# Integrating Tagging into the Web of Data: Overview and Combination of Existing Tag Ontologies

*Hak-Lae Kim[1], Simon Scerri[2], Alexandre Passant[2], John G. Breslin[3], Hong-Gee Kim[4]*
[1]*Samsung Electronics Co., Korea*
[2]*Digital Enterprise Research Institute, National University of Ireland, Ireland*
[3]*School of Engineering and Informatics, National University of Ireland, Galway, Ireland*
[4]*Biomedical Engineering Lab, Seoul National University, Korea*
*haklae.kim@gmail.com, {simon.scerri,  alexandre.passant}@deri.org, john.breslin@nuigalway.ie, hgkim@snu.ac.kr*

## Abstract

As the number of social websites offering tagging facilities increases, tagging has become not only a common basis for user participation, but also an important aspect of social content. Tagging is primarily based on the user's participation and interaction, including the sharing and the exchange of their interests. However, even though users can collect and share tagging data with conventional technologies, this data is fragmentary and remains disconnected. Therefore, discovering content across independent sites remains a challenge. On the other hand, technologies of the Web of Data, and especially ontologies, help towards this goal of integrating tagged data.

In this paper, we investigate the core features of existing tag ontologies that attempt to represent a common conceptualization of tagging and that provide well-defined semantics for tagging data, allowing for the sharing and interlinking of tagging data across independent sites or communities.

**Keywords:** Tagging, Folksonomy, Ontology, Semantic Web, Linked data, State of the art.

## 1   Introduction

A culture of mass participation on Web 2.0 has created large numbers of resources such as photos, bookmarks, video clips, blog posts, etc., which have become a medium for social interaction among people. Most social sites resort to tagging as a means of organization and classification of shared data, allowing the users themselves to assign a set of tags to the resources [8][20]. Thus tags can be a common basis for searching various types of resources across different social sites, since they are used for describing social content and are attached to these resources. However, despite the popularity of tagging, many issues with tagging arise: the inherent ambiguity of language in the meaning of terms, given a proliferation of synonyms, variations of spelling and phrasing [17][21][30]. Although RSS syndication and public APIs attempt to expose and share this data, they cannot make sense of the tagging context.

The context, in general, consists of all tagging entities (e.g., users, tags, and resources) as well as the relationships among these entities via a collective and aggregative manner [11][28]. Presently, this context is available only on the host sites and it cannot be shared or used externally [6]. Thus although we can collect tagging data from heterogeneous social sites, it still remains a challenge to create new knowledge by sharing these data. These limitations are primarily associated with the representation of tagging data in an inconsistent manner. Since there is no standardized data format to describe tagging data and practices, tagging data from heterogeneous sites cannot easily be aligned. This makes it difficult to compare, connect, interlink, or integrate the data with its own context across these sources.

In this paper, we propose the combination of existing tag ontologies. Since tagging data is ill-structured and lacks semantic control, a consistent way of exposing and accessing tagging data is a key to support interoperation amongst heterogeneous sources [28]. Within this federation, user-contributed content is expressed in an appropriate representation by using RDF(S)/OWL vocabularies. This allows users to share and discover tagging data across sites.

This paper is organized as follows. We start by providing formal conceptual modeling of the different objects we are focusing on within this paper including tags, tagging actions and folksonomies. Section 3 provides an overview of work that defines semantics for tags and tagging. Here we also present a number of available tag ontologies, focusing on the given common framework and guidelines for their use. After presenting the intended purpose of each of these ontologies, in Section 4 we propose the federation and alignment of a number of these ontologies in order to be used for the modeling of the tagging process. We finally conclude the paper.

## 2   Conceptualising Tagging and Folksonomies

One of our main goals is the portability of tagging data across different platforms, i.e., allowing users to move their tagging data from one site to another and to make

*Corresponding author: Hak-Lae Kim; E-mail: haklae.kim@gmail.com

them interoperate. In order for this to be possible, there must be a common understanding of what tags, tag clouds, folksonomies and the tagging process itself entails. In this section we formalize this common understanding in the form of a number of topologies of a tagging model at different levels. Three entities (users, tags, and resources) in a *tripartite model* make up a tagging activity [13][19-20][22]. This is often called either a tagging or folksonomy model, depending on an interpretation of tagging features.

Gruber [12] defined *Tagging* as the action of associating tags with an object or item and indeed proposed the conceptual model for tagging activities. In this model *Tagging* is a combination of the following core concepts:

• *Object*, i.e., the resource to be tagged. For example, a bookmark, a picture, a blogpost etc.
• *Tag*, i.e., the tag associated with the resource
• *Tagger*, i.e., the agent - usually a person - that creates the link between the *Tag* and the *Object*
• *Source*, i.e., the space where the tagging action has been performed, e.g.,: Flickr, Delicious
• *Polarity*, i.e., a vote for or against the assertion of the tagging for the purpose of solving spamming issues in tagging systems

We will now provide formal definitions modeling different aspects of tagging practices.

A simple tagging activity is modeled as follows:

$$T_s = (u, t, r, s) \tag{1}$$

where $T$ is a single instance of tagging, u, t, r, and s refers to a single user, a tag, a resource, and a tag space respectively. That is, a tagging activity occurs when a single user assigns a single tag to a particular resource in a specific space or tagging system (e.g., Flickr, Delicious, etc.).

