Title	A URI is worth a thousand tags: From tagging to linked data with MOAT
Author(s)	Passant, Alexandre; Laublet, Philippe; Breslin, John G.; Decker, Stefan
Publication Date	2009
Publication Information	Passant, A., Laublet, P., Breslin, J. G., & Decker, S. (2009). A URI is Worth a Thousand Tags: From Tagging to Linked Data with MOAT. International Journal on Semantic Web and Information Systems (IJSWIS), 5(3), 71-94. doi:10.4018/jswis.2009081904
Publisher	IGI Global
Link to publisher's version	http://dx.doi.org/10.4018/jswis.2009081904
Item record	http://hdl.handle.net/10379/7231
DOI	http://dx.doi.org/10.4018/jswis.2009081904

IGI PUBLISHING
701 E. Chocolate Avenue, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.igi-global.com

ITJ 5386

This paper appears in the publication, International Journal on Semantic Web and Information Systems, Volume 5, Issue 3
edited by **Amit Sheth © 2009, IGI Global**

A URI is Worth a Thousand Tags:
From Tagging to Linked Data with MOAT

Alexandre Passant, National University of Ireland, Ireland

Philippe Laublet, Université Paris-Sorbonne, France

John G. Breslin, National University of Ireland, Ireland

Stefan Decker, National University of Ireland, Ireland

ABSTRACT

Although tagging is a widely accepted practice on the Social Web, it raises various issues like tags ambiguity and heterogeneity, as well as the lack of organization between tags. We believe that Semantic Web technologies can help solve many of these issues, especially considering the use of formal resources from the Web of Data in support of existing tagging systems and practices. In this article, we present the MOAT—Meaning Of A Tag—ontology and framework, which aims to achieve this goal. We will detail some motivations and benefits of the approach, both in an Enterprise 2.0 ecosystem and on the Web. As we will detail, our proposal is twofold: It helps solve the problems mentioned previously, and weaves user-generated content into the Web of Data, making it more efficiently interoperable and retrievable.

Keywords: *Enterprise 2.0, Linked Data, MOAT, SIOC, Social Web Tagging, User-Generated Content*

INTRODUCTION

The Social Web, or Web 2.0 (O'Reilly, 2005), has become an important trend during the last few years. While end-users of the Web were previously considered as being only consumers of content, the paradigms that the Social Web introduced has led them to become producers as well. For instance, blogs allow anyone to publish and share their thoughts on the Web whereas wikis are used to collaboratively build consensual information within a community. In the meantime, social network services have allowed people to define acquaintance networks and to keep in touch with each other on the Web. Moreover, apart from providing a means to create discussions and to define or manage social networks, an important feature of social Web sites is the ability to share content with one's peers. On many social Web sites, this data can be shared either with whoever is subscribed to (or just browsing) the Web site or else it can be shared within a restricted community. Also, not only textual content can be shared, but various types of media or other content objects: pictures (Flickr), videos (YouTube), slides (Slideshare), trips (Dopplr), and so forth. To make this content more easily discoverable,

DOI: 10.4018/jswis.2009081904

most of these websites allow users to add free-form keywords, or tags, that act like subjects or categories for anything they wish to share. For example, this article could be tagged with "semanticweb" and "socialweb" on a scientific bibliography management system such as Bibsonomy or Connotea.

Although tags can be generally considered as a type of metadata, since they provide additional information about a tagged item, it is important to keep in mind that they are user-driven. Indeed, while a blog engine may automatically assign a creation date to any blog post or a photo sharing service could use embedded EXIF information to display the aperture of a camera, tags are added voluntarily by users themselves. To that extent, they clearly reflect the needs and the will of the user who assigns the tags. In this way, tags focus on what a user considers as important regarding the way he or she wants to share and present information. The main advantage of tagging for end users is that one can use the keywords that fit exactly with his or her needs and they do not have to learn a pre-defined vocabulary scheme (such as a taxonomy). Tags and tagging actions lead to what is generally called a folksonomy (VanderWal, 2007), an open and user-driven classification scheme that evolves during time thanks to the tagging actions of the community itself, contrary to pre-defined and authoritative classification directories, which are generally fixed.

Yet, in spite of its advantages when annotating content items, tagging leads to various issues regarding information retrieval, which makes the task of retrieving tagged content sometimes quite costly. Mathes (2004) estimates that a "folksonomy represents simultaneously some of the best and worst in the organization of information." Indeed, even if dedicated algorithms like FolkRank (Hotho, Jäschke, Schmitz, & Stumme, 2006) and clustering techniques can be used to improve retrieval of tagged-content—in spite of the shortcomings we will discuss later—tag-dedicated search engines are generally simply based on plain-text strings, that is, a user types a tag and gets only the content that has been tagged with that particular keyword. Therefore, this can lead to various issues, since such an engine only considers a set of characters that it cannot interpret which consequently introduces some noise and silence issues.

In the Semantic Web domain, the Web of Data is considered a more pragmatic vision of the Semantic Web, focused mainly on exposing data in RDF and interlinking it, that is, providing Linked Data on the Web, rather than on using formal ontologies and inference principles that form the complete Semantic Web vision. Interlinking user-generated content with URIs of well-known and unambiguous resources from the Semantic Web would help to solve the aforementioned issues, as user-generated content would be then interlinked with well-defined and unambiguous identifiers. Moreover, it offers a way to weave such content into the Semantic Web, hence considering Web 2.0 and the Web of Data not as disjoint domains but as being beneficial to each other.

In this article, we describe the MOAT framework that aims to provide an intuitive and lightweight way to bridge this gap between free-tagging and Linked Data, in what we consider a twofold approach with strong benefits for both the Social Web and the Semantic Web communities. The article is organized as follows. In the first section, we describe some of the main issues of free-tagging systems regarding data querying and information retrieval. We also emphasize, based on a corporate-blogging use-case, why current tag-based clustering algorithms may not be enough to solve these issues. Then, we introduce our proposal, MOAT, beginning with its theoretical background in which we extend the usual tripartite model of tagging to a quadripartite one, taking into account the meaning of tags. We then describe the related OWL ontology and continue by reviewing the MOAT framework architecture, combining the "architecture of participation" principles of Web 2.0 together with Semantic Web technologies and RDF(S)/OWL data representation principles to let people intuitively bridge this gap between tagging and Linked Data. We then detail two use-cases for the approach. The first relates to the corporate blogging platform that initially motivated the MOAT approach. The second describes LODr, an application based on MOAT dedicated to weaving existing user-generated content from well-known services like Flickr or Delicious into the Web of Data. The analysis of these two use-cases helps evaluate the approach, both in terms of how it can be used to solve tagging issues and how it weaves user-generated content into the Semantic Web in a twofold approach. We then present an overview of related work and detail our position in relation to it, before concluding the article.

Common Issues with Free-Tagging Systems

In this section, we give an overview of current issues in free-tagging systems, based on some observations and an analysis both of the Web and of corporate blogging systems. Interestingly, the issues below have parallels in the world of libraries and are one reason why librarians use classification schemes like thesauri or taxonomies, such as the Dewey Decimal Classification or the ACM Taxonomy. Therefore, we may consider how to find a smoother transition between the openness of tagging systems and the rigidity of such classification schemes, and we will later describe how our proposal aims to solve this.

Tag Ambiguity

Since tags are text-strings only, without any semantics or obvious interpretation (rather than a set of characters) for a software program that reads these tags, ambiguity is an important issue. Although a person knows that the tag "apple" means something different when it is used to tag a blog post about a laptop, a picture of a bag of fruit, or a review of a Beatles record, a tag-based information system cannot distinguish between them. Indeed, the only thing it understands is that the content is tagged with a text string composed of the characters "a-p-p-l-e" in this particular order. Hence, a tag-based query engine will retrieve various items for a search on "apple" even if the user had the computer brand in mind: items about fruits will be mixed up with iPod-related ones. Consequently, users must sort out themselves what is relevant or not regarding their expectations. Depending on the context and the number of retrieved items, it can be a costly task.

For example, the following Figure shows the result of a search for the most relevant items tagged "apple" on Flickr, mixing pictures of fruits and Apple devices. Similar issues can be observed on Delicious, for example with the "swig" tag, since the acronym identifies both the "Semantic Web Interest Group" and "Simplified Wrapper and Interface Generator." Both are unrelated, but unfortunately a user subscribed to the related RSS feed have to face a noise and information overload issue, as they will be delivered unrelated content.

Tag Heterogeneity

Tag ambiguity refers to when the same tag it used to refer to different things, but a parallel issue is that different tags can also be used to refer to the same thing. Such heterogeneity is mainly caused by the multilingual nature of tags (e.g., "semanticweb" in English and "websemantique" in French), but also due to the fact that people use acronyms or shortened versions ("sw" and "semweb"), as well as linguistic and morpho-syntactic variations (synonyms, plurals, case variations, etc.). As an example, the following table lists some of the various tags used on Delicious to identify the concept of "Semantic Web," not taking into account related tags like RDFa, SPARQL, and so forth, as we will describe later. In this case, one must use various queries to get Semantic Web related content and, most importantly, one must know that each tag exists, which sometimes requires serendipitous discovery.

