

# Identity of Resources and Entities on the Web

Valentina Presutti and Aldo Gangemi

Laboratory for Applied Ontology  
ISTC National Research Council (CNR)  
Rome, Italy

**Abstract.** One of the main strengths of the web is that it allows any party of its global community to share information with any other party. This goal has been achieved by making use of a unique and uniform mechanism of identification, the URI (Universal Resource Identifiers). Although URIs succeed when used for retrieving resources on the web, their suitability as a way for identifying any kind of *things*, for example resources that are not on the web, is not guaranteed.

In this article we investigate the meaning of *identity* of a web resource, and how the current situation as well as existing and possible future improvements can be modeled and implemented on the web.

In particular, we propose an ontology, IRE, which provides a formal way to model both the problem and the solution spaces. IRE describes the concept of resource from the web viewpoint, by reusing an ontology of Information Objects, built on top of DOLCE+ and its extensions. In particular, we formalize the concept of *web resource*, as distinguished from the concept of a generic *entity*, and how those and other concepts are related e.g. by different *proxy for* relations. Based on the analysis formalized in IRE, we propose a formal pattern for modeling and comparing different solutions to the identity problem.

## 1 Introduction

The web is an information space realized by computationally accessible resources, each embedding some information, which is encoded in some language, and expresses some meaning. One of the successful achievements of the web is allowing different parties of its global communities to share information [21]. Typically, typing an address in a web browser is enough in order to visualize or download an object, the meaning of which can be then understood by a human agent. The web address is a URI (Universal Resource Identifier) [4]. The URI mechanism is key to the web success. However, another ambitious goal of the web is that of referencing *things* in general. For example, consider the World Wide Web Consortium (W3C) [10]: it should be possible to distinguish (on the web) the reference to the organization from that to its web site.

The simple association of a URI to a thing or real world entity is very powerful. On one hand, it has already demonstrated its effectiveness with regard to the identification of objects that are accessible through the web, e.g., web pages. On the other hand, there is no complete consensus on how to manage identification of things that are not on the web. Reducing the ambiguity of identifying the entities a web resource refers to is essential for information sharing, interoperability, and reasoning on the web [24]. In order to propose solutions to this issue, it is crucial to analyze and properly describe the problem space.

The problem space can be expressed in terms of the impact that identification of (generalized) resources has on the web. In this paper we analyze the state of art related to this problem, and from this analysis we make it emerge five distinct issues. We propose that, in order to describe these issues and to compare the respective solutions, we need to analyze the reason why a URI can be associated with an entity. We carry out such analysis based on an ontology called *Identity of Resources and Entities on the web (IRE)*.

IRE focuses on four main classes: *URI*, *web resource*, *information object*, and *entity*, which encompass the things in the domain of discourse of the web referencing problem.

Once the problem domain has been analyzed, the solution domain can be approached. We discuss how the current evolution of web science from web to web 2.0 and semantic web has affected the solution domain. We also consider some proposed and envisaged solutions, and discuss them in terms of IRE.

The rest of the paper is organized as follows: section 2 tells a story about the existing literature on the problem of identifying a web resource. Section 3 discusses how the problem of resource identification impacts on the web. Section 4 informally presents the IRE ontology. Section 5 deals with the solution space, and section 5.1 presents an extension of IRE in order to represent it., Section 6 summarizes the main arguments presented. Finally, the appendix (7) contains a first-order logic formalization of IRE. The OWL version of IRE can be downloaded from <http://wiki.loa-cnr.it/index.php/LoaWiki:IRE>.

## 2 Some history

The identification of resources is an important task to use them on the web [25]. Currently, there is a diffuse feeling that resource identification procedures suffer from a lack of consensus about how to handle them. This lack of consensus partially finds its root from normative documents where the concept “resource” has been defined in the context of the web. However there are also other motivations underlying the identification problem, which we discuss in this article.

The term “resource” is generally used for all things that might be identified by a URI [21]. In the literature, we find several definitions for the term “resource” used in the context of world wide web. In particular we quote here three normative documents, [3, 4, 21]<sup>1</sup> and discuss about the way and consequences of the definition they provide for “resource”. In [3] the concept of resource is defined as follows:

A resource can be anything that has identity. Familiar examples include an electronic document, an image, a service (e.g., “today’s weather report for Los Angeles”), and a collection of other resources. Not all resources are network “retrievable”; e.g., human beings, corporations, and bound books in a library can also be considered resources. The resource is the conceptual mapping to an entity or set of entities, not necessarily the entity which corresponds to that mapping at any particular instance in time. Thus, a resource can remain constant even when its content—the entities to which it currently corresponds—changes over time, provided that the conceptual mapping is not changed in the process.

The following definition of “resource” is given by [4], which updates [3]:

This specification does not limit the scope of what might be a resource; rather, the term “resource” is used in a general sense for whatever might be identified by a URI. Familiar examples include an electronic document, an image, a source of information with a consistent purpose (e.g., “today’s weather report for Los Angeles”), a service (e.g., an HTTP-to-SMS gateway), and a collection of other resources. A resource is not necessarily accessible via the Internet; e.g., human beings, corporations, and bound books in a library can also be resources. Likewise, abstract concepts can be resources, such as the operators and operands of a mathematical equation, the types of a relationship (e.g., “parent” or “employee”), or numeric values (e.g., zero, one, and infinity).

In [21] the concept of resource is used with a twofold meaning: either whatever might be identified by a URI, or anything that can be the subject of a discourse, such as cars, people etc. Furthermore, the concept of *Information resource* is defined as a resource whose essential characteristics can be conveyed

---

<sup>1</sup> Note that [3] has been replaced by [4], however we decided to quote and discuss about it because it helps realizing the historical motivations why there is confusion on the sense of “resource” on the web

in a message. [21] also defines the principle of *opacity* of a URI, which promotes the independence between an identifier and the state of the identified resource.

Given that, at least four possible interpretations of the term “resource” can be singled out.

- **computational object:** if a resource is a computational object, e.g. an electronic document [4, 3] - in this context we define “computational object” such as (i) the physical realization of an information object, (ii) something that can participate in a computational process. Examples of computational objects are: a database, a digital document, a software application - then its identity would not be equivalent to a virtual localization, because a computational object is a physical entity and realizes (is the support for) a certain information object. Neither physical entities nor information objects can be reduced to regions in a virtual space, especially if that space should be uniquely identifiable through URIs. For example, the personal home page of Aldo Gangemi is a document which exists on the web and is reachable through the dereferencing of its URI, but it does continue to exist also if it changes its location or if the server it is stored on becomes offline.
- **conceptual mapping:** if a resource is intended as a “conceptual mapping” [3] then its identity is purely formal. For this reason it cannot be also intended as a “computational object”. As a conceptual mapping, a resource can be characterized as a location in the virtual space of the combinatorial regions that are identified by the URIs. Consequently, the identity of a resource in this sense is equivalent to a localization in that space. As a matter of fact, without that space, it would not exist, and its URI is sufficient to identify it unambiguously.
- **proxy:** considering the principle of *opacity* [21], the sense of resource can be that of a ‘proxy’ that is localized in a region of the virtual space identified by the URIs. In this case, the resource is actually intended as a computational object, and its identity is given by the set of elements composing the proxy. For example, an English text, a picture, a metadata schema, etc. According to this meaning of “resource”, its identity goes beyond its location. A resource does exist beyond its location, and its identity holds over its presence on the web.
- **entity:** by defining “resource” with the meaning of an entity [4] - being it either a computational object or not - is problematic because the relationship that holds between a resource and a URI would be the same for addressing computational objects and physical or abstract objects. This approach is problematic, because it attempts to address entities (i.e., physical and abstract objects) that are not addressable in principle.

