Semantic Web meets Web 2.0 (and vice versa): The Value of the Mundane for the Semantic Web

mc schraefel, Daniel Alexander Smith, Alistair Russell, Max Wilson

IAM Group, Electronics and Computer Science,
University of Southampton, Southampton UK
{ar05, das05r, mw105r, mc+iswc} @ ecs.soton.ac.uk

Abstract. Web 2.0, not the Semantic Web, has become the face of "the next generation Web" among the tech-literate set, and even among many in the various research communities involved in the Web. Perceptions in these communities of what the Semantic Web is (and who is involved in it) are often misinformed if not misguided. In this paper we identify opportunities for Semantic Web activities to connect with the Web 2.0 community; we explore why this connection is of significant benefit to both groups, and identify how these connections open valuable research opportunities "in the real" for the Semantic Web effort.

Keywords: Web 2.0, Semantic Web, tagging, trackback, blogging

Introduction

In a recent blog entry on the tech-savvy O'Reilly site, the author queries "Is Web 2.0 killing the Semantic Web?"[1] By way of background, Web 2.0 is elsewhere on O'Reilly described as a service delivered through the Web in a Web way, delivering a "rich user experience" [ref][2]. Exemplars offered are BitTorrent, Google and Blogs. The blog writer who asks about Web 2.0 killing the Semantic Web sets his argument as Web 2.0's instant superficial gratification of people vs the Semantic Web's deep, meaningful and lasting relationship with data. Web 2.0, he says, is indifferent to technology and just wants to "'give power to the people', quickly and efficiently" for superficial things like sharing files, opinions and photos. The author then goes on to state that the Semantic Web "is the polar opposite" being all about data and machine readability of that data and "sav[ing] lives." It is because of Web 2.0's flash, held against the SW's deep but unsexy worthiness that has caused resources that would be better spent on the SW to go to Web 2.0 projects:

> Users are seeing all this cool, flexible new Web 2.0 stuff, and it's making the SW look even more complex, rigid and unnecessary. Both technologies

[1] http://www.oreillynet.com/xml/blog/2005/10/is_web_20_killing_the_semantic.html. All following quotes which reference this entry can be found on this page

[2] O'Reilly has claim to defining the area since Tim O'Reilly initiated the first Web 2.0 conference to bring (web2con.com) together people to discuss and clarify the Web 2.0 concept.

> appear similar to the outside world - share and aggregate data – but Web 2.0 has a pretty interface, and is here and now. And thus the (finite) budgets of organisations are being spent on wikis and blogs, rather than RDF database converters.

In the following paper we suggest that the above characterization is partially right – that Web 2.0 may make the Semantic Web appear complex, rigid and unnecessary – but that the above is also wrong in suggesting that both technologies appear the same to the outside world. Our finding has been that the Semantic Web does *not* appear to the "outside world" at all, and that when it does, it is not seen in a positive light. Indeed the comments in reply to this blog post are indicative: while there is one brave mention of SPARQL, the majority of comments resonate with the following quotes, "Whenever I see articles about the Semantic Web, I keep getting reminded of the AI hype from the 1960s." Or, "It's about time some real-world pressures were exerted on the Semantic Web (SW) activities. SW was entirely conceived in an ivory tower, and it really shows in results." Or, most poignantly, the header of one post reads "Cannot kill what's not alive."

While these views may be put down to those of the unwashed, they are not restricted to it. In our own recent presentations within a variety of fields that should be deeply interested in the Semantic Web such as Information Retrieval and Human Factors, there is deep suspicion of the Semantic Web project. One prominent researcher in facetted browsing recently stated during a public course that the pedigree of the Semantic Web was "all those old AI researchers – good for you for working with them, but I'm keeping away." The comment was representative of other, equally respected leading researchers in IR, and is ironically reminiscent of how we have heard Semantic Web types speak about the "failed project of Hypertext" (you know who you are).