Lack of Organization and Relationships Between Tags

Since a folksonomy is essentially a flat bundle of tags, the lack of relationship between them

Figure 1. Tag ambiguity on a Flickr search for pictures tagged "apple"

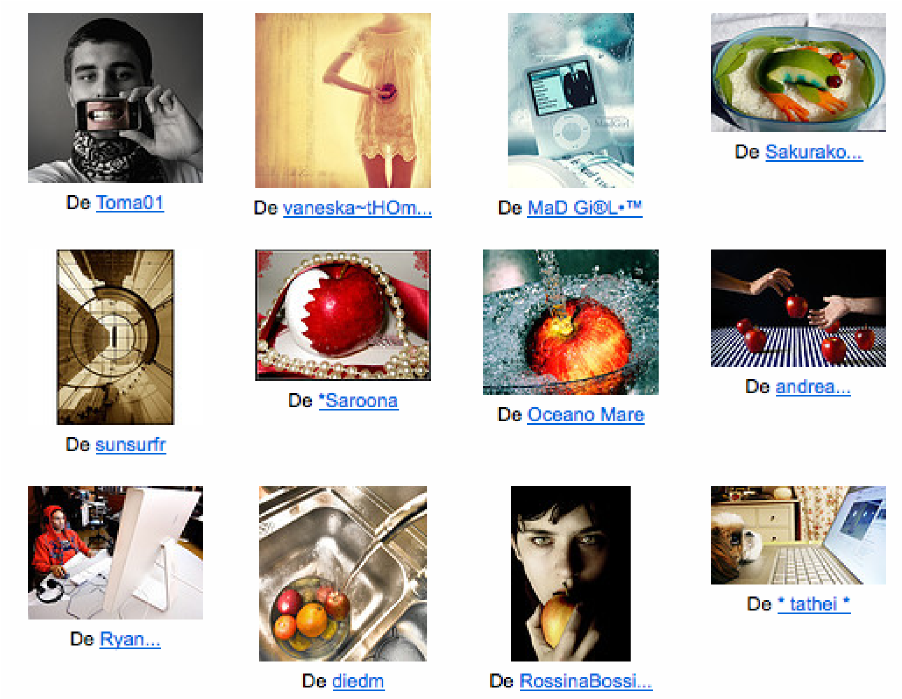


Table 1. Example of tag heterogeneity for the "Semantic Web" concept on Delicious

Tag	URL of related content
Semanticweb	http://delicious.com/tag/semanticweb
semantic-web	http://delicious.com/tag/semantic-web
Semaweb	http://delicious.com/tag/semaweb
Semweb	http://delicious.com/tag/semweb
Websemantic	http://delicious.com/tag/websemantic
web-semantic	http://delicious.com/tag/web-semantic
Websemantique	http://delicious.com/tag/websemantique
Websemantica	http://delicious.com/tag/websemantica
web-semantica	http://delicious.com/tag/web-semantica
Websemantico	http://delicious.com/tag/websemantico
web-semantico	http://delicious.com/tag/web-semantico
Websem	http://delicious.com/tag/websem

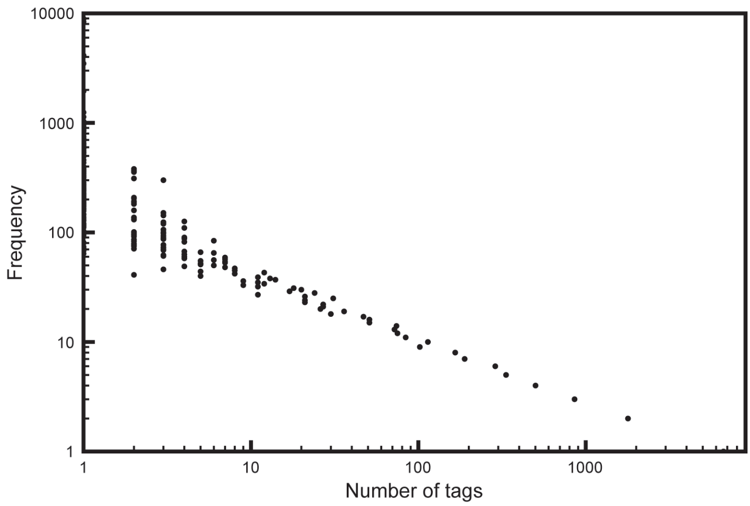
makes difficult to find information if one is not directly looking at the right tag, in addition to the previously-mentioned issues. This is clearly a problem in the practice of tagging, especially if people use different tags depending on their level on expertise or if they search for broader or narrower tags, as noted by (Golder & Huberman, 2006) when analysing Delicious. Indeed, although we mentioned the tags "semanticweb" or "socialweb" regarding this article, an expert on Semantic Web technologies may not use those terms (as they will be too broad for him) but instead would prefer tags like "moat," "linkeddata," or "sparql" to better classify the article. Then, someone simply looking at items tagged "semanticweb" will not be able to retrieve this article even though there is a clear relationship between these tags in terms of the technological domain. This relates, as Golder and Huberman (2006) noted, to a more generic issue regarding how different people consider different things as being the "basic level" for a knowledge domain, depending on their cognitive background and expertise in a field (Tanaka & Taylor, 1991). To that extent, the "semanticweb" tag might be considered to be a basic-level term for someone not involved in the domain, whereas it will be too broad for an expert or researcher, which may consider "sparql" or "linkeddata" as basic-level terms, depending on their research field. Since tags are unrelated, neither hierarchically nor by other means, bridging these basic levels between people and communities is hence an inherent issue of such systems.

Why Is Clustering Not Enough?

To overcome the issues we described, tag-based clustering algorithms have been proposed to identify similar and related tags (Begelman, Keller, & Smadja, 2006). However, their success depends on the tagging distribution, that is, if there is a strong co-occurrence between tags or not, which may not be the case in some folksonomies, even for tags that identify related concepts. In relation to this, an analysis of a corporate blogging system at Electricité de France R&D (http://retd.edf.fr), part of a general Enterprise 2.0 ecosystem in the company, raised some interesting issues. Enterprise 2.0 (McAfee, 2006) defines a corporate information system in which Web 2.0 tools and paradigms are used as a means to engage discussions and carry out knowledge sharing internally in an organisation. Firstly, we noticed that most of the tags used in this platform were used only a few times. In a total of 12,257 tags used on 21,614 blog posts, more than 68% were used twice or less, while only 10% were used more than ten times (Figure 2). As Hayes and Avesani (2007) reported, tag-based clustering may not be adapted for this kind of distribution, unless it is combined with other techniques such as reusing background information extracted from the tagged content itself.

Moreover, another interesting lesson that came out of from our analysis is that knowledge workers tag differently depending on their level of expertise, as we have already mentioned. The observations of Golder and Huberman (2006) regarding Delicious were confirmed by our study of corporate tagging. For example, experts in solar energies used tags like "TF" (an acronym for Thin Film, a particular kind of solar cell), whereas non-experts used generic ones like "solaire" (English for solar). Furthermore, experts often did not use any broader terms. Only 1% of the 194 items tagged with "TF" were tagged together with "solaire," while less than 0.5% of the 704 items tagged with "solaire" tagged with "TF." Therefore, tag-based clustering algorithms cannot be used to find related tags since they are too weakly related, as discussed in (Begelman et al., 2006). Consequently, non-experts cannot retrieve blog posts written by experts, even if there is a clear link between the different concepts. In such corporate contexts, that issue is clearly a problem: experts will write knowledgeable blog posts that non-experts will not be able to retrieve since they cannot be connected to broader concepts. These posts lie in the "long tail" where they generally contain high-value information. This gap regarding expertise and tagging behaviours is hence an important limitation for knowledge management in organizations that use such tagging systems.

Figure 2. Tag distribution in a corporate blogging platform



INTRODUCING MOAT

Tagging as a Quadripartite Model

Various theoretical definitions have been proposed to model tagging activities (Marlow, Naaman, Boyd, & Davis, 2006; Mika 2005). A widely agreed way is to represent a tagging action as a tripartite model between a User, a Resource, and a Tag and is defined as follows:

Tagging(User, Resource, Tag)

Hence, the following figure represents three different tagging actions (T1, T2, T3) made by two different users (U1, U2) on a particular photo. It also emphasizes on the social aspect of tagging existing in many applications, that is, different users tagging the same item, using the same tags or not.

The three tagging actions on this figure can then be represented as:

```
T1(U1, photo, apple)
T2(U2, photo, apple)
T3(U2, photo, laptop)
```

Yet, in our opinion, an important aspect of tagging is missing here, that is, the representation of the meaning of the tag used. As we explained, tags do not have any machine-readable semantic information, being simple text strings. However, there is generally a clear and unambiguous meaning associated with a tag by a user in a particular tagging action. Considering the previous example, it is clear that both users have in mind the computer brand when using the keyword "apple" to tag the picture. Hence, our first proposal is to extend the usual tripartite model of tagging to a quadripartite one as follows:

```
Tagging(User, Resource, Tag, Meaning)
```

The previous tagging actions can then be represented as:

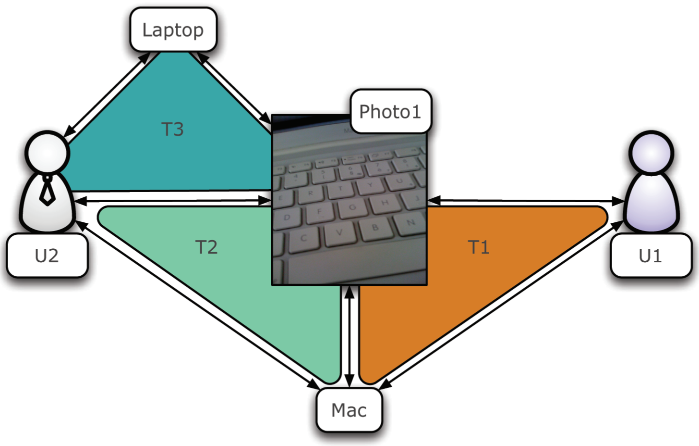
```
T1(U1, photo, apple, Apple computers)
T2(U2, photo, apple, Apple computers)
T3(U2, photo, laptop, a particular
kind of computer)
```

While a picture about fruits will be tagged as

```
T4(U3, photo, apple, a fruit)
```

