A Conceptual Model for a Web-scale Entity Name System*

Paolo Bouquet[1], Themis Palpanas[1], Heiko Stoermer[1], and Massimiliano Vignolo[1]

University of Trento, Trento, Italy
{bouquet, themis, stoermer, vignolo}@disi.unitn.it

Abstract. The problem of identity and reference is receiving increasing attention in the (semantic) web community and is emerging as one of the key features which distinguish traditional knowledge representation from knowledge representation *on the web* with respect to data interlinking and knowledge integration on a large scale. As part of this debate, the OKKAM project proposed the creation of an Entity Name System which provides **rigid identifiers**, named `OKKAMid`s, for any type of concrete and particular entities, and links `OKKAMid`s to existing identifiers which have been created elsewhere for the same entity. The introduction of these identifiers raises some practical and conceptual concerns. In this paper we address them by extending two proposed ontologies (IRE and IRW) to accomodate the notion of `OKKAMid`, describe their formal properties, illustrate why they may play an important role in the construction of the Semantic Web and discuss how they can be integrated with other approaches for mapping URIs onto each others.

1 Introduction

One of the most ambitious visions of the Semantic Web is to create an open, decentralized space for sharing and combining knowledge, like the web did for hypertexts. In a note from 1998, Tim Berners-Lee described this vision as follows:

> Knowledge representation is a field which currently seems to have the reputation of being initially interesting, but which did not seem to shake the world to the extent that some of its proponents hoped. It made sense but was of limited use on a small scale, but never made it to the large scale. This is exactly the state which the hypertext field was in before the Web [...]. The Semantic Web is what we will get if we perform the same globalization process to Knowledge Representation that the Web initially did to Hypertext.

* This work is partially supported by the FP7 EU Large-scale Integrating Project **OKKAM – Enabling a Web of Entities** (contract no. 215032). For more details, visit `http://www.okkam.org/`. We are grateful to Stefano Bocconi for his helpful comments on an early draft of this paper.

As a contribution to this vision, the EU-funded OKKAM project[1] has started the design and development of a so-called *Entity Name System* (ENS) [2], a web-scale, open service which supports users and applications in the systematic reuse of global and stable identifiers for entities which are named and described in distributed collections of data and content. The main goal of the project is to consolidate the information space of the web of data by reducing the number of URIs which are used for referring to the same entity in different datasets, making the integration and fusion of RDF data and content much easier and faster.

However, the very idea of an ENS has raised some theoretical and practical concerns which need to be addressed before such a service may be adopted by the community at large. This paper aims at solving some of these issues by proposing a conceptual model in which we draw a clear distinction between the meaning and the role of the HTTP URIs which are used as identifiers by the ENS (`OKKAMids`) and the standard RDF URIs which are used in RDF datasets. The underlying intuition is that `OKKAMids` provide a form of *direct reference* to real-world entities, whereas RDF URIs provide a description-based reference to entities (which means that different RDF URIs may be needed to publish different representations of the "same" entity). We will discuss why, in our view, both types of URIs are needed for building the Semantic Web, and show how they can easily cohexist in practical methods for publishing RDF data.

The OKKAM conceptual model (OCM) distinguishes these two views clearly at a foundational level by providing a formal definition of `OKKAMids` and RDF URIs. The underlying intuition is that these two types of identifiers need not be perceived as mutually exclusive. On the contrary, they serve different (complementary) purposes, and therefore they should be used together to bring knowledge representation on the web to its full potential. In section 2 we clarify some important conceptual issues concerning the fundamental relations between URIs, the Semantic Web and the real world. Section 3 introduces the OCM model formalized in first order logic. In section 4 we give justification to the distinction between the linguistic function of `OKKAMids` as rigid and direct identifiers and the linguistic function of RDF URIs as identifiers based on descriptions. In section 5 we discuss some practical consequences of the difference between the two linguistic functions.

2 Identity and Reference in the Semantic Web

Similarly to what happened for the hypertext web with URLs and `HREF` references, one of the key factors for realizing the vision we quoted above is to enable a global and uniform naming space for the "entities" which are named in a piece of data and content, so that people and machines can always refer unambiguously to whatever entity they need to name. The proposal is to exploit the Web architecture and use URIs (more precisely HTTP URIs[2]) as such naming mech-

[1] `http://www.okkam.org/`
[2] `http://www.w3.org/DesignIssues/LinkedData.html`

anism. Indeed, an important feature of URIs is that the same URI is always dereferenced in the same way, no matter where it appears.

The question is whether the idea works in the other direction as well: does it make sense to request that **the same entity is always referred to by the same (HTTP) URI**? The answer to this question is not so straightforward. The most important objection is conceptual, and has to do with the following issue: does a URI make direct reference to an entity, or is it "equivalent" to a *description* of that entity? The current trend is to view URIs as basically equivalent to descriptions, namely the sets of RDf statements which we obtain when the URIs are dereferenced. In this view, the two identifiers `http://www.w3.org/People/Berners-Lee/card#i` and `http://dblp.l3s.de/d2r/resource/authors/Tim_Berners-Lee` provide different (and potentially inconsistent) information about a person (Tim Berners-Lee), and since these two descriptions should not be confused, it makes sense to have two different URIs for the same person. From a different perspective, however, there are researchers who stress that the above issue can be understood only if one assumes that the two URIs are indeed about the same entity. Therefore it must be the case that a name is somehow "attached" directly to an entity without the mediation of any particular description. In short, in the first view *reference is essentially mediated by description*, and the latter is more fundamental than the former; in the second view, reference is not necessarily mediated by any particular description, as *reference is a primitive and direct relation* between a real world entity and its identifier.[3]