However, besides these interpretations, the identity of entities referenced on the web is *de facto* implemented as the location at which a resource is placed. This implicit assumption is very confusing when we want to use a URI to reference entities that are not web resources. In other words, there is a need for an explicit distinction between the identity of entities, the reference of a resource

and its identifier. For example, the `http://www.w3.org` URI has its own identity as an identifier (a string), the web location it is associated to has its own identity as a(n abstract) place, the web document has its own identity as a computational object (a file), and the subject of the document has its own identity (the W3C organization as a social object). Now, a question like the following can arise: when used in a resource, does the URI “`http://www.w3.org`” turn up identifying the web document that is placed at that web location, or the W3C organization?

There have been many proposals suggesting different approaches to the aim of addressing the issue. A brief summary of some significant ones is summarized here.

Alistar Miles [30] describes his perception of the problem by identifying a possible obstacle: the creation of a same URI for representing different concepts. This has also been named *URI collision* [21]. Miles proposes an interesting ‘low level’ approach as a best practice, that of using HTTP URIs to address entities that are not accessible on the web. He proposes to manage the problem at the server side by means of a negotiation on how to resolve the URI. For example, if one creates the URI `http://foo.com/me` to describe himself/herself, then it could be resolved by the server as the URI `http://foo.com/me.html` or `http://foo.com/me.xml` or other, depending on a sort of configuration of the browser.

Steve Pepper [34] expresses a similar difficulty about the use of URIs for identifying all kinds of entities. In particular, he proposes to associate a resource to a document, whose content describes the subject of the resource (i.e., a subject indicator). Nevertheless, this solution leaves the responsibility of interpreting the identity of a resource to a human agent, and there is no way to ensure that the subject indicator refers to a single subject.

Kendall Clark [9] discusses the ‘tidiness’ of web specifications, and the importance to clarify the conceptual assumptions upon which the web is built, and the semantic web is being built.

David Booth [6] proposes an informal categorization of what can be identified by a URI, suggesting the definition of different conventions for each of the four uses he has identified.

John Black [5] suggests to create a sort of machine-oriented Wikipedia, which shares knowledge through the construction of web sites such as `del.icio.us` [11].

Parsia and Patel-Schneider [32] deeply analyze the issue of defining meaning in the SW. They propose to determine the meaning of a document as the result of an entailment. In this sense, “only documents explicitly mentioned in constructs like the OWL importing mechanism contribute to the meaning of that document.”

Bouquet et al. [7] propose to build a system, “OkkaM”, to implement a catalog of URIs that 1:1 reference entities. Those URIs should be reused as much as possible, supported by tools, and advised as a good practice to refer to entities.

Another good suggestion comes from Pat Hayes [20] who underlines the difference between *access* and *reference*. Both are relationships between names and things, but they are inherently different and the fact that [21] does not distinguish between the two contributes to cause confusion.

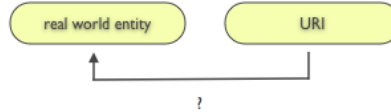
Recently, in the context of a W3C working group, an effort on how to embed RDF triples in HTML has produced a working draft with a proposal for a syntax, RDFa [1], for typing html links. This is discussed in section 5.

All the above proposals are important contributions to solve the “identity” problem. However, none of them provides a comprehensive analysis of the aspects involved in the “identification of resources” problem domain, and how they impact on the web. What is more, no proposal contains a formal semantic model that describes a common ground to situate solutions at either the syntactic or operational levels. Our goal is to cover this lack, while giving justice to the existing solutions that have been devised for the web identity problem.

### 3 Issues in the problem space

The story we have told shows that the problem of web resource identification has been approached from different perspectives. In this section we want to answer the following question: what issues and needs are involved in the identification of resources? And how they impact on web science? From a critical analysis of the state of art presented above, and the preliminary distinctions drawn between URIs, web resources, and entities, at least five different issues emerge:

- I. **Web semantics.** How to clarify the semantics of the web: what are its basic notions, and how can we formalize them [18, 9]?
- II. **Sense of referencing.** How to clarify what it is meant by *referencing things* [20, 32]?
- III. **Multiplicity of referencing.** How to clarify whether (or when) a reference to something is unique or multiple (this is related to the so-called *uniqueness* principle in [23]), and whether only one identifier is admitted for it (related to the *singularity* principle in [23, 14, 7, 33])?
- IV. **Coupling between web and real world.** How to make explicit the relations between web elements and objects in the real world [21, 4, 18]?
- V. **Resolvability of references.** How to clarify when and how a reference is resolvable [6, 30]?



**Fig. 1.** The URI-entity relation

In order to understand the above issues, which characterize our problem space, and possibly to improve on the current situation, we need to analyze the reason why a URI can be associated with an entity. In other words, we need to understand the nature of the apparently simple relation that is informally depicted in Figure 1.

The next section presents an ontology named *Identity of Resources and Entities on the web (IRE)*. IRE allows us to formally describe the nature of the relation between a URI and (one or more) entities, as well as to express the five issues characterizing the space of the web referencing problem.

## 4 The IRE metamodel

In this section we present the IRE metamodel. We firstly provide an informal description of the rationale behind the metamodel.

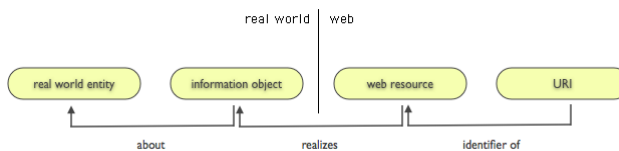
The relation in Figure 1 is directly connected to a general assumption of computer science, usual in web science too: *virtual world is made of symbols while real world is made of things*.<sup>2</sup> This makes it impossible for machines to recognize (or “resolve”, or “refer”) entities “as such”, unless they are symbols as well. Typically, computational reference to entities implies either that humans will interpret it, e.g. when a web page includes the string “W3C”, an image of downtown Prague, etc., or that computational simulations of those entities substitute real world entities, e.g. when dice are thrown in a virtual casino application.

Most problems of web referencing are due to this assumption, therefore we need to analyze in more detail how URIs can be interpreted as references to entities.

Referencing is analyzed in the IRE design by assuming four layers. These layers distinguish the types of things in the domain of the web referencing problem: *URI*, *web resource*, *information object*, and *entity*, as shown in Figure 2.

An example of layering is the following: the URI “http://www.w3.org” identifies a file (a web resource), stored on a W3C server that is accessed when the above URI is resolved; the file is made up of e.g. linguistic or XHTML information (a set of information objects); that information is about the actual W3C organization (a real world entity).

The general assumption mentioned above (in the context of web science) can be now rephrased: *the web is made up of URIs and web resources. Real world is made up of entities whatsoever, including information objects, humans, substances, cables, etc.* Real world can only be processed by agents that have adequate recognition and processing capabilities. The topmost problem is then: *how to encode the real world parts on the web, and in a way that approximates intelligent agents’ recognition and processing of those parts?* Answering this question is part of the solution space (cf. section 5), while in the rest of this section we detail the IRE layers and their formalization.



**Fig. 2.** Four layers of web referencing

<sup>2</sup> Whatever we assume to exist in the real world: either physical or social, mental or abstract, possibly including symbols themselves.



