OntoWeb

State of the Art in Content Standards

Authors: Martin Dörr, Nicola Guarino, Mariano Fernández López, Ellen Schulten
Milena Stefanova, Austin Tate
Editors: Nicola Guarino, Milena Stefanova
Affiliations: FORTH, Greece; LADSEB-CNR, Italy; University of Madrid, Spain; Heloise Ontologies
Associates, AIAI Institute - University of Edinburgh, UK.

OntoWeb Consortium

This document is part of a research project funded by the IST Programme of the Commission of the European Communities as project number IST-2000-29243.

Vrije Universiteit Amsterdam (VU)-Coordinator
Faculty of Sciences,
Division of Mathematics and Computer Science
De Boelelaan 1081a
1081 HV Amsterdam, the Netherlands
Fax and Answering machine: +31-(0)84-872 27 22
Mobil phone: +31-(0)6-51850619
Contactperson: Dieter Fensel
E-mail: dieter@cs.vu.nl

Revision Information

Revision date	Version	Changes

Acknowledgements

We would like to acknowledge several people which helped us with providing content to this document. In particular, would like to thank the members of the OntoWeb Content Standards SIG/Process Standards WG (see http://www.aiai.ed.ac.uk/project/ontoweb/). These include Michael Gruninger, Howard Smith, Ismael Ghalimi, Adam Pease, Stuart Aitken, Ben Rode, Jerry Hobbs, David Martin and Martin Bryan.

Austin Tate's work was partially supported by the UK EPSRC-funded Advanced Knowledge Technologies (AKT) project.

We would also like to thank all the people which read the document and provided constructive comments for improving its content.

Table of Contents

1 Introduction

The OntoWeb SIG on Content Standards focuses on one of the primary goals of OntoWeb: promoting the development of ontology-based metadata standards and content harmonization/interoperability across different standards for the creation, communication and sharing of such things as information, knowledge, products, capabilities, and process descriptions on the Web.

The main goal of WP3 and the related SIG on Content Standards is to promote the role of well-founded ontologies in standards harmonization, in order to facilitate standards interoperability – or at least mutual understanding – across different communities operating in similar areas. As specified in the project workplan, this goal is to be achieved in collaboration with interested partners and will be focused on a limited number of selected standard clusters, suitably identified within the most relevant application areas on the basis of actual feasibility. The work to be done is organized into the following main phases:

1. *Preliminary classification:* A large set of potentially relevant content standards will be identified and classified on the basis of a preliminary classification scheme. Most relevant content standardization areas will be also identified.
2. *Standard clusters selection:* small clusters of standards with related content will be isolated, on the basis of actual interest from OntoWeb participants to ontology-based harmonization, comparison, or interoperability work.
3. *Comparison and harmonization:* For each cluster, ontology compatibility issues and harmonization requirements will be analyzed in a mixed top-down and bottom-up fashion. The more general issues will be discussed in a separate Working Group focused on foundational issues in ontology-based harmonization.
4. *Comparison and evaluation framework.*: this will be the final output of WP3 and the SIG on Content Standards. On the basis of the experience gained on selected clusters, a general comparison and evaluation framework for content standards will be developed, which shall be tested on a sample of standards selected from those analyzed in the first phase.

This document describes the results of phase 1 above. It will be updated on a regular basis to include important missing standards and new emerging standards, as well as further suggestions concerning the classification scheme.

The document is organized as follows. We first present a preliminary scheme for classifying content standardization efforts along two main dimensions, *content type* and *semantic depth*. The choice of these dimensions is quite natural: all what is the SIG about is *content*. Thus aspects such as representation format, languages used, development and deployment status will be not considered.

Among the many different content standards, we list first all those we believe have some relevance for e-commerce and the semantic web, and that are publically accesssible. These standards are presented in alphabetic order according to their acronym, in a large roadmap table.

We present then in more detail those standards that are more relevant to the OntoWeb members. These have been been described and organized by the three Working Groups that have been created within the SIG: *product description standards*, *process description standards*, and *cultural repositories standards*.

2 Classifying Content Standardisation Efforts

We have analyzed a number of content standardization efforts and defined what are, for our purposes, relevant properties we wish to associate with each one. The following sections describe our initial classification scheme. The initial assumptions for the selection of these standards are the following:

1. Content relevance: the content of all the selected standards is relevant to either
  (a) e-commerce, or
  (b) harmonization of e-commerce standards

In the first category we include standards that describe in some way electronic business transactions or their participants. By "participants" we mean all kind of entities which can potentially participate in a business transaction: this includes entities such as information, products, services, as well as parties and organizations which may participate in business transactions. In the second category we consider efforts which can be of some help for achieving harmonization of standards belonging to the first category. These can be upper-level ontologies, reference models, etc.

2. Standard accessibility: We restrict ourselves to open, non-proprietary standards. Proprietary standards are protected by trademark, patent, or copyright, and are made, produced or distributed by somebody having exclusive rights. Non-proprietary standards do not require a license for use, are free for distribution, and adopt an open development procedure. The development is often organized through equal membership status.

2.1 Main classification

All standard efforts are classified along two principle dimensions, *content typ*e (i.e., the standard's subject domain), and *semantic depth*. Note that we focus *only* on content standards, formatting standards (e.g., XML, ASCII, etc.) and representation standards (e.g., RDF, RDFS, etc.) are out of the scope of this work.

Content Type. There are four kinds of standards hat have distinguished in our analysis according to this dimension:

1. *Artifact-centric:* The content in such standards is organized around a set of (material or immaterial) artifacts. Examples of artifacts are various products, information, money and securities, etc. Such standards may include only content related to artifacts (e.g., UNSPSC) or describe artifacts also from points of view of their use, creation processes, parties involved, etc). Usually, artifact-centric standards which include additional information about processes or parties involved do not consider the latter two categories in a general way - the additional content is restricted to the scope of the artifacts considered. Of particular importance for e-commerce are artifact-centric standards which describe goods participating in electronic business transactions. Typically such artifacts are organized in electronic catalogues.
2. *Process- and activity-centric:* In most cases, these standards also include content about agents, as well as content about objects and artifacts. Some interesting e-commerce standards in this domain are those describing all aspects of electronic business transactions or those focusing on value transfer processes.
3. *Agent-centric*: These standards focus on people and organizations, rather than on processes or artifacts, though they may describe a party in terms of processes or activities to which it participates. Currently we have not considered standards falling in this category, we include it for the sake of completeness.
4. *General :* those standards combine some (or all) of the groups above and also establish relations between them. Here we may consider various upper- and core-level ontologies.

Semantic Depth. The standards we looked at use modelling primitives of different richness and expressivity. We distinguish several levels of semantic depth, according to the semantic primitives adopted:

> Level 0:　　　　　　*Dictionaries*, describing incormal definitions associated to concept

Level 1:	*Taxonomies*, describing specialization relationships between concepts;
Level 2:	*Thesauri*, adding to taxonomies various lexical relationships (hyperonimy, synonimy, partonomy, etc…)
Level 3:	*Reference models*, combing many of the relations above and trying to capture other more complex relations between concepts.

The vocabularies are on the lowest level of semantic depth. Taxonomies define static inclusion relations between concepts and contain implicitly or explicitly a vocabulary. Thesauri define equivalence relations between concepts. Usually, thesauri make use of a taxonomy, that is why we consider them one level higher than taxonomies. Reference models try to capture a domain by identifying the most important concepts and relations between them, with the purpose of making clear the intended meaning and allowing useful inferences.

2.2 Other Attributes

In addition to the categories described above, and obvious attributes like *acronym*, *name*, and *URL*, we have also included in our classification scheme other information useful to assist in understanding standardization efforts.

Meta and upper-level content. Some standardization efforts specifically address meta-content issues, as opposed to simply describing a domain. This is basically a Boolean attribute, however some efforts such as the Dublin Core deal with meta-content in such a simplistic fashion that we choose to differentiate it from others with a "some" value.

Business content. Most efforts contain business knowledge encapsulated in the standard, while others (normally the efforts classified as general or general for a domain) attempt to be independent. This is also a Boolean attribute.

Focus. Names the specific sub-domain of interest, if any, motivating the standardization effort. For example, OFX is a standard in the bank domain, with a focus on bank-customer interface services.

3 A Content Standards Roadmap

The table below lists content standardization efforts classified according to the main criteria and additional attributes discussed in the previous sections. Due to the difficulties of getting accurate and uptodate information about some these standards, we make no claims of completeness nor accuracy, so suggestions and improvements are mostly welcome. Blank cells in the table indicate that we have not yet determined a value of that attribute. A "?" indicates that we have tried to determine a value but were unable to, and "some" indicates an intermediate value between "yes" and "no". [NOTE: Not all these standards are described in detail in the next section].

Acronym	Full Name	Semantic Depth	Domain Type	Meta Content	Business Content	Focus
AAT	Art and Architecture Thesaurus	Thesaurus	Artifact		Yes	Art and architecture
ANX	Automobile Network eXchange	Thesaurus	Artifact	No	Yes	Automotive
BizTalk			Process		Yes	Business documents handling
BPML	Business Process Modeling Language	Model (some)	Process	Some	Yes	e-business
CBL	Common Business Language	Vocabulary			Yes	
CDIF	Common Data Interchange Framework	Model	General (some)	Yes	No	
CIDOC CRM	CIDOC Conceptual Reference Model	Model	General			Museums
CIMI profile	Z39.50 Profile for Cultural Heritage Information	Thesaurus	Artifact	No	Yes	Museums
CPR	Core Plan Representation	Model	Process			
CPV	Common Procurement Vocabulary	Vocabulary	Artifact	No	Yes	Public procurement
CWM	Common Warehouse Model	Model	Artifact	No?	No	Data Warehousing
cXML	Commerce XML	Model	Process	No	Yes	e-commerce
CYC		Model	General			
DAML-S	DAML for Services	Model	Process			Web services
DCMI	Dublin Core Metadata Initiative	Model	Artifact	No	Yes	Libraries/Web catalogs
ebXML	Electronic Business XML	Model	Process	Some	Yes	e-business

Acronym	Full Name	Semantic Depth	Domain Type	Meta Content	Business Content	Focus
e-cl@ss		Taxonomy	Artifact		Yes	Product catalogues
EGAS	ECCMA Global Attribute Schema	?	General	Yes	No	Product classifications
EPISTLE	European Process Industries STEP Technical Liaison Executive	Model	General	Yes	Yes	Process industry
fpML	Financial Products Markup Language	Model	Process/Artifact	No	Yes	Finance
FRBR	Functional Requirements for Bibliographic Records	Model	General			Libraries
HL7	Health Level 7	Model	Artifact	Yes	Yes	Healthcare
ICE	Information and Content Exchange	Model	Process	?	Yes	Exchange of online assets
INDECS	Interoperability of Data in e-commerce systems	Model	Artifact	No	Yes	Intellectual Property
IOTP	Internet Trading Protocol	Model	Process	No	Yes	Retail trading
ISITC	International Securities Association for Institutional Trade Communication	Model	Process	No	Yes	Security Industries Transactions
ISO 2789	Documentation guidelines for the establishment and development of monolingual thesauri	Vocabulary	General	Yes	No	Thesauri
ISO IEC 11179	Metadata Registry Coalition	Model	General	Yes	No	Modeling
ISO 1087	Terminology Vocabulary	Vocabulary	General	Yes	No	Terminology
ISO 5964	Documentation and establishment of multilingual thesauri	Vocabulary	General	Yes	No	Thesauri

