Trading digital information goods based on semantic technologies

Wolfgang Maass¹, Wernher Behrendt² and Aldo Gangemi³

¹ Furtwangen University, wolfgang.maass@hs-furtwangen.de
² Salzburg Research Forschungsgesellschaft, wernher.behrendt@salzburgresearch.at
³ Institute of Cognitive Sciences and Technology (CNR), Rome aldo.gangemi@istc.cnr.it

Received 19 December 2006; received in revised form 23 September 2007; accepted 15 October 2007

Abstract

Digital information goods constitute a growing class of economic goods. During decision making for a purchase a buyer searches for information about digital information goods, such as information about the content, price and trading information, usage information, how it can be presented, and which legal restrictions apply. We present a logical container model for knowledge-intensive digital information goods (knowledge content object - KCO) that directly references formalised semantic descriptions of key information types on information goods. Key information types are formalised as plug-in slots (facets). Facets can be instantiated by semantic descriptions that are linked with domain ontologies. We have identified six logically congruent facet types by which a user can interpret information goods. KCOs are mediated and managed by a technical middleware, called Knowledge Content Carrier Architecture - KCCA. Based on the technical and logical structure of a KCO we will discuss five economic implications that drive further research.

Key words: information goods, electronic markets, semantic technologies, distributed architecture
1 Introduction

One of the assumptions of the Semantic Web is that structured meta-data about information resources provides a better means for human actors and software agents to access, manipulate, delete and create new information resources via digital networks [10]. Digital information environments, such as Intranet and Internet-applications, have traditionally been based on an underlying network metaphor that is driven by intrinsic features of free digital contents where information is perceived as a huge reservoir that can be mixed and used independently of economic interests.

By contrast, the concept of an electronic market for commercial digital information goods is based on an object metaphor, which places a product at the centre that can be appropriated on the seller’s side while buyers want to assess the product’s quality based on product information during purchase decision making [58], [13]. Digital information goods suffer from poor interoperability because the components that make up the digital information good often stem from different sources (i.e. applications) which have widely differing underlying assumptions about describing the content, its usage and what they regard as meta-data. This leads to a plethora of heterogeneous languages and semantics for description and subsequently, to diverging interpretations of the meta-data.

As a consequence, electronic markets require that digital information goods (1) interact with services in electronic markets by defined interfaces and (2) carry directly accessible product information, i.e. information that describes content usage scenarios according to various attributes. These requirements are at odds with the network metaphor of the World Wide Web and even the Semantic Web is not sufficiently helpful yet for modeling economically viable applications for commercial digital information goods [13].

The remainder of this paper is organised as follows: First we will discuss the requirements from an economic viewpoint (section 2). Next we describe a core ontology for electronic markets (section 3) and digital information goods (section 4) that are based on foundational ontologies. Based on this, we introduce in section 5 a generic representation framework for information goods called knowledge content object (KCO) which can be deployed in open, loosely coupled digital information infrastructures, such as the World Wide Web. In section 6, the overall information exchange architecture - which has been used within three application cases - is described. Section 7 describes how the KCO model and its exchange infrastructure may impact on the future of digital information goods and section 8 summarises and concludes the paper.

2 Economic view on information goods

Network-based digital content environments can be perceived as huge information markets where supply and demand meet. If content has sufficiently high value for actors representing the demand side, it will generate market prices. This kind of content is generally termed “paid content” as a special form of information goods and is viewed as a digital product [13]. Shapiro and Varian [58] define the term information good very broadly. "Essentially, anything that can be digitised - encoded as a stream of bits - is information. [...] Baseball scores, books, databases, magazines, movies, music, stock quotes, and Web pages are all information goods". Based on the definition of [13] anything one can send and receive over the Internet has the potential to be a digital product. "Information is a primary example of a digital product, for example knowledge-based goods that can be digitised and transferred over a digital network".

Research on the economics of information distinguishes between search products and experience products [11]. Search products are goods or services for which the most essential attributes can easily be evaluated prior to a purchase and provide a basis for an informed buying decision because consumers can verify claims before purchase [23]. Society is very much accustomed to buying search products such as cars, houses and computers. Experience products are goods or services for which the cost to evaluate the most essential attributes is so high that direct experience is often the evaluation method with the lowest costs in terms of time, money, cognitive effort, or other resources [23]. Because of the difficulty involved in evaluating claims for experience products, consumers will be more sceptical of claims for experience products in comparison with search products.

Information goods are immaterial goods which require carriers for implementation. Information goods are restricted three constraints: (1) creation constraints, (2) access constraints, and (3) usage constraints. Creation constraints mean that the origination of information is limited by an author’s capabilities, knowledge and expertise. Therefore the creation of information goods has typically linear scalability with a quantity and quality trade-off. Access constraints are used to limit access to an information good [31], [48]. Digital information goods are easy to copy and access [59], [36], so that the actual usage is limited by usage constraint technologies [21], [20], such as Digital Rights Management (DRM) systems [14], [25]. Access and usage limitations are artificially designed and limit the choice of runtime environments on which an information object can be used which, in turn, limits the potential market size [22].
2.1 Referential and self-referential information goods

Information goods can be either referential or self-referential. Referential information goods are representations of entities or situations in user-perceivable worlds. News, product descriptions, user manuals and discussions are representatives of this class. But information goods can also solely origin from the author’s mental conceptions so that resulting information goods refer to non-perceivable worlds. This class of information goods, such as poems and novels is called self-referential. Self-referential information goods are complete in the sense that they do not refer to user-perceivable entities or situations.

