

An Ontology of Resources: Solving the Identity Crisis

Harry Halpin and Valentina Presutti
H.Halpin@ed.ac.uk valentina.presutti@cnr.it

School of Informatics
University of Edinburgh
2 Buccleuch Place
EH8 9LW Edinburgh
Scotland, UK
and
Semantic Technology Laboratory
ISTC-CNR
Via Nomentana 56 00161
Rome, Italy

Abstract. The primary goal of the Semantic Web is to use URIs as a universal space to name anything, expanding from using URIs for webpages to URIs for “real objects and imaginary concepts,” as phrased by Berners-Lee. This distinction has often been tied to the distinction between information resources, like webpages and multimedia files, and non-information resources, which are everything from real people to abstract concepts like ‘the integers.’ Furthermore, the W3C has recommended not to use the same URI for information resources and non-information resources, and several communities like the Linked Data initiative are deploying this principle. The definition put forward by the W3C, that non-information resources are things whose “essential nature is information” is a difficult distinction at best. For example, would the text of Moby Dick be an information resource? While this problem could safely be ignored up until recently, with the rise of Linked Data and projects like OKKAM, it appears that this problem should be modelled formally. An ontology called IRW (Identity and Reference on the Web) of various types of resources and their relationships, both for the hypertext Web and the Semantic Web, is presented. It builds upon Information Object Lite (an extension of DOLCE Ultra Lite for describing information objects) and IRE (an earlier ontology of and aligns with other work in this area. This ontology can be used as a tool to make the Semantic Web more self-describing and to allow inference to be used to test for membership in various classes of resources.

Keywords: *Web architecture, resource, ontology, linked data*

1 Introduction

The Semantic Web constitutes a truly new scenario for knowledge representation that may have a strong impact on existing models due to its use of URIs as a mechanism of identification of resources. The key feature of the Semantic Web is not its use of knowledge representation technologies like ontologies and inference per se, but the introduction of these technologies to operate over the Web, as defined as a space of URIs. On the Semantic Web, we can use URIs in order to name anything, which is in contrast to the hypertext Web, which primarily used URIs just to access webpages. Due to this bifurcation, there are now several types of resources on the Web, and the W3C has prescribed different behaviors for their identification and access. For this reason it is of key importance to make it clear what kinds of resources can be identified on the Semantic Web, and how to distinguish between them. Yet, there is still debate within Web circles as to what the definition of a ‘resource’ is, what kinds of resources there are, and how these relate to the Semantic Web and to the pre-Semantic Web hypertext Web. We model these terms and the debates around them using a lightweight formal ontology in OWL-DL, which we call *IRW*, for ‘Identity of Resources on the Web.’ *IRW* is meant to be an helpful formal tool for resolving conflicting arguments about identity and URIs, and as a consequence, it provides a supporting vocabulary for implementing practical solutions in a variety different scenarios.

One of the most difficult tasks of Web architecture is stating the definitions of terms in natural language, the obvious first step to formalization. First, what is a resource? The W3C TAG state in their *Architecture of the Web* that ‘resource’ is used in a general sense for whatever might be identified by a URI [19]. Previously, a resource was thought of as strictly network-accessible objects such as webpages, since the term ‘resource’ is defined by Fielding in the first HTTP RFC as “a network data object or service, identified by a URI”¹. However, Berners-Lee broadened the concept of resource in IETF RFC 2396, stating that “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” [4].

One distinction that has been upheld by Hayes and others is the distinction between reference and access [17]. Making an analogy between URIs and names, *access* means “that the name provides a causal pathway to the thing, perhaps mediated by the Web” while *reference* means that “the name is being used to mention the thing,” which may or may not coincide with access [17]. Some thing is then ‘Web-accessible’ if it can accessed via the use of HTTP. This use of the term ‘resource’ for both referring to non-Web accessible things and for naming Web-accessible things is continued in URI RFC 3986, the current IETF RFC,

¹ <http://tools.ietf.org/html/draft-fielding-http-spec-00>

which states that “this specification does not limit the scope of what might be a resource ... 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)” [5]. It seems that *very different* kinds of things are being described by the notion of a resource.

