



HAL
open science

The artifactualization of reference and "substances" on the Web.

Alexandre Monnin

► **To cite this version:**

Alexandre Monnin. The artifactualization of reference and "substances" on the Web.: Why (HTTP) URIs do not (always) refer nor resources hold by themselves (post-print). APA Newsletters, 2012, 11 (2). hal-00672301v2

HAL Id: hal-00672301

<https://paris1.hal.science/hal-00672301v2>

Submitted on 17 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The artifactualization of reference and “substances” on the Web.

Why (HTTP) URIs do not (always) refer nor resources hold by themselves.

Alexandre Monnin

Université Paris 1 Panthéon-Sorbonne (PhiCo, EXeCo)
Institut de Recherche et d’Innovation (IRI) du Centre Pompidou
INRIA (Wimmics)
CNAM (DICEN)

“we now have to pay our way in order to subsist”¹ (B. Latour)

Introduction.

From an architectural point of view, the Web can be conceived as an information space full of URIs – Web identifiers. Contrary to popular belief it is *not* a traditional hypertext linking documents or “pages” to one another. Indeed, to account for all the situations encountered on the Web (Web services, dynamic pages, applications, feeds, content negotiation, etc.), a more encompassing theory was needed. According to the latter (the REST style of architecture), Web identifiers have to be treated as *dereferenceable proper names* – URIs (Uniform resource Identifiers), instead of the more well-known URLs (Uniform Resource Locators).

URIs are especially interesting for philosophers. Like proper names, a concept central both to the philosophy of language and metaphysics, they seem to refer to an object. If the architecture of the Web retains some of their characteristics, then philosophers are no longer facing a *terra incognita* but rather a familiar landscape. Unlike proper names, however, URIs also give access to Web contents. As such, they betoken an important change, from a symbolical dimension, where proper names are bestowed certain functions and used to solve philosophical conundrums regarding identity, to a technological one, to quote the late German media theorist Friedrich Kittler, where they earn new functionalities and act as the pillar of a world-wide information system².

¹ Latour & al (2007), p. 110.

² We argue elsewhere that concepts, as semiotic constructs, are also technical tools. Artifactualization is thus always a re-artifactualization.

This shift is what we call *artifactualization*³, the becoming-artifact of philosophical concepts. Our first goal in this paper is to show that reference, the frail symbolic relation between a sign and its referent, is turned into something entirely different on the Web, the space *between* referent and reference, the relation itself, being *adjusted* so as to warrant that reference doesn't fail.

Our second goal is to deal in the same movement with the correlate of URIs, "resources". About ten years after the birth of the Web, it was understood/decided, after careful analysis, that its architecture was a *resource-oriented* one. A very paradoxical move inasmuch as resources are *not* accessible *per se*. But a most important one since it provided the URIs a means to identify "anything at all". Things on the Web, outside of the Web, chairs, people, rates, square circles, etc. The introduction of resources can be seen as a potent way to reopen the ontological question afresh.

Yet, it must also be understood that while resource can be anything, they also share very specific characteristics which have not been properly identified. Drawing from Kittler once again, we could say that the concept of an object for philosophers from Goclenius, Lohardus and Suarez to Kant to Bretano, Twardowski and Meinong, belonged to the symbolic realm while the very notion of a resource belongs to the technical realm as well, born as it was out of an effort to restore consistency to a technical project.

As the Web is spreading and becoming more ubiquitous day after day, we witness an interesting change whereby objects are becoming resources. From an online document to a person or an RFID-enhanced product or device, they are everywhere – or *everyware*, to borrow designer Adam Greenfield's portmanteau word.

Interestingly, on the surface resources share many aspects with what used to be the dominant ontological conception of objects for centuries: substance. However, unlike substances, the category of resource is no longer a natural one. The function of substances was to explain how things like people, organisms or artifacts persisted over time. Without such an ontological background, the issue remains open. We will see that on the Web, resource persistence has a *cost* which has to be assumed by a publisher and depends on protocols and standards. Overall, this will lead to a completely different ontological framework. One that is gaining more and more traction insofar as the network expands.

I. From Web pages to resources.

It has been said that the new digital continent opened new perspective for ontology. "Not since the first work of fiction was produced have philosophers been confronted with such an impressive and so totally unexplored new realm of ontological inquiry as is presented by cyberspace" says David Koepsell in the opening pages of his book, *The Ontology of Cyberspace*. In a similar vein, Luciano Floridi prefers to speak of a process of re-ontologization⁴ but the idea is roughly the same.

The issue is that on specific questions such as "what exactly is a Web page?", philosophers – except for a few exceptions worth mentioning like Harry Halpin – haven't taken into account the work of Web architects. Thus, up until now, a lot more has been done to understand the fundamentals of the

³ Monnin (2009).

⁴ Floridi (2005).

Web inside standardization bodies like the W3C⁵. Koepsell, for instance, in the already quoted book, explains the “retrieval” of a Web page the following way:

Web pages are just another form of software. Again, they consist of data in the form of bits which reside on some storage medium. Just as with my word processor, my web page resides in a specific place and occupies a certain space on a hard drive [sic] in Amherst, New York. When you “point” your browser to <http://wings.buffalo.edu/~koepsell>, you are sending a message across the Internet which instructs my web page’s host computer (a Unix machine at the university of Buffalo) to send a copy of the contents of my personal directory, specifically, a HTML file called “index.html,” to your computer. That file is copied into your computer’s memory and “viewed” by your browser. The version you view disappears from your computer’s memory when you no longer view it, or if cached, when your cache is cleaned. You may also choose to save my web page to your hard drive in which case you will have a copy of my index.html file. My index.html file remains, throughout the browsing and afterward, intact and fixed.⁶

While the default view of the Web is conform to the paragraph quoted, a more general theory was needed to account for cases not covered in this picture:

- The dynamic Web which is also, incidentally, becoming the default Web (services⁷, constantly changing “pages” like newspapers homepages, blogs, etc.)
- “Content negotiation” (abbreviated as “conneg”). A feature of the HTTP protocol accounting for the fact that that users may specify the form of the information they gets access to according to such criteria as languages, accessibility, formats, etc. This means that it is not possible to generalize on the basis of a single case that of retrieving a single HTML page on a server. After all, what gets sent to a browser may take many different forms. It may even be generated on the fly and thus nowhere to be found “on a server” before a request is even sent. In which cases, what is identified by a URI can simply no longer be a single (HTML) file.
- URIs without addressable content (temporarily or not)⁸.
- The lack of a file versioning system⁹ (WebDAV could be used as a counter-example but it never really scaled)

Further examination of the intricate history of Web identifiers is needed to understand why the naïve picture of how the Web works is no longer tenable. Before the creation of the W3C, the Web’s implementation and principles were not thoroughly distinguished. The Web existed in the guise of programming libraries, software and the likes, but no agreed upon standards defined the very principles to which these libraries had to stick. This led to many a conceptual difficulty when the first Web standards were devised around 1994-1995.