Equation (1) does not take into account the *Polarity* aspect introduced by Gruber's model [12]. In Figure 1, $T_2$ and $T_3$ are subsets of $T_s$. $T_1$ has the tag "car" and "engine" to describe "Photo#1" in *Flickr*. Using this model we can derive a *personomy P* which is a collection of tagging instances for a particular user u as follows:
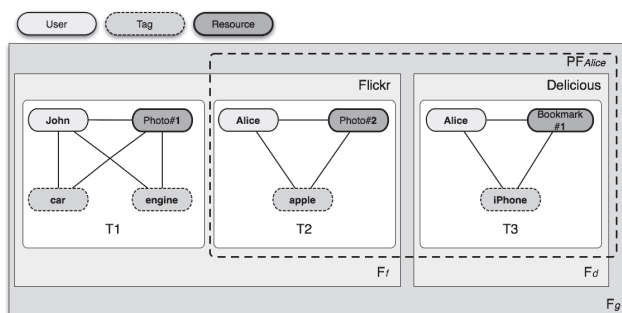


Figure 1 The Topologies of Tagging and Folksonomy

$$P_u = (u, T, R, s, Y) \tag{2}$$

where $T$ refers to a set of tags, $R$ refers to a set of resources, and $Y$ refers to an assignment of tags to a resource (i.e., $Y \subseteq U \times T \times R$). Thus this is an aggregation of tagging activities for the user u in a single space s. While this model does not assume any social interaction amongst users, it could be extended to describe social features of tagging, where multiple users apply a set of tags to a single resource (e.g., a bookmark in Delicious).

A folksonomy is comprised of a collection of personomies (i.e., $\sum P_u$). Therefore a straightforward model [12] for a folksonomy $F$ could be defined as follows:

$$F = (U, T, R, s, Y) \tag{3}$$

where $F$ is a folksonomy such that a number of users participating in a tagging activity may share multiple tags in a single site. Most of the current tagging systems can be described given this model. In Figure 1, $F_f$ of Flickr and $F_d$ for Delicious are subsets of $F$.

However, the tag space is unique since folksonomies can emerge not only from a single platform but also from distributed sources or applications such as $F_{d,f}$ -- which is a platform-independent folksonomy. Thus, the platform-independent-folksonomy model $F_I$ is obtained by changing the fourth parameter in Equation (3) to represent multiple sites or spaces. $F_I$ may be built through merging and interlinking among tagging entities across sites or applications, even if there are no existing systems to satisfy this model at present. Using this model, user-specific folksonomies $F_u$ can be driven by merging tagging activities of the same user $u$ in different sources. For example, $F_{Alice}$ is consolidated with both $P_{Alice}f$ in Flickr and $P_{Alice}d$ in Delicious.

A particular aspect that is not explicitly included in those models is that of the meaning of each tag, as used in a particular tagging context. Indeed, when users assign a particular keyword to a document, they have a particular associated meaning in mind. For instance, when adding the tag '*apple*' to a picture, the user might think either of a computer or a fruit. Hence, the first definition can be extended to take this particular aspect into account as follows:

$$T_{sm} = (u, t, r, s, m) \tag{4}$$

where m identifies the meaning of the tag t in the tagging action.

Consequently, the previous definitions can also be extended to take into account this aspect when defining folksonomies and personomies.

# 3   Overirew of the Semantic Web

The Semantic Web is the extension of the World Wide Web that enables people to share content across heterogeneous sources [1]. It has also been described as "Web of Data [2]" that not only makes links to be explored for both humans and machines, but also integrate documents and data shared by the links.

The basic parts of the Semantic Web are expressed in formal specifications, which aim to represent a wide range of metadata, and syntax for serializing and exchanging that model. For example, The Resource Description Framework (RDF) is a framework for representing information on the Web, and allows interoperability among applications exchanging machine-understandable information on the Web. There are some RDF vocabularies that are becoming prevalent:

- The Dublin Core Metadata Element Set is a standard for cross-disciplinary resource description [32]. The RFC (Request for Cooperation) 2413 describes the semantics of the 15 elements of the Dublin Core (title, subject, description, creator, publisher, contributor, date, type, format, identifier, source, language, relation, coverage and rights)[1]. These elements make use of XML or RDF mixed by other standards.
- The Friend of a Friend (FOAF) is one of the earliest and most popular projects on the Semantic Web. FOAF vocabularies allow users to describe their social networks that who know each other in communities or Websites (e.g., foaf:knows between foaf:Person). Using this ontology, user profiles in many social networking websites can be produced by Semantic Web profiles [9].
- The Semantically Interlinked Online Communities (SIOC) project aims to "enable integration of online community information [5]." The SIOC provides Semantic Web ontology for interconnecting data from online communities (e.g., blogs, forums and mailing lists, etc.) in RDF. SIOC is used in conjunction with the FOAF vocabulary for expressing personal profile and social-networking information [4].

According to Breslin and Decker [6], the Semantic Web provides the required representation and linking mechanisms to enable navigation between social networks across different domains. Online resources can be semantically represented by linking these vocabularies. In addition, a number of concepts and properties in these vocabularies can be reused for describing folksonomies at a semantic level.

# 4   Tag Ontologies

After formally defining the common conceptual groundwork for tagging systems, in this section we review different models that represent tags, folksonomies and related tagging activities thanks to Semantic Web technologies, in the light of the conceptualizations just presented.

A lot of scholarly work has been done on the topics of folksonomies and the Semantic Web, but relatively few studies have been carried out on interlinking among existing tag ontologies. Having Semantic Web models to represent tags and their related objects such as tagging actions, taggers or tagclouds, is indeed a key requirement to make them part of the Web of Data [2-3]. As such these models do not consider tags to be simple keywords, but elements of the Semantic Web represented by proper URIs, that can be interlinked (especially across heterogeneous and distributed tagging applications) and linked to (and from) any other kind of RDF data.

In our review we do not discuss approaches that try to extract taxonomies or ontologies from tagging systems, as proposed for instance by Halpin [13] or Mika [22].

### 4.1   The Tag Ontology

Newman [24] defined an ontology of tags and tagging, simply called the Tag Ontology, that describes the relationship between an agent, an arbitrary resource, and one or more tags (see Figure 2). Thus, in his ontology, the three core concepts *Taggers*, *Tagging*, and *Tags* are used to represent the tagging activity. Contrary to the previous model [12], it does not represent the source of the tagging action. Yet, it has been implemented (in OWL) and is available on the Web[2].