Acronym	Full Name	Semantic Depth	Domain Type	Meta Content	Business Content	Focus
MDCOIM	MDC Open Information Model	Model	General	Yes	Yes	
MOF	Meta Object Facility	Model	General	Yes		Modeling
NAICS	North American Industry Classification System	Taxonomy	Artifact			Products and services
NIIIP	National Industrial Information Infrastructure Protocol	Model	Artifact, Process	Some	Yes	Virtual Enterprise
NIIIS	National Industrial Information Infrastructure	Model				EXPRESS models
OBI	Open Buying on the Internet	Model	Process	No	Yes	B2B electronic procurement
OCF	Online Catalog Format	Model	Artifact	No	Yes	Product Catalogs
OFX	Open Financial Exchange	Model	Process	No ?	Yes	retail banking
PDDL	Planning Domain Description Language	Model	Process			Planning
POSC-CAESAR	Petrotechnical Open Software Corporation - CAESAR	Model	Artifact	No	Yes	Petroleum
PSL	Process Specification Language	Model	Process	Yes	Some	Interoperability for manufacturing and business process software
RosettaNet	RosettaNet Catalog Interoperability Proposal	Taxonomy (some), vocabulary	Artifact	No	Yes	Electronic equipment
SCTG	Standard Classification of Transported Goods	Taxonomy	Artifact		Yes	Transported goods
SPAR	Shared Planning and Activity Representation	Ontology	Process	Yes		
SUO	Standard Upper Ontology	Ontology	General	Yes		
SHIC	Social, Historical, Industrial Classification	Thesaurus	Artifact		Yes	Museums

Acronym	Full Name	Semantic Depth	Domain Type	Meta Content	Business Content	Focus
SWIFT		Vocabulary	Process	No	Yes	Finance transactions
TGN	Thesaurus of Geographic Names	Thesaurus	Artifact/Agent			Art and architecture places
tpaML	Trading Partners Agreement ML	Model	Process		Yes	B2B transactions
UDDI	Universal Description Discovery and Integration	Vocabulary	Process/Artifact	No	Yes	Web-services
UDEF	Universal Data Element Framework	Vocabulary	General	No	Yes	Business
ULAN	The Union List of Artist Names	Vocabulary	Agent	No	Yes	Art and architecture
UML Meta Model	Universal Modeling Language Meta Model	Model	General	Yes		
UMLS		Thesaurus	General	Yes	Yes	Healthcare
UN/EDIFACT	United Nations Electronic Data Interchange for Administration, Commerce and Transport	Vocabulary (model entry definitions)	Process	No	Yes	e-business transactions
UN/SPSC	United Nations Standard Products and Services Codes	Taxonomy	Artifact	No	Yes	Product catalogues
UNSPSC	Universal Standard Products and Services Classification	Taxonomy	Artifact, Process (some)	No	Yes	Product catalogues
WPDL	Workflow Process Definition Language	Model	Process	Some	Yes	Workflow
WSDL	Web Services Description Language	Model				Web-services
WSFL	Web Services Flow Language	Model	Process			Web-services
X.12	EDI standard	Vocabulary (model entry definition)	Process	No	Yes	e-business transactions
xCBL	XML Common Business Library	Model (?)	Process	Some	Yes	Business documents handling

4 Most Relevant Content Standards

We present here in more detail those standards – from the table above – that are more relevant to the OntoWeb members. These have been been selected, described and organized by the three Working Groups that have been created within the SIG: *product description standards*, *process description standards*, and *cultural repositories standards*.

4.1 Product Description Standards

We can distinguish between *global standards* and *regional standards* (inspired on [Cor2001]). A global standard address the whole domain of products, while regional standards are created for specific domains.

4.1.1 Global Standards

United Nations Standard Products and Services Codes (UN/SPSC) and   Universal Standard Products and Services Classification (UNSPSC)
URLs: http://www.un−spsc.net, http://www.unspsc.org

In 1999, the United Nation's Common Coding System (UNCCS) and Dun & Bradstreet's Standard Product and Service Codes (SPSC), merged into the by the United Nations Standard Products and Services Codes (UN/SPSC) owned by the United Nations Development Program (UNDP).

Initially, the UNDP handled the management of the code to the Electronic Commerce Code Management Association ECCMA (www.eccma.org). This partnership has recently been withdrawn, leading to two *different* versions of the UNSPSC; the United Nations Standard Products and Services Codes owned by the UNDP and the Universal Standard Products and Services Classification managed by the ECCMA.

UNSPSC is a general standard for description of products and services. Its coding system is organized as a five-level taxonomy of products, each level containing a two-character numerical value and a textual description. These levels are defined as follows:

Segment. The logical aggregation of families for analytical purposes.
• *Family.* A commonly recognized group of inter-related commodity categories.
• *Class.* A group of commodities sharing a common use or function.
• *Commodity.* A group of substitutable products or services.
• *Business Function.* The function performed by an organization in support of the commodity. This level is seldom used.

A recent version of the UNSPSC classification ((ECCMA V 8.0 released Oct 1, 2001) contains around 12000 products organized in 55 segments. Segment 43, which deals with computer equipment, peripherals and components, contains around 300 kinds of products.

The segments of UNSPSC are currently:

[10] Live Plant and Animal Material and Accessories and Supplies
[11] Mineral and Textil and Inedible Plant and Animal Materials
[12] Chemicals including Bio Chemicals and Gas Materials
[13] Resin and Rosin and Rubber and Foam and Film and Elastomeric Materials
[14] Paper Materials and Products
[15] Fuels and Fuel Additives and Lubricants and Anti corrosive Materials
[20] Mining Machinery and Accessories
[21] Farming and Fishing and Forestry and Wildlife Machinery and Accessories
[22] Building and Construction Machinery and Accessories
[23] Industrial Manufacturing and Processing Machinery and Accessories
[24] Material Handling and Conditioning and Storage Machinery and their Accessories and Supplies
[25] Commercial and Military and Private Vehicles and their Accessories and Components

[26] Power Generation and Distribution Machinery and Accessories
[27] Tools and General Machinery
[30] Structures and Building and Construction and Manufacturing Components and Supplies
[31] Manufacturing Components and Supplies
[32] Electronic Components and Supplies
[39] Lighting and Electrical Accessories and Supplies
[40] Distribution and Conditioning Systems and Equipment and Components
[41] Laboratory and Measuring and Observing and Testing Equipment
[42] Medical Equipment and Accessories and Supplies
[43] Communications and Computer Equipment and Peripherals and Components and Supplies
[44] Office Equipment and Accessories and Supplies
[45] Printing and Photographic and Audio and Visual Equipment and Supplies
[46] Defense and Law Enforcement and Security and Safety Equipment and Supplies
[47] Cleaning Equipment and Supplies
[48] Service Industry Machinery and Equipment and Supplies
[49] Musical Instruments and Recreational Equipment and Supplies and Accessories
[50] Food Beverage and Tobacco Products
[51] Drugs and Pharmaceutical Products
[52] Domestic Appliances and Supplies and Consumer Electronic Products
[53] Apparel and Luggage and Personal Care Products
[54] Timepieces and Jewelry and Gemstone Products
[55] Published Products
[56] Furniture and Furnishings
[70] Farming and Fishing and Forestry and Wildlife Contracting Services
[71] Mining and Oil and Gas Drilling Services
[72] Building and Construction and Maintenance Services
[73] Industrial Production and Manufacturing Services
[76] Industrial Cleaning Services
[77] Environmental Services
[78] Transportation and Storage and Mail Services
[80] Management and Business Professionals and Administrative Services
[81] Research and Science Based Services
[82] Editorial and Design and Graphic and Fine Art Services
[83] Public Utilities and Public Sector Related Services
[84] Financial and Insurance Services
[85] Healthcare Services
[86] Education and Training Services
[90] Travel and Food and Lodging and Entertainment Services
[91] Personal and Domestic Services
[92] National Defense and Public Order and Security and Safety Services
[93] Politics and Civic Affairs Services
[94] Organizations and Clubs

An example of a part of the structure of UNSPC is reported below:

```
[43] Communications, Computer Equipment, Peripherals, Components and
Supplies
 -family-[4316] Software
  -class-[431615] Database systems
    -commodity-[43161501] Database software
  -class-[431616] Operating systems
    -commodity-[43161601] Mainframe operating system software
    -commodity-[43161602] Personal computer (PC) operating system software
  -class-[431617] Business transaction and personal business software system software
    -commodity-[43161701] Investment management software
    -commodity-[43161702] Tax preparation software
    -commodity-[43161703] Facilities management software
    -commodity-[43161704] Software suites
    -commodity-[43161705] Inventory management software
    -commodity-[43161706] Financial analysis software
    -commodity-[43161707] Accounting software
    -commodity-[43161708] Time accounting or human resources software
```

NAICS (North American Industry Classification System)
URL: http://www.naics.com

NAICS was created by the Census Office of USA in cooperation with the Economic National Classification Committee of USA, Statics of Canada, and the Instituto Nacional de Estadística, Geografía e Informática de Méjico. It describes products and services in general and is used in USA, Canada and Mexico. NAICS was created after revising the Standard Industrial Classification (SIC) standard. SIC was originally developed in the 1930's to classify establishments by the type of activity in which they are primarily engaged and to promote the comparability of establishment data describing various facets of the U.S. economy.

NAICS industries are identified by a 6-digit code, in contrast to the 4-digit SIC code. The longer code accommodates the larger number of sectors and allows more flexibility in designating subsectors. It also provides for additional detail not necessarily appropriate for all three NAICS countries. The international NAICS agreement fixes the first five digits of the code. The sixth digit, where used, identifies subdivisions. NAICS industries that accommodate user needs in individual countries. Thus, digit U.S. codes may differ from counterparts in Canada or Mexico, but at the-digit level they are standardized.he New Hierarchical Structure. The general structure is the following:

X   Industry Sector (20 broad sectors up from 10 SIC)
XX   Industry Subsector
XXX   Industry Group
XXXX   Industry
XXXXX   U.S., Canadian, or Mexican National specific

The following are the 20 broad sectors (up from the 10 divisions of the SIC system):
•   Agriculture, Forestry, Fishing and Hunting
•   Mining
•   Utilities
•   Construction
•   Manufacturing
•   Wholesale Trade
•   Retail Trade
•   Transportation and Warehousing
•   Information
•   Finance and Insurance
•   Real Estate and Rental and Leasing
•   Professional, Scientific, and Technical Services
•   Management of Companies and Enterprises
•   Administrative and Support and Waste Management and Remediation Services
•   Education Services

- Health Care and Social Assistance
- Arts, Entertainment, and Recreation
- Accommodation and Food Services
- Other Services (except Public Administration)
- Public Administration

Many of the new sectors reflect recognisable parts of SIC divisions, such as the Utilities and Transportation sectors, broken out from the SIC division Transportation, Communications, and Utilities. Similarly, the SIC division for Service Industries has been subdivided to form several new sectors

An example of this classification is reported below:

```
11    Agriculture, Forestry, Fishing
111   Crop Production
1111  Oilseed and Grain Farming
11111 Soybean Farming
11112 Oilseed (except Soybean)
11113 Dry Pea and Bean Farming
11114 Wheat Farming               E
11115 Corn Farming                R
11116 Rice Farming                E
11119 Other Grain Farming
111191 Oilseed and Grain Combination   u  N
111199 All Other Grain Farming        u  R
1112  Vegetable and Melon Farming
11121  Vegetable and Melon Farming
111211 Potato Farming            c  E
111219 Other Vegetable (except       c  R
1113  Fruit and Tree Nut Farming
11131  Orange Groves              N
11132  Citrus (except Orange) Groves     R
11133  Noncitrus Fruit and Tree Nut
111331 Apple Orchards             u  N
111332 Grape Vineyards            u  E
111333 Strawberry Farming          u  N
111334 Berry (except Strawberry)     u  R
111335 Tree Nut Farming           u
111336 Fruit and Tree Nut Combination  u  N
111339 Other Noncitrus Fruit Farming   u  R
1114   Greenhouse, Nursery
```

There exist crosswalk files between the UNSPSC and the SIC codes, and between the UNSPSC and the NAICS codes (see for example www.eccma.org /unspsc/crosswalk.html).

SCTG (Standard Classification of Transported Goods)
URL: http://www.bts.gov/programs/cfs/sctg/welcome.htm

The Bureau of Transportation Statistics (BTS) sponsored the development of a new product classification for collecting and reporting future Commodity Flow Survey (CFS) data. SCTG was created by the U.S. Department of Transportation's (DOT), Volpe National Transportation Systems Center (Volpe Center), Standards and Transportation Divisions of Statistics Canada, U.S. Bureau of the Census (BOC), and the U.S. Bureau of Economic Analysis (BEA). The classification has four levels, each of which follows two important principles. First, each level covers the universe of transportable goods. Second, each category in each level is mutually exclusive.

