# Towards an OWL ontology for identity on the web

Valentina Presutti
Laboratory for Applied Ontology
National Research Council (ISTC-CNR)
Roma, Italy
Email: valentina.presutti@istc.cnr.it

Aldo Gangemi
Laboratory for Applied Ontology
National Research Council (ISTC-CNR)
Roma, Italy
Email: aldo.gangemi@istc.cnr.it

***Abstract*— One of the main strength of the web is that it realizes the goal of allowing any party of its global community to share information with any other party. This has been achieved by making use of a unique and uniform mechanism of identification, the URI (Universal Resource Identifiers). Web applications such as search engines have been built up on this mechanism. 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. We investigate the meaning of identity of a web resource and how it can be modeled in order to be implemented on the web. In particular, we propose an ontology that models the problem, and discuss some possible solutions. We describe the concept of resource from the web domain point of view, using an ontology of Information Objects, built on top of DOLCE and of some of its modular extensions. In particular, we formalize some concepts of a *web resource*, and distinguish them from the concept of a generic *entity*. We finally propose a formal pattern for resource modelling.**

## I. 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 that of allowing different parties of its global communities to share information [17]. 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. Such web address is a URI (Universal Resource Identifier) [4]. There is no doubt about the effectiveness of the URI mechanism for the referencing of resources that are realized on the web.

Nevertheless, there is something more ambitious that the web is supposed to allow than just referencing web resources, that is referencing things in general. That ambitious goal requires a software agent to identify a resource unambiguously, in order to perform the appropriate operations on it. Tim Berners Lee et al. in [18] mention identification of resources is essential for information sharing, interoperability, reasoning, and elaboration in general on the web. The semantic web (SW) has as main goal that of make such scenarios possible.

Currently, URIs are used as the uniform mechanism for identifying heterogeneous entities, e.g., documents, metadata, ontologies, abstract concepts, physical things, events. But there is no clear categorization of which are the possible ways to identify and reference those entities on the web. This sort of confusion has led to lack of consensus on which is the most suitable way to solve the problem of handling the recognition of the identity of an element that is referenced by a URI, and consequently there is not a defined operational semantics associated with each of these different sorts.

We propose an OWL [30] ontology named IRE (Identifiers, Resources, and Entities) that was originally formalized in first order logic, containing concepts that compose the architecture of the web. IRE is based on an ontology of Information Objects [2], built on top of DOLCE and on some of its modular extensions [23], [13]. For a complete report the reader can refer to [12], [22], [25], [2], while all OWL files are available at http://www.loa-cnr.it/ontologies. IRE provides the basis for defining a categorization of the kinds of resources that can be referenced on the web. We are confident that, based on this categorization, it is possible to study the most suitable way of handling the operational semantics that can be applied to different references.

The rest of the paper is organized as follows: in section I-A we discuss briefly the existing approaches to the problem of identifying a resource on the web, in section II we discuss the definitions of the concept of resource and its relation to the URI mechanism of identification that can be found in normative documents such as [4] and [17]. In particular, we underline the ambiguity of such definitions, showing the need for a rationale. Finally, in III we present the design rationale of IRE ontology, the most specific contribution of this paper. Section V summarizes the proposal underlining our contribution.

### A. Related work

The identification of resources is an important task to use them on the web. Currently, there is a diffuse feeling that the identification procedures suffer from a lack of consensus about how to handle them. A typical example is the URI http://www.w3.org: does it identify the web document that is placed at that web address, or the W3C organization? There have been many proposals suggesting different approaches to the aim of addressing the issue. In this section we briefly summarize some significant ones.

Alistar Miles [26] describes his perception of the problem by isolating a possible obstacle: the creation of a same URI for representing different concepts. This has also been named *URI collision* [17]. 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 [33] 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 [7] 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 going to be 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, machines that share knowledge through the construction of web sites such as del.icio.us [8].

Parsia and Patel-Schneider [31] 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."

Another good suggestion comes from Pat Hayes [16] who underlines the difference between *access* and *reference*. Both are relationships between names and things, but they are inherently different and the fact that [17] does not distinguish between the two contributes to cause confusion.