The latter had to do both with the nature of the objects available on the Web and their identifiers. At first, the notion of a document (or page) seemed to prevail. The obvious conclusion was that Web identifiers had to be addresses (URLs for *Uniform Resource Locators*) allowing for documents

⁵ See in particular the “Technical Architecture Group” (TAG). Cf. Thompson (2007) for a presentation. Our guess is that while they constitute a pre-requisite for any philosophy of the Web, we should not on the other hand accept non-critically every observation from Web architects.

⁶ Koepsell (2003).

⁷ “Second, there exist many addresses that corresponded to service rather than a document—authors may be intending to direct readers to that service, rather than to any specific result from a prior access of that service.” Fielding & Taylor (2002).

⁸ “Finally, there exist addresses that do not correspond to a document at some periods of time, as when the document does not yet exist or when the address is being used solely for naming, rather than locating, information.” *Ibidem*.

⁹ “First, it suggests that the author is identifying the content transferred, which would imply that the identifier should change whenever the content changes.” *Ibid*.

retrieval in a hypertextual environment. Pages evolving over time (even in the so-called web 1.0 – forums being a good example of the latter), the identification of stable entities as exemplified through library identifiers like ISBNs for books or ISSNs for journals, was transferred to URNs (for *Uniform Resource Names*) – proper names referring to objects not available on the Web. The only problem of these identifiers is that the Web’s main feature is to provide information about a range of entities, whatever status (“inside” or “outside” of the Web) they have. URNs no longer giving access to anything, their value became disputable. The contradiction regarding addressing, on the other hand, became flagrant in one official document, RFC¹⁰ 1736¹¹:

« Locators may apply to resources that are not always or not ever network accessible. Examples of the latter include human beings and physical objects that have no electronic instantiation ».

This is no mere contradiction, rather the renegotiation, *in media res*, of the most fundamental features of a technical project. It is precisely this non-sense that was corrected three years later, in 1998, when the notion of a resource first appeared (elsewhere than is acronyms such as “URIs”, “URLs”, “URNs” or “URCs”). Merely as a correlate of URIs, the latter being established as the new Web identifier after having been sundered in URNs and URLs. URIs are peculiar inasmuch as they add a technical dimension to identification, namely, access¹². They have the status of “dereferenceable proper names” for this reason; being, in other words, proper names that *identify* a resource and *give access* to its representations.

Why resources instead of Web pages, a concept everyone is acquainted with? Simply said, because what is aimed at here is a stable entity whose « representation » can nevertheless vary over time or at a given moment (with conneg). The homepage of the newspaper *The Guardian* I access at time *t* is different from the same homepage I access at *t'*. Likewise, accessing it from a mobile phone or a textual browser will yield different results. These various representations are subject to « synchronic » and « diachronic » modifications¹³. Albeit not the least identical to one another, they must be somehow “faithful” to a given resource (*The Guardian* homepage, not accessible *per se*). Such a notion is especially important with regards to the fact that it allows reference not only to documents (“page”) but also services, physical objects, etc. Overall, it is of paramount importance to restore the Web’s coherence as a technological project beyond the technical changes it underwent with the evolution from a Web of documents to a Web of data or things (also known as the “Semantic Web”).

II. The Ontological status of resources:

Up until now with the resource/representation duality, paralleled in the identification/access one, the debate mainly focused on URIs and their referring prowess. Here, one needs to distinguish between the URI *minter* – the person or institution that create a single URI based of the possession of a domain name, from the *service provider*, the person, institution (generally a company) that will

¹⁰ RFCs (Requests for Comments) are documents of the IETF (Internet Engineering Task Force) where most Internet standards are consigned. In spite of the creation of the W3C, where Web recommendations are produced, IETF continues to publish standards for Web identifiers.

¹¹ Kunze (1995).

¹² See Hayes and Halpin (2008), Halpin (2009), Halpin and Presutti (2009).

¹³ Monnin (2011).

work towards maintaining the access to a given set of representations. The issue at stake was to understand how URIs refer to resources and to find a suitable explanation accounting for their referential stability.

It looks like we currently lack an explanation of the URI/resource binding. But this seems us to be a profoundly misguided way of begging the question. Indeed, at first sight, it seems to be the case that URIs, save for access, behave like philosophical proper names. However, the remainder of this paper will be dedicated to showing that after close examination such an assumption cannot be taken for granted.

Resources as abstract artifacts.

Let us begin by reminding the reader how Web architecture defines resources. A resource, says RFC 2396, can be just about anything: the homepage of *The Guardian*, Tim Berners-Lee, founder of the Web, the number of married people in the US, a square circle, etc.

The URI directly identifies a resource which, in turn, can be (among many things):

- The rigid designation of “the Moon” (Kripkean style)
- The current satellite of Earth (Russellian definite description)
- One of the entities to which we have not present access or knowledge about (an “indefinite description” to quote Pierre Livet¹⁴).

Each time, we find a different way to identify or pick up an object. We follow eminent Web architect Roy Fielding who states that identification, picking up an entity, does in fact give the true account of the resource. A resource, in Husserlian terms, is the intentional act of picking up something, and by doing so, aiming at an object. It has a content (the object identified) and a form (the action of identifying something). The distinction at stake is reminiscent of the *hulè/morphè* distinction in Husserl’s *Ideen*. Unfortunately, the Husserlian vocabulary is tied to a somewhat mentalistic approach to the mind that is not entirely suitable to explain a *socio-technical system* like the Web (something outside of the scope of Husserl’s phenomenological investigation up until his later books, particularly the *Krisis* and the *Origin of geometry*).