Level	Structure	# Codes
I	XX	43
II	XX.X	145
III	XX.XX	331
IV	XX.XXX	597

The SCTG is organised in a hierarchical structure:

[Two digit codes] --Codes for aggregate CFS data reporting--links to 2-digit SICs
[Three digit codes]--Codes for reporting intermediate CFS data detail
[Four digit codes]--Codes for reporting intermediate CFS data detail
[Five digit codes]--Codes and wording for CFS data collection and detailed data reporting
[Full SCTG Hierarchy]--Basics for creation of 2-digit, 3-digit, 4-digit, and 5-digit codes, with detailed information about contents of each code

An example of this classification is the following:

01   LIVE ANIMALS AND LIVE FISH
010   Live animals and live fish
0100   Live animals and live fish
      01001   Bovine animals
      01002   Pigs
      01003   Poultry
      01004   Other live animals including horses, sheep, goats, fur-bearing animals, honey bees, leaf-cutter, bee larvae, bait, pet or song birds, cats, and dogs
      01005   Live fish including aquarium
02   CEREAL GRAINS
021   Wheat
0210   Wheat
      02100   Wheat
022   Corn (except sweet)
0220   Corn (except sweet)
      02200   Corn (except sweet)
029   Cereal grains n.e.c.
0291   Rye
      02910   Rye
0292   Barley
      02920   Barley
0293   Oats
      02930   Oats
0299   Other
      02990   Other including rice, grain sorghum, buckwheat, millet, and canary seed
03   AGRICULTURAL PRODUCTS EXCEPT CEREAL GRAINS

Ecl@ss
URL: http://www.eclass.de

E-cl@ss is a German initiative to create a standard classification of material and services for information exchange between suppliers and their customers. In its core e-cl@ss is an alternative to UNSPC. It is currently used by companies like BASF, Bayer, Volkswagen-Audi, SAP, etc. The e-cl@ss classification consists of four levels of concepts (called *material classes*), with a numbering code similar to the one used in UNSPSC (each level has two digits that distinguish it from the other concepts). The four levels are: *Segment, Main group*, *Group* and *Commodity Class*. Inside the same commodity class we can have several products (in this sense, several products can share the same code, and this could lead to a fifth level with all of them). It also contains around 12000 products organized in 21 segments. Segment 27, which deals with *Electrical Engineering*, contains around 2000 products. Finally, the main group 27-23, which deals with *Process Control Systems*, together with the main groups 24-01 to 24-04, which deal with *Hardware*, *Software*, *Memory* and other computer devices, contain around 400 concepts. An example of this classification is reported below:

Klasse/Class	Klassenbezeichnung	Class description	E
24-01-99-00	Reparatur (Hardware)	Repair (hardware)	s
24-01-99-01	Großrechnersystem (Reparatur)	Mainframe system (repair)	4
24-01-99-01	Instandsetzung (Großrechnersystem)	Repair (Mainframe system)	s
24-01-99-01	Reparatur (Großrechnersystem)		s
24-01-99-02	Serversystem (Reparatur)	Server system (repair)	4
24-01-99-02	Instandsetzung (Serversystem)	Repair (Server system)	s
24-01-99-02	Reparatur (Serversystem)		s
24-01-99-03	PC-System (Reparatur)	PC system (repair)	4
24-01-99-03	Instandsetzung (PC-System)	Repair (PC system)	s

The aim of ecl@ss is to provide a set of Attributes at every classification end point. The set of Attributes is an aggregation of indicidual characteristis describing the related commodity. This set of attributes distinguishes ecl@ss from UNSPSC and is offering a solution to the shallowness of the UNSPSC. For example, at 24-01-99-03 PC System (Reperatur), one can add the Attribute-Set:

AAA001001- Hersteller
AAA002001- Produkt Typ
AAA003001- Produkt Name
AAA252001- Hersteller-Artikelnummer
AAA889001- EAN Code

EGAS: ECCMA Global Attribute Schema
URL: http://www.eccma.org/egas/

The Electronic Commerce Code Management Association ECCMA is a not-for-profit membership organization. The mission of the ECCMA is 'to provide an open, transparent and efficient process for the development of open source code standards over the Internet.' The ECCMA has over 1300 members from 47 countries worldwide. Next to the management of the Universal Standard Products and Services Classification, mentioned above, ECMMA manages a variety of data dictionaries, such as the International Address code. Interesting for our purposes is the recently announced maintenance of the Global Attribute Schema (EGAS). The code will be used to define a standard set of attributes to help users more closely describe the characteristics of a commodity found in the UNSPSC. One of the partners that will join this initiative is the Universal Content Extended Classification Organization UCEC.org. There are currently no specifications available.

CVP: Common Procurement Vocabulary
URL: http://simap.eu.int/EN/pub/src/cpv98.htm

The CPV was created in 1993 as a tool for improving transparency and efficiency in the field of public procurement. Use of standard terms in the CPV makes it easier for potential suppliers to identify the procurement contracts in which they are interested. The CPV also facilitates fast and accurate translation of contract notices for publication in the EC Official Journal, and makes it easier to establish procurement statistics. The CPV is urged by the EC to be utilized by public authorities and economic operators when drafting public procurement notices. However, its use remains optional, though it is being considered to make it compulsory.

The CPV consists of 8 digits for classification and a ninth digit, in order to check if the eight others are correct. These codes and the definitions of products (goods and services) in all the languages of the European Union constitute the main vocabulary of the CVP.

There exist crosswalks between the CPV and the UNSPSC
(http://www.eccma.org/unspsc/crosswalk.html)

4.1.2 Regional Standards

RosettaNet
URL: http://www.rosettanet.org

RosettaNet is a self-funded, non-profit consortium of more than 400 companies in Electronic Components, Information Technology, Semiconductor Manufacturing and Solution Provider companies. Originated in the IT industry, Rosettanet is currently expanding to other verticals, notably the automotive, consumer electronics and telecommunications industries. RosettaNet classification does not use a numbering system, as UNSPSC does, it is based on the names of the products it defines. This classification is related to the UNSPSC classification by providing the UNSPSC code for each product defined in it. RosettaNet has just two levels in its taxonomy of concepts:

- *RN Category*. A group of products, such as *Video Products*.
- *RN Product*. A specific product, such as *Television Card*, *Radio Card*, etc.
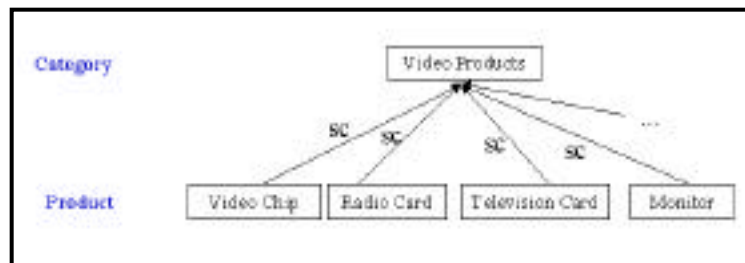

Figure 1. A snapshot of the classification of video products of the RosettaNet taxonomy. [Cor2001]

The RosettaNet classification consists of 14 categories and around 150 products. It must be taken into account (in relationship with UNSPSC) that RosettaNet is more specific than the UNSPSC classification. Figure 1 shows a small part of the RosettaNet classification, related to video products for computer equipment.

An example of this classification is reported below:

RNProductName	RNCategoryName	Notes
Analog Modem	Communications	
Bar Code Reader	Imaging Products and Accessories	
Battery	Power Equipment	Vendors are encouraged to choose a more specific UN/SPSC code in this class for particular product instances.
Battery Charger	Power Equipment	
Bridge	Network Hardware	Include "brouters" (bridge/router hybrids).
Cable	Accessories	
Cable Tester	Accessories	
Carrying Case	Accessories	
Central Processing Unit	Memory and Processors	
Central Processing Unit Card	Memory and Processors	
Channel Service Unit	Network Hardware	(UN/SPSC will change the name to include CSU.)
Chassis	Accessories	
Communication Switch	Communications	
Computer Cabinet	Computer Systems and Components	Vendors are encouraged to choose a more specific UN/SPSC code in this class for particular product instances.
Connector	Accessories	
Controlled Access Unit	Network Hardware	

4.2 Process Description Standards

Here some of the emerging process standards are presented. They are grouped according to their content. They are divided into general process standards which will be considered in the WG on foundational issues and e-business process standards.

4.2.1 General Process Standards

In this section we consider several standardisation efforts containing concepts which can be used for describing of classes of processes, activities and plans. Many of these standards serve as languages for business process definitions, where the stress is on internal processes. They do not need an immediate harmonization for providing interoperability, rather they can serve as semantical harmonization framework for concrete process standards, such as trading, financial, etc. standards. The effort considered are:

- Business Process Modeling Language (BPML);
- Process Definition Metamodel and Workflow Process Definition Language (WPDL);
- Process Specification Language (PSL);
- Core Plan Representation (CPR);
- Shared Planning and Activity Representation (SPAR)
- Planning Domain Description Language (PDDL).

Business Process Modeling Language
URL: http://www.bpmi.org
Sources: BPMI.org, *Business Process Modeling Language (BPML)*, Working Draft 0.4, August, 2001

The Business Process Modeling Language (BPML) is an effort by the Business Process Management Initiative, a non-profit corporation that empowers companies of all sizes, across all industries, to develop and operate business processes that span multiple applications and business partners. BPMI.org defines open specifications such as the Business Process Modelling Language (BPML) and the Business Process Query Language (BPQL) that will enable the standards-based management of business processes with forthcoming Business Process Management Systems (BPMS), in much the same way SQL enabled the standards-based management of business data with off-the-shelf Database Management Systems (DBMS). Whereas the enterprise wide management of data using DBMS is based on the relational data model, the enterprise wide management of processes using BPMS is based on process calculus.

BPML is a standardization effort for modeling of entire value chains of many partners, its current version is an working draft released in August 2001.

BPMI.org considers an e-Business process conducted among two business partners as made of three parts: a Public Interface and two Private Implementations (one for each partner). The Public Interface is common to the partners and is supported by protocols such as ebXML, RosettaNet, and BizTalk. The Private Implementations are specific to every partner and are described in any executable language. BPML is one such language.

BPML defines a business process as an interaction between participants and the execution of activities according to a defined set of rules in order to achieve a common goal. Although BPML does not rely on an explicit meta-model (like WPDL) or does not specify more formally the relations between its primitives (???), it seems to make use of a rich set of concepts (richer than WPDL) with clear ontological content. For example, there are informal definitions (like in a glossary) of concepts like participants, processes, activities, transactions. The notion of transaction makes BPML suitable for modeling value chains (i.e., entire process chains between different business partners each of which adds some value to the final product or service). Further, BPML defines ways of composing processes out of other processes. For example it tackles the problems of nesting and parallelism. The representation format of these is defined in a XML Schema, which with all its advantages also implies lack of more precise (e.g., formal) semantics of the concepts and may hinder the understandability of the language.

Process Definition Metamodel and Workflow Process Definition Language (WPDL)

URL: http://www.wfmc.org

Sources: Workflow Management Coalition, *Interface 1: Process Definition Interchange Process Model,* Document Number WfMC TC-1016-P, Version 1.1 (Official release), October 29, 1999

This is a standard effort by the Workflow Management Coalition (WfMC) which is a non-profit organization whose aim is the creation of standards for workflow management systems.

> *What is workflow management*: Workflow management is the computerized facilitation or automation of business processes, in whole or in part. Hence the automated business processes are the central issue. Workflow is often realized by means of workflow management system - a system that completely defines, maintains, and executes the workflow based on a computer representation called process definition. Workflow management is used in the office environments in staff intensive operations such as insurance, banking, legal al, and general administration. It is also applicable is some classes of industrial and manufacturing applications. Many software vendors have WFM products available today and there is a continuous introduction of more products into the market. However there is still no standard defined to enable different WFM products to work together.