All these proposals are important contributions to solve the "identity crisis" but none of them provides a formal model, which could help in finding a comprehensive solution at both syntactic and operational levels. The identity of entities referenced on the web is often intended as the location at which a resource is placed; in other words, there is a need for an explicit distinction between the identity of a resource and its identifier. For example, recalling the W3C web site, the http://www.w3.org URI has its own identity as identifier (i.e., a string), the web location it is associated to has its own identity as a place, the web document has its own identity as a computational object (i.e., file), and the subject of the document has its own identity (i.e., the W3C organization).

Recently in the context of W3C working groups a work effort on how to embed RDF triples in HTML is being conducted and there is already a working draft containing a proposal for a syntax, namely RDFa [1]. RDFa complies to our approach although it does not cover all aspects.

## II. URI AND RESOURCES

The term "resource" is generally used for all things that might be identified by a URI [17].

In literature, we find several definitions for the term "resource" used in the context of world wide web. In particular we quote here three authoritative documents, [4], [3], [17][1] 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.

Moreover, in the same document the mechanism for identifying resources (i.e., URI) is specified, and it is also said that:

> Having identified a resource, a system may perform a variety of operations on the resource, as might be characterized by such words as 'access', 'update', 'replace', or 'find attributes'.

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).

On the other hand, in [17] the concept of resource is used with a twofold meaning: that of whatever might be identified by a URI, and that of anything that can be the subject of a discourse, such as cars, people etc. Furthermore, the concept of *Information resources* is defined as those resources which

[1]Note that [3] has been replaced by [4] but we decided to quote and discuss about it here because it is still used as reference from many web users

essential characteristics can be conveyed in a message. [17] also defines the principle of *opacity* of a URI, which promotes the independence between an identifier and the state of the identified resource.

Given the above, at least four possible interpretations of the intended meaning of the term "resource" emerge. Even though it is not our principal aim, it could be useful to establish what meaning is the most suitable in the web domain.

- **computational object:** if a resource is a computational object, something on which one can perform operations [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* [17], 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.

Based on this rationale, we give a formal description of the different meanings associated to the term 'resource', and contextualize those meanings in a formal pattern. The definition of "resource" in [4] corresponds to that of concept "Entity" used in IRE. We reserve the name 'web resource' to one meaning, which seems more intuitive from a commonsense viewpoint, even though our principal aim is conceptual, not terminological.

## III. The IRE ontology

As outlined above, the definitions of resource that can be found in literature [17], [3], [4] show ambiguity, making the issue of handling the identification of a web resource very problematic.

Our approach restricts the nature of the web resource to that of a computational object. This choice is motivated by the fact that a resource is something that has to be addressable, and things like cars and people are not addressable for their nature. Hence, it is wrong in principle to use the same mechanism of addressing for entities that have such different sorts. In [15] IRE has been formalized in first order logic, here we approach its formalization in OWL, and introduce two possible ontology design patterns for expressing IRE content.

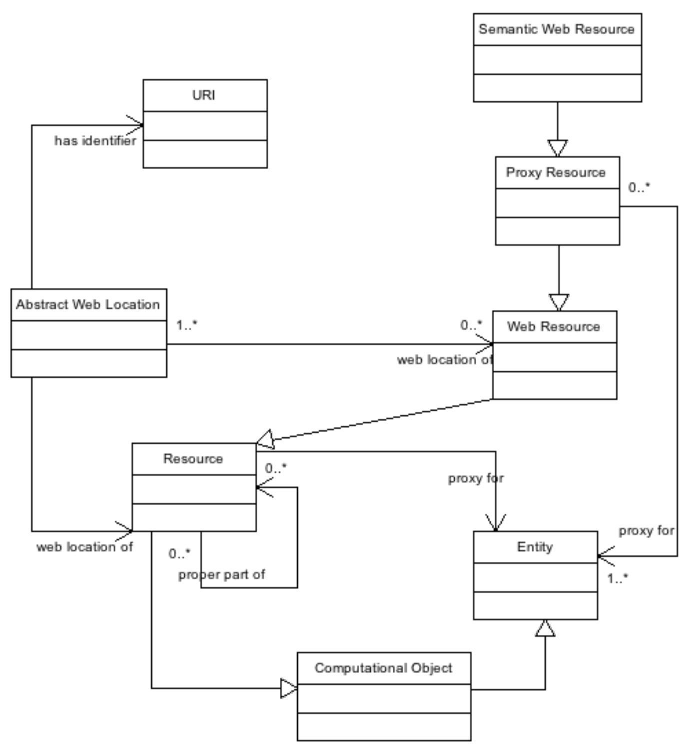


Fig. 1. The IRE conceptual pattern

Figure 1 depicts the IRE (Identifiers, Resources, and Entities) conceptual pattern. It includes the main concepts related to resources and their identification.