Another way of putting things would be to conceive of a resource as a *rule for identification*. It presents the advantage of allowing for different ways of identifying an object. In the example above, the rule can be of a Russellian nature (“the current King of France” is relatively similar to “the homepage of the Guardian” yielding different – including the possibility of no – results over time) or a Kripkean one (“the Moon”¹⁵) – among infinite possibilities. Anyone is entitled to choosing any rule. When the standards explain that a resource can be anything, this is precisely what they mean: this choice is completely free. We’re led back to Roy Fielding’s definition, undoubtedly the most precise ever given. The Web Architectural style REST (for Representational State Transfer) he authored¹⁶ indeed defines “a resource to be the semantics of what the author intends to identify, rather than

¹⁴ “Web ontologies as renewal of classical ontology”, to be published in a forthcoming special issue of *Metaphilosophy* dedicated to the Philosophy of the Web.

¹⁵ Though, properly speaking, in Web architecture there is no room for possible worlds. This, however, is not sufficient to prevent the institution that mints a URIs from referring in a way that presupposes possible worlds.

¹⁶ Fielding (2000) & Fielding & Taylor (2002).

the value corresponding to those semantics at the time the reference is created.”¹⁷. Precisely what gives enough room to distinguish between a rule and the result of its *application(s)* at a specific time (here, the date a resource was created).

This is not to say resources have no properties of their own: they’re also arguably *abstract*, the way a concept is. This is central to Fielding’s account of resources in REST. REST provided the Web its *post hoc* theory. Its influence on Web standards is not just a known fact but what made possible the transition from the first standards of 1994-1995 to those of 1998 where resources were first defined. Fielding’s definition is the first hint that resources have got some specific properties distinct from those of its representations. This is generally not well understood and standards bear the mark of this difficulty. Especially when the existence of “physical resources” such as chairs, rocks or even online documents (collections of bits) is assessed. How, then, are we to make sense of this dual notion, torn between conflicting requirements?

An entity is identified in contrast with some features of resources. If we are to distinguish between two sets of properties, those of the entity identified and those of the resource, Edward Zalta’s account of fictions¹⁸ might prove immensely useful. He indeed distinguishes between properties that are *exemplified* or *encoded* by a fiction.

For instance, Sherlock Holmes is as much a well-known drug addict whose genius elder brother, Mycroft, works for the Queen, as he is the creation of Conan Doyle, **his brother** mentioned **—** according to Wikipedia **—** in four stories authored by Doyle ("The Greek Interpreter", "The Final Problem", "The Empty House" and "The Bruce-Partington Plans"), **etc.**

Similarly, it seems plausible to split resources in two comparable (which is not to say identical) sets of properties, as illustrated in the figure below:

Exemplification	Encoding
<p>Type properties shared by all resources:</p> <ul style="list-style-type: none"> - Abstract - Published at a given date, - Published by a given person, institution, etc., 	<p>Type properties of a rule.</p> <p>The way an entity is identified: through rigid designation, definite descriptions, etc. The result will change according to the chosen rule.</p>
<p>Token properties of the above types:</p> <p>What distinguishes sets of exemplifying properties of resources from each other’s.</p>	<p>Tokens properties of a token resource.</p> <p>The elements according to which an entity is identified. Properties linked to the content can be used to formulate any rule of identification.</p>

To better distinguish between the two, Zalta has chosen to vary the direction of predication. *Fx* designates exemplifying properties while *xF* is used for encoding one. This helps to ground the

¹⁷ Fielding & Taylor (2002), p. 135.

¹⁸ Zalta (2003).

data/metadata distinction in ontology and the more intuitive notion that while my resource may be Tim Berners-Lee and may thus encode the property of being a man, English, creator of the Web, etc. (xF), Tim Berners-Lee himself is no resource (Fx). Like a fiction, a resource can be anything *and* an abstract artifact (to borrow an expression from contemporary metaphysician Amie Thomasson¹⁹) at the same time. We may now make more sense of the little conundrum to which standards had no answers for: all resource are abstract (type exemplifying properties) and yet a given resource may be physical as well (token encoding property).

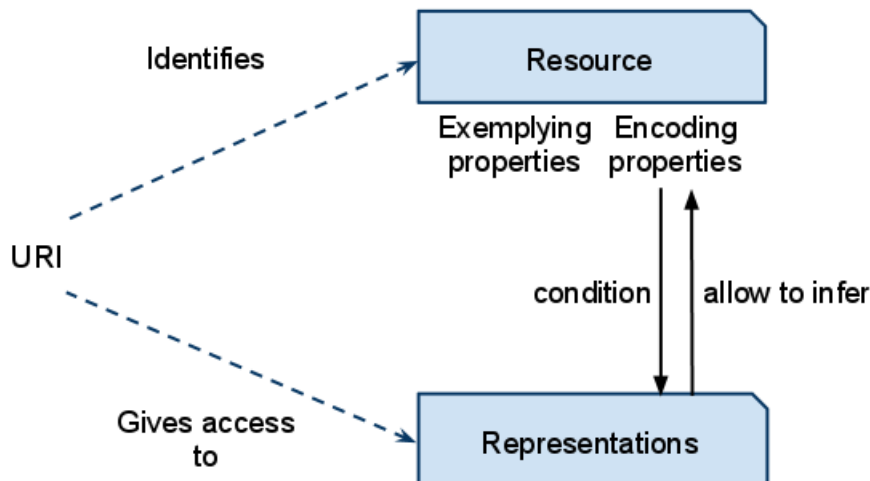


Figure 1 Once the two sets of properties of a resource are clearly separated, it becomes easy to expand the Webarch picture with additional ontological details.

III. The demise of reference under mutual adjustment between identification and access.

I hereafter understand ontology as being what I call the *science of reference* or a *theory of objects*. It is always useful to mention that the word “ontology” appeared in the XVIIth century, thanks to Jacobus Lohardus and Rudolphus Goclenius. Two thousand years had passed since Aristotle’s *Metaphysics*. The shift is perceptible, especially with Francisco Suarez²⁰, from an understanding of Being dominated by the divine **and/or substances (onto-theology)**, to a more general definition of Being as *aliquid*, something in general, that needs not be actual. Actuality being no longer an necessary condition, Being came to be understood as the possibility of an object, its conceivability on the mind. Between actual objects and mere fictions (*ens rationis*) a new realm was discovered: metaphysics thus turned into the *a priori* science of the possible (*ens reale*). The problem is that for objects not to fall into fictions, they had to be properly distinguished. Suarez’s answer relied on the principle of non-contradiction: for a truly thinkable object to merely be, its very concept had to be free of any contradiction. This is exactly what later led to Kant’s critique of logical possibilities,

¹⁹ Thomasson (1999)

²⁰ Courtine (1990).

understood in terms of compatibility of concepts, in the name of his own solution that dealt with the transcendental possibility of things through the synthesis of the knowing mind.