Some in the Semantic Web community may brush off such largely undeserved derision, suggesting that we are making steady gains in the imaginations of those who count. As Tim Berners-Lee suggested at his closing keynote to ISWC in 2003, when he gets asked by businesses about the Semantic Web now, it is no longer about "what it is" but about how it can be (4). Indeed, IBM's WebFountain (21) and Hewlett-Packard's JENA triplestore are examples of big industry seriously exploring the capability of the Semantic Web. In the past three years, the Semantic Web and Grid have also been used in many demonstrations for large-scale Science (22). The 2005 closing keynote of ISWC (the barometer of the state of the Semantic Web) reflected the importance of growing these demonstrators. Interestingly, however, that keynote also acknowledged two newer points: the value of *individual* involvement in activities like growing ontologies, and the importance of activities in interface research for the Semantic Web (13). Considering the absence of any presentation layer from the Semantic Web layer cake (13), this latter point is particularly compelling, especially for those of us who have been advocating the importance for the Semantic Web community to see that invisible layer as a core component of the Semantic Web activity (32)(27). Web 2.0's focus on the "rich user experience" as O'Reilly puts it, may well become the glue between innovative interfaces for the Semantic Web and technological combinations that blend Web 2.0 services with Semantic Web power.

In their consideration of the "Killer App and the Semantic Web", Alani *et al* suggest that breakthrough technologies are not always made of the leading, bleeding edge, but of effective re-packaging of more mature technologies (2). From our own work in what we call this everyday space of a *mundane* rather than Research or Enterprise Semantic Web, we see that there is considerable opportunity – low hanging fruit – for the Semantic Web to make a compelling breakthrough into the "real" and by this, advance its own research goals as well as improve understanding and interaction with them from related disciplines. In the following sections therefore we describe some of the areas where we and others have already been looking at Web 2.0/Semantic Web hybrids. Our goal is that by these examples we will demonstrate opportunities to exploit both existing Semantic Web tools and be a platform for innovative new Semantic Web research.

The Great Semantic Web + Web 2.0 Mashup

One of the *modes du jour* for seeing Web 2.0 applications in action is in the concept of the "mashup." The term mashup started in the audio domain, referring to artists remixing two (or more) recordings into a new entity. The founding example is the Grey album – a mashup of The Beatles' White Album and Jay-Z's black album (37). Where the music industry has had mixed reactions to mashups (suit was predictably brought against the Grey album. Meanwhile artists like David Bowie encourage mashups to be made of their music) the main reaction within the Web community has been to welcome mashups, usually understood as a combination of Web services. A grid at (31) lists mashup names and what services they combine. One of the most sited examples are those using Google Maps combined with an additional data source. A list can be found at (19). Google provides an API to their map service so that data sources can use the framework to present their data sources visualized in a geographical context. A compelling example is "http://mapsexoffenders.com/" which maps sex offenders in the US with their home addresses. Questions like "where are schools relative to density of offenders" suddenly become answerable at a glance. Suddenly this mashup takes on the grandeur of the "life saving" potential X dismissed from Web 2.0 consideration, attributing that only to the Semantic Web.

The simple act of presenting textual data in a spatial representation – one that is relevant and immediate to the people who use it – makes Web 2.0 services vital, meaningful. It is not hard to imagine how spreading just a little Semantic Web dust on Web 2.0 would take such already potent applications to the next level. Indeed, that has been our discovery with some semantic Web/Web 2.0 combinations we have been trying: that *a little bit of semantics goes a long way*.

In design work it is common to work from scenarios of use to explore how technological intervention may improve a process. In the following scenario, pulling back from the awful to the mundate, we look at how the combination of Web 2.0 and the Semantic Web can better support immediate but critical human activities such as finding food[3]. Let us say that Shumin and his colleagues are in a new town for a

[3] While mundane, food and shelter are at the top of Maslow's pyramid of human needs (11).

conference and they want to find a Japanese restaurant near their current location. Services such as Google Local (18) which search on keywords for businesses, like restaurants, can match location information from addresses to street maps to make it relatively easy to see whether restaurants that meet the criteria of Japanese cuisine meet the proximity. The query becomes more difficult, however, once Shumin asks for Japanese restaurants that serve vegetarian or vegan food and have some lactose free versions of their entrees and have a local beer that his boss might like. If he then wants to find a *good* restaurant (from Shumin's perspective) within this spec, this compound query means that we enter the realm of searching multiple sites over increasing time spent on what is actually a simple activity *if* the data is available.

