Chapter 26

Understanding Online Communities by Using Semantic Web Technologies

Alexandre Passant
National University of Ireland-Galway, Ireland

Sheila Kinsella
National University of Ireland-Galway, Ireland

Uldis Bojars
National University of Ireland-Galway, Ireland

John G. Breslin
National University of Ireland-Galway, Ireland

Stefan Decker
National University of Ireland-Galway, Ireland

ABSTRACT

During the last few years, the Web that we used to know as a read-only medium shifted to a read-write Web, often known as Web 2.0 or the Social Web, in which people interact, share and build content collaboratively within online communities. In order to clearly understand how these online communities are formed, evolve, share and produce content, a first requirement is to gather related data. In this chapter, we give an overview of how Semantic Web technologies can be used to provide a unified layer of representation for Social Web data in an open and machine-readable manner thanks to common models and shared semantics, facilitating data gathering and analysis. Through a comprehensive state of the art review, we describe the various models that can be applied to online communities and give an overview of some of the new possibilities offered by such a layer in terms of data querying and community analysis.

INTRODUCTION

Social media is now a part of the everyday lives

of people who are using Web technologies. People read and comment on blogs, participate in editing wiki pages, use social networking to interact with their friends (or to get new ones), and share pictures, memories and more via services such as

Flickr or YouTube: the whole paradigm often being known as Web 2.0 (O'Reilly, 2005). Moreover, this phenomenon goes further, for example, impacting research communities with services such as the Nature Network - http://network.nature. com/ - and enterprise information systems in a shift known as Enterprise 2.0 (McAfee, 2006). The more that both data and people interact and connect via Web 2.0, the more scientists (both from Social Science and Computer Science) try to understand how online communities are formed, how they evolve, what do they share, and what valuable information can be extracted from these analyses. Yet, the diversity of tools, communities and services makes the process of gathering the data, and consequently understanding these communities, a complex task. For each ecosystem, new algorithms must be built, new links must be mined, new applications must be designed, etc.

Nevertheless, another trend from the research community during the last ten years, the Semantic Web (Berners-Lee et al., 2001), aims to provide models for interoperable data between applications and can be of great interest for communities from the Social Web. By relying on standard models to represent data as well as shared semantics between applications, it offers a means to better integrate and query data from various systems, as well as creating links between them. Using Semantic Web technologies can help us to better understand these online communities, by providing common means to represent, link and mine information from various distributed systems and heterogeneous data sets, as emphasised by Figure 1.

Thus, the goal of this chapter is to provide readers - especially advanced undergraduate and graduate students in Computer Science, Social Science and more generally in Web Science (a term that we will describe later) – with a comprehensive state-of-the-art study on how Semantic Web technologies can be used to model, export and analyse virtual communities in distributed environments such as the Web or Web-based Information System (e.g. corporate organisations).

Figure 1. Using the Semantic Web to facilitate community analysis
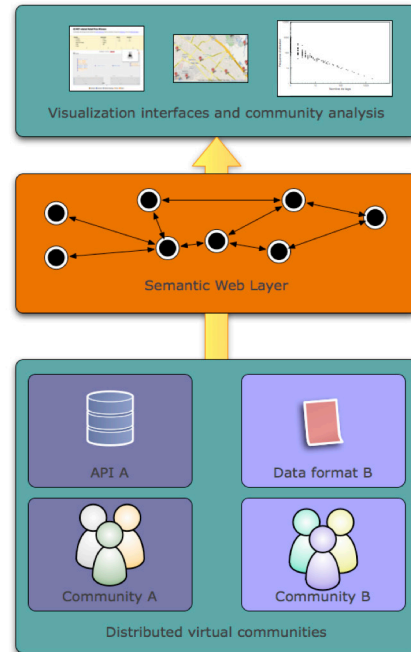
The chapter is structured as follows. In the first part, we will focus on current practices to understand and model virtual communities and their related content as well as describing the shortcomings of these approaches, such as relying on vendor-specific APIs. It will hence provide us with incentives to introduce the core of this chapter, i.e. the need for Semantic Web technologies for modelling virtual communities and identifying the advantages they offer regarding data and content analysis as well as interoperability between social applications. In the second section, we will then introduce Semantic Web principles and provide a comprehensive state–of-the-art review of existing models from the Semantic Web that are dedicated to Social Web data. In the third part, we will then discuss use cases on how to use these technologies to better understand communities. We will thus give the reader an overview of possibilities that are offered by such methods: querying communities, mining profiles from distributed social

networks, browsing social data, etc. In the fourth section, we will discuss some upcoming challenges and we will focus both on how Semantic Web technologies can be used to solve some of them and at the same time which challenges are still faced by the Semantic Web community in the context of understanding virtual communities, especially at Web scale. Finally, we will conclude the chapter with an overall discussion on how the Semantic Web could help to understand not only virtual communities, but also the Web in general. We will discuss the recent Web Science Research Initiative and will discuss how, in our opinion, the Semantic Web and the Web Science agenda relate to each other, and how Semantic Web technologies could help people to better understand the evolution and complexity of the Web and of Web-based information systems.

UNDERSTANDING VIRTUAL COMMUNITIES

From a Web to a Social Web

Since it was established, the Web has been used to enable communication not only between computers but also between people. Usenet newsgroups, mailing lists and web-based forums allowed people to connect with each other and thereby enabled communities to form, often around specific topics of interest. The social networks formed via these technologies were not explicitly stated, but were implicitly defined by the interactions of the people involved (e.g. by replying to each other). Later, technologies such as IRC (Internet Relay Chat), instant messaging and blogging continued the trend of using the Internet to build communities of interest.

One of the most visible trends on the Web is the emergence of Web 2.0-style services. The term Web 2.0 (O'Reilly, 2005) refers to a perceived second-generation of Web-based communities and hosted services. Although the term suggests a new version of the Web, it does not refer to an update of the World Wide Web technical specifications and architecture (TAG, 2004), but rather to new structures and abstractions that have emerged on top of the ordinary Web. Although it is difficult to define the boundaries of what structures or abstractions belongs to Web 2.0, there seems to be an agreement that services and technologies like blogs, wikis, folksonomies, podcasts, many-to-many publishing, social networking sites (SNSs), Web APIs, web standards and online Web services are part of Web 2.0. Web 2.0 has not only been a technological but also a business trend: according to Tim O'Reilly: *"Web 2.0 is the business revolution in the computer industry caused by the move to the Internet as platform, and an attempt to understand the rules for success on that new platform"* (O'Reilly, 2006).

In addition to participation features (through blogging, wiki participation, etc.), an important feature of the Web 2.0 meme is the online social networking aspect. Social networking sites such as Friendster (an early SNS previously popular in the US, now widely-used in Asia), orkut (Google's SNS), LinkedIn (an SNS for professional relationships) and MySpace (a music and youth-oriented service) - where explicitly-stated networks of friendship form a core part of the website - have become part of the daily lives of millions of users, and have generated huge amounts of investment since they began to appear around 2002. Since then, the popularity of these sites has grown hugely and continues to do so. (Boyd and Ellison, 2007) recently described the history of social networking sites, and suggested that in the early days of SNSs, when only the SixDegrees service existed, there simply were not enough users: *"While people were already flocking to the Internet, most did not have extended networks of friends who were online"*. According to Internet World Stats, between 2000 (when SixDegrees shut down) and 2003 (when Friendster became the first successful SNS), the number of Internet users had doubled.

Web 2.0 content-sharing sites with social networking functionality such as YouTube (a video-sharing site), Flickr (for sharing images) and Last.fm (a radio and music community site) have enjoyed similar popularity. The common features of a social networking site include personal profiles, friends listings, commenting, private messaging, discussion forums, blogging, and media uploading and sharing. Many content-sharing sites, such as Flickr and YouTube also include some social networking functionality. In addition to SNSs, other forms of social websites include wikis, forums and blogs. Some of these publish content in structured formats enabling them to be aggregated together.

A common property of Web 2.0 technologies are that they facilitate collaboration and sharing between users with low technical barriers – although usually on single sites or with a limited range of information. In this book we will refer to this collaborative and sharing aspect as the "Social Web", a term that can be used to describe a subset of Web interactions that are highly social, conversational and participatory, whereby social media content is being created and augmented on a variety of social media platforms. The Social Web may also be used instead of Web 2.0 as it is clearer what feature of the Web is being referred to, and we will use both in this chapter.

Finally, it is worth noticing that this social vision of the Web is actually closely aligned to the original vision of the Web, as Tim Berners-Lee noted in an interview with the BBC: *"The idea was that anybody who used the web would have a space where they could write and so the first browser was an editor, it was a writer as well as a reader. […] What happened with blogs and with wikis, these editable web spaces, was that they became much more simple. When you write a blog, you don't write complicated hypertext, you just write text, so I'm very, very happy to see that now it's gone in the direction of becoming more of a creative medium."* Since the beginning of the Web, that participation aspect was enabled. For example, the first Web browser Amaya was not just a read-only browser, but it also allowed one to edit pages from the browser (similar to methods now popularised by wiki interfaces).

Current Approaches for Data Mining and Analysis

The field of social network analysis (SNA) gives us a methodology for gaining insight into the structure of communities. Social network analysis uses methods from graph theory to study networks of individuals and the relationships between them. The individuals are often referred to as nodes or actors, and they may represent people, groups, countries, organisations or any other type of social unit. The relations between them can be called edges or ties, and can indicate any type of link, for example, acquaintance, friendship, co-authorship or information exchange. Ties may be undirected, in which case the relationship is symmetric, or directed, in which case the relationship has a specific direction and may not be reciprocated. Social network analysis enables us to discover information such as the key people in a network, the distinct communities in a network, and the different types of roles which occur in a network.