Later, the question known as the “problem of representations without objects”, raised by Bernard Bolzano (how do you distinguish between the square circle and the golden mountain), had a tremendous influence on thinkers such as Brentano, Twardowski, Meinong and Husserl. After all, it created a breach between what is thinkable, and the possibility of an object. Meinong’s answer was to explore those very possibilities which Suarez had avoided: to include fictions and impossible object (impossible to conceive) in order to somehow free objectivity from conceivability. This became known as the theory of (non-existent) objects (*Gegenstandstheorie*). Ontology itself was transcended by something broader than a description of the furniture of our world, even broader than a science of the possible. Something that encompasses the impossible itself: a theory of objects. And what those objects lacked in reality and possibility they made up in identity. How? Because it was still possible to *refer* to such objects as they came to be though as correlates of naming. What is remarkable is that objects were thus free from actuality and possibility. Naming, positing a referent, was acknowledged as the most fundamental (and trickiest!) ontological operation. Husserl later reframed the whole issue but that is outside of the scope of the present study.

The Web can be perceived as a technological equivalent of a *general system of reference*, owing to the paramount importance of URIs. This is also why the ontological question of the status of the referent is so prevalent. But unlike traditional ontology, where objects were substances, and theories of objects, which had no regard to technology, we will see that the Web deals with *non-natural*, even *technical* objects, with the surprising consequence that reference (and thus naming) is no longer a suitable concept in this context.

a) The Ontology of resources as a theory of *non-natural* objects:

Resources, it has been said, can be “anything at all”. This is strongly reminiscent of the ontological notion of an object: “something in general” (*aliquid*). The difference here, and this is why we need to taking into account the technological aspect is very important, is that objects (resources) on the Web are no longer *mental*. They have been *artifactualized*. One can consider them intentional object, but only the way books by contrast to mental states are intentional since books are *published* and thus follow a very complicated chain comprising events, people, institutions, book shops, etc. Hence, instead of intentional objects we prefer to speak of *institutional objects*.

Resources are also context-dependent, not unlike facts in Latour’s account (see his *Science in action* for instance). We can establish a very good analogy between both: to produce a fact as science does, requires that a statement be reused by as many people as possible while, at the same time, remaining stable. The same goes for resources. Either through tagging and hyperlinking or with RDF (the knowledge representation language of the Semantic Web), resources are just nodes in a graph whose identity depend on their ever-changing position inside that same graph since anyone can say anything about anything. As referent, they’re *not* context independent – and the important point is precisely here to mix ontology and epistemology otherwise one would end up with some form of nominalism as in Saussure’s *arbitraire du signe* whereby signifiers are distinguished only through differences and the position it entails: semiotics always had a dissolving effect on ontology. If we don’t want to lose the referent and end up with mere signs or representations because only the

latter can traditionally be treated as context-dependent, we must attribute other identity conditions to referent (resources) *themselves*.

Finally, there is the importance of *trust*: the Semantic Web is no mechanism to help one decide how to link words to the world (reference!) or how to end up with a well-founded theory of truth. After all, it is dominated by *trust*, the highest layer on the famous Semantic Web “cake”²¹. Take Dbpedia²² for example, the Semantic Web’s most successful application that semantize Wikipedia and treats its entries as entities about which we have knowledge. Entities found in the semantic version of Wikipedia are always the result of the contributions of thousands of users all around the planet. The latter is subject to peer-review, contradiction, improvement, re-writing, etc. It’s no longer “given”. Only once the activities that support our ontology are negated and forgotten, can it appear natural (in a very paradoxical way: here it is the artifactual activity of machines that produces naturalness). This is partly what currently happens with Dbpedia: human discussions as well as machines extractions disappear from the final product, giving it an uncanny pristine appearance that doesn’t corresponds the reality of the complex and muddled lifecycle that made it possible. These entities or resources are always the result of collective choices and evaluations.

Recently, Halpin, Hayes, McCusker, McGuinness and Thompson²³ have shown that many resources that seemed identical, already distinguished entities that only need to be picked-up, are in fact different under close examination (of course, using our framework, we could always distinguish resources by their *token exemplifying properties* but let’s just assume this was not possible for the sake of the argument). For instance, the sodium element both on the CYC²⁴ knowledge-base and on Dbpedia appear to genuinely correspond to the same entity or referent. However, they do not share the same definition (*token encoding properties*)²⁵. Sometimes the problem will seem circumscribed with regards to a common reality beneath these “representations” and distinct from them, well-established enough to warrant that dissimilarities are only symptomatic of a lack of precision. Two and a half millennia of philosophy should temper such optimism – also because this so-called underneath level, call it nature or society depending on your discipline, has always proved quite hard to find.

Does it mean that all we need to get rid of all the problems we’re facing on the Web is a good social epistemology? We’d rather keep up with the ontological inquiry, not lose the worlds, and go back to the Latin language for whom, as French historian of philosophy Jean-François Courtine puts it,

²¹ Many versions of it have been proposed, the canonical one now seems to be: [http://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/#\(24\)](http://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/#(24))

²² <http://dbpedia.org/About>

²³ Halpin & al (2010).

²⁴ CYC is a longstanding AI project run by Douglas Lenat first at MCC then at Cycorp, whose goal is to construct “a foundation of basic “common sense” knowledge--a semantic substratum of terms, rules, and relations--that will enable a variety of knowledge-intensive products and services” (see: <http://www.cyc.com/cyc/technology/whatisicyc>).