Table 1: Categorisation of digital information goods

<table>
<thead>
<tr>
<th>Type</th>
<th>Static</th>
<th>Dynamic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referential</td>
<td>Content centered</td>
<td>Content centered</td>
</tr>
<tr>
<td></td>
<td>product description</td>
<td>sensor-based information</td>
</tr>
<tr>
<td></td>
<td>report</td>
<td>live show</td>
</tr>
<tr>
<td></td>
<td>news</td>
<td>recorded show</td>
</tr>
<tr>
<td>Self-referential</td>
<td>music</td>
<td>genetic algorithm</td>
</tr>
<tr>
<td></td>
<td>book</td>
<td>intelligent agent</td>
</tr>
<tr>
<td></td>
<td>multimedia object</td>
<td>online computer game</td>
</tr>
<tr>
<td></td>
<td>web site / blog</td>
<td>chat service</td>
</tr>
</tbody>
</table>

Furthermore information goods are either static or dynamic. Static information goods do not maintain mechanisms that allow content modification. Instead, modifications are achieved by external applications. This case is the role model for database applications. Dynamic information goods maintain intrinsic mechanisms and logics that allow self-modification of contents. This gives information goods capabilities of intelligent agents [70], [28] and autopoietic systems [44] by obtaining internal and external behaviour such as self-modification, reproduction, termination and interaction with environments.

2.2 Anomalies

Digital information goods as offered on the internet provide a huge information market with offer and demand patterns. Information goods with sufficient value generate a market price [6]. Digital information goods exhibit three anomalies: (1) buying anomaly, i.e., information goods have trust and experience features, so they cannot be evaluated by consumers before buying otherwise he will not buy it anymore [57], [50], [63], (2) price anomaly, i.e., pricing of an digital information good cannot be determined by margin costs because they tend to be negligible [58], and (3) copy anomaly, i.e., copy and original of an information object cannot be distinguished.

In competing markets individuals know that their buying decisions are based on restricted information [58] which typically results in information asymmetries between offering and demanding actors. This, in turn, can influence the relationship between price, quality and demand and might lead to market failure [2]. In the following we will describe possibilities for enriching digital information goods with metadata that help consumers to reduce information asymmetries and increase consumer’s convictions to buy the right information good which supports the effectiveness of electronic markets for information goods. Buying anomalies can be eliminated by methods of two categories:

1. Signalling of secondary attributes [36], [61], [50], [16]
   - Quality ratings [58], [13]
   - Reputation [29], [5], [27]
   - Trust [2], [32]

2. Content projections:
   - Static content projections (abstracting, previewing, browsing) [66]
   - Dynamic adaptive content projections [39]

Signalling is a method to reduce information asymmetries in markets [60]. In relation to information goods signalling is achieved by offering of associative features and by variation of product and price presentation [58], [66]. Content projections are descriptions about information goods. Static content projections allow partial or time limited access on contents of information goods [58]. Examples for partial access are Ebsco’s summaries of scientific articles,
Amazon’s “search inside” excerpts, or movie trailers. Dynamic adaptive content projections process contents according to external requirements, for instance, user models [68], [45] or situational representations [1].

Stuart showed that content projections result in sustainable effects on pricing for high-valued goods if consumers have trust in their correctness [65] which can be achieved by provable valuations of independent trusted individuals [47]. Integration of content, metadata and ontological descriptions by digital representations are the basis for self-describing digital information goods [40], [8].

3 Ontological framework for trading information goods on electronic markets

Electronic markets are online locations that support information exchanges between agents. The basic ontology of digital media consists of five concepts, i.e. channel system, coding system, logical space, role system and protocol [53]. The channel system (C) provides connections by which agents can exchange messages. Messages are syntactically represented by a coding system (L) that must be learned by participating agents. The meaning is extracted by interpretation of the message against semantics (W) that is shared between market agents. The orchestration of an electronic market (and also digital media in general) is based on a role system (R) which defines the rules, rights, obligations and prohibitions related to a particular role while protocols (P) characterize dynamic behaviours and interactions between agents. Protocols are integrated sets of rules, rights, obligations, prohibitions and associated processes [53]. Schmid abstractly defines media as follows: medium = C + (L + W) + (R + P) [53], [54]. Media can be implemented by information and communication technologies that provide services towards the organisational level. These services are orchestrated according to applied protocols [52].

Complete implementations of electronic markets require services for four transaction phases: knowledge, signalling, contracting and execution [52]. In the knowledge phase, agents access and evaluate information, for instance, about product features, ratings, prices, and trading conditions. In the signalling phase a potential buyer signals his interest to buy a product while the seller signals his interest to sell it. If both sides signal trading interests they will enter the contracting phase. The contracting phase is governed by negotiation protocols and results in binding contracts. Contracts are implemented by the execution phase. Most of all this encompasses financial logistics and product logistics but also invocation of support, training and disposal services.