Defining the meaning of tags using a simple text string leads to the same issues as before, since one user can describe apple as "Apple

Figure 3. Representing different tagging actions related to the same content



computers" and another user as "Apple Inc." Hence, our proposal is to consider each meaning being represented not by a text string but by a URI that defines it. Thanks to efforts conducted via the Linking Open Data (LOD) community project (Bizer, Heath, Ayers, & Raimond, 2007), millions of URIs for representing items as varied as places, brands, companies, people, and so forth, are now available on the Web from sources, such as DBpedia (Auer, Bizer, Lehmann, Kobilarov, Cyganiak, & Ives, 2007), Geonames, DBTune (Raimond & Sandler, 2008), and so forth, and can be efficiently used as identifiers as shown by (Hepp, Siorpaes, & Bachlechner, 2007). Adding this URI as a fourth element and not forcing users to directly use a URI when tagging content, permits them to keep their existing free-tagging habits, for example, using acronyms or multilingual tags, and selecting the exact tag they want, keeping intact their "desire lines" as Merholz(2004) called them. Hence, one can rely on those URIs to represent the meaning of each tag in tagging actions in a non-ambiguous and machine-readable way, as in the following extension to our example and using URIs provided by DBpedia.

```
T1(U1, photo, apple, <http://dbpedia.
org/resource/Apple_Inc.>)
T2(U2, photo, apple, <http://dbpedia.
org/resource/Apple_Inc.>)
T3(U2, photo, laptop, <http://dbpedia.
org/resource/Laptop>)
T4(U2, photo, apple, <http://dbpedia.
org/resource/Apple>)
```

This proposal solves both the ambiguity and heterogeneity issues with tagging. Regarding ambiguity, a user can now tag a fruits picture using "apple" with the meaning being defined differently to a laptop photo with the same tag? This can be done using a URI representing the apple as a fruit, for example, <http://dbpedia.org/resource/Apple>, as identified in T4 above. Considering the heterogeneity issue, another tag (e.g., "apple_computers") can be used and linked to the same meaningful URI (i.e., <http://dbpedia.org/resource/Apple_Inc.>) in a tagging action, solving the issue when retrieving information. Multi-lingual issues of tagging are taken into account in a similar way. Indeed, someone tagging a picture with "manzana" would be able to link it to the same <http://dpedia.org/resource/Apple> URI. To that extent, it is important to mention that the meanings of tags are defined thanks to URIs of entities, and not URLs of documents (as these would be as ambiguous as free-tags), conforming to the vision of an (ongoing) Web of Data in addition to the (current) Web of Documents.

In some cases, different URIs can be used for the same meaning, for example, <http://dbpedia.org/resource/Paris> and <http://sws.geonames.org/2988507/> for the city of Paris. Here, systems should take into consideration any owl:sameAs links that may exist between such resources (such as the two previous URIs) to identify that, in spite of different URIs, they represent the same entity. Such links may not yet exist and hence must be introduced separately. It is important to consider issues related to the use of owl:sameAs, which has a strong semantic meaning regarding identity, and other techniques could be considered to identify relatedness between entities (Jaffri, Glaser, & Millard, 2008). In addition, it may happen that an entity is being considered from a different point of view with different meanings, for example a city as a populated place and as an administrative division, their meaning being different. In that case, different URIs must be employed in the tagging action, for example <http://sws.geonames.org/2988507/>, for the city of Paris and <http://sws.geonames.org/6455259/> for Paris as an administrative division, both being defined in Geonames.

Moreover, in some cases, there may be no URI to represent the desired concept, for example, in the case where it is a very specific topic. In these cases, users should rely on external applications like Semantic Wikis to create a new URI for the concept. Creating such URIs is, in general, a good practice, as Jacobs and Walsh (2004) suggest, "To benefit from and increase the value of the World Wide Web, agents should provide URIs as identifiers for resources" and as emphasized by the Linked Data principles "Use URIs as names for things" (Berners-Lee, 2006). We will also see later how some MOAT clients can ease the process of creating new URIs when tagging content.

Finally, one important thing to consider is that these URIs are not isolated, but linked together to build a single Giant Global Graph (Berners-Lee, 2007) of structured knowledge. Hence, a system can infer that a blog post tagged (via MOAT) with the URI <http://dbpedia.org/resource/Apple_Inc.> is somehow related to a picture tagged with <http://dbpedia.org/resource/iPhone> as both are related thanks to DBpedia, following the Linked Data principles, by including links to related URIs as well as to other relevant information. We will later on give some more example of how such interlinks can be used in real-world applications and for querying purposes, but we will first focus on how we represent this quadripartite model in a formal way, that is, using a dedicated OWL ontology.

THE MOAT ONTOLOGY

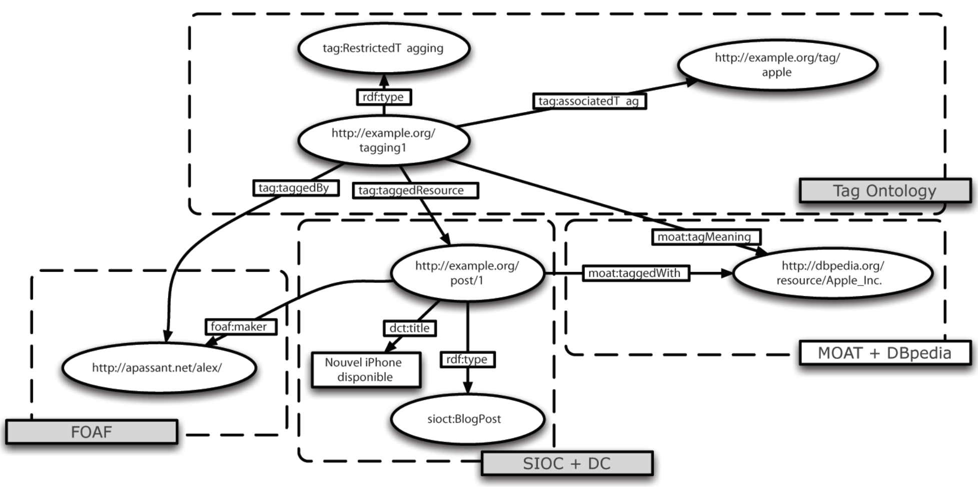
To model our proposal in a formal way, allowing software agents to represent and to query tagged items taking into account their links to entities from the Web of Data, we designed the MOAT project—Meaning Of A Tag (Passant & Laublet, 2008) (http://moat-project.org)—consisting of (1) a lightweight ontology and (2) a related collaborative framework. The ontology is based on prior work on tagging ontologies and reuses the Tag Ontology (http://www.holygoat.co.uk/projects/tags).

First, the MOAT ontology introduces a Tag class (as a subclass of the Tag one defined in the Tag Ontology) to define the concept of Tag, allowing each tag to get a proper URI, being linked to the tag (as a keyword) with a name property. This class addresses one of the problems of the Tag Ontology, since in this model, an instance of Tag can be assigned different labels without any restriction. This can lead to tags labelled with both "RDF" and "Ireland," which does not make any sense from a user point of view, but no software can detect this inconsistency since it is not defined in the model. Hence, MOAT introduces an OWL cardinality constraint so that an instance of Tag can have a single name. In addition, MOAT reuses the RestrictedTagging class defined in the Tag Ontology to model the tripartite action of tagging and simply introduces a tagMeaning property in order to link to the URI of the tag meaning in a tagging action. The following snippet of code and the related figure (Figure 4) hence represent how to model that, in a particular tagging context, the tag “apple” means <http://dbpedia.org/resource/Apple_Inc./>, that is, the computer brand. As one can see, in addition to MOAT and the Tagging Ontology, we use FOAF to represent the agent that realised the tagging action, SIOC (Breslin, Harth, Bojārs, & Decker, 2005) and DublinCore to represent the tagged item, whereas DBpedia is used to define the meaning of the tag in that example.

```
@prefix moat: <http://moat-project.
org/ns#> .
@prefix foaf: <http://xmlns.com/
foaf/0.1/> .
@prefix sioc: <http://rdfs.org/sioc/
ns#> .
@prefix dct: <http://purl.org/dc/
terms/> .
<http://example.org/post/1> a
sioc:Post ;
foaf:maker <http://apassant.net/alex>
;
dct:title “Browsing Linked on iPhone”
;
moat:taggedWith <http://dbpedia.org/
resource/Apple_Inc.> .
<http://example.org/tagging/1> a
tags:RestrictedTagging ;
tag:associatedTag <http://example.org/
tag/apple> ;
tag:taggedBy <http://apassant.net/
alex> ;
tag:taggedResource <http://example.
org/post/1> ;
moat:tagMeaning <http://dbpedia.org/
resource/Apple_Inc.> .
```

As we can also see in this figure, the vocabulary uses a taggedWith property to model a direct link between the tagged item and the meaning URI. This can be used when the tripartite relationship is not needed, providing a shorter path for querying data. Although properties like dc:subject from DublinCore or skos:subject from SKOS (while recently deprecated) could have been used here, their semantics specifically indicate that the related object is a subject of the annotated item, which may not be the case. Tags can indeed be seen not only as descriptive metadata but also as structural or administrative metadata, considering the digital libraries terminology regarding metadata (Taylor, 1999). Hence, there is a need to model that a URI is linked to an item via a tagging action, but is not a subject, for example <http://dbpedia/GNU_Free_Documentation_License> could be used to identify that the annotated work is licensed under GNU FDL but is not about GNU FDL, and this is the goal of the taggedWith property.

Figure 4. Modelling the meaning of a tag in a particular tagging action with MOAT



In this quadripartite model representing tagging actions, we consider the meaning of tag to be local, that is, depending on the tagging action itself, and we call it the *local meaning* of a tag. However, taken out of context, the same tag can have multiple meanings, that is, the tag apple can refer to various things. This is a particular feature of tags that that we also want to model in MOAT and that we named the *global meanings* of a tag. To model it, we defined by the following theoretical model, in which {User} represents the set of users that assign a particular meaning to this tag.