A *Resource* is a computational object that may have a web location. We define a *web resource* as a resource that is placed in one or more *abstract web locations*. In this sense to be a web resource is a particular state of a resource. This means that the identity of the resource is something that goes beyond its location. An *abstract web location* is a place (i.e., a point) in the combinatorial regions that are identified by the URI addressing mechanism (i.e., each URI identifies one and only one abstract web location) [2]. To this aim the datatype property *has identifier* is specified as functional. Although we cannot directly address an entity that is not a computational object, we need to be able to assert facts about it on the web. We can do this by using a web resource whose functions as a proxy for that entity (i.e., a *ProxyResource*). In order to make the model clearer, we give a prose description for each element that has been defined:

- **Abstract Web Location**: a point in the combinatorial regions identified by the URI metric.
- **Resource**: a computational object that can be composed of other resources (i.e., *proper part of*). It might have a location (i.e.,*Abstract Web Location*), the address of which is a URI. If the resource is a composed resource the identifier of its abstract location is also an approximate identifier for its parts.
- **Web Resource**: a resource that is associated with a web location, hence potentially accessible on the web.
- **Proxy Resource**: a web resource which functions as a proxy for whatever entity (e.g., a personal home page, a set of metadata describing a person).
- **Semantic Resource**: a web resource that realizes an information object through a codification in a formal language for the web (e.g., OWL [21]) which functions as a proxy for whatever entity.

The IRE pattern from Figure 1 suggests a categorization of the resources when their role is functioning as a proxy for an entity. Starting from this particular feature of a resource we noticed that the relation *proxy for* can be of four types, and that each of them can be treated in a different way. Each kind of *proxy for* relation may correspond to a different computational approach, or more specifically to a different operational semantics associated to the resolution of the proxy's URI. The four kinds of proxy relations can be described informally as follows:

- **Resolvable proxy for**: a relationship between a proxy resource and a web resource where the proxy resource allows the access to the web resource it is proxy for. For example, a 'href' in a HTML document is a resolvable proxy for the web resource it allows to access by clicking the corresponding link.
- **Approximate proxy for**: a relationship between a resource and a collection of entities, where the resource is about some of those entities. In this case the resource represents something within the collection, resulting in approximate reference.
- **Informal exact proxy for**: a relationship between a resource and exactly one entity, where the resource is about that entity. For example a document whose subject of discourse is the entity that the resource is proxy for.
- **Formal exact proxy for**: a relationship between a semantic resource and exactly one entity, where the semantic resource is about that entity, and describes it by means of a semantic structure. For example, a set of metadata associated to an individual from an OWL ontology.

[2]Notice that IRIs (Internationalized Resource Identifier) [19] are supposed to replace URIs in next future. Given this, IRIs involvement in the IRE pattern is the same that URIs have.

We remark that the relations described above are mappable to already existing or proposed concrete solutions. As a proof of concept of this claim, let's consider the skos property `skos:subjectIndicator` [27] and the topic maps element `subjectIndicatorRef` [24]. The web resource that is the value of such properties would function as a proxy resource for an entity, and this means that the two properties are either approximate proxies or informal exact proxies. Although this can be a way of identifying the entity of interest, the interpretation of the content of the web resource remains a responsibility of a human agent. As a matter of fact, when expressed informally, there is currently no way to automatically understand the meaning of the content of a web resource, with a decent precision. The situation is slightly different if the web resource is a semantic resource. In that case it functions as a formal exact proxy for an entity, and it is possible to enable a machine to derive the identity of the entity of interest. For example, this is the case of a set of metadata asserting facts about an individual of a given web ontology, which a software agent could be able to perform some inferences on.

### *A. An example: IRE pattern and Topic Maps*

In order to give an example of the applicability of this approach also to techniques other than RDF, we informally describe how to classify the Topic Maps concepts related to identity issues in terms of IRE. Based on [20]:

> "..subject - anything whatsoever, regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever..".

The Topic Maps concept of "subject" is equivalent to that of "Entity". More precisely, the terms "subject" and "entity" are synonyms. From the same document [20] we have the following definition of topic:

> A topic is a symbol used within a topic map to represent one, and only one, subject, in order to allow statements to be made about the subject through the assignment of characteristics to the topic.

