
versions of symbolic logic are included under Arrangement.")
```

As was the case with the all-encompassing lexical approach proposed by EAGLES, the proposed Standard Upper-level Ontology is designed to cover all knowledge, and therefore starts with concepts that are at much too high a level for the integration of business processes. It would be more correct to call it the Standard Top-level Ontology as it is designed to encompass all ontologies, rather than provide an upper level for a set of ontologies that cover specific areas, of the type proposed for the Multilingual Upper-Level Electronic Commerce Language.

Note: MULECO is not designed to integrate all existing ontologies, or to provide a meta-schema for describing ontologies. It is strictly limited to providing a means of identifying the relationships between existing ontologies by providing them with a set of addressable shared terms that they can link their top-levels to.

9.5.2 DAML+OIL

The Ontology Inference Language (OIL) that has been adopted as part of the DARPA Agent Markup Language (DAML) is an application of the W3C Resource Description Framework (RDF). DAML+OIL (http://www.daml.org/2001/03/reference.html) divides the world up into objects, which are elements of DAML classes, and datatype values, i.e., values that come from XML Schema datatypes, like the integer 4.

In DAML+OIL an ontology is recorded using a set of definitions that define classes, subclasses, properties that connect classes and individual instances. Classes have names, descriptive documentation, statements of which class it creates a subclass of, and one or more constraining facets. Classes are allowed to have multiple superclasses, which are deemed to be conjunctive unless specifically defined as being disjoint. DAML+OIL properties are divided into two sorts, those that relate objects to other objects and those that relate objects to datatype values. The former belong to daml:ObjectProperty while the latter belong to daml:DatatypeProperty. Properties are defined as having ranges of permitted values. Multiple ranges can be applied to aproperty but then the value of the property must satisfy all range statements (they are conjunctive rather than disjoint, with only the intersection of all the statements being valid). Properties, but not their values, can be defined as being the inverse of each other

DAML Class definitions can be defined in multiple statements, as the following parts of a March 2001 DAML Class definition example illustrate:

```
<daml:Class rdf:ID="Person">
  <rdfs:subClassOf rdf:resource="#Animal"/>
  <rdfs:subClassOf>
    <daml:Restriction>
      <daml:onProperty rdf:resource="#hasParent"/>
      <daml:toClass rdf:resource="#Person"/>
    </daml:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <daml:Restriction daml:cardinality="1">
      <daml:onProperty rdf:resource="#age"/>
    </daml:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <daml:Restriction>
      <daml:onProperty rdf:resource="#shoesize"/>
      <daml:minCardinality>1</daml:minCardinality>
    </daml:Restriction>
  </rdfs:subClassOf>
</daml:Class>
<daml:Class rdf:about="#Person">
  <rdfs:comment>Every person is a man or a woman</rdfs:comment>
  <daml:disjointUnionOf rdf:parseType="daml:collection">
    <daml:Class rdf:about="#Man"/>
    <daml:Class rdf:about="#Woman"/>
  </daml:disjointUnionOf>
</daml:Class>
...
<daml:Class rdf:about="#Person">
  <rdfs:subClassOf>
    <daml:Restriction daml:maxCardinality="1">
      <daml:onProperty rdf:resource="#hasSpouse"/>
    </daml:Restriction>
  </rdfs:subClassOf>
</daml:Class>
...
<daml:Class rdf:about="#Person">
  <rdfs:subClassOf>
    <daml:Restriction daml:maxCardinalityQ="1">
      <daml:onProperty rdf:resource="#hasOccupation"/>
      <daml:hasClassQ rdf:resource="#FullTimeOccupation"/>
    </daml:Restriction>
  </rdfs:subClassOf>
</daml:Class>
```

DAML classes are a subset of the RDF Schema (RDFS) Class construct. The rdfs:SubclassOf element that forms its first level contents is extended by the use of the daml:Restriction definition. Whilst this leads to a more detailed definition of DAML classes it does mean that there is a confusion between classes of the type used for defining schemas in RDF and the types of categorization used to define an ontology.[9]

An instance of the DAML Class shown above might take the form:

```
<Person rdf:ID="Peter">
  <rdfs:comment>
  Peter is an instance of Person. Peter has shoesize 9.5 and age 46
  </rdfs:comment>
  <shoesize>9.5</shoesize>
  <age><xsd:integer rdf:value="46"></age>
</Person>
```

Each DAML ontology can have associated with it metadata that identifies what the ontology is about, the version of DAML being used, and other information relevant to the management of the ontology. Ontologies can import part or all of another ontology.

A typical DAML+OIL header takes the form:

[9] The classes used in programming are typically additive in nature, properties at a lower level being added to those at higher levels. Categories in ontologies, in contrast, are restrictive in nature, the properties at one level distinguishing subsets of the properties applicable at a higher level.

```
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
 xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
 xmlns:xsd ="http://www.w3.org/2000/10/XMLSchema#"
 xmlns:dex ="http://www.daml.org/2001/03/daml+oil-ex#"
 xmlns:exd ="http://www.daml.org/2001/03/daml+oil-ex-dt#"
 xmlns     ="http://www.daml.org/2001/03/daml+oil-ex#" >
<daml:Ontology rdf:about="">
 <daml:versionInfo>$Id: daml+oil-ex.daml,v 1.9 2001/05/03 16:38:38
                  mdean Exp $</daml:versionInfo>
  <rdfs:comment>
    An example ontology, with data types taken from XML Schema
  </rdfs:comment>
  <daml:imports rdf:resource="http://www.daml.org/2001/03/daml+oil"/>
</daml:Ontology>
```

9.5.3 A Thesaurus Interchange Format in RDF

The IST LIMBER project has prepared a paper defining *A Thesaurus Interchange Format in RDF* (http://www.limber.rl.ac.uk/External/SW_conf_thes_paper.htm) which is based on the techniques defined in ISO 2788:1986 *Documentation--Guidelines for the establishment and development of monolingual thesauri* (2nd ed., 1986) and ISO 5964:1985 *Documentation--Guidelines for the establishment and development of multilingual thesauri* (1985). The model defines RDF classes for Thesaurus Object, Concept, Top Concept, Term, Scope Note and Scope Note Type. Concepts can have the following properties: Classification Code, In Language Of, Has Scope Note, Is Indicated By, Concept Relation and Concept Equivalence. The Is Indicated By property has subproperties of Preferred Term and Used For. The Concept Relation property has subproperties of Broader Concept, Narrower Concept, Top of Hierarchy, Top Concept and Is Related To. The Concept Equivalence property has subproperties of Exact Equivalent, Inexact Equivalent, Partial Equivalent and One-to-many Equivalent. Scope Notes have the property of In Language Of and Has Type Of, where permitted types are General, Hierarchy, Translation, Editor and History. The following example illustrates how these terms are applied:

```
<rdf:Description rdf:ID="EN620">
 <rdf:type rdf:resource=
    “http://www.data-archive.ac.uk/Limber/RDF/thesaurus#Concept”/>
 <thes:ClassificationCode>EN620</thes:ClassificationCode>
 <thes:inLanguageOf rdf:resource=
            "http://www.data-archive.ac.uk/Limber/ISO639#en"/>
 <thes:PreferredTerm>
  <rdf:Description>
   <rdf:type rdf:resource=
    "http://www.data-archive.ac.uk/Limber/RDF/thesaurus#Term"/>
    <rdf:value>friends</rdf:value>
  </rdf:Description>
 </thes:PreferredTerm>
 <thes:UsedFor>
  <rdf:Description>
   <rdf:type rdf:resource=
      "http://www.data-archive.ac.uk/Limber/RDF/thesaurus#Term"/>
   <rdf:value>mates</rdf:value>
  </rdf:Description>
 </thes:UsedFor>
 <thes:hasScopeNote>
  <rdf:Description>
   <rdf:type rdf:resource=
    "http://www.data-archive.ac.uk/Limber/RDF/thesaurus#ScopeNote"/>
   <thes:inLanguageOf rdf:resource=
     "http://www.data-archive.ac.uk/Limber/ISO639#fr"/>
   <thes:hasTypeOf rdf:resource=
     "http://www.data-archive.ac.uk/Limber/RDF/thesaurus#Regular"/>
    <rdf:value>To be used only for Platonic relationships.
    </rdf:value>
 </rdf:Description>
 </thes:hasScopeNote>
 <thes:BroaderConcept rdf:resource="#EN983"/>
 <thes:NarrowerConcept rdf:resource="#EN329"/>
 <thes:RelatedConcept rdf:resource="#EN243"/>
 <thes:TopConcept rdf:resource="#EN345"/>
 <thes:ExactEquivalent rdf:resource="#FR620"/>
</rdf:Description>
```

9.5.4 XML Representation of ISO 13250 Topic Maps

The XML Topic Maps (XTM) specification provides a model and grammar for representing the structure of information resources used to define *topics*, and the associations (relationships) between topics. Names, resources, and relationships are said to be *characteristics* of topics. Topics can have their characteristics defined within *scopes* that limit the contexts within which the names and resources are regarded as meaningful. One or more interrelated documents employing this grammar is called a “topic map”.

A minimal topic, consisting of a base name and a single resource identified as an occurrence of the topic, could be defined as:

```
<topic id="hamlet">
  <instanceOf><topicRef xlink:href="#play"/></instanceOf>
  <baseName>
    <baseNameString>Hamlet, Prince of Denmark</baseNameString>
  </baseName>
  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#plain-text-format"/>
    </instanceOf>
    <resourceRef
     xlink:href="ftp://www.gutenberg.org/pub/1ws2610.txt"/>
  </occurrence>
</topic>
```

An association representing the relationship between Shakespeare and the play *Hamlet* might look like this:

```
<association>
  <instanceOf><topicRef xlink:href="#written-by"/></instanceOf>
  <member>
    <roleSpec><topicRef xlink:href="#author"/></roleSpec>
    <topicRef xlink:href="#shakespeare"/>
  </member>
  <member>
    <roleSpec><topicRef xlink:href="#work"/></roleSpec>
    <topicRef xlink:href="#hamlet"/>
  </member>
</association>
```

Within topic maps, scopes establish the contexts in which a name or an occurrence is assigned to a given topic, and the context in which topics are related through associations. Any topics having the same base name in the same scope implicitly refer to the same subject and therefore should be merged.

XTM, unlike the underlying ISO standard, privileges two types of association: class-instance, and superclass-subclass. It fails, however, to follow the ISO standard in permitting the assignment of user-defined facets to provide multi-dimensional views of topic maps.

9.5.5 The Unified Modeling Language (UML)

UML is the main technique used for modelling business processes. It forms the basis of the UN/CEFACT Modeling Methodology (UMM), Version 10 of which can be found at http://www.gefeg.com/tmwg/n090r10.htm. UMM forms the basis for modelling business processes within the ebXML/ebTWG initiative to establish a new generation of business messaging services that is compatible with XML.

An overview of the ebXML process is provided in the following UML diagram:

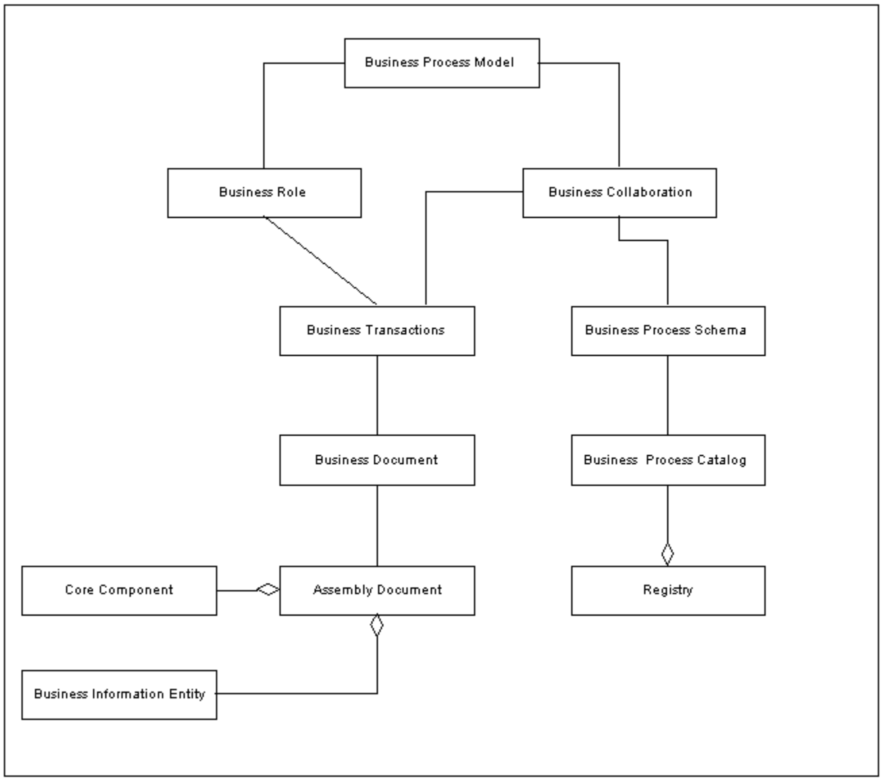
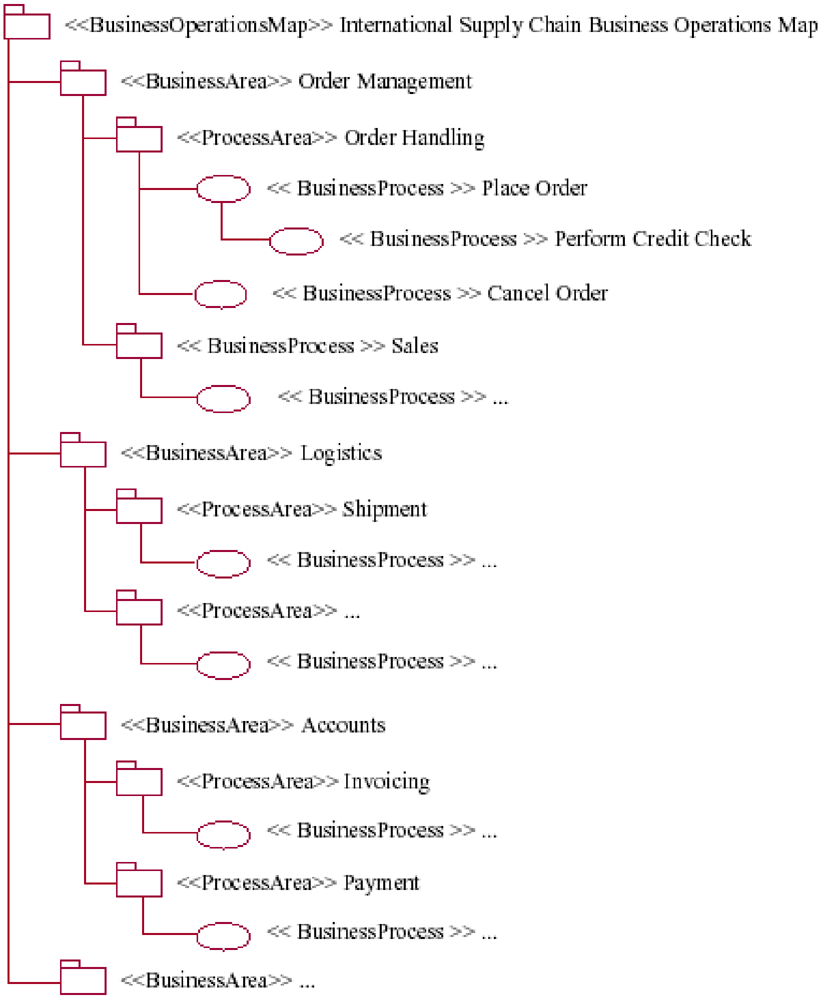


Figure 20 Core ebXML concepts

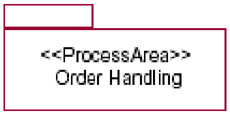
UML has problems with the correct definition of sequence, an important feature of business process management, but does allow choice and multiple triggering inputs to be clearly identified.

From a linguistic point of view a UML model can be thought of as a set of *entities* (things that have names which are nouns) that are linked by *relationships* (lines whose names are verbs). Entities have *properties* (whose values provide adjectives that qualify the nouns) and relationships have *qualifications* (whose values provide adverbs that qualify the verbs).

A UMM Business Operations Map (BOM), which forms the basis for ebXML business models, has the following typical representation:


<<BusinessOperationsMap>> International Supply Chain Business Operations Map
<<BusinessArea>> Order Management
<<ProcessArea>> Order Handling
<< BusinessProcess >> Place Order
<< BusinessProcess >> Perform Credit Check
<< BusinessProcess >> Cancel Order
<< BusinessProcess >> Sales
<< BusinessProcess >> ...
<<BusinessArea>> Logistics
<<ProcessArea>> Shipment
<< BusinessProcess >> ...
<<ProcessArea>> ...
<< BusinessProcess >> ...
<<BusinessArea>> Accounts
<<ProcessArea>> Invoicing
<< BusinessProcess >> ...
<<ProcessArea>> Payment
<< BusinessProcess >> ...
<<BusinessArea>> ...

where the following example definitions are used to define the Order Handling Process Area and Place Order Business Process:



CategorySchema:	UN/CEFACT's Ontology
Category:	UN.CEFACT.SupplyChain.OrderManagement.OrderHandling
Objective:	Facilitate key activities relative to order handling between buyer/seller for products and services.
Scope:	Limited to order handling activities, excluding sales.
Business Justification:	Facilitation of order handling activities.

Precondition:	A contract or a framework agreement exists
BeginsWhen:	The customer recognises a need for a products and contacts a seller.
Definition:	The customer recognises a need for a product and places an Order under an established contact or a framework agreement. The supplier receives order and responds.
EndsWhen:	Ends when order conditions have been met or order is cancelled.
Exceptions:	• Seller is not able to deliver the goods. • Seller is not allowed to sell the goods (trade embargoes, government regulations, etc.) • …
Postcondition:	The customer has received positive acknowledgement of product scheduled delivery.

The Centre for User-oriented IT Design (CID) at the Swedish Royal Institute of Technology (KTH) have developed a techniqu e for generalizing UML models to provide Unified Language Modeling (ULM) that allows formal models to be expressed in terms that are easily understood by businesses. The following diagrams summarize this technique:

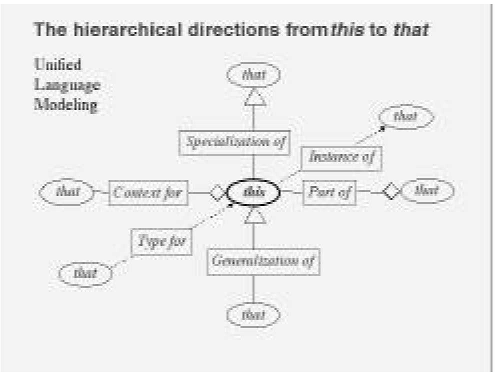


Figure 21 The basic principles for Unified Language Modeling

Using this technique you can understand that:

- The concept called car represents kind of vehicle
- The concept called vehicle is an abstraction of the concept called car
- The concept called wheel forms a part of a car
- A car has one or more wheels
- A specific car (:car) is an instance of the car concept
- A specific wheel (:wheel) is an instance of the wheel concept
- A specific wheel is a part of a specific car
- A specific car is a kind of vehicle

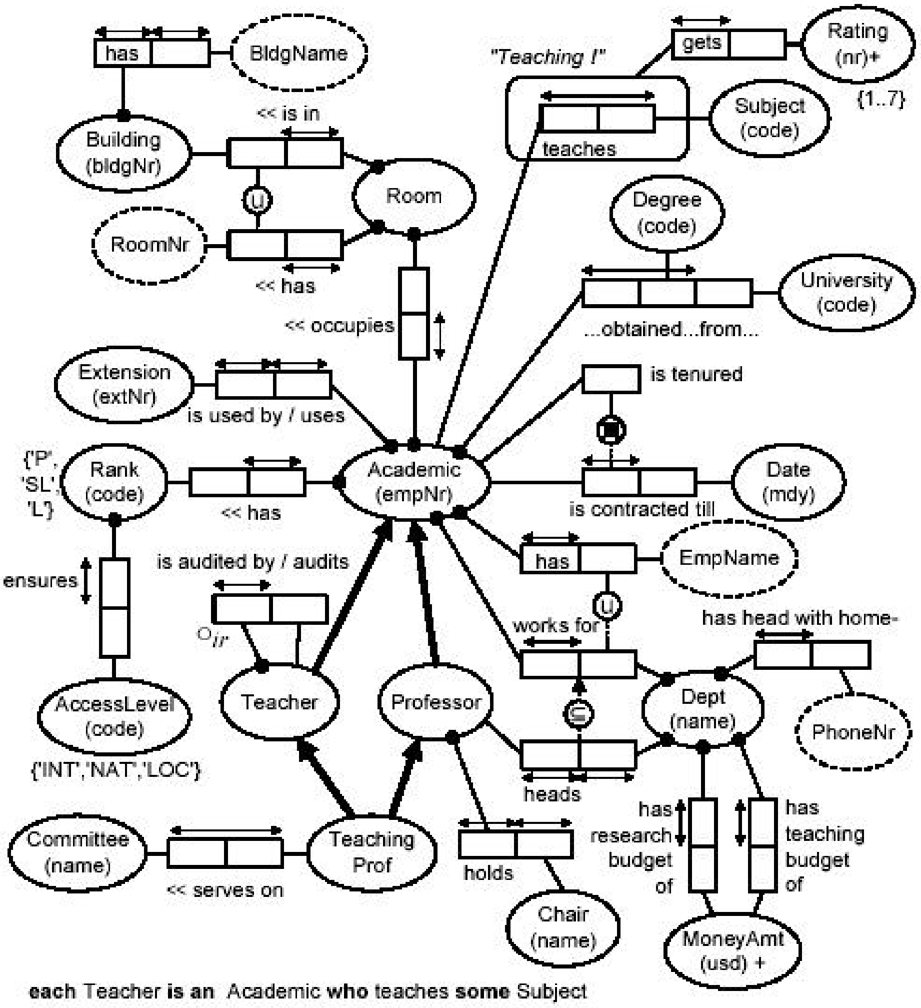
9.5.6 The Object-Role Modeling (ORM)