 The WfMC has standardized a language for describing process definitions: the Workflow Process Definition Language (WPDL). WPDL provides a formal language for the definition and exchange of a process definition using the objects and attributes defined within a *meta-model*. The Meta-Model describes the top level entities contained within a Workflow Process Definition, their relationships and attributes (including some which may be defined for simulation purposes rather than workflow enactment). It also defines various conventions for grouping process definitions into related process models and the use of common definition data across a number of different process definitions or models. Some of the main concepts in the meta-model are *activities*, *transitions*, *workflow relevant data*, and *participants*. A particular processes is defined with the help of such concepts

> *What is a process definition:* the representation of a business process in a form that supports automated manipulation, such as modeling, or enactment by a workflow management system. The process definition consists of a network of activities and their relationships, criteria to indicate the start and termination of the process, and information about the individual activities, such as participants, associated IT applications and data, etc. (WfMC Glossary - WfMC-TC-1011)

Subsequent versions of the meta-model will consider also concepts for describing sophisticated organizational models. Currently these are not standardized.

Process Specification Language (PSL)

URL: http://www.mel.nist.gov/psl

The Process Specification Language (PSL) is a standard developed within NIST and standardized by ISO. It has been designed to facilitate correct and complete exchange of process information among manufacturing and business software systems. Included in these applications are scheduling, process modeling, process planning, production planning, simulation, project management, workflow, and business process reengineering. PSL consists of an ontology for process-related terminology together with a syntax for specifying process descriptions.

The PSL Ontology is organized as a set of theories: PSL-Core (which incorporates the earlier work from the Process Interchange Format project) and a partially ordered set of extensions to PSL-Core. All axioms and definitions for the ontology are written in KIF (the Knowledge Interchange Format). All theories within the PSL Ontology that are currently being standardized have been proven to be sound and complete with respect to the intended semantics of their terminology.

PSL is a project (ISO 18629) within Joint Working Group 8 of Sub-committee 4 (Industrial data) and Sub-committee 5 (Manufacturing integration) of Technical committee ISO TC 184 (Industrial automation systems and integration). Part 1 of the standard has been accepted as a Committee Draft

Core Plan Representation (CPR)
URL: http://projects.teknowledge.com/CPR2

CPR is a model that expresses information common to many plan, process, and activity models. The goal of this effort is to leverage common functionality and facilitate the reuse and sharing of information between a variety of planning and control systems. The CPR embodies a standard that is general enough to cover a spectrum of domains from planning and process management to workflow and activity models. The representation is powerful enough to support complex, hierarchical plan structures. The initial application of the CPR is in addressing plan interchange requirements of several military planning systems, but the model goes beyond military planning and presents a more general plan representation.

Shared Planning and Activity Representation (SPAR)
URL: http://www.aiai.ed.ac.uk/project/spar

SPAR was developed by a group concerned with military plan representation who also had experience of the development of initial standards for plan, process and activity description. A large review team was also involved. This work accounted for the experience gained on the PIF, CPR and NIST PSL standards efforts and involved many of the same people involved in these groups. SPAR deepened and extended the core upper ontologies of PIF and NIST PSL and built upon the experience of applying CPR to a range of military needs. It can be viewed as providing a second level of shared model or ontology beyond what is offered by NIST PSL and is compatible with NIST PSL.

SPAR 0.1a is a very detailed model that also addresses engineering issues of changing ontologies in future. SPAR 0.2 is a much simpler "sentence" level description of the core model (essentially the same as NIST PSL Core) along with a number of terminology and model extensions grouped into packages based on their function. Note that SPAR 0.2 is not a direct replacement for 0.1a and this earlier document could still offer ideas for future ontologies underlying process standards on the web.

Planning Domain Description Language (PDDL)
URL: http://cs-www.cs.yale.edu/homes/dvm

PDDL was developed by the AI Planning Systems (AIPS) Competition Committee for use in defining planning problem domains. It provides a widely-used syntax for expressing STRIPS-like planning operators and other elements of the domain, plus the start and goal states of planning problems.

4.2.2 E-Business Process Standards

These efforts aim at standardization of various business transactions and entire value chain processes between different business partners. Among the main concepts in many of these efforts are business transaction, partner, agreement, contract. Some of these concepts could be seen as extensions of concepts in efforts presented in section 4.2.1, while other are typical for the domain of business transactions (e.g., contract, channel, etc.). We consider e-business standards and initiatives, which describe the individual transaction processes between businesses or provide reference models for entire value chains of business transactions. Since there is a vast collection of possible transactions and transaction services according to the market sector of transactions, we will consider only standards applicable in multiple sectors or such which have cross-sector use (e.g., trading transactions). In future versions of this document, if there is explicit interest from OntoWeb members we will include also vertical standards.

The standards considered are classified into standards for electronic transactions, e-business frameworks for electronic transactions, and web-services. The following standardization efforts are considered.

1. Electronic transactions
   - Electronic Data Interchange (EDI) standards
   - Internet Open Trading Protocol (IOTP)
   - Open Buying on the Internet (OBI)

- Trading Partners Agreement Mark-up Language (tpaML)
2. E-business frameworks for electronic transactions
   - Electronic Business XML (ebXML)
   - National Industrial Information Infrastructure Protocol (NIIIP)
3. Web-services
   - DARPA Agent Mark-up Language for Services (DAML-S)
   - Web Services Description Language (WSDL)
   - Web Services Flow Language (WSFL)
   - Universal Description, Discovery and Integration (UDDI)

4.2.2.1   Electronic Transactions

These efforts aim at standardizing single transactions between business partners. Some of them provide elaborated descriptions of transactions and parts of transactions but without explicitly standardizing the overall process in which these transactions occur. For, example traditional EDI standards provides only lists of documentation of transactions. Other standards try to document transactions describing steps of transaction protocols (e.g., IOTP) or describing the overall business process (value chain) in which a transaction occurs (e.g., new approaches to EDI). The last one often includes some harmonization efforts. Below we list some of the most important developments in such standardization efforts.

Electronic Data Interchange standards (EDI)

URL: http://www.unece.org/trade/untdid/welcom1.htm
          http://www.x12.org
          http://www.xmledi.com
          http://www.cenorm.be/isss/workshop/ebes

> *What is EDI:* Electronic Data Interchange is the exchange of structured messages between business applications. Usually the term EDI refer not only to the content of the messages, but also to the format and even to the underlying computer networks which provide the transport of the messages. Traditional applications of EDI are purchase orders, bills of lading, invoices, shipping orders and payments. However, the development of standards and the widespread use of computers has encouraged the use of EDI in many new arenas including health care insurance and management, record-keeping, financial services, government procurement, and transactions over the Internet.

The content of traditional EDI is meant to directly replace paper documents. As a practical consequence of this, EDI were implemented at a high cost and only between limited number of partners. New initiatives, such as OpenEDI try to define the content of documents according to the workflow in which these documents are used. The hope is that such new approaches to standardization of EDI can be used open environments including many participants. Traditional EDI correspond to very complex vocabularies in the sense adopted in the present paper. They contain non-structured list of business messages (formalizing particular kinds of transactions or parts of transactions). However, the definition of each message is a sort of a reference model characterizing the transaction the message describes. The problem of this traditional approach is that its lacks an overall reference model of the process, with established links between the models of conceptually connected transactions.

The most important and widely used EDI standards are UN/EDIFACT and the ANSI X.12 standards. UN/EDIFACT is supported by the United Nations Center for Trade and Electronic Business (UN/CEFACT) and the e-business board for European Standardization (eBES). UN/CEFACT was established in 1996 with a goal to facilitate international transactions, through the simplification and harmonization of procedures and information flows. eBES is the European entry point for the EDIFACT process. ANSI X.12 is supported by the American National Standards Institute and was created in order to develop uniform standards for inter industry electronic interchange of business transactions. Both EDIFACT and X.12 are standards which claim achieving interoperability between different business sectors, since they are used widely and because they provide means for formalization of big variety of business processes and transactions. However, the interoperability between them is doubtful, not only because of the fact that they use different formats, but also because of semantics considerations. Both the standards are very big and they do not make use of upper level and meta concepts. Recently ANSI and UN/CEFACT have established a program to harmonize the standards through the use of business process modeling techniques compatible with those proposed by the

ebXML initiative. They aim at defining a core set of business processes supported both by EDIFACT and X.12 standards. These core components are to be used in the ebXML. These new developments in EDIFACT and X.12 move them from traditional repository-like EDI toward reference models. A related initiative is XML/EDI initiative whose aim is not the creation of a new EDI standard but rather finding ways of changing the formats used by current EDI standards into XML.

Internet trading protocol (IOTP)
URL: http://www.ietf.org/html.charters/trade-charter.html
Sources: Internet Open Trading Protocol - IOTP, RFC 2801, version 1.0

This protocol is developed by the TRADE WG of IETF. Currently version 1.0 is standardized, while there are some requirements for a forthcoming version 2 of the protocol. IOTP defines reference models for trading transactions based on various trading events. Here is a short overview of it taken from the IOTP specification:

> "The developers of OTP seek to provide a virtual capability that safely replicates the real world, the paper based, traditional, understood, accepted methods of trading, buying, selling, value exchanging that has existed for many hundreds of years. The negotiation of who will be the parties to the trade, how it will be conducted, the presentment of an offer, the method of payment, the provision of a payment receipt, the delivery of goods and the receipt of goods. These are events that are taken for granted in the course of real world trade. OTP has been produced to provide the same for the virtual world, and to prepare and provide for the introduction of new models of trading made possible by the expanding presence of the virtual world. The other fundamental ideal of the OTP effort is to produce a definition of these trading events in such a way that no matter where produced, two unfamiliar parties using electronic commerce capabilities to buy and sell that conform to the OTP specifications will be able to complete the business safely and successfully. In summary, OTP supports: 1) Familiar trading models; 2) New trading models; 3) Global interoperability."

Despite that the trading models defined by IOTP are general enough to capture many individual trading transactions, IOTP, like many of the standards using XML format, lacks precise semantics of its concepts and is difficult for human understanding.

Open Buying on the Internet (OBI)
URL: http://www.openbuy.org

This standard was defined by the Internet Purchasing Roundtable, a forum consisting of big companies and suppliers of indirect materials. After the definition of the standard the OBI consortium was formed to support it. The consortium consists of buying and selling organizations, technology companies, financial institutions and others. The standard consists of an architecture, technical specifications and guidelines. It is based on a model for business-to-business e-commerce procurement process in which the participants are: requisitioner, selling organization, Buying organization, payment authority. The basic principle of the OBI architecture is that process owners are responsible for all the information associated with their business processes.

Trading partners agreement ML (tpaML)
URL: http://xml.coverpages.org/tpa.html

This is a proposal by IBM and is intended for use in ebXML framework . tpaML is based on a model for multi-party e-commerce. The aim is setting-up and maintaining distributed, long running business deals spanning multiple autonomous business organizations. The foundation of tpaML is the Trading Partner Agreement (TPA), which defines how trading partners will interact at the transport, document exchange and business protocol layers. A TPA contains the general contract terms and conditions, participant roles (buyers, sellers), communication and security protocols and business processes, (valid actions, sequencing rules, etc.). The information in a TPA is used to automatically generate configuration information and interaction rules which must be executable by each party's system.