²⁵ Ibidem: “In particular, this issue comes into play when different agents describe the world at different levels of granularity. For example, different sources of Linked Data may make subtly different claims about some common-sense term like ‘sodium.’ This occurs in the case of the concept of sodium in DBPedia, which has a sameAs link to the concept of sodium in OpenCyc. The OpenCyc ontology says that an element is the set of all pieces of the pure element, so that sodium in Cyc has a member which is a lump of pure metallic sodium with exactly twenty-three neutrons. On the other hand, sodium as defined by DBPedia includes all isotopes, which have different number of neutrons than ‘standard’ sodium, and in this particular case are unstable. So, one should not state the number of neutrons in DBPedia’s use of sodium, but one can with OpenCyc. At least in web settings with little inference or reliance on detailed structures, it is unlikely that most deployers of Linked Data actually check whether or not all the properties and their associated inferences are shared amongst linked data-sets.”

substantia (the translation of the Greek *ousia*) was the answer to the question *an sit?* “How about the reality of a given fact?”, “can we make something out of it?”, “is there something sure and solid (*res certa et solida*)?”. What Aristotle, after Plato named in terms of presence (*ousia*), became the topic of debate, a moot point, a fact in dire need to be established. Latin rhetoric explicitly **aimed** at convincing (*facere fidem* – to produce trust), stabilizing, what **was** always given as a matter of doubt (*Res dubia*) at first.²⁶ Following Bruno Latour we should say that realities (instead of only representations) are fallible and need to be constantly adjusted²⁷.

b) The cost of maintaining a resource: why reference once artifactualized is no longer truly reference.

Substances used to hold by themselves.

“A “resource””, writes Tim Berners-Lee, “is a conceptual entity (a little like a Platonic ideal).”²⁸ . More than a clarification *per se*, this sentence indicated that its exact status remains to be investigated. Berners-Lee never really explains this reference to Plato. Furthermore, one would be in dire pain to find a proper theory of Ideas in Plato’s writings. That is why we will now turn to Jules Vuillemin’s *a priori* deduction of philosophical systems²⁹, including Plato’s idealist position, from what he calls the “scheme of pure predication” (“linguistic universals” in opposition to “perception universals” as in Aristotelianism).

In nominal sentences such as “Humility is a virtue”, sentences that asserts conceptual truths, explains Vuillemin, two linguistic universals are associated, one positioned as an argument, the other one as a function to use Frege’s terminology. In both cases this kind of predication (but can we really speak of predication and at the same time mobilize Frege’s function/argument dichotomy?) is not liable to change – such states of affairs will not vary since their arguments are neither located in space nor time and as such are completely intangible. Such predicative functions then hold or don’t independently from circumstances. For the same reason, nothing is displayed to the sense, not even the referent. To borrow from French linguist François Rastier, the Idea belongs to the distal anthropic zone. In the same fashion, the Web might be seen as a technical device aiming at bringing the distal (the identified resource), what is never accessible, under the guise of the proximal (the accessed representation). Since no translation is available that would make these entities break out from their respective zones, doing so makes it mandatory to pick up entities from a zone we can have access to, the proximal one. **The gap between the resource and its representation is thus never overcome. Rather, what important is not so much the so-called gap between heterogeneous zones but the mediation that takes place through makeshifts** (descriptions, information realizations, etc.). From this point of view, Berners-Lee was not so wrong to identify resources to a Platonic idea rather than an Aristotelian substance, despite the rather obvious analogy between resources and substances on the one side and accidents and http-representations on the other side.

Yet, the issue remains to determine what ties the core-content to what gets attached to it (substance and accident with regards to the canonical objects of the Western metaphysical tradition, resources

²⁶ The remainder of this paragraph is adapted from Courtine (2003): “Les traductions latines d’ΟΥΣΙΑ et la compréhension romano-stoïcienne de l’être”.

²⁷ Latour (2007).

²⁸ Berners-Lee (1996).

²⁹ Vuillemin (1986).

and representations in Webarch) – the difficulties are of a very different nature in both cases. A) How, on the one hand is the rule, in other words the content of a resource, being communicated to those who ignore what a URI is supposed to identify and only have a *de facto* access to potentially constantly changing representations?³⁰ B) On the other hand, as long as the resource doesn't change, it nevertheless undergoes modifications, but internal ones. How then are we to define a rule and all its applications *ab initio*? The stability of URIs is directly tied to this capacity. Apache Software Foundation President Justin Erenkrantz³¹ is right to equate resources to “network continuity” in his PhD thesis, though it doesn't explain *how* this continuity of subsistence is produced insofar as we chose to abstain from adopting substances as an explanation – a notion whose value resides precisely in its capacity to beg the issue at stake (as a metaphysical concept, its specific « agency » instead of making things hold together like ontological cement might simply be to prevent the real problem from coming under scrutiny).

Nelson Goodman's paradoxes of predicate projectibility described in *Facts, Fiction, and Forecast* here come to mind. To infer the content of a resource from a finite number of representations (or what we may call local “projections” on a screen) is not very different from the proverbial problem of induction. Projectibility paradoxes were later reactivated from a wittgensteinian perspective as rule paradoxes by Saul Kripke (*Wittgenstein on Rules and Private Language*). This undoubtedly makes him the philosopher whose works were the most influential on philosophical engineering³² – and thus on the Web, from the idea of “baptism” to the paradox of rules (rigidity is another matter since the Web was not conceived with possible worlds in mind but rather from a universalist point of view that is reminiscent of Frege and the early XXth century logicians as opposed to algebraists like De Morgan, Boole or Schröder who put forward the notion of a universe of discourse, prompting such thinkers as Jean van Heijenoort and later Jaakko Hintikka to draw a distinction between “logic as language” and “logic as calculus”).

Many Web architecture discussions borrow from Kripke's idea of baptism but only in order to underline the fact that the publisher of a resource is to decide of its “content”. This, in a sense, is very true. After all, *access* is what first and foremost distinguishes URIs from philosophical proper names. Apart from URIs understood as common names (the string of character considered as meaningful as in <http://www.w3.org/People/Berners-Lee/>) you only get a glimpse of a resource by being acquainted **causally** to its representations. Users can only fathom the meaning of a resource from its representations. This opens the door to Kripke's skeptic **paradox described in his later work and exemplified** in his famous “quaddition” argument: one can infer that a resource is X from its various representations until one gets a result that no longer seems consistent with the rule first postulated (and for which there is no standard way on the Web to make it explicit).

Rules and resources as virtual trajectories.

As a consequence, the issue at stake concerns not only the synthesis which binds resources and representations, but also the way it is realized and its cost. This includes understanding how the rule/resource constrains the supply of representations and, in turn, how obstacles to the application of the said rule, or the fact that it may be put on hold, lead to modifying it in return.

³⁰ See (Booth 2007b) et (Halpin 2008).

³¹ Erenkrantz (2009).

³² See, on this notion, (Halpin 2008b) and, in French, (Monnin 2012).

The framework recently proposed by Pierre Livet and Frédéric Nef for the analysis of social beings³³ offers a number of conceptual tools to think about such syntheses. Its starts as ontology of *processes* bearing on a coupling of the actual and the virtual – the startup stage of a process going from the actual to the virtual, to end up with an actualization of the virtual and a virtualization of the actual.