The discussion is far from being a mere academic debate on the theory of reference. Indeed, it has a deep practical impact on how people are developing the Semantic Web, and in particular the so-called Web of Data. The first view offers a very elegant DNS-based method for publishing and accessing sets of statements about an entity (and to evaluate their level of trust). In addition, it provides the technical basis for enabling the web-style exploration of semantic data via RDF browsers, like the Tabulator[4], Disco[5] or the OpenLink RDF browser.[6] The second enables very powerful forms of URI-based data retrieval (e.g. through semantic search engines, like Sigma[7], Sindice[8] or Falcons[9]), makes semantic mashups very easy and straightforward, and enables web-scale distributed reasoning. In order to unleash the full power of the technologies that have been developed by the community, we propose a formal model that distinguishes both views at a foundational level.

[3] See [8] for a philosophical discussion of this thesis. In [7], the concept of direct reference is presented in a slightly different way as the idea that "on the Web, the resource identified by a URI is whatever was intended by the owner". We'll explain later on why the two definitions have different consequences on our argument.

[4] `http://dig.csail.mit.edu/2007/tab/`

[5] `http://www4.wiwiss.fu-berlin.de/bizer/ng4j/disco/`

[6] `http://demo.openlinksw.com/DAV/JS/rdfbrowser/index.html`

[7] `http://sig.ma/`

[8] `http://sindice.com/`

[9] `http://iws.seu.edu.cn/services/falcons/objectsearch/index.jsp`

3 The OKKAM Conceptual Model

3.1 Basic Concepts

The architecture of the web forces a subdivision of the universe into things that exist or might exist on the web and things that cannot exist on the web. Given that existence on the web amounts to accessibility on the web through dereferencing URIs, most real world entities, indeed all those that are not computational objects, are things that cannot exist on the web because we cannot access them directly but only their representations. Accordingly, the OCM draws the distinction between computational objects and non-web resources.

A computational object is defined as (i) the physical realization of an information object and (ii) something that can participate in a computational process that ensures the resolution of a URI (see [5]). All digital documents, databases, electronic services, files, applications, are computational objects. Once a computational object is assigned a URI that gives it a location on the web, and thereby makes it a web-accessible entity, the computational object becomes a web resource. Non-web resources are all those entities that are not computational objects. The class of OKKAM entity is a subclass of the class of non-web resources, more precisely the class of all particular and concrete entities (events included). This means that classes, properties and abstract concepts do not count as OKKAM entities. It is worth noting that the class of non-web resources is not the complement of the class of web resources. Indeed, a computational object that does not possess a location on the web is neither a web resource nor a non-web resource. Figure 1 shows the relation of inclusion between the above classes of entities. One objective of the Semantic Web is to allow people to talk about

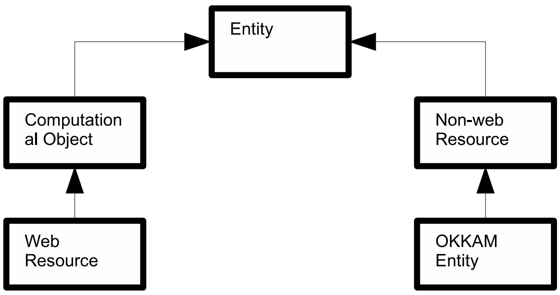

Fig. 1. Relations between classes of entities

things that do not exist on the web and nevertheless to talk about them by using URIs. There is a tension between the objective of the Semantic Web and the idea of using URIs for talking about entities that do not exist on the web. Indeed, from a linguistic point of view, URIs work as descriptions. A URI describes a certain entity as that entity that can be accessed at a certain location on the web. For example, the URI `http://www.dit.unitn.it/~bouquet/` denotes by description a web resource, i.e. Paolo Bouquet's homepage. The linguistic function of that URI is the same as the linguistic function of the definite description *the web resource accessible by resolving the URI* `http://www.dit.unitn.it/~bouquet/`. The idea that the linguistic function of URIs is that of definite descriptions denoting web resources gives rise to two main difficulties. The first is that only computational objects can be accessed on the web. Moreover, while web *locations* persist identical over time, the web resources located there can change. For example, Paolo Bouquet's homepage might change over time due to the updating of his publications, teaching activities, academic appointments etc. The second difficulty, then, is that URIs, as descriptions denoting web resources, are not *rigid designators*,[10] because there is no guarantee that by employing the same URI we will always be talking about the same entity, while in talking about non-web resources we would like to use rigid designators that are guaranteed not to change their referents. Therefore, the very idea of employing URIs as names for entities seems to require the distinction between two separate linguistic functions of URIs. One is the function of denoting by description and the other is the function of naming by reference [10]. The Semantic Web needs URIs that can be used as names that *denote* web resources by description and URIs that can be used as names that *refer* to non-web resources.