Each layer can be related to the next one as follows (see Figure 2).

**A URI identifies a web resource.** The URI mechanism creates a combinatorial space made of what we call *abstract web locations*. Each abstract web location e.g., the one localized by “http://www.w3.org”, can “contain” a computational object, e.g. a digital file stored on a W3C server. If it is accessible through the web, i.e., the URI is resolvable, then the computational object is a *web resource*.

This first step allows us to reduce the problem space from analyzing the nature of relations between URIs and real world objects to that of analyzing the relations between web resources and real world objects.

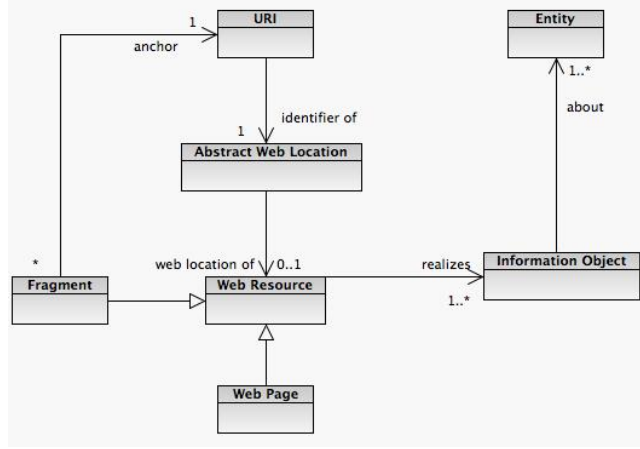
Web resources contain information that has been produced and has entered a life-cycle, e.g., the information encoded in English or XHTML within the W3C web resource. We say that **a web resource realizes some information object**. The *realizes* relation is the same that holds for example between a poem (an information object) and the printed book containing it (its realization). Consider also that same poem as realized by a web document: it would be a different occurrence of the realizes relation for that same information object.

This relation crosses the line between the web and the real world. Now, if we accept the *opacity* principle (web resources can change for a same URI) *for all URIs*, the uniqueness of reference would be already broken at this point. On the contrary, if we admit that at least some URIs are unique, our problem space is reduced to analyzing the nature of the relations between information objects and real world entities.

The assumption that certain URIs are unique lets us going out of the computational world, and entering issues that are traditionally covered by philosophy, linguistics, and logic. In this article we do not enter the details of the literature from those domains, but we use some results of those disciplines, in order to provide a rationale that gives room to a formal description of the problem space that is addressed by solutions like the semantic web and several web 2.0 applications.

We generalize the relations between information objects and real world entities as “being about”: **an information object is about some real world entity**. For example, the linguistic description of W3C from the W3C pages *is about* W3C. The being about relation requires that information objects are interpreted by someone that is able to conceive a “reference” from information objects (either those contained in a resource, or others that can be associated with them), to a set of circumstances, in which real world entities are “situated” (see [17]).

How can we express and operationalize the being about relation, so analyzed, on the web? In order to answer this question we need to enter the solution domain, which will be done in section 5. Figure 3 depicts a UML [19] class diagram containing the basic elements and relations of the IRE model. In the remainder of this section, we still concentrate on the problem domain and on the formalization of IRE, which is summarized as follows. **URIs identify Abstract Web Locations**. A *URI* is the *identifier* of an *Abstract Web Location*.



**Fig. 3.** The IRE basic model

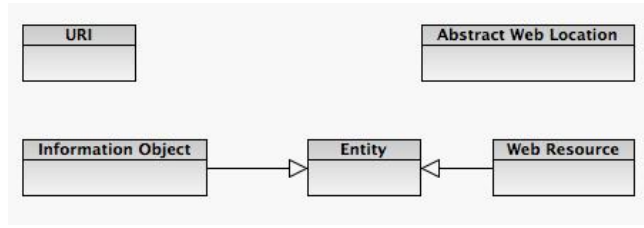
**Abstract Web Locations are locations of Web Resources.** Each *Abstract Web Location* can be the *web location of* at most one *Web Resource*. On the contrary, a *Web Resource* can be placed in one or more *Abstract Web Location*(s). This means that the identity of a web resource is something that goes beyond its location. An abstract web location is a point in the combinatorial space that is created by the URI addressing mechanism (i.e., each URI identifies one and only one abstract web location)<sup>3</sup>. To this aim, the datatype relation *identifier of* and its inverse are functional (have a 1:1 cardinality).

**Anchors relate fragments of web resources to URIs.** A URI can also be used to create an association between two web locations, as with *anchor* relations, e.g., HTML “href”. Notice however that anchor relations hold between a *fragment* of a web resource (around which the href is placed), and a URI, even if the intention of the anchor creator is to link two web resources. As a matter of fact, if the URI cannot be resolved anymore, or if the linked resource does not contain anymore the original information that the anchor creator intended to be reachable, the link is not affected by that.

**Web Resources realize Information Objects that are about Entities.** A *Web Resource* realizes one or more *Information Objects*, which in turn are *about* some real world *Entity*/ies. In order to make the model clearer, Figure 4 depicts the IRE basic taxonomy. We give a prose description for each element that has been introduced:

**URI:** a string that satisfies syntactical rules defined in [4].

<sup>3</sup> Notice that IRIs (Internationalized Resource Identifier) [26] are supposed to replace URIs in the next future. Modulo that replacement, IRIs involvement in the IRE model is the same that URIs have.



**Fig. 4.** The IRE basic taxonomy

**AbstractWebLocation:** a point in the combinatorial space identified by the URI mechanism.

**Information Object:** an information unit that has been created by some agent at some time for some reason. Information objects range from texts to pictures, from poems to logical formulas, from diagrams to sounds, and are independent from their physical realization.

**WebResource:** a computational object<sup>4</sup> that is made available on the web, hence accessible through a web protocol (e.g., a document, a web service).

**Entity:** anything in the real world (material, social, cognitive, etc.).

With regard to the five issues listed in section 3, they can be now redescribed in terms of the following IRE model elements:

- I:** Web Resource, URI, and Anchor provide the basic primitives for expressing the issues raised by web semantics.
- II:** The four layers and the three corresponding relations allow to model the web referencing problem.
- III:** The issue of multiple referencing can be represented as a path thorough IRE relations (and cardinalities), as constructed from the class URI to the class Entity.
- IV:** IRE describes elements and relations that are both in the web and in the real world, hence providing what is needed in order to express the “coupling” issue.
- V:** URI and Web Resource are the primitives involved in the resolvability issue.

Moreover, not only does IRE allow to formally model the problem space of web referencing issues, but it also provides means to understand the relations between those issues. This additional feature of formalizing web semantics will be furtherly exploited when we deal with the solution space in section 5.

<sup>4</sup> In this context, we define “computational object” as (i) the physical realization of an information object, (ii) something that can participate in a computational process. Examples of computational objects are: a database, a digital document, a software application

## 5 The Solution space

In the previous sections, we have analyzed the problem space of web referencing.

In this section, we discuss how the solution space is populated by existing technology.

The **web** as it was originally designed only deals with relations between URIs and web resources. The URI mechanism provides a tool for denoting the location of web resources.

Recently, so-called **web 2.0** applications, such as Flickr [15], Wikipedia [36], del.icio.us [11], etc., have exploited *tagging* mechanisms as a solution to clarify web referencing. Actually, it was possible since early versions of HTML to add “meta” tags to the header part of web pages. They were intended mainly as an aid to search engines, and only very experienced users used to annotate web pages.