Meaning(Tag) = {(Meaning, {User})}

Based on this model, the following snippet of code and the related figure (Figure 5) show how to represent two different global meanings for the apple tag in a given folksonomy, respectively <http://dbpedia.org/resource/Apple_Inc.> by one user and <http://dbpedia.org/resource/Apple> by two of them. To represent these global meanings with MOAT, we introduced a particular Meaning class and hasMeaning and meaningURI properties, allowing us to reifying these relationships, that is, taking into account the different users that assign a particular meaning to a tag.

```
@prefix moat: <http://moat-project.
org/ns#>.
@prefix foaf: <http://xmlns.com/
foaf/0.1/>.
<http://example.org/tag/apple> a
moat:Tag;
moat:hasMeaning <http://example.org/
meaning/apple/1>;
moat:hasMeaning <http://example.org/
meaning/apple/2>.
<http://example.org/meaning/apple/1> a
moat:Meaning;
moat:meaningURI <http://dbpedia.org/
resource/Apple_Inc.>;
foaf:maker <http://apassant.net/alex/>
<http://example.org/meaning/apple/2> a
moat:Meaning;
moat:meaningURI <http://dbpedia.org/
resource/Apple>;
foaf:maker <http://example.org/alice>;
foaf:maker <http://example.org/bob>.
```

Then, an overview of local and global meanings of tags defined in MOAT can then be represented as follows (Figure 6), with the complete ontology being available at http://moat-project.org/ns.

Figure 5. Representing two different meanings for the tag "apple"

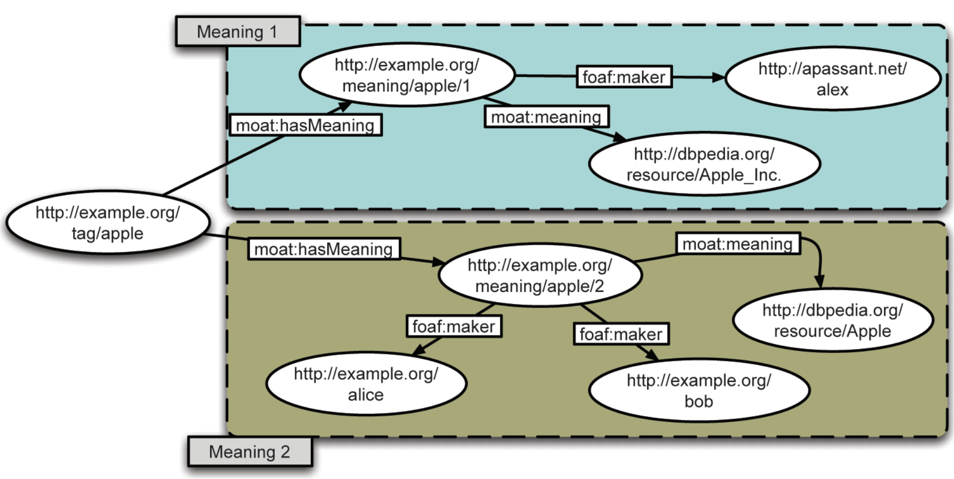
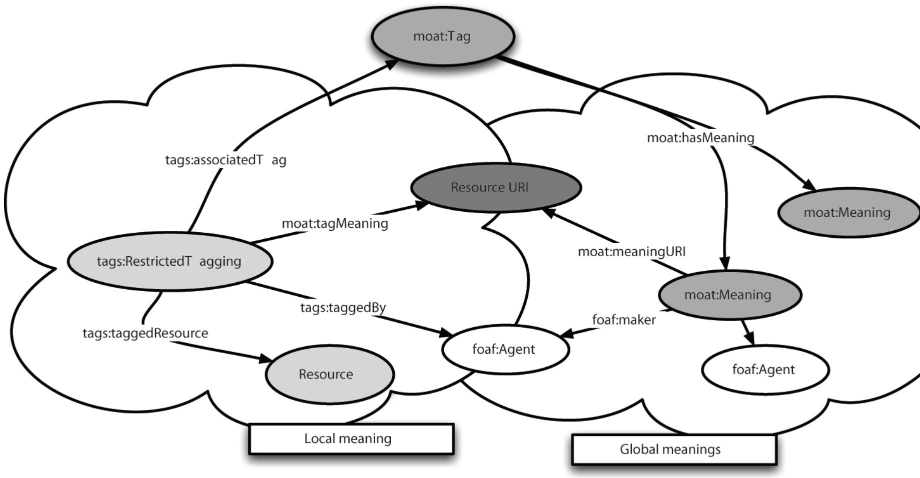


Figure 6. The MOAT ontology



THE MOAT FRAMEWORK

To the apply these principles of semantically enhanced tagging and to allow people to assign meaning to their tags, we designed a complete framework associated with the MOAT ontology that consists of: a MOAT server, which people can be subscribed to—as they can do with Annotea (http://www.w3.org/2001/Annotea/) (Kahan & Koivunen, 2001)—and that stores global meanings of tags for a given community of users;

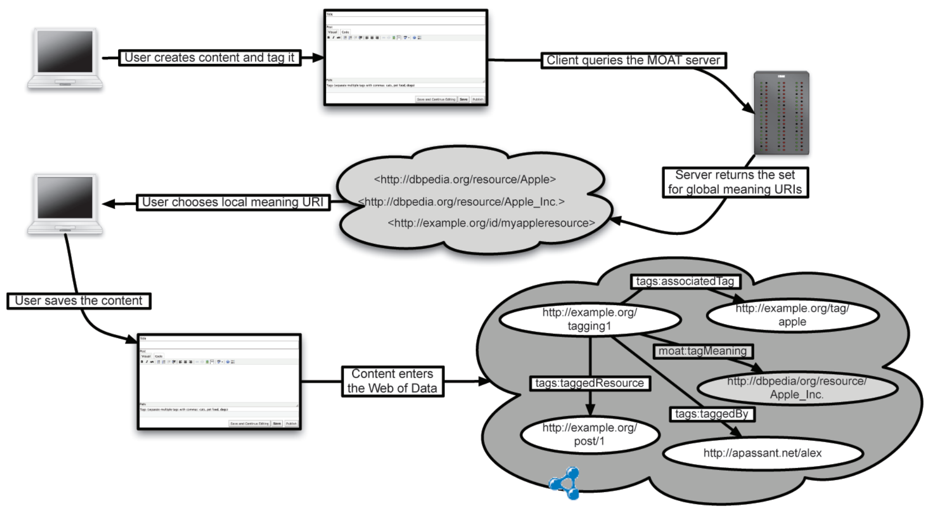
MOAT clients that interact with servers to identify global meanings when users tag content to let them choose a local meaning for their tags. If needed, new URIs can be added by the user through the client. Clients also generate the related RDF data once the content has been semantically tagged.

The MOAT framework and its related workflow are depicted in the following picture (Figure 7), and it is worth mentioning that the client and server simply interact by exchanging RDF graphs via HTTP.

Since the community can add new meanings, this framework combines the architecture of participation principles of Web 2.0 (i.e., sharing and adding meanings within a community) and the knowledge representation paradigms from the Semantic Web (i.e., providing RDF data for tagging actions). To model users within this architecture, we rely on FOAF (as previously mentioned) to ensure the uniqueness of one's identity and in a distributed manner if required (Bojārs, Passant, Breslin, & Decker, 2008). The use of FOAF can also be combined with authentication schemes like OpenID or FOAF+TSL (Story, Harbulot, Jacobi, & Jones, 2009) in the future.

The previous architecture has been implemented as open-source framework, available at http://moat-project.org. A MOAT server is available in PHP and can be used in combination with various triple stores, to ease its integration in existing architectures. It also provides Linked Data for any tag URI. For example, one can browse the tag <http://tags.moat-project.org/paris> to get the list of global meanings, retrieving RDF/XML or HTML depending on the user agent. It can also deliver JSON to help Web 2.0 developers to build MOAT-based applications without learning Semantic Web principles. A Drupal module has been designed to interact with such servers. To let users add new URIs when nothing relevant is retrieved from the server, we rely on the Sindice (Tumarrello et al., 2007) search widget. The following picture (Figure 8) displays the use of the MOAT client

Figure 7. Workflow associated with the MOAT framework



for three different tags: at the bottom, for the tag "sparql" a single URI has been suggested by the server and selected by the user; in the middle, three URIs were suggested (and one selected) for the tag "paris" while on the top, the Sindice widget is used to find a new URI for the tag "barcamp." In addition to these public client and server implementations, a MOAT client and server has been integrated in the OpenLink Data Space platform, a complete Web 2.0 suite built on Semantic Web technologies (Idehen & Erling, 2008). While the current Drupal implementation displays URIs, a user-friendly way would be to expose human-readable labels, as we have recently carried out in a corporate environment and will soon detail. On the Web, a solution would be to query each URI to retrieve its label, or one can use the recent SPARCool service (http://sparcool.net) that has a more elaborate interface to make such queries easier.