The most obvious distinction is between a resource that could *in principle* be Web-accessible, like a webpage, and a resource that is not in principle Web-accessible, like the Eiffel Tower itself. This distinction is given by the W3C TAG as the distinction between an information resource and something that may not be an information resource [19]. The W3C TAG then define an *information resource* as something “whose essential characteristics can be conveyed in a message,” which is a controversial definition [19]. As noted by the Linked Data tutorial, this implies there is another kind of resource, *non-information resources*, for things that are not possibly Web-accessible, like a URI whose primary purpose is to refer to the Eiffel Tower [7]. Furthermore, one can distinguish ‘Web resources’ (a subset of information resources) that *are* usually Web-accessible, such as web-pages, from things that simply carry information, like the text of Moby Dick, regardless of whether it is on the Web or not. Again, let us emphasize that some find these distinctions very intuitive, while others do not. Lastly, in order to distinguish URIs for non-accessible things on the Semantic Web (the ‘Cool URIs for the Semantic Web’) from the normal use of URIs on the hypertext Web, we call the former *Semantic Web URIs* [26]. In Web architecture circles, what are typically called ‘webpages’ are called instead ‘representations’ of a resource, since multimedia and other not-hypertext pages can be retrieved via the Web [19]. In order to distinguish the use of the word ‘representation’ in Web architecture circles from its normal usage, the word *Web Representation* is used in this paper to designate a more encompassing notion of web-page as any set of bits that is ‘coming down the wire’ on the Web.

2 The Identity Crisis

What does a URI identify? For multiple agents to exchange knowledge representations on the Semantic Web, they must be able to identify what *precise* kinds of resource a URI identifies if they are to make any non-trivial inferences. For example, how can an agent determine whether or not a URI for the Eiffel Tower is used to refer to the Eiffel Tower in Paris? If so, does the aforementioned URI access a Web-accessible representation? This cluster of questions has been dubbed the *Identity Crisis* of the Semantic Web.

2.1 Two Viewpoints

There are at least two distinct positions on this question. The first position, the *direct reference position*, holds that on the Web a URI identifies whatever it

was intended by the owner to identify. The owner of the URI should be able to unambiguously declare and communicate the meaning of any URI, including a Semantic Web URI. In this position, the referent is generally considered to be some individual unambiguous *single* thing, like the Eiffel Tower or the concept of unicorns. The URI is given its referent by fiat by the owner or creator of the URI, who then communicates this via some causal chain, in a similar manner to Kripke’s causal theory of proper names [21]. This viewpoint is the one generally held by many Web architects, like Berners-Lee, who imagine it holds not just for the Semantic Web, but the entire Web. Of course this seems true for the hypertext Web, since a URI naturally identifies a web-page by virtue of providing access to the web-page.

The second position, the *logician position*, holds that for the Semantic Web, a URI refers to whatever model(s) – including actual things – that satisfy the formal semantics of the Semantic Web. Adherents of this position hold that the referent of a URI is almost always ambiguous, as many different models can satisfy an interpretation of a RDF graph. This position has been championed extensively against Berners-Lee by Hayes, as Hayes believed that the direct reference position “doesn’t make sense, that it isn’t true, and that it could not possibly be true” as it contradicts the standard interpretation of Tarski-style formal semantics [16]. A URI has no identity in and of itself, but only in the context of its use in a graph or, in a minor variation argued for by Parsia and Patel-Schneider, the explicit use of `owl:imports` [23]. This position is generally held by those who claim that the Semantic Web is entirely distinct from the hypertext Web. Others like Booth [9] hold a hybrid position that an agent should be able to access logical descriptions from the URI itself so that whatever things satisfy these accessible logical descriptions are the referents.

2.2 The TAG’s Resolution

The TAG officially resolved *httpRange-14* by saying that the 303 **See Other** HTTP header can serve to disambiguate between information resources and possible non-information resources. The official resolution to Identity Crisis by the TAG is given below as [13]:

- If an HTTP resource responds to a GET request with a 2xx response, then the resource identified by that URI is an information resource;
- If an HTTP resource responds to a GET request with a 303 (**See Other**) response, then the resource identified by that URI could be any
- If an HTTP resource responds to a GET request with a 4xx response, then the nature of the resource is unknown.

One concrete example would be an agent is trying to access a URI that refers to the Eiffel Tower itself, `http://dbpedia.org/resource/Eiffel_Tower`. Upon attempting to access that resource with a HTTP GET request on a URI, since the Eiffel Tower itself is not an information resource, no Web representations

are directly available. Instead, the agent gets a 303 *See Other* that in turn redirects them to an information resource that hosts Web representations about the Eiffel Tower, such as `http://dbpedia.org/page/Eiffel.Tower`. When this URI returns the 200 status code in response to an HTTP GET request, the agent can infer that `http://dbpedia.org/page/Eiffel.Tower/` is actually an information resource. The Semantic Web URI used to refer to the Eiffel Tower itself, `http://dbpedia.org/resource/Eiffel.Tower`, could be any kind of resource and so *could* be a non-information resource [13]. This example is illustrated in Figure 1, using terms from the IRW ontology introduced in Section 5. An alternative to the 303 redirection is the *hash convention*, in which one