Web 2.0 applications allow users to *tag* web resources and are usually associated with a social community that share a common interest. Tags are used as hints on the intended meaning that is assumed by a user to the information contained in a web resource. For example, consider the Flickr community web site. Each Flickr user has his/her own web space where he/she can put on pictures and share them with the Flickr community of users. Each picture can be commented and users can *tag* them either by creating a new tag, or by using so called *machine tags*. On Flickr web site [15] you can find the following suggestions about *tags*:

You can give your photos a “tag”, which is like a keyword or category label. Tags help you find photos which have something in common. You can assign up to 70 tags to each photo.

Following IRE, tags are information objects that “are about” entities, types of entities, or even relations between entities that are also referenced by the tagged information objects. In other words, web 2.0 adds new information objects to the web, in order to suggest useful hints on how the information contained in web resources (web pages, images, videos, bookmarks) can be classified and retrieved.

Lately, web 2.0 trend turned to use controlled vocabularies for structuring tags (e.g. with Flickr “machine tags”). However, the interpretation of tags is still entirely on humans; therefore tagging remains within the traditional web metamodel, and is not sufficient to clarify the *about* relation between information objects and entities.

For example, figure 5 shows a typical example of tags associated with a picture on the Flickr web site. The picture shows Aldo Gangemi during his visit in Prague. The first two tags are about the picture’s content. The second two are about the technical details on the picture seen as a web resource i.e., the camera with which it has been taken, and its resolution. Furthermore, there are also two machine tags taken from the namespace identified by the prefix `geo:`. They are about the relation between the picture and the place at which it has been taken. The first indicates its longitude, while the second indicated the latitude.



**Fig. 5.** An example of *tagging* (1)

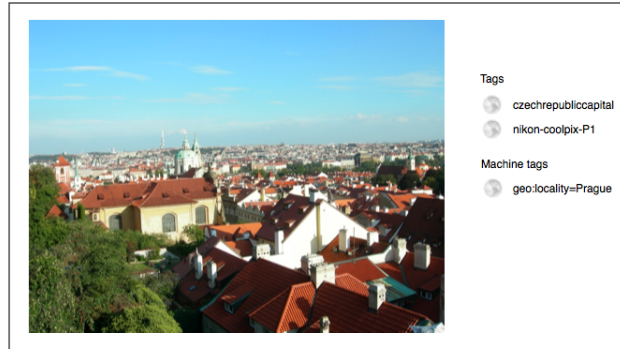
Figure 6 shows another example of Flickr-style tagging. The picture depicts a view of Prague that is referred by the machine tag `geo:locality`, and by the `czechrepubliccapital` tag. There is no way to relate the `geo:lon` and `geo:lat`, to the `geo:locality` machine tags, as well as `Prague`, and `czechrepubliccapital` appear to be unrelated to a machine.

Additional solutions have been introduced by the **semantic web** initiative. The semantic web architecture [21] provides a stack of languages that allows to express web referencing in more explicit and sophisticated ways. RDF [35] allows to create graphs of URIs with a relational syntax, and allows the declaration of explicit schemas (RDFS, [8]) to create those graphs with reference to the names included in the schemas. OWL [31] allows to add a *formal semantics* to those graphs and names, which become ontology elements. The formal semantics used for OWL is *set-theoretic semantics*.

In this article, we create a correspondence between those names and *tags* used in web2.0 applications, assuming that tags can be given a formal interpretation similarly to names used for OWL ontology elements.

Formal semantics allows a rigorous interpretation of the names included in an *ontology*. An ontology is a logical theory<sup>5</sup>, which is composed of (briefly and

<sup>5</sup> The term ontology has been used also to refer to many data structures: thesauri, lexicons, folksonomies, database schemas, logical theories, etc., independently from how they are encoded, but we focus here on its use within the semantic web, where *ontology* indicates either RDF schema or OWL models).



**Fig. 6.** An example of *tagging* (2)

informally speaking): predicate (e.g. OWL classes and properties) names, individual (e.g. OWL individuals) names, and axioms (what is asserted in a particular ontology, e.g. OWL class subsumption and disjointness, `rdf:type` assertions, relationships between OWL individuals, etc.).

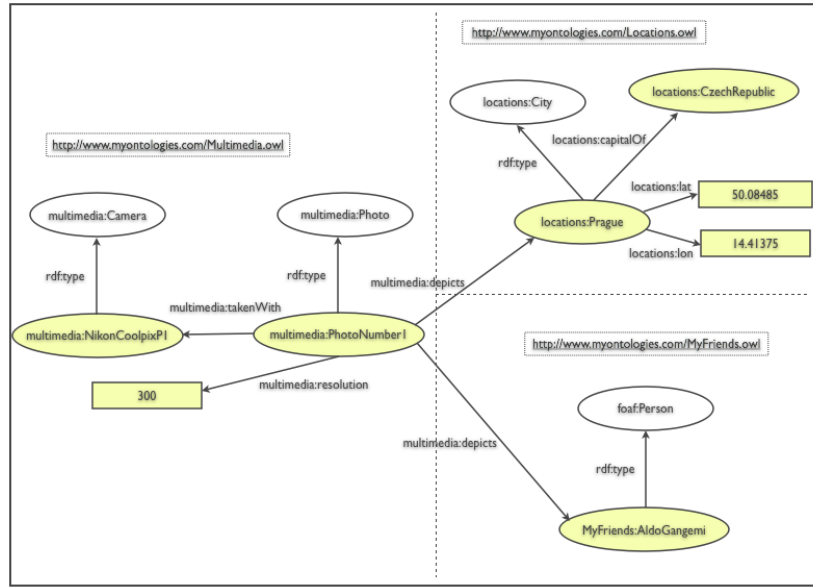
Predicate and individual names are identified by URIs in web ontologies. On the contrary, web 2.0 tags are not usually identified by URIs. This is important from the viewpoint of web semantics, but it is less important when we consider tags as potential names for ontology elements, once they are given a formal semantics. However, this difference is not crucial in the rest of this section.

In this article, we are interested in how ontology element names (and their context, provided by the axioms within an ontology) help making web referencing more precise, and possibly more precise than in web 2.0 applications. In order to do that, we need to know firstly how the names used in ontologies refer to real-world entities. This is straightforward: within set-theoretic semantics, individual names are assumed to refer to individuals in the real world<sup>6</sup>, while predicate names are assumed to refer to sets of individuals or to sets of relationships between individuals in the real world. Individuals are assumed as elements of the *domain of interpretation* of an ontology.