This ontology can be considered a basic model for representing tagging data, and in fact other models have extended it, as we will see later. An important feature of this ontology is that tags are represented as instances of the *tag:Tag* class which is assigned custom labels, i.e., the string representing the tag as seen by the user.
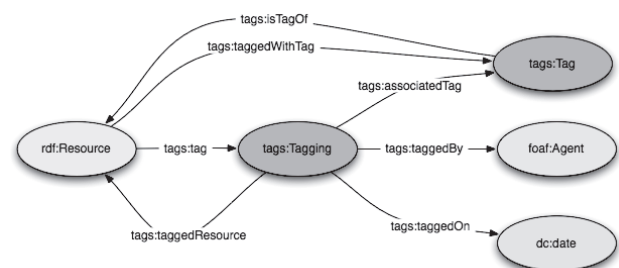


Figure 2 A Tagging Activity as Defined by the Tag Ontology

Being instances of a class means that they are assigned a URI. URIs are a key feature of the Semantic Web since, contrary to simple literals, they can be used as subject of triples, whereas literals can be only used as objects. Therefore this ontology serves a first requirement to make tags and tagging part of the Web of Data. Moreover, thanks to these URIs, tags can be linked together and people can semantically represent connections and similarities between tags in order to organize them. For this purpose the ontology introduces a *tag:related* property, defined as a subproperty of *skos:semanticRelation* from the SKOS vocabulary [23].

Yet, this relation does not have much semantics, since it does not define the nature of the relation, e.g., if two tags are related because of a linguistic variation or because they identify similar topics. Another limitation is that the ontology does not define any cardinality constraint on the number of labels a Tag can have. This can raise problems since it allows a Tag instance to have two completely disjoint labels (i.e., a Tag instance with labels "RDF" and "Paris"), which makes no sense from a tagging point of view.

Still, this ontology reuses pre-defined Semantic Web vocabularies, making it compliant with existing standards. As we mentioned, the property to define related tags extends what was proposed in SKOS and the Tag class itself inherits from *skos:Concept*. DublinCore is used to represent the date of a tagging action, with subproperties of *dc:date*. Finally, the ontology relies on FOAF to identify the tagger of a tagging action thanks to the *foaf:Person* property.

## 4.2 SCOT

SCOT[3] -- Social Semantic Cloud of Tags -- describes folksonomic characteristics in order to provide social interoperability of semantic tag data across heterogeneous sources [14]. This model can express the structure, features of and relationships between tags and users. It allows the exchange of semantic tag metadata for reuse in social applications and enables interoperation amongst data sources, services or agents in a tag space. In general, users may have a number of tagging instances across independent sites. As we discussed, folksonomies can be created by aggregating these activities. To semantically represent them, tagging activities are described in a collective model, aggregating all tagging instances and their relationships.

SCOT offers a collection of basic terms. Both Tagcloud and Tag class in SCOT play a role in their presentation of the social and semantic context of tagging, since both classes include users, tags, and resources and additional information to clarify tags' semantics. *scot:TagCloud* has

properties that describe a certain user (*scot:hasUserGroup*), number of tags (*scot:totalTags*), posts (*scot:totalPosts*) and co-occurrences (*scot:totalCooccurringTags*) and their frequencies, as well as updated information (*scot:updated*).

Notably, this defines the source (*scot:tagspace*) as the scope of namespaces or universe of quantification for objects. This allows one to differentiate between tagging data from different systems and is the basis for collaborative tagging across multiple applications. The property *scot:contains* links *scot:TagCloud* to a set of *scot:Tag* instances. *scot:Tag*, as a subclass of tag:Tag from the TagOntology, describes a tag that is aggregated from individual tagging activities.

It is important to note that SCOT uses concepts and properties from Newman's model. As shown in Figure 2, the Tagging class represents tags themselves, the resources that are being tagged, and the users that create these tags(*tag:taggedBy*). The *scot:TagCloud* class connects *tag:Tagging* instances via the property *scot:taggingActivity*. In SCOT, the range values of tagging properties are defined more specifically. For instance, *tag:taggedResource* has *sioc:Item* as a range value whereas *tag:associatedTag* has *scot:Tag* as its range. Individual tags in *tag:Tagging* are mapped to a source with *scot:Tag* instance and then these tags are represented by a collection of tags underlying a *scot:TagCloud*. The property *scot:taggingAccount* represents user accounts within online services. Figure 3 illustrates the SCOT ontology model integrated with Newman's model.
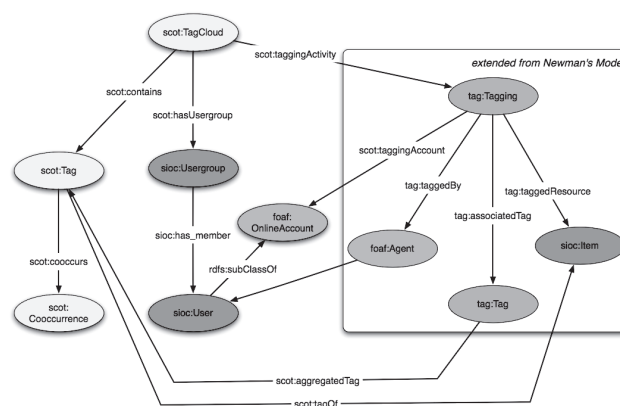


Figure 3 A Simplified Folksonomy Model in SCOT

## 4.3 MOAT

MOAT's -- Meaning Of A Tag[4] -- goal is to provide a Semantic Web model to define the meaning of tags in a machine-readable way [25].