The opportunity for the Semantic Web to connect with a Web 2.0 service like Google Local to support the above kind of compound query becomes readily apparent: restaurants publish RDF versions of their menus. These can be associated with food ontologies such that what foods are vegetarian (no meat) and which lactose free (no dairy) can be inferred without the restauranter having to specify these attributes explicitly. Value is added. Web 2.0 interfaces query these sources to enable exploration of restaurants which meet the necessary criteria. Such integrated services are reminiscent of both Haystack's flight browsing and booking system (1) and mSpace mobile's concurrent location-mapping of multiple categories of interest against locations of interest (39)[4].

With respect to recommendations, one may note that Google Local already matches reviews it finds to restaurants it maps. With the Semantic Web it is possible to use social network analysis as modeled by Jennifer Golbeck's FilmTrust work to go beyond recommendations to trust (16). Thus, rather than presenting undifferentiated recommendations, these can be filtered to those which meet, in this case, Shumin's profile for such recommendations. Indeed, this system can be adapted to help Shumin find new local beers his boss may like. According to Golbeck, the system would look at his preferences and his trust network to get that information. Of course, there is a security/privacy issue to be addressed. If the boss doesn't want someone to know her preferences, then there is potentially a problem. She may however be more willing to make her beer preferences known rather than allowing access to her trust network. If she has not made this preference explicit, a work around could be to create a sample trust network where the only person Shumin trusts in the network is his boss. Then Shumin could ask for recommendations for himself, and all of them would come from his boss's preferences.[5]

While the above scenario sounds a straight forward win for mashing up Semantic Web capabilities with Web 2.0 presentation services, it also implicitly foregrounds several oft-cited Semantic Web stumbling blocks: the lack of available RDF; the lack of ways for (small businesses and individuals in particular) to produce that RDF; the potential complexity of mechanisms to evaluate all that RDF. These problems are themselves opportunities for Semantic Web research, development and deployment into the mundane, but they require Semantic Web researchers either to blend their systems interests with consideration of usability, or they require Semantic Web

[4] A version of this kind of Semantic Web + Web 2.0 location explorer can be tried at http://montreal.mspace.fm

[5] Email conversation with Golbeck, May 14, 2006.

researchers to collaborate with HCI researchers. For instance, while few would want to produce multiple sets of Web site data – one in HTML, one in RDF – blogging has shown that people have no problem releasing multiple versions of their data if their service supports it. Blog software automatically outputs data in RSS 1, RSS 2, ATOM and HTML. The interfaces for blogging make the process of producing multiple data formats transparent. Likewise, effective Semantic Web production services that make it simple for proprietors to publish their menus (or their catalogues) as easily in RDF as they do their current Web sites would surely go a long way to making the range of rich details about their services easily publishable. Tools like Piggy Bank (23) already create templates for *post hoc* Web page scraping into RDF via manually authored site-specific templates. Turning those tools from post-Page creation to integrated with the site generation process itself would help not only produce more RDF, but would make the process of RDF production far more tractable. A well-learned lesson from HCI is to work *with* existing strategies and improve them. Thus, integration of RDF publishing solutions with popular Web creation tools and workflows like Dreamweaver would mean lower costs of entry. This is not to say developing such tools is trivial. If it were not a significant challenge, we would have such tools today. If we are not, however, prepared to spend the energy to create these tools to produce the outputs, can we complain that the outputs are so sparse?