Therefore, from the viewpoint of IRE, we can assume that OWL ontologies introduce new names, distinct for either individuals or predicates, which are the formal counterpart of real world entities. Figure 7 depicts, by using a graph notation<sup>7</sup>, three simple OWL ontologies. The “Multimedia” ontology contains classes and properties for describing multimedia objects like, photos, videos, digital cameras, etc. The “Location” ontology addresses the description of geo-

<sup>6</sup> Notice that we are using “real world entity” here to mean any entity except the symbols used on the web

<sup>7</sup> White ovals are used for OWL classes, yellow ovals are used for OWL individuals, rectangles are used for datatype values e.g., strings, float, etc., while arrows are used for object and datatype properties



**Fig. 7.** An example of *tagging* with OWL

graphical places and their characteristics. Finally, the “MyFriends” ontology is about people that are friends of the ontology author<sup>8</sup>. By instantiating these simple ontologies, the pictures of Figure 5 and 6 can be tagged with OWL individuals and “property values”. In this way, it is possible to formally express what we are referencing to, and e.g. to relate the two pictures, even in case different terms for tagging are used. For example, consider the case of the photo of Figure 5 that is represented by the OWL individual `multimedia:PhotoNumber1` of Figure 7. With only this assertion i.e., `multimedia:PhotoNumber1 rdf:type multimedia:Photo`, we are saying something about it i.e., it is a photo. Moreover, we can relate the photo to the camera we used to take it, which is represented by an OWL individual with a specific type as well. The photo is then related to `MyFriends:AldoGangemi`, which is a Person, and to `locations:Prague`, which is formally defined as a city, capital of the Czech republic in the Location ontology. Of course, analogously to having different web 2.0 tags with the same sense, we could have different ontologies for the same domain, defining individuals and predicates that have the same sense in those different ontologies. However, in this last case there are other formal constructs that can express these equivalences.

Currently, the semantic web provides the best approach for addressing web referencing (issue II), because it allows to create specific symbols, usable within web resources, which are associated with information objects and real world entities.

<sup>8</sup> Notice that we re-use here the class `foaf:Person`

In practice, the semantic web provides languages to create new symbols with associated URIs for information objects (individuals and predicate tags), which are assumed to have a unique reference (called “denotation” by logicians) to the individuals and concepts that exist in the real world. Such symbols are in fact additional web resources, which are available to a machine in order to “recognize” entities, thus gathering a more explicit processing of web reference. For the sake of clarity, we call “semantic” the URIs that identify these additional web resources.

There is something that is still missing from the picture. While formal semantics allows to establish a unique relation between semantic URIs and entities, the *reason* why those entities are referenced is not addressed by formal semantics alone, since it has to do with the motivations, goals, and rationales of the humans, communities, and agents that use the (semantic) web. Such motivations can be extremely varied, and while formal semantics ensures the uniqueness of reference within a formal domain of discourse, it cannot ensure that the structural context provided by an ontology (its axioms) is enough to unambiguously reference an entity (i.e. that the domain of discourse actually reflects the world as intended by a human interpreter). Ontologies can make different assumption (different axioms) on a same set of real world entities, so that their domains of interpretation may need to be mapped in non-trivial ways.

This last observation makes us move from the problem of the *sense* of web referencing (issue II.) to the problem of the *multiplicity* of referencing (issue III.), which is another way to deal with the *identity* problem.

According to [23], in order to have an effective “object identification mechanism”, identifiers have to satisfy the following principles:

**Immutability:** an object’s identifier should be the same at any point in time and everywhere (globally recognizable)

**Uniqueness:** two objects cannot be represented by the same identifier

**Singularity:** two different identifiers cannot represent the same object.

Assuming URIs as identifiers (immutable by definition) and entities as objects, a specific web ontology (by which reference is unique and singular, as explained above) can satisfy the above principles. This could be effective if it was reasonable to assume that ontologies are univocal in referencing real world entities. But we know this is not reasonable.

The names of logical elements are useful to give hints on how (the URI of) a tagged web resource references entities, but this hint is obtained by putting those names in the namespace of a specific ontology, therefore the same entity can appear in different namespaces with either the same or different names. For example, most web users know that Italy is an existing entity, whatever existing may mean for them and whatever interpretations those users may want to attach to Italy as an existing entity. When we put Italy in some ontology domain of interpretation, the result is that Italy is being assumed as a new entity, and an owl:sameAs axiom is required to assert the identity of Italy in the domain of ontology A as the same as that of Italy in the domain of ontology B.



Moreover, our web referencing problems deal with how URIs of *any* web resource reference entities, and not only with how web 2.0 tags or the names of web ontologies reference those entities. Consequently, the fact that ontology URIs references are unique and singular does not ensure that the reference of (the URI of) the web resources tagged with those URIs will be unique and singular as well.

The Italy example suggests us a workaround that can lead us to a new breed of solutions that can fit very well in an evolving space of web ontologies.

Consider the possibility that there exists a collection of entities in the real world, each element of which can be assigned a unique and singular URI on the web, and with no dependence on any specific context. Let's call the resulting URI collection: *web catalogue of entities*.

An example of such a catalog is [7], which presents an application for "identity management support". The authors propose a catalog of URIs that uniquely and singularly reference entities. Those URIs should be reused as much as possible, supported by tools, and advised as a good practice to refer to entities.

Now, if we fill such catalogue with semantic URIs, those URIs will identify semantic web resources, e.g., OWL individuals. The semantic catalog workaround consists then in relaxing the interpretation of referencing to the point of having an ontology with an empty set of predicate names, but with as many individual names as possible. Such an ontology would only include individual names: no predicate names or axioms.

Contrary to current practices for ontology design and exploitation, which firstly create predicate names and axioms, and then populate the namespace with individual names, the semantic catalog suggests to populate an ontology with the individual names that are already realized in the catalog, and then to add axioms for the tagged entities. Each ontology could provide a view on the same set of individuals, thus avoiding the difficult and arbitrary practice of introducing owl:sameAs axioms in order to map differently viewed individuals.

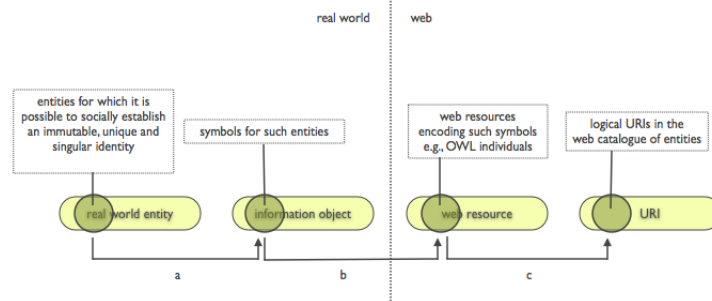
The potential advantages of reusing URIs in different ontologies is intuitive, as well as the fact that [23]'s principles applied to such URIs would allow their easy reuse, and provide a solution to the identification of entities as referenced on the web.

Solving the identity problem by means of a catalog seems quite simplistic, but it's a matter of fact that our social behavior makes a lot of assumptions of existence, without caring about what is the reason why we assume that existence, or that reason is so shared and relevant to most of us that we can do without asking about it. Examples include countries, geographic locations, institutions, organizations, persons, well known objects (buildings, roads), historical facts, etc.

Although it is true that assertions of existence for entities are dependent on a specific context, which is characterized by certain motivations, goals, and rationales, there are entities whose existential conditions are *socially recognized* i.e., their existence is a social statement. In order to understand those conditions, the usage of the web needs to be observed.

The web is growing as a virtual simulation of social processes. It is not much related to physical processes as instead it is the case with “the internet of things” [14]<sup>9</sup>[14]. It is reasonable to assume that the criteria by which entities should be included in the catalogue are the same that govern the social agreement over the way entities are publicly shared. A similar intuition is described in [29], where the relations between social networks and semantics are discussed. [17] also discusses the contextual nature of identity conditions for entities. Social agreement on what should be assumed in a universe of discourse seems much more powerful than philosophical discussions on what should be assumed to exist.

For example, many entities that physically exist in the real world e.g., a stone, only acquire value for the web when they are put in a specific context e.g., collaborative geological cataloguing. This is in line with the fact that associations between URIs and entities depend on a given context. On the other hand, there are entities that have no physical existence, but whose existence is universally recognized because it satisfies well known social conditions: large organizations (e.g. United Nations), corporations (e.g. Daimler Benz), famous fiction characters (e.g. Mickey Mouse, Jupiter), concepts (e.g. force, beauty, being a friend), etc.