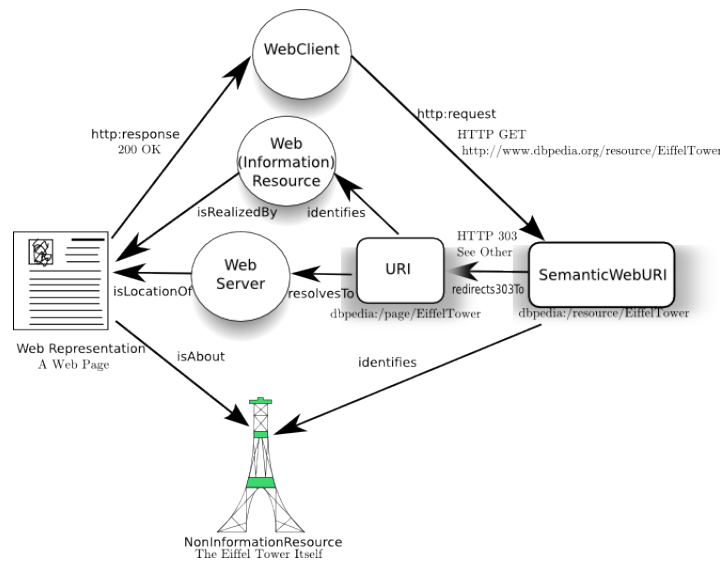


Fig. 1. 303 Redirection for Semantic Web URIs

uses the fragment identifier of a URI to get redirection ‘for free’ with smaller RDF vocabularies. If one wanted a Semantic Web URI that referred to the Eiffel Tower itself without the hassle of a 303 redirection, one would use the URI `http://www.tour-eiffel.fr/#it` to refer to the Eiffel Tower itself. Since browsers either dispose of or treat the fragment identifier as a fragment of a hypertext document or some other Web representation, if an agent tries to access via HTTP GET a Semantic Web URI that uses the hash convention, the server will not return a 404 *Not Found* status code, but instead will resolve to the URI before the hash, `http://www.tour-eiffel.fr`, which can then be an Web resource capable of returning Web representations, which is called an ‘associated description’ in the Linked Data community [7]. In this way, Semantic Web inference engines can keep the Semantic Web URI that refers to the Eiffel Tower and an associated description about the Eiffel Tower separate by taking advantage of

the predefined behaviour in web browsers. However, practically the 303 redirection of the W3C TAG and the hash convention leave the question of whether a resource is an information resource or non-information resource indeterminate, since there is nothing to prevent 303 redirection from being used to redirect from one information resource to another information resource, and the hash convention is dependent on media types, being more often used for named parts in the document in HTML instead of as a shortcut for distinguishing non-information resources and their associated descriptions.

3 Related Work

There has been some related work in this area. Mogul has suggested that there are fundamental disagreements about what precisely the difference between an HTTP entity and a “representation of a resource” are, and that this leads to widespread problems with caching implementations in HTTP [22]. David Boorh has proposed an informal categorisation of what can be identified by a URI, noticing the confusion between ‘naming’ and ‘identifying’ and even ‘describing’ [8]. Hayes has long attempted to elucidate the fundamental difference between the use of resources to access webpages and the use of a URI to refer to some non-Web accessible thing [17]. Furthermore, the use of URIs to refer to physical entities and the subsequent clarification of the direct reference position has led to the OKKAM project, a project to build a catalogue of ‘entity’ URIs that is supposed to directly refer to physical entities [10]. This general line of thinking has led to a number of workshops at conferences such as the World Wide Web Conference and the European Semantic Web Conference devoted to this topic [11, 12].

Within the W3C, there is an informal activity of the W3C TAG called the ‘Architecture of the Semantic Web’ (AWWSW) that has for over a year attempted to decipher Web architecture formally, in part prompting by the need to model HTTP in RDF directly in order for HTTP transactions to be validated via EARL, the RDF-based *Evaluation and Report Language* used by the W3C to validate new W3C standards and describe test-cases [1, 20]. Yet, HTTP in RDF currently does not model the notion of ‘resource’ except with a misuse of `rdf:Alt`, so it must be corrected by integrating an ontology of resources like IRW. While both EARL and the AWWSW are attempting a much more detailed and low-level description of HTTP transactions than we attempt, the lightweight IRW ontology described in this paper should allow specifications like HTTP in RDF to directly address the notion of a ‘resource.’

4 The Use of a Formal Ontology

The primary use of a formal ontology in the context of Web architecture and the arguments around the Identity Crisis is to allow those involved in debates to model formally their positions using a common ontology as a starting point,

and so clarify disagreements, as shown in Section 7. To this aim, IRW can be discussed, reviewed, and comment on the [ontologydesignpatterns.org](http://ontologydesignpatterns.org/wiki) wiki². To serve the aim of elucidating arguments, additional modules of IRW have been developed and are briefly introduces in Section 5.