With respect to filtering for trusted recommendations, effectively exploiting Semantic Web-based social networks the FilmTrust project has lead the way (16). The cost of entry however is not nothing: participants muse take time to rank a set of 50 films, from which the system inferences about other movies about which the participant is interested. Is such ranking necessary for all domains? How might inferences be applied across a few domains to generalize decisions about other domains? And if that is possible, what attributes are to be assessed? For instance, how match food preferences or environment attributes from which to develop particular recommendations for restaurants, hotels, historic cites, books, movies, friends and so on? Perhaps more particularly, once these methods for cross domain inferred trust are developed, we again come to the question of how expose the data needed to feed the algorithms to associate with great services? Again, there has been considerable work on post hoc semantic extraction of information from existing Web documentation, such as GATE (12), and Armadilo(11), to name two. By tuning some of this work from the the Research space to the mundane, we would have a compelling test case for the viability of Semantic Web technologies in the wild: either it will be, as we predict, readily easy to add value to such a Web2.0+ SW scenario, or it will not. Both cases offer interesting and compelling feedback to the Semantic Web community. We suspect the former will be the case, and reflect on three sites where we are investigating Semantic Web opportunities for Web 2.0 engagement.

A Little Bit of Semantics Can Go a Long Way

In this section we focus on how adding Semantic Web technologies to Web 2.0 applications can both extend the effectiveness of Web 2.0 applications and provide new spaces for Semantic Web research and evaluation. In particular we consider

technologies related to blogging, (24) one of the major Web 2.0 successes. Blogging enables people to readily create Web pages, but more than this, blogs make it easy to add metadata and to publish that metadata in ways that make blog entries readily linkable, via trackback (33), discoverable via tags and categorization, and subscribable via RSS (26). It is this gestalt of relatively lightweight technologies to support discovery, sharing and networking that has contributed to the success of blogs. It is not just the number of people creating blogs, but the variety of contexts in which they are found, from the casual social network to individual reports from either war regions [ref] or areas where even search services are censored (29) to blogs associated with formal, established news services like the BBC (8) that demonstrate the power of this mechanism for publication.

Despite the established effectiveness of these Web 2.0 technologies, it is easy for a Semantic Web researcher to see how Semantic Web technologies have the potential so improve the effectiveness of these applications while providing a platform for evaluating the roll of the Semantic Web in the mundane. In the following sections we consider Semantic Blogging, Semantic Tagging, Semantic Trackback, and an intermeidatry storage and query support service called RDF Garage.

Semantic Blogging

An approach to semantic blogging has been demonstrated focusing on bibliographic management (10). In the demonstrator, a person uses the Semantic Blog (based on the open-source blojsom blog software) to note down papers that they have read and papers they intend to read in the future. For instance, the person may see a relevant paper on the ACM Digital Library, downloads the PDF, and then creates a new blog entry, associating the bibtex metadata from the ACM portal to the post. The metadata about the paper, such as Author, Publications, etc. is then associated with the blog entry through importing a citation format like bibtex, or by using a web form and can be exported as RDF, using ontologies such as Dublin Core(35). The main value proposed from the service comes from being able to easily find out if other people have semantically blogged about the same paper, or similar papers. Users can search the metadata, known as “Query-by-Entry”, where values of a specific metadata fields can be searched again of other blog entries and use that metadata as the basis of a new entry. The semantic nature of the data means that exporting to alternate formats, such as those used by a citation tool, is relatively straightforward.

One of the obvious rationales of a semantic-based blogging system is not only discovery of people who have also blogged about a paper but for the inference possible via such semantics: inferring papers which may be about similar work but tagged differently to each other. Likewise, the semantics improves the potential as modeled by systems like FilmTrust to infer trustworthiness of findings or opinions in a given blog entry. In other words, an overall goal of a semantically enabled blog network would be to reduce what has been referred to as “information smog.” Systems such as FilmTrust (16) will work more effectively, and with less user interaction when there is more RDF. Right now, as we have seen, FilmTrust requires people to rate a set of movies to act as a benchmark for determining trust of reviews of new movies. Lightweight mechanisms to enable capture representations of opinions, reviews, or guidance integrated into blog software can help generate clear

data that can be reused by trust services, reducing the load on the individual to complete a separate and explicit task of benchmarking multiple domains.