**Fig. 8.** The procedure for creating the web catalogue of entities.

Figure 8 informally depicts the procedure by which a web catalog of entities is built from the perspective of the four layers from IRE (see section 4). Specifically, by relation *a* an entity, whose identity is socially recognized regardless to any specific interpretation, is associated with a symbol, i.e. an information object that is about it, complying e.g. to the semantics of an OWL individual name. Relation *b* holds between such symbol and its realization on the web,

<sup>9</sup> The internet of things could be an additional component to be exploited on the web for supporting identification of entities; for example, to realize the so-called ubiquitous computing ([22]). Nevertheless, this aspect is out of scope of this paper.

i.e. a (semantic) web resource for that name. Finally, relation  $c$  represents the association of a web location with that web resource, hence a URI.

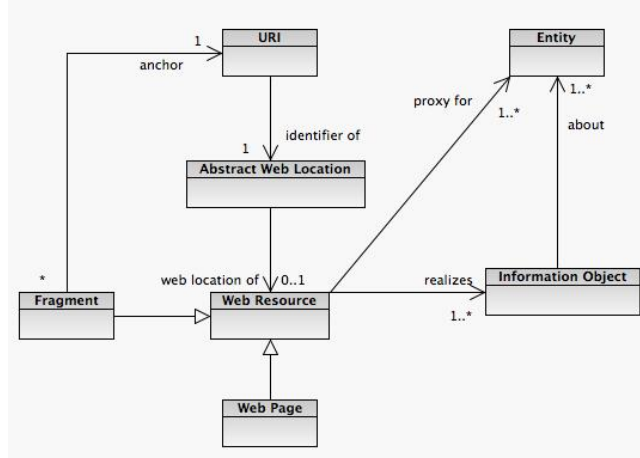
The catalogue idea can be stretched, in order to cover associations too. An association catalogue can also be built, if the semantics of association between entities is also relaxed, but based on evidence coming e.g. from statistical methods (such as those currently used by search engines), from NLP techniques (e.g. named-entity recognition and relation extraction), and from existing association repositories (factbooks, organizational thesauri, etc.). An additional merit of these ideas is to disclose a new strategy to integrate top-down work from ontology engineering and the semantic web, with bottom-up work from NLP, information retrieval and extraction, and folksonomy-based web applications. A formal model of the association catalogue idea is presented in the next section.

## 5.1 The solution domain in terms of IRE

In this section we present the IRE model extended in order to cover the needed expressivity for representing the solution domain. Although on the web we can not resolve the reference to an entity of the real world that is not a computational object, we need to be able to assert facts about it on the web. As mentioned above, we can approach this issue in different ways depending on the technology we use.

In general we say that: a web resource *functions as a proxy for* an entity, at a given time. This association between a web resource and an entity means that the web resource realizes an information object, which is about some entity/ies at a given time.

Figure 9 depicts the IRE diagram, which includes the relation *proxy for* between a web resource and an entity.



**Fig. 9.** The IRE model for identification of resources on the web

A web resource can be a proxy for one or more than one entity. This aspect suggests a categorization for the *proxy for* relation. Notice that each specialization of *proxy for* can correspond to a different computational approach, or more specifically to a different operational semantics associated with the resolution of the web resource's URI.

The kinds of *proxy for* relations can be described informally as follows:

**Approximate proxy for** : is a relationship between a web resource and more than one entity at a given time, where the web resource realizes some information objects, which are about those entities. In this case the web resource approximately represents all the entities.

**Exact proxy for** : a relationship between a semantic resource and one entity at a given time, where the semantic resource realizes an information object, which is about only that entity, and describes it through a semantic structure. For example, an individual of an OWL ontology.

For example, consider the web catalogue of entities we discussed in section 5. It contains, among the others, two individuals: one for “Prague”, the other for “Aldo Gangemi”. These two individuals are exact proxy for the city of Prague and the person Aldo Gangemi, respectively. While the photo shown in Figure 5 is an approximate proxy for the person Aldo Gangemi and the city of Prague.

## 6 Conclusion and remarks

The *identification of resources on the web* has been mentioned as one of the essential issues to be addressed by web science in order to successfully perform information sharing, interoperability, and reasoning on the web.

In this paper we have firstly analyzed the state of the art related to this problem and from this analysis we made it emerge five distinct issues. We have described these issues and compared the respective solutions through the investigation of the reason why a URI can be associated with an entity. We have proposed an ontology called *Identity of Resources and Entities on the web (IRE)*, which encompass the things in the domain of discourse of the web referencing problem. Furthermore, we have discussed the solution domain in terms of how it has been affected by the current evolution of web science from the web to web 2.0 and to the semantic web. We have also considered some proposed and envisaged solutions and shown how they can be modeled in terms of IRE. IRE has been formalized in *first order logics*<sup>10</sup>, and in OWL(DL)<sup>11</sup>.

---

<sup>10</sup> The FOL formalization can be found in the Appendix (see section 7)

<sup>11</sup> The OWL version of IRE is available at <http://wiki.loa-cnr.it/index.php/LoaWiki:IRE>

## **Acknowledgment**

We are grateful to the members of the NeOn consortium who contributed to the NeOn vision being funded by the European Commission 6th IST Framework Programme. Further information on NeOn is available on <http://www.neon-project.org>.

## 7 Appendix: FOL Formalization of IRE

The IRE model specializes the DOLCE+ (in the Ultralite variety, abbreviated as DUL) reference ontology [12, 13], and some of its modular extensions, namely Information Objects (IOLite) and Knowledge Content Objects (KCO) modules (the last two have been developed in the EU Metokis project [28, 2]). All modules are available in OWL at <http://www.loa-cnr.it/ontologies/>. For a complete report the reader can refer to [27, 16].

The IRE ontology specializes or reuses the following predicates (classes have capitalized names, here we only show the basic taxonomic axioms; for full axiomatization see indicated URLs and [27]).

IRE uses the following predicates<sup>12</sup>:

$\{dul:Entity, dul:Region, dul:AbstractRegion, dul:isRegionFor,$   
 $dul:InformationObject, dul:InformationRealization, dul:realizes,$   
 $kco:ComputationalObject, dul:TimeInterval,$   
 $dul:Method, dul:describes, dul:about,$   
 $URI, AbstractWebLocation, ResolutionMethod, ,$   
 $WebResource, anchor, hasIdentifier, webLocationOf,$   
 $iol:FormalLanguage, iol:hasRepresentationLanguage, ProxyFor,$   
 $ApproximateProxyFor, SemanticResource, ExactProxyFor.\}$

The following axioms either characterize or define the above predicates:

$$(1) \quad hasIdentifier(x, y) \rightarrow dul:Region(x) \wedge xsd:Datatype(y)$$

Axiom (1) introduces a relation between identifiers (specializing the class *Region* from DOLCE) and datatypes as encoded in XSD. In OWL, this is a so-called *DatatypeProperty*.

$$(2) \quad URI(x) \rightarrow xsd:Datatype(x)$$

A URI is characterized in (2) as an XSD datatype, since this is the current practice for xml-based languages like OWL. The possible integration between DOLCE regions and datatypes consists in assuming a datatype structure as a metrics for DOLCE regions; for example,  $xsd : Date$  can be assumed as a metrics for a subset of 'time intervals', which are regions in DOLCE.