There have been previous attempts to model at least a subset of the notions outlined beforehand in a formal ontology, but all lack coverage of some crucial concepts. For example, while the ontology given by RDF Schema touches upon the vocabulary of resources via its term `rdfs:Resource`, it does not cover the distinction between information and non-information resources. The IRE (Identifiers, Resources, and Entities), based on Dolce Ultra Lite (DUL),³ a light version of the widely-known DOLCE foundational ontology and its extension for describing information objects⁴ (IOL, described in [14]), attempted to model some of these concepts earlier [25]. However, many aspects were not included in IRE, such as the distinctions between resources and their Web representations, or the concept of accessing a web-page via a web server, that are crucial to the efforts within the W3C, while many of the distinctions drawn by DUL+IOL were found to be too ‘heavy-weight’ for these communities [15]. In response to these concerns, the IRE ontology has been evolved into the IRW ontology.

5 The IRW Ontology

The prefix `irw:` is for the namespace <http://purl.org/NET/irw/> of the IRW ontology. The stable version of the ontology can also be accessed via its PURL. The latest version of the IRW ontology may be accessed at:

<http://ontologydesignpatterns.org/ont/web/irw.owl>. The prefix `rdfs:` is used for the RDF(S) namespace <http://www.w3.org/2000/01/rdf-schema#>. Note `ir:` is

<http://www.ontologydesignpatterns.org/cp/owl/informationrealization.owl>.

While the IRW ontology in full can not explicated due to lack of space, the primary classes and properties are given in Figure 2. The IRW-related elements needed for the example of 303 redirection are given in Figure 1. The IRW ontology starts with `irw:Resource`. While this class expresses the same intuition as `rdfs:Resource`, we have defined it because this version of IRW is within OWL-DL expressivity. In OWL Full, this class is equivalent to `rdfs:Resource`. Now, we move to modelling the debates around the Identity Crisis.

Identification and reference. The notion of a URI is modeled as a class, `irw:URI` that has exactly one value for the datatype property `irw:hasURI` allowing to specify its value. Modelling URIs as a class allows us to talk about different kinds of URIs, such as IRIs (Internationalized Resource Identifiers) and Semantic

² <http://ontologydesignpatterns.org/wiki/Submissions:IRW>

³ <http://www.loa-cnr.it/ontologies/DUL.owl>

⁴ <http://www.loa-cnr.it/ontologies/IOLite.owl>

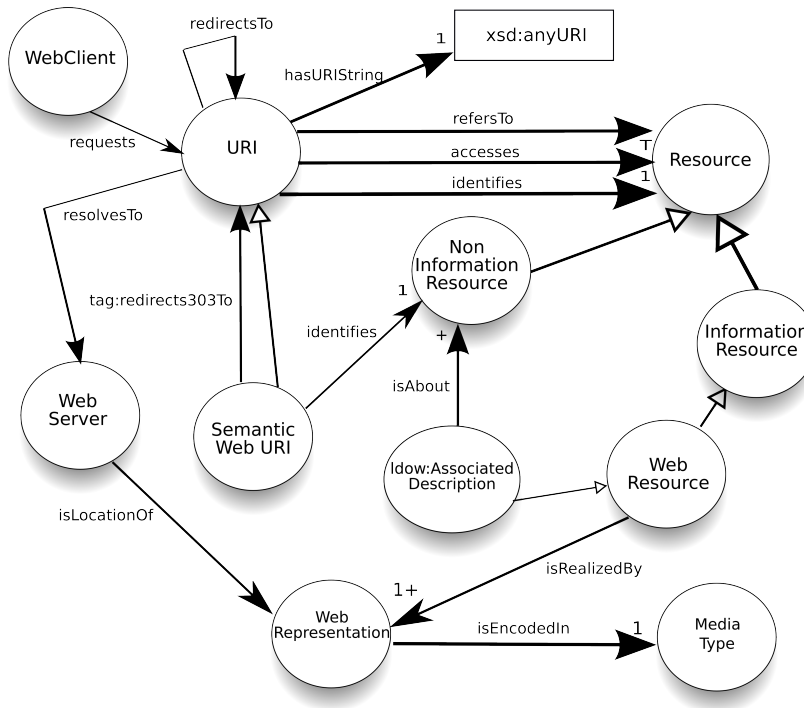


Fig. 2. The IRW ontology illustrated as a graph. Rounded nodes are classes, while rectangular ones are datatypes. Arcs ending with an empty triangle are `rdfs:subClassOf` relationships. Arcs ending with a filled triangle are either object properties or datatype properties depending of the range node. Arcs' direction indicates the domain and range of the property. A '1' associated to a property means it is functional, a 'T' means it is transitive, '1+' means 'at least one'. Prefixes are indicated only if different from `irw:`.