In addition to processes, the two authors introduce the notion of *quality*, whose peculiarity is to rest at the crossroads between the epistemic and the ontological; whence the impossibility to describe it through a single process. Qualification is indeed defined as the articulation of two processes, the second being the qualifying one: “The first corresponds to what is always actual whereas the second one is what binds this actual to the virtual”³⁴. The first is the actualization of the qualification expected from the second one: “In whatever sense we take things, a qualifying process requires the following coupling: the introduction of an expectation, a virtuality, and the accomplishment of this virtuality by another process so that coexists in actuality the process of expectation or initial reception of another process and the accomplishment of this virtuality by the second process.”³⁵

The expectation, here, is due to the identified rule/resource, a rule whose application constrains representations to be accurate. The process with which this second “qualifying process” (noted “Re” for “resource” regarding its virtual aspect and “StRe” for “state of the resource” with regards to its actual aspect) is assembled is one that allows to generate representations of the resource (noted “Dref”, for “dereferencing”). Representations derive their quality from the resource, understood *as a requirement of the virtual bearing on the actual*. Qualification results from an actualization of the virtual and a concomitant virtualization of the actual. It entails a tight *coupling* between the two processes that will prove crucial for our investigation.

Following the notation³⁶ used by Livet and Nef to represent the coupling of processes, we would get:

$$\begin{array}{l}
 \text{Dref } V_1^1 \quad \text{Dref } A_2^1 \\
 \mathbf{Dref } A_1 \rightarrow (\text{Dref } V_2^1) \quad \textit{StRe } A_1^2 \\
 \textit{StRe } A_2^2 \rightarrow \textit{Re } V_1^2
 \end{array}$$

Such an approach sheds light on the canonical examples of the W3C Technical Architecture Group, used to explain the difference between a resource and its representations:

“While planning a trip to Mexico, Nadia reads “Oaxaca weather information: 'http://weather.example.com/oaxaca'” in a glossy travel magazine. Nadia has enough experience with the Web to recognize that "http://weather.example.com/oaxaca" is a URI and that she is likely to be able to retrieve associated information with her Web browser.

(...)

Dirk would like to add a link from his Web site to the Oaxaca weather site. He uses the URI <http://weather.example.com/oaxaca> and labels his link “report on weather in Oaxaca on

³³ Livet & Nef (2009).

³⁴ Ibidem.

³⁵ *Ibid.*

³⁶ The arrow stands for the coupling typical of a process; the characters between parentheses, the aspect that is going to be replaced; in italic, the aspect that will take its place; in bold, the process qualified; the subscript indicate the initial (1) or final (2) aspect of a process; the superscript, process number 1 or 2. “V” stands for Virtual, “A” for Actual.

1 August 2004". Nadia points out to Dirk that he is setting misleading expectations for the URI he has used. The Oaxaca weather site policy is that the URI in question identifies a report on the current weather in Oaxaca—on any given day—and not the weather on 1 August. Of course, on the first of August in 2004, Dirk's link will be correct, but the rest of the time he will be misleading readers. Nadia points out to Dirk that the managers of the Oaxaca weather site do make available a different URI permanently assigned to a resource reporting on the weather on 1 August 2004."³⁷

The first resource being a daily report of the weather in Oaxaca in contrast with a report of the weather in Oaxaca on 1 August 2004, Nadia's expectations are very different: she knows that actual representations are constrained on a virtual level by the resource. Any two apparently similar representations, if states of different representations, are in fact *moments belonging to heterogeneous virtual trajectories* that are part of their respective identities. Hence, despite the actual outward similarity between the two, they are in fact completely unlike – this becomes obvious once our gaze is no longer solely focused on the actual (similar remarks could be made with regards to a range of cases, including mirror representations hosted on a server with a different domain name for instance).

The (ontological and technical) coupling between resources and representations means URIs do not refer.

Such a coupling also makes it possible to conceive of the dependency between the rule and the results of its applications. Widespread (but not in any way less troubling) phenomena, often shunned in standards owing to their normative grasp of the Web, may finally get an explanation, starting with "non-cool URIs", in other words, URIs that do change contrary to the stability requirement imposed on "cool-URIs"³⁸ – a somewhat inappropriate name since, all things considered, URIs never change unless they disappear or are discarded in favor of new ones. Only the URI/resource pairing may evolve over time³⁹, for the simple reason that the resource (rule) identified by a URI will have underwent modification first. As Livet and Nef put it, it is indeed impossible

"to foresee all the obstacles to the application of a rule and anticipate all the subsequent right revisions they entail. The spirit of the rule only makes sense in retrospect, when an obstacles lead to revising the rule in a satisfying way, albeit not one that complies with previous trajectories. Prospectively, we cannot pretend to know in advance how to follow a rule for every new situation that presents obstacles. It is because such forward-looking knowledge is out of reach that rules seem overhanging. But this, in a sense, is a mere appearance for the application of the rule rests upon past successes, thus only in retrospect."⁴⁰

³⁷ Jacobs & Walsh (2004).

³⁸ Berners-Lee (1998), Sauermann & Cyganiak (2008).

³⁹ In a recent communication (Arwe 2011), John Arwe mentions four of them:

1. "Needs evolve: successful proof-of-concept systems are pressed into wider use; department-level systems grow into corporate systems with different quality of service needs that require different deployments.
2. Domain name ownership changes. Acquired organizations' names are retired, and there is a (small but non-zero) economic incentive to release unneeded domain names. Sometimes there is a legal requirement to do so.
3. Some URLs are poorly constructed to begin with, in that they include components that lead people to want to change them over time. They include brand names, or version numbers, that are really mutable properties of the identified resource. Some organizations are simply used to being able to change URLs because their current consumers are human-attended user agents rather than fully automated/autonomous processes.
4. Organizations out-source control of their network environments; exerting control of the sort required to control DNS entries or issue 301 redirects essentially becomes a legal process, ill-suited to solving technical problems."

⁴⁰ Livet & Nef (2009), p. 201.