![Figure 1: Key concepts of the DOLCE ontology](image)

If an electronic market is implemented on a service-oriented architecture (SOA) it will deploy loosely coupled services for each phase that communicate via open technical protocols. This means that in general services are provided by more than one software vendor, i.e., exchanges between services need to be transparently coordinated. This generates a trade-off decision between data representations of information goods vs. protocol representations. In one extreme, all information that is required to trade information goods is implemented at the protocol level, i.e., the protocol carries and coordinates data exchanges between market services. The other extreme assumes light-weight services but extended data structures of information goods. In the latter case, relevant information is stored in the information goods so that services are coordinated by the information good itself. For instance, if an execution service needs information about contractual constraints it directly requests this information from the information good.
In the traditional case, the execution service would query contract monitoring services that store particular contract information for this information good.

Next, we will introduce a core ontology for digital media as a general ontological framework for electronic markets before we present a semantically expressive data structure for information goods. The core ontology for digital media is a semantic representation of the generic context that can be set up by digital media. Electronic markets are specialisations of digital media in which information goods are traded. The core ontology for digital media specialised for electronic markets provides explicit and machine processable terms of reference that can be used by market services.

3.1 Core ontological model based on foundational ontologies

Foundational ontologies such as DOLCE [43] make it possible to not only to describe distinctions between generic concepts but to root them in fundamental communicative acts. In other words, in on-line systems of the future, there is likely to be legal requirements for systems to be ontologically aware or else, their owners may be held responsible for their systems’ "lack of intelligence".

Basic DOLCE top-level includes the following categories and relations (Figure 1):

- **Endurants** (objects or substances) and **perdurants** (events, states, or processes) are distinct categories linked by the relation of participation (e.g., a group of people participate in an expedition).

- **Endurants** are localized in space, and get their temporal location from the **perdurants** they participate in. **Perdurants** are localized in time, and get their spatial location from the **endurants** participating in them.

- **Qualities** inhere in either **endurants** (as physical or abstract qualities) or in **events** (as temporal qualities), and they correspond to “individualized properties”, i.e. they inhere only in a specific entity, e.g. “the color of this red herring”, “the depth of the water at this point”, etc.

- Each kind of **quality** is associated to a **quality space** representing the space of the values that qualities can assume (e.g. a metric space).

- **Quality spaces**, as all **abstracts** (the fourth category), are neither in time nor in space.

DOLCE is extended towards the representation of non-physical objects, especially social and content objects [24]. In more detail, DnS is based on a fundamental distinction between **descriptions** (for instance, in the legal domain, legal descriptions, or conceptualizations, which encompass laws, norms, regulations, crime types, etc.) and **situations** (again, in the legal domain, legal facts or cases, which encompass legal states of affairs, non-legal states of affairs that are relevant to the Law, and purely juridical states of affairs). In fact, its very first formulation was a design pattern represented by means of a UML class diagram (see Figure 2).

Figure 2: Extension of DOLCE by the Description-and-Situation (DnS) ontology
• A description is a (non-agentive) social object which represents a conceptualization, hence it is generically dependent on some agent and communicable [33]. Examples of descriptions are regulations, plans, laws, diagnoses, projects, plots, techniques, etc. Like physical objects, social ones have a lifecycle, can have parts, etc. Unlike physical objects, social (like all non-physical) ones are generically dependent on some agentive physical object. Hence, a description generically depends on some agent which is (at some time) able to conceive it. Agent is introduced here as a primitive (subclass of endurant).

• A situation is a non-agentive social object which represents a state of affairs or relationship, or tuple, or fact, under the assumption that its components ‘carve up’ a view (a setting) on the domain of an ontology by virtue of a description. A situation aims at representing the referent of a “cognitive disposition” towards a world, thus reflecting the willingness, expectation, desire, belief, etc. to carve up that world in a certain way. Consequently, a situation has to satisfy a description.

\[
\text{Situation}(x) = \text{df} \quad \text{NonAgentiveSocialObject}(x) \land (\exists y. \text{Description}(y) \land \text{Satisfies}(x,y)) \land (\exists z. \text{Particular}(z) \land \neg \text{Situation}(z) \land \text{Setting}(z,x))
\]

\[
\text{Situation}(x) \rightarrow \forall y. \text{Part}(x,y) \rightarrow \text{Situation}(y)
\]

The setting relation holds between situations and particulars from the ground ontology. At least a perdurant must exist in the situation setting:

\[
\text{SettingFor}(x,y) \rightarrow \text{Situation}(x) \land \text{Particular}(y) \land \neg \text{Situation}(y)
\]

\[
\text{SettingFor}(x,y) \rightarrow \exists z. \text{Perdurant}(z) \land \text{SettingFor}(x,z)
\]

\[
\text{Setting}(x,y) = \text{df} \quad \text{SettingFor}(x,y)
\]

The satisfies relation holds between situations and descriptions, and implies that at least some concept in a description must classify at least some particular in the situation setting:

\[
\text{Satisfies}(x,y) \rightarrow \text{Situation}(x) \land \text{Description}(y)
\]

\[
\text{Satisfies}(x,y) \rightarrow \exists z. \text{Concept}(z) \land \text{Uses}(y,z) \land \exists w.t. \text{SettingFor}(x,w) \land \text{Classifies}(z,w,t)
\]

• A plan is a description that is conceived by a cognitive agent, defines or uses at least one task (a kind of course of actions) and one role (played by agents), and has at least one goal as a proper part. Examples of plans include: the way to prepare an espresso in the next five minutes, a company’s business plan, a military air campaign, a car maintenance routine, a plan to start a relationship, etc.