Web URIs. According to the *direct reference position*, URIs identify exactly one resource. This is modeled in IRW by the functional property `irw:identifies`, having range `irw:Resource` (and inverse property, `irw:isIdentifiedBy`). Hayes would disagree with `irw:identifies` being functional and would not use it at all, but instead would restrict the vocabulary to what he believes are the more precise `irw:accesses` and `irw:refersTo` [17]. The viewpoint of the *logician position*, namely that a URI can refer to more than one resource, is modeled by the object property `irw:refersTo` (and inverse property, `irw:isReferencedBy`). One aspect of this property is that the object of reference can be “immediately causally disconnected” from its subject [17]. This is important, as reference is used as a relationship between URIs and resources, including not only information resources but also non-information resources like the Eiffel Tower or integers.

Access and redirection. Distinct from reference is the `irw:accesses` relationship, which is a causal connection to the thing identified. This is modelled again as a relationship between URIs and resources, although it is transitive, unlike `irws:refersTo`. If one can access *a* and *a* accesses *b* then *a* accesses *c* (via *b*). Although a wide notion, access allows us to model the typical HTTP request-response Web transactions between a Web client and a server. A URI may also have a `irw:redirectsTo` property, a sub-property of `irw:accesses`, that we can use to model HTTP redirection. However, since redirection can be used between just information resources that have nothing to do with the Semantic Web, their domain and range say nothing about the type of resource. In order to model explicitly the redirection solution to the “Identity Crisis” by the W3C TAG, two distinct sub-properties of this have been added in a TAG-specific module of IRW⁵ that contains `tag:redirects303To` property and a `tag:redirectsHashTo` property. Obviously, `tag:redirects303To` models the TAG’s ‘solution’ to *httpRange-14* while `tag:redirectsHashTo` represents the hash convention.

Types of resources. Having defined reference and redirection, we can now categorize resources. There are two main disjoint sub-classes of `irw:Resource`. The first subclass is given as `irw:InformationResource`, which is an information object, such as a musical composition, a text, a word, or a picture. An information object is an object defined at a level of abstraction, independently from how it is concretely realized. So an `irw:InformationResource` expresses the same intuition and is an equivalent class to the DUL+IOL information object [14]. This means an information resource has, via the `ir:realizes` property (with inverse `ir:isRealizedBy`), at least one `ir:InformationRealization`, a concrete *realization*. This term is again imported from DUL+IOL [14]. So an information resource’s “essential characteristics can be conveyed in a single message” implies that everything from a bound book to an HTTP message can be a realization of an information resource [19]. Furthermore, the property `irw:isAbout` (and inverse property, `irw:isTopicOf`) expresses the relationship of an information resource to a resource or resources the information is ‘about.’ Examples of this are descriptions of a resource using natural language or depictions of a resource using images. Information resources also can, but not necessarily, be identified (either accessed or referred to) with a URI. In this manner, the text of Moby Dick can be an information resource since it could be conveyed as a single message in English, and can be realized by both a particular book or a webpage containing that text.

Note `irw:NonInformationResource` complements `irw:InformationResource` from which it is disjoint with. Such class represents things that can not themselves – for whatever reason – be realized as a single digitally encoded message. A number of different kinds of things may be `irw:NonInformationResources`.

⁵ <http://www.ontologydesignpatterns.org/ont/web/tag2irw.owl> associated with prefix `tag:.`

Since this concept is the cause of much confusion and debate, it is detailed with three disjoint sub-classes. These kinds of IRW distinctions are not normative, as there are other possible plausible, more detailed modeling choices. Our aim here is to communicate the intuition behind the concepts of information and non-information resources without entering the philosophical debate about top-level ontologies. IRW contains three sub-classes of `irw:NonInformationResources`:⁶ `irw:PhysicalEntityResource`, is a resource that is ‘touchable’ like physical people, artifacts, places, bodies, chemical substances, biological entities; `irw:ConceptualResource`, which refer to resources that are created in a social process that can not be completely realized digitally, such as legal entities, political entities, social relations, as well as the concept of horse and imaginary objects like unicorns; and finally `irw:AbstractResource`, which refers to abstract combinatorial spaces that cannot be located in space-time such as formal entities like functions or the integers as well as more mundane resources like the infinite set of names that constitute the resource identified by URIs themselves.

A sub-class of `irw:InformationResource` is `irw:WebResource`, which is an information resource identified by at least one URI and realized by at least one `irw:WebRepresentation`, so that a Web resource is just an information resource that is realized by at least one accessible Web representation like a web-page. `irw:WebRepresentation` is a sub-class of `irw:InformationRealization` with constraints added to make the cardinality of `ir:isRealizedBy` and `irw:isIdentifiedBy` both at least 1. In this way IRW can distinguish between a resource for the text of ‘Moby Dick’ in general and a webpage about ‘Moby Dick.’