Semantic Tagging

Tagging in Web2.0 is the ability to associate a keyword like "food", or set of concatenated keywords like "mycooking" with a resource such as a blog entry or a photograph. The benefit of doing this is that the interface then can allow browsing by tag, and can show the popularity of certain tags in a "tag cloud."[6] A tag cloud is a list of all the tags used in the current context, such as one person's blog, where the size of the individual tag's font size is proportional to its usage. For example, a blogger may frequently write about food and restaurants, so every blog entry about food is tagged with "food." The blogger, however, may want to specify when the food described is their own creation, so they use the tag "myfood". They can then quickly tell, by looking at the size of the tags in their tag cloud the proportion of posts about "food" compared to the posts about "myfood". While this knowledge may be useful, the ability for a computer to automatically utilize the tag "myfood" with other data is limited. However, if one were to describe the meaning of this tag, by referencing ontologies, the meaning, to both computer and humans alike becomes clearer. This not only means that ambiguity of terms is reduced, but that the users are in control of their own vocabulary.

We propose that a Semantic Tag is one where the tag itself is backed up by an RDF graph, functioning therefore as a useful shortcut to link knowledge to content. This is better than a keyword search as the Semantics are backed up with an Ontology, allowing the knowledge to not only be integrated easily with other Semantic data, but also the inference over this Ontology. For example, someone may tag something with "Beef Steak", and this infers, from the relevant food ontology, that another person is describing with the tags Beef, Meat, and Food. This means that someone who then searches on non-vegetarian cuisine will find this item. Semantic Tags should also be subsumptive, for example, the tag "mSpace" also infers the tags "Semantic Web Research" and "HCI Research" and the tag "Semantic Web Research" infers the tag "Computer Science Research". The RDF behind these tags depends on the usage, and therefore needs to be alterable by users. Likewise, users should be allowed to alter the meaning of a tag, by adding, removing and altering triples before they publish their use of the tag.

In related work (3), Beckett describes a practical way for tags to be formatted for semantics in current tag applications. This is a way to add semantics to tags without alteration of current applications, by specifically namespacing tags. For example, if a user was to specify that a file's type is MP3, they would tag it with "system:filetype:mp3". The problem of how to describe the underlying concepts and how to resolve these is left open. Alternately, our approach requires a small alteration to tagging interfaces, while also providing a method to resolve the knowledge where it has been previously described, and a way to allow the bootstrapping of knowledge in the event that it has yet to be properly described in an ontology, after the tagging

[6] An example of a Tag Cloud can be seen at http://en.wikipedia.org/wiki/Image:Flickr_Tag_Cloud.gif Accessed on: May 22, 2006.

has occurred. The parameters of these edits presents an interesting problem for shared meaning of tags since spam, vandalism and point-of-view differences give people reason to alter the meaning of a tag in a way that is not shared by all people who use that tag. We suggest a system where all possible permutations of the knowledge of a tag are all stored, and whenever a user specifies the tag, they are given the choice of which meaning to use, ordered potentially by how many times that meaning has been used. For example, the use of the tag TonyBlair could have many people tagging with the meaning of "Prime Minister" and "Politics", but there may also be a number of users that also wish this tag to infer "Liar". This option should not be suppressed[7], however it may be more convenient for the most utilized tag meanings to be at the top of a selection list, while of course enabling the preferred and perhaps less popular selected meanings of a tag to be immediately apparent when reading the meanings as they appear in a blog post which uses them.[8] The tag "TonyBlair" may well likely be vandalized (associating the tag with something insulting), however not all tag users would share this assessment, or want to include it in their use of the tag even if they did, so the alteration of the tag's meaning by lightweight interaction, such as deselecting this attribute of a tag can act as a vote against that meaning, and would help to reduce the occurrence of that meaning being associated with that tag. Wikipedia has faced a similar problem – It allows editing of its articles by anybody, and as such has to deal with problematic entries, Spam and difference of opinion on a regular basis. In spite of this, it reigns as a resounding success, based mainly on its policies for dispute resolution (38) that offers the ability for users to vote on whether article should be deleted, altered, or how a particular topic is described. With enough votes against a piece of knowledge behind the semantics of the tag, that vandalism will be removed.