There are two ways of solving such linguistic ambiguity. One is to use different URIs according to the linguistic function performed, the other is to use the context of use for disambiguating the two linguistic functions. According to the first solution, RDF URIs are used to make reference to non-web resources. RDF URIs are so configured that when they are dereferenced the web server returns a 303 redirection code redirecting to another URI that might resolve into a web resource or start a further process of redirection. This view makes the distinction between the RDF URIs' linguistic function of making reference to non-web resources and the URIs' linguistic function of denoting web resources by description. It tells that being denoted by description consists in being referenced by a URI, while being referred to consists in being named by a RDF URI which redirects to another URI. Thus, RDF URIs working as names referring to non-web resources are distinguished from URIs working as definite descriptions denoting web resources.

The aim of OCM presented in this paper is to enlarge that way of solving the linguistic ambiguity of URIs to a global scale – though restricted to concrete and particular non-web resources. The main idea underlying OKKAM is that `OKKAMids` are fundamental tools for enlarging and integrating the use of RDF URIs as identifiers of non-web resources. As is represented in IRW, some RDF URIs can be treated as identifiers of non-web resources. In fact, the `irw:identifies` property, i.e. the linguistic function of reference defined in IRW over RDF URIs and non-web resources, is functional. The problem with this picture is that the meachanism by which RDF URIs make reference to entities is still based on description. RDF URIs identify their referents as those entities

[10] An identifier is a so-called rigid designator if in all possible worlds it denotes the same object. See e.g. [8] for Kripke's introduction to this notion.

that satisfy the information conveyed in the web resources that are referenced by the URIs to which the RDF URIs redirect. It follows that we might have – and indeed this is already the case with many RDF URIs – different RDF URIs expressing different descriptions of the same entity in different contexts. `OKKAMids` allow us to make it explicit in the web community that two or more RDF URIs identify one and the same entity, though from different points of view. Figure 2 shows the URIs' linguistic function – defined in OCM – of denoting web resources, the RDF URIs' linguistic function – defined in IRW – of designating non-web resources by description and the `OKKAMids`' linguistic function – defined in OCM – of referring *rigidly* and *directly* to OKKAM entities.

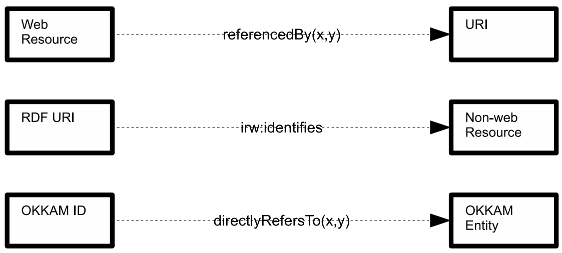

Fig. 2. URI's linguistic functions

3.2 Formalization

The conceptual model we present is built on top of the ontologies for identity and reference on the Web (IRE, IRW) which have been presented in [6, 5, 10, 7]. OCM adds new concepts which are specifically related to OKKAM in order to model the relations between OKKAM, the real-world and the web. IRE specializes the DOLCE ontology and some of its modular extensions, namely Spatial Relations, DnS with Information Objects, and Knowledge Content Objects (KCO) and Ontology Design Ontology (ODO) modules. Figure 3 shows the relation of inclusion between the concepts defined below and the relations of **`directlyRefersTo`**(x, y) defined over `OKKAMids` and OKKAM entities, **`referencedBy`**(x, y) defined over web resources and URIs, and **`irw:identifies`**(x, y) defined over RDF URIs and non-web resources. In the following we present the definitions and axioms that form the OKKAM Conceptual Model[11]:

$$\texttt{OkkamEntity}(x) =_{def} \texttt{Entity}(x) \wedge \\ \texttt{Particular}(x) \wedge \texttt{dol}:\texttt{Concrete}(x) \wedge \neg\texttt{od}:\texttt{ComputationalObject}(x) \quad (1)$$

[11] The OWL specification of the OCM can be found at `http://models.okkam.org/OKKAM-conceptual_model.owl`. The name space for the OCM objects is `http://models.okkam.org/OKKAM-conceptual_model.owl#`

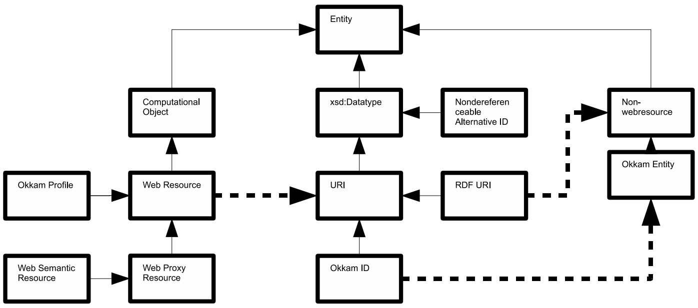

Fig. 3. Map of OCM concepts and relations

Definition 1 states that only concrete and particular entities are OKKAM entities, i.e. entities apt for being assigned `OKKAMids`. Note that `Entity`(x) and `Particular`(x) are implicit in DOLCE. Here we use them for the sake of exposition.

$$\texttt{ire}:\texttt{URI}(x) \rightarrow \texttt{xsd}:\texttt{Datatype}(x) \tag{2}$$

Axiom 2 gives a characterization of the concept of URI in terms of XSD datatype as is usual practice.