Hypertext Web transactions. The typical hypertext Web transaction can be modelled by IRW. We begin with `irw:WebClient`, which is some client in the context of the Web that can have a `irw:requests` relationship to a URI (note that `irw:requests` serves as an hook to the alignment of IRW with *HTTP in RDF* [20]), as exemplified by a typical HTTP GET request). The `irw:requests` property is a sub-property of `irw:access`. A `irw:WebClient` then `irw:requests` a `irw:URI`. We also introduce the class `irw:WebServer`, which has a `irw:isResolutionOf` property that relates a URI to a concrete Web server (inverse property `irw:resolvesTo`). This `irw:resolvesTo` property is currently implemented by mapping a URI to an IP address or addresses. So each `irw:WebServer` is the resolution of at least one `irw:URI`. Additionally, a `irw:WebServer` has a `irw:isLocationOf` property with at least one `irw:WebRepresentation` (inverse property, `locatedOn`), indicating the Web server concretely can respond to an HTTP request with a particular Web Representation.

Semantic Web (Linked Data) transactions. The typical Semantic Web transaction is also modeled. A new sub-class of `irw:URI`, `SemanticWebURI` is

⁶ Note that the three classes does not constitute an exhaustive partition.

given, where the Semantic Web URI has a constraint that it must have at least one `irw:redirects` property.

In the Linked Data Initiative, another important kind of resource is “associated descriptions,” which is just an Web resource that can be accessed via redirection from a Semantic Web URI [7]. For example, in DBPedia⁷ the resource `dbpedia:/resource/Eiffel_Tower` redirects to some RDF/XML at `dbpedia:/data/Eiffel_Tower`, and to an HTML page at `dbpedia:/page/Eiffel_Tower` depending on the requested media type [3]. This Linked Data typical scenario can be generalized: a `irw:WebClient` `irw:requests` a `irw:SemanticWebURI` x and the request is redirected (e.g. via hash or 303 redirection) to another URI, where this second URI identifies an `ldow:AssociatedDescription`,⁸ which has one `irw:isAbout` property to a non-information resource. We model `ldow:AssociatedDescription` as a subclass of `irw:WebResource`.

6 Resolving the Identity Crisis

One purpose of this ontology is to describe, in formal detail, the exact nature of the conflicts between the various sides of the Identity Crisis debate. The main conflict between the *logician* position of Hayes and the *direct reference* position of Berners-Lee can then be cast as an argument over three IRW properties. Berners-Lee’s slogan that “URIs identify one thing” is modelled by having the `irw:identifies` property be *functional*, i.e. a URI can only identify one resource. Furthermore, he would also hold that a `irw:SemanticWebURI` `irw:refersTo` exactly one `irw:NonInformationResource`.

Hayes’s response would be that `irw:identifies` should be eliminated and there can be no constraints whatsoever on `irw:refersTo` and thus no constraints on the usage of URIs for referring to things on the Semantic Web, while typical hypertext Web transactions can be modelled functionally with `irw:accesses`. Although IRW models Berners-Lee’s more general notion of identification via `irw:identifies`, IRW also captures Hayes’s perspective with the properties `irw:refersTo` and `irw:accesses`. Lastly, the criticisms of redirection modeled with `irw:redirectsTo` has mainly to do with the fact that the domain can only be simply `irw:URI` rather than `irw:SemanticWebURI`, which we also explicitly model. Thus, there is no way to ever definitely be sure that a URI is a Semantic Web URI and so one can never be sure that a URI identifies a non-information resource. We show how IRW can solve this problem in Section 8.

⁷ Prefix `dbpedia:` is used for the namespace `http://dbpedia.org`

⁸ Typical Linked Data terminology is represented in a specific module of IRW represented here by the prefix `ldow:` referring to the namespace `http://ontologydesignpatterns.org/ont/web/ldow2irw.owl`

7 Aligning IRW to other Ontologies

In this section, we present a number of suggested alignments, as given in Table 2. The alignments are to the three primary other ontologies, the *RDF in HTTP* ontology [20], and the IRE ontology as well as an ontology for HTTP used by the Tabulator Browser [25, 6]. The namespaces for `ont` is <http://www.w3.org/2007/ont/http>. IRE, due to its modular construction and re-use of terms from DUL+IOL patterns, uses many namespaces, but they can be found at <http://www.ontologydesignpatterns.org/cpont/ire.owl>. The `http` namespace is <http://www.w3.org/2006/http#>.