Based on IRE pattern a topic in a topic map is a *formal exact proxy* (i.e., a semantic resource) for a given entity. Again from [20] we have the following definitions about identification of entities:

> "A subject indicator is an information resource that is referred to from a topic map in an attempt to unambiguously identify the subject of a topic to a human being. Any information resource can become a subject indicator by being referred to as such

from within some topic map, whether or not it was intended by its publisher to be a subject indicator. A subject identifier is a locator that refers to a subject indicator. Topic maps contain only subject identifiers (and not the corresponding subject indicators)...." "...A subject locator is a locator that refers to the information resource that is the subject of a topic. The topic thus represents that particular information resource; i.e. the information resource is the subject of the topic... "

In terms of IRE pattern, a "subject indicator" is a web resource which acts as *approximate proxy for* a given entity. In order to identify an entity by a subject indicator, a human interpretation is needed, and it is not guaranteed that such identification is unambiguous. Furthermore, a subject identifier is an abstract web location (i.e. a URI), which locates a subject indicator on the Web. On the other hand, when the entity is addressed by a "subject locator", it can be considered a computational object available on the web, i.e., a web resource.

## IV. Designing IRE in OWL

The IRE (Identity of Resources and Entities) conceptual pattern has been specified in first order logic in [15]. In order to make it available on the SW, we have expressed it in the OWL(DL) language [30].

The following sections are organized as follows: in section IV-A we present some reused and specialized predicates from other ontologies, in section IV-B we describe how we engineered IRE in OWL showing two possible design approaches. The figures included in next sections are drawn by using UML notation and assuming an OWL profile. Specifically, UML classes map to OWL classes, UML associations map to OWL object properties, and UML class attributes map to datatype properties. We use a DL syntax notation for expressing axioms, following [32].

### A. Imported predicates in IRE

The IRE pattern specializes the DOLCE reference ontology [9], [10], and some of its modular extensions, namely Spatial Relations, DnS with Information Objects, and Ontology Design Ontology (ODO) modules. All modules have been developed within EU projects Metokis [25], [2], WonderWeb [22], [12], and NeOn [28], and (including IRE ontology) are available in OWL at http://www.loa-cnr.it/ontologies/. For the sake of readability we abbreviate the path "`http://www.loa-cnr.it/ontologies`" with "`http://[*]`" while the following prefixes are used in the following definitions that correspond to the associated namespaces:

$$\texttt{xsd} = \texttt{http://www.w3.org/2001/XMLSchema\#}$$
$$\texttt{dol} = \texttt{http://[}\star\texttt{]/DOLCE-Lite.owl\#}$$
$$\texttt{edns} = \texttt{http://[}\star\texttt{]/ExtendedDnS.owl\#}$$
$$\texttt{inf} = \texttt{http://[}\star\texttt{]/InformationObjects.owl\#}$$
$$\texttt{od} = \texttt{http://[}\star\texttt{]/OD/OntologyDesign.owl\#}$$

The IRE ontology specializes or reuses the following predicates. For a complete axiomatization see the indicated URLs and [22]. From DOLCE [10] IRE imports:

$$\texttt{dol:particular, dol:social-object,}$$
$$\texttt{dol:region, dol:abstract-region,}$$
$$\texttt{dol:time-interval, dol:proper-part-of}$$
$$\texttt{dol:social-object} \sqsubseteq \texttt{dol:particular}$$
$$\texttt{dol:region} \sqsubseteq \texttt{dol:particular}$$
$$\texttt{dol:abstract-region} \sqsubseteq \texttt{dol:region}$$
$$\texttt{dol:time-interval} \sqsubseteq \texttt{dol:region}$$

DOLCE ontology makes basic distinctions between objects, events, and abstract entities. While objects and events like computers and software crashes are spatio-temporally localized, abstract entities like sets and value spaces have no space-time (they are purely formal entities). DOLCE also axiomatizes basic relations such as *part* and *location-of*.