$$\begin{aligned}&\texttt{OkkamID}(x) =_{def} \texttt{ire}:\texttt{URI}(x)\wedge\\ &(\texttt{Pattern}(x) =' http://www.okkam.org/ens/id < UUID >')\end{aligned} \tag{3}$$

$$\begin{aligned}&\texttt{URIokkamProfile}(x) =_{def} \texttt{ire}:\texttt{URI}(\texttt{x})\wedge\\ &(\texttt{Pattern}(x) =' http://www.okkam.org/ens/id < UUID > /about.rdf')\end{aligned} \tag{4}$$

Definitions 3 and 4 define the concepts of `OKKAMid` and URI for OKKAM profiles by specifying their patterns.[12] The OKKAM web server returns a 303 redirection code response for a request of an `OKKAMid` and gives the URI for the OKKAM profile as the new location of the document.

$$\texttt{ire}:\texttt{hasIdentifier}(x,y) \rightarrow \texttt{dol}:\texttt{Region}(x) \wedge \texttt{xsd}:\texttt{Datatype}(y) \tag{5}$$

Axiom 5 characterizes the relation of having an identifier between regions and datatype identifiers.

$$\begin{aligned}&\texttt{ire}:\texttt{AbstractWebLocation}(x) =_{def}\\ &\texttt{dol}:\texttt{AbstractRegion}(x) \wedge \exists y(\texttt{ire}:\texttt{URI}(y) \wedge \texttt{ire}:\texttt{hasIdentifier}(x,y)\wedge\\ &\neg\exists z(\texttt{ire}:\texttt{URI}(z) \wedge y \neq z \wedge \texttt{ire}:\texttt{hasIdentifier}(x,z)))\end{aligned} \tag{6}$$

[12] $<$ `UUID` $>$ as defined in `http://java.sun.com/j2se/1.5.0/docs/api/java/util/UUID.html`

$$\begin{aligned}&\texttt{ire}:\texttt{AbstractWebLocation}(x) \rightarrow\\ \neg\exists y,z(&\texttt{ire}:\texttt{URI}(y) \wedge \texttt{ire}:\texttt{AbstractWebLocation}(z) \wedge x \neq z\\ &\wedge \texttt{ire}:\texttt{hasIdentifier}(x,y) \wedge \texttt{ire}:\texttt{hasIdentifier}(z,y))\end{aligned} \tag{7}$$

Definition 6. and axiom 7. state that an abstract web location is a point in the combinatorial regions identified by the URI metric such that it is identified by at most one URI and cannot be identified by any other URI already employed to identify another abstract web location.

$$\begin{aligned}&\texttt{ire}:\texttt{webLocationOf}(x,y,t) =_{def}\\ &\texttt{dol}:\texttt{eAbstractLocationOf}(x,y,t) \wedge \texttt{ire}:\texttt{AbstractWebLocation}(x)\\ &\wedge \texttt{od}:\texttt{ComputationalObject}(y) \wedge \texttt{dol}:\texttt{Time}(t)\end{aligned} \tag{8}$$

Definition 8 defines a relation between abstract web locations, computational objects and times and specializes the relation `dol:eAbstractLocationOf`(x,y) imported from the Spatial Relations module.

$$\texttt{ire}:\texttt{ResolutionMethod}(x) \rightarrow \texttt{edns}:\texttt{Method}(x) \tag{9}$$

$$\texttt{RedirectionMethod}(x) \rightarrow \texttt{edns}:\texttt{Method}(x) \tag{10}$$

$$\begin{aligned}&\texttt{ire}:\texttt{WebResource}(x) =_{def}\\ &\exists m(\texttt{ire}:\texttt{ResolutionMethod}(m) \wedge \texttt{edns}:\texttt{involves}(m,x))\\ &\wedge \exists y,t(\texttt{ire}:\texttt{webLocationOf}(y,x,t))\end{aligned} \tag{11}$$

Axioms 9 and 10 and definition 11 state that web resources are computational objects accessible on the web by dereferencing a URI. We add `time`(x) – which might be an interval of time – because there can be a computational object that lacks an abstract web location at time t, but gets one at time t' or viceversa a computational object that has an abstract web location at t and then it loses it at t'.

$$\texttt{NonWebResource}(x) =_{def} \texttt{Entity}(x) \wedge \neg\texttt{od}:\texttt{ComputationalObject}(x) \tag{12}$$

Definition 12 defines the concept of non-web resource. `OkkamEntity`(x) is a subclass of `NonWebResources`(x).

$$\begin{aligned}&\texttt{OkkamProfile}(x) =_{def}\\ &\exists u,y,t(\texttt{URIokkamProfile}(y) \wedge \texttt{ire}:\texttt{hasIdentifier}(u,y)\wedge\\ &\texttt{ire}:\texttt{webLocationOf}(u,x,t))\end{aligned} \tag{13}$$

Definition 13 defines the notion of OKKAM profile. An OKKAM profile is a web resource that is accessible on the web by dereferencing an `OKKAMid`.

$$\begin{aligned}&\texttt{referencedBy}(x,y) =_{def}\\ &\texttt{ire}:\texttt{WebResource}(x) \wedge \texttt{ire}:\texttt{URI}(y)\wedge\\ &\exists m(\texttt{ire}:\texttt{ResolutionMethod}(m) \wedge \texttt{edns}:\texttt{involves}(m,x))\wedge\\ &\exists z,t(\texttt{ire}:\texttt{webLocationOf}(z,x,t) \wedge \texttt{ire}:\texttt{hasIdentifier}(z,y))\end{aligned} \tag{14}$$