Apart from comprehensive textbooks in this area (Wasserman and Faust, 1994), there are many academic tools for visually examining social networks and performing common SNA routines. For example, the tool Pajek - http://vlado.fmf.uni-lj.si/pub/networks/pajek/ - can be used to drill down into various social networks. A common method is to reduce the amount of relevant social network data by clustering. One can choose to cluster people by common friends, by shared interests, by geographic location, by tags, etc. visualisations. Alternatively, a library like JUNG - http://jung.sourceforge.net/ - which provides analysis and visualisation methods, can be used to develop custom analytic or visual tools. In any case, before loading the data into one of these analysis tools, the relevant data must first

be converted to an appropriate representation, which is dependant on the tool used. For more on social network analysis, and on a Semantic Web framework for carrying out SNA, see the chapter titled "Semantic Social Networks Analysis, a Concrete Case".

Another approach for analysing online communities is using Natural Language Processing (NLP) algorithms to extract entities, topics and relationships from textual content generated by users. However when dealing with social media sites, performing NLP can be particularly difficult due to the typically informal nature of user posts, which tend to contain a lot of slang and context-dependant terms, with little attention given to spelling and grammar (Gruhl et al., 2009). Thus, while NLP algorithms are potentially very useful tools for investigating SNSs, there are challenges particular to user-generated content which must be handled.

When dealing with typical SNSs, even just acquiring the relevant data can require a lot of effort. A typical approach is to start from the profiles of a seed user or set of users, and follow the links to their friend's profiles, and friend-of-friends' profiles, and so on. Often it is necessary to download each user's profile as a HTML page, and then scrape the desired information. This process is time consuming and sometimes difficult. The code requires updating every time the structure of a page is changed, and needs to be completely rewritten for every new website which one wishes to investigate, since information is represented differently depending on the website.

As an alternative to scraping, many Web 2.0 sites provide APIs, for example LiveJournal, Twitter, Flickr and YouTube. The main motivation for providing APIs is to facilitate the integration of services into new applications or mashups. One can send requests to an API about a particular user, content item, or other resource, and the results are returned in a structured, easy-to-parse format. This makes the process of data acquisition much easier, at least when the data of interest is limited

to one site. However, if analysis requires data to be collected from multiple sites, integration can be problematic.

For example, let us consider two major applications for the Social Web, the first one being Twitter, the microblogging service, and the second one being Flickr, the photo sharing system. Both provide a public API that can be used by third-party developers to make their own applications, or to simply gather some data to analyse the communities (e.g. identifying social networks, groups the users belong to, etc.). Yet, these APIs are different both in terms of sending the request and parsing the results. While both of them are based on HTTP calls and provide common formats for the API output response (such as XML and JSON – JavaScript Object Notation, a popular format for exchanging structured data within Web 2.0 applications), they use different parameters and return values. For example, to identify all the people who are connected to a particular user on Twitter (and to get results using XML), one must call a URL pattern such as http://twitter.com/friends/ids/terraces.xml which retrieves results shown in Table 1.

On Flickr, a similar query is performed by calling a pattern such as http://api.flickr.com/services/rest/?method=flickr.contacts. getPublicList&api_key=f4c67b996f01077cf2e1 d1469a7e790f&user_ id=33669349%40N00&api_sig=c8c0fd49fe472 72e1410e9574c98096c and the results are shown in Table 2.

As one can see, there is no obvious relationship between these two models: while they share common properties in terms of data artefacts, such as the user id, this one is represented as an XML tag by itself using id on Twitter, but as an attribute nsid on Flickr, which makes integration complex. Moreover, the parameters that have to be passed to the API are also different. In the next chapter, we will see how Semantic Web technologies, providing uniform description of resources using RDF and ontologies, can be used to provide a

Table 1. Example of a social network retrieved using the Twitter API

```
< ?xml version="1.0" encoding="UTF-8"?>
<ids>
<id>30076775</id>
<id>59870857</id>
<id>19927027</id>
<id>39321347</id>
  [...]
</ids>
```

Table 2. Example of a social network retrieved using the Flickr API

```
<rsp stat="ok">
<contacts page="1" pages="1" per_page="1000" perpage="1000" total="6">
<contact nsid="32233977@N00" username="captsolo" iconserver="3223" iconfarm="4"
ignored="0"/>
<contact nsid="43184127@N00" username="cygri" iconserver="8" iconfarm="1" ig-
nored="0"/>
<contact nsid="14027651@N04" username="scorlosquet" iconserver="3291" icon-
farm="4" ignored="0"/>
    [...]
</contacts>
</rsp>
```

common layer of representation over such data, from heterogeneous APIs to standardised representation models (both in terms of how to get the data and how to understand it).

SOCIAL SEMANTICS TO THE RESCUE

The Semantic Web: An Introduction

When looking at the initial proposal that led to the World Wide Web (Figure 2) by (Berners-Lee, 1989), we can see that it links typed objects (people, projects, software, etc.) using various types of properties (describes, refers to, etc.). However, in spite of this initial proposal, so far the Web we are using is mainly a Web of documents (either text files or multimedia ones), linked together by untyped and unidirectional hyperlinks. While this is probably enough for human readers (who can interpret the content of these documents and the meaning of the links between these documents), the situation is far more complex for software agents. For example, one person reading the Wikipedia page about Paris can understand that this is a city and that a link to a page about France identifies that Paris is located in France, but there is no means for a software agent to understand anything about the nature of the objects described in these pages, despite the evolution of NLP algorithms that can be used to extract named entities and relationships from such pages.

Figure 2. The architecture proposal by Tim Berners-Lee that gave birth to the Web
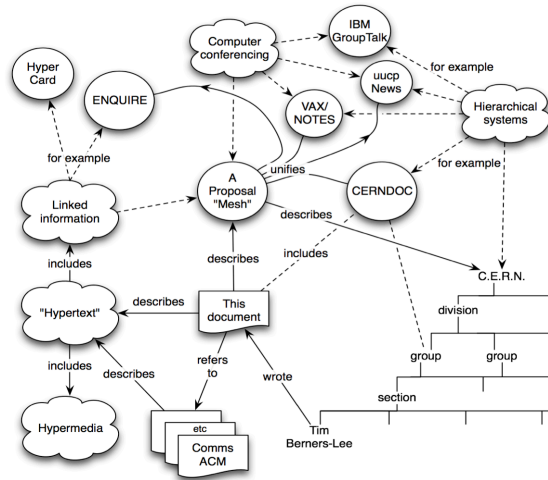

The Semantic Web vision aims to solve these issues by providing a Web of machine-readable information, with well-defined structure and semantics. The W3C (World Wide Web Consortium) recently termed this as "a Web of Data", in contrast to the Web of Documents, i.e. a Web in which data (not just documents) can be represented, exchanged and understood in a meaningful way. The Semantic Web is not a new Web, disconnected from the current one, but an *"extension of the current Web"* (Berners-Lee et al., 2001). While most of the current standardisation efforts around the Semantic Web have occurred via the W3C within their Semantic Web activity - http://www. w3.org/2001/sw/ - some older projects such as SHOE (Helfin and Hendler, 2000) have focused on similar ideas leading towards a machine-readable Web.

In order to achieve the Semantic Web goal, different technologies are needed, that form the complete Semantic Web layer cake, depicted in Figure 3. While we do not aim to provide a complete description of this stack, some particular elements must be understood before going further in this chapter.

A first component that enables the Semantic Web is the use of URIs – Uniform Resource Identifiers (Berners-Lee et al., 2005) - as identifiers for everything that is described on the Web: people, cities, communities, etc. These URIs act as Web-scaled identifiers for naming resources. For example, <http://dbpedia.org/resource/Galway> can be used to identify the city of Galway, (while <http://dbpedia.org/page/Galway> would identify a page about it). There can be multiple identifiers for the same resource, and we will

Figure 3. The Semantic Web "layer cake" from the W3C
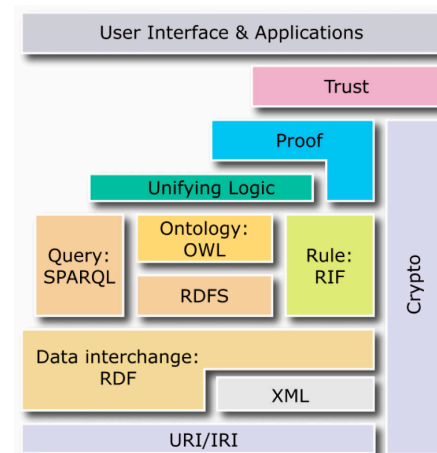
detail this in the context of online identity in a later section of this chapter.

Another requirement is to define facts or assertions about these URIs, for example, to say that Galway is a city. RDF – Resource Description Framework – provides a way to do so by defining a triples-based model in the form of <subject> <predicate> <object>, with <subject> and <predicate> being URIs, and <object> being either a URI or a literal. RDF is composed of both an abstract graph model and of various serialisations that map this abstract model to a machine-readable form. These serialisations include RDF/XML, N3, Turtle (being a subset of the previous one) and more recently RDFa for embedding RDF annotations directly inside XHTML documents. For example, the snippet of code (Turtle) shown in Table 3 represents that Galway is a city and that DERI is based in Galway, both having human-readable labels.