From the DnS and Information Objects modules [34] the imports are:

$$\texttt{edns:situation, edns:information-object,}$$
$$\texttt{edns:information-realization,}$$
$$\texttt{edns:formal-language, edns:method,}$$
$$\texttt{edns:realizes, edns:about,}$$
$$\texttt{edns:ordered-by, edns:involves.}$$

$$\texttt{edns:information-object} \sqsubseteq \texttt{edns:social-object}$$

$$\texttt{edns:information-realization} \equiv \texttt{dol:particular} \sqcap \exists\, \texttt{edns:realizes.edns:information-object}$$

$$\texttt{edns:method} \sqsubseteq \texttt{edns:social-object}$$

DnS and Information Objects ontologies make basic distinctions between 'descriptive' and 'ground' entities, where the descriptive entities include social objects, like the 'student' or 'professor' roles, the 'being active' task, juridical persons, methods, and also information objects like the text of this paper. Descriptive entities have a lifecycle differently from pure information, which is an abstract entity.

Information objects have a core conceptual pattern, by which they are `edns:ordered-by` a representation language, they are `edns:realized-by` physical information realizations (physical objects, events, etc.), and can be `edns:about` any other entity. Situations are reifications of states of affairs i.e., reifications of n-ary relationships.

From the ODO module [35] IRE imports:

$$\texttt{od:computational-object}$$
$$\texttt{od:computational-object} \sqsubseteq \texttt{edns:information-realization}$$

The OD ontology specializes the Information Objects ontology in order to build a conceptual schema for digital and analog content to be exchanged during collaborative activities for ontology design. The concept `od:computational-object` specializes `edns:information-realization` for the computational world. Any physical document, electronic service, file, application, etc. are considered here computational objects.

*B. OWL IRE*

IRE express concepts about "referencing" and "accessing" of a web resource, according to [16].

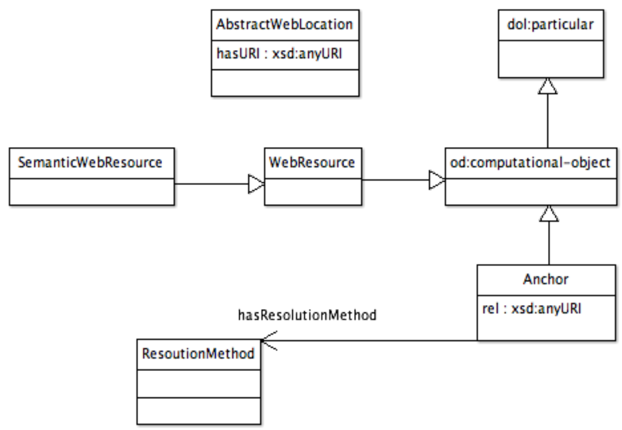


Fig. 2. The IRE taxonomy

Figure 2 shows the OWL classes that we have defined in order to express the foundations of IRE. Specifically, a `WebResource` is a computational object that have at least one assigned web location i.e., `AbstractWebLocation`. An `AbstractWebLocation` is an abstract region i.e., `dol:abstract-region`, a place in the combinatorial space identified by URIs. A `SemanticWebResouce` is a web resource that is represented i.e., `dol:realized-by` with a formal language for the SW e.g., OWL. The class `Anchor` represents computational objects that are embedded in web resource and that may allow access to web resources through a `ResolutionMethod`, e.g., an HTML link.

The core of IRE theory is in the proxy relations between resources and entities of the world. Those relations are temporally indexed i.e. a resource is a proxy for a certain entity at some time. We identified two possible design patterns to represent that in OWL: the first treats temporal aspects explicitly by reifying the relation with the n-ary relation pattern [29], [11], while the second treats temporal aspects as implicitly assumed by binary relations. In this section we show these two ontology design patterns for IRE.

*C. IRE with binary relations*

Figure 3 shows IRE in OWL. The figure depicts the taxonomy introduced in IV-B, and the object properties expressing the *proxy for* relation kinds. In this case the time aspect

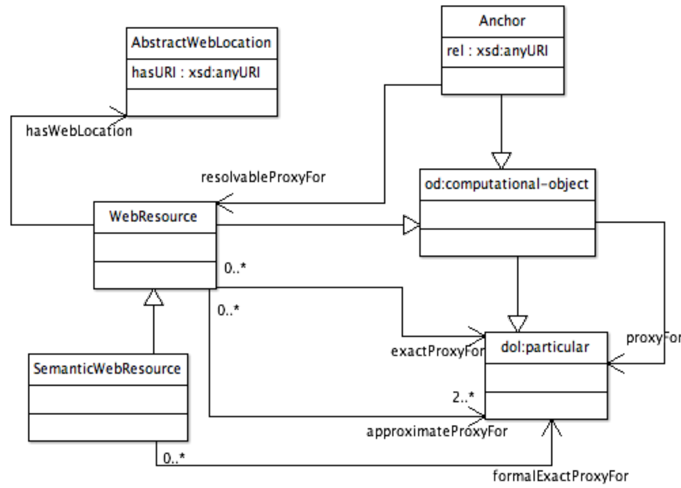


Fig. 3. IRE with binary relations

is implicit. In particular we introduce the following object properties:

- **ApproximateProxyFor**: is an object property having `WebResource` as its domain, and `dol:particular` i.e., any entity, as its range. We define a cardinality restriction on `WebResource` for this object property. More specifically, for a web resource to be an approximate proxy for, it has to be related with at least two `dol:particulars` by means of the `approximateProxyFor` object property
- **ExactProxyFor** is an object property having `WebResource` as domain and `dol:particular` i.e., any entity as range. We define a cardinality restriction on `WebResource` for this object property: a web resource can be an exact proxy for only one `dol:particular`
- **FormalExactProxyFor** is a subproperty of `exactProxyFor`, where the domain is restricted to `SemanticWebResource`
- **ResolvableProxyFor** is an object property where the domain is `Anchor`, and excludes `WebResource`. `Anchor` has a universal restriction on this object property that restricts the range to `WebResource`, and a cardinality restriction "equal to 1": an anchor is a resolvable proxy for one and only one web resource
- **ProxyFor** is an object property having `od:computational-object` as domain and `dol:particular` as range. All previously defined object properties are subproperties of this one.

*D. IRE with situations*

In this section we show how to represent IRE by means of `edns:situation` [14]. Situations (i.e., instances of `edns:situation`) are reifications of states of affairs i.e., reifications of n-ary relationships. This approach is compliant to the n-ary relation pattern suggested by [29], for which it provides an explicit vocabulary and general axioms. Situations allow us to explicitly treat temporal indexing. Referring to

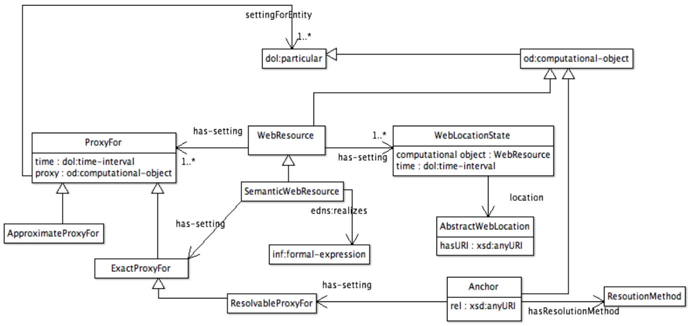


Fig. 4. IRE with n-ary relations

section III, the *proxy for* relation is intended as the "situation" for a web resource to "being a proxy for some entity at a given time", while the addressing of a web resource, which potentially allows its accessing, is intended as the "situation" for a web resource of "having a web location at a given time". Figure 4 depicts the IRE design pattern in terms of this engineering approach. More specifically, we define the following relation classes:

- **WebLocationState**: is a situation of a web resource that is associated to a point in the combinatorial regions identified by the URI metric i.e., a web location, at a given time.
- **ProxyFor**: a situation of a web resource which functions as a proxy for whatever entity (e.g., a personal home page, a set of metadata describing a person), at a given time.

The kinds of proxy situations can be informally described as follows:

- **ApproximateProxyFor**: a relationship between a resource and more than one entity at a given time, where the resource is about those entities. In this case the resource represents all the entities approximately.
- **ExactProxyFor**: a relationship between a semantic resource and one entity at a given time, where the semantic resource is about only that entity, and describes it through a semantic structure. For example, a set of metadata associated to an individual of an OWL ontology.
- **ResolvableProxyFor**: a relationship between an anchor and a web resource at a given time, where the anchor might allow the access to the web resource it is proxy for.

## V. Conclusion

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. There is something rather ambitious that the web is supposed to allow than just referencing web resources; that is, referencing things in general. On the other hand, that goal requires a software agent in order to identify a resource in an unambiguous way, and to perform the appropriate operations on it.