4.2.2.2 E-business frameworks for electronic transactions

These efforts aim not only at standardizing single transactions, but also at providing means for entire value chains or parts of those. In doing this some of them try to develop also reference models of value chain processes in which there are various transactions included which form a part of a production or service process. Some of them (e.g., ebXML) are making use of more general concepts for describing business processes and value chains. Most of the initiatives are in a very preliminary phase of development with respect to the business content standardization and reference models for value chains. However, the OntoWeb interest in such initiatives is obvious, because they need to harmonize transaction standards which will be part of possible value chains.

ebXML (Electronic Business XML)

URLs: http://www.ebXML.org
          http://www.ebtwg.org/projects/bpimeschema.html
          http://www.ebtwg.org/projects/bpschema.html

Sources: ebXML Business Process Specification Schema Version 1.01, Business Process Project Team, May 2001;
Collaboration-Protocol Profile and Agreement Specification, Version 1.0 6, ebXML Trading-Partners Team, May 2001

ebXML's goal is the standardization of XML business specifications and development of a technical framework which enables interoperability, security and consistent use of business information. This has been initiated by UN/CEFACT (United Nations Center for Trade and e-Business) and OASIS (Organisation for the Advancement of Structured Information Standards). The initiative has started in 1999 and was formally closed in May 2001. Currently its activities have been taken up by eBTWG (e-Business Transition ad-hoc Working Group of UN/CEFACT), which is a transition group for a permanent working group on e-Business (see section 5).

The Core Components and Business Models Group of ebXML have been busy with representing a generic semantics of concepts and business processes from different market sectors, while allowing applications to use XML or EDI syntax. The ebXML has not started from scratch: it builds on the long term experience with EDI syntax and semantics and on more recent initiatives for translating EDI to XML. We give a short overview of the final status of the business models and core concepts part of ebXML.

The Business Process team of ebXML has released the first version of a Business Process Specification Schema. This schema provides a standard framework by which business systems may be configured to support execution of business collaborations consisting of business transactions. It is based upon prior UN/CEFACT work, specifically the metamodel behind the EDIFACT (UN/CEFACT Modeling Methodology (UMM) defined in the N090R9.1 specification). This metamodel is based on the REA ontology (Resource, economic Events and economic Agents), which was developed in order to support Generalized Accounting Systems and Shared Data Environments. It is also aimed at Supply Chain Integration. Recently REA has been extended with the concept of economic agreement which is of importance for electronic transactions (in the context of ebXML). The Business Process Specification Schema is available in two formats: XML Schema and UML class diagram.

> "The Specification Schema supports the specification of Business Transactions and the choreography of Business Transactions into Business Collaborations. Each Business Transaction can be implemented using one of many available standard patterns. These patterns determine the actual exchange of Business Documents and business signals between the partners to achieve the required electronic commerce transaction. The current version of the specification schema addresses collaborations between two parties (Binary Collaborations). It is anticipated that a subsequent version will address additional features such as the semantics of economic exchanges and contracts, more complex multi-party choreography, and context based content. "

In addition to the business process specifications there is a specification protocol for an agreement process between trading partners fitting into the general business process schema. It is based on the tpaML protocol discussed above. Further, a lot of work has been done on identifying and classifying common e-commerce processes and specifying per-market core process components. These are

documented in a series of technical reports. These are based on previous works done by the EDI-community.

Current ebXML-related activities are done in a number of projects of the eBTWG Business Proces Modelling Group. For example a Business Process Information Exchange Schema is under development, and the Business Process Specification Schema is being further developed.

National Industrial Information Infrastructure Protocol (NIIIP)
URL: http://www.niiip.org/about-NIIIP.html
Sources: NIIIP Reference Architecture - Concepts and Guidelines, 1996, report nr. NTR96-01

The NIIIP initiative regards not only the process but also the product standards. It is interesting for the OntoWeb network because one of its goals is harmonizing existing standards. Another important component of the consortium work is the development of horizontal technologies that are not yet available, mainly in the area of the work and knowledge management technologies.

The NIIIP work is based on the following assumptions:
- cooperation that is often unscheduled
- both sophisticated and un-sophisticated users
- wide range of systems, environments and processes
- transparency for users
- using de facto and the jure standards, instead of developing new ones

Existing protocols enabled by NIIIP are Mercator, AutoCAD, SPDS, ProductManager, STEP, CATIA. The NIIIP work and knowledge management is based on the work of the CAD Framework Initiative, the ARPA Intelligent Integration of Information, and the WfMC. This work is in turn based on the key communication, object and information technologies, represented by the Internet Community, OMG and STEP.

The NIIIP consortium is formed by a number of organizations that agreed to develop open industry software protocols that allow manufacturers and their suppliers to co-operate in a Virtual Enterprise. The consortium, led by IBM consists of end users, hardware and software suppliers, academia and standards development organizations.

4.2.2.3 WEB-services for B2B e-commerce
There are a number of initiatives which are tailored at describing the processes for delivering of web-based services. Some of the initiatives provides means of describing the process on software level, other include also the business level of services. Below we discuss initiatives which has some contribution on the business content level.

DAML for Services (DAML-S)
URL: http://www.daml.org/services

DAML-S is an ontology for describing Web services, currently under development by a collection of researchers funded by the DAML (Darpa Agent Markup Language) program. The ontology is layered atop the DAML+OIL (http://www.daml.org/language/) ontology language.

The aim is to supply Web service providers with a core set of markup language constructs for describing the properties, capabilities, and behavior of their Web services in unambiguous, computer-interpretable form. DAML-S is designed to facilitate the automation of Web service tasks including discovery, execution, composition and interoperation. In addition to the specification of processes, DAML-S provides for the declaration of "profiles" (advertisements) and "groundings" (bindings to specific communication protocols) of Web services.

Web Services Description Language (WSDL) and Web Services Flow Language (WSFL)
URL: http://www.w3c.org/TR/wsdl
          http://www-106.ibm.com/developerworks/library/ws-ref4/
          http://www-4.ibm.com/software/solutions/webservices/pdf/WSFL.pdf

WSDL, by itself, is not a process or workflow description language, and is also impoverished with respect to describing services, in its current form. But it does become more interesting when coupled with either XLANG or WSFL (both of which are for describing processes). Both of these have some interesting features, both are meant to be used with WSDL, and there is some talk that they may eventually be merged.

Universal Description, Discovery and Integration (UDDI)
URL: http://www.uddi.org

UDDI specifications define a way to publish and discover information about Web-services. The term "web-service" describes specific business functionality exposed by a company, usually through an internet connection, for the purpose of providing a way for another company or software program to use the service. Initially introduced by Ariba, IBM, and Microsoft, UDDI is gaining wide support from org

4.3 Cultural Repositories Standards

Cultural Repository standards fall into two large groups: The terminologies aiming at the standardization of concepts and factual knowledge commonly referred to in data fields of descriptions of objects of cultural heritage, people, communities, activities, places and periods. The second are more or less elaborate schemata and documentation guidelines for the structuring of information and its completeness for certain information goals. The guidelines are based on informal ontologies. The concepts expressed in schemata and guidelines have been recently subject to formalization and harmonization efforts.

4.3.1 Terminologies

The overwhelming predominance of the English language in international electronic communication puts all non-English terminologies on a second place. The investment of the Paul Getty Trust in cultural terminology has allowed for the creation of three terminological systems, which exceed all others in size and coverage. Those are the Art & Architecture Thesaurus (AAT), The Union List of Artist Names (ULAN) and the Thesaurus of Geographic Names (TGN). Particularly the AAT and the TGN can be regarded as a quasi standard due to the numerous applications and translation efforts of the AAT to other languages.
The English Heritage Thesaurus (http://www.rchme.gov.uk/thesaurus/thes_splash.htm) about sites and monuments and the "mda Archeological Objects Thesaurus" (http://www.mda.org.uk/archobj/archcon.htm) are two other important thesauri going partially into more detail than the AAT, but not reaching its coverage.

A unique resource of different kind is the Social Historical and Industrial Classification (SHIC), used in many British museums to classify objects by their relevance to human activities, a concept deserving more attention. AAT, the English Heritage Thesaurus, the mda Archeological Thesaurus and SHIC are structured more or less according to ISO2788. SHIC uses decimal codes characteristic for library classification systems. There are ongoing harmonization efforts between the AAT, the English Heritage Thesaurus, the French MERIMEE thesaurus about architecture and others in the European-funded HEREIN project (http://www.european-heritage.net/en/index.html).

In the wider sense, subject headings of library catalogues play a role in the cultural sector and the growing e-commerce with books and music. The Conference of European National Librarians now harmonizes the American LCSH (Library of Congress Subject Headings), the French RAMEAU and the German Schlagwortnormdatei (SWD) via ISO5964-like equivalence expressions in a project called MACS (http://infolab.kub.nl/prj/macs).

In the following we describe in more details the AAT, TGN, ULAN and SHIC.

Art & Architecture Thesaurus (AAT)
URL: http://www.getty.edu/research/tools/vocabulary/aat/index.html
Sources: Getty AHIP (1994) *Introduction to the Art & Architecture Thesaurus*. Published on behalf of The Getty Art History Information Program, Oxford University Press, New York, 1994.

Soergel, D., 1995. *The art and architecture thesaurus (AAT): A critical appraisal.* Visual Resources, X, 369-400.

The AAT is a structured vocabulary containing around 125,000 terms and other information about concepts. Terms in AAT may be used to describe art, architecture, decorative arts, material culture, and archival materials. Its broad coverage of products as they appear in whatever kind of art object or as architectural element and the general purpose attributes make it useful for product descriptions as well.

The coverage of the AAT ranges from Antiquity to the present, and the scope is global. The Art & Architecture Thesaurus is a compiled resource; it is not comprehensive. The AAT grows through contributions. Information in the AAT was compiled by the Vocabulary Program in collaboration with many institutions, including the following projects:

- Avery Index to Architectural Periodicals
- Bibliography of the History of Art / Bibliographie d'Histoire de l'Art
- Columbia College Library
- Census of Antique Art and Architecture Known to the Renaissance
- Foundation for Documents of Architecture
- Getty Conservation Institute
- Getty Research Library Photo Archive
- Getty Research Library IRIS records
- The J. Paul Getty Museum
- Bildarchiv Foto Marburg
- Mystic Seaport Museum
- National Archives and Records Administration
- National Gallery of Art, Washington
- Smithsonian Institution, National Museum of African Art
- Provenance Index
- Ransom Center, University of Texas, Austin
- University of California, Berkeley
- Getty Vocabulary Program

The AAT addresses the following communities: Archives and Special Collections, Libraries Museums, Visual Resources Collections, Conservation Agencies.

It is a monohierarchical thesaurus compatible with ISO2788. Secondary "BT" relations are denotes at "RT" relations. The polyhierarchy will be physically realized in 2001. The focus of each AAT record is a concept. Each concept is identified by a unique numeric ID. Linked to each concept are terms, related concepts, and one "Broader Term" concept. Terms for any concept can include the plural, singular, natural order, inverted order, spelling variants, various forms of speech, and synonyms that have various etymological roots. The AAT provides not only the terminology to generically name art objects and architecture, but the vocabulary necessary to describe them as well. This supporting terminology includes the materials and techniques relating to their construction and conservation (such as deacidification), their physical attributes (such as shape and colour), terminology associated with their production and study (such as the roles of persons), vocabulary indicating their style or period, and concepts relating to their history, theory, criticism, and purpose.

Facets constitute the major subdivisions of the AAT hierarchical structure. A facet contains a homogeneous class of concepts, the members of which share characteristics that distinguish them from members of other classes. The AAT is "postcoordinated", i.e. terms from different facets can be combined to "compound terms". The relationships between the combined terms is thought to be defined per default from the nature of the facet itself. E.g. "factory + grinding" is understood as "factory which does grinding", a usual practice in subject catalogues. Object terms combined with materials are understood as being made of that material, etc.