Definition 14 defines the relation of being referenced over web resources and URIs. The property `referencedBy`(x,y) in OCM is different from the property `irw:isReferencedBy`(x, y) and equivalent to the inverse of the `irw:accesses`(x, y) property.

$$\begin{aligned}&\texttt{redirectsTo}(x,y) =_{def}\\ &\texttt{ire}:\texttt{URI}(x) \wedge \texttt{ire}:\texttt{URI}(y) \wedge \exists m(\texttt{RedirectionMethod}(m)\wedge\\ &\texttt{edns}:\texttt{involves}(m,x) \wedge \texttt{edns}:\texttt{involves}(m,y))\wedge\\ &\exists z(\texttt{referencedBy}(z,y))\end{aligned} \tag{15}$$

Definition 15 defines the relation of redirection over URIs.

$$\begin{aligned}&\texttt{assignsTo}(x,y,z) =_{def}\\ &\texttt{OkkamProfile}(x) \wedge \texttt{OkkamID}(y) \wedge \texttt{OkkamEntity}(z) \wedge \exists t,s(\texttt{dol}:\texttt{Time}(t)\\ &\wedge\texttt{edns}:\texttt{InformationObject}(s) \wedge \texttt{edns}:\texttt{realizes}(x,s,t)\wedge\\ &\texttt{edns}:\texttt{about}(s,z,t) \wedge \exists u,w(\texttt{ire}:\texttt{URIokkamProfile}(w)\wedge\\ &\texttt{ire}:\texttt{hasIdentifier}(u,w) \wedge \texttt{ire}:\texttt{webLocationOf}(u,x,t)\wedge\\ &\texttt{RedirectsTo}(y,w)))\end{aligned} \tag{16}$$

Definition 16 captures the idea that an OKKAM profile does not describe an entity but is used to perform a baptism of an entity with an `OKKAMid`. (More on the idea of baptism in the following section).

$$\exists x,y(\texttt{assignsTo}(x,y,z) \wedge \texttt{assignsTo}(x,y,w)) \rightarrow z = w \tag{17}$$

Axiom 17 states that two different OKKAM entities cannot have the same `OKKAMid`.

$$\begin{aligned}&\texttt{directlyRefersTo}(x,y) =_{def}\\ &\texttt{OkkamID}(x) \wedge \texttt{OkkamEntity}(y) \wedge \exists z(\texttt{OkkamProfile}(z)\\ &\wedge\texttt{assignsTo}(z,x,y))\end{aligned} \tag{18}$$

Definition 18 defines the relation of direct reference over `OKKAMid`s and OKKAM entities. The `directlyRefersTo`(x, y) property in OCM is a functional property and aims to capture the Berners-Lee's *direct reference position*. The property `directlyRefersTo`(x, y) is distinct from the `irw: refersTo`(x, y) property, since the latter is not a functional property. The `irw:identifies` property, too, is a functional property defined over RDF URIs and non-web resources. However, `OKKAMid`s make *rigid* and *direct* reference to entities, whereas RDF URIs refer to entities through the mediation of a description.

$$\begin{aligned}&\texttt{Okkamised}(x) =_{def}\\ &\texttt{OkkamEntity}(x) \wedge \exists y(\texttt{OkkamID}(y) \wedge \texttt{directlyRefersTo}(y,x))\end{aligned} \tag{19}$$

Definition 19 defines the concept of having an `OKKAMid` and states that only Okkam entities can be assigned `OKKAMid`s.

$$\texttt{webProxyFor}(x,y,t) =_{def}$$

$$\texttt{ire}:\texttt{WebResource}(x) \wedge \neg\texttt{OkkamProfile}(x) \wedge \texttt{Entity}(y) \wedge \texttt{dol}:\texttt{Time}(t)$$
$$\wedge\exists z(\texttt{edns}:\texttt{InformationObject}(z) \wedge \texttt{edns}:\texttt{realizes}(x,z,t)\wedge$$
$$\texttt{edns}:\texttt{about}(z,y,t)) \tag{20}$$

Definition 20 states that in order for x to bear the `webProxyFor`(x, y, t) relation, x must be a resource that realizes an information object about the entity y at t and cannot be an OKKAM profile.

$$\texttt{WebProxyResource}(x) =_{def} \exists y,t(\texttt{webProxyFor}(x,y,t)) \tag{21}$$

Definition 21 states that a web proxy resource x is a web resource that stands in the webProxyFor relation to an entity y at time t.

$$\texttt{WebSemanticResource}(x) =_{def}$$
$$\exists y,t,z,w(\texttt{webProxyFor}(x,y,t) \wedge \texttt{edns}:\texttt{InfomationObject}(z)\wedge$$
$$\texttt{edns}:\texttt{FormalLanguage}(w) \wedge \texttt{edns}:\texttt{realizes}(x,z,t) \wedge \texttt{edns}:\texttt{about}(z,y,t)\wedge$$
$$\texttt{edns}:\texttt{orderedBy}(z,w)) \tag{22}$$