Thirdly, there is a need for shared vocabularies to represent the meaning of these URIs. Ontologies (Gruber, 1993) provide this additional layer of semantics, allowing us to define classes and properties as well as more advanced axioms. For example, one could define (in a formal way), that <http://example.org/Person> represents the concepts of Person and <http://example.org/name> represents a name property. Ontologies can be developed using desktop clients, such as Protégé - http://protege.stanford.edu/ - or using Web-based tools, like Neologism (Basca et al., 2008). Ontologies are usually published on the Web in order to be shared. Moreover, they can be extended remotely, i.e. one can create his or her own ontology extending existing ones. It is also important to mention that the goal of the Semantic Web is not to build a wide and unique Web-scale ontology, but rather to let everyone define their own and agree on a set of core ontologies, some of them being described in (Bizer et al., 2007). The two main languages used to represent ontologies on the Web are RDFS (Brickley and Guha, 2004) and OWL (Bechhofer et al., 2004), the latter being more expressive than the first one (e.g. providing cardinality constraints on properties).

Finally, SPARQL provides a graph-based query language (Prud'hommeaux and Seaborne, 2008) as well as a protocol (Clark et al., 2008) to retrieve information from RDF graphs, and it can be used to identify content created and shared in online communities as we will later describe. In the next sections of this chapter, we will describe how such languages can be used both to represent and to query online communities on the Web.

The Two Sides of the Social Semantic Web

While the Semantic Web vision has long been disconnected from the Social Web, an important trend of these last few years has been towards

Table 3. Example of RDF data (in N3 syntax)

```
@prefix dbpedia: <http://dbpedia.org/resource/>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
dbpedia:DERI foaf:based_near dbpedia:Galway ;
  rdfs:label "DERI" ;
  rdfs:label "Digital Enterprise Research Institute".
dbpedia:Galway a dbpedia:City ;
  rdfs:label "Galway" ;
  rdfs:label "Gaillimh"@ga.
```

considering how they could be integrated. Indeed, not only are they not meant to be disconnected but they can be efficiently combined, in a trend generally known as the Social Semantic Web (Breslin et al., 2009), representing the convergence of these two fields as the Web evolves. Thus, many researchers have demonstrated the usefulness of this convergence (Heath and Motta, 2007) (Ankolekar et al., 2008) and new trends have emerged such as Semantic Wikis (Buffa et al., 2008), Semantic Microblogging (Passant et al. 2008) or Semantic Social Network Analysis (Erétéo et al., 2009a). NLP also plays a part in this convergence, enabling relationships extracted from free text to enhance a knowledgebase, such as in the SOFIE system (Suchanek et al., 2009).

More generally, two different ways for combining the Semantic Web and the Social Web can be considered:

- On one hand, some efforts focus on using Semantic Web technologies to model social data. With models such as FOAF – Friend Of A Friend (Brickley and Miller, 2004) - and SIOC – Semantically-Interlinked Online Communities (Breslin et al., 2005) - that we will describe in the next subsection, Social Web data can be represented using shared and common models, and then it becomes more interoperable and portable between applications;

- On the other hand, leveraging the wisdom of the crowds from Web 2.0-based services is a perfect opportunity for creating a large amount of Semantic Web data. As pointed out by (Berners-Lee, 2005) *"I think we could have both Semantic Web technology supporting online communities, but at the same time also online communities can also support Semantic Web data by being the sources of people voluntarily connecting things together"*.

Combined together, these two paths allow us to overcome the *chicken-and-egg* problem with Semantic Web technologies: tools are needed to showcase the values of these technologies to end users, but data is also required to make these tools work properly. Hence, the integration of these two sides is twofold, and leads to the Social Semantic Web, bridging the gap between the Social Web and the Semantic Web (Figure 4).

Figure 4. The social Semantic Web
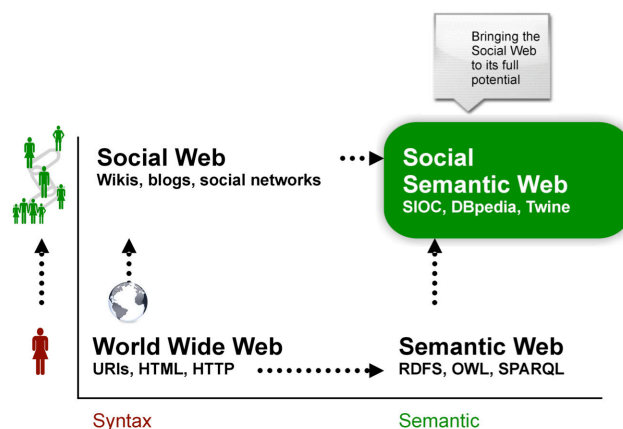
ONTOLOGIES FOR THE SOCIAL WEB: STATE OF THE ART

Social Networking

In order to represent social networking information using Semantic Web technologies, FOAF – Friend Of A Friend (Brickley and Miller, 2004) – http://foaf-project.org – is probably the most well-known ontology. It provides a model to represent people (with a foaf:Person) class, their properties and attributes (ranging from foaf:name to foaf:schoolHomepage) as well as a foaf:knows relationship which is used to represent social networking aspects. This last relationship is semantically weak, and to overcome this lack of precise semantics, the RELATIONSHIP vocabulary - http://vocab.org/relationship/ - provides a set of subproperties such as rel:colleagueOf or rel:lifePartnerOf to describe more precise relationships between people. Moreover, since ontologies can be extended in a distributed manner as we mentioned in the first section, anyone can create his or her own property, for example, wroteAChapterWith could be used to identify people in a social network as being co-authors.

In FOAF, each person is generally represented by its own URI, and information about a person is put on the Web in what is generally called a FOAF profile, i.e. a set of assertions about the person that put it online, acting as an online and semantic ID document. For example, the snippet shown in Table 4 identifies that Alex knows Sheila and that he works at NUI Galway:

As one can see in the previous example, URIs identifying people can come from different services, which provides a distributed and cross-platform way to define social networks, one key factor of FOAF and of Semantic Web technologies themselves. That is, we enrich the network value of a graph by linking information together at Web scale (Hendler and Golbeck, 2008) (Passant et al., 2009a). Merged together, different FOAF profiles can then be combined for an integrated view of one's social network, using profiles from different people and even different applications, as shown in the following picture. Here, we can see that three social networks have been defined on three different platforms, but are generally interlinked thanks to the use of common URIs to identify the people belonging to these social networks. Practically, this means that three different services, e.g. a weblog using WordPress, a Content Management System based on Drupal and a wiki using MediaWiki can unify their social networks, while being managed by three distinct applications that natively store their information using various heterogeneous models. Then, on the top of this unified social network, new services can be deployed, as we will see in the next sections.

Various services natively export their data using FOAF, such as LiveJournal - http://livejournal.com/ -, FriendFeed - http://friendfeed.com - or the Apache project directory - http://projects.apache.org/ - the last one in combination with DOAP - Description of a Project -, an ontology used to describe software projects. In addition, various exporters have been built for open-source applications, such as Drupal and WordPress, as well as for major Web 2.0 services, by mapping their vendor APIs to the FOAF ontology. There also exist exporters for Flickr, Twitter and even Facebook, providing common machine-readable

Table 4. Example of FOAF information

```
<http://apassant.net/alex> foaf:name "Alexandre Passant" ;
  foaf:knows  <http://sw.deri.org/~sheila/foaf.rdf#me> ;
  foaf:workplaceHomePage <http://www.nuigalway.ie/>.
```

Figure 5. Integrating FOAF data from different profiles
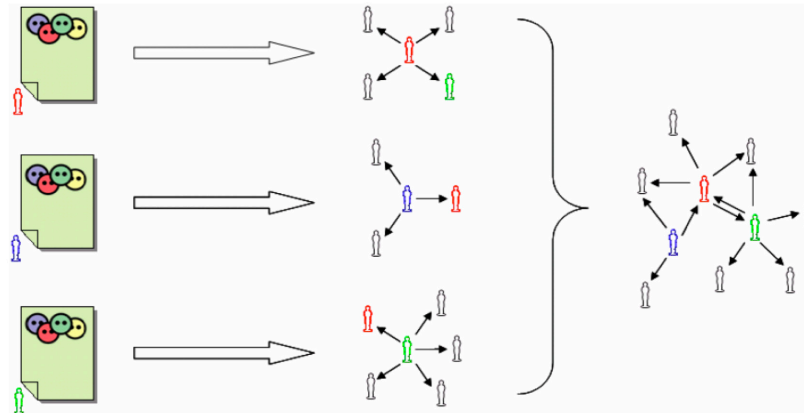
information about social networks from these services that can be uniformly browsed and queried, as we will see later. The snippet shown in Table 5 shows some Flickr social network information from the FOAF exporter, in RDF/XML.