An alternative to UML is the Object-Role Modeling (ORM) promoted by Microsoft. ORM's *conceptual schema design procedure* (CSDP) focuses on the analysis and design of data. The conceptual schema specifies the information structure of the application: the *types of fact* that are of interest; *constraints* on these and *derivation rules* for deriving certain facts from others. The stages involved are:

1. Transform familiar information examples into elementary facts, and apply quality checks
2. Draw the fact types, and apply a population check
3. Check for entity types that should be combined, and note any arithmetic derivations
4. Add uniqueness constraints, and check arity of fact types
5. Add mandatory role constraints, and check for logical derivations
6. Add value, set comparison and subtyping constraints
7. Add other constraints and perform final checks

Object-Role Modeling is so-called because it views the world in terms of objects playing roles. Facts are assertions that objects play roles. An *n*-ary fact has *n* roles. It is not necessary that the roles be played by different objects.

A typical ORM diagram has the form:



each Teacher **is an** Academic **who** teaches **some** Subject
each Professor **is an** Academic **who** has Rank 'P'
each TeachingProf **is both a** Teacher **and a** Professor

* Dept employs academics of Rank in Quantity **iff** Quantity =
count each Academic **who** has Rank **and** works for Dept

* **define** Extension provides AccessLevel **as**
Extension is used by **an** Academic **who** has **a** Rank **that** ensures AccessLevel

Figure 22 ORM diagram.

9.5.7 The Common Warehouse Metamodel (CWM) Business Nomenclature Package

The following diagram summarizes the parts of the Open Management Group’s Common Warehouse Metamodel (OMG CWM) Business Nomenclature Package:

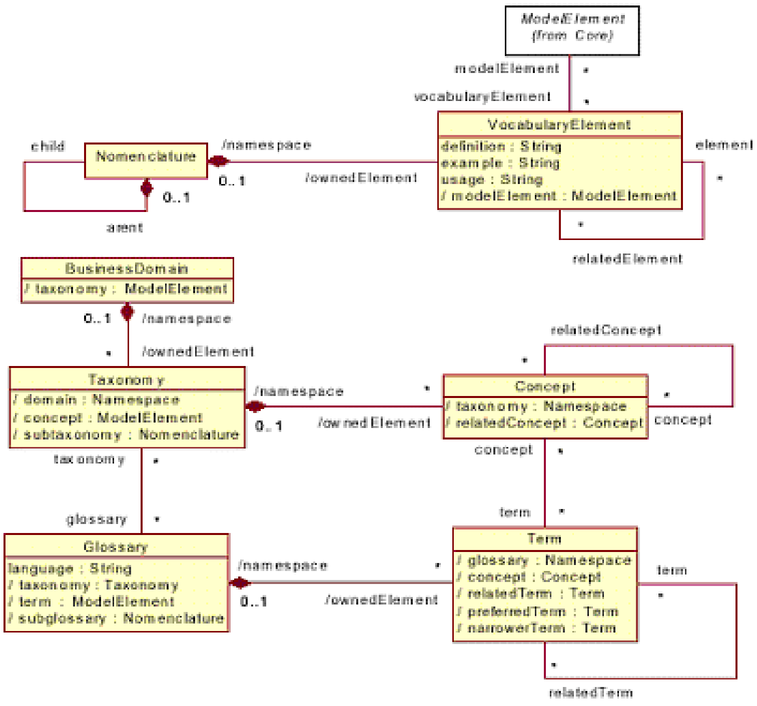


Figure 23 OMG's CWM core concepts.

The OMG model considers Taxonomies to consist of a number of Concepts, which may or may not have Related Concepts. A Taxonomy may be related to a Glossary, which contains one or more Term, which may have a Preferred Term, a number of Related Terms and one or more Narrower Terms. Terms can be related to Concepts in a Taxonomy.

9.5.8 ISO 11179: Specification and Standardization of Data Elements

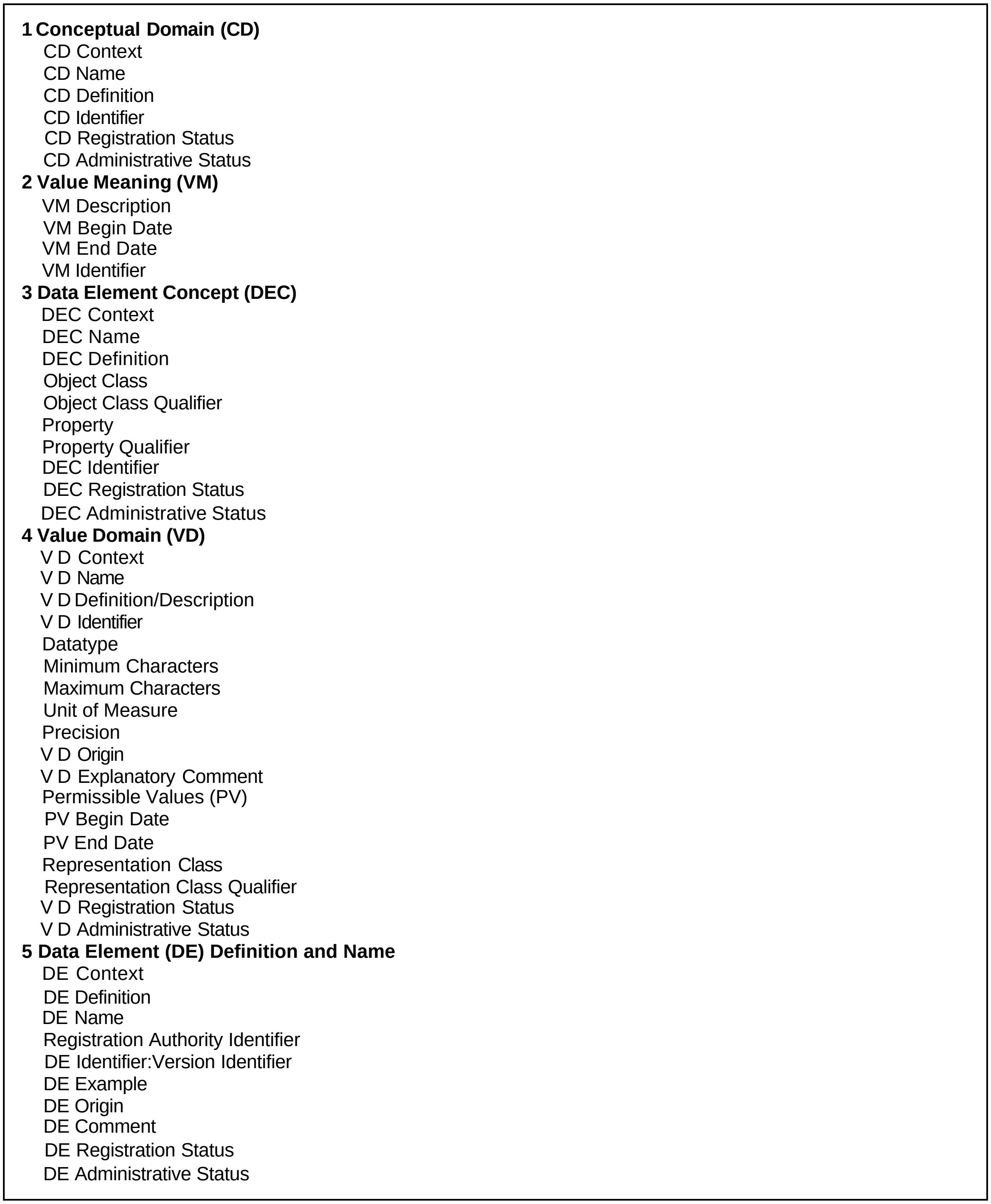
As noted in ISO DTR 20943, *Procedures for achieving metadata registry (MDR) content consistency*, the two types of abstraction of most interest to data element development are *specialization/generalization* and *decomposition/aggregation*.

Specialization/generalization is a relationship between two classes, where all items in one (subclass) are also in the other (superclass). Decomposition/aggregation relates an item to its parts. Decomposition may be described as "x is a part of y," or the *part-of* relationship. The reverse, aggregation, shows that y may be composed of x among other items.

The ISO/IEC 11179-3 metamodel does not provide for linking of data elements. A registration authority recording such data elements, however, might choose to extend the model to link data elements based on their layers of abstraction, including generalization to specialization, and other relationships. Linkages can occur in both vertical relationships (e.g., from general to more specific) and horizontal relationships (e.g., with equivalent layers of specialization). They can also be linked according to other relationships (e.g., data elements that are always used together). Vertical relationships are those where a specialized data element that has been registered for a particular purpose is related to a generalized data element that is intended for a general purpose. Horizontal relationships are those where data elements with different names have equivalent definitions that represent the same layer of specialization, with equivalent data value domains.

When using a top-down approach to ISO 11179 data element registration, small amounts of data are often added to a registry in groups or rather than as individual data elements. When a classified group of data elements is to be added to the registry, the analyst might choose to identify the conceptual domains that are relevant to the group, consider their value meanings, and work down to data elements.

Figure below shows the Order of Registering Components for Top Down Registration of a Data Element defined in ISO DTR 20943.



1 Conceptual Domain (CD)

- CD Context
- CD Name
- CD Definition
- CD Identifier
- CD Registration Status
- CD Administrative Status

2 Value Meaning (VM)

- VM Description
- VM Begin Date
- VM End Date
- VM Identifier

3 Data Element Concept (DEC)

- DEC Context
- DEC Name
- DEC Definition
- Object Class
- Object Class Qualifier
- Property
- Property Qualifier
- DEC Identifier
- DEC Registration Status
- DEC Administrative Status

4 Value Domain (VD)