• Plan executions are situations that proactively satisfy a plan, meaning that the plan anticipates its execution:

\[
\text{PlanExecution}(x) = \text{df} \quad \text{Situation}(x) \land \exists y. \text{Plan}(y) \land \text{Satisfies}(x,y) \land \exists t. \text{PresentAt}(y,t) \land \neg \text{PresentAt}(x,t)
\]

• Tasks are courses that are (mostly) used to sequence activities, or other perdurants that can be under the control of a planner. They are defined by a plan, but can be used by other kinds of descriptions.

The previous distinctions are supported by a large axiomatisation (http://dolce.semanticweb.org). In the next section, we introduce a conceptual model of electronic markets based on the foundational ontology DOLCE. Therefore market concepts are aligned with ontological concepts of DOLCE.

3.2 Towards a core ontology of economic markets

For software agents or services to interact reliably there is a need for well defined operational semantics particularly in those areas where the physical and the virtual world meet. For instance, while it is acceptable for a machine to make poor recommendations on what book to buy (as long as the human user makes the buying decision, ultimately) it will not be acceptable for a machine to spend a significant amount of its owner’s money to buy useless goods due to a semantic “misunderstanding”. It is likely that a mix of legal, organisational and technological provisions will be required to safeguard the operation of autonomous software services in the future. We suggest that one way to safeguard such operations at the technological level is the use of well-founded, explicitly described, and formally bounded ontologies.

As already introduced, an electronic market is one of the dominant metaphors for an economic transaction which is characterised by a sequence of four basic performative communication acts: (1) informing, (2) signalling, (3) contracting and (4) executing [33], [34]. These performative communication acts are operationalised by electronic services offered by the electronic market application environment [54]. Informing acts require information about goods that describe characteristics of the good in question and which match the buyer’s preference set [49], his/her level of expertise [9], as well as on the net value of the benefits and costs of both the good and the processes of
obtaining it [30]. For signalling their willingness to negotiate sellers and buyers use domain specific signalling acts which are sometimes the first step towards a contracting act, e.g., placing a good into a shopping cart. Signalling can be governed by specific business protocols, e.g., auctions for works of art. Contracting acts specify binding and enforceable procedures [34], [55] for subsequent exchanges of goods that are part of execution acts that control digital and physical logistic procedures [3]. Performative communication acts of electronic markets are modelled as plans that can be satisfied by situations (see Figure 3).

The core media ontology consists of nine concepts that are derived from the media model [52], [54]:

1. Market description: following DOLCE and its extension DnS [24], a market description is a specialisation of a description and represents a conceptualisation of an economic market. Core concepts of a market description are market role, market plan, market tasks and associated parameters. It is required for a market description that role-taking agents conceive a market description, i.e. they have knowledge about the organisation, processes and entities that are part of market settings. Market descriptions are satisfied by market situations and its components market endurants, market activities and regions.

2. Market role: a role is social in the sense that it is intentionally created and agreed by a community [58]. The distinction between roles and agents helps to represent that one agent might take different roles. Role-taking agents are restricted in their rights, obligations and prohibitions according to situations. In DOLCE, the concept role is conceptualised as a social-object. For electronic markets, we use the more specialised concept commerce role.

3. Market plan: a market plan is a description that is conceived by agents participating in markets. A market plan defines or uses at least one market task and has at least one market-oriented goal as a proper part (for details cf. [24]).

4. Parameters: a parameter is a concept that classifies (in particular, it is ‘valued by’) regions, as defined by some description. Parameters are the descriptive counterpart of regions, and, as regions represent the qualities of perdurants or endurants, they can be requisites for some role or course. A parameter has at least one region that is a value for it. For instance, a reward value is valued by money measure that is a region of type measurement-unit.

5. Market task: market tasks are defined by plans. They are courses that are used to sequence market activities. DDPO provides simple logical operators to express relationships between market tasks, such as
successor relation and control tasks based on a reification of control structures. Due to their logical complexity control structures on market tasks are typically implemented at the application level. For instance, a control structure is required to model that either a negotiation task is repeated until a mutually agreeable contract can be found or the negotiation was terminated without contract.

6. Market endurants: market endurants are physical or non-physical endurants. Endurants are defined as particulars in space that participate in at least one perdurant. A perdurant, in turn, are particulars in time which have at least one participant. Examples for market endurants are market role-taking agents, realised contracts, or money that can be either physical, i.e., in space or virtual, i.e., realised by, for instance, digital representations.

7. Market activity: a market activity is an action in a market that is generically constantly dependent on a shared market plan adopted by participants. This implies that a market action must be sequenced by a market task.

8. Region: a region defines attribute-value relationships. For instance, MoneyMeasure is related to a quality region of float numbers.

9. Market situation: situations describe circumstances. According to Masolo et al. [43], situations are social objects constituted by entities of a circumstance and their relations that are defined by descriptions. Situations highlight entities and relations of a domain that satisfy descriptions, i.e. expected conceptualisations of circumstances. At least one role-taking agent participates in a situation. Situations can be settings for courses, such as needed for describing different trading phases of electronic markets. Situations are also used to characterise communication situations, such as needed in the information phase of electronic markets.