Once seen as a rule, and analyzed from the perspective of a coupling of processes associating the actual and virtual dimensions, resources can no longer be conceived outside of their representations: the latter only make sense against the rule that gave them birth, as well as they may modify it according to the borderline cases that force to renegotiate the application of the **said** rule. Many a creator of conference website know that after the first installment, it may become necessary to turn the original resource (for instance PhiloWeb 2010 symposium homepage) into something more generic (PhiloWeb symposiums homepage), should it prove somewhat successful. The “ontologisation” of the rule/resource that an analysis in terms of actual and virtual modalities warrants, offers an explanation that escapes the Kripkean skeptic paradoxes and could help to find a more careful treatment of such phenomena as non-Cool URIs and mutable resources.

Let us add that for users, and probably for a majority of the institutions that publish a resource, representations provide (with **the exception of** connotative URIs, URIs that function not only as proper names but also common ones), though there is no consensus regarding the way such URIs should be written to avoid obsolescence⁴¹) the most efficient means to infer its content. This, as we have seen, has a cost. As Brian Cantwell Smith⁴² once said, “*reference to succeed doesn’t need adjustment to its target*”. On the Web we witness exactly the contrary. Only URIs with no dereferenceable function may be said to refer although one may rather consider that they simply do not identify a resource, nor give access to its representation – which is admittedly quite different. We may call the aforementioned adjustment the *editorial commitment* made by the publisher of a resource to ensure “network” (and service⁴³) “continuity”, to borrow Justin Erenkrantz’ definition of the resource⁴⁴.

Nowhere do we need to ask ourselves whether or not URIs refer to something permanently and how. URIs do *not* refer for the aforementioned reasons. We rather publish Web resource identified by the latter and, if given enough resources (in the traditional meaning of the word), then we maintain a positive feedback loop between Web resources and accessible representations; all in all, a very different story.

URIs are undoubtedly the result of the artifactualization⁴⁵ of proper names. “Web proper names”, to quote Thompson and Halpin⁴⁶, are no longer conceptual, philosophical or semiotic objects, but rather *technical* ones. The consequences of this simple yet decisive truth are yet to be properly measured. While regular (philosophical) proper names may possess the distinctive feature to refer if used accordingly, neither identification nor access on the Web have to do with reference. In other words,

⁴¹ It is even quite the contrary as a result of the much discussed “axiom of opacity” which states that the “meaning” of a URI should only be inferred from dereferenced representations instead of any connotative aspect of the string of characters that constitutes the URI itself.

⁴² “Reference and Identity on the Internet”, presentation given at the “Philosophy of the Web” seminar organized **by** the author and Harry Halpin at La Sorbonne on January 28 2012.

⁴³ Ensuring a continuity of service constitutes no less than a necessary condition of the dereferencing process. Often monitored by third-parties (instead of the institution that published the resource and guarantees that compliance of the representations is implemented in the long run), thus adding an additional line of expenditure when it comes to summarizing the efforts required to maintain the resource over time – or, to be more precise, the coupling between processes of dereferencing and qualification that we have previously analyzed.

⁴⁴ The lack of substances brings this very issue to the foreground: “why do things subsist? Once [enduring] substance has been excluded, subsistence comes to the fore, and then the big question is how many ways there are for the entities to graze their subsistence in the green pastures.” Latour & al. (2011), p. 48.

⁴⁵ For an introduction to this concept, see (Monnin 2009). More recently, Luciano Floridi has been using a similar line of argument.

⁴⁶ Halpin & Thompson (2005).

the artifactualization of proper names is tantamount to replacing reference with other (technical) processes, coupled to one another. Thus the explanation of the binding between URI and resources rests upon entirely new principles. The issue at stake is no longer to *point* at or *designate*, but rather to *maintain* a coupling between two kinds of processes through socio-technical means⁴⁷.

Conclusion.

Despite the above reasons, according to the title of this paper, URIs do not *always* refer. Indeed, the Semantic Web foundational language, RDF, a knowledge representation syntax, functions as an additional layer to the existing pile of standards that govern the Web. According to RDF and its semantics⁴⁸, URIs *are* indeed proper names, interchangeable props or tags with no meaning whatsoever. As a corollary, URIs once moved to this context do indeed revert to the perennial definition of proper names in logic (and, we may add, in philosophy, though the emphasis then is less on fining an “interpretations controlled by the pure semantic power of the axioms that use them” – see below). Therefrom, it can be argued that URIs keep referring since from the perspective of RDF all the previously observed intricate details just vanish into the background. What we must now think are the different semantic (and ontological) commitments across the layers⁴⁹ formed by the heap of accumulated standard, formal languages, logics, that characterizes the Web. Equally necessary is a theory describing how such layers *assemble* and how the properties of objects shift from layer to layer, from logical proper name to genuine Web proper names – and back, for instance. No one better described the situation with regards to names and URIs than Patrick J. Hayes in a keynote where he advanced the idea of Web logic or “Blogic”. Let us quote him at length:

“Names are central in blogic. They are global in scope. They have structure. They link blogical content to other meaningful things, including other blogical content. They embody human/social meanings as well as being conduits and route maps for information transfer. In many ways, the Web is constituted by the links which are the blogic names, and the logical content which we write using those names is only one component, perhaps a minor one, of the whole social and technical structure which determines their meanings. And yet seen from the perspective of the logic, these IRIs are merely “logical names”, elements of an arbitrary set of meaningless character strings. In AI/KR, we teach our students that the names are irrelevant, because one can replace them all with gensyms without changing the logical meaning.

Clearly, there is something unsatisfactory about this picture, a serious disconnect between the classical logical view of names as simply uninterpreted strings waiting in a kind of blank innocence to have their possible interpretations controlled by the pure semantic power of the axioms that use them, and the reality of the almost unrestricted referential power that these names actually have in the dynamics of the Web. *Think of the concern and attention that is devoted to their choice, who owns them, who is responsible for maintaining and controlling them, and the ways they are decomposed and used in the planet-wide machinery called the Internet,*

⁴⁷ The fact that reference is no longer the issue sits well with the Web’s reluctance to deal with the notion of truth. As already said, the epistemology of the Web is one of trust. Content providers, including resource publishers, must thus ensure that the definition they give of a resource (its encoding properties) are *trustful*. See also Henry Thompson, member of the TAG, who has dedicated a lot of thought to the analysis of URI persistence: “persistent identifier efforts can and should save *huge* amounts of fuss by focussing (sic) on the *non*-technology substrate issues involved in producing persistence” (Thompson 2007).

⁴⁸ <http://www.w3.org/TR/rdf-mt/>

⁴⁹ Specifically referring to RDF, Patrick Hayes called this problem “Death by layering”, thus summarizing the issue at stake in a most fitting way: “names have a different logical status at different levels”. What’s true within the framework of the Semantic Web is all the more true within the broader framework of the Web itself here examined.

none of which has very much at all to do with logical assertions. Another way to put it: IRIs are *identifiers*, not mere logical names. Unfortunately, nobody seems to be able to say what in God's name that can possibly mean. [our emphasis]"

In a sense, this paper can be construed as an attempt to shed some light on this conundrum by looking directly at the "the whole social and technical structure which determines [URIs] meanings".

References:

- Arwe, J. (2011), "Coping with Un-Cool URIs in the Web of Linked Data".
http://www.w3.org/2011/09/LinkedData/ledp2011_submission_5.pdf (last accessed January 31, 2012).
- Berners-Lee, T. (1998), "Cool URIs don't change". <http://www.w3.org/Provider/Style/URI> (last accessed April 23, 2011).
- Courtine, J.-F. (1990), *Suarez et le système de la métaphysique*. Presses Universitaires de France - PUF.
- Courtine, J.-F. (2003), *Les catégories de l'être : Etudes de philosophie ancienne et médiévale*. Presses Universitaires de France, Paris, PUF.
- Erenkrantz, J. R. (2009), "Computational REST: A New Model for Decentralized, Internet-Scale Applications". PhD Thesis. University of California Irvine.
- Fielding, R. T. (2000), "Architectural Styles and the Design of Network-based Software Architectures". PhD Thesis. University of California, Irvine.
- Fielding, R. T. & Taylor R. N. (2002), "Principled design of the modern Web architecture". *ACM Transactions on Internet Technology (TOIT)* 2(2): 115–150.
- Floridi, L. (2005), "The Ontological Interpretation of Informational Privacy". *Ethics and Information Technology* 7(4): 185-200.
- Halpin, H. (2008), "The Principle of Self-Description: Identity Through Linking". *Proceedings of the 1st IRSW2008 International Workshop on Identity and Reference on the Semantic Web*, Paolo Bouquet and al. (ed.), Tenerife, Spain. CEUR Workshop Proceedings. <http://ceur-ws.org/Vol-422/irsw2008-submission-13.pdf>.
- Halpin, H. (2008b). Philosophical Engineering: Towards a Philosophy of the Web, *APA Newsletter on Philosophy and Computers*, 7(2).
- Halpin, H., (2009). *Sense and Reference on the Web*, PhD Thesis. Institute for Communicating and Collaborative Systems, School of Informatics, University of Edinburgh.
<http://www.ibiblio.org/hhalpin/homepage/thesis/>
- Halpin, H. & Presutti, V. (2009), "An Ontology of Resources: Solving the Identity Crisis". *The Semantic Web: Research and Applications*, Aroyo & al. (ed.), Springer-Verlag Berlin, Heidelberg, p. 521–534.
- Halpin, H. & Thompson, H. S. (2005), "Web Proper Names: Naming Referents on the Web". Chiba, Japan. <http://www.instsec.org/2005ws/papers/halpin.pdf> (last accessed May 25, 2009).
- Halpin, H. & al. (2010), "When owl:sameAs Isn't the Same: An Analysis of Identity in Linked Data". *The Semantic Web—ISWC 2010*: 305–320.
- Hayes, P. J., (ed.) (2004), "RDF Semantics (W3C Recommendation 10 February 2004)".
<http://www.w3.org/TR/rdf-mt/> (last accessed February 20, 2012).
- Hayes, P. J. (2009), "BLOGIC or Now What's in a Link?"
http://videolectures.net/iswc09_hayes_blogic/ (Last accessed February 20, 2012).

- Hayes, P. J., and H. Halpin (2008), "In defense of ambiguity". *International Journal on Semantic Web & Information Systems* 4(2): 1–18.
- Husserl, E., *passim*.
- Jacobs, I. & Walsh N. (2004), "Architecture of the World Wide Web, Volume One (W3C Recommendation 15 December 2004)". <http://www.w3.org/TR/webarch/#formats> (last accessed February 1 2009).
- Koepsell, D. R. (2003), *The Ontology of Cyberspace: Philosophy, Law, and the Future of Intellectual Property*. New edition. Open Court Publishing Co, U.S.
- Kunze, J. (1995), "RFC 1736 - Functional Recommendations for Internet Resource Locators". <http://www.rfc-editor.org/rfc/rfc1736.txt> (Last accessed February 20, 2012).
- Latour, B. (2007), "Quel cosmos ? Quelle cosmopolitiques" in Lolive, J. & Soubeyran, O. (ed.) *L'émergence des cosmopolitiques*, Colloque de Cerisy, Recherches, La Découverte, Paris, pp. 69-84.
- Latour, B., Harman, G., Erdelyi, P. (2011), *The Prince and the Wolf: Latour and Harman at the Lse*. Zero Books.
- Livet, P. & Nef F. (2009), *Les êtres sociaux : Processus et virtualité*. Hermann.
- Monnin, A. (2009), "Artifactualization: Introducing a new concept". Southampton, United Kingdom. http://hal-paris1.archives-ouvertes.fr/hal-00404715_v1/ (Last accessed September 21, 2009).
- Monnin, A. (2011), "La ressource et l'ontologie du Web" (to be published in *Intellectica*). <http://hal-paris1.archives-ouvertes.fr/hal-00610652> (Last accessed, February 5 2012)
- Monnin, A. (2012), "L'ingénierie philosophique comme design ontologique », in *Archéologie des nouvelles technologies, Réel-Virtuel : enjeux du numérique*, 3.
- Sauermann, L., & Cyganiak, R. (2008), "Cool URIs for the Semantic Web (W3C Interest Group Note 03 December 2008)", <http://www.w3.org/TR/cooluris/> (Last accessed February 1, 2009).
- Thomasson, A.L. (1999), *Fiction and Metaphysics*, Cambridge University Press.
- Thompson, H. S. (2007), "URIs and Persistence: How long is forever?" http://www.ltg.ed.ac.uk/~ht/UKOLN_talk_20070405.html (Last accessed May 31, 2010).
- Thompson, H. S. (2012), "An introduction to naming and reference on the Web" http://www.ltg.ed.ac.uk/~ht/PhilWeb_2012/ (Last accessed February 5 2012)
- Vuillemin, J. (1986), *What are philosophical systems?*, Cambridge University Press.
- Zalta, E. N. (2003), "Referring to Fictional Characters". *Dialectica* 57(2): 243-254.