<sup>12</sup> We use the prefixes *dul:*, *iol:*, *kco:*, and *xsd:* for predicates imported from DOLCE-Ultralite (<http://www.loa-cnr.it/ontologies/DUL.owl>), Information-Objects-Lite (<http://www.loa-cnr.it/ontologies/IOLite.owl>), Knowledge Content Objects (<http://www.loa-cnr.it/ontologies/KCO/KCO.owl>), and XML Schema (<http://www.w3.org/2001/XMLSchema>) respectively.



$$\begin{aligned}
& \text{AbstractWebLocation}(x) =_{def} \text{dul:SpaceRegion}(x) \wedge \\
& \quad \exists y(\text{URI}(y) \wedge \text{hasIdentifier}(x, y) \wedge \\
(3) \quad & \quad \neg \exists z(\text{URI}(z) \wedge y \neq z \wedge \text{hasIdentifier}(x, z))
\end{aligned}$$

Abstract web locations are defined in (3) as DOLCE abstract regions that have exactly one URI as an identifier.

Furthermore, in (4), we state that no URI can identify more than one abstract region. This ensures the isomorphism between URI-like encoding and the abstract web space generated by it (the abstract web space being the maximal sum of all abstract web locations).

$$\begin{aligned}
& \text{AbstractWebLocation}(x) \rightarrow \neg \exists (y, z) \\
& (\text{URI}(y) \wedge \text{AbstractWebLocation}(z) \wedge x \neq z \wedge \\
(4) \quad & \quad \text{hasIdentifier}(x, y) \wedge \text{hasIdentifier}(z, y))
\end{aligned}$$

(5) defines a relation between an abstract web location and a computational object at a time interval. The definition specializes the relation *eAbstractLocationOf* (imported from the Spatial Relations ontology) for abstract web locations and computational objects. *eAbstractLocationOf* is a generic relation holding between regions and physical objects.

$$\begin{aligned}
& \text{webLocationOf}(x, y, t) =_{def} \\
& \quad \text{dul:isRegionFor}(x, y, t) \wedge \\
& \quad \text{AbstractWebLocation}(x) \wedge \\
(5) \quad & \quad \text{kco:ComputationalObject}(y) \wedge \text{dul:TimeInterval}(t)
\end{aligned}$$

Notice that being a web location of a computational object does not imply the successful resolution of the URI for the abstract web location into the computational object (see (7) below), but only the assignment of an address to the resource.

$$(6) \quad \text{ResolutionMethod}(x) \rightarrow \text{dul:Method}(x)$$

$$\begin{aligned}
& \text{WebResource}(x) =_{def} \text{kco:ComputationalObject}(x) \wedge \\
& \quad \exists (m, y, t)(\text{ResolutionMethod}(m) \wedge \text{dul:TimeInterval}(t) \wedge \\
(7) \quad & \quad (\text{webLocationOf}(y, x, t) \wedge \text{dul:describes}(m, x, t)))
\end{aligned}$$

(7) defines web resources as computational objects that have an assigned abstract web location. In operational terms, web resources have been here restricted to those computational objects that are involved in a computational

method that ensures the resolution of a URI, given certain circumstances (e.g. an abstract web location is an assigned web location of the resource, the server is switched on, the connection is active, etc).

$$(8) \quad \textit{WebPage}(x) \rightarrow \textit{WebResource}(x)$$

$$(9) \quad \textit{Fragment}(x) \rightarrow \textit{WebResource}(x)$$

$$(10) \quad \textit{Fragment}(x) \rightarrow \neg \textit{WebPage}(x)$$

(8) and (9) state that both *WebPages* and *Fragments* are *WebResources*, while (10) states that a *Fragment* is not a *WebPage*.

$$(11) \quad \textit{anchor}(x, y, t) \rightarrow \textit{Fragment}(x) \wedge \textit{URI}(y) \wedge \textit{dul:TimeInterval}(t)$$

(11) defines a relation between a fragment and a URI at a given time, e.g., the HTML 'href'.

$$(12) \quad \textit{kco:ComputationalObject}(x) \rightarrow \textit{dul:InformationRealization}(x)$$

The KCO ontology specializes the Information Objects ontology in order to build a conceptual schema for digital and analog content. The concept *ComputationalObject* specializes *InformationRealization* for the computational world, and includes any physical document, electronic service, file, application, etc.

$$(13) \quad \textit{dul:InformationObject}(x) \rightarrow \textit{dul:Entity}$$

$$(14) \quad \begin{aligned} & \textit{dul:realizes}(x, y, t) \rightarrow \\ & \textit{WebResource}(x) \wedge \textit{dul:InformationObject}(y) \wedge \\ & \textit{dul:TimeInterval}(t) \end{aligned}$$

A web resource is the realization of an information object (13) at a given time, as formalized by axiom (14).

$$(15) \quad \begin{aligned} & \textit{dul:about}(x, y, z) \rightarrow \\ & \textit{dul:InformationObject}(x) \wedge \textit{dul:Entity}(y) \wedge \\ & \textit{dul:TimeInterval}(t) \end{aligned}$$

Finally, axiom 15 formalizes the reference relation between information objects and entities at a time.

Below we report the formalization in FOL of the *proxy for* relations. Specifically we introduce the following predicates:

$$(16) \quad \begin{aligned} proxyFor(x, y, t) =_{def} & WebResource(x) \wedge dul:Entity(y) \wedge \\ & dul:TimeInterval(t) \wedge \exists z(dul:InformationObject(z) \wedge \\ & dul:realizes(x, z, t) \wedge dul:about(z, y, t)) \end{aligned}$$

16 defines a relation between a web resource  $x$  and any entity  $y$  the web resource can be a proxy for, at a time interval. The definition says that for  $x$  to be a proxy for  $y$ , it must *realize* an information object that is *about*  $y$ .

In the following definitions, we introduce a typology of proxy relations, independently on the available technology.

$$(17) \quad \begin{aligned} approximateProxyFor(x, y, t) =_{def} \\ proxyFor(x, y, t) \wedge \exists z(dul:Entity(z) \wedge y \neq z \wedge \\ proxyFor(x, z, t)) \end{aligned}$$

A resource is an approximate proxy for an entity (17) when it is a proxy for at least two entities, e.g. an html page can contain more than one link, or a document (text, image, etc.) can refer to several entities, either computational or not.

$$(18) \quad \begin{aligned} SemanticResource(x) =_{def} & WebResource(x) \wedge \\ \exists(y, t, z, w)( & proxyFor(x, y, t) \wedge dul:InformationObject(z) \wedge \\ & iol:FormalLanguage(w) \wedge dul:realizes(x, z, t) \wedge \\ & dul:about(z, y, t) \wedge iol:hasRepresentationLanguage(z, w)) \end{aligned}$$

The definition 18 introduces semantic resources, i.e. the web resources that realize information objects that are ordered (namely, encoded) by means of a formal language, e.g. OWL [31].

$$(19) \quad \begin{aligned} exactProxyFor(x, y, t) =_{def} & SemanticResource(x) \wedge proxyFor(x, y, t) \wedge \\ & \neg \exists z(dul:Entity(z) \wedge y \neq z \wedge proxyFor(x, z, t)) \end{aligned}$$

A resource is an exact proxy for an entity (19) when it is a semantic resource which functions as a proxy for only one entity. For example, the URI-referenced entry of an Italian legal ontology:  
<http://fooformleg.it/ontology.owl#Controparte> is a formal exact proxy for the concept *Controparte* in the Italian Law.