Next we describe how information goods can also be represented by formal conceptualisation. This is later merged with the core ontology of electronic markets.

4 An ontological framework for describing digital information goods

Having described the core ontology for electronic markets, we now discuss how information goods can be ontologically embedded. A central aspect of this discussion is the distinction between information as a set of mental entities conceived by humans, information realised by artefacts (content objects), and information used as descriptors of other information.

4.1 An Ontology of Information Objects

A specific usage context of a content object may require us to talk about the digital reproduction of a painting that is owned by an institution, and such institution is willing to commercialize the reproduction at certain conditions that include differentiation for users, pricing, regulations to be followed, inclusion of content metadata, explanations,
interpretations, ways of rendering it, etc. This context is complex, and requires a subtle differentiation of the various entity types involved in it.

According to DDPO (an extended foundational ontology encompassing DOLCE, descriptions, situations and plans [24]), a content (information) transferred in any modality is a kind of social object called information object (IO). Information objects are spatio-temporal reifications of pure (abstract) information as described e.g. in Shannon's communication theory, hence they are assumed to be in time, and realized (materialized) by some entity. Information objects are the core notion of a semiotic ontology design pattern, which employs typical semiotic relations [24].

We present the axiomatization of KCOs in OWL Abstract Syntax. We firstly present the definition of DnS:information-object, which encodes the basic axioms of an ontology of semiotics extending the basic DOLCE ontology (see Figure 4):

\[
\text{Information-object}(x) = \text{df} \text{social-object}(x) \land (\forall y. \text{particular}(y) \land \text{about}(x, y)) \land (\exists z. \text{information-realization}(z) \land \text{realized-by}(x, z)) \land (\forall k. \text{agent}(k) \land \text{interpreted-by}(x, k)) \land (\forall i. \text{description}(i) \land \text{expresses}(x, i)) \land (\exists j. \text{information-encodings-system}(j) \land \text{ordered-by}(x, j))
\]

The definition says that information objects are necessarily encoded by some information encoding system, must be realized by some particular, can express a description, and, if that description is satisfied by a situation, can be about that situation, or some entity in its setting and can be interpreted by agents that can conceive of the description expressed by said IOs.

For example, Jack Kerouac’s novel "On The Road" is an information object, is ordered by modern American English language (the information encoding system), is realized by, e.g., a digital copy in PDF format, expresses a certain plot on the Beat Generation and its related meaning, is interpreted by an agent in the role of a reader with average knowledge on American sociology, and it is about certain entities and facts (see Figure 5).

These semiotic relations constitute a typical ontology design pattern, so that any composition of relations can be built starting from any node in the pattern or in an application of the pattern.
5 Semantic Modelling of Knowledge Content Objects

The information object design pattern describes the content and context of an information object on a concept and instance level, i.e. information objects are mental conceptions while information object realisations are represented by physical entities, such as a book, a CD or a digital file. Information object realisations and in particular content objects are increasingly annotated with meta-data, such as Dublin Core, NewsML, MPEG or Adobe’s XMP. Meta-data can be inspected by applications and by web services that are enabled for a particular meta-data schema. This is used in applications such as data warehouses, catalogue integration and information integration [18]. In these applications matching of data and data schemas are core functionalities that require meta-data and embedding into ontologies. Automatic data processing in heterogeneous environments depends on the degree of formalisation of the meta-data and the ontologies. In the following, we introduce an interchange format for content objects, called Knowledge Content Object (KCO), that on one hand provides an ontology based container structure as required by service-oriented business applications and on the other hand allows flexibility so that contents from different sources can be integrated. Next we discuss the general logical structure of a Knowledge Content Object with a special focus on the integration of information about the electronic market environment. We conclude the section with a specific example.

5.1 Logical Structure

A Knowledge Content Object (KCO) is a specialisation of a content object (cf. Figure 6). In electronic markets it realises a design object that in turn can take the market role of a product. The purpose of a KCO is to hold a maximum of data or information that can be used for automatic processing in web service infrastructures.

The logical structure of the KCO builds on requirements of performative communication acts used in electronic markets, on previous approaches to multimedia and hypermedia document models [12], [71], [66] and on an analysis of several hundred existing paid content business models [62]. The KCO format encompasses five key information types, called facets, that support information needs concerning digital information goods during different phases of their life cycle [38], [8]. Several of these facets are subdivided into further sub-facets which support a better logical clustering of meta-data. At the lowest level, it is intended that each of the leaf elements is associated with well-defined operational semantics by formalised ontologies, in order to enable organisations to quickly deploy KCOs as part of their information infrastructure.

The facet structure is derived by the following requirement set for tradable content objects:
- Content description: a content object shall carry information about its content which enables automatic abstracting, content syndication, semantic search and other kind of content-dependent services.

- Community description: tradable content objects can take the role of products. Therefore a content related product shall be aware of the other roles and entities that are present and required in particular situations.

- Business description: content objects are traded in business contexts, such as electronic markets. Content objects shall represent appropriate negotiation schemes and shall carry contract information at their instance level.

- Presentation description: contents that are carried by content objects generally differ in their content projections according to output devices and user needs. Therefore KCOs shall carry specifications of how its content is presented.