Popback. There is also the potential for exploration of similar blog posts by ontology inference. Unpacking the TonyBlair tag both for its explicit and its inferred associations can be used by a service to present posts associated with those tags the Labour Party, UK politics, UK and US politics and so on. The facility to explore the granularity of the tag's meaning would be useful to people either with little knowledge about a particular domain and for whom such associated tags may help connect them with more familiar touchstones of meaning, or for people who simply wish to explore the associated context of a particular information artefact. We call this tag expansion and contraction *popback*.

Tag Forager. Semantic web services like Armadilo (11) and GATE (12) extract knowledge from documents. There is an opportunity to connect such services to help suggest potential tags to users from post-processing of their text. Mentioning Tony Blair would therefore suggest the above TonyBlair tag, and perhaps a "Labour Party." Richard M. Stallman might either suggest the tag Open Source, or the Open Source tag may simply be part of the tag Richard M. Stallman. These suggested tags could be

[7] Indeed, the possibility for so clearly and persistently exposing the polymorphic attributes of signifieds and signifiers (29) has philosophical, linguistic and social implications beyond the scope of this paper.

[8] There are further social and political implications in being able to construct a visualization services which would enable one to readily collect and peruse the writings of whomever used more "subversive" meanings of a tag

provided as a service for pre-publication and possible inclusion in blog posts, or for post-hoc processing of others' blog entries. This associated information can be ascertained from semantic clusters of tags and subjects that other bloggers have used.

Interestingly the plethora of Web 2.0 tags challenges the usual assumption that people hate to create metadata. While Google Co-Op pays users to add tags to data right now (17), there seems to be good evidence that people are willing to tag at least their own data: one of the most popular tagging sites is del.icio.us, which has 300,000 participants (20). Del.icio.us is a website that allows URLs to be submitted and tagged. Del.icio.us (14) allows for the subscription, via RSS of particular tags. A similar idea could be utilized here, with the user being able to subscribe to tags, but also particular classes, predicates and ontologies, so that even if new tags are used, the fact that the particular semantics described by that tag are interesting to the user means they should receive those updates too. In del.icio.us one may subscribe to the tag "computer", but will not receive items that have been tagged "PC". How to make subscription to predicates as palatable and sensible as subscription to tags is yet another Semantic Web UI challenge. The trust system could then potentially be employed to stem the information smog by only delivering trusted blogs.

Semantic Trackback

Trackback (33) is a system whereby one blog entry can reference another, and automatically have this reference placed on both blog entries, thereby creating a binary link. This is useful to bloggers, so they can find out easily who is blogging about their content. The web does not have a mechanism itself to permit the creation of binary links, and regular hyperlinks on the web are unary in nature.

The idea of a Semantic Trackback system is that instead of simply creating a binary link between two pages, as trackback currently enables, that link could carry knowledge and meaning with it. The meaning of a link at present, at least to a PageRank (7) system, is a vote for the other page, an assumption that publishing a link to somewhere not preferred by the posters would not happen. This behaviour has changed slightly in recent time, with the use of a rel=nofollow (30) to signify a link that should not be considered a positive vote of that page, mainly to curb the effect of "comment spam" on blogs. A Semantic Trackback system could apply more knowledge than a yea or nea about the linked artefact via an RDF graph to the link. For example, if somebody blogged about going to a Zoo, another blogger could create a Semantic Trackback link supported by an RDF graph that describes the fact that that person went with them, all using the same interaction as bloggers currently use to create ordinary trackbacks. This kind of acton-confirming information is useful to trust systems, as well as friend systems, since when the other user confirms the link, it asserts into the system that both parties agree they went to the Zoo on that day. Such confirmation brings the mundane into the Semantic Web trust layer.Semantic Trackback might also be used in a way similar to the "seeAlso" predicate is used in the RDF Syntax ontology (5) to create a link to more data about a particular resource.