Class or Property	Alignments
<code>irw:WebRepresentation</code>	<code>owl:equivalentClass http:Message</code> <code>owl:equivalentClass ont:ResponseMessage</code> <code>rdfs:subClassOf ire:InformationRealization</code> <code>rdfs:subClassOf ir:InformationRealization</code>
<code>http:Content</code>	<code>rdfs:subClassOf ir:InformationRealization</code>
<code>http:MessageHeader</code>	<code>rdfs:subClassOf ir:InformationRealization</code>
<code>irw:InformationResource</code>	<code>owl:equivalentClass ir:InformationObject</code>
<code>irw:SemanticWebURI</code>	<code>ire:SemanticWebURI</code>
<code>irw:identifies</code>	<code>ire:isExactProxyFor</code>
<code>irw:isAbout</code>	<code>ire:about</code>

Table 1. Mapping of IRW to Other Ontologies

8 The Self-Describing Semantic Web

The main application of this ontology is to clarify debates over Web architecture, in particular, the ‘Identity Crisis’ debate. However, the IRW ontology can also solve the problem noted earlier that currently it is impossible to describe whether or not some resource describes some non-Web accessible thing, such that there is no “definition, description, some other kind of indication of what the identifier is intended to identify” [24]. Solving this can be done on via adding IRW statements to associated descriptions accessible via Semantic Web URIs. There would be a number of advantages if webpages that have RDF content could distinguish themselves as such, in the same way that HTML ‘valid’ documents are currently validated by W3C Validators. This can be done by embedding a IRW statement in RDF/XML documents, RDF returned from SPARQL endpoints, and RDFa or GRDDL statement in XHTML or XML documents [2]. Ideally, this would be in conjunction with some sort of graphical logo to distinguish the page as ‘Semantic Web Enabled,’ much as current web-pages can be marked up with a logo for ‘XHTML 1.0 Valid.’ This is useful because detecting RDF ‘in the wild’ on the Web, such as embedded RDFa, can be difficult for humans. The main

problem is that an `irw:NonInformationResource` has no Web representation to embed such a statement in. Take for example the Semantic Web URI created by Pat Hayes for himself: `www.ihmc.us/users/phayes/PatHayes.html`. While originally a stand-alone web-page, currently Hayes has the URI use 303 redirection to `http://www.ihmc.us/users/phayes/PatHayesAbout.html`.⁹ This latter web-page could easily use a combination of RDFa as IRW to mark itself up as a representation of a non-information resource by including the statement that `phayes:PatHayes.html rdf:type irw:NonInformationResource` and adding `phayes:PatHayes.html tag:redirects303To phayes:PatHayesAbout.html`. The IRW ontology can help explicitly model and make available to the rest of the Semantic Web the often subterranean details of Web architecture.

9 Conclusion and Future Work

Overall, the IRW ontology is a beginning, yet it should serve as foundational contribution of modelling Web architecture, the “Dark Side of Semantic Web” that Hendler believes may give the Semantic Web a crucial advantage over previous efforts in knowledge representation [18]. What is surprising is that it has taken so long for this problem to be solved in a relatively straightforward manner. However, as the problem space is highly contentious and the various documents that describe this problem are spread throughout many informal and semi-formal notes and standards (and arguments over W3C lists), systematising the problem and modelling it formally was perhaps more difficult than would be expected. Such problems that are closest the Semantic Web are often the most difficult to model objectively, such as meta-modelling in general. Future work needs to be done to standardise IRW or a descendant thereof through the W3C, which will doubtless result in refinements to IRW, and to encourage its use within the Linked Data community in the context of various validators, debuggers, and search engines. By developing a consistent vocabulary for describing the identity of resources in IRW, the first step has been taken.

10 Acknowledgements

We would like to thank Aldo Gangemi for his insightful comments. Harry Halpin was partially supported by a Microsoft ‘Beyond Search’ award.