Facets and Hierarchies in the AAT:

ASSOCIATED CONCEPTS FACET
Hierarchy: Associated Concepts

PHYSICAL ATTRIBUTES FACET
Hierarchies: Attributes and Properties , Conditions and Effects , Design Elements , Colour

STYLES AND PERIODS FACET
Hierarchy: Styles and Periods

AGENTS FACET
Hierarchies: People , Organizations

ACTIVITIES FACET
Hierarchies: Disciplines , Functions , Events , Physical, Activities , Processes and Techniques

MATERIALS FACET
Hierarchy: Materials

OBJECTS FACET
Hierarchies: Object Groupings and Systems , Object Genres, Components; Built Environment: Settlements and Landscapes , Built Complexes and Districts , Single Built Works, Open Spaces and Site Elements; Furnishings and Equipment: Furnishings , Costume, Tools and Equipment , Weapons and Ammunition, Measuring Devices , Containers , Sound, Recreational Artefacts , Transportation Vehicles; Visual and Verbal Communication: Visual Works, Exchange Media

Thesaurus of Geographic Names (TGN)
URL: http://www.getty.edu/research/tools/vocabulary/tgn/index.html

The TGN is a structured vocabulary containing around 1,000,000 names and other information about places. It is a gigantic work and one of the largest resources of its kind. The TGN includes all continents and nations of the modern political world, as well as historical places. The emphasis in TGN is on places important for art and architecture. Nevertheless it has achieved a remarkable coverage of the world, making it a unique link between textual references to places and geometric referencing or GIS, useful far beyond art and architecture. Being about factual knowledge, it cannot be regarded as an ontology, but it plays a similar practical role.

Its notion of "place" is rather intuitive. Places in the TGN can be either physical or political entities. They include physical features such as continents, rivers, and mountains; and political entities, such as empires, nations, states, districts, townships, cities, and neighbourhoods. The "place type" in the TGN is a term that characterizes a significant aspect of the place, including its role, function, political anatomy, size, or physical characteristics. Place types are indexing terms chosen from the structured vocabulary of the AAT. It represents the spatial inclusion as a monohierarchy representing the current political and physical world and normally goes down to the level of larger settlements. The polysemy of "place" makes the notion of unique inclusion fairly ambiguous. The TGN tries to accommodate for that by secondary, "non-preferred", inclusion relationships. Another weakness is the use of centres for spatial coordinates instead of minimal rectangles covering the respective area.

The focus of each TGN record is a place, represented by a unique numeric ID in the database. Linked to the record for the place are names, a "parent" or position in the hierarchy, other relationships, geographic coordinates, notes, sources for the data, and "place types," which are terms describing the role of the place (e.g., "inhabited place" and "state capital"). Names for any place can include the vernacular, English, other languages, historical names, natural order, and inverted order. Among these names, one is flagged as the preferred name, or "descriptor."

Contributions: The Getty Thesaurus of Geographic Names is a compiled resource; it is not comprehensive. The TGN grows through contributions. Information in the TGN was compiled by the Vocabulary Program in collaboration with many institutions, including the following projects:

- Art Libraries Society of North America
- Avery Index to Architectural Periodicals
- Bibliography of the History of Art / Bibliographie d'Histoire de l'Art
- Foundation for Documents of Architecture
- Getty Conservation Institute
- Getty Research Library Photo Archive
- The J. Paul Getty Museum
- Provenance Index
- Getty Vocabulary Program

Social, Historical, Industrial Classification (SHIC)

URL: http://www.holm.demon.co.uk/shic.htm

Sources: " Social, Historical, Industrial Classification (SHIC), a subject classification for museum cataloguing", second edition published for the SHIC Working Party 1993 by The Museum Documentation Association, Cambridge, UK. ISBN 0 905963 91 1

The Social History and Industrial Classification (SHIC) is a subject classification for museum cataloguing widely used by UK museums. It was created nearly 20 years ago by the SHIC Working Party which continues to develop it. The first published edition appeared in 1983 and a revised second edition was published in 1993 with further updates in 1996.

It is designed to make links between a wide variety of museum material - objects, photographs, archival material, tape recordings, information files - according to the sphere of human activity with which they are primarily associated. The aim of the Classification is to make relationships between objects, etc. via their context. A carpenter's hammer, for example, is used with other carpenter's tools, and is therefore classified alongside those related items. SHIC does not group material according to generic type or family similarities. A carpenter's hammer could be classified with all other percussive tools regardless of the trades and industries in which they were used but this is not the aim of SHIC. Other classification systems can be used in parallel to
achieve this.

SHIC can cope with abstract concepts and material of a very general nature in addition to more precisely defined items. It has a hierarchical structure with levels that run from the general to the specific. As a rule the higher levels are more conceptual whilst the lower levels are more directly object based. There are four primary headings (Sections) covering all aspects of human activity:

1. COMMUNITY LIFE
2. DOMESTIC AND FAMILY LIFE
3. PERSONAL LIFE
4. WORKING LIFE

These four Sections are considered equal in importance and are not mutually exclusive. Each heading within these levels can be divided into ten subdivisions between 0 and 9. Of these subdivisions the first (0) is always reserved for general or unprovenanced material and the last (9) is usually used to group together minor miscellaneous categories which cannot otherwise be accommodated within
the other headings. They can then be separated at the next level down if necessary. The SHIC hierarchy has a characteristic depth of five to eight levels, e.g.:

4. Working Life

Division	4.3 Minerals and chemicals
Class	4.33 Chemical Industry
Group	4.335 Soap and toilet preparations
Subgroup	4.3351 Soap and synthetic detergents
Activity Subdivision	4.3351.6 External operations
	4.3351.62 Advertising and publicity

The Classification is capable of expansion and additional headings can be created as required. Normally no attempt will be made to provide a new heading until an accumulation of material warrants this. The SHIC Working Party will continue to meet to review the Classification.

Union List of Artist Names (ULAN)
URL: http://www.getty.edu/research/tools/vocabulary/ulan/index.html

The ULAN is a structured vocabulary that contains around 220,000 names and other information about artists. The coverage of the ULAN is from Antiquity to the present, and the scope is global. The scope of the ULAN includes any identified individual or "corporate body" (i.e., a group of people working together) involved in the design or creation of art and architecture. The focus of each ULAN record is the artist. Artists in the ULAN generally represent creators involved in the conception or production of visual arts and architecture. Some performance artists are included (but not actors, dancers, or other performing artists). Its commercial utility is mainly in the trading with art objects. It is interesting from a methodological point of view, and could be widely extended. Being about factual knowledge, it cannot be regarded as an ontology, but it plays a similar practical role

The ULAN is compiled from artists' names and biographical information that has been collected by various Getty projects and by other institutions outside the Getty. When multiple contributors have submitted information about the same artist or corporate body, all the names and information about this person or corporate body have been merged into a single record.

Linked to each record are names, relationships (including student-teacher relationships), locations (for birth, death, and activity), important dates (such as for birth and death), notes, and sources for the data. Names for any artist can include the vernacular, English, other languages, natural order, inverted order, nicknames, and pseudonyms. Among these names, one is flagged as the preferred name, or "descriptor."

Terms and concepts in the ULAN are linked by three types of relationships: Equivalence, associative, and hierarchical relationships.
 All relationships between names within the same ULAN record are equivalence relationships. One name is flagged as the preferred name, which is the indexing form of the name most often found in scholarly or authoritative publications.
 Associative relationships may exist between and among people and corporate body records in ULAN. For example, an artist may have a student/teacher relationship with his master. Family relationships are noted only if the family member was an artist or other important historical figure. Also, corporate bodies and other groups of individuals may be related to single individuals, as a workshop or architectural firm should be related to its members. An architectural firm that has reorganized with new partners may have a relationship with the original firm. There may be hierarchical relationships between corporate bodies in ULAN. Corporate bodies may have hierarchical administrative structures, and these hierarchical relationships are recorded by using separate ULAN records that are linked.

Contributions: The Union List of Artist Names is a compiled resource; it is not comprehensive. The ULAN grows through contributions. Information in the ULAN was compiled by the Vocabulary Program in collaboration with many institutions, including the following projects:
   • Avery Index to Architectural Periodicals
   • Bibliography of the History of Art / Bibliographie d'Histoire de l'Art
   • Canadian Centre for Architecture
   • Census of Antique Art and Architecture Known to the Renaissance
   • Getty Conservation Institute
   • Foundation for Documents of Architecture
   • Frick Art Reference Library
   • Getty Research Library Photo Archive
   • Getty Research Library IRIS records JPGM
   • The J. Paul Getty Museum
   • Provenance Index
   • Getty Vocabulary Program
   • Witt Checklist of Painters c1200-1976
   • Witt Computer Index

4.3.2 Data structure standards

The term "culture" and "cultural heritage" is virtually unlimited with respect to human activities. As e.g. via science museums, history of science, science becomes an object of cultural considerations. The most prominent common denominator about cultural repository data seems to be the orientation to history and objects from the past, but performing arts and other cultural events are also subject to future-oriented information. This makes the creation of common schemata for cultural repositories a nearly impossible task. There are some dozens more important formats, and may be a thousand idiosyncratic ones in use (RLG talk). Among the first ones are:

- The CIDOC Relational Data Model (http://www.cidoc.icom.org/pub1.htm)
- Consortium for the Computer Interchange of Museum Information (CIMI) DTD, (http://www.cimi.org/publications.html).
- mda SPECTRUM data fields (http://www.mda.org.uk/spectrum.htm).
- Council for the Prevention of Art Theft Object ID (http://www.object-id.com).
- Research Libraries Group (RLG) Cultural Materials Initiative DTD (http://www.rlg.org).
- Consortium for the Computer Interchange of Museum Information (CIMI) Z39.50 Profile (http://www.cimi.org/old_site/downloads/ProfileFinalMar98/cimiprofile1.htm).
- Dublin Core (http://www.dublincore.org).
- The International Committee for Documentation of the International Council of Museums (CIDOC): The International Core Data Standard for Archaeological and Architectural Heritage (http://natmus.min.dk/cidoc/archsite/coredata/arch1/htm).
- Core Data Index to Historic Buildings and Monuments of the Architectural Heritage (http://www.object-id.com/heritage/core.html).
- English Heritage MIDAS - A Manual and Data Standard for Monument Inventories (http://www.rchme.gov.uk/midas/index.html).
- Art Museum Image Consortium (AMICO) data dictionary (http://www.amico.org).

Among those, mda SPECTRUM has become the de facto standard for museum documentation software, a very elaborate document describing documentation processes and an unstructured list of recommended data fields. Dublin Core, on the other side, is may be the format mostly promoted as standard - only 15 data fields or properties thought to be the common denominator of many formats. It reflects quite well library applications, but its application in museum documentation is not regarded convincing by enough experts. Nevertheless many cultural repositories use it to define structured finding aids. Non of the above are ontologies, but implementation oriented data structures. The above list does not even comprise Natural History or the needs of science museums. Related but not overlapping is the OPENGIS standardisation effort (http://www.oprngis.org/) for interoperability of GIS through the Internet, as many cultural applications run on GIS or with the assistence of GIS.

In addition to that, there are some prescription with considerable elaboration of the implied notions. Those are:

- mda SPECTRUM – the definition of museum documentation processes.
- The International Guidelines for Museum Object Information: The CIDOC Information Categories.
- International Federation of Library Associations and Institutions (IFLA) Functional Requirements for Bibliographic Records (FRBR) (http://www.ifla.org/VII/s13/frbr/frbr1.htm).

Those documents have become very influential in the domain as sources for the related concepts. FRBR defines also an E-R diagram.

Since 1996 efforts began to solve the interoperability problem of data structures in cultural repositories via conceptual reference models or ontologies. Those are: The CIDOC CRM (http://cidoc.ics.forth.gr), the IndeCS Model (http://www.indecs.org), and the ABC Harmony Model (http://metadata.net/harmony/Results.htm). Whereas the first comes from the museum world, the other two come from a library and multimedia market perspective. IndeCS is a project, which addressed the problem of tracing intellectual property rights in the media market. All three models have in common

that they regard the explicit modelling of events as the element that allows to integrate data about material and immaterial objects, sites, people, activities and history in a large scale. Closest to a real standard and richest in semantics is the CIDOC CRM, which is currently in the Committee Draft Stage at ISO TC46. There are ongoing harmonization efforts between the CIDOC CRM and ABC Harmony.

The CIDOC CRM can be regarded on one side as an ontology to describe the semantics behind data structures of cultural repositories, and on the other side in can be seen as a top-level ontology for integrating the various terminologies. There hasn't however been any effort so far to actually harmonize specific terminologies and the CIDOC CRM.