To achieve it, MOAT defines:
- The *global meanings* of a tag, i.e., the list of all meanings than can be related to a tag in a complete folksonomy;

3 http://scot-project.org

4 http://moat-project.org

• The *local meaning* of a tag, i.e., the meaning of a tag in a particular tagging action.

For instance, the tag "*apple*" can mean -- depending on the user, the context and other factors -- a computer brand, a fruit or a record label. Hence, taken out of context, this tag has different *global meanings*. Yet when someone uses it in a tagging action, it has a particular *local meaning*, for example the computer brand.

Using MOAT, these meanings (both *global* and *local*) can be defined without ambiguity by the tagger. MOAT provides a machine-readable format, in the form of an OWL-Lite ontology, to allow computers to understand these meanings, relying on URIs of existing concepts from knowledge bases as DBpedia[5], GeoNames[6], or even corporate knowledge bases to define it.

Figure 4 shows how MOAT models these meanings and reuses the Tag Ontology. MOAT introduces a Tag class as a subclass of Newman's Tag class. These subclass addresses one of the problems of the Tagging Ontology we referred to earlier, and through an OWL cardinality constraint it is only allowed to have one unique label for a given Tag instance. Each tag is linked to one or more *moat: Meaning* instances, which represent the meaning(s) of a tag without any context. Each meaning must have one unique URI identifying it (e.g., http://dbpedia.org/resource/Paris), and be linked to the agents that defined this meaning, relying on FOAF.

To represent the context of a tag in a certain tagging action, using the quadripartite model defined before, MOAT relies on the *tag:RestrictedTagging* class from Newman's ontology, and introduces a *moat:tagMeaning* property. This allows linking to the meaning of the tag in this particular context.

Thanks to this framework and its model, MOAT aims to provide an easy way to bridge the gap between free-tagging and semantic indexing. While users can still benefit from the simplicity of free-tagging when annotating content, linking to URIs offers a way to solve tagging ambiguity (a single tag can be related to different URIs) and heterogeneity (various tags can be related to a single URI). Moreover, using MOAT, tagged content can be linked to URIs of reference datasets, leveraging tags and tagged content to the Web of Data not only by modeling tagging actions in RDF, but by linking those actions to other resources. Subsequently relationships defined in those datasets can be used to suggest relevant content, e.g., suggesting posts tagged "paris" (using dbpedia:Paris) from posts tagged "france" (using dbpedia:France) since related concepts are interlinked in DBpedia, solving the problem of lack of organisation in tag systems.

MOAT also defines a *moat:taggedWith* property so that a direct link can be established between the tagged resource and the concept, i.e., the meaning of the tag used. This property is purposely not a subproperty of *tag:taggedWithTag*, as it can be used to model links to various types of tags (for instance a 'GPL' to mention the license of a project), as we will explain later in this section.
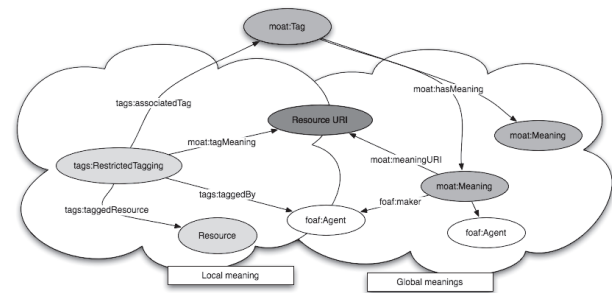


Figure 4 Tags' Local and Global Meaning in MOAT

### 4.4 Other Related Models

In addition to the different models introduced before, other vocabularies can be considered to model tags or tagging activities.

Knerr [16] defines the concept of tagging in the "Tagging Ontology," where the core element of the ontology is in fact 'Tagging.' The ontology provides concepts for the representation of the tagging time, user, domain, visibility, tag, resource, and type. Although his ontology covers different aspects (e.g., visibility), the main ideas and concepts largely overlap with those in Newman's model. Moreover, we were not able to find any tool using the ontology. Echarte et al. [7] proposed an OWL-DL model of folksonomies by extending concepts defined by Knerr above[7]. As for the previous model, we did not find any indication of their use at the time being.

SKOS can also be directly used to model tags using the *skos:Concept* class and relationship to the tagged content using the *skos:subject* property. Yet, this property was present in the SKOS Core module, which has recently been deprecated in favor of the new SKOS Vocabulary[8]. In SIOC, the *Tag* class can also be used to represent tags, and the *sioc:topic* property can be used to indicate the keywords assigned to any instance of *sioc:Item*, hence allowing to link to any kind of Web 2.0 content. Moreover, SIOC defines a *sioc:User* class, that can be used to model the user that created the given tagged item [4].

While modeling tags via RDF vocabularies is a fairly recent idea, it is worth mentioning the Annotea project and especially its Annotea Bookmark Schema[9]. This

---

[5] http://dbpedia.org

[6] http://geonames.org

[7] http://eslomas.com/tagontology-1.owl

[8] http://www.w3.org/2004/02/skos/vocabs

[9] http://www.w3.org/2003/07/Annotea/BookmarkSchema-20030707

schema proposed a *bookmark:hasTopic* property and a *bookmark:Topic* class to establish a link between any annotated resource and its annotating topics. Moreover, it also provided a *bookmark:subTopicOf* property so that tags can be organized, a long time before SKOS.

Similarly, the Taxonomy module of RSS1.0[10] also allows the representation of taxonomic annotations related to RSS items, using a single *taxo:topic* property. However it does not define any particular class to represent the tag itself. Microformats rely on a *rel-tag*[11] microformat that can be used to mark-up keywords on HTML documents and define them as tags. Search engines, e.g., Technorati, can then extract these. The semantics in microformats are less powerful in comparison to what can be done with the previously mentioned RDF models, and with RDF in general. However data annotated with microformats can easily be converted to RDF to form part of the Web of Data thanks to GRDDL [29].