- Trust and security description: the business facet particularly requires access restrictions. Therefore a KCO shall carry information about authorisation and authentication issues. Only authorised actors might get access to, for instance, business descriptions.

The whole facetted KCO structure is an interchange format that can be instantiated by different meta-data formats and ontologies. We now describe all facets in more detail followed by an example based on a formal ontological representation with DOLCE.

Table 2: KCO facet structure

<table>
<thead>
<tr>
<th>Facets</th>
<th>Elements</th>
<th>Short Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Description</td>
<td>Propositional</td>
<td>Central information about the content itself is formally described in propositional formats. This facet might be instantiated by descriptions, such as the NewsML format. The most sophisticated propositional descriptions will be based on a fully ontology-based knowledge representation language.</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>Content Classification</td>
<td>Keywords</td>
<td>Keywords and concepts assigned to the content object based on a classification schema. The IPTC thematic thesaurus (<a href="http://www.iptc.org/NewsCodes/">http://www.iptc.org/NewsCodes/</a>) and the ICON Class classification system (<a href="http://www.iconclass.nl/">http://www.iconclass.nl/</a>) are other examples of controlled vocabularies.</td>
</tr>
<tr>
<td></td>
<td>Classification</td>
<td></td>
</tr>
<tr>
<td>Multimedia</td>
<td>Characterization</td>
<td>This description includes information about the content format, such as encoding, storage, and location. This sub facet can be instantiated by, for instance, MPEG7 descriptions.</td>
</tr>
<tr>
<td>Business Description</td>
<td>Negotiation</td>
<td>Business descriptions define requirements on trading situations of KCOs. Negotiation protocols describe organisational and process requirements on business interactions, such as in electronic markets. Pricing schemes are internal representations of pricing strategies of KCOs that cannot be accessed by potential buyers. Finalised contracts are formalised by contract descriptions, such as IPROnto [17].</td>
</tr>
<tr>
<td></td>
<td>protocol</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pricing scheme</td>
<td>A plan that describes how a price is determined. A pricing model is a key element of an information object’s business model.</td>
</tr>
<tr>
<td></td>
<td>Contract</td>
<td>A contract is a digital representation of mutually agreed constraints that govern the use of an information object. Contract can have informal, semi-formal or formal representations.</td>
</tr>
<tr>
<td>Community Description</td>
<td>Organisation</td>
<td>Description of the organisation (roles, rights, obligations, prohibitions) and processes (plans, tasks and workflows). Both provide a context model in which KCOs can be used.</td>
</tr>
<tr>
<td></td>
<td>Processes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Usage history</td>
<td>List of activities performed with the KCO during its lifecycle.</td>
</tr>
<tr>
<td>Presentation Description</td>
<td>Spatio-temporal</td>
<td>Descriptions of how the content of a KCO is presented to users. Presentation includes the rendering, rendition as well as interaction models. An example is descriptions in SMIL format (<a href="http://www.w3.org/TR/2005/REC-SMIL2-20050107/">http://www.w3.org/TR/2005/REC-SMIL2-20050107/</a>).</td>
</tr>
<tr>
<td></td>
<td>rendition</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interaction-based</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rendition</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Services</td>
<td>Service descriptions are specifications on computational services that are required by KCOs in usage situations. Examples for semantic service descriptions are WSMO or OWL-S.</td>
</tr>
<tr>
<td>Trust &amp; Security</td>
<td>none</td>
<td>Enforcement of authorisation and evaluation of certifications are processed on application level.</td>
</tr>
<tr>
<td>Self-description</td>
<td>none</td>
<td>Specification of the (inner) structure of the KCO (e.g., active facets, ontologies used) in machine-interpretable form.</td>
</tr>
</tbody>
</table>
The **content description facet** contains three sub-facets: propositional description, content classification, and multimedia characterization (Table 2). The propositional description contains a formalised model of a content and it either includes the content or it references the content. Various semantic content description formats have been proposed, such as Situational Calculus [46], Rhetorical Structure Theory [42], Conceptual Dependency Theory [51], Conceptual Graphs [59], Story Telling Theory [19], or DOLCE D&S [24]. In the following we will use DOLCE D&S because it is grounded on a foundational ontology and can be used directly on web-based application environments.

The **business description** contains a description of the business semantics associated with the KCO. This comprises the facet element negotiation protocol which describes the business scripts by which a contract is being negotiated. A negotiation protocol is described as a DDPO plan [24]. The pricing scheme is used in restricting the price policies that can be applied during the negotiation. The pricing scheme is grounded in DDPO as a regulation concept. In the simple case of a fixed-price scheme, the negotiation is reduced to a simple over-the-counter (OTC) purchase [69]. The pricing scheme is required for price differentiation strategies that are defined by the seller on the basis of a differentiating factor such as age, quantity discount or date of content origin (see [67],[62]). Again, the resulting contract is represented by a DDPO plan [24]. Alternative plan representation formats are simple license schemes such as Digital Rights Management formats, e.g., XRL, ODRL, or IPROnto. They can be used to describe the situation and the agentive roles that can be taken by agents which can then act by using defined tasks.