RDF Garage – where everybody knows your ~~name~~ URI

In order to be able to reason over RDF whether to determine trust or association or similarity, the RDF needs to be available. Building ontologies is no mean feat, and ontologies have been referred to as the bottle neck of the Semantic Web (15).We propose RDF Garage[9] as a mechanism to aggregate sufficient RDF so that a sense of "good" predicates to use can be developed by consensus, and passed back to services like blogs for reuse. For instance, if a person wants to blog about the movie they just saw, the user selects "movie" from their category list or types in movie as a keyword or carries out some similar lightweight activity. The fact that people use tags, categories and keywords frequently as part of blog creation indicates that by leveraging this practice, we can extract data and add value without adding to a person's interaction load. Once this keyword has been selected or added, a service searches RDF Garage to see if any classes are being used in the RDF repository that (fuzzy) match their term. If they do, then that class can be used; where there is ambiguity, the service can ask the writer to *unselect* the meanings that do not apply. In the case where there is no pre-existing term match, the user is prompted to confirm their term, and to specify a meaning, to matched to specific ontologies. If ontological matches are found, these are associated, if they are not, they are saved, for later matching to an ontology that may exist in the future. When another blogger creates a movie review, then these classes can be rapidly rediscovered.

The key to this integration is use of Uniform Resource Identifiers (URIs). A URI is a short string that identifies resources: documents, images, services, people, places and other resources. One of the problems that the Semantic Web faces is that of co-reference. When two different documents or data sources refer to the same resource, are they using the same URI to refer to it? If they are, no problem, however if they are using different URIs, then the full benefit of the Semantic Web is lost, the data from these sources are kept disparate and never the twain shall meet. This problem is one that RDF Garage is attempting to solve, at least with respect to the application of resource tagging. By performing a lookup to a repository, the user can determine whether there exists a URI that describes the resource they are referencing, and thus the knowledge they are expressing about this resource is combined with the existing knowledge.

As RDF garage grows, it can be used to identify ontology gaps in domain areas of use. As such, better defined predicates and formally evolved ontologies can emerge "folksonomically" (36). RDF Garage will therefore provide a service to allow software to check for updated predicate and class mappings automatically. In terms of the philosophy "a little bit of semantics can go a long way," this approach to RDF Garage's re/use of RDF descritptions as available with mapping to ontologies when available means that people can gain the benefit of shared understandings of things now (my movie review concept can map to your film review concept), and in the future when formal ontologies are introduced to describe these concepts, their data will utilise them.

The use of RDF garage as both an implicit and explicit mechanism to support the growth of rdf and to refine ontologies speaks to Berners-Lee's call to individuals to help build ontologies. It supports this call by making that development transparent

[9] http://rdfgarage.mspace.net

for people who do not need to know anymore about the workings of RDF than they do about RSS: that RDF grows is inconsequential to a group who experiences value from richer discovery by others of their work, or easier discovery of others' work on a topic of interest. For instance, when the blogger has identified "movie" as a subject of their blog post, a service can likewise reach out to find movie reviews already published by trusted sources. As with the Semantic Blogging demonstrator's discovery of others who have read the same paper, discovery of other sources, in our case, trusted sources, provides the opportunity to research those reviews either before or while writing one's own piece, in order to reference or simply associate those reviews to one's own post. This approach of using the Semantic Web to discover resources while writing is explored in the WiCK project (9).

It is worth making explicit another advantage of RDF publication for trust determined across semantic networks. Such networks can cross heterogeneous Web 2.0 systems. The Internet is currently made up of many "blog islands", where connection over to the other items is hard. For example, a user on LiveJournal cannot add a person on MySpace to their friends list, however a system like FOAF [ref] handles "friending" behaviour very well. The RDF that the service described above creates when a person reviews and rates a movie takes advantage of FOAF data to trust two-way friend relationships: the fact that Shumin says Wendy is a friend of his becomes trustable when combined with data where Wendy says Shumin is a friend of hers. As such, Semantic Web services surmount system boundaries.

Discussion, Conclusions: Qui Bono?