⁹ Let `phayes:` stand for `http://www.ihmc.us/users/phayes.`

Bibliography

- [1] S. Abou-Zahra. Evaluation and Report Language (EARL) 1.0 Schema. W3C Working Draft, W3C, 2007. <http://www.w3.org/TR/EARL10-Schema/>.
- [2] B. Adida, M. Birbeck, S. McCarron, and S. Pemberton. RDFa in XHTML: Syntax and Processing. W3C Recommendation, W3C, 2008. <http://www.w3.org/TR/rdfa-syntax/>.
- [3] S. Auer, C. Bizer, J. Lehmann, G. Kobilarov, R. Cyganiak, and Z. Ives. DBpedia: A nucleus for a web of open data. In *Proceedings of the International Semantic Conference and Asian Semantic Web Conference (ISWC/ASWC2007)*, pages 718–728, Busan, Korea, 2007.
- [4] T. Berners-Lee, R. Fielding, and L. Masinter. IETF RFC 2396 Uniform Resource Identifier (URI): Generic Syntax, 1998. <http://www.ietf.org/rfc/rfc2396.txt> (Last accessed on Sept. 15th 2008).
- [5] T. Berners-Lee, R. Fielding, and L. Masinter. IETF RFC 3986 Uniform Resource Identifier (URI): Generic Syntax, January 2005. <http://www.ietf.org/rfc/rfc3986.txt> (Last accessed on April 2th 2008).
- [6] T. Berners-Lee, J. Hollenbach, K. Lu, J. Presbrey, E. Prud'hommeaux, and mc schraefel. Tabulator Redux: Browsing and Writing Linked Data. In *Proceedings of the WWW2007 Workshop on Linked Data on the Web*, 2008.
- [7] C. Bizer, R. Cygniak, and T. Heath. How to publish Linked Data on the Web, 2007. <http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/> (Last accessed on May 28th 2008).
- [8] D. Booth. URIs and the myth of resource identity. In *Proceedings of Identity, Reference, and the Web Workshop at the WWW Conference*, 2006. <http://www.ibiblio.org/hhalpin/irw2006/dbooth.pdf>.
- [9] D. Booth. URIs declaration versus use. In *Proceedings of Identity, Reference, and the Semantic Web Workshop at the European Semantic Web Conference*, 2008.
- [10] P. Bouquet, H. Stoermer, and D. Giacomuzzi. OKKAM: Enabling a Web of Entities. 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.*, CEUR Workshop Proceedings, ISSN 1613-0073, May 2007. online http://CEUR-WS.org/Vol-249/submission_150.pdf.
- [11] P. Bouquet, H. Stoermer, G. Tummarello, and H. Halpin, editors. *Proceedings of the WWW2007 Workshop I³: Identity, Identifiers, Identification, Entity-Centric Approaches to Information and Knowledge Management on the Web, Banff, Canada, May 8, 2007*, CEUR Workshop Proceedings. CEUR-WS.org, 2007.
- [12] P. Bouquet, H. Stoermer, G. Tummarello, and H. Halpin, editors. *Proceedings of the ESWC2008 Workshop on Identity, Reference, and the Web, Tenerife, Spain, June 1st, 2008*, CEUR Workshop Proceedings, 2008.
- [13] D. Connolly. A pragmatic theory of reference for the web. In *Proceedings of Identity, Reference, and the Web Workshop at the WWW Conference*, 2006. <http://www.ibiblio.org/hhalpin/irw2006/dconnolly2006.pdf> (Last accessed November 22nd 2008).
- [14] A. Gangemi. Norms and plans as unification criteria for social collectives. *Journal of Autonomous Agents and Multi-Agent Systems*, 16(3), 2008.

- [15] A. Gangemi, N. Guarino, C. Masolo, R. Oltramari, and L. Schneider. Sweetening ontologies with DOLCE. In *Proceedings of International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web*, pages 166–181. Springer, 2002.
- [16] P. Hayes. Message to www-rdf-comments@w3.org, 2003. <http://lists.w3.org/Archives/Public/www-tag/2003Jul/0198.html>.
- [17] P. Hayes and H. Halpin. In defense of ambiguity. *International Journal of Semantic Web and Information Systems*, 4(2):1–18, 2008.
- [18] J. Hendler. The Dark Side of the Semantic Web. *IEEE Intelligent Systems*, 22(1):2–4, 2007.
- [19] I. Jacobs and N. Walsh. Architecture of the World Wide Web. Technical report, W3C, 2004. <http://www.w3.org/TR/webarch/> (Last accessed Oct 12th 2008).
- [20] J. Koch, C. A. Velasco, and S. Abou-Zahra. HTTP Vocabulary in RDF. W3C Working Draft, W3C, 2008. <http://www.w3.org/TR/EARL10-Schema/>.
- [21] S. Kripke. *Naming and Necessity*. Harvard University Press, Cambridge, Massachusetts, USA, 1972.
- [22] J. Mogul. Clarifying the fundamentals of HTTP. In *Proceedings of the 11th International World Wide Web Conference*, pages 444–457, 2002.
- [23] B. Parsia and P. F. Patel-Schneider. Meaning and the Semantic Web. In *Proceedings of Identity, Reference, and the Web Workshop at the WWW Conference*, 2006. <http://www.ibiblio.org/hhalpin/irw2006/bparsia2006.pdf>.
- [24] S. Pepper. The case for published subjects. In *Proceedings Identity, Reference, and the Web Workshop at the WWW Conference*, 2006. <http://www.ibiblio.org/hhalpin/irw2006/spepper2.pdf>.
- [25] V. Presutti and A. Gangemi. Identity of resources and entities on the web. *International Journal of Semantic Web and Information Systems*, 4(2):49–72, 2008.
- [26] L. Sauerman and R. Cygniak. Cool URIs for the Semantic Web. W3C Note, W3C, 2008. <http://www.w3.org/TR/cooluris/>.