Therefore, identifying some people that are known by someone else can be carried out using a single SPARQL query, no matter where the data comes from. The example in Table 6 shows how this can be done, and such a query can be applied to any dataset of RDF data using the FOAF ontology, provided for instance using the aforementioned exporters:

Two things must be understood from these examples, especially regarding the limits we identified in the previous chapter:

• Firstly, following Linked Data principles, each user is represented by their own URI and information about himself or herself (in this case, their social network) can be automatically delivered when dereferencing this URI with an RDF-aware client. That information is then provided in a machine-readable format, using one of the RDF serialisation that we previously mentioned. For instance http://apassant.net/alex iden-

Table 5. FOAF data generated from the Flickr exporter

```
<foaf:Person rdf:about="http://apassant.net/home/2007/12/flickrdf/peo-
ple/33669349@N00">
<foaf:name>Alexandre Passant</foaf:name>
<foaf:mbox_sha1sum>80248cbb1109104d97aae884138a6afcda688bd2</foaf:mbox_sha-
1sum>
<geonames:locatedIn rdf:resource="http://sws.geonames.org/3038213/"/>
<foaf:holdsAccount rdf:resource="http://apassant.net/home/2007/12/flickrdf/
user/33669349@N00"/>
<foaf:knows rdf:resource="http://apassant.net/home/2007/12/flickrdf/peo-
ple/86846122@N00"/>
</foaf:Person>
```

Table 6. Example of a basic SPARQL query to identify friends of a person using FOAF

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>.
SELECT ?knows
FROM <http://example.org/user>
WHERE {
  <http://example.org/user> foaf:knows ?knows
}
```

tifies one of the authors of this chapter, and that URI, when being dereferenced, redirects to http://apassant.net which provides information about himself, in RDFa. Hence, there is no need to use different API calls, with various parameters, to retrieve the social network of a particular user, and standard tools can be used.

• Secondly, this information is available using the same format (RDF) and semantics (using the FOAF ontology and the foaf:knows relationship for the network definition) whatever the original format is, and it can then be queried with any SPARQL engine, provide a unified query scheme on the top of social data from different services.

However, since exporters define their own URI for each person, there is a need to mention that two URIs refer to the same person. That can be done thanks to the owl:sameAs property, which specifies that two URIs identify the same resource, as described in (Bojars et al., 2008). For example, the statements shown in Table 7 declare that these different URIs (from different exporters for Social Web services) identify the same person.

Finally, while we will focus on Semantic Web technologies (i.e. we mainly consider RDF(S)/OWL solutions), it is also worth mentioning the microformats community – http://microformats.org/ - which follows a similar idea for adding structured information to the Web. While less powerful (i.e. less extensibility, no reasoning over microformatted data), they can be translated to RDF using GRDDL – Gleaning Resource Descriptions from Dialects of Languages (Connolly, 2007).

Social Media Contributions

One of the first things that comes to mind when considering a common model for representing social media contributions (from different services) in a uniform way is RSS (an acronym with various definitions including Really Simple Syndication and RDF Site Summary). It provides a simple, but widely-deployed format for representing the recent contributions from a social media system, e.g. posts on a weblog or the latest edits in a wiki. Among the different RSS variations, RSS 1.0 is based on RDF - http://web.resource.org/rss/1.0/ - and it has been extended using different modules, such as the Content Module - http://web.resource.

Table 7. Unifying identities using owl:sameAs

```
:me owl:sameAs flickr:33669349@N00 ;
  owl:sameAs twitter:terraces#me ;
  owl:sameAs facebook:foaf-607513040.rdf#me.
```

org/rss/1.0/modules/content/. Similar to RSS is the Atom Syndication Format - http://tools.ietf.org/rfc/rfc4287.txt - an XML format and recent IETF standard that is also commonly used for syndicating web feeds (e.g. from Blogger.com).

Generally, RSS and Atom feeds consist of a container (the website itself, a channel in RSS) and some items, both having a description, link, etc. as shown in the following picture. In terms of related formats, it is also worth mentioning NewsML – http://newsml.org/ - mainly used for exchanging information between news agencies, since 2000.

In spite of their large-scale deployment and their ease of use, several issues arise with RSS and Atom, especially if one wants to take advantage of it to analyse a given community:

- First, they are only syndication formats. Hence, they only expose the latest items produced in a given community; generally the last 10 or 20. Therefore, one cannot get a complete overview of the community unless they have crawled all RSS feeds since the creation the community, which generally does not happen as SNA is done *a posteriori*.
- Then, there is no fine-grained representation of the items that have been produced. For example, comparing an RSS feed from Wikipedia with one from Twitter: one is about wiki pages and edits, the other is about status updates, but nothing can

be used to differentiate these two kinds of contributions in the RSS feed. Hence, specificities of social media contributions (such as multi-author editing in wikis) are not taken into account in RSS.
- Finally, RSS does not expose the complete structure of a community such as versioning of items, user groups, etc. as it is mainly limits one to data about the item.

To overcome these limitations, various models have been deployed. For example, with regards to wikis, WIF – Wiki Interchange Format – and WAF – Wiki Archive Format – have been developed (Völkel and Oren, 2006) as common models to exchange and archive data between different wikis, as well as the WikiOnt vocabulary (Harth et al., 2005), with a more complete list of wiki-based models being available in (Orlandi and Passant, 2009). Other specific models include SAM (Franz and Staab, 2005) and NABU (Osterfeld et al., 2005) for instant messaging, as well as mle (Rehatschek and Hausenblas, 2007) and SWAML (Fernández et al., 2007a) for mailing list representation using Semantic Web technologies.

In addition to these specific models, and in order to provide interoperability in online communities between different types of applications, the SIOC project was created – http://sioc-project.org/. SIOC – Semantically-Interlinked Online Communities (Breslin et al., 2005) (Bojars et al., 2008) – provides a complete ecosystem, and
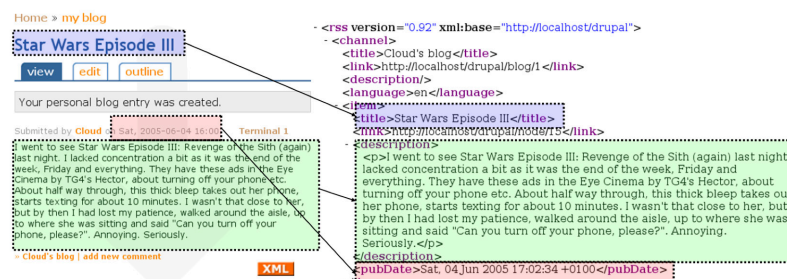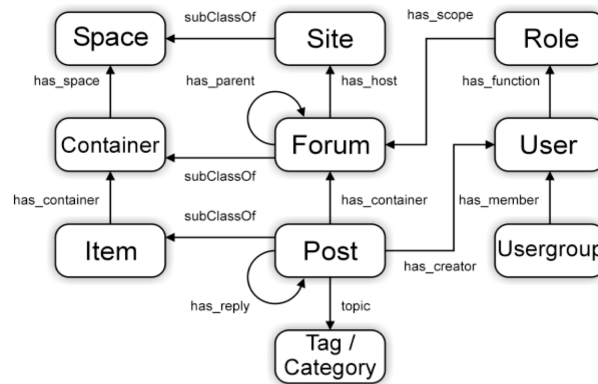
Figure 6. RSS syndication

Figure 7. The SIOC Core ontology

is comprised of an ontology and a set of related tools to enable semantics in online communities.

The SIOC ontology consists of the SIOC Core ontology (Figure 7) and five different modules. While created to be lightweight, the classes and properties contained in the SIOC Core ontology are powerful enough for representing many types of conversation that can occur in online communities. For example, it can be used to state that Alice has created a Post in a particular Forum (a sioc:Forum being a general discussion space, not specifically a bulletin board) and that Bob posted a reply Post in another Forum, as Table 8 shows.

As one can see in the previous example, modelling metadata for community-created content is carried out by combining SIOC with other lightweight ontologies such as Dublin Core (in this example, it is used to represent the creation date and title of a blog post).

The SIOC modules collection includes an argumentation module (for describing argumentative discussions), an access module (for access rights and permissions), a services module (for basic information about Social Web services), a types module (that provides fine-grained classes such as sioct:Wiki and sioct:WikiArticle to describe various content item types and containers) and a quotes module to represent quoting patterns on the Social Web and in systems such as bulletin boards and email discussions. There is also a module providing alignments with the SWAN (Ciccarese et al., 2008) ontology, in order to enable a complete

Table 8. Example of SIOC data for a distributed conversation