The **presentation description facet** is the specification of time-based spatial presentation of, and interaction with, complex content. Given some media tokens, we specify on one or more temporal "tracks" which describe when the associated media data will be rendered, and where they will be rendered (in terms of spatial arrangements). The second facet element deals with interaction and dialogue. Here, the semantic annotation specifies whether the presentation is entirely pre-programmed, whether it is entirely open (e.g. web based navigation) or whether it follows some dialogue pattern where humans and system take conversational turns in order to navigate the knowledge/information structure. This description defines one or more discourse structures that can be associated with the content for its rendering.

The **community description** describes the organizational context in which contents can be used. This covers four sub facets: organisation, that is formally described by reference to an ontology of roles (rights and obligations) that users would take in order to manipulate or consume the content; process that describes on abstract level how role-taking actors are allowed to interact with one another. In electronic markets the organisation is described by market roles and market endurants that are defined to play the market roles. The process is described by a market plan and its markets tasks which sequence a set of market activities. The market plan and market tasks define the protocol of an electronic market while the market activities actually implement the protocol. The service sub-facet contains service descriptions and their relationships to market tasks. A possible representation candidate is the WSML language [15]. Finally, the usage history, keeps traces of previous use in order to support workflow systems as well as collaborative filtering systems. The latter can be achieved by keeping track of user data when the KCO is being "touched" by that user.

Finally, the Trust and Security facet was not developed in detail in the METOKIS project, but the intention is to regard this facet as an interface which ensures the rights and needs of the potential customer (trust) as well as those of the potential seller (security). For example, eBay's information to customers about vendors and their quality of service is an important trust-inducing factor which probably influences many purchase decisions. However, many more trust inducing schemes are imaginable, but there is a lack of conceptual machinery to even describe such schemes and their metrics. For example, each purchase which went well and for which the purchasing agent was given positive feedback by its user, should percolate back into the system, as a small measure of "customer satisfaction" without necessarily being traced back to the individual.

### 5.2 A KCO example

Imagine that that a person "Hans Meyer" owns an electronic copy of Jack Kerouac's novel "On the Road" and he intends to sell it to another person "Karl Schneider". This snap shot of a transaction will be discussed in the following with a special focus on the business description facet (see Figure 7).

This scenario requires all five roles applicable to business transactions: seller, buyer, product, commitment and reward. Seller and buyer are played by Hans and Karl. For clarification purposes, we have omitted instances in Figure 7. The product is an information object that depends on the author, Jack Kerouac. This novel is a design object that is realized by a KCO owned by Hans. Buyer and Seller conceive of a market plan, i.e. the general pattern of how to trade a book which is described by a selling plan. This plan is set up by a market situation in which seller and buyer participate. A selling plan describes a buying activity that eventually helps to generate a contract which, in turn, plays the role of a commitment. Beside other information, a contract is part of the KCO's container model. The rest of this model, i.e. plans, situations, activities, and commerce roles, is part of the KCO's community description.

Facet information of a KCO directly support transaction oriented queries. For instance, queries on ownership, contract status, currently active selling situations, and defined commitments. Changes of values in a KCO's data
model are either implicitly made by runtime applications or explicitly by application independent procedure systems, such as rule based models, e.g., based on SWRL or Jena rules (for an example cf. [37]).

![Diagram of KCO classes and relationships]

6 Electronic markets based on KCCA enabled infrastructures

In order to make use of the semantic richness that can be expressed with KCOs we need an infrastructure whose components support the functionality afforded by the KCO. The Knowledge Content Carrier Architecture (KCCA) does this in the shape of services which are logically clustered by KCCA's components arranged in a 3-tier architecture. This gives rise to the following structural core components: (1) KCO Service API - offering the functions described by the facets in table 1, (2) KCCA Registry and Manager - managing a federation of KCO-aware nodes, (3) KCTP Service - a protocol to exchange service requests across KCCA nodes and (4) KCCA Profiles - Services for the wrapping and integration of external data sources (see Figure 8).

One of the assumptions of our work is that eventually, most information systems will make use of two further components: firstly, reasoning services based on ontologies and secondly, a task execution environment that will support the definition and execution of flexible workflows. KCOs are designed to support such an architecture through content descriptions (this is where reasoning services can access the KCO), community description (describing the tasks for which this KCO is useful and the roles of actors that would do the tasks) and business descriptions.

We envisage future publishing environments to use an integrated framework consisting of the components described. This will leave the application builder to focus on application and domain specific adaptations, and on the tailoring of the presentation /interaction layer to the needs of the customer.

The following architectural overview shows the full picture combining KCCA components, reasoning and task execution environment, as well as domain specific adaptations and the application layer.

The KCO services offer access to the operational semantics of the KCO facets. The KCCA Registry and Manager component keeps track of how a federation of KCO aware information systems is set up. The KCCA environment keeps information about information sources, wrappers and maintains state in user sessions that may span requests and transactions across the federation. The KCTP Services define a stateful protocol that allows communication between KCCA nodes by exchanging serialised RDF graphs. The KCCA Integration Services give assistance in binding non-KCCA resources to a KCO aware system. This is done by a two-stage mapping process. The external information source is first mapped into an equivalent RDF schema which we call "context profile". This can be a
"naive" mapping to RDF. Next, a view is defined over the context profile and this view is made KCO compliant ("view profile"). The provider of an external information source needs to write a wrapper which provides the context profile for the resource. The KCCA integrator uses the context profile to create the view profile.