USING MOAT IN ENTERPRISE 2.0

Background Context and Related Use-case

As we mentioned earlier, one of our first motivations for MOAT originated in the use and analysis of a corporate blogging platform at Electricité De France R&D, in the context of a project in which we studied how Semantic Web technologies could improve Enterprise 2.0 ecosystems (Passant, 2008). In this context, and while not directly related to MOAT, we reused data from Geonames in a Semantic Wiki to build runtime semantic mash-ups combining internal and external data sources (see Figure 9). We believe that these mash-ups can be the future of Enterprise 2.0 applications: similar to how RSS allows companies to benefit from public information, reusing publicly available Linked Data allows us to take advantage of large-scale knowledge about different topics for relatively minor cost. Hence, we believe that Linked Data—particularly data available using open licences—has an important role to play in business information systems and could be a key feature for the Web of Data and related

Figure 8. Using the Drupal MOAT client

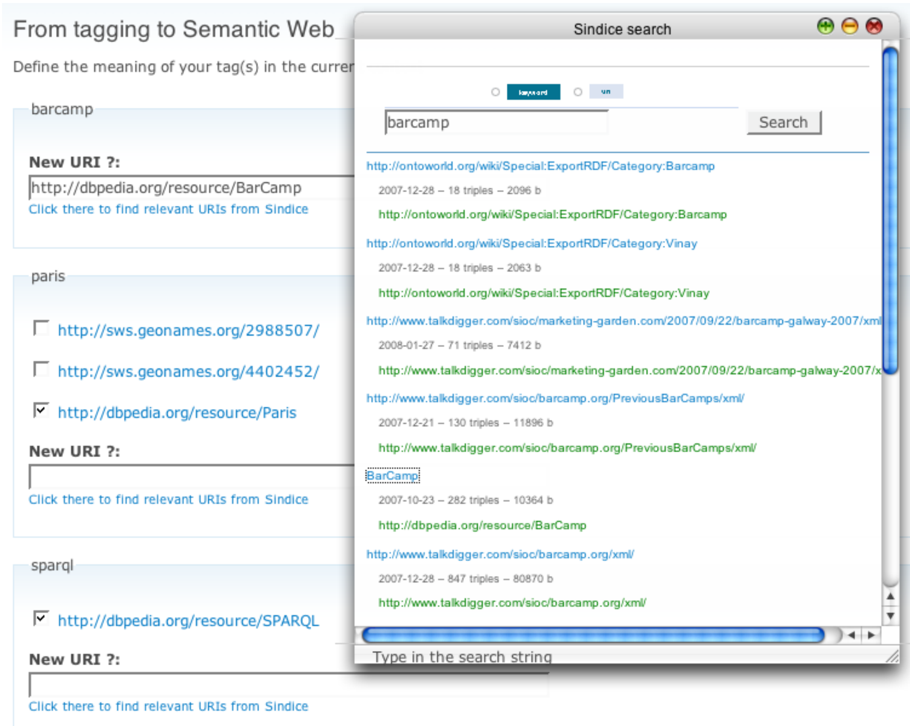
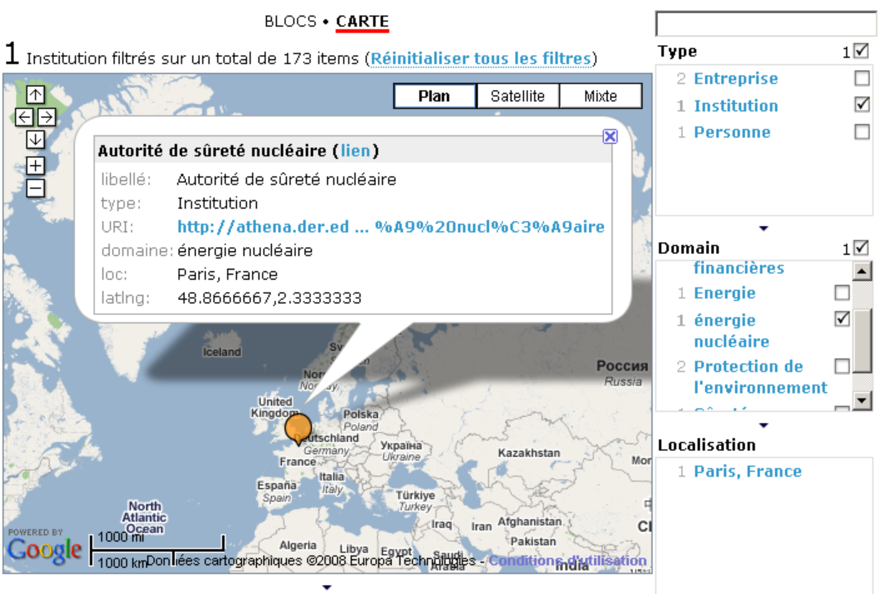


Figure 9. A semantic mashup with Exhibit, combining internal and external RDF data from the LOD-cloud



technologies in corporate contexts, as also demonstrated recently by the BBC (Kobilarov et al., 2009).

To embed MOAT in our architecture (since we did not want end users to be faced with URIs in order to define the meaning of their tags), we updated the generic Drupal client to display not URIs, but rather human-readable labels (based on the rdfs:label property) of resources from our internal knowledge base (populated mainly via our Semantic Wiki). To add global meanings to tags, our client allows users to simply browse our internal knowledge base and choose the right resource to assign the tag, or create a new one, without having to face any RDF(S)/OWL data and using a simple Flash interface (Figure 10). This interface also allows us to see which tags are related to any resource.

BENEFITS OF MOAT IN ENTERPRISE 2.0

In order to derive benefit from the semantic tagging process, we integrated MOAT in a semantic search engine that we built internally, aggregating RDF data from various internal sources (Passant, Laublet, Breslin, & Decker, 2009). The engine uses MOAT to:

- Suggest relevant and appropriate resources based on a searched term. Hence, a user searching for the term "france" will be asked if he or she wants to retrieve information about the resource "France," "Electricité de France," or "Gaz de France," using the links between tags and related URIs;
- Once the relevant resource has been identified, the system retrieves all content

Figure 10. Browsing the internal knowledge base to create a new resource from a given tag

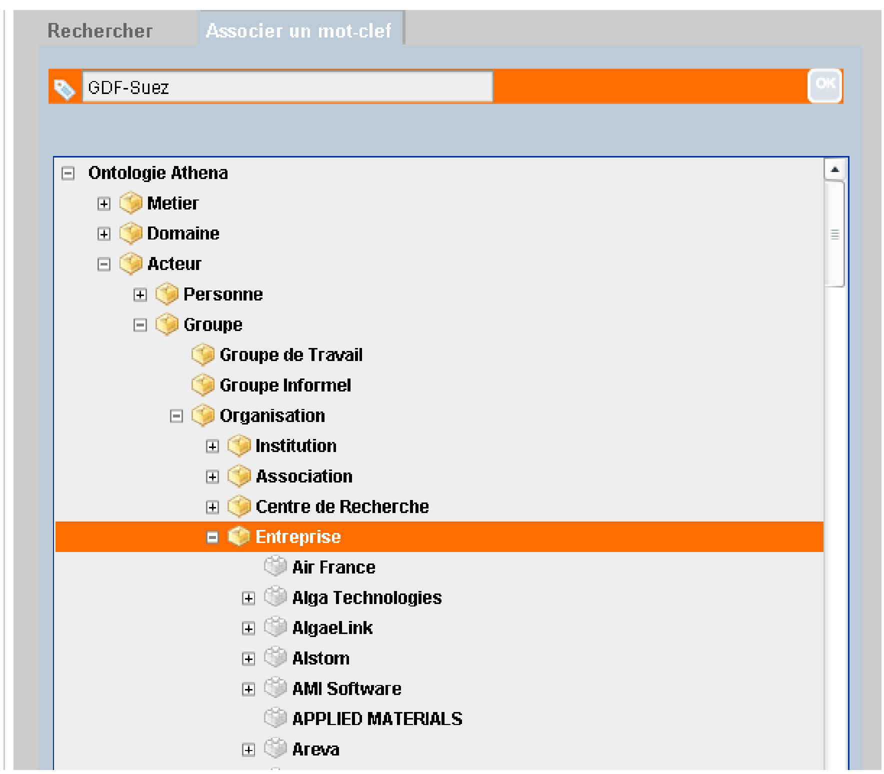


Figure 11. Semantic Search engine taking advantage of MOAT

linked to that URI, also identifying its source (i.e., Wiki or blogging platform). As the search is based on URIs and not on keywords anymore, it solves both ambiguity and heterogeneity issues.

Although the engine relies completely on RDF(S)/OWL data and SPARQL queries, these are hidden from users as the goal (Figure 11) is to showcase the benefits of Linked Data technologies without any complex interfaces.

We previously mentioned an important issue related to differing expertise levels and its effect on tagging behaviour, leading to some content that is difficult to find for some users. Hence, our search engine also suggests concepts related to the one a user is currently searching for, by analysing the underlying knowledge base and displaying related concepts based on some rules (e.g., using the skos:broader relationship which links to a broader concept).

EVALUATING THE APPROACH

From a total of 12,257 tags used in our platform, 1176 of them were related to 715 different URIs, both from our internal knowledge base and from GeoNames. Analysing these relationships showed that while only one tag was subject to ambiguity issues, heterogeneity was important. As the following table shows, a total of 205 resources (i.e., URIs) were subject to heterogeneity with more than one tag assigned to each URI. Specifically, 39 were assigned at least five or more tags. For example, "Supercapacitor" (a component used in electrical engineering) was related to the five following tags: "supercapacité," "supercondensateur," "ultracapacité," "ultracapacitor," and "ultracondensateur." As expected, it emphasises the usual heterogeneity issues of tagging systems, such as synonymy ("supercondensateur," "ultracondensateur," etc.) and multi-lingual issues ("ultracapacité," "ultracapacitor").

Using MOAT and URIs instead of simple tags helped to solve this heterogeneity issue since our engine retrieves information because of these URIs. Then, a single query is needed to retrieve information about "Supercapacitor," instead of the five related tag-based queries that it originally implied.