The CIDOC Conceptual Reference Model (CIDOC CRM)
URL: http://cidoc.ics.forth.gr

The "CIDOC object-oriented Conceptual Reference Model" (CRM), was developed by the ICOM/CIDOC Documentation Standards Group since 1996. Since September 2000, the CRM is being developed into an ISO standard in a joined effort of the CIDOC CRM SIG and ISO/TC46/SC4/WG9. It represents an 'ontology' for cultural heritage information i.e. it describes in a formal language the explicit and implicit concepts and relations relevant to the documentation of cultural heritage. The primary role of the CRM is to serve as a basis for mediation of cultural heritage information and thereby provide the semantic 'glue' needed to transform today's disparate, localised information sources into a coherent and valuable global resource.

The intended scope of the CRM may be defined as all information required for the scientific documentation of cultural heritage collections, with a view to enabling wide area information exchange and integration of heterogeneous sources. This definition requires some explanation:

a) The term scientific documentation, is intended to convey the requirement that the depth and quality of descriptive information which can be handled by the CRM should be sufficient for serious academic research into a given field and not merely that required for casual browsing. This does not mean that information intended for presentation to members of the general public is excluded, but rather that the CRM is intended to provide the level of detail and precision expected and required by museum professionals and researchers in the field.

b) The term cultural heritage collections is intended to cover all types of material collected and displayed by museums and related institutions, as defined by ICOM (1). This includes collections, sites and monuments relating to natural history, ethnography, archaeology, historic monuments, as well as collections of fine and applied arts. The exchange of relevant information with libraries and archives, and the harmonisation of the CRM with their models, fall within the CRM's intended scope.

c) The documentation of collections is intended to encompass the detailed description both of individual items within collections as well as groups of items and collections as a whole. The scope of the CRM is the curated knowledge of museums. Information required solely for the administration and management of cultural heritage institutions, such as information relating to personnel, accounting, and visitor statistics, falls outside the intended scope.

d) The CRM is specifically intended to cover contextual information: the historical, geographical and theoretical background in which individual items are placed and which gives them much of their significance and value.

e) The goal of enabling information exchange and integration between heterogeneous sources determines the constructs and level of detail of the CRM. It also determines its perspective, which is necessarily supra-institutional and abstracted from any specific local context.

f) The CRM aims to leverage contemporary technology while enabling communication with legacy systems.

 The initial practical scope of the CRM was defined by the International Guidelines for Museum Object Information: The CIDOC Information Categories , published in June 1995. This document, edited by a joint team of the CIDOC Data and Terminology and the Data Model Working Groups, resulted from the consolidation of two parallel initiatives: the Information Categories for Art and Archaeology Collections, 1992 and the CIDOC Relational Data Model 1995, both of which had been in gestation since 1980. The Guidelines thus represent the fruit of many years of collective effort and reflection concerning museum information. The CIDOC CRM intends to cover the semantics of all formats and guidelines referred in this section above, to the degree they fall into the intended scope.

The CIDOC CRM Version 3.2 has been proposed as Committee Draft to ISO TC46 in October 2001. This version consists of 75 high-level entities and 108 properties. It employs the knowledge representation language TELOS. Characteristic is the use of multiple inheritance for entities (classes) and properties (attributes). Properties (attributes) are objects that connect two entities, and which can have properties (attributes) of their own. The CRM foresees the use of multiple instantiation for entity and property instances. An RDFS version of the CRM is provided.

Fundamental entities are:
1. Temporal Entity - Phenomena limited in time
2. Stuff - Material and immaterial objects and features
3. Actor - People and organisations that can act in a legal sense.
4. Appellation - Names and identifiers used in the universe of discourse
5. Types - Concepts from terminologies suitable to refine the CRM Entities
6. Place - Geometric place on earth or some object independent from temporal changes.
7. Time - The temporal dimension in the sense of physics.

The CRM declares properties describing the relative roles of the above entities and specializations of them. In particular about:
1. identification
2. part-whole decomposition
3. participation in temporal entities, begin and end of existence
4. location and physical properties
5. ownership
6. motivations of activities

Its structure is tailored for the integration of information about the past. It tries to accommodate for different degrees of knowledge about the same facts in different sources. It does not contain models for description of planned events.

5 Standardisation bodies and other organisations

In this section descriptions of some important standards organizations are presented. We have used the typology of standards organizations adopted by the Diffuse project (see below). According to this typology there are formal, consortium and other standard-related organizations. The formal and consortium organizations are standardization bodies, which develop or/and approve standard specifications. They have a regulatory character - usually standards adopted by them should be implemented in the countries which are members of such organizations. Formal standardization bodies can consist of governmental or private organizations. The consortium type standardization bodies consists of industrial organizations and produce the de-facto standards. These standards can be further adopted, modified, and approved by formal standardization bodies on national and international levels. In the category "other" we consider initiatives and organizations which monitor, harmonize, maintain, or facilitate the application of already adopted standards.

5.1 Formal Standardisation Bodies

5.1.1 ISO - International Standards Organisation

URL: http://www.iso.ch/iso/en/ISOOnline.frontpage

ISO is a non-governmental organisation established in 1947. It is a worldwide federation of about 140 national standards bodies. Its mission is to promote the development of standardization and related activities in the world with a view to facilitating the international exchange of goods and services, and to developing cooperation in the spheres of intellectual, scientific, technological and economic activity. ISO's work results in international agreements that are published as International Standards. The scope of ISO is not limited to any particular industry; it covers all areas of standardization except electrical and lectronic engineering which is responsibility of IEC.

The technical work of ISO is highly decentralized, carried out in a hierarchy of some 2 850 technical committees, subcommittees and working groups. In these committees, qualified representatives of industry, research institutes, government authorities, consumer bodies, and international organizations from all over the world come together as equal partners in the resolution of global standardization problems. Although the greater part of the ISO technical work is done by correspondence, there are, on average, a dozen ISO meetings taking place somewhere in the world every working day of the year. Some 30 000 experts participate in meetings each year.

There are three main phases in the ISO standards development process.

1. *Defining the scope of a standard.* The need for a standard is usually expressed by an industry sector, which communicates this need to a national member body. The latter proposes the new work item to ISO as a whole. Once the need for an International Standard has been recognized and formally agreed, the first phase involves definition of the technical scope of the future standard. This phase is usually carried out in working groups which comprise technical experts from countries interested in the subject matter.
2. *Detailed specification.* In this phase countries negotiate the detailed specifications within the standard.
3. *Formal approaval:* this is the final phase in which the draft specification is formally approved by voting follwoig previously established voting scheme. The approved draft is published as an ISO International Standard.

Some of the ISO Technical Committees which are of interest to the OntoWeb SIG on Content Standards are the follwoing.
- *TC46 (Information and documentation)* standardizes practices relating to libraries, documentation and information centers, indexing and abstracting services, archives, information science and publishing. It is coordinated by AFNOR, the French National Standardization Body.

- *TC 154 (Processes, data elements and documents in commerce, industry and administration)* is coordinated by SNV, the Swiss Association for Standardization. Its activities relate to international standardization and registration of business and administration processes, supporting data used for information interchange between and within individual organizations, and supporting standardization activities in the field of industrial data.
- *TC 184 (Industrial automation systems and integration)* is coordinated by AFNOR. Its scop[e is standardization in the field of industrial automation and integration concerning discrete part manufacturing and encompassing the application of multiple technologies, i.e. information systems, machines and equipment, and telecommunications.
- *TC 213 (Dimensional and Geometrical Product Specifications).*
- *TC 215 (Health informatics)* is coordinated by ANSI, the American National Standards Institute. Standardization in the field of information for health, and Health Information and Communications Technology (ICT) to achieve compatibility and interoperability between independent systems. Also, to ensure compatibility of data for comparative statistical purposes (e.g. classifications), and to reduce duplication of effort and redundancies.

5.1.2 CEN - European Committee for Standardisation

URL: http://www.cenorm.be

CEN is an international association set up to manage cooperation between European National Standards Bodies (NSBs). The objective of CEN is to produce (create or adopt) standards that are relevant throughout Europe.

The Standards Program is controlled by the Technical Board of CEN. The standardization activities of CEN are done as part of ISO standardization, adopting existing ISO standards or are organized in Technical Committees which often base their work on existing ISO standards. The Technical Board coordinates the following technical bodies:
- *CEN Technical Committees* are responsible for the planning and the management of the standards making process.
- *CEN Workshops* are new open environments for producing specifications on a consensus basis, as CEN Workshop Agreements, pre-Standards, guidance or other material. The aim of the CEN Workshops is bridging the gap between industrial consortia that produce de facto standards, and the formal European standardization process which produces standards through consensus under the authority of the CEN member bodies.
1. *Associated standards bodies.* When CEN feels it appropriate that in specific fields the preparatory work can be done by an existing organization, it can decide to work in collaboration with Associated Standards Bodies.
1. *Other organizations.* CEN also has agreements with a number of organizations which either prepare specifications which are then processed through CEN as European Standards or with whom it works in tight cooperation.

CEN has established the Information Society Standardization System (ISSS) as a single unit for CEN activities in the ICT field. CEN/ISSS has established a wide range of workshops, including the ones on Electronic Commerce (EC) and e-Business Board for European Standardization (eBES). The current activities of EC workshop make it be of a special interest for the Ontoweb SIG on Content Standards.

5.1.2.1   CEN/ISSS Electronic Commerce Workshop

CEN/ISSS Electronic Commerce Workshop   offers a coherent and cohesive focus for EC standardization at a European level, within the context of global EC standardization activities. The basic scope of the EC Workshop's technical projects is the core elements of Electronic Commerce, together with strategy/awareness activities which promote and complement the technical work.

Collaborative and liaison activities are very important in this all-encompassing field to ensure that no duplication of work, which might waste resources, occurs. The Workshop takes into consideration relevant work in other organizations, and maintains appropriate liaison with a range of CEN/ISSS Workshops, as well as with other external activities.

A particularly interesting CEN/ISSS EC Project is MULECO - Multilingual Upper-Level Electronic Commerce Ontology. It has been approved on the last meeting of the Workshop (in October 2001).

5.2 Industrial and Public Consortiums

5.2.1 World Wide Web Consortium

URL: http://www.w3c.org
Sources: information on http://www.diffuse.org

The W3C was founded to develop common protocols for the evolution of the World Wide Web. It is an international industry consortium, jointly hosted by the Massachusetts Institute of Technology Laboratory for Computer Science [MIT/LCS] in the United States, the French Institut National de Recherche en Informatique et en Automatique [INRIA] in Europe, and the Keio University Shonan Fujisawa Campus in Asia. Currently it has more than 500 member organizations.

W3C provides a vendor-neutral forum for its Members to address Web-related issues. Working together with its staff and the global Web community, the Consortium aims to produce free, interoperable specifications and sample code. Funding from membership dues, public research funds, and external contracts underwrite these efforts. W3C's long term goals for the Web are:

2. **Universal Access:** To make the Web accessible to all by promoting technologies that take into account the vast differences in culture, education, ability, material resources and physical limitations of users on all continents;
3. **Semantic Web:** To develop a software environment that permits each user to make the best use of the resources available on the Web;
4. **Web of Trust:** To guide the Web's development with careful consideration for the novel legal, commercial, and social issues raised by this technology.

W3C concentrates its efforts on three principle tasks:

(c) **Vision:** W3C promotes and develops its vision of the future of the World Wide Web.
(d) **Design:** W3C designs Web technologies to realize this vision, taking into account existing technologies as well as those of the future. The fundamental design principles of the Web as an application built on top of the Internet are: Interoperability, Evolution and Decentralization.
(e) **Standardization:** W3C contributes to efforts to standardize Web technologies by producing specifications (called "Recommendations") that describe the building blocks of the Web. W3C makes these Recommendations (and other technical reports) freely available to all.

W3C Activities are generally organized into groups: Working Groups (for technical developments), Interest Groups (for more general work), and Coordination Groups (for communication among related groups).

There are five Domains: *Architecture, Document Formats, Interaction, Technology and Society,* and the *Web Accessibility Initiative*. Each Domain is responsible for investigating and leading development in several Activity Areas which are critical to the Web's global evolution and interoperability.