Dublin Core metadata provides a general property (i.e., *dc:subject*) to describe tags related to a particular resource. Many properties such as *tag:taggedWith*, *sioc:topic* and *taxo:topics* are a sub-property of *dc:subject*; the last two implying that the tags are considered as topics for the tagged resource. Yet, a tag can also refer to other kinds of metadata. For example, the tag '*Flickr*' is not a subject, but is about the service provider or website. In this perspective,

the *dc:relation* property would be more appropriate for describing tags[12] as well as the *moat:taggedWith* property introduced before. More recently, the NiceTag ontology has been defined, modeling tags as RDF named graphs, and focusing on representing the kind of link that exists between a tag and a resource, going further than a simple 'tagged' link (e.g., author, provenance, opinion, etc.) [18].

### 4.5 Comparaison of Tag Ontologies

Figure 5 summarizes the different characteristics of each tagging ontology given the following criteria [26]:

- *Tag*: the model defines a way to represent a tag object;
- *Tagging (simple)*: the model defines a way to represent a simple tagging relationship between a tag and a tagged resource;
- *Tagging (tripartite)*: the model defines a way to represent a complete tagging action, as a tripartite model between a tag, a resource and an agent;
- *Tagger*: the model defines a way to represent the agent responsible for the tagging action;
- *Tagcloud*: the model defines a way to represent a tagcloud with its characteristics (co-occurrence, etc.);
- *Tag meaning*: the model defines a way to represent the meaning of the tags formally (i.e., not as a textual string).

For each ontology, we acknowledge the presence of a criterion only if the model itself allows to model such an

| Ontology | Format | Tag | Tagging (simple) | Tagging (tripartite) | Tagger | Resource | Tag Cloud | Tag Meaning |
|---|---|---|---|---|---|---|---|---|
| Gruber's model | N/A | √ | x | √ | √ | √ | x | x |
| Newman's model | OWL (Full) | √ | √ | √ | x | √ | x | x |
| SCOT | OWL (Full) | √ | | √ | √ | √ | √ | - |
| MOAT | OWL (Lite) | √ | √ | √ | x | √ | x | √ |
| NAO | RDFS | √ | x | x | x | √ | x | x |
| Knerr's model | OWL | √ | √ | √ | √ | √ | x | x |
| Echarte's model | OWL (DL) | √ | √ | √ | √ | √ | x | x |
| SKOS Core | OWL (Full) | √ | x | x | x | x | x | x |
| SIOC | OWL (Lite) | √ | x | x | √ | x | x | x |
| Annotea | RDFS | √ | x | x | x | x | x | x |
| Taxonomy module | RDFS | √ | x | x | x | x | x | x |
| SIOC+SCOT+MOAT+ Tag Ontology | OWL-Full | √ | √ | √ | √ | √ | √ | √ |

Figure 5 Representative Features of Surveyed Tag Ontologies

---

[10] http://web.resource.org/rss/1.0/modules/taxonomy/

[11] http://microformats.org/wiki/rel-tag

[12] http://librarytechnz.natlib.govt.nz/2008/09/adding-tags-to-dc-metadata.html

object or relationship. For instance, the Tag Ontology uses the *foaf:maker* property to relate the tagger to a tagging action but does not define the Tagger object itself (which is done in FOAF). Hence we do not consider it to satisfy this criterion.

To the same extent, we do not consider that SCOT or MOAT model the tripartite tagging action as they rely on the Tag Ontology to do so. Although each ontology was designed to achieve a different goal, if combined together, SIOC, SCOT and MOAT provide a complete framework to model tags, tagging actions and related objects in the Web of Data.

In the next section, we will discuss their combination and its benefits for the end-user.

# 5   Combining Tag Ontologies

Folksonomies tend to be understood as a global information space consisting not just of collective sets of tags, but also of linked data among the objects. However at present, tags and folksonomies do not provide enough meaningful metadata to build a linked tag space. They are just strings denoting some type of concept for resources in a particular system. As a result, we need to help organize and represent them in a format so that the tagging data can be interlinked across different services or users.

The tag ontologies discussed in this paper could be considered as a solution to this problem. However as we stressed in the previous section, each ontology was designed for a specific use case. A single ontology does not satisfy some tagging process and queries to support all full-fledged aspects in terms of social tagging. Thus, there is a need to combine these ontologies, so that tagging data created using the different tagging ontologies can be integrated.

In addition, extensive vocabularies such as such as SIOC, FOAF, SKOS and DC need to be adapted to this federation. Since these vocabularies provide appropriate semantics for specific resources, they enable tagging data to link resources across domains and services. Combining RDF vocabularies has the advantages of providing an open, standardized access mechanism to enable people to share their data on the Web [14].

Summarizing, the goal of this federation is not only to create links between tag ontologies but also to build a linked tag space between different ontologies with appropriate semantics, using URIs via HTTP protocol. This approach allows us to offer effective and efficient ways of publishing, deploying, and exposing semantic tag metadata. For example, in a linked tag space, current isolated tag sets in a particular system can be integrated within the Web of Data and navigated from a blogger in WordPress

to her bookmarks in Delicious, or from a photo in Flickr to her videos provided by YouTube. Part of our work thus focused on how to federate these ontologies, and defining a use case for each stage of the process given this federation. We will consider the federation among SCOT, MOAT and SIOC, since the ontologies have been used and updated in communities and in various related applications.