In addition, we noticed that even for a single user, different tags were used for the same concept. For example, only three users were involved in the previous example, with one of them using three different tags. More generally, we particularly noticed that "personal heterogeneity" issues regarding tags referring to people (i.e., tags used for the full name versus last name only) and locations (e.g., "USA" and "Etats-Unis"), as well as some technologies

Table 2. Statistics of tag heterogeneity

Number of tag(s)	Related URI
1	510
2	96
3	70
4 or more	39

(multilingual issues and acronyms), displaying some interesting behaviour regarding personal tagging habits.

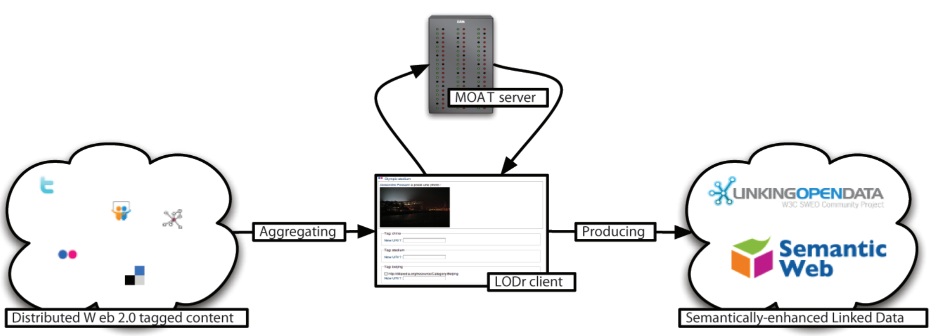
We also interviewed six users and asked them to rate both their interest in the system and the complexity of the approach. Although they found this approach to be more complex than simple tagging, all of them agreed that it was interesting, with an average rate of 2.83/5, whereas the search engine was rated 3.5. Interestingly, three users mentioned that this search interface with links to related items helped them to discover new content. In addition, four also acknowledged that they used the advanced interface to create new meanings for their tags. One outcome is also that incentives (such as our search engine) must be given to end users to make them go through this additional step and to make them understand that this is worth doing.

LODR: WEAVING HETEROGENEOUS USER-GENERATED CONTENT INTO THE WEB OF DATA

Goals and Principles

To apply MOAT principles on the Web, we implemented LODr (http://lodr.info), a personal application that allows one to re-tag their existing Web 2.0 content and to weave it into the Web of Data thanks to the aforementioned principles. Its main objective is to provide a simple way to create RDF and interlinked content from existing Web 2.0 tools, so that queries like "list all SlideShare items tagged with a topic related to the Semantic Web" can be answered. LODr is an open-source application written in PHP5 using an object-oriented model, and although it is completely RDF-based, it simply uses a generic LAMP setup thanks to ARC2 (http://arc.semsol.org). LODr is based on a set of wrappers translating user-generated content from various services into RDF, featuring wrappers for major

Figure 12. The LODr architecture



Web 2.0 services (Twitter, SlideShare, Flickr, Delicious, Bibsonomy) while new wrappers can be easily written. One motivation for writing a standalone application is that we did not want to create another tagging service but rather produce a system offering users with a way to enrich existing data to avoid social network fatigue (Fitzpatrick & Recordon, 2007), and this allows users to keep using their existing applications and tagging habits.

Once original content have been translated to RDF because of these wrappers, it is immediately available in RDFa, using notably FOAF and SIOC. This first step also allows us to get over the issue of isolated data silos since the social data is then considered via a unified semantic layer. In a second step, users can interact with a MOAT server to add meaning to their tags and hence interlink this data with existing URIs, as described in Figure 12. Moreover, LODr allows us to get meanings suggested from public SPARQL endpoints, which can make the process easier in some cases as a user can choose an endpoint corresponding to his or her particular interests.

BENEFITS OF THE APPROACH

LODr provides advanced interfaces to browse semantically tagged data, as Figure 13 depicts. By analysing RDF data corresponding to the chosen meaning (in that case via the SPARQL URI on DBpedia), it displays additional information about it as well as suggesting relevant URIs based on direct and indirect relationships. For example. GRDDL is suggested when browsing SPARQL as both share a similar value for the skos:subject property, that is, dbpedia:Category:World_Wide_Web_Consortium_standards.

More generally, by being interlinked to other data sources, this user-generated content becomes more valuable than the original, since it is no longer (1) locked in proprietary data silos, and (2) is not just related to simple meaningless free-text tags. As Metcalfe's law defines, the value of a network is proportional to the number of nodes in the network. Hence, by providing new links, we augment the network effect and hence the value of this user-generated data. For example, content tagged on Flickr and re-published using MOAT can be interlinked

Figure 13. Browsing items related to a particular URI with LODr

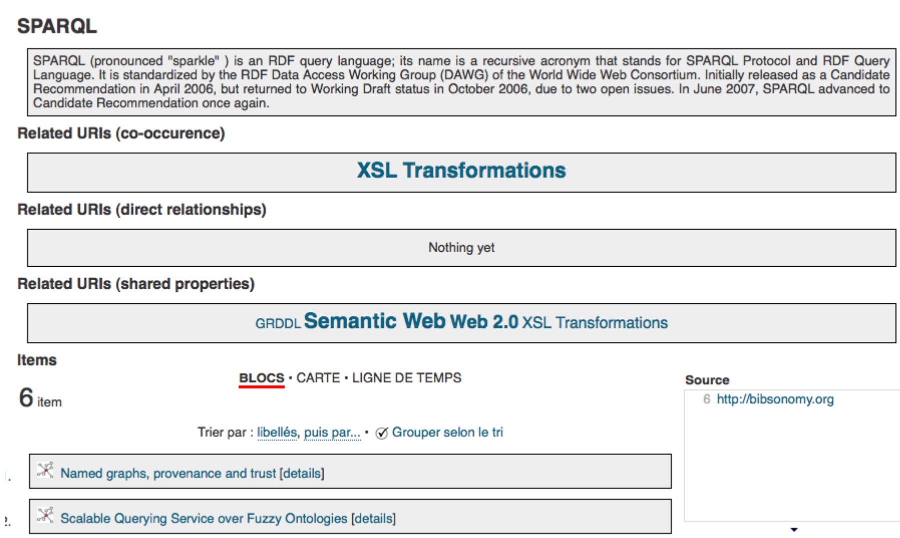
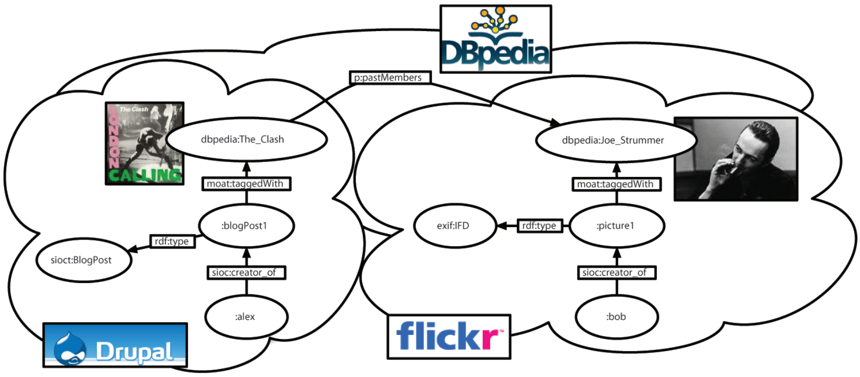


Figure 14. Interlinking user-generated content through various paths thanks to MOAT



to Drupal content thanks to DBpedia as seen in Figure 14.

Furthermore, advanced queries can be answered once these links have been provided. For example, the following query (with prefixes omitted) identifies the last five SlideShare items related to Semantic Web technologies.

```
SELECT DISTINCT ?item ?author ?date
?tag ?meaning
WHERE {
 ?item a sioc:Item ;
 dct:created ?date ;
 sioc:has_space <http://slideshare.
net> ;
 foaf:maker ?author .
 [] a tags:RestrictedTagging ;
  tags:taggedResource ?item ;
  tags:associatedTag[
   tags:name ?tag .
  ] ;
  moat:tagMeaning ?meaning .
 ?meaning ?p <http://dbpedia.org/re-
source/Category:Semantic_Web> .
}
ORDER BY DESC(?date) LIMIT 5
```

The next SPARQL query provides a similar use case that retrieves pictures related to a particular place, identified by its GeoNames URI, with results being displayed in Figure 15.

```
SELECT DISTINCT ?item ?author ?date
?tag ?meaning
WHERE {
 ?item a sioc:Item ;
 dct:created ?date ;
 sioc:has_space <http://flickr.com> ;
 foaf:maker ?author .
 [] a tags:RestrictedTagging ;
  tags:taggedResource ?item ;
  tags:associatedTag [
   tags:name ?tag .
  ] ;
  moat:tagMeaning ?meaning .
 ?meaning foaf:based_near <http://sws.
geonames.org/2522437/> .
}
ORDER BY DESC(?date) LIMIT 5
```

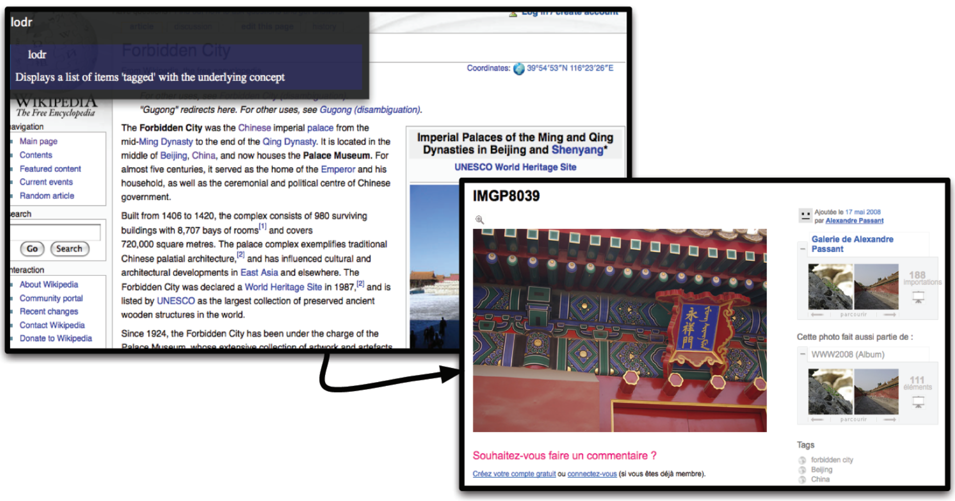
What is of particular interest in this query is how Linked Data can be leveraged to enhance information discovery. The original picture does not contain any geolocation-related tag. However, it has been linked to the <http://data.semanticweb.org/conference/eswc/2008> URI using LODr. This URI represents the ESWC 2008 conference and delivers lots of associated information including its location (identified as <http://sws.geonames.org/2522437/>), which allows us to answer the previous query.