```
http://example.org/blog/post/1 rdf:type sioc:Post ;
  dct:created "2009-08-07T09:33:30Z" ;
  dct:title "Where's Wally?" ;
  sioc:has_creator http://example.org/alice ;
  sioc:has_container http://example.org/blog/1 ;
  sioc:has_reply http://foobar.org/blog/post/3.
http://foobar.org/blog/post/3 rdf:type sioc:Post ;
  dct:created "2009-08-07T10:43:55Z" ;
  dct:title "Here he is!" ;
  sioc:has_creator http://foobar.org/user/bob ;
  sioc:has_container http://foobar.org/blog/bob.
```

Figure 8. Interlinking Social Web data with SIOC
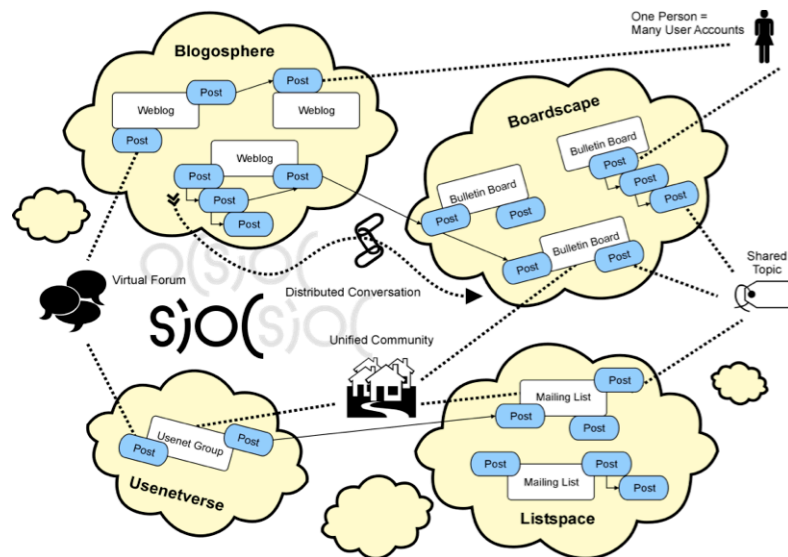
framework for modelling argumentaative discussions in scientific online communities, especially in the context of biomedical information (Passant et al., 2009c).

In addition to the ontology, various SIOC applications have been created: producers (for open-source applications such as Drupal or WordPress as well as exporters for major services like Flickr), crawlers and dedicated browsers, with some of them described in a future section of this chapter. All together, these applications and the ontology then enable interoperability on the Social Web at Web scale (Figure 8), providing better ways to gather, analyse and browse such data as the next section will emphasise.

Finally, another interesting feature of social media applications is tagging, i.e. using free-text keywords to provide user-driven indexing of Web 2.0 content. Various vocabularies have also been designed to enable interoperability between tagging systems, including the Tag Ontology, SCOT, MOAT and CommonTag. A more detailed description of these models for semantic tagging is available in (Passant, 2009) and in another chapter from this book (Erétéo et al., 2009b).

Integrating Social Networks and Social Data in Online Communities

As we explained in previous sections, different ontologies can be used to represent both social networks and social data. More specifically, FOAF and SIOC have strong ties together, and one physical person (with FOAF) can be linked to several user accounts (with SIOC), with each of them related to his or her various contributions within online communities. In this way, it can provide a complete "semantic social graph", overcoming the original silos of information from the original websites, as depicted in the following picture.

"Social network portability" is a related term that has been used to describe the ability to reuse one's own profile across various social networking sites and applications. The founder of the LiveJournal blogging community, Brad Fitzpatrick, wrote an article in August 2007 from a developer's point of view about forming a "decentralised social graph" - http://bradfitz.com/social-graph-problem/ - which discusses some ideas for social network portability and aggregating one's friends across sites. Dan Brickley, the co-creator of the FOAF vocabulary, wrote a re-

Figure 9. Social data portability using FOAF and SIOC
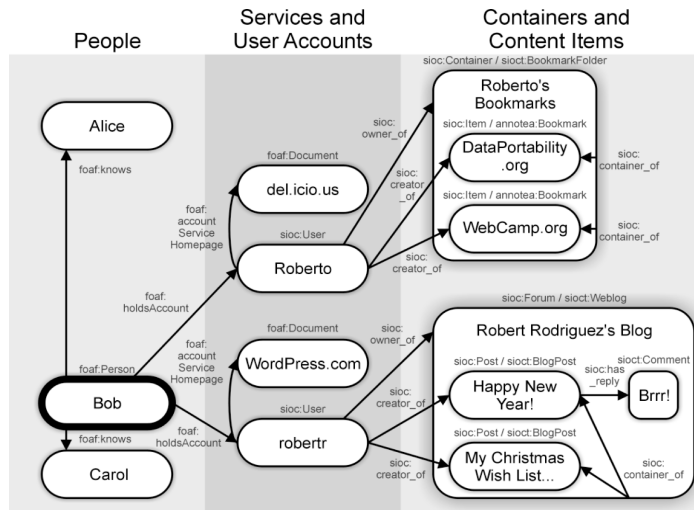
lated article entitled "The World is Now Closed" which talked about how SNSs should not define one's relationships in absolute terms and that even an aggregate social graph cannot be so clearly defined - http://danbri.org/words/2007/09/13/194. In parallel with this, a social network portability mailing list was established discussing many interesting topics including centralisation versus decentralisation, FOAF, XFN, hCard, OpenID, Bloom filters, ownership of your published content, categorizing friends and personas, the Open-FriendFormat, SNAP (Social Network Application Platform), aggregation and privacy, and XMPP (Extensible Messaging and Presence Protocol).

Use-Cases and Case Studies

In this section, we will describe some use-cases and case-studies of what can be achieved thanks to the previous models when analysing information from online communities. More advanced examples of social networking analysis and querying can be found in the chapter by (Erétéo et al., 2009b) in this book.

Querying Information from Online Communities

As we mentioned earlier, SPARQL allows us to run unified queries on any dataset of RDF-ised social data for analysis purposes. For example, one can retrieve a list of 10 people that claim to know Tim Berners-Lee, ordered by name, using the query in Table 9. This query, as with the others in this section, can be run using the SPARQL endpoint available at http://lod.openlinksw.com/sparql that hosts a replica of the Linked Data cloud, the result of the Linking Open Data project, providing billions of RDF statements from sources as diverse as DBpedia (Auer et al., 2008) or biomedical information systems, thanks to a set of good practices for publishing and interlinking data on the Web (Berners-Lee, 2006) (Bizer et al., 2007).

More complex queries can also be done, in the context of understanding communities, for example, to identify people interested in a topic and their relationships, which can be useful for topics such as expert finding on the Web. The query in Table 10 then retrieves a list of people that know someone interested in a topic containing the string semantic, and the figure below corresponds to a

Table 9. Retrieving people using SPARQL and FOAF

```
SELECT DISTINCT ?name
WHERE {
  ?s foaf:knows <http://www.w3.org/People/Berners-Lee/card#i> ;
  foaf:name ?name.
}
ORDER BY ASC(?name)
LIMIT 10
```

part of the answer. While provided using a simple tabular view, more advanced analysis can be done with such queries, such as ranking people by interest or mining subcommunities depending on the relationships that the topics share together, for instance considering that Semantic Web Services is a subtopic of the Semantic Web and then identifying related subcommunities from that relationship.

Another advanced area in which this kind of queries can be applied concerns the geolocation aspects of social networks. For example, the query in Table 11 identifies all the persons known by Sheila that are based near a location identified by its coordinates (using the foaf:based_near property). Such queries are used in the FOAFMap application (Passant, 2006) that provides a geolocation mashup of social networking information thanks to Google Map, as depicted in Figure 11.

Table 10. Identifying people around a particular topic

```
SELECT DISTINCT ?pname ?fname ?interest
WHERE {
  ?p foaf:name ?pname ;
     foaf:knows ?k.
  ?k foaf:name ?fname ;
     foaf:topic_interest ?interest.
  FILTER (REGEX(?interest, "semantic", "i"))
}
```

Figure 10. Results of the previous SPARQL query

pname	fname	interest
Quentin Reul	Sean Bechhofer	Knowledge Representation Languages for the Semantic Web
Quentin Reul	John Domingue	Semantic Web Services
Tom Heath	Michele Pasin	semantic web
Mischa Tuffield	Michele Pasin	semantic web
Tom Heath	Michele Pasin	semantic web
Afraz Jaffri	Tom Heath	semantic web
Michael G. Hausenblas	Tom Heath	semantic web

Figure 11. Geolocation of social networks using FOAFMap
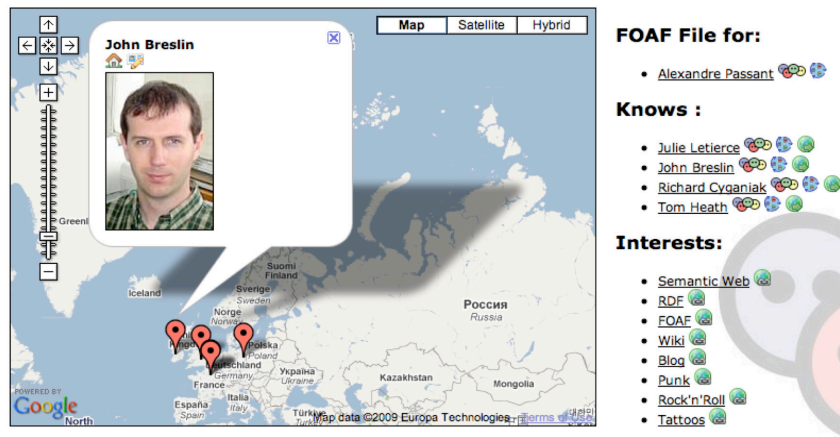
Moreover, using advanced SPARQL features (San Martin and Gutierrez, 2009) such as aggregates that are currently being standardised in the W3C SPARQL Working Group (http://www.w3.org/2009/sparql/wiki/Main_Page), Semantic Social Network Analysis features can be provided on the top of RDF data represented for instance using FOAF and SIOC.

Browsing Distributed Social Graphs

As we saw in the previous section, using FOAF to model social networking information provides a uniform representation layer for networks wherever or whatever the original application is. Moreover, one can centralise his or her various profiles using owl:sameAs statements, hence providing a single entry point to browse one's own data from several applications.

In order to take advantage of such uniform modelling, we developed a simple application called FOAFGear (based on the Flash GraphGear API), available at http://apassant.net/home/2008/01/foafgear that permits one to browse a distributed social graph in a coherent way. Thanks to the use of FOAF, and more generally of Semantic Web technologies, the script that produces the graph contains just 100 lines of code and uses only SPARQL to identify friends across networks, while original applications would have required various APIs to do something similar, leading developers to learn each API separately and to mash-up results together. The figure below shows the application being used to browse the networks for a particular

Table 11. Combining social networking and geolocation