The design and implementation of KCCA coincided with the development of the WSMO Semantic Web Services approach [15] which was pioneered in several large European research projects. While we were aware of WSMO there was no implementation of the semantic web services framework available in 2004 and 2005. Therefore, a lightweight implementation using available formalisms (FIPA) and frameworks (SOAP) was chosen to implement a basic KCCA infrastructure. However, in an Austrian national project [http://grisino.salzburgresearch.at], we are currently re-designing and implementing KCCA on the basis of the WSMO framework. One of the challenges of combining semantic web services with knowledge content objects lies in understanding where to draw the line between static modeling of parameters (in the KCO) and the description of processes (web services). There is some debate at present that (semantic) web services may require so much descriptive modelling that the description is equal in effort to straightforward implementation, but to our knowledge, there is at present no study which could provide supporting evidence for such a claim. Our view is that semantically rich data structures (KCO) allow the use of simpler processes (services) whereas poor data structures require more intelligence in the services, hence our hypothesis that the two combined should lead to a balanced infrastructure.

7 Implications for the use of digital products

From an economic point of view five hypotheses can be derived from the KCO carrier model for digital information goods. These hypotheses are discussed in turn.

7.1 Reduction of transaction costs

KCO-contained digital information goods are designed to be shared and traded by electronic markets because they can leverage the standardised format of KCOs. Any kind of digital product is wrapped into a homogeneous semantically enhanced format that can be carried by any KCO-supporting infrastructure. Facet information can be perceived as semantic content plug-ins delivered by providers. Rich facet information can be used by consumers in electronic markets to reduce their information costs, i.e. information about prices and product characteristics [7], [41].
7.2 Value increase

As a consequence of the first hypothesis, users will be able to evaluate the value of digital information goods for lower cost, which is likely to speed up diffusion and adoption processes [62]. In general it can be assumed that high-quality digital information goods will extend their reach and life time, i.e. increase of revenue.

7.3 Increased competition

If semantically annotated digital information goods reduce transaction costs and increase their reach then the size of markets might increase as well. Because KCO-contained digital information goods provide a modular but sufficiently complete structure, consumers gain improved means for product comparison. Obviously, providers will reduce this effect by differentiation of provided semantic information. Commoditised digital information goods, such as stock or weather information, are likely to become the target of intensified competition [7], [41], which results in additional consumer economic surplus.

7.4 Reduction of operational costs

If hypotheses 1 to 3 hold, then it can be assumed that consumers might be more willing to use electronic markets as knowledge channels, i.e., they outsource knowledge competencies to external providers. Organising and archiving knowledge coded in information sources is handed over to external services if it is more effective and efficient than internal provisioning [4]. Container models, such as KCO, enable this outsourcing process for complex digital information goods. This process starts with simple information goods such as bookmarks (e.g. del.icio.us or shadows.com), pictures (flickr.com, bubbleshare.com), videos (videoegg.com and jumpcut.com), feeds (feedblender.com), wikis, blogs and email conversations (9cays.com) and may well extend to complex information goods such as software repositories (sourceforge.net, collab.net).

7.5 Influence on consumer’s decision making

Finally, recent results on consumer behaviour indicate that information given by electronic recommender systems positively influence consumers’ buying decision and furthermore reduce the number of alternative goods considered [26]. In particular facets 1 to 5 contain information which is likely to be used by consumers for making their buying decision for physical and digital goods in general. However, empirical tests will be required to evaluate this hypothesis.

Sellers and producers of digital information goods can leverage semantic enhancements by positive signalling effects. Producers of high quality products are interested in investing into information that signals higher quality while those with lower quality are tempted to avoid this additional effort. Hence, semantic enhancement might support a separation by signalling. Additionally sellers and intermediaries can use semantic descriptions as the basis for added services that adapt product characteristics and prices to buyers’ needs. This enables innovative product and pricing strategies.

8 Summary and open issues

We have discussed how digital information goods can be semantically described by a KCO container model. Straight forward approaches such as Adobe’s XMP follow similar goals. It is most likely that several other architectures will appear before a solution is accepted by a large audience. The KCO/KCCA approach paves the way for ontologically grounded models and explores application areas and technological solutions in the realm of the semantic web.

In the future we will investigate how KCO-embedded digital information goods can be used to generate new kinds of information goods. This can be either done manually as part of an editorial process or automatically based on plans and situations. For instance, new information goods can be aggregated based on user preferences, willingness-to-pay or user locations to name a few. How and by which means KCO facet information can be processed is also part of future research.

At present, the realisations of KCOs are still immature because the infrastructure which can make full (semantic) use of KCO features is still being developed. We predict the following possible scenarios: if KCOs or similar knowledge and content structures are adopted, there will be a paradigmatic shift in the way content and knowledge applications will be built, because more effort will go into tailoring the KCO ontology to the needs of the application domain and less effort will have to be spent on developing complex and difficult-to-maintain applications. The alternative scenario (which seems more likely at present) is that no paradigmatic shift will occur and that traditional content models will slowly get more “sematicised”, but partly by way of extensive programming. It may take several years (as it did for relational data bases) before proper KCO based systems with attendant reasoning facilities will appear.
Acknowledgments

This article is based on results of the IST-FP6 project METOKIS (http://metokis.salzburgresearch.at/) that has been co-funded by the European Commission.

References