## 5.1  Use Case: Functional Level

In order to represent how these ontologies, once aligned, could be used together, Figure 6 represents the following use case, and shows which ontologies are used at each stage:

- Alice and Bob both add the resources about a Semantic Web conference on Delicious and Flickr respectively
- Both of them assign different tags -- 'sw' and 'sparql' for Alice and 'semantic web' for Bob
- They both add a meaning to their tag, which, in spite of different tags is the same URI, i.e., *<http://dbpedia.org/resource/Semantic_Web>*
- Alice decides to export her tagcloud to her photo manager to add pictures of the conference
- Alice and Bob want to mix their tagging data to make their own customized folksonomy from the two different services

This scenario shows how tagging can be 'translated' from one site to another, while keeping all its properties thanks to the combination of the Tag Ontology, SCOT and MOAT; in order to annotate any SIOC-described content.
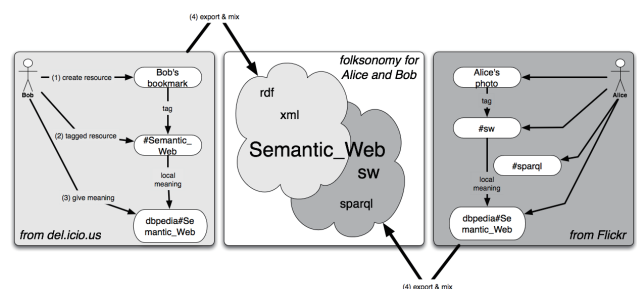


Figure 6 Use Case for Defining the Meaning of Tags and for Linking Resources

## 5.2  Use Case: Technical Level

Many social media sites now offer public APIs for free, meaning that it is very simple to retrieve or gather data on the sites using an application. However since most of them support RSS descriptions, the data as it is is not a good source of linked data. Beyond RSS, a simple way to create and expose semantic metadata from these sites is to use SIOC and FOAF, the latter being one of the most successful Semantic Web vocabularies. For example in FOAF the *foaf:weblog* property is linked to the weblog described in SIOC. But it is still a challenge to describe

semantic information for tags, although both vocabularies have properties describing categories or concepts (i.e., *sioc:topic*, *skos:concept*) which depict the topic of the resource. Therefore, tag ontologies we mentioned in this paper can be a candidate for describing semantic tag metadata that links to other RDF vocabularies. Following the results of our comparisons, Newman's model provides a nice basis for describing tagging ontologies. In fact, SCOT and MOAT are currently using it to represent Tags and Tagging. Moreover, SCOT and MOAT are suitable for linking SIOC and FOAF, because they are using some classes and properties from the ontologies.

As illustrated in Figure 6, Bob and Alice will create a resource, assign a tag(s) to the resource, give a meaning to this tag in that tagging context, and then export their tag cloud to another service using various ontologies that link to each other. This process follows the process to represent resources in SIOC, to define a meaning of tags in MOAT, and to export and merge all of them in SCOT. All of Alice and Bob's resources could be represented by SIOC instances, and they can be defined by a meaning of a certain tag underlying MOAT. It might be a candidate for providing linked tag data because such RDF vocabularies provide a method to interlink different sources at a semantic level.

After producing the metadata in the publishing and tagging stage, Alice and Bob might be able to semantically link each other's underlying tag meaning. This meaning would be shared between their bookmark and photo. However, these objects are just one instance where items tagged by Alice and Bob can share the same meaning. It is thus useful to expose each individual user's tagging

data semantically. These semantics, in the form of MOAT meanings, can then be shared between users with a social connection, supported by SCOT.

Figure 7 illustrates how the information in the use case can be realised. An individual tagging instance is represented by *tag:Tagging* class, and this class is mapped to the *sioc:User* and *sioc:Item* to describe taggers (i.e., Bob and Alice) and the resources (bookmark and photo). The tags 'semantic web,' 'sw,' and 'sparql' are linked to *moat:Meaning* to specify a certain meaning using the *moat:hasMeaning* property.

A TagCloud class in SCOT aggregates all tagging instances with their relevant information. At this level, tagging entities are represented with their collective features underlying their relationships. The *scot:composedOf* property is then used to merge Alice and Bob's TagCloud. The linked tag space for Alice and Bob consists of all tagging instances and the meaning of the tags, and each entity is interlinked within the space. Also, a platform-independent folksonomy can be obtained by mixing their personomies. Based on this environment, Alice and Bob can share their resources including users, resources, tags, and the meaning of tags.

## 6 Conclusion

In this article we discussed the need to integrate tagging data into the Web of Data, brought about by the fact that existing data sharing systems that utilise tagging to classify their data remain isolated at the knowledge representation level. Therefore tagging data remains fragmented across the
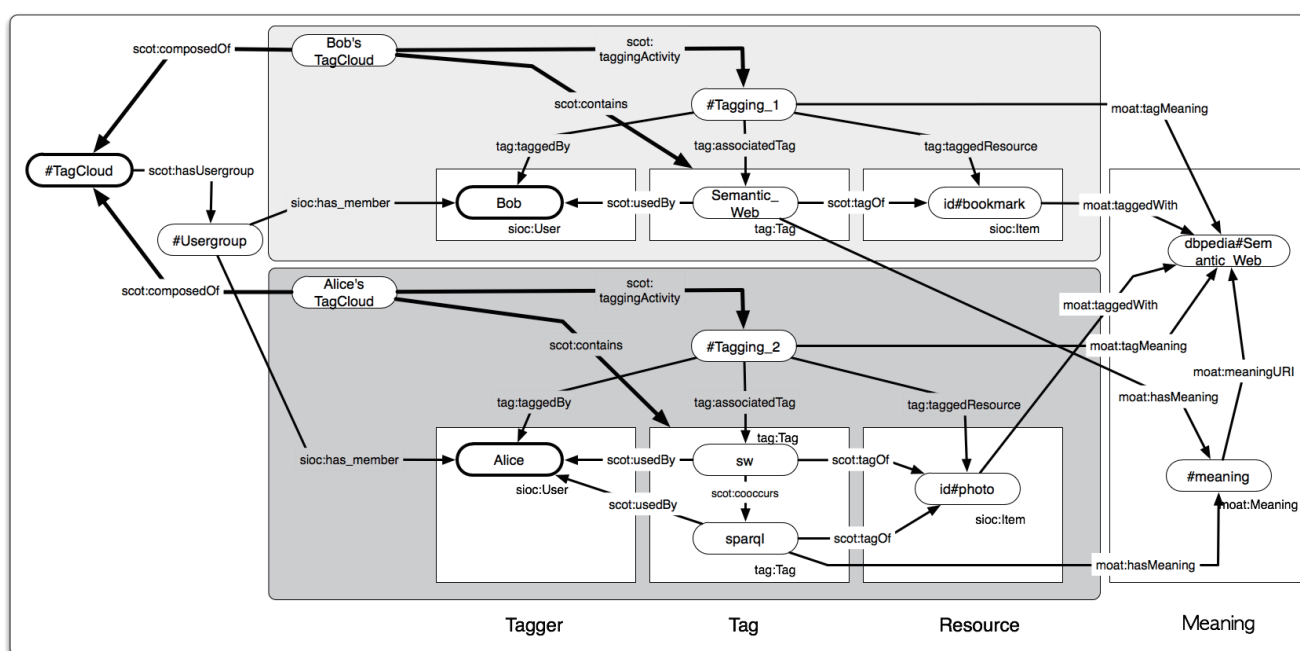