Definition 22 states that a web semantic resource is a web proxy resource that realizes an information object about an entity by a codification in a formal language for the web. Example: `http://dbpedia.org/page/Eiffel_Tour`

$$\texttt{RDFURI}(x) =_{def}$$
$$\texttt{ire}:\texttt{URI}(x) \wedge \exists w,y,u,t(\texttt{WebSemanticResource}(y) \wedge \texttt{ire}:\texttt{URI}(w)\wedge$$
$$\texttt{ire}:\texttt{hasIdentifier}(u,w) \wedge \texttt{ire}:\texttt{webLocationOf}(u,y,t)\wedge$$
$$\texttt{redirectsTo}(x,w)) \tag{23}$$

Definition 23 states that a RDF URI is a URI that redirects to the URI of a web semantic resource.[13]

$$\texttt{dereferenceableAlternativeIDOf}(x,y) =_{def}$$
$$\texttt{RDFURI}(x) \wedge \texttt{okkamised}(y) \wedge \exists w,z,u,t(\texttt{ire}:\texttt{URI}(w)\wedge$$
$$\texttt{ire}:\texttt{hasIdentifier}(u,w) \wedge \texttt{ire}:\texttt{webLocationOf}(u,z,t)\wedge$$
$$\texttt{webProxyFor}(z,y,t) \wedge \texttt{redirectsTo}(x,w)) \tag{24}$$

Definition 24 defines the relation of being a dereferenceable alternative ID of an OKKAM entity that has an `OKKAMid`. Dereferenceable alternative IDs of OKKAM entities are RDF URIs.

$$\texttt{corefer}(x,y) =_{def}$$
$$\texttt{RDFURI}(x) \wedge \texttt{OkkamID}(y) \wedge \exists z(\texttt{Okkamized}(z)\wedge$$
$$\texttt{directlyRefersTo}(y,z) \wedge \texttt{dereferenceableAlternativeIDOf}(x,z)) \tag{25}$$

[13] It must be noted that the content negotiation might ask for "text/html", so what is here presented as a RDF URI might redirect to a URI that retrieves a HTML page. Therefore, replacing the term "RDFURI" with another term like, say, "Linked-DataURI" might seem more appropriate. We leave such terminological question aside. Nothing conceptually important follows if one uses "LinkeDataURI", since the redirection is to only one representation per media type.

Definition 25 defines the relation of coreference between RDF URIs and `OKKAMids` that holds when the entity identified by an RDF URI is the same entity directly referred to by an `OKKAMid`. Of course, more than one RDF URI can bear the `corefer`(x,y) relation to the same `OKKAMid`.

$$\begin{gathered}\texttt{coidentify}(x, y) =_{def} \\ \texttt{RDFURI}(x) \wedge \texttt{RDFURI}(y) \wedge \exists z(\texttt{OkkamID}(z) \wedge \\ \texttt{corefer}(x, z) \wedge \texttt{corefer}(y, z))\end{gathered} \tag{26}$$

Definition 26 defines the relation of coidentification between RDF URIs. The raltion can be inferred from the fact that two RDF URIs are deferenceable alternative IDs of the same OKKAM entities.

4 OKKAM IDs and RDF URIs

One aspect of OCM above others deserves special clarification. The fact is that one might consider the objection that `OKKAMids` and RDF URIs are not really distinguished. Indeed, URIs of both types can be dereferenced and the act of dereferencing them triggers a process of redirection to URIs for web resources. `OKKAMids` redirect to URIs for OKKAM profiles, whereas RDF URIs redirect to URIs for other web resources. OKKAM profiles give information about the OKKAM entities referred to. Therefore, the mechanism of reference of `OKKAMids`, too, seems to be mediated by description. The objection, then, is that there is no structural and web architectural difference between `OKKAMids` and RDF URIs to the effect that the linguistic distinction between them looks arbitrary and unjustified. Why do `OKKAMids` make *rigid* and *direct* reference to OKKAM entities, whereas RDF URIs identify non-web resources by descriptions?

We reply to this objection by granting the indiscernibility of `OKKAMids` and RDF URIs from the structural and web architectural point of view. However, the ground of the distinction can be found elsewhere, namely in the purpose of using such URIs. Our reply is that the linguistic distinction between `OKKAMids` and RDF URIs has a pragmatic ground. The purpose of creating and using `OKKAMids` is to give the start to a linguistic practice by an act of baptism and by following acts of subscription to that linguistic practice. Such a practice is not assessed in terms of truth and falsity, which amounts to saying that the information contained in an OKKAM profile need not be true of the entity being assigned the `OKKAMid`. It is sufficient that the web community converges on that information in order to fix the referent of that `OKKAMid`.

To make the point clear it might be helpful to adapt a famous example (from [3]) in philosophy of language to our case. Imagine that Jane and John are enjoying a party. They give a look at a man holding a martini glass. For some reasons they are willing to assign a proper name to that man and agree on the following convention: *lets us call the man drinking martini "Jack"*. In order for their convention to be successful it is not necessary that the liquid in the glass be martini. However, even if the description *the man drinking martini* does not denote the man standing in front of Jane and John – because, say, the liquid in the glass is water – the act of baptism is successful provided that both Jane and John share the belief that that man is drinking martini, no matter how false that belief is. In fact, the description *the man drinking martini* is not used to express information about the man standing in front of Jane and John, but to fix the referent of the newly introduced name "Jack". Jane and John are not so much concerned as to whether that man is drinking martini or not, as to the fact that they both share that belief and use it to fix the referent of the name "Jack".