- V D Context
- V D Name
- V D Definition/Description
- V D Identifier
- Datatype
- Minimum Characters
- Maximum Characters
- Unit of Measure
- Precision
- V D Origin
- V D Explanatory Comment
- Permissible Values (PV)
- PV Begin Date
- PV End Date
- Representation Class
- Representation Class Qualifier
- V D Registration Status
- V D Administrative Status

5 Data Element (DE) Definition and Name

- DE Context
- DE Definition
- DE Name
- Registration Authority Identifier
- DE Identifier:Version Identifier
- DE Example
- DE Origin
- DE Comment
- DE Registration Status
- DE Administrative Status

Order of Registering Components for Top Down Registration of a Data Element

9.5.9 Terminological Markup Framework (TMF)

ISO TC37/SC3 has published a Draft International Standard (DIS 16642, December 2001) that defines a Terminology Markup Framework. This International Standard specifies a model that has been designed for the purpose of providing guidance on the basic principles for representing terminological data, as well as for describing specific terminological markup languages. It is based on the principles laid down in ISO 704:2000, *Principles and methods of terminology* and provides an annex that shows how the framework can be applied to ISO 12620, *Computer applications in terminology - Data categories*.

The relationship of the component parts of the TMF model is shown in Figure below:

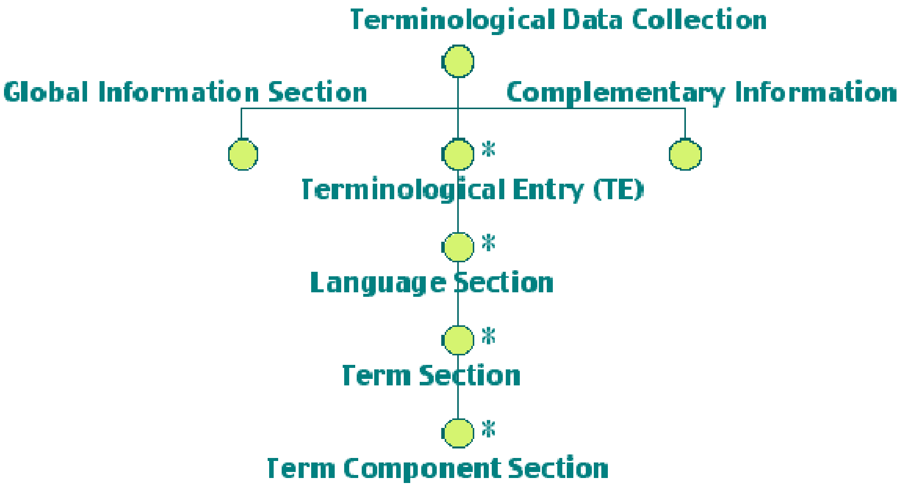

Terminological Data Collection
Global Information Section
Complementary Information
*
Terminological Entry (TE)
*
Language Section
*
Term Section
*
Term Component Section

Figure 24 Terminological Markup Framework (TMF) Metamodel

The component parts of this diagram a re:

- **TDC (Terminological Data Collection):**
  Top level *container* for all information contained in a terminology system. Generally used as a *container* for other containers in the system - may contain descriptive information such as, in XML, the validating schema that would be used.
- **GIS (Global Information Section):**
  Information that applies to all elements represented in a file, as opposed to information that may pertain to some, but not to all components of the file.
  Usually contains, for example, the title o f the (XML) file, the institution or individual originating the file, address information, copyright information, update information, etc.
- **CI (Complementary Information):**
  Usually contains, for example, textual bibliographical or administrative information residing in or external to the file, static or dynamic graphic images, video, audio, or virtually any other kind of binary data. Might also include references to other terminological resources or contextual links to related text corpora or to ontologies.
- **TE (Terminological Entry):**
  Information that pertains to a single concept. Usually contains, for example, the terms assigned to a concept, descriptive information pertinent to a concept, and administrative information concerning the concept. Can contain one or more language sections.
- **LS (Language Section):**

Contains all the terminological sections (TS) for a terminological entry that are used in a given language, as well as information such as definitions, contexts, etc. associated with that language or the te rms in that language.

- **TS (Term Section):**
  Information about terms, including definitions, contexts, etc, associated with the term.
- **TCS (Term Component Section):**
  Information about morphemic elements, words, or contiguous strings from which a polynomial term is formed. In languages such as French or Spanish it is important to be able to include information such as gender for the individual words used in constructing a multiword term because this information is necessary when using the term in texts.

The standard also defines a Generic Mapping Tool (GMT) using XML as the description language. The meta-model can be represented by means of a generic element <struct> (for *structure*) which can recursively express the embedding of the various representation levels of a terminological data collection. The role of each structural node within the meta -model is identified by means of a *type* attribute associated with the <struct> element, i.e., TDC, GIS, TE, CI, LS, TS, TCS. Basic information units associated with a stru ctural skeleton are represented using the <feat> (for *feature*) element. Compound information units are represented using the <brack> (for *bracket*) element, which can itself contain a <feat> element followed by any combination of <feat> elements and <brack> elements. The content model of the <feat> element can contain annotations expressed by means of an <annot> (for *annotation*) element. A *type* attribute can be used to reference an ISO 12620 data category or an equivalent user-defined data category.

An entry in the GMT format might take the form:

```
<?xml version="1.0" encoding="iso-8859-1"?>
<tmf>
 <struct type="TE">
  <feat type="entry identifier">ID67</feat>
  <feat type="subject field">manufacturing</feat>
  <feat type="definition">A value between 0 and 1 used in ...</feat>
    <struct type="LS">
     <feat type="language identifier">en</feat>
      <struct type="TS">
       <feat type="term">alpha smoothing factor</feat>
       <feat type="term type">fullForm</feat>
      </struct>
    </struct>
    <struct type="LS">
     <feat type="language identifier">hu</feat>
       <struct type="TS">
         <feat type="term">Alfa simítási tényezõ</feat>
       </struct>
    </struct>
 </struct>
</tmf>
```

An annex in the standard shows how the meta -model can be used to repersent an ISO 12200 *MARTIF-compatible format with specified constraints (MSC)* using the TMF Typed Element Style. Using this method the above entry takes the format:

```
<?xml version='1.0'?>
<!DOCTYPE martif SYSTEM "./MSCcoreStructureDTD-v-1-0.DTD.TXT">
<martif type='MSC' xml:lang='en' >
  <martifHeader>
    <fileDesc>
```

```
      <sourceDesc>
       <p>from an Oracle corporation termBase</p>
      </sourceDesc>
    </fileDesc>
    <encodingDesc>
     <p type='DCSName'>MSCdefaultXCS-v-1-0.XML</p>
    </encodingDesc>
  </martifHeader>
  <text>
   <body>
    <termEntry id='eid-Oracle-67'>
     <descrip type='subjectField'>manufacturing</descrip>
     <descrip type='definition'>A value between 0 and 1 used in ...
     </descrip>
     <langSet xml:ang='en'>
       <tig>
        <term tid='tid-Oracle-67-en1'>alpha smoothing factor</term>
        <termNote type='termType'>fullForm</termNote>
      </tig>
     </langSet>
     <langSet xml:lang='hu'>
      <tig>
       <term tid='tid-Oracle-67-hu1'>Alfa simítási tényezõ</term>
      </tig>
     </langSet>
    </termEntry>
   </body> </text>
</martif>
```

9.5.10 ISO 704: Principles and methods of terminology

The “Principles of term formation” defined in this standard includes the following:

- A term should be attributed to a single concept.
- Terms should be “transparent”: a term is considered to be transparent when the concept it designates can be inferred, at least partially, without a definition.
- The terminology of any subject field should provide a coherent terminological system corresponding to the concept system.
- Terms should adhere to familiar, established patterns of meaning within a language community.
- Terms should be as neutral as possible: avoid negative connotations.
- Terms should be as concise a s possible.
- Term formations that allow derivatives should be favoured.
- Terms should conform to the mrophological norms of the language.
- Native language expressions should be given preference over direct loans.

9.5.11 The International Standard for Industrial Classification (ISIC)

ISIC Version 3.0 (ISIC3) is the primary scheme used by governments throughout the world to classify business activity. It forms the basis of the Euopean NACE classification of EU economic activity. ISIC uses the following top-level hierarchy:

- A - Agriculture, hunting and forestry
- B - Fishing
- C - Mining and quarrying
- D - Manufacturing
- E - Electricity, gas and water supply

- F - Construction
- G - Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods
- H - Hotels and restaurants
- I - Transport, storage and communications
- J - Financial intermediation
- K - Real estate, renting and business activities
- L - Public administration and defence; compulsory social security
- M - Education
- N - Health and social work
- O - Other community, social and personal service activities
- P - Private households with employed persons
- Q - Extra-territorial organizations and bodies

Each of these subdivisions is further subdivided. For example, the Manufacturing subdivision is further subdivided into:

- 15 - Manufacture of food products and beverages
- 16 - Manufacture of tobacco products
- 17 - Manufacture of textiles
- 18 - Manufacture of wearing apparel; dressing and dyeing of fur
- 19 - Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear
- 20 - Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
- 21 - Manufacture of paper and paper products
- 22 - Publishing, printing and reproduction of recorded media
- 23 - Manufacture of coke, refined petroleum products and nuclear fuel
- 24 - Manufacture of chemicals and chemical products
- 25 - Manufacture of rubber and plastics products
- 26 - Manufacture of other non-metallic mineral products
- 27 - Manufacture of basic metals
- 28 - Manufacture of fabricated metal products, except machinery and equipment
- 29 - Manufacture of machinery and equipment
- 30 - Manufacture of office, accounting and computing machinery
- 31 - Manufacture of electrical machinery and apparatus
- 32 - Manufacture of radio, television and communication equipment and apparatus
- 33 - Manufacture of medical, precision and optical instruments, watches and clocks
- 34 - Manufacture of motor vehicles, trailers and semi-trailers
- 35 - Manufacture of other transport equipment
- 36 - Manufacture of furniture
- 37 – Recycling

These upper levels of the ISIC scheme show the typical problems that occur when you try to group together subjects into a single hierarchy. For example, the topmost level A shows that activities related to land use other than those associated with property have been grouped together, but the name applied to them does not reflect the reason for this grouping, but instead is simply the sum of the lower-level activities that make up the group.

Similar problems occur in the relationships between different levels. For example, the Manufacture of food products and beverages is part of the Manufacturing set of activities (D15) and not in any way associated with the production of the raw materials used therein, which generally comes under the heading Agriculture. Similarly the Manufacture of basic metals is not associated with the Mining and quarrying needed to obtain the raw materials for the processes.

These problems could be overcome by adopting a polyhierarchical system, which allowed the fact that the inputs from one chain came from the lower levels of a second chain, but to date most industrial classification schemes rely on single levels of nested classes.

It is unlikely that people will use the terms used in the ISIC headings as the basis for asking questions about classification schemes. It is, therefore, necessary to consider what terms users are likely to use to identify each of these terms. A more natural set of simple-to-understand top-level headings might include:

- Food
- Energy
- Raw materials
- Manufactured Goods
- Retailing
- Financial Services
- Transport
- Recreation
- Personal Services
- Property
- Civil Engineering
- Education
- Health and Social Services
- Public Administration
- Non-profit organizations

It should be noted that the ISIC listing is only available in three languages, English, French and Spanish. Translations into other languages would be needed to provide a truly multilingual classification scheme.

9.6 Proposed Approach

The ontology representation language should be expressed in XML so that individual components of it can be referenced as component parts of either a Unique Resource Indicator (URI), XML Path definition or XML Query.

The underlying structure of the XML should be based on the concepts described in the EAGLES framework, but with alternative forms of element names based on typical business renditions of technical terms (e.g. BroaderTerm in place of Hypernym). The terms to be adopted form EAGLES, and their equivalent business terms are shown in the following table:

Linguistic Terminology	Ontological Terminology	Business Terminology
Phrase	Concept	Term/Name
Hypernym	Superclass	Broader Term
Holonym	Subclass	Narrower Term
Synonym	Synonym	Alternative Term

Near-Synonym		Partially Matches / Includes
		No Equivalent*
Holonym		Forms Part Of
Meronym		Has Part / Has Process
Antonym		Opposite
	Restriction	Constraint Excludes
	Association	UsedBy RelatedTo ReverseOf
Measurement		MeasurementsRequired MeasurementType PermittedUnit
Time		TimesRequired TimeType

* Indicates that a particular language has no matching term

Entries should be provided with metadata which is defined by reference to existing sources of information or by use of standardized metadata descriptors. Each term must be assigned to at least one subject domain, ideally by linking it to a standardized domain identified within ISIC.

These terms can be used to create the following XML DTD:

```
<?xml version="1.0" encoding="UTF-8"?>
<!--DTD for defining the Multilingual Upper-Level Electronic Commerce Ontology
(MULECO)
© Martin Bryan of The SGML Centre (www.sgml.u-net.com) 25th January 2002 -->
<!--Entity used to indicate that attribute value should be expressed as a valid
URL (or as a pointer to an ID assigned in the ontology).-->
<!ENTITY % URL "CDATA">
<!ELEMENT Ontology (Term|TimeType|MeasurementType|PermittedUnit)+>
<!ATTLIST Ontology
  Region %URL; #IMPLIED
  Industry %URL; #IMPLIED
  Process %URL; #IMPLIED
>
<!ELEMENT Term (SubjectDomain+, Definition+, Name+, NoEquivalent*,
 (AlternativeTerm | Synonym)*, (PartiallyMatches | Includes |
 NearSynonym)*, (BroaderTerm | Hypernym)*, (NarrowerTerm | Hyponym)*,
 (FormsPartOf | Holonym)*, (HasPart | HasProcess | Meronym)*,
 (Opposite | Antonym)*, (Restriction | Constraint | Excludes)*,
 (UsedBy | RelatedTo | ReverseOf)*, (MeasurementsRequired |
 TimesRequired)*)>
<!ATTLIST Term
 ID ID #REQUIRED
 RecordedBy CDATA #REQUIRED
 Organization CDATA #IMPLIED
 WhenRecorded CDATA #REQUIRED
>
<!--Pointer to subject domains that this term is used within.
 (Subject domains may be defined in a separate onotology.)-->
<!ELEMENT SubjectDomain (Label*)>
<!ATTLIST SubjectDomain
 xlink:href  %URL; #REQUIRED
 LabelsID     ID    #IMPLIED
 LabelsRef    IDREF #IMPLIED
>
```

```
<!--Formal definition of term. Must be at least one. If more than one
 the values of their xml:lang attributes must differ.-->
<!ELEMENT Definition (#PCDATA | html:p)*>
<!ATTLIST Definition
 xml:lang CDATA #IMPLIED
 RecordedBy CDATA #IMPLIED
 Organization CDATA #IMPLIED
 WhenRecorded CDATA #IMPLIED
>
<!--Name by which term is known within specified SubjectDomains-->
<!ELEMENT Name (#PCDATA)>
<!ATTLIST Name
 xml:lang CDATA #IMPLIED
 RecordedBy CDATA #IMPLIED
 Organization CDATA #IMPLIED
 WhenRecorded CDATA #IMPLIED
>
<!--Identifies languages for which there is no equivalent term-->
<!ELEMENT NoEquivalent EMPTY>
<!ATTLIST NoEquivalent
 xml:lang CDATA #REQUIRED
 RecordedBy CDATA #IMPLIED
 Organization CDATA #IMPLIED
 WhenRecorded CDATA #IMPLIED
>
<!--Alternative name by which term is known in an identified domain
 (must share definition used for name exactly, otherwise it is a
 NearSynonym). (Technically these are referred to as Synonyms, but
 for convenience of use the AlternativeTerm element is defined as an
 equivalent.)-->
<!ELEMENT AlternativeTerm (#PCDATA)>
<!ATTLIST AlternativeTerm
 xml:lang CDATA #REQUIRED
 RecordedBy CDATA #REQUIRED
 Organization CDATA #IMPLIED
 WhenRecorded CDATA #REQUIRED
 SubjectDomain %URL; #IMPLIED
>
<!ELEMENT Synonym (#PCDATA)>
<!ATTLIST Synonym
 xml:lang CDATA #REQUIRED
 RecordedBy CDATA #REQUIRED
 Organization CDATA #IMPLIED
 WhenRecorded CDATA #REQUIRED
 SubjectDomain %URL; #IMPLIED
>
<!--Name applied in different language (or domain) that is near to
 current term but not exactly synonymous (i.e. does not share exactly
 the same definition). If the term overlaps with the related term
 then it is said to Partially Match it. If it is a subset of the term
 (without overlapping into other areas) then the parent term Includes
 the near synonym.-->
<!ELEMENT PartiallyMatches (#PCDATA)>
<!ATTLIST PartiallyMatches
 xml:lang CDATA #REQUIRED
 RecordedBy CDATA #REQUIRED
 Organization CDATA #IMPLIED
 WhenRecorded CDATA #REQUIRED
 SubjectDomain %URL; #IMPLIED
>
```