## References

1. B. Adida and M. Birbeck. RDFa Primer 1.0 Embedding RDF in XHTML. Technical report, World Wide Web Consortium, May 2006. <http://www.w3.org/TR/xhtml-rdfa-primer/>.
2. W. Behrendt, A. Gangemi, W. Maass, and R. Westenthaler. Towards an Ontology-Based Distributed Architecture for Paid Content. In *The Semantic Web: Research and Applications: Second European Semantic Web Conference, ESWC, May 29 - June 1, 2005. Proceedings*, volume 3532, pages 257–271, Heraklion, Crete, Greece, 2005. Springer-Verlag GmbH.
3. T. Berners-Lee, R. T. Fielding, and L. Masinter. Uniform Resource Identifier (URI): Generic syntax. Internet Draft standard RFC2396, The Internet Society, August 1998. <http://rfc.net/rfc2396.html>.
4. T. Berners-Lee, R. T. Fielding, and L. Masinter. Uniform Resource Identifier (URI): Generic syntax. Internet Draft standard RFC3986, The Internet Society, January 2005. <http://rfc.net/rfc3986.html>.
5. J. Black. Creating a Common Ground for URI Meaning Using Socially Constructed Web sites. Architecture and Philosophy of the Web Identity, Reference, and the Web (IRW2006), WWW2006 Workshop, Edinburgh, Scotland, May, 23rd 2006. <http://www.ibiblio.org/hhalpin/irw2006/jblack.pdf>.
6. D. Booth. Four Uses of a URL: Name, Concept, Web Location and Document Instance. <http://www.w3.org/2002/11/dboothnames/dboothnames.clean.htm>, 2003.
7. P. Bouquet, H. Stoermer, M. Mancipopi, and D. Giacomuzzi. OkkaM: Towards a Solution to the “Identity Crisis” on the Semantic Web. In *In Proceedings of SWAP 2006, the 3rd Italian Semantic Web Workshop*, Pisa, Italy, December 18-20 2006. CEUR Workshop Proceedings.
8. Brickley D. and Guha R. V. RDF Vocabulary Description Language 1.0: RDF Schema. Technical report, W3C, 2004. <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>.
9. K. Clark. Identity Crisis. <http://www.xml.com/pub/a/2002/09/11/deviant.html>, 2002.
10. W. W. W. Consortium. W3C Web Site. <http://www.w3c.org>.
11. del.icio.us. <http://del.icio.us/>.
12. DOLCE - Project Home Page. <http://dolce.semanticweb.org>.
13. DOLCE Ontology. <http://www.loa-cnr.it/ontologies/DOLCE-Lite.owl>.
14. R. A. Dolin. Deploying the “internet of things”. In *SAINT '06: Proceedings of the International Symposium on Applications on Internet*, pages 216–219, Washington, DC, USA, 2006. IEEE Computer Society.
15. Flickr. <http://flickr.com/>.
16. A. Gangemi, S. Borgo, C. Catenacci, and J. Lehmann. Task Taxonomies for Knowledge Content. Deliverable D07 of the Metokis Project. [http://www.loa-cnr.it/Papers/D07\\_v21a.pdf](http://www.loa-cnr.it/Papers/D07_v21a.pdf), 2005.
17. A. Gangemi and P. Mika. Understanding the Semantic Web through Descriptions and Situations. In *Meersman R. et al. (eds.) - Proceedings of ODBASE03 Conference*. Springer, 2003.
18. A. Gangemi and V. Presutti. The bourne identity of a web resource. Architecture and Philosophy of the Web Identity, Reference, and the Web (IRW2006), WWW2006 Workshop, Edinburgh, Scotland, May, 23rd 2006. <http://www.ibiblio.org/hhalpin/irw2006/presentations/-vpresutti.pdf>.

19. O. group. The unified modeling language (uml). <http://www.omg.org/uml>.
20. P. Hayes. In defense of Ambiguity. Architecture and Philosophy of the Web Identity, Reference, and the Web (IRW2006), WWW2006 Workshop, Edinburgh, Scotland, May, 23rd 2006. <http://www.ibiblio.org/hhalpin/irw2006/presentations/-HayesSlides.pdf>.
21. I. Jacobs and N. Walsh. Architecture of the World Wide Web. Technical report, W3C, 2004. <http://www.w3.org/TR/webarch>.
22. T. Kawamura, K. Ueno, S. Nagano, T. Hasegawa, and A. Ohsuga. Ubiquitous service finder discovery of services semantically derived from metadata in ubiquitous computing. In Y. et al. [37], pages 902–915.
23. W. Kent, R. Ahmed, J. Albert, M. Ketabchi, and M.-C. Shan. Object Identification in Multi-Database Systems. In *Proc IFIP TC2.6 DS-5 Conference on Semantics of Interoperable Databases*, Lorne, Victoria, Australia, November 1992.
24. T. B. Lee, W. Hall, J. A. Hendler, K. O'Hara, N. Shadbolt, and D. J. Weitzner. A Framework for Web Science. *Foundations and Trends in Web Science*, 1:1–130, 2006.
25. T. B. Lee, W. Hall, J. A. Hendler, N. Shadbolt, and D. J. Weitzner. Web Science. *Science*, 313, August 2006.
26. D. M. and S. M. Internationalized Resource Identifiers (IRIs). Internet Draft standard RFC3987, The Internet Society, January 2005. <http://www.ietf.org/rfc/rfc3987.txt>.
27. C. Masolo, A. Gangemi, N. Guarino, A. Oltramari, and L. Schneider. Wonderweb eu project deliverable d18: The wonderweb library of foundational ontologies. <http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf>, 2004.
28. Metokis Project. <http://metokis.salzburgresearch.at>.
29. P. Mika. Ontologies are us: A unified model of social networks and semantics. In Y. et al. [37], pages 522–536.
30. A. Miles. Working Around the Identity Crisis. <http://esw.w3.org/topic/SkosDev/IdentityCrisis>, 2005.
31. OWL Web Ontology Language Family of Specifications. <http://www.w3.org/2004/OWL>, 2004.
32. B. Parsia and P. F. Patel-Schneider. Meaning and the Semantic Web. Architecture and Philosophy of the Web Identity, Reference, and the Web (IRW2006), WWW2006 Workshop, Edinburgh, Scotland, May, 23rd 2006. <http://www.ibiblio.org/hhalpin/irw2006/bparsia.pdf>.
33. S. Pepper. The Case for Published Subjects. Architecture and Philosophy of the Web Identity, Reference, and the Web (IRW2006), WWW2006 Workshop, Edinburgh, Scotland, May, 23rd 2006. <http://www.ibiblio.org/hhalpin/irw2006/spepper2.pdf>.
34. S. Pepper and S. Schwab. Curing the Web's Identity Crisis: Subject Indicators for RDF. Technical report, Ontopia, 2003. <http://www.ontopia.net/topicmaps/-materials/identitycrisis.html> (2003).
35. Resource Description Framework (RDF). <http://www.w3.org/RDF>.
36. Wikipedia. Wikipedia Home Page. <http://www.wikipedia.org>.
37. G. Y., M. E., B. V.R., and M. M. A., editors. *The SemanticWeb - ISWC 2005, 4th International SemanticWeb Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005, Proceedings*, volume 3729 of *Lecture Notes in Computer Science*. Springer, 2005.