Figure 7 Federating Existing RDF Vocabularies to Achieve the Required Level of Representation

web. Tagging can be seen as a collective effort by multiple users to organise shared objects by topic. By limiting tagging data to the platform within which it was generated, we have been missing a great opportunity for interlinking users' collective knowledge across heterogeneous platforms. In order for this to become possible

● there needs to be a common understanding of the tagging process and its components

● tagging data must be represented in a format that enables its usability across the web

Our solution to the first problem was the formalization of tagging and folksonomy models, based on the continuation of studies dedicated to the conceptualization of tagging processes. We addressed the second problem by federating a number of existing tagging ontologies so as to provide a universal semantic representation for tagging data. On both functional and technical levels, user-contributed tagging data is grounded into existing Semantic Web standards and other established RDF vocabularies such as SIOC, FOAF, and DC. To support these requirements, a semantic tagging platforms needs to offer functionalities not only to describe tagging data in the form of ontological semantics but also to share them. These platforms need to combine technologies from both the Social and the Semantic Web.

## Acknowledgements

## References

[1]   Tim Berners-Lee, James Hendler and Ora Lassila, *The Semantic Web*, *Scientific American*, Vol.285, No.5, May, 2001, pp.34-43.

[2]   Tim Berners-Lee, *Linked Data*, 2006, http://www.w3.org/DesignIssues/LinkedData.html

[3]   Chris Bizer, Richard Cyganiak and Tom Heath, *How to Publish Linked Data on the Web*, 2007, http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedData Tutorial

[4]   Uldis Bojars and John G. Breslin, *SIOC Core Ontology Specification*, March, 2010, http://rdfs.org/sioc/spec/

[5]   Uldis Bojars, John G. Breslin, Vassilios Peristeras, Giovanni Tummarello and Stefan Decker, *Interlinking the Social Web with Semantics*, *IEEE Intelligent Systems*, Vol.23, No.3, 2008, pp.29-40.

[6]   John G. Breslin and Stefan Decker, *The Future of Social Networks on the Internet: The Need for Semantics*, *IEEE Internet Computing*, Vol.11, No.6, 2007, pp.86-90.

[7]   Francisco Echarte, José Javier Astrain, Alberto Córdoba and Jesús Villadangos, *Ontology of Folksonomy: A New Modeling Method*, *Proc. Semantic Authoring*, *Annotation and Knowledge Markup Workshop (SAAKM2007)*, *4th International Conference on Knowledge Capture (KCap 2007)*, Whistler, Canada, October, 2007.

[8]   Wai-Tat Fu, *The Microstructures of Social Tagging: A Rational Model*, *Proc. 2008 ACM Conference on Computer Supported Cooperative Work*, San Diego, CA, November, 2008, pp.229-238.

[9]   Jennifer Golbeck and Matthew Rothstein, *Linking Social Networks on the Web with FOAF: A Semantic Web Case Study*, *Proc. 23rd National Conference on Artificial Intelligence*, Chicago, July, 2008, pp.1138-1143.

[10]  Scott Golder and Bernardo A. Huberman, *The Structure of Collaborative Tagging Systems*, *Journal of Information Science*, Vol.32, No.2, 2006, pp.198-208.

[11]  Tom Gruber, *Collective Knowledge Systems: Where the Social Web Meets the Semantic Web*, *Web Semantics*, Vol.6, No.1, 2008, pp.4-13.

[12]  Tom Gruber, *Ontology of Folksonomy: A Mash-up of Apples and Oranges*, *International Journal on Semantic Web & Information Systems*, Vol.3, No.1, 2007, pp.1-11.

[13]  Harry Halpin, Valentin Robu and Hana Shepherd, *The Complex Dynamics of Collaborative Tagging*, *Proc. 16th International Conference on World Wide Web*, Banff, Canada, May, 2007, pp.211-220.

[14]  Hak-Lae Kim, *Leveraging a Semantic Framework for Augmenting Social Tagging Practices in Heterogeneous Content Sharing Services*, Ph.D. thesis, National University of Ireland, Galway, 2010.

[15]  Hak-Lae Kim, Stefan Decker and John G. Breslin, *Representing and Sharing Folksonomies with Semantics*, *Journal of Information Science*, Vol.36, No.1, 2010, pp.57-72.