```
<!ELEMENT Includes (#PCDATA)>
<!ATTLIST Includes
 xml:lang CDATA #REQUIRED
 RecordedBy CDATA #REQUIRED
 Organization CDATA #IMPLIED
 WhenRecorded CDATA #REQUIRED
 SubjectDomain %URL; #IMPLIED
>
<!ELEMENT NearSynonym (#PCDATA)>
<!ATTLIST NearSynonym
 xml:lang CDATA #REQUIRED
 RecordedBy CDATA #REQUIRED
 Organization CDATA #IMPLIED
 WhenRecorded CDATA #REQUIRED
 SubjectDomain %URL; #IMPLIED
>
<!--Pointer to elements used to define terms whose broader meaning includes this
term. (Technically these are referred to as Hypernyms, but for convenience of
use the BroaderTerm alternative is a defined equivalent.)-->
<!ELEMENT BroaderTerm (Label*)>
<!ATTLIST BroaderTerm
 xlink:href %URL; #REQUIRED
 LabelsID     ID    #IMPLIED
 LabelsRef    IDREF #IMPLIED
>
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the term pointed to in an external ontology
should be entered as the contents of this element. If labels used here are
referenced from elsewhere the LabelsID  attribute must be used to name the set.
If labels used elsewhere are to be applied to this element the labelsRef
attribute must reference the relevant LabelsID.-->-->
<!ELEMENT Hypernym (Label*)>
<!ATTLIST Hypernym
 xlink:href %URL; #REQUIRED
 LabelsID     ID    #IMPLIED
 LabelsRef    IDREF #IMPLIED
 >
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the term pointed to in an external ontology
should be entered as the contents of this element. If labels used here are
referenced from elsewhere the LabelsID  attribute must be used to name the set.
If labels used elsewhere are to be applied to this element the labelsRef
attribute must reference the relevant LabelsID.-->
<!--Pointer to elements used to define terms whose narrower meaning is included
in this term. (Technically these are referred to as Hyponyms, but for
convenience of use the NarrowerTerm alternative is a defined equivalent.)-->
<!ELEMENT NarrowerTerm (Label*)>
<!ATTLIST NarrowerTerm
 xlink:href %URL; #REQUIRED
 LabelsID     ID    #IMPLIED
 LabelsRef    IDREF #IMPLIED
>
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the term pointed to in an external ontology
should be entered as the contents of this element. If labels used here are
referenced from elsewhere the LabelsID  attribute must be used to name the set.
If labels used elsewhere are to be applied to this element the labelsRef
attribute must reference the relevant LabelsID .-->
<!ELEMENT Hyponym (Label*)>
<!ATTLIST Hyponym
```

```
 xlink:href %URL; #REQUIRED
 LabelsID     ID    #IMPLIED
 LabelsRef    IDREF #IMPLIED
>
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the term pointed to in an external ontology
should be entered as the contents of this element. If labels used here are
referenced from elsewhere the LabelsID  attribute must be used to name the set.
If labels used elsewhere are to be applied to this element the labelsRef
attribute must reference the relevant LabelsID.-->-->
<!--Pointer to elements used to define terms that provide components used to
create the information set defined by term. (Technically these are referred to
as Meronyms, but for convenience of use the HasPart alternative is a defined
equivalent.)-->
<!ELEMENT HasPart (Label*)>
<!ATTLIST HasPart
 xlink:href %URL; #REQUIRED
 LabelsID     ID    #IMPLIED
 LabelsRef    IDREF #IMPLIED
>
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the term pointed to in an external ontology
should be entered as the contents of this element. If labels used here are
referenced from elsewhere the LabelsID  attribute must be used to name the set.
If labels used elsewhere are to be applied to this element the labelsRef
attribute must reference the relevant LabelsID.-->
<!ELEMENT HasProcess (Label*)>
<!ATTLIST HasProcess
 xlink:href %URL; #REQUIRED
 LabelsID     ID    #IMPLIED
 LabelsRef    IDREF #IMPLIED
>
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the term pointed to in an external ontology
should be entered as the contents of this element. If labels used here are
referenced from elsewhere the LabelsID  attribute must be used to name the set.
If labels used elsewhere are to be applied to this element the labelsRef
attribute must reference the relevant LabelsID.-->
<!ELEMENT Meronym (Label*)>
<!ATTLIST Meronym
 xlink:href %URL; #REQUIRED
 LabelsID     ID    #IMPLIED
 LabelsRef    IDREF #IMPLIED
>
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the term pointed to in an external ontology
should be entered as the contents of this element. If labels used here are
referenced from elsewhere the LabelsID  attribute must be used to name the set.
If labels used elsewhere are to be applied to this element the labelsRef
attribute must reference the relevant LabelsID.-->
<!--Pointer to elements used to define terms for which  this term forms a
component of the definition. (Technically these are referred to as Holonyms, but
for convenience of use the FormsPartOf alternative is a defined equivalent.)-->
<!ELEMENT FormsPartOf (Label*)>
<!ATTLIST FormsPartOf
 xlink:href %URL; #REQUIRED
 LabelsID     ID    #IMPLIED
 LabelsRef    IDREF #IMPLIED
>
```

```
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the term pointed to in an external ontology
should be entered as the contents of this element. If labels used here are
referenced from elsewhere the LabelsID  attribute must be used to name the set.
If labels used elsewhere are to be applied to this element the labelsRef
attribute must reference the relevant LabelsID.-->
<!ELEMENT Holonym (Label*)>
<!ATTLIST Holonym
 xlink:href %URL; #REQUIRED
 LabelsID     ID    #IMPLIED
 LabelsRef    IDREF #IMPLIED
>
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the term pointed to in an external ontology
should be entered as the contents of this element. If labels used here are
referenced from elsewhere the LabelsID  attribute must be used to name the set.
If labels used elsewhere are to be applied to this element the labelsRef
attribute must reference the relevant LabelsID.-->
<!--Pointer to elements used to define the exact opposite of this term.
(Technically these are referred to as Antonyms, but for convenience of use the
Opposite alternative is a defined equivalent.)-->
<!ELEMENT Opposite (Label*)>
<!ATTLIST Opposite
 xlink:href %URL; #REQUIRED
 LabelsID     ID    #IMPLIED
 LabelsRef    IDREF #IMPLIED
>
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the term pointed to in an external ontology
should be entered as the contents of this element.-->
<!ELEMENT Antonym (Label*)>
<!ATTLIST Antonym
 xlink:href %URL; #REQUIRED
 LabelsID     ID    #IMPLIED
 LabelsRef    IDREF #IMPLIED
>
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the term pointed to in an external ontology
should be entered as the contents of this element. If labels used here are
referenced from elsewhere the LabelsID  attribute must be used to name the set.
If labels used elsewhere are to be applied to this element the labelsRef
attribute must reference the relevant LabelsID.-->
<!--Pointer to documents that define constraints on the use of this term.
(Technically these are referred to as Restrictions, but for convenience of use
the Constraint alternative is a defined equivalent.)-->
<!ELEMENT Constraint (Label*)>
<!ATTLIST Constraint
 xlink:href %URL; #REQUIRED
 LabelsID     ID    #IMPLIED
 LabelsRef    IDREF #IMPLIED
>
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the term pointed to in an external ontology
should be entered as the contents of this element. If labels used here are
referenced from elsewhere the LabelsID  attribute must be used to name the set.
If labels used elsewhere are to be applied to this element the labelsRef
attribute must reference the relevant LabelsID.-->
<!ELEMENT Restriction (Label*)>
<!ATTLIST Restriction
 xlink:href %URL; #REQUIRED
 LabelsID     ID    #IMPLIED
 LabelsRef    IDREF #IMPLIED
>
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the term pointed to in an external ontology
should be entered as the contents of this element. If labels used here are
referenced from elsewhere the LabelsID  attribute must be used to name the set.
If labels used elsewhere are to be applied to this element the labelsRef
attribute must reference the relevant LabelsID.-->
<!--Pointer to components that may not be included in associated products or
processes. (Technically these are also Restrictions, but they are more specific
in application that the more generalized constraints definitions, for which no
formal language has yet been defined.)-->
<!ELEMENT Excludes (Label*)>
<!ATTLIST Excludes
 xlink:href %URL; #REQUIRED
 xml:lang CDATA #IMPLIED
 LabelsID     ID    #IMPLIED
 LabelsRef    IDREF #IMPLIED
 >
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the excluded item as defined in an external
ontology should be entered as the contents of this element. Where more than one
term applies the language of each term must be identified using the xml:lang
attribute. If labels used here are referenced from elsewhere the LabelsID
attribute must be used to name the set. If labels used elsewhere are to be
applied to this element the labelsRef attribute must reference the relevant
LabelsID.-->
<!--Pointer to term that makes use of this term.-->
<!ELEMENT UsedBy (Label*)>
<!ATTLIST UsedBy
 xlink:href %URL; #REQUIRED
>
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the term pointed to in an external ontology
should be entered as the contents of this element. If labels used here are
referenced from elsewhere the LabelsID  attribute must be used to name the set.
If labels used elsewhere are to be applied to this element the labelsRef
attribute must reference the relevant LabelsID. -->
<!--Pointer to term related to this term, together with name that defines the
relationship between this term and the current term.-->
<!ELEMENT RelatedTo (RelationshipName+, RelatedTerm)>
<!ATTLIST RelatedTo
 ID ID #REQUIRED
>
<!--Name used to describe relationship between two or more related terms.-->
<!ELEMENT RelationshipName (#PCDATA)>
<!ATTLIST RelationshipName
 xml:lang CDATA #IMPLIED
 RecordedBy CDATA #REQUIRED
 Organization CDATA #IMPLIED
 WhenRecorded CDATA #REQUIRED
 SubjectDomain %URL; #IMPLIED
>
<!--Where more than one relationship name is defined for a relationship the
language of each name must be identified using the xml:lang attribute. -->
<!--Term to which the current term is to be related.-->
<!ELEMENT RelatedTerm (#PCDATA)>
<!ATTLIST RelatedTerm
 xlink:href %URL; #REQUIRED
 xml:lang CDATA #IMPLIED
>
<!--Element contents can be empty if the URL pointed to is within the same
ontology. Otherwise the name of the excluded item as defined in an external
ontology should be entered as the contents of this element. -->
<!--Relationship (defined as part of another term) that this relationship is the
reverse of.-->
<!ELEMENT ReverseOf (RelationshipName+)>
<!ATTLIST ReverseOf
 RelatesTo  IDREF #REQUIRED
>
<!--Identification of measurements required to fully define term, process or
part-->
<!ELEMENT MeasurementsRequired (MeasurementTypeRef+)>
<!--Pointer to definition of measurement type required.-->
<!ELEMENT MeasurementTypeRef EMPTY>
<!ATTLIST MeasurementTypeRef
 MeasurementType  IDREF #REQUIRED
>
<!--Definition of measurement type consisting of the names by which the
measurement type is referred together with a list of units that are allowed to
be used to specify the measurement.-->
<!ELEMENT MeasurementType (MeasurementName+, PermittedUnitRef+)>
<!ATTLIST MeasurementType
 ID ID #REQUIRED
>
<!--Name used to request measurement.-->
<!ELEMENT MeasurementName (#PCDATA)>
<!ATTLIST MeasurementName
 xml:lang CDATA #IMPLIED
 RecordedBy CDATA #REQUIRED
 Organization CDATA #IMPLIED
 WhenRecorded CDATA #REQUIRED
 SubjectDomain %URL; #IMPLIED
>
<!--Where more than one measurement name is defined the language of each name
must be identified using the xml:lang attribute. -->
<!--Pointer to definition of the unit type required.-->
<!ELEMENT PermittedUnitRef EMPTY>
<!ATTLIST PermittedUnitRef
 Unit IDREF #REQUIRED
>
<!-- Definition of unit that can be used for measurement, consisting of the
names by which the unit type is referred together with the string(s) that is
used to qualify numbers defined using this unit (e.g. kg or lbs and ozs).-->
<!ELEMENT PermittedUnit (UnitName+,(UnitIdentifier|PermittedValue)+)>
<!ATTLIST PermittedUnit
 ID ID #REQUIRED
>
<!--Name used to request measurement.-->
<!ELEMENT UnitName (#PCDATA)>
<!ATTLIST UnitName
 xml:lang CDATA #IMPLIED
 RecordedBy CDATA #REQUIRED
 Organization CDATA #IMPLIED
 WhenRecorded CDATA #REQUIRED
 SubjectDomain %URL; #IMPLIED
>
<!--Where more than one name is assigned to the unit the language of each name
must be identified using the xml:lang attribute. -->
```