(f) **Architecture:** Enhancing the infrastructure of the Web and increasing its automation. Includes: Document Object Model (DOM); Jigsaw server; Uniform Resource Identifier (URI); Extensible Markup Language (XML); XML Protocol.
(g) **Document Formats:** Improving the technology that allows Web users to effectively perceive and express information. Includes: Amaya browser; Graphics; including Portable Network Graphics (PNG), Web Computer Graphics Metafile (WebCGM), and Scalable Vector Graphics (SVG); Hypertext Markup Language (HTML); Internationalization (I18N); Maths, including Mathematical Markup Language (MathML); Style Sheets, including Cascading Style Sheets (CSS) and eXtensible Style Sheet Language (XSL).
(h) **Interaction:** Exploring new ways to access Web information. Includes: Device Independence; Synchronized Multimedia, including Synchronized Multimedia Integration Language (SMIL); Voice Browser.

(i) **Technology and Society:** Understanding the social impact of the Web and reaching out to affected communities. Includes: Privacy, including Platform for Privacy Preferences (P3P); Semantic Web; XML Encryption; XML Signature.

(j) **Web Accessibility Initiative** (WAI). Improving accessibility to web resources for those with disabilities. This work is split into a WAI International Program Office (IPO) and a WAI Technical Activity.

In addition to the five Domains, the **Quality Assurance** (QA) Activity has been launched with a Working Group and Interest Group whose primary mission is to improve the quality of W3C specification implementation in the field.

5.2.2 United Nations Centre for Trade and Electronic Business; UN/CEFACT

URL: http://www.unece.org/cefact

UN/UNCEFACT is a consortium consisting of Member States, governmental and industrial organizations recognized the Economic and Social Council of the United Nations. Within the United Nations, UN/CEFACT is located in the Economic Commission for Europe (UN/ECE), which is part of the United Nations network of regional commissions. The activities of UN/CEFACT are related to the worldwide facilitation of international transactions, through the simplification and harmonization of procedures and information flows.

The technical structure of UN/CEFACT is based on permanent and ad-hoc working groups. Some of the permanent working groups which are of potential interest to the OntoWeb SIG on content standards are:

(a) Business Analysis Working Group (BAWG). The purpose of the Business Analysis Working Group (BAWG) is to analyze current business processes , to identify constraints that adversely impact on the mission and objectives of CEFACT, and to propose appropriate changes to those business processes.

(b) UN/EDIFACT Working Group (EWG). This Working group develops the UN/EDIFACT standard. It is divided into sub-groups according to vertical industry sectors. For example there are subgroups for finance, transport, healthcare, etc.

(c) Techniques and Methodologies Working Group (TMWG)

Currently, there are three ad-hoc working groups, two of which are of potential interest for the OntoWeb SIG on Content Standards. These are:

4. *Electronic Commerce Ad-hoc Working Group (EC AWG)*

5. *eBusiness Transition Ad-Hoc Working Group (eBTWG)*. The purpose of this group is to serve as a bridge from the recently completed ebXML initiative to a future permanent working group on eBusiness. The mission of the eBTWG is to identify specific work items to facilitate the completion of the activities related to the ebXML Business Process and Core Components Projects and to oversee the further development of those items. In addition, this group would be responsible for developing and maintaining the UN/CEFACT eBusiness architecture to ensure consistency with the ebXML architecture specification.

5.3 Other organisations

5.3.1 National Institute of Standards and Technology- NIST

URL: http://www.nist.gov

National Institute of Standards and Technology (NIST) is a non-regulatory federal agency within the U.S. Commerce Department's Technology Administration. NIST's mission is to develop and promote measurement, standards, and technology to enhance productivity, facilitate trade, and improve the quality of life. NIST carries out its mission through four interwoven programs:

- *NIST Laboratories* that provide technical leadership for the technology infrastructure needed by U.S. industry to continually improve its products and services;

- *A quality program* that recognizes quality achievement by U.S. manufacturers, service companies, educational organizations, and health care providers;

- *the Manufacturing Extension Partnership*, a nationwide network of local centers offering technical and business assistance to smaller manufacturers; and

- *the Advanced Technology Program,* accelerating the development of innovative technologies for broad national benefit through R&D partnerships with the private sector.

NIST is not a formal standards body, rather it participates in the technical work of such bodies and industrial consortium providing to them its technical expertise.

Of particular interest to the OntoWeb SIG on Content Standards is the Manufacturing Engineering Lab and especially its Manufacturing Systems Integration Division (MSID) of which a short account is presented below.

Manufacturing Systems Integration Division (MSID)
URL: http://www.mel.nist.gov/msid/standard.htm

"MSID promotes developing technologies and standards that lead to implementing information-intensive manufacturing systems. Such systems can be integrated into a national network of enterprises working together to make U.S. industry more productive. To accomplish its mission, MSID works closely with industry (both as individual companies and as consortia), other government agencies, standards bodies, and universities."

Among the technical standardization activities of MSID are those related to the domains of intelligent agents, enterprise modeling, Enterprise Resource Planning, Product Data Management. MSID has participated actively in the development of standards such as STEP and PSL. It collaborates with formal standardization bodies as well as industrial consortia, such as ISO, OMG, OAG to name just a few.

5.3.2 Diffuse project
URL: http://www.diffuse.org

The Diffuse project has been set up to provide neutral reporting on developments relating to standards and specifications in support of Key Action II (New Methods of Work and Electronic Commerce) and Key Action III (Multimedia Content and Tools) of the European Commission's Information Society Technologies (IST) programme. The project outputs are primarily targeted at potential and actual IST participants. Whilst the emphasis of the project is focused on the needs of the Research and Technologies Development (RTD) communities, it also has a broader perspective of serving the information requirements of industry and public sector in general. The Diffuse project builds on the accomplishments of the European Commission's Open Information Interchange (OII) initiative, which concluded in December 1999.

The Diffuse project is implemented by the Diffuse consortium consisting of TIEKE (Finnish Information Society Development Centre) in the role of coordinator and IC Focus and the SGML Center in the role of partners.

Among the Diffuse services are the following:

- **Business Guides.** This service provides information of how to apply existing standards in key technology areas.
- **Standards and Specifications Reference Data** . This service provides information on existing standards (standards and specifications list), on the activities of standardization organisations (standard fora list), and on existing European RTD projects.
- **Standards and Specifications News**. News regarding Electronic Commerce, Information Management, RTD Project, and Conferences are being published in regular intervals.

6 References

General References

[Boer00] Boertien, N., Bijlsma M., Stefanova M., *State of the art in e-business services and components,* Giga TS deliverable, Janssen, W., van der Stappen P. editors, Telematica Institute, The Netherlands, https: //extranet.telin.nl, 2000

[BPMI01] BPMI.org, *Business Process Modeling Language (BPML)* , Working Draft 0.4, August, 2001

[Cor2001] Corcho, O. Gómez-Pérez, A. "Integrating e-commerce standards and initiatives in a multi-layered ontology". *IJCAI'01 Workshop on Ontologies and Information Sharing*.

[ebBP01] ebXML Business Process Project Team, *ebXML Business Process Specification Schema Version 1.01*, May 2001

[ebTP01] ebXML Trading Partners Team, *Collaboration-Protocol Profile and Agreement Specification*, Version 1.0 6, May 2001

[Guar00] Guarino, N., Welty, C., and Partridge, C., *Towards Ontology-based harmonization of Web content standards,* In S. Liddle, H. Mayr and B. Thalheim (eds.), Conceptual Modeling for E-Business and the Web: Proceedings of the ER-2000 Workshops. Springer Verlag: 1-6, 2000

[Jan00] Janssen, W., Koolwaaij J., Stefanova M., *The Extensible Markup Language, hype or hope?*, Giga TS deliverable, Telematica Institute, The Netherlands, 2000

[IETF] *Internet Open Trading Protocol - IOTP*, RFC 2801, version 1.0

[NIIIP96] NIIIP, *NIIIP Reference Architecture - Concepts and Guidelines*, 1996, report nr. NTR96-01

[WfMC95] Workflow Management Coalition, *The Workflow Reference Model*, Document Number TC00-1003, January 1995

Standards web-pages

AAT, Art and Architecture Thesaurus, http://www.getty.edu/research/tools/vocabulary/aat/index_html

ANX, Automobile Network eXchange, http://www.anxo.com

BizTalk, http://www.biztalk.org

BPML, Business Process Modeling Language, http://www.bpmi.org

CBL, Common Business Language, http://www.commerceone.com/news/us/cbl2.htm

CDIF, Common Data Interchange Framework, http://www.eigroup.org/cdif/index.html

CIDOC CRM, Conceptual Reference Model, http://cidoc.ics.forth.gr

CIMI Profile, Z39.50 Profile for Cultural Heritage Information, http://www.cimi.org

CPR, Common Plan Representation, http://projects.teknowledge.com/CPR2

CWM, Common Warehouse Model, http://www.cwmforum.org

cXML, Commerce XML, http://www.cxml.org

CYC, http://www.cyc.com, http://www.opencyc.org

DAML-S, DAML for Services, http://www.daml.org/services

DCMI, Dublin Core Metadata Initiative, http://www.dublincore.org

EbXML, Electronic Business XML, http://www.ebXML.org

EPISTLE, European Process Industries STEP Technical Liaison Executive, http://www.stepcom.ncl.ac.uk/epistle/epistle.htm

FpML, Financial Products Mark-up Language, http://www.fpml.org

HL7, Health Level 7, http://www.hl7.org

ICE, Information and Content Exchange, http://www.oasis-open.org/cover/ice.html

IDEF5, Integration Definition for Function: Modeling Ontology Description Capture, http://www.idef.com

INDECS, Interoperability of Data in e-Commerce Systems, http://www.indecs.org

IOTP, Internet Open Trading Protocol, http://www.ietf.org/html.charters/trade-charter.html

ISITC, International Securities Association for Institutional Trade Communication, http://www.isitc.org

ISO IEC 11179, Metadata Registry Coalition, http://www.sdct.itl.nist.gov/~ftp/18/other/coalition/Coalition.html

MDCOIM, MDC Open Information Model, http://www.mdcinfo.com
MOF, Meta Object Facility, http://www.omg.org
NIIIP, National Industrial Information Infrastructure Protocol, http://www.niiip.org/about-NIIIP.html
NIIIS, National Industrial Information Infrastructure, http://www.niiis.org
OBI, Open Buying on the Internet, http://www.openbuy.com
OCF, Online Catalog Format, http://www.martsoft.com/ocp
OFX, Open Financial Exchange, http://www.ofx.net/ofx
PDDL, Planning Domain Description Language, http://cs-www.cs.yale.edu/homes/dvm
POSC-CAESAR, Petrotechnical Domain Description Language, http://www.posccaesar.com
PSL, Process Specification Language, http://www.mel.nist.gov/psl
RosettaNet, http://www.rosettanet.org
SPAR, Shared Planning and Activity, http://www.aiai.ed.ac.uk/project/spar
SUO, Standard Upper Level Ontology, http://suo.ieee.org, http://ontology.teknowledge.com
SWIFT, http://www.swift.com
tpaML, Trading Partners Agreement Mark-up Language, http://xml.coverpages.org/tpa.html
UDDI, Universal Description, Discovery and Integration, http://www.uddi.org
UDEF, Universal Data Element Framework, http://www.udef.com
UN/EDIFACT, United Nations Electronic Data Interchange for the Administration, Commerce and Transport, http://www.unece.org/trade/untdid/welcom1.htm
UN/SPSC, United Nations Standard Products and Services Code, http://www.un-spsc.net
UNSPSC, Universal Standard Products and Services Classification, http://www.unspsc.org
WPDL, Workflow Process Definition Language, http://www.wfmc.org
WSDL, Web Services Description language, http://www.w3.org/TR/wsdl
WSFL, Web Services Flow language,
http://www-4.ibm.com/software/solutions/webservices/pdf/WSFL.pdf
X.12, EDI standard, http://www.x12.org
XCBL, XML Common Business Library, http://www.commerceone.com/xml/cbl/index.html