The use of the information conveyed in an OKKAM profile is the same as the use of the description *the man drinking martini* in the above scenario: it is not used to express information about an entity but to fix the referent of a name, i.e. an `OKKAMid`. Very likely, most of the information conveyed in an OKKAM profile will be true *de facto* of the entity to which that profile assigns the `OKKAMid`. That circumstance does not alter the fact that users need not endorse such information as true. To make the point clear, consider the above scenario again. Imagine Clark, too, is at the party and comes to know Jane and John's convention of naming the man standing in front of them "Jack" and that they believe that that man is drinking martini. Clark, however, knows that the liquid in the glass is water and not martini. Nevertheless, Clark can appeal to Jane and John's false belief to disambiguate utterances of the name "Jack", although Clark does not endorse such belief as true. For example, if Clark says "Jack is a computer scientist" and Jane or John replies "Jack who?", Clark might answer "the man who is drinking martini" to fix the referent of his utterance of the name "Jack" and to say of the man named "Jack" by Jane and John that he is a computer scientist.

The idea underlying our view is that OKKAM profiles are not about OKKAM entities in the same way as the information conveyed in other web resources is about non-web resources. More precisely, an OKKAM profile is not a description of an entity but constitutes the virtual context for the assignment of an `OKKAMid` to an entity. One should think of the assignment of an `OKKAMid` to an entity as a baptism that dubs that entity with a name. A baptism is a performative speech act. Performatives, unlike constatives, which are assessed in terms of truth or falsity, can only be assessed as felicitous or infelicitous (see Austin's felicity conditions [1]). A baptism is a speech act with its own felicity conditions. One of them is the existence and the salience of the entity being dubbed. No baptism can take place if there is no entity to be dubbed and if that entity is not cognitively available as the most salient to the persons who have the authority to make the baptism. An OKKAM profile serves exactly to make the entity to be dubbed salient, and its purpose is not to provide a description of that entity. On the other hand, accessing an OKKAM profile by dereferencing an `OKKAMid` amounts to the speech act of subscribing to the linguistic practice of using that `OKKAMid` as a name for a certain non-web resource. The creator of the OKKAM profile for an entity is the producer of that linguistic practice, whereas the users who access that OKKAM profile since its creation are the consumers of that linguistic practice(see [4] Ch. 11). It is not necessary that the information in the OKKAM profile be true of a non-web resource in order for the baptism to be successful. For the accomplishment of the baptism it is sufficient that the web community shares or converges on that (mis)information. Consider the following example, borrowed from [4]. Take the poet known to his contemporaries as "Homer" (or known by some name from which "Homer" descends); we think of the claim "Homer wrote the *Iliad*" as a substantial hypothesis about the authorship of the poem. But suppose the hypothesis is false. We might still use that piece of (mis)information to create an OKKAM profile assigning an `OKKAMid` to the *Iliad*, and saying that the *Iliad* was written by Homer. So long as a community converges on that piece of (mis)information, the baptism is felicitous.

OCM mirrors the semantic distinction between `OKKAMid`s and RDF URIs by the stipulation that the former redirect to URIs for OKKAM profiles and the latter to URIs for web proxy resources, and that OKKAM profiles are not web proxy resources. The justification of that distinction is not fully expressed by the definitions and the axioms in OCM. Indeed, the axioms by themselves simply stipulate that there is a linguistic difference between `OKKAMid`s and RDF URIs. Nevertheless, the distinction can be justified from pragmatic reflections on the purpose of using `OKKAMid`s and RDF URIs. RDF URIs are used to express and endorse information about entities, whereas `OKKAMid`s are used to fix the referent of RDF URIs within the whole web community and eventually to make it explicit that two or more different RDF URIs are different names of the same entity independently of the information retrievable by dereferencing those RDF URIs.

5 Conclusions

Practice shows that RDF URIs are commonly used for three different things:

1. Redirecting to a set of assertions *about* a non-web resource. As mentioned in the first part of this paper, dereferencing a RDF URI usually results in the retrieval of RDF triples describing non-web resources.
2. Linking from one set of assertions to another. Employing the `owl:sameAs` construct, a link can be established between one RDF URI and another. The semantics of this will be further addressed in this section.
3. Providing a surrogate/substitute/proxy for non-web resources. This is the typical case for the notion of "identifier for an individual" in a Description Logics knowledge base.

Cases (1) and (2) form a vital mechanism of the Linked Data approach. From our point of view, case (2) implies some very important semantics that have to be respected. First of all, there is a certain mismatch of the use of the `owl:sameAs` property in the Linked Data approach, and its intended semantics in the OWL specification [9]: collapsing all equivalent RDF nodes into a single one and thus joining the set of all axioms about these equivalent nodes onto the collapsed node, thus losing the ability to distinguish which nodes the statements were about in the first place, is the defined semantics of `owl:sameAs`. The actual use of this construct today however is one of linkage, i.e. the author of such a statement rather intends semantics of pointing to, or even endorsing, more axioms about the same real-world entity provided by another source. And indeed, losing the provenance of the axioms is not only undesirable, but also not commonly practiced. In Semantic Web applications, the `owl:sameAs` property is often directly translated to a hypertext link which the user can click to navigate to another set of assertions, as for example in the Tabulator application. Or the assertions are retrieved following case (1) and presented in an aggregated view, but preserving provenance, in order for example to gather feedback from users which is fed into a trust model about data sources, as practiced e.g. in the sig.ma application.