```
<!--String used to identify the use of number as a measurement or time unit.-->
<!ELEMENT UnitIdentifier (#PCDATA)>
<!ATTLIST UnitIdentifier
 Separator CDATA #IMPLIED
 Purpose CDATA #IMPLIED
>
<!--Where two or more sub-units share the same separator they must each be
assigned a different value for the Purpose attribute. Where it has not been
specified, the separator is presumed to be a space or line ending following the
contents of the element.-->
<!--String used to indicate specific value of measurement-->
<!ELEMENT PermittedValue (#PCDATA)>
<!ATTLIST PermittedValue
 xml:lang CDATA #REQUIRED
 RecordedBy CDATA #REQUIRED
 Organization CDATA #IMPLIED
 WhenRecorded CDATA #REQUIRED
 SubjectDomain %URL; #IMPLIED
>
<!--Identification of times required to fully define term, process or part-->
<!ELEMENT TimesRequired (TimeTypeRef+)>
<!--Pointer to definition of time type required.-->
<!ELEMENT TimeTypeRef EMPTY>
<!ATTLIST TimeTypeRef
 TimeType IDREF #REQUIRED
>
<!--Definition of time type consisting of the names by which the time type is
referred together with a list of units that are allowed to be used to specify
the time.-->
<!ELEMENT TimeType (TimeName+, PermittedUnitRef+)>
<!ATTLIST TimeType
 ID ID #REQUIRED
>
<!--Name used to request time.-->
<!ELEMENT TimeName (#PCDATA)>
<!ATTLIST TimeName
 xml:lang CDATA #IMPLIED
 RecordedBy CDATA #REQUIRED
 Organization CDATA #IMPLIED
 WhenRecorded CDATA #REQUIRED
 SubjectDomain %URL; #IMPLIED
>
<!--Where more than one time name is defined the language of each name must be
identified using the xml:lang attribute. -->
<!--Element used to label a link to a service that is not defined
multilingually.-->
<!ELEMENT Label (#PCDATA)>
<!ATTLIST Label
 xml:lang CDATA #REQUIRED
 RecordedBy CDATA #IMPLIED
 Organization CDATA #IMPLIED
 WhenRecorded CDATA #IMPLIED
 SubjectDomain %URL; #IMPLIED
>
<!--Element inherited from HTML specification.-->
<!ELEMENT html:p (#PCDATA)>
```

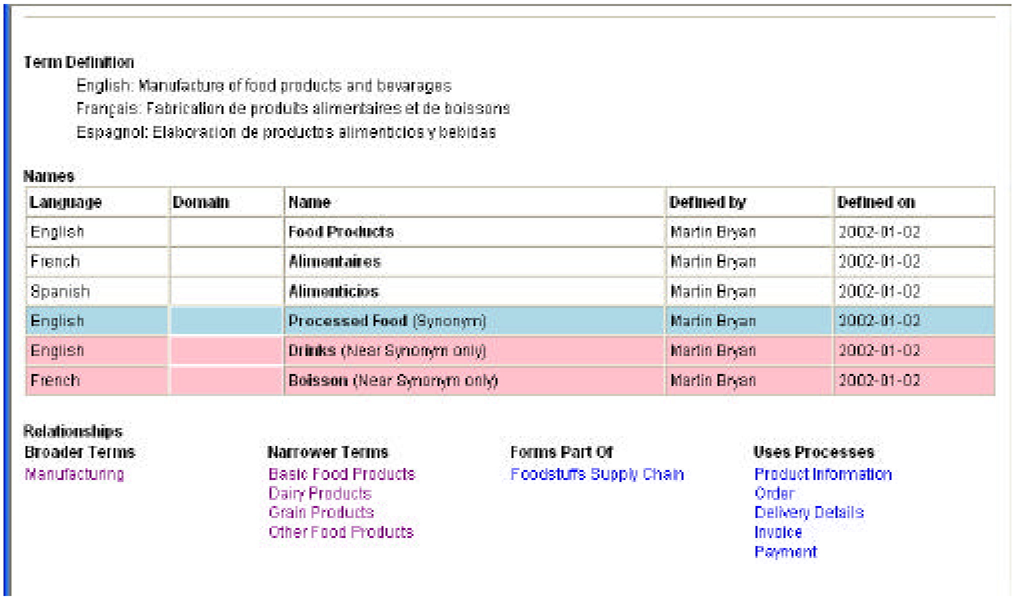
A simplified example of the use of this DTD might have the following form:

```
<Ontology
 Region="http://www.iso.org/ISO639/EU"
 Industry=”http://esa.un.org/unsd/registry/ISIC3/”>
 ...
<Term ID="ISIC3-D15"
 RecordedBy="Martin Bryan" WhenRecorded="2002-01-02">
  <SubjectDomain xlink:href="#ISIC3-D"/>
  <Definition xml:lang="EN">
   Manufacture of food products and bevarages</Definition>
  <Definition xml:lang="FR">
   Fabrication de produits alimentair&eacute;s et de boissons
  </Definition>
  <Definition xml:lang="ES">
   Elaboraci&oacute;n de productos alimenticios y bebidas
  </Definition>
  <Name xml:lang="EN">Food Products</Name>
  <Name xml:lang="FR">Alimentaires</Name>
  <Name xml:lang="ES">Alimenticios</Name>
  <AlternativeTerm xml:lang="EN"
   RecordedBy="Martin Bryan" WhenRecorded="2002-01-02">
   Processed Food</AlternativeTerm>
  <Includes xml:lang="EN"
   RecordedBy="Martin Bryan" WhenRecorded="2002-01-02">
   Drinks</Includes>
  <Includes xml:lang="FR"
   RecordedBy="Martin Bryan" WhenRecorded="2002-01-02">
   Boisson</Includes>
  <BroaderTerm xlink:href="#ISIC3-D"/>
  <NarrowerTerm xlink:href="#ISIC3-D151"/>
  <NarrowerTerm xlink:href="#ISIC3-D152"/>
  <NarrowerTerm xlink:href="#ISIC3-D153"/>
  <NarrowerTerm xlink:href="#ISIC3-D154"/>
  <FormsPartOf xlink:href="SupplyChain.xml#Food">
    <Label xml:lang="EN">Foodstuffs Supply Chain</Label>
    <Label xml:lang="ES">Comestibles suministro</Label>
    <Label xml:lang="FR">Alimentaire</Label>
  </FormsPartOf>
  <HasProcess xlink:href="BusinessProcesses.xml#ProductInformation">
    <Label xml:lang="EN">Product Information</Label>
    <Label xml:lang="ES">Productos</Label>
    <Label xml:lang="FR">Produits</Label>
  </HasProcess>
  <HasProcess xlink:href="BusinessProcesses.xml#Order">
    <Label xml:lang="EN">Order</Label>
    <Label xml:lang="ES">Orden</Label>
    <Label xml:lang="FR">Ordre</Label>
  </HasProcess>
  <HasProcess xlink:href="BusinessProcesses.xml#Delivery">
    <Label xml:lang="EN">Delivery Details</Label>
    <Label xml:lang="ES">Reparto</Label>
    <Label xml:lang="FR">Livraison</Label>
  </HasProcess>
  <HasProcess xlink:href="BusinessProcesses.xml#Invoice">
    <Label xml:lang="EN">Invoice</Label>
    <Label xml:lang="ES">Factura</Label>
    <Label xml:lang="FR">Facturer</Label>
  </HasProcess>
  <HasProcess xlink:href="BusinessProcesses.xml#Payment">
    <Label xml:lang="EN">Payment</Label>
    <Label xml:lang="ES">Pago</Label>
    <Label xml:lang="FR">Paiement</Label>
  </HasProcess>
</Term>
...
<MeasurementType ID=”weight”>
  <MeasurementName xml:lang=”EN” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>Weight</MeasurementName>
  <MeasurementName xml:lang=”FR” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>Poid</MeasurementName>
  <MeasurementName xml:lang=”ES” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>Peso</MeasurementName>
  <PermittedUnitRef Unit=”kg”/>
  <PermittedUnitRef Unit=”lb”/>
  <PermittedUnitRef Unit=”lbs-and-ozs”/>
  <PermittedUnitRef Unit=”tonnes”/>
  <PermittedUnitRef Unit=”hundredweight”/>
</MeasurementType>
...
<PermittedUnit ID=”kg”>
  <UnitName xml:lang=”EN ES FR” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>kilo</UnitName>