[16]  Tonkin Knerr, *Tagging Ontology -- Towards a Common Ontology for Folksonomies*, 2006, http://tagont.googlecode.com/files/TagOntPaper.pdf

[17]  Ellyssa Kroski, *The Hive Mind: Folksonomies and User-Based Tagging*, 2005, http://infotangle.blogsome.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/

[18]  Freddy Limpens, Alexandre Monnin, David Laniado and Fabien Gandon, *Nicetag Ontology: Tags as*

*Named Graphs*, *Proc. International Workshop in Social Networks Interoperability, 4th Asian Semantic Web Conference*, Shanghai, China, December, 2009.

[19] Caimei Lu, Xin Chen and E. K. Park, *Exploit the Tripartite Network of Social Tagging for Web Clustering*, *Proc. 18th ACM Conference on information and Knowledge Management*, Hong Kong, China, November, 2009, pp.1545-1548.

[20] Cameron Marlow, Mor Naaman, Danah Boyd and Marc Davis, *HT06*, *Tagging Paper*, *Taxonomy*, *Flickr*, *Academic Article, to Read*, *Proc. Seventeenth Conference on Hypertext and Hypermedia*, Odense, Denmark, August, 2006, pp.31-40.

[21] Adam Mathes, *Folksonomies -- Cooperative Classification and Communication Through Shared Metadata*, 2004, http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html

[22] Peter Mika, *Ontologies Are Us: A Unified Model of Social Networks and Semantics*, *Web Semantics*, Vol.5, No.1, 2007, pp.5-15.

[23] Alistair Miles and Dan Brickley, *SKOS Core Vocabulary Specification*, 2005, http://www.w3.org/TR/swbp-skos-core-spec

[24] Richard Newman, Danny Ayers and Seth Russell, *Tag Ontology*, 2005, http://www.holygoat.co.uk/owl/redwood/0.1/tags/

[25] Alexandre Passant, Jean-David Sta and Philippe Laublet, *Folksonomies*, *Ontologies and Corporate Blogging*, *Proc. 4th Blogtalk Conference (Blogtalk Reloaded)*, Vienna, Austria, October, 2006.

[26] Alexandre Passant, *Semantic Web Technologies for Enterprise 2.0*, Ph.D. thesis, Universite Paris-IV Sorbonne, France, 2009.

[27] Alexandre Passant, Philippe Laublet, John G. Breslin and Stefan Decker, *A URI is Worth a Thousand Tags: From Tagging to Linked Data with MOAT*, *International Journal on Semantic Web & Information Systems (IJSWIS)*, Vol.5, No.3, 2009, pp. 71-94.

[28] *TagCommons Project*, *Ontologies Vs. Formats Vs. Schema Vs. APIs*, 2007, http://tagcommons.org/2007/03/02/ontologies-vs-formats-vs-schema-vs-apis/

[29] Harry Halpin and Ian Davis, *GRDDL Primer: W3C Working Group Note*, 2007, http://www.w3.org/TR/2007/NOTE-grddl-primer-20070628/

[30] Xian Wu, Lei Zhang and Yong Yu, *Exploring Social Annotations for the Semantic Web*, *Proc. 15th international Conference on World Wide Web*, Edinburgh, UK, May, 2006, pp.417-426.

[31] Zhijun Yin, Rui Li, Qiaozhu Mei and Jiawei Han, *Exploring Social Tagging Graph for Web Object Classification*, *Proc. 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris, June, 2009, pp.957-966.

[32] Stuart Weibel, *The State of the Dublin Core Metadata Initiative: April, 1999*, *D-Lib Magazine*, Vol.5, No.4, 1999, http://www.dlib.org/dlib/april99/04weibel.html

## Biographies

**Hak-Lae Kim** is a senior engineer of Media Solution Center at Samsung Electronics Co., LTD. He was a researcher at the Digital Enterprise Research Institute at NUI Galway, and his research efforts focused on creating social spaces based on tagging activities. He is founder of the SCOT (Social Semantic Cloud of Tags) ontology and project. He completed his BS in Business Administration, and MS and PhD in Management Information System from Dankook University in 2004 and 2008, and PhD in Computer Science from National University of Ireland, Galway in 2010.

**Simon Scerri** is a researcher at the Digital Enterprise Research Institute (DERI), within the National University of Ireland, Galway. His PhD research at this institute strived to structure ongoing communication processes in email-based digital collaborative work, in order to support the participants with their shared management. Also within DERI, he has participated in a number of projects, such as /Nepomuk: The Social Semantic Desktop/ (2006-2008) and the ongoing /Digital Me: Userware for the Semantic Personal Sphere/. Research topics he interested in included Web2.0, Semantic Web and Applications, Knowledge Modelling, Ontology Engineering, Text Mining, Information Extraction, Computational Linguistics, Text Analytics, Workflow Management, Information Visualisation, Human-Computer Interaction, Intelligent User Interfaces, Digital Collaboration and Communication Support Systems.

**Alexandre Passant** is a researcher at DERI, NUI Galway, leading the Social Software Unit. His research focuses on how the Social Web and the Semantic Web interact with and benefit from each other to provide a socially-enabled machine-readable Web, leading to new services and paradigms for end-users. He holds a PhD from Université Paris-Sorbonne.

**John G. Breslin** is a lecturer at NUI Galway. He is also an associate researcher at DERI, where he is researching the Social Semantic Web. Breslin is the founder of the SIOC project, which aims to interlink online communities with semantic technologies. He is co-founder of the community website boards.ie.

**Hong-Gee Kim** is an associate professor of medical informatics and the director of Biomedical Knowledge Engineering Lab at Seoul National University. His primary research efforts are in building real world applications by using ontology and semantic technologies in the e-Health domain. He is involved in various research projects including a semantic terminology server for electronic health record systems, a semantic search engine on a cloud computing environment, and social semantic digital library. Hong-Gee Kim completed his PhD and MS in Artificial Intelligence and Philosophy from the University of Georgia and his BS in Psychology from Korea University.