We have singled out some ambiguities that could prevent successful solutions to the web identity crisis., and proposed a way to clarify and formalize the different meanings of resources in a unique modelling framework. Finally, we have suggested some extensions to the model that can help classifying syntactic and operational solutions, and verifying their completeness and consistency.

Currently, URIs are used as the uniform mechanism for identifying heterogeneous entities, e.g., documents, metadata, ontologies, abstract concepts, physical things, events. But there is no clear categorization of which are the possible ways to identify and reference those entities on the web. We support the use of a formal model for the categorization of entities that can be referenced on the web. To this aim, we have defined IRE (Identifiers, Resources, and Entities), a conceptual pattern based on an ontology of Information Objects [2], built on top of DOLCE and on some of its modular extensions [23], [13], [12], [22]. In [15] we formalized IRE with first order logic, while here we have proposed an OWL ontology based on those formal definitions. IRE describes "referencing" and "accessing" of a web resource according to [16]. We have shown in this paper that there are two ways of engineering IRE in OWL. The first provides an ontology pattern based on binary relationships, the way OWL natively allows to express relations, and the second provides a pattern based on reification of n-ary relations according to [29], [14]. The latter pattern allows us to also express temporal indexing.

## 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.

## 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/dbooth-names_clean.htm, 2003.

[7] K. Clark. Identity Crisis. http://www.xml.com/pub/a/2002/09/11/deviant.html, 2002.

[8] del.icio.us. http://del.icio.us/.

[9] DOLCE - Project Home Page. http://dolce.semanticweb.org.

[10] DOLCE Ontology. http://www.loa-cnr.it/ontologies/DOLCE-Lite.owl.

[11] A. Gangemi. Ontology Design Patterns for Semantic Web Content. In *M. Musen et al. (eds.): Proceedings of the Fourth International Semantic Web Conference*, Galway, Ireland, 2005. Springer.

[12] 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, 2005.

[13] A. Gangemi, N. Guarino, C. Masolo, and A. Oltramari. Sweetening wordnet with dolce. *AI Mag.*, 24(3):13–24, 2003.

[14] 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.

[15] 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.

[16] 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.

[17] I. Jacobs and N. Walsh. Architecture of the World Wide Web. Technical report, W3C, 2004. http://www.w3.org/TR/webarch.

[18] 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.

[19] 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.

[20] G. L. M. and M. G. ISO/IEC 13250-2: Topic Maps – Data Model, 2005. http://www.isotopicmaps.org/sam/sam-model/.

[21] M. K. Smith, C. Welty, and D. L. McGuinnes. OWL Web Ontology Language Guide. Technical report, W3C, 2004. http://www.w3.org/TR/2004/REC-owl-guide-20040210/.

[22] 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.

[23] C. Masolo, L. Vieu, E. Bottazzi, C. Catenacci, R. Ferrario, A. Gangemi, and N. Guarino. Social roles and their descriptions. In D. Dubois, C. A. Welty, and M.-A. Williams, editors, *KR*, pages 267–277. AAAI Press, 2004.

[24] Members of the Topic Maps.Org Authoring Group. Xml topic maps (xtm 1.0). Technical report, TopicMaps.org, 2001. http://www.topicmaps.org/xtm/1.0/.

[25] Metokis Project. http://metokis.salzburgresearch.at.

[26] A. Miles. Working Around the Identity Crisis. http://esw.w3.org/topic/SkosDev/IdentityCrisis, 2005.

[27] A. Miles and D. Brickley. SKOS Core Vocabulary Specification. Technical report, World Wide Web Consortium (W3C), November 2005. http://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20051102/.

[28] NeOn EU project home page. http://neon-project.org.

[29] Noy N. and Rector A. Defining N-ary Relations on the Semantic Web: Use With Individuals. Technical report, W3C, 2005. http://www.w3.org/TR/swbp-n-aryRelations/ (2004).

[30] OWL Web Ontology Language Family of Specifications. http://www.w3.org/2004/OWL, 2004.

[31] 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.

[32] P. Patel-Schneider. Working Version (Draft): Description-Logic Knowledge Representation System Specification from the KRSS Group of the ARPA Knowledge Sharing Effort.

[33] 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).

[34] The Description and Situation ontology. http://www.loa-cnr.it/ontologies/ExtendedDnS.owl.

[35] The Ontology Design ontology. http://www.loa-cnr.it/ontologies/OD/OntologyDesign.owl.