  <UnitName xml:lang=”EN FR” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>kilogramme</UnitName>
  <UnitName xml:lang=”EN” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>kilogram</UnitName>
  <UnitIdentifier>kg</UnitIdentifier>
</PermittedUnit>
<PermittedUnit ID=”lbs-and-ozs”>
  <UnitName xml:lang=”EN” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>Pounds and ounces</UnitName>
  <UnitIdentifier>lb</UnitIdentifier>
  <UnitIdentifier>oz</UnitIdentifier>
</PermittedUnit>
...
<TimeType ID=”DeliveryDate”>
  <MeasurementName xml:lang=”EN” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>Delivery Date</MeasurementName>
  <MeasurementName xml:lang=”FR” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>Temp de deliverance</MeasurementName>
  <MeasurementName xml:lang=”ES” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>Fecha da reparto</MeasurementName>
  <PermittedUnitRef Unit=”day”/>
  <PermittedUnitRef Unit=”date-and-time”/> >
</MeasurementType>
<PermittedUnit ID=”date-and-time”>
  <UnitName xml:lang=”EN” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>Date and Time</UnitName>
  <UnitIdentifier Separator=”-” Purpose=”year”/>
  <UnitIdentifier Separator=”-” Purpose=”month”/>
  <UnitIdentifier Separator=”T” Purpose=”date”/>
  <UnitIdentifier Separator=”:” Purpose=”hour”/>
  <UnitIdentifier Separator=”:” Purpose=”minute”/>
</PermittedUnit>
<PermittedUnit ID=”day”>
  <UnitName xml:lang=”EN” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>Day and Time</UnitName>
  <PermittedValue xml:lang=”EN” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>Monday</PermittedValue>
  <PermittedValue xml:lang=”EN” RecordedBy=”Martin Bryan”
```

```
   WhenRecorded=”2002-01-06”>Tuesday</PermittedValue>
  <PermittedValue xml:lang=”EN” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>Wednesday</PermittedValue>
  <PermittedValue xml:lang=”EN” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>Thursday</PermittedValue>
  <PermittedValue xml:lang=”EN” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>Friday</PermittedValue>
  <PermittedValue xml:lang=”EN” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>Saturday</PermittedValue>
  <PermittedValue xml:lang=”ES” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>lunes</PermittedValue>
  <PermittedValue xml:lang=”ES” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>martes</PermittedValue>
  <PermittedValue xml:lang=”ES” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>miércoles</PermittedValue>
  <PermittedValue xml:lang=”ES” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>jueves</PermittedValue>
  <PermittedValue xml:lang=”ES” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>viernes</PermittedValue>
  <PermittedValue xml:lang=”ES” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>sábado</PermittedValue>
  <PermittedValue xml:lang=”FR” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>lundi</PermittedValue>
  <PermittedValue xml:lang=”FR” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>mardi</PermittedValue>
  <PermittedValue xml:lang=”FR” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>mercredi</PermittedValue>
  <PermittedValue xml:lang=”FR” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>jeudi</PermittedValue>
  <PermittedValue xml:lang=”FR” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>vendredi</PermittedValue>
  <PermittedValue xml:lang=”EN” RecordedBy=”Martin Bryan”
   WhenRecorded=”2002-01-06”>samedi</PermittedValue>
</PermittedUnit>
...
</Ontology>
```

Using the XSL Transformation Language (XSLT), this file can be converted into an HTML file for display on a web browser in the following format:



Term Definition

English: Manufacture of food products and bevarages
Français: Fabrication de produits alimentaires et de boissons
Espagnol: Elaboracion de productos alimenticios y bebidas

Names

Language	Domain	Name	Defined by	Defined on
English		**Food Products**	Martin Bryan	2002-01-02
French		**Alimentaires**	Martin Bryan	2002-01-02
Spanish		**Alimenticios**	Martin Bryan	2002-01-02
English		**Processed Food** (Synonym)	Martin Bryan	2002-01-02
English		**Drinks** (Near Synonym only)	Martin Bryan	2002-01-02
French		**Boisson** (Near Synonym only)	Martin Bryan	2002-01-02

Relationships

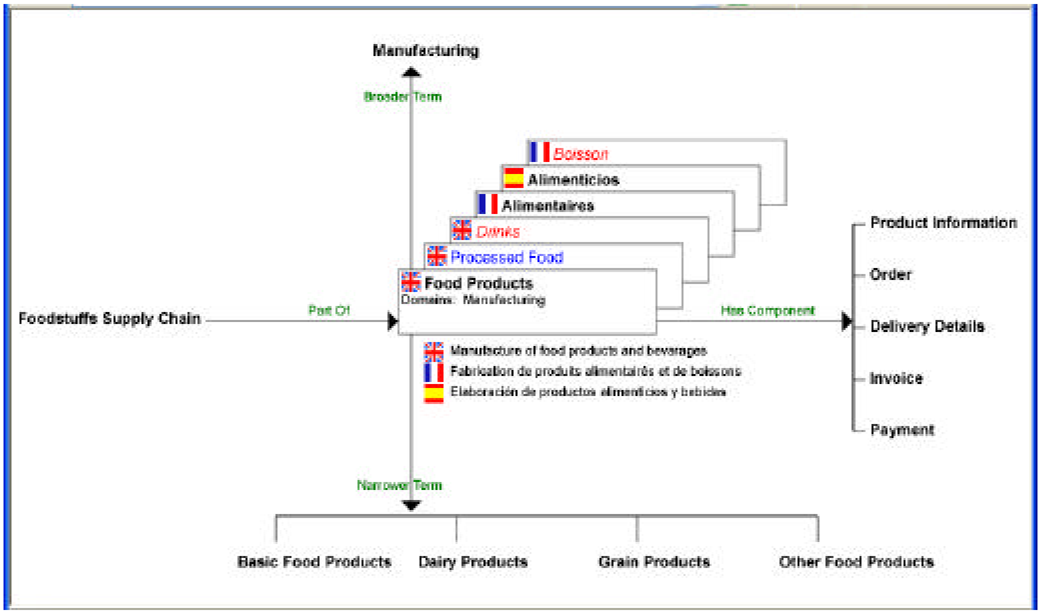
Broader Terms
Manufacturing

Narrower Terms
Basic Food Products
Dairy Products
Grain Products
Other Food Products

Forms Part Of
Foodstuffs Supply Chain

Uses Processes
Product Information
Order
Delivery Details
Invoice
Payment

Alternatively the term could be presented graphically as:



Manufacture of food products and bevarages
Fabrication de produits alimentairés et de boissons
Elaboración de productos alimenticios y bebidas

9.7 Current Status

MULECO is an on-going project, and as yet no formal set of definitions, or accompanying DTD/Schema, has been agreed. Areas of ongoing study include those currently being undertaken by European research projects such as MILES, CLAMOUR and OntoWeb, and by international e-commerce initiatives such ebXML/ebTWG, related to:

- Formal languages for describing ontologies
- Formal languages for describing multilingual word sets
- Formal models for the maintaining industrial classification schemes
- Formal languages for modelling business processes
- Techniques for the creation and maintenance of process-based ontologies

If you would like to take part in the project please contact our website at http://www.cenorm.be/isss/Workshop/ec/New_Projects.htm#MULECO

Martin Bryan
The SGML Centre
28th January 2002

Annex A: Pictoral representation of MULECO Schema

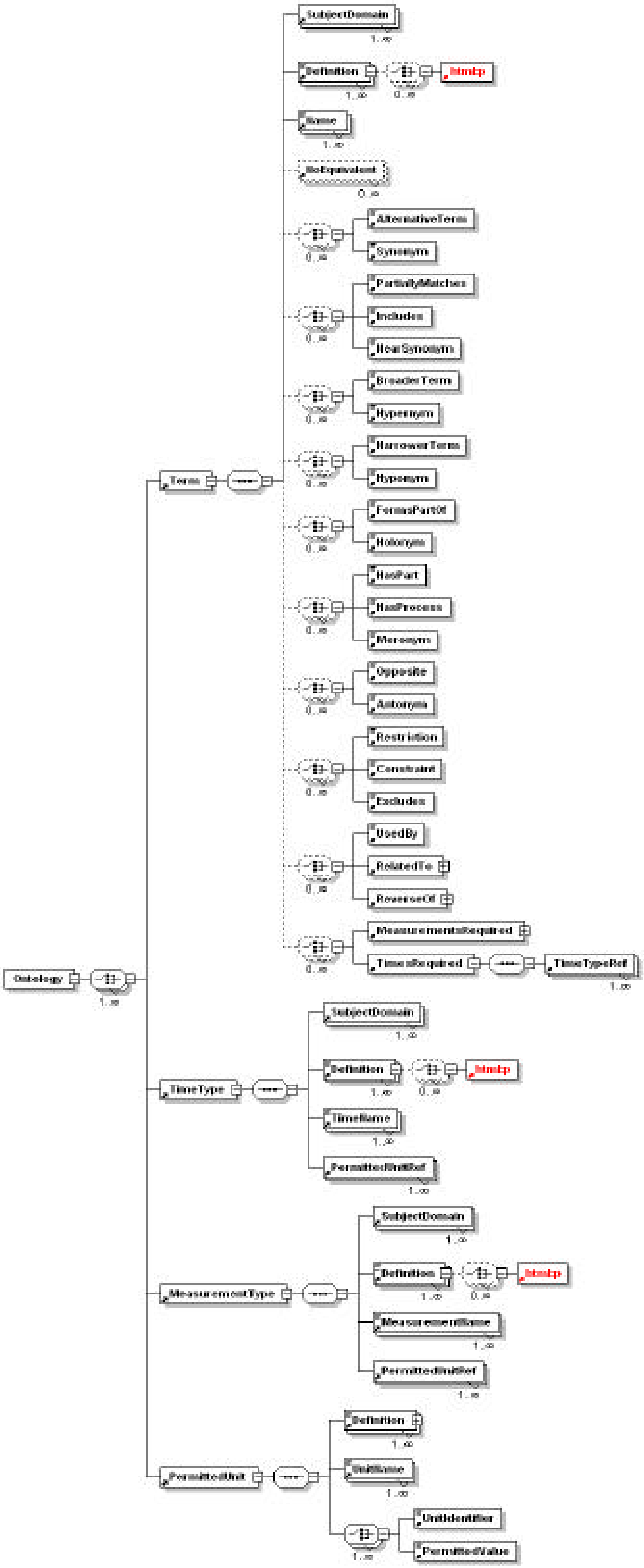


Figure 25 MULECO Schema