```
SELECT DISTINCT ?o ?lat ?long
WHERE {
      <http://sw.deri.org/~sheila/foaf.rdf#me> foaf:knows ?o.
    ?o foaf:based_near ?place.
    ?place geo:lat ?lat.
    ?place geo:long ?long.
}
```

Figure 12. Browsing distributed social networks
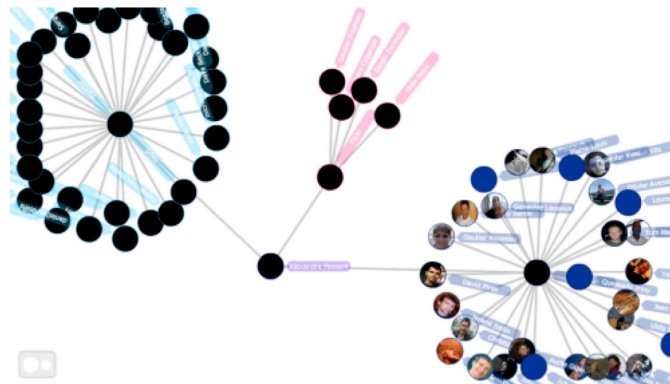
user from Twitter, Flickr and Facebook, centralised around the person's profile.

Browsing Topics and Contributions from a Community

The SIOC Explorer (Heitmann and Oren, 2007) allows users to browse and explore Social Web content from disparate online community sites in an integrated manner, as long as the browsed data has been represented using SIOC. The core of this application is BrowseRDF – a faceted navigation system for RDF data which is domain independent and provides a generic view of all RDF data associated with SIOC. The application aggregates Social Web content from various sources into a local RDF store and provides various ways to view the content and associated data.

When viewing posts from an individual forum or a group of forums, the user is presented with the list of posts in a reverse chronological order. Each post is summarised (see Figure 13) and can be expanded in order to read the full content. Clicking on the creator of a post shows all posts (including comments and replies) written by this person, across all forums; clicking on a topic shows

Figure 13. Faceted exploration of SIOC RDF data in the SIOC Explorer. The facets on the left allows to filter information by author in a given community

all posts tagged with this topic, again across all forums. In contrast to ordinary feed readers, such lateral browsing works across all different types of community forums that can be described in SIOC: clicking on a name (e.g. "Elias Torres") will not only show this user's blog posts, but also his microblog posts and IRC conversations.

A generic faceted navigation interface is offered on the left-hand side, displaying relevant facets that are not already shown as a part of the default browsing interface. Facets are built dynamically at view time and will show the properties and values derived from the actual data, while also displaying properties which may not be known at the system design time. Some facets (like the year) contain only "simple" values while complex facets, such as maker or topic, can be further expanded to see subsequent sub-facets (as shown on the bottom left of Figure 13).

Discovering Social Networks from SIOC Data

The Social SIOC Explorer (Bojars et al., 2007) builds upon the SIOC Explorer described earlier and allows users to see and explore social relations on the Social Web manifested via user-generated content. The motivation for this application comes from the observation that the social context related to content and its creator is as important as the content itself. An interesting concept in this regard is the notion of object-centred sociality as described by Jyri Engestrom - http://www.zengestrom.com/blog/2005/04/why_some_social.html - (Knorr-Cetina, 1997) - which provides evidence that people are connected via the content they create, co-annotate, and reply to. These collaborations uncover the implicit relations between people which are typically ignored by other Social Web applications such as feed readers.

What makes the Social SIOC Explorer different is the usage of FOAF in addition to SIOC data, the addition of user profile pages and the social context analysis component. Using this component, the application can extract two types of social contextual information from online community sites: the social neighbourhood (i.e. social network) of each site member and the indicators of their online reputation. As a result, this prototype allows users to browse and explore all this disparate information in an integrated manner.

In addition to views provided by the SIOC Explorer itself, this application provides views that let users explore person's social context and relations with other content creators. Figure 14 shows an example of a person's description, including information from his FOAF profile such as his picture, homepage and interests, and also his extracted social context. The screenshot shows a summary of these relations, with more details and links to actual people in the social neighbourhood available. It also shows the user's online reputation metrics. For example, we can see that this user has written 338 posts and made 115 comments, and knows 634 people through shared discussions – activity that could have taken place across several different sites, but has been unified thanks to the use of SIOC data.

A person's interests could be extracted using NLP algorithms from content that they publish online. Alternatively the use of Semantic Tagging ontologies could help in mining user interests, notably MOAT and CommonTag since these models focus on using URIs to represent tags, instead of simple strings, so that tagged content becomes meaningful and is linked to structured information.

Advanced Navigation Interfaces

Last year, the Digital Enterprise Research Institute (DERI) at the National University of Ireland, Galway and boards.ie, the largest Irish message board site, ran a competition in which entrants were asked to submit interesting creations based on discussion posts created on boards.ie between 1998 and 2008 (approximately 9 million documents), represented in RDF, mainly using the FOAF and

Figure 14. An example user profile with extracted social relations
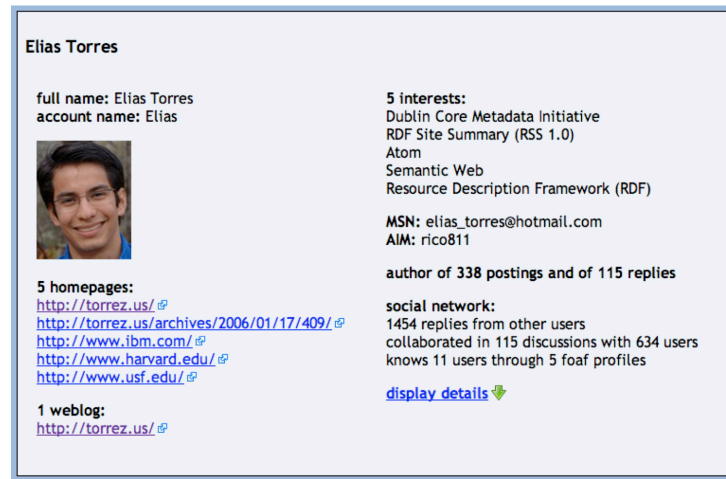
SIOC ontologies. Among the various submissions, the three winning ones provided advanced browsing and visualisation interfaces:

- The top winning submission was entitled "SIOC.ME" and illustrates how 3-D visualisations may be harnessed to not only provide an interactive means of presenting or browsing data but also to create useful data analysis tools, especially for manipulating the "semantic" (meaningful) data from online communities and social networking sites.

- In second place was a visualisation application called "boardsview", providing an interactive, real-time animation where one can watch the historical content from many discussion forums changing in real or compressed time (see Figure 15). Such application can be used to evaluate the wealth of a community and how such community evolve among time. Once again, a main advantage is that this service is "application agnostic" as it just uses SIOC data and does not consider the original applications that have been used to model such data.

- Finally, the third prize was awarded to the "Forum Activity Graph", a visualisation service showing the popularity of forums on boards.ie as represented by coloured rivers of information, which were drawn as SVG graphics and then rendered and displayed using Google Maps.

As we have already mentioned, as well as the applications themselves, an interesting aspect of these services is that since they rely on data based on known formats, the methods can be adapted to any application. Hence, the use of Semantic Web technologies can be used to provide adaptable browsing interfaces for online communities, as we demonstrated with the previous example of FOAFGear.

Challenges and Research Agenda for the Social Semantic Web

Before concluding the chapter, we would like to give an overview of some important challenges that still need to be considered in the context of social data on the (Semantic) Web. Indeed, while the technologies we have described so far solve various issues with regards to understanding

*Figure 15. Visualising evolving conversations over time*
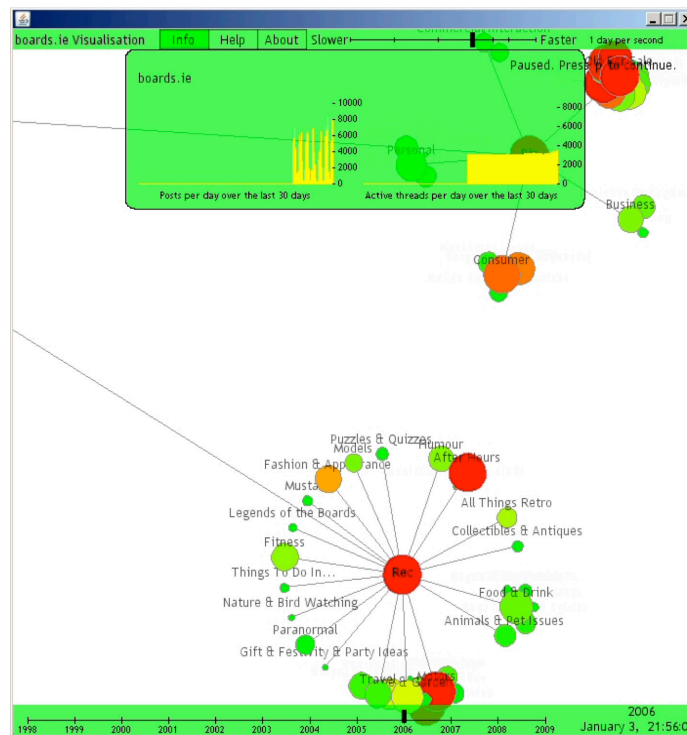
online communities, some other issues still have to be considered.

Trust and Privacy in Online Communities

When dealing with online community analysis, and especially with social data, whether it is personal information, social networking data or social media contributions, privacy is an important issue to consider. Policy languages can be used to allow or deny access to some social resources, as well as to provide restricted access to SPARQL endpoints (Abel et al., 2007). In addition, these policies can be finely defined thanks to the amount of RDF data now available on the Web, especially social data. For example, one could use information from his or her social network to allow access to their CV. This could be based on the people with whom one shares information on the LinkedIn business networking service (Passant et al., 2009b).

However, an interesting aspect of social networking and media-sharing websites is that most people use various websites because they want to deliberately fragment their online identity: uploading pictures of friends on MySpace, forming business contacts on LinkedIn, etc. (Figure 16). Under each persona, a user may reveal completely different facets of their personality. People may wish to share many of their identities with certain contacts, but retain more privacy when dealing with others. For example, many people are careful to keep their personal life distinct from their professional life. Yet, as we saw in the previous sections, inference and querying capabilities of the Semantic Web could enable better computation and reasoning over social data, sometimes breaking this voluntary fragmentation (especially when reasoning over Inverse Functional Properties). Hence, while the Semantic Web provides some solutions to privacy issues, it also introduces new ones that must be taken into account.