Case (3) has been the cause for lengthy discussions especially within W3C, which started from the opinion that a URI cannot identify a non-web resource and a web resource at the same time. The agreed recommendation [11] on how to solve this conflict is to use status codes of the underlying HTTP protocol to inform an agent whether a URI is identifying a web resource or not, and use a redirection mechanism that provides a web resource.

While this approach solves the problem of knowing *what kind of resource* a URI identifies, it does not address the question of *which* non-web resource such a URI identifies, and does not guarantee that it identifies always the same one. This fact makes mere RDF URIs problematic, even if they are well-implemented (i.e. providing the right status codes and redirection mechanism).

In OCM we are devising a way to add precision to the management and interpretation of identifiers on the Semantic Web. While RDF URIs satisfy cases (1) and (2), and can be implemented technically to conform to W3C recommendations, `OKKAMid`s add the possibility to refer *rigidly* and *directly* to a non-web resource. This means that to become a "cool URI", we recommend one of the following solutions: (i) using directly `OKKAMid`s for non-web resources, whenever possible; or (ii) adding a `corefer` statement for each non-web resource named in the dataset (in OKKAM, this is called "OKKAMization"); or (iii) makes sure that applications aiming at aggregating different RDF datasets make a runtime call the ENS for retrieving the `OKKAMid`s of the non-web resources named in a dataset.

The adoption of such an approach has three important benefits; first, RDF URIs will maintain their intended use of being interpreted into a set of triples. Secondly, as a consequence, RDF URIs are perfectly suited to implement *Linked Data*, preserving provenance and context. Finally, *ad-hoc* solutions for calculating the transitive closure over `owl:sameAs` statements can be often avoided because identity is syntactically evident. If transitive closure is required, the ENS accomodates for the notion of *dereferenceable alternative ID* and provides the community with the practical solution of maintaining these closures in a defined location.

To sum up, our recommendation is that the `owl:sameAs` statements are reserved to cases in which one intends to express a strong semantic link of compatibility between two different *descriptions* of the same non-web resource. Whereas coidentification statements, i.e. statements about the fact that two or more RDF URIs identify the same non-web resource, should be inferred from the fact that different RDF URIs are mapped onto the same `OKKAMid` in the ENS through the `corefer`(x,y) relation defined in 25. In this picture, one can think of an OKKAM profile as a **gateway** to information about a non-web resource existing on the Web. It turns out that the relation of redirection should be thought of as performing two distinct functions for RDF URIs and `OKKAMids`. The processes of redirection and resolution that connect RDF URIs to pieces of information should be thought of as functional in the following sense: a RDF URI – via redirection and the `owl:sameAs` relation – should be connected to one and only one *coherent* piece of information about a non-web resource and keep tracks of its sources. On the contrary, the process of redirection and resolution that connect `OKKAMids` to pieces of information in general is not functional, as the OKKAM profile of an entity should not be interepreted nor used as an additional piece of information about the entity, but only as information which a community agrees to use in order to fix the referent of that OKKAMid.

References

[1] John Langshaw Austin. *How to Do Things with Words: The William James Lecture.* Oxford University Press, Oxford, 1962.

[2] Paolo Bouquet, Heiko Stoermer, Claudia Niederee, and Antonio Mana. Entity Name System: The Backbone of an Open and Scalable Web of Data. In *Proceedings of the IEEE International Conference on Semantic Computing, ICSC 2008*, number CSS-ICSC 2008-4-28-25, pages 554–561. IEEE Computer Society, August 2008.

[3] Keith S. Donnellan. Reference and definite descriptions. *The Philosophical Review*, 77:281–304, 1966.

[4] Gareth Evans. *The Varieties of Reference.* Oxford University Press, Oxford, 1982.

[5] Aldo Gangemi and Valentina Presutti. Towards an OWL Ontology for Identity on the Web. In *Semantic Web Applications and Perspectives (SWAP2006)*, 2006.

[6] Aldo Gangemi and Valentina Presutti. A grounded ontology for identity and reference of web resources. In *i3: Identity, Identifiers, Identification. Proceedings of the WWW2007 Workshop on Entity-Centric Approaches to Information and Knowledge Management on the Web, Banff, Canada, May 8, 2007.*, 2007.

[7] Harry Halpin and Valentina Presutti. An ontology of resources: Solving the identity crisis. In Lora Aroyo *et al.*, editor, *Proceedings of ESWC2009*, Studies in Logic and Computation, pages 121–140. Research Studies Press/Wiley, 2009.

[8] S. Kripke. *Naming and necessity.* Harvard University Press, 1972.

[9] P.F. Patel-Schneider, P. Hayes, and I. Horrocks. Web Ontology Language (OWL) Abstract Syntax and Semantics. Technical report, W3C, February 2003. `http://www.w3.org/TR/owl-semantics/`.

[10] Presutti V. and Gangemi A. Identity of resources and entities on the web. *International Journal on Semantic Web and Information Systems*, 4(2), 2008.

[11] W3C. *Cool URIs for the Semantic Web. W3C Interest Group Note 03 December 2008*, December 2008. online `http://www.w3.org/TR/cooluris/`.