It is clear that Web 2.0 application users and developers will benefit from less flaky data from successful application of Semantic Web approaches. We suggest that there may be a more profound effect from a more deliberate effort on the part of the Semantic Web community engaging with Web 2.0. As for the opening blogger's premise that Web 2.0 is encroaching on resources that would otherwise be spent on the more worthy and long term values the Semantic Web promises, one may also postulate that this is only a short term phenomenon as the science and research universe of the Semantic Web has high requirements for well-structured data that the ad hoc'ness of folksonomies may not be able to address fully. In other words, there is a case to be made that the Semantic Web is doing just fine, thank you, in places like e-Science and e-Research, and it is only a matter of time before that success includes (more of) e-Business as well. In a sense, however, such large-scale deployments are protected from the complete wilderness of being if not "web scale" then "web wild" applications. Data sources on the Semantic Grid (13) in Chemistry or Bioinformatics are if not perfectly curated then at least largely persistently identified and better managed than Web content in the raw. This is not to minimize the challenges for Semantic Web research through all aspects of its current layer cake which the work in e-Science and the Grid highlights. On the contrary. It does suggest, however, that there may be different and equally valuable and rewarding lessons to be learned by working in the messy wild of the mundae Web proper, where such lessons may translate back into the more formal constraints of the more structured exercises of the

current Semantic Web agenda in the large, were the Semantic Web to be more ecclesiastical and consider the value of the mundane in the construction of that agenda. Indeed, this year's ISWC conference has made a first gesture in this direction: the call for papers recognizes that "the Semantic Web requires new infrastructure on all levels;"[10] "Semantic Web 2.0" also appears as a category on its submission site where "catching research in the intersection [of Web 2.0 and the SW] was the purpose of this category."[11]

The effect of the Semantic Web research community in turning some of its attention from the grand challenge of domains like e-Science and Enterprise to the challenge of the mundane may be more providing a a new domain for the same agenda. For the Semantic Web purists, if there are such, then Web 2.0 involvement will provide a potential push to explore its current research agenda in terms of the existing Semantic Web layers of the cake. For instance, the Web 2.0 energy for mashups and communal publishing and tagging may provide sufficient enthusiasm to push through a number of the Semantic Web's bottlenecks (getting folksonomies to kickstart ontologies; leveraging a source like RDF Garage to identify knowledge gaps for further ontology refinement; using semantic tags to create concept networks which can also enable new research for inferring trust, and so on). Alternatively, our proposal for more explicit connexions to be established between Web 2.0 (and researchers in other fields drawn to Web 2.0) and Semantic Web communities may be seen as a way to fundamentally open the agenda, and allow different voices to be heard in the construction of the very formalisms of the SW (for example, the suggestion of what may be seen as heretically putting in new layers – like a presentation layer - in the layer cake), whether these voices come from HCI, IR, Web 2.0 or elsewhere. One effect may be that peer review of Semantic Web/Web2.0 papers may need to be ready to see not only more of the same Semantic Web approaches as applied in a new (mundane) domain, but rather to see also the reciprocal effects of that domain contributing to the construction of what Semantic Web research is, itself.

Our sense, however, is that there's enough cake to go around, and that the community is strong enough to be open to the possibilities that supporting engagement with other than its traditional research (and development) communities could only be a boon. In such an exchange, everyone benefits.

In this paper we have sketched out both why the Semantic Web community might want to consider more explicitly joining the Web 2.0 effort. There is value in the Web-scale potential of mundane Web activities to explore all the layers of the Semantic Web layer cake. We have suggested several opportunities where with a little extra semantics and lightweight UI's, we can leverage existing Web 2.0 interactions to increase the available pool of RDF, help people engage (implicitly) in ontology development, and most important perhaps, be able to point to compelling and directly relevant cases of the ways in which the Semantic Web is practically making a *meaningful* difference for people using the Web *today*.

[10] http://iswc2006.semanticweb.org/submissions/res_ac_track.htm seen from May 10-22, 2006.
[11] Email from Stefan Decker, program co-chair, providing category definition, May 21, 2006.

References

[1] Adar, E., Kargar, D., Stein, L.A.: Haystack: per-user information environments. In: CIKM '99: Proceedings of the eighth international conference on Information and knowledge management, New York, NY