Figure 16. Identify fragmentation on the Social Web (from http://www.flickr.com/photos/fredcavazza/278973402/)
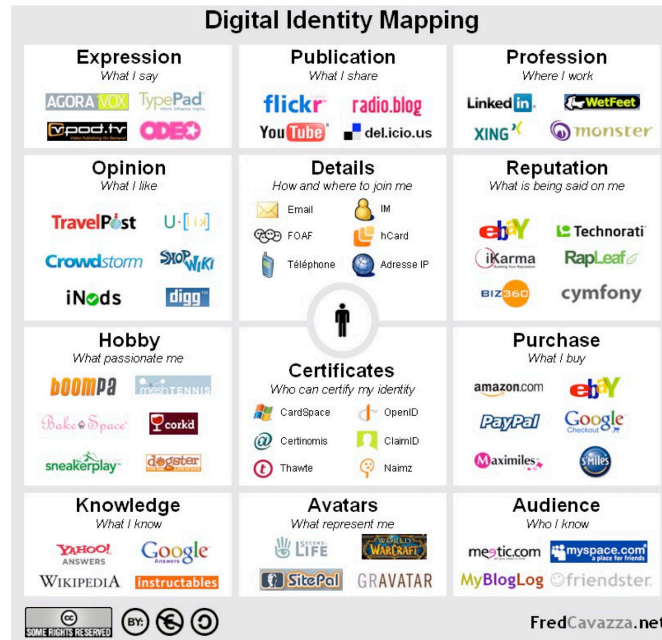
In addition, another related area is trust, i.e. ensuring that the information provided on the Web is trustworthy and comes from relevant sources. Models such as FOAF can have an important role to play, combined with authentication techniques such as FOAF+SSL (Story et al., 2009) or OpenID. In addition (Hartig, 2009) recently showed how trust can be enabled when querying information in SPARQL, a first step towards the top level of the Semantic Web layer cake.

Querying the Web as a Database: Scalability and Architectures

While the Semantic Web aims to provide advanced querying capabilities, considering the Web as a giant database, we must keep in mind that this is not a database per se: the Web is distributed, content is very dynamic, it uses various heterogeneous schemas, etc. Hence, various issues arise.

For example, how does one keep the distributed nature of the Web but also allow efficient query processing on this data? Soundness and completeness are new means for evaluating information retrieval on the Semantic Web, e.g. one can include approximate querying defined by time constraints as proposed by MaRvin (Oren et al., 2009). Another question is how does one enable querying over a set of heterogeneous schemas? For this, rule languages such as RIF - Rule Interchange Format (Boley et al., 2009) -, that enable rules and interoperability between these rules on the Web, have a role to play.

In addition to these considerations about the scalability of distributed computing for the Semantic Web, certain issues arise regarding application architectures. (Heitmann et al., 2009) showed, by analysing more than 90 Semantic Web applications, that most of them use a common set of components (Figure 17) such as data interfaces or crawlers. Yet, there is still a need to better understand the implications of the Semantic Web for developers.

Figure 17. Components of Semantic Web applications from (Heitmann et al., 2009)
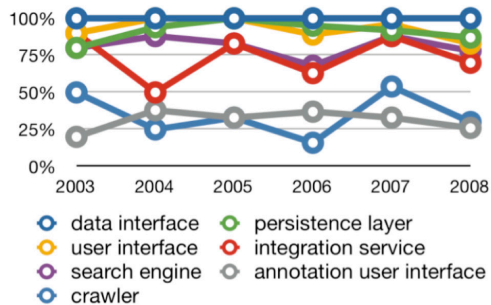
The Social Semantic Web and Web Science

We would like to conclude this chapter with some thoughts regarding online communities, the Social Semantic Web and the recent Web Science Research Initiative. The Web has permeated our daily lives. In a short space of time, much of our banking, commerce, and information dissemination needs have become dependent on the Internet. Therefore an understanding of the processes which play out on the Web is vital. At the moment, however, these properties are not fully understood. Understanding the Web is not as simple as understanding the technology behind the routers and servers which power it. The Web has evolved into a vast and complex network. Billions of interlinked documents are available online, and Web 2.0 technologies are enabling a significant fraction of the world's population to communicate with each other and create and share content. Simple interactions between people, documents and content result in emergent properties on a macro scale. This is a phenomenon similar to those observed in networks studied previously by physicists (e.g. of interacting particles) and biologists (e.g. of protein interactions). The growth of the Web and the network effects which have emerged present many opportunities in areas such as advertising, science, and healthcare, where more and more activities are taking place online. There

are also threats becoming apparent: espionage, cyber war and identity theft, for example. These opportunities and threats show that the Web is not an artefact solely of interest to computer scientists: there are also legal, economical and social issues which are essential to consider if we are to truly comprehend the Web.

Web Science is a new interdisciplinary branch of science (Berners-Lee et al., 2006) (Hendler et al., 2008) which aims to shed light on the phenomena which are emerging on the Web, and to engineer its future so that it evolves in a way which is beneficial for society. It was formally proposed in November 2006 when the foundation of the Web Science Research Initiative (WSRI) - http://webscience.org/ - was announced. The WSRI aims to bring together researchers from different disciplines to study the World Wide Web in order to gain understanding which can help guide its future use and design. The diverse fields from which Web Science draws include mathematics, physics, computer science, sociology, psychology, law, political science, economics, and ecology. The issues included within the scope of Web Science are broad-ranging and include technical issues (e.g. architecture, languages), social issues (privacy), legal issues (intellectual property) and many more.

There is a strong relationship between the Social Web and Web Science (Passant et al., 2009a). Online communities using Web 2.0 technologies have contributed vast amounts of data and complexity to the Web, and are of great interest to sociologists and researchers interested in understanding the motivations and behaviour of citizens of the Web. On the other hand, Web Science can also support online communities, by providing solutions to problems of trust and privacy, for example. The Social Web and Web Science have a lot to contribute to each other. Virtual communities are prime objects of study for scientists of the Web, and Web Science can be a valuable source of solutions for virtual communities.

The Semantic Web can similarly benefit from Web Science. Much effort has been put into building the foundations of the Semantic Web, and making semantically-described data available. However, there is still a need to understand the consequences for users of the Web which have resulted from these technologies and the increased exposure of data. Standards for the Web of Data are still in their formative stages, and perhaps Web Science can provide an insight into the best ways to proceed with these.

The future of the Social Semantic Web and the future of Web Science are very much interlinked. By studying the Social Semantic Web, where the interconnections between people and resources are richly described, we can gain deep insight into the online activities of communities. What we learn can contribute to developing solutions to the challenges identified in this chapter.

ACKNOWLEDGMENT

REFERENCES

Abel, F., De Coi, J., Henze, N., Koesling, A., Krause, D., & Olmedilla, D. (2007). Enabling Advanced and Context-Dependent Access Control in RDF Stores. *In Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007).*

Ankolekar, A., Krötzsch, M., Thanh Tran, D., & Vrandecic, D. (2008). The Two Cultures: Mashing up Web 2.0 and the Semantic Web. *Journal of Web Semantics*, *6*(1), 70–75.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2008). DBpedia: A Nucleus for a Web of Open Data. *In Proceedings of the 6th International Semantic Web Conference (ISWC 2007),* Busan, Korea.

Basca, C., Corlosquet, S., Cyganiak, R., Fernández, S., & Schandl, T. (2008). Neologism: Easy Vocabulary Publishing. In *Proceedings of the ESWC2008 Workshop on Scripting for the Semantic Web (SFSW2008).*

Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D., Patel-Schneider, P. F., & Stein, L. A. Dean, M., & Schreiber, G. (Eds.) (2004). OWL Web Ontology Language Reference. W3C Recommendation 10 February 2004. http://www.w3.org/TR/owl-ref/

Berners-Lee, T. (1989). *Information Management: A Proposal*. Technical report, CERN. http://www.w3.org/History/1989/proposal.html

Berners-Lee, T. (2005). Podcast at ISWC2005. Retrieved from http://esw.w3.org/topic/IswcPodcast

Berners-Lee, T. (2006). Linked Data. W3C Design Issue. Retrieved from http://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T., Fielding, R., Irvine, U., & Masinter, M. (2005). Uniform Resource Identifiers (URI): Generic Syntax. Request for comments: 3986, Internet Engineering Task Force. Retrieved from http://www.ietf.org/rfc/rfc3986.txt

Berners-Lee, T., Hall, W., Hendler, J., Shadbolt, N., & Weitzner, D. (2006). Creating a Science of the Web. *Science*, *313*(5788), 769. doi:10.1126/science.1126902

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, *284*(5), 34–43. doi:10.1038/scientificamerican0501-34

Bizer, C., Cyganiak, R., & Heath, T. (2007). How to publish Linked Data on the Web. Retrieved from http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/

Bojars, U., Breslin, J. G., Peristeras, V., Timmarello, G., & Decker, S. (2008). Interlinking the Social Web with Semantics. *IEEE Intelligent Systems*, *23*(3), 29–40. doi:10.1109/MIS.2008.50