As these SPARQL queries are geared toward advanced users, we also deployed a Mozilla Ubiquity command to allow everyone benefit from this method. This command retrieves tagged-data linked to the concept behind a browseable page. The following picture describes a related use-case: someone browsing the German Wikipedia page about the Forbidden City calls the command that will get, via DBpedia, the related URI and then, via LODr, the related tagged item. Then, the

Figure 15. Identifying pictures thanks to MOAT and Linked Data

item	author	date	tag	meaning
http://www.flickr.com/photos/terraces/2924436932/	http://apassant.net/alex	2008-10-08T14:26:06+01:00	eswc2008	http://data.semanticweb.org/conference/eswc/2008

Figure 16. Ubiquity command to retrieve user-generated content within the Linked Data Web



user can browse the original pictures in Flickr (Figure 16).

Thanks to this service as well as the previous SPARQL query, one can see the benefits of linking Web 2.0 content to URIs with MOAT rather than using simple free-text tags alone. By following the Linked Data principles, these various links enable increased integration between data originally locked into independent and proprietary silos, which become better connected thanks to common representation models and the interlinks between them. Such services also provide incentives for people to do enrich existing data with semantics.

Related Works

We mainly consider related works in terms of (1) ontologies used to represent tagged data, (2) mining ontologies from folksonomies to solve tagging issues and (3) providing users with means to organise their tags, MOAT being at the frontier of these three approaches.

Firstly, since MOAT defines a particular ontology dedicated to tagging activities, it is worth mentioning that various models have been already defined to achieve this goal. The Tag Ontology (Newman, Ayers, & Russell, 2005), on which MOAT is built, is then the first ontology of this kind that became available on the Web, is provided in OWL-Full, and is used in various applications such as Revyu.com (Heath & Motta 2007). This vocabulary is based on the theoretical foundations defined by Gruber (2007) and provides a representation of both tags and tagging actions. It relies on FOAF for modelling taggers as well as using SKOS to model tags and to allow people to organise them. SCOT—Semantic Cloud Of Tags (Kim, Yang, Breslin, & Kim, 2007)—provides a comprehensive model dedicated to model tag clouds and related objects such as tags co-occurrences, aiming to provide interoperability and portability between tagging applications. Other models that can be used to represent tags include the Nepomuk Annotation Ontology (NAO), SIOC, and the Annotea annotation and bookmark schemas. Both NAO and SIOC define a new Tag class, the latter one subclassing skos:Concept. SIOC also defines a topic property to link a resource to some of its topics. Although not explicitly using the "tag" word in its definition, the Annotea bookmark model provides a Topic class and a hasTopic property to link an item to some related keywords, and provides the ability to model a hierarchy of topics thanks to subTopicOf relationships. However, these vocabularies do not consider the tripartite tagging model of tagging (but simply the relationship between an item and its tags) and consequently cannot capture the complete essence of folksonomies. Although each of the previous ontologies focuses on a particular aspect of tagging, none of them takes the meaning of tags into account. Combined together, SCOT, SIOC, MOAT, and the Tag Ontology provide a complete framework for tagged data.

Other approaches have been considered to help solve the issues with free tagging, especially by analysing folksonomies to create taxonomies or ontologies from them, based on the ideas that emergent semantics appear. Among others, Halpin, Robu, and Shepard (2006) used an approach based on related co-occurrences of tags to extract hierarchical relationships between concepts. (Mika, 2005) defined a socially aware approach for automatically building ontologies by combining social network analysis and clustering algorithms based on folksonomies. Schmitz (2006) describes how to create hierarchical models from Flickr tags while FolksOntology (Van Damme, Hepp, & Siorpaes, 2007) provides another method to bridge the gap between folksonomies and ontologies. More recently, the FoLksonomy Ontology enRichment (FLOR) technique has provided a completely automated approach to semantically enrich tag spaces by mapping tags to Semantic Web entities (Angeletou, 2008). By enriching tag spaces with semantic information about the meaning of each tag, some issues of tagging in relation to information retrieval (such as tag ambiguity as mentioned earlier) can be solved.

It is worth noticing that these two domains are not disjoint and can be combined together. For example, MOAT can be used as a background model in support of automated approaches like FLOR. (Abel, 2008) uses the MOAT ontology in combination with an automated method to enrich existing tagging spaces in the GroupMe application. Such improvements may be considered in the future to make the MOAT process simpler for end-users.

Finally, we must also consider other manual approaches and tools used to solve the issues of tags. For example, tools like Gnizr or Bibsonomy allow users to define manually hierarchical relationships between tags and then provide some personal tagging organisation schemes. Although the MOAT approach does not take into account this personal contextualisation aspect as it relies on shared knowledge bases for tag meanings (such as DBpedia), we believe it can be more beneficial, especially as we have noticed that most of the relationships defined in these tools are widely known relationships, such as "france" defined as a subtag of "europe," and so forth. Moreover, this way of manually organising hierarchies of tags does not solve the ambiguity issue. In addition, machine tags, as introduced by Flickr, can also be considered. Due to their "prefix:property=value" approach, these are mainly dedicated to advanced users or automated-tagging systems, such as applications for GPS-enabled camera phones. Finally, Faviki uses a similar approach that relies only on DBpedia URIs and does not consider the free-tagging aspect; that is, it asks users to directly use DBpedia URIs and does not consider that users will have their own ways to tag content.

CONCLUSION

In this article, we demonstrated how Web 2.0 content and Linked Data principles could be combined in order to solve usual tagging issues. We showed how that integration allows to envisioning a better convergence between those two visions of the Web, leading to a Web of structured, interoperable and user-driven data, also known as the Social Semantic Web (Breslin & Decker, 2007).

We described some common issues with free-tagging systems, including tag heterogeneity and ambiguity, and a lack of relationships amongst tags. We introduced the MOAT ontology, which is based on a quadripartite tagging model, and demonstrated its general usefulness through two use cases in a corporate intranet and on the public Web.

The methods described here will help machines and humans to work more closely, by having people voluntary publishing large sets of tagged user-generated content as RDF so that it can be more efficiently reused for information discovery and navigation through attractive mash-ups and query interfaces. However, we must also keep in mind that while technology and especially the Linked Data principles may help to achieve this goal, a key component to its success is the social aspects and people themselves.

ACKNOWLEDGMENT

The work presented in this article has been funded in part by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2).

REFERENCES

Abel, F. (2008). The benefit of additional semantics in folksonomy systems, In *Conference on Information and Knowledge Management: Proceedings of the 2nd PhD Workshop on Information and Knowledge Management,* Napa Valley, CA (pp. 49-56). New York: ACM Publishers.

Angeletou, S. (2008, October 26-30). Semantic enrichment of folksonomy tagspaces. In A. Sheth, S. Staab, M. Dean, M. Paolucci, D. Maynard, T. Finin, et al. (Eds.), *The Semantic Web-ISWC 2008,* Karlsruhe, Germany (LNCS 5318, pp. 889-894).

Auer, S., Bizer, C., Lehmann, J., Kobilarov, G., Cyganiak, R., & Ives, Z. (2007, November 11-15). DBpedia: A nucleus for a web of open data. In K. Aberer, K. S. Choi, N. Noy, D. Allemang, K. I. Lee, L. J. B. Nixon, et al. (Eds.), *The Semantic Web: Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007),* Busan, Korea (LNCS 4825, pp. 715-728).

Begelman, G., Keller, P., & Smadja, F. (2006, May 22-26) *Automated tag clustering: Improving search and exploration in the tag space*. Paper presented at the Collaborative Web Tagging Workshop, 15th International World Wide Web Conference (WWW 2006), Edinburgh, Scotland.

Berners-Lee, T. (2007, November 21). Giant global graph. Retrieved May 8, 2009, from http://dig.csail.mit.edu/breadcrumbs/node/215

Berners-Lee, T. (2006, June 27). *Linked data-design issues*. Retreived May 8, 2009, from http://www.w3.org/DesignIssues/LinkedData.html

Bizer, C., Heath, T., Ayers, D., & Raimond, Y. (2007). *Interlinking open data on the Web*. Poster session presented at the 4th European Semantic Web Conference (ESWC 2007), Innsbruck, Austria.

Bojārs, U., Passant, A., Breslin, J. G., & Decker, S. (2008, May 6) *Social network and data portability using Semantic Web technologies*. Paper presented at the 2nd Workshop on Social Aspects of the Web (SAW 2008), Innsbruck, Austria.

Breslin, J. G., & Decker, S. (2007). The future of social networks on the Internet: The need for semantics. *IEEE Internet Computing*, *11*, 86–90. doi:10.1109/MIC.2007.138

Breslin, J. G., Harth, A., Bojārs, U., & Decker, S. (2005, May 29-June 1). Towards semantically-interlinked online communities. In A. Gómez-Pérez & J. Euzenat (Eds.), *The Semantic Web: Research and Applications: Proceedings of the 2nd European Semantic Web Conference (ESWC 2005),* Heraklion, Crete (LNCS 3532, pp. 500-514).