Bojars, U., Heitmann, B., & Oren, E. (2007). A prototype to explore content and context on social community sites. In *Proceedings of the International Conference on Social Semantic Web (CSSW)* (pp. 47–58).

Bojars, U., Passant, A., Breslin, J. G., & Decker, S. (2008). Social Networks and Data Portability using Semantic Web technologies. In *Proceedings of the 2nd Workshop on Social Aspects of the Web (SAW 2008)*.

Boley, H., Hallmark, G., Kifer, M., Paschke, A., Polleres, A., & Reynolds, D. (2009). RIF Core Design. W3C Working Draft 3 July 2009. http://www.w3.org/TR/rif-core/

Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, *13*(1).

Breslin, J. G., Harth, A., Bojars, U., & Decker, S. (2005). Towards Semantically-Interlinked Online Communities. In *Proceedings of the 2nd European Semantic Web Conference (ESWC2005)*. Springer.

Breslin, J. G., Passant, A., & Decker, S. (2009). *The Social Semantic Web*. Springer. doi:10.1007/978-3-642-01172-6

Brickley, D., & Guha, R. V. (Eds.). (2004). RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation 10 February 2004, World Wide Web Consortium. Retrieved from http://www.w3.org/TR/rdf-schema/

Brickley, D., & Miller, M. (2004). FOAF Vocabulary Specification. Namespace Document 2 Sept 2004. http://xmlns.com/foaf/0.1/

Buffa, M., Gandon, F., Erétéo, G., Sander, P., & Faron, C. (2008). SweetWiki: A semantic wiki. *Journal of Web Semantics*, *6*(1), 84–97. doi:10.1016/j.websem.2007.11.003

Ciccarese, P., Wu, E., Wong, G., Ocana, M., Kinoshita, J., Ruttenberg, A., & Clark, T. (2008). The SWAN biomedical discourse ontology. *Journal of Biomedical Informatics*, *41*(5), 739–751. doi:10.1016/j.jbi.2008.04.010

Clark, K., Feigenbaum, L., & Torres, E. (Eds.). (2008). SPARQL Protocol for RDF. W3C Recommendation 15 January 2008, World Wide Web Consortium. Retrieved from http://www.w3.org/TR/rdf-sparql-protocol/

Connolly, D. (Ed.). (2007). Gleaning Resource Descriptions from Dialects of Languages (GRDDL). W3C Recommendation 11 September 2007, World Wide Web Consortium. Retrieved from http://www.w3.org/TR/grddl/

Erétéo, G., Buffa, M., Corby, O., Gandon, F., Leitzelman, M., Limpens, F., & Sander, P. (in press). Semantic Social Network Analysis, a Concrete Case. In *Handbook of Research on Methods and Techniques for Studying Virtual Communities*. IGI Global.

Erétéo, G., Gandon, F., Buffa, M., & Corby, O. (2009). Semantic Social Network Analysis. In *Proceedings of WebSci'09*. Athens, Greece: Society On-Line.

Fernández, S., Berrueta, D., & Labra, J. E. (2007). Mailing Lists Meet The Semantic Web. In *Proceedings of the BIS 2007 Workshop on Social Aspects of the Web (SAW2007)*, volume 245 de CEUR Workshop Proceedings. CEUR-WS.org

Franz, T., & Staab, S. (2005). SAM: Semantics Aware Instant Messaging for the Networked Semantic Desktop. In *Proceedings of the 1st Workshop on The Semantic Desktop, 4th International Semantic Web Conference*, volume 175 of CEUR Workshop Proceedings. CEUR-WS.org

Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, *5*, 199–220. doi:10.1006/knac.1993.1008

Gruhl, D., Nagarajan, N., Pieper, J., Robson, C., & Sheth, A. (2009). Context and Domain Knowledge Enhanced Entity Spotting in Informal Text. In Proceedings of the 8th International Semantic Web Conference (ISWC 2009), Chantilly, VA, USA. Springer.

Harth, A., Gassert, H., O'Murchu, I., Breslin, J. G., & Decker, S. (2005). WikiOnt: An Ontology for Describing and Exchanging Wikipedia Articles. In *Proceedings of Wikimania 2005 – The First International Wikimedia Conference*.

Hartig, O. (2009). Querying Trust in RDF Data with tSPARQL. In *Proceedings of the 6th European Semantic Web Conference (ESWC 2009)*.

Heath, T., & Motta, E. (2007). Revyu.com: A Reviewing and Rating Site for the Web of Data. In *Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007)*. Springer.

Heflin, J., & Hendler, J. A. (2000). Searching the Web with SHOE. In Artificial Intelligence for Web Search. *Papers from the AAAI Workshop. WS-00-01*. (pp. 35–40). AAAI Press.

Heitmann, B., Hayes, C., & Oren, E. (2009). Towards a reference architecture for Semantic Web applications. In *Proceedings of WebSci'09*. Athens, Greece: Society On-Line.

Heitmann, B., & Oren, E. (2007). Leveraging existing web frameworks for a SIOC Explorer to browse online social communities. In *Proceedings of the ESWC Workshop on Scripting for the Semantic Web* (pp. 52–61). June 2007.

Hendler, J. A., & Golbeck, J. (2008). Metcalfe's law, Web 2.0, and the Semantic Web. *Journal of Web Semantics*, *6*(1), 14–20. doi:10.1016/j.websem.2007.11.008

Hendler, J. A., Shadboldt, N., Hall, W., Berners-Lee, T., & Weitzner, D. (2008). Web Science: An interdisciplinary approach to understanding the World Wide Web. [cover story]. *Communications of the ACM*, (July): 2008.

Knorr-Cetina, K. D. (1997). Sociality with objects: Social relations in postsocial knowledge societies. *Theory, Culture & Society*, *14*(4), 1–30. doi:10.1177/026327697014004001

Mcafee, A. P. (2006). Enterprise 2.0: The Dawn of Emergent Collaboration. *MIT Sloan Management Review*, *47*(3), 21–28.

O'Reilly, T. (2005). O'Reilly Network: What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. Retrieved from http://www.oreillynet.com/lpt/a/6228

O'Reilly, T. (2006). O'Reilly Radar: Web 2.0 Compact Definition: Trying Again. Retrieved from http://radar.oreilly.com/archives/2006/12/web-20-compact.html

Oren, E., Kotoulas, S., Ganadiotis, G., Siebes, R., Ten Teije, A., & van Harmelen, F. (2009). A platform for large-scale analysis of Semantic Web data. In *Proceedings of WebSci'09: Society On-Line*. Athens, Greece: Marvin.

Orlandi, F., & Passant, A. (2009). Enabling cross-wikis integration by extending the SIOC ontology. In *Proceedings of the Fourth Workshop on Semantic Wikis (SemWiki2009)*.

Osterfeld, F., Kiesel, M., & Schwarz, S. (2005). Nabu – a semantic archive for xmpp instant messaging. In *Proceedings of the 1st Workshop on the Semantic Desktop at the 4th International Semantic Web Conference (ISWC 2005)*. CEUR-WS.org

Passant, A. (2006). FOAFMap: Web2.0 meets the Semantic Web. Challenge Submission in the 2nd Workshop on Scripting for the Semantic Web (SFSW2006), Budva, Montenegro.

Passant, A. (2009). *Technologies du Web Semantique pour l'Entreprise 2.0* (Semantic Web Technologies for Enterprise 2.0). PhD Thesis.

Passant, A., Ciccarese, P., Breslin, J., & Clark, T. (2009c) SWAN/SIOC: Aligning Scientific Discourse Representation and Social Semantics. In *Proceedings of the 1st International Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009) at the 8th International Semantic Web Conference (ISWC 2009),* Washington, DC, USA.

Passant, A., Hastrup, T., Bojars, U., & Breslin, J. (2008). Microblogging: A Semantic Web and Distributed Approach. In *Proceedings of the 4th Workshop on Scripting for the Semantic Web (SFSW2008), Workshop at ESWC 2008*.

Passant, A., Kärger, P., Hausenblas, M., Olmedilla, D., Polleres, A., & Decker, S. (2009b) Enabling Trust and Privacy on the Social Web. In *Proceedings of the W3C Workshop on the Future of Social Networking*.

Passant, A., Samwald, A., Breslin, J., & Decker, S. (2009a). Federating Distributed Social Data to Build an Interlinked Online Information Society. In *Proceedings of WebSci'09*. Athens, Greece: Society On-Line.

Prud'hommeaux, E., & Seaborne, A. (Eds.). (2008). SPARQL query language for RDF. W3C Recommendation 15 January 2008, World Wide Web Consortium. Retrieved from http://www.w3.org/TR/rdf-sparql-query/

Rehatschek, H., & Hausenblas, M. (2007). Enhancing the Exploration of Mailing List Archives Through Making Semantics Explicit. In *Semantic Web Challenge 2007 at the 6th International Semantic Web Conference (ISWC 2007)*.

San Martín, M., & Gutierrez, C. (2009) Representing, Querying and Transforming Social Networks with RDF/SPARQL. In *Proceedings of the 6th European Semantic Web Conference (ESWC 2009)*. Springer.

Story, H., Harbulot, B., Jacobi, I., & Jones, M. (2009). FOAF+TLS: RESTful Authentication for Distributed Social Networks. In *Proceedings of the first International Workshop on Trust and Privacy on the Social and Semantic Web (SPOT2009)*.

Suchanek, F. M., Sozio, M., & Weikum, G. (2009). SOFIE: a Self-Organizing Framework for Information Extraction. In *Proceedings of the 18th International Conference on World Wide Web (WWW 2009)*. ACM.

TAG - W3C Technical Architecture Group (2004). W3C Technical Architecture Group, Ian Jacobs and Norman Walsh, editors (2004). Architecture of the World Wide Web, Volume One. W3C Recommendation 15 December 2004, World Wide Web Consortium. http://www.w3.org/TR/webarch/.

Völkel, M., & Eyal Oren, E. (2006). Towards a Wiki Interchange Format (WIF) – Opening Semantic Wiki Content and Metadata. In *Proceedings of the First Workshop on Semantic Wikis - From Wiki to Semantics (SemWiki-2006), volume 206 of CEUR Workshop Proceedings*. CEUR-WS.org

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.