Fitzpatrick, B., & Recordon, D. (2007, August 17). *Thoughts on the social graph*. Retrieved May 8, 2009, from http://bradfitz.com/social-graph-problem/

Golder, S., & Huberman, B. A. (2006). The structure of collaborative tagging systems. *Journal of Information Science*, *32*(2), 198–208. doi:10.1177/0165551506062337

Gruber, T. (2007). Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Semantic Web and Information Systems*, *3*(1), 1–11.

Halpin, H., Robu, V., & Shepard, H. (2006, November 6). The dynamics and semantics of collaborative tagging. In *Proceedings of the 1st Semantic Authoring and Annotation Workshop (SAAW 2006), 5th International Semantic Web Conference (ISWC 2006)*, Athens, GA (Vol. 209) CEUR Workshop Proceedings.

Hayes, C., & Avesani, P. (2007, March) *Using blog tags to identify topic authorities*. Paper presented at the 1st International Conference on Weblogs and Social Media (ICWSM 2007), Boulder, Colorado.

Heath, T., & Motta, E. (2007, November 11-15). Revyu.com: A reviewing and rating site for the Web of Data. In K. Aberer, K. S. Choi, N. Noy, D. Allemang, K. I. Lee, L. J. B. Nixon, et al. (Eds.), *The Semantic Web: Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007),* Busan, Korea (LNCS 4825, pp. 895-902).

Hepp, M., Siorpaes, K., & Bachlechner, D. (2007). Harvesting Wiki consensus: Using Wikipedia entries as vocabulary for knowledge management. *IEEE Internet Computing*, *11*(5), 54–65. doi:10.1109/MIC.2007.110

Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006, June 11-14). Information retrieval in folksonomies: Search and ranking. In Y. Sure & J. Domingue (Eds.), *The Semantic Web: Research and Applications: 3rd European Semantic Web Conference (ESWC 2006),* Budva, Montenegro (LNCS 4011, pp. 411-426).

Idehen, K., & Erling, O. (2008, April 22). Linked data spaces & data portability. In *Proceedings of the Linked Data on the Web Workshop (LDOW 2008),* Beijing, China (Vol. 369). CEUR Workshop Publishers.

Jacobs, I., & Walsh, N. (2004, December 15). *Architecture of the World Wide Web, volume one*. Retrieved May 8, 2009, from W3C Recommendation's Web site: http://www.w3.org/TR/webarch/

Jaffri, A., Glaser, H., & Millard, I. (2008, June 2). *Managing URI synonymity to enable consistent reference on the Semantic Web*. Paper presented at the Workshop on Identity and Reference on the Semantic Web (IRSW 2008) at the 5th European Semantic Web Conference (ESWC 2008), Tenerife, Spain.

Kahan, J., & Koivunen, M.-R. (2001, May 1-5). Annotea: An open rdf infrastructure for shared Web annotations. In *Proceedings of the 10th international conference on World Wide Web (WWW 2001)*, Hong Kong, China (pp. 623-632). New York: ACM Publishers.

Kim, H. L., Yang, S. K., Breslin, J. G., & Kim, H. G. (2007, November 2-5). Simple algorithms for representing tag frequencies in the SCOT exporter. In *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, Silicon Valley, CA (pp. 536-539). IEEE Computer Society.

Kobilarov, G., Scott, T., Raimond, Y., Oliver, S., Sizemore, C., Smethurst, M., et al. (2009, May 31-June 9). *Media meets Semantic Web-how the BBC uses DBpedia and linked data to make connections*. Paper presented at the 6th Annual European Semantic Web Conference (ESWC 2009), Heraklion, Greece.

Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006, August 23-25). *HT06, tagging paper, taxonomy, Flickr, academic article, to read.* Paper presented at the 17th Conference on Hypertext and Hypermedia, Odense, Denmark.

Mathes, A. (2004). *Folksonomies: Cooperative classification and communication through shared metadata* (Computer Mediated Communication, LIS590CMC). [Graduate School of Library and Information Science, University of Illinois Urbana-Champaign.]. *Urbana (Caracas, Venezuela)*, IL.

McAfee, A. (2006). Enterprise 2.0: The dawn of emergent collaboration. *MIT Sloan Management Review*, *47*(3), 21–28.

Merholz, P. (2004, October 19). Metadata for the masses. Retrieved May 8, 2009, http://www.adaptivepath.com/ideas/essays/archives/000361.php

Mika, P. (2005, November 6-10). Ontologies are us: A unified model of social networks and semantics. In Y. Gil, E. Motta, V. R. Benjamins, & M. A. Musen (Eds.), *The Semantic Web: ISWC 2005: 4th International Semantic Web Conference,* Galway, Ireland (LNCS 3729, pp. 522-536).

Newman, R., Ayers, D., & Russell, S. (2005). *Tag ontology*. Retrieved May 8, 2009, from http://www.holygoat.co.uk/owl/redwood/0.1/tags/

O'Reilly, T. (2005, September 30). *What is Web 2.0: Design patterns and business models for the next generation of software*. Retrieved May 8, 2009, from http://www.oreillynet.com/lpt/a/6228

Passant, A. (2008, June). Case study: Enhancement and integration of corporate social software using the Semantic Web. *Semantic Web use cases and case studies*. Retrieved May 8, 2009, from the W3C Semantic Web Education and Outreach Interest Group's Web site: http://www.w3.org/2001/sw/sweo/public/UseCases/EDF

Passant, A., & Laublet, P. (2008, April 22). Meaning Of A Tag: A collaborative approach to bridge the gap between tagging and linked data. In *Proceedings of the Linked Data on the Web Workshop (LDOW 2008),* Beijing, China (Vol. 369). CEUR Workshop Publishers.

Passant, A., Laublet, P., Breslin, J. G., & Decker, S. (2009, April 21). Semantic search for Enterprise 2.0. Paper presented at the Workshop on Semantic Search (SemSearch 2009) at the 18th International World Wide Web Conference (WWW 2009), Madrid, Spain.

Raimond, Y., & Sandler, M. (2008, September 14-18). *A web of musical information*. Paper presented at the Ninth International Conference on Music Information Retrieval (ISMIR 2008), Philadelphia.

Schmitz, P. (2006, May). *Inducing ontology from Flickr tags*. Paper presented at the Workshop on Collaborative Web Tagging at the 15th International World Wide Web Conference (WWW 2006), Edinburgh, Scotland.

Story, H., Harbulot, B., Jacobi, I., & Jones, M. (2009, June 1). FOAF+TLS: RESTful authentication for the social Web. In *Proceedings of the First Workshop on Trust and Privacy on the Social and Semantic Web (SPOT 2009)*, Heraklion, Greece (Vol. 447). CEUR Workshop Proceedings.

Tanaka, J., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder. *Cognitive Psychology*, *23*(3), 457–482. doi:10.1016/0010-0285(91)90016-H

Taylor, A. G. (1999). *The organization of information*. Santa Barbara, CA: Libraries Unlimited.

Van Damme, C., Hepp, M., & Siorpaes, K. (2007). FolksOntology: An integrated approach for turning folksonomies into ontologies. In *Proceedings of the Workshop on Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007),* Innsbruck, Austria (pp. 55-70). University of Kassel.

VanderWal, T. (2007, February 2). *Folksonomy coinage and definition*. Retrieved May 8, 2009, from http://www.vanderwal.net/folksonomy.html

Alexandre Passant is a postdoctoral researcher at the Digital Enterprise Research Institute, National University of Ireland, Galway. His research activities focus around the Semantic Web and social software; in particular, how these fields can interact with and benefit from each other in order to provide a socially-enabled machine-readable Web, leading to new services and paradigms for end-users. Prior to joining DERI, he was a PhD student at Université Paris-Sorbonne and carried out applied research work on "Semantic Web technologies for Enterprise 2.0" at Electricité De France. He is the co-author of SIOC, a model to represent the activities of online communities on the Semantic Web, the author of MOAT, a framework to let people tag their content using Semantic Web technologies, and is also involved in various related applications as well as standardization activities.

Philippe Laublet worked as senior researcher at the French Aerospace Research Institute (ONERA) in fields relating to artificial intelligence (1986-1996). The main topic of his PhD was automatic theorem proving. He has been a Senior Lecturer (associate professor) at the University of Paris-Sorbonne (Paris IV) since 1994 and is a member of the LaLIC laboratory. His research topics and interests have successively encompassed object-oriented methods and

languages, knowledge engineering and modelling, ontologies, semantic annotation, Semantic Web and Web 2.0.

John Breslin is currently a lecturer in electronic engineering at the National University of Ireland, Galway. He is also an associate researcher and leader of the Social Software Unit at the Digital Enterprise Research Institute at NUI Galway, researching semantically-enabled social networks and community portals. He is the founder of the SIOC project, which aims to semantically- interlink online communities. He has received a number of awards for Web site design, including two Net Visionary awards from the Irish Internet Association for the Irish community website boards.ie, which he co-founded in 2000. Dr. Breslin is a member of the IEI, IET and IEEE.

Stefan Decker is a professor at the National University of Ireland, Galway, and director of the Digital Enterprise Research Institute, an institute with more than 130 people working on the Semantic Web. Previously he worked at ISI, University of Southern California (2 years, research assistant professor and computer scientist), Stanford University, Computer Science Department (Database Group) (3 Years, postdoc and research associate), and Institute AIFB, University of Karlsruhe (4 years, PhD student and junior researcher). He is one of the most widely cited Semantic Web scientists, and his current research interests include semantics in collaborative systems, Web 2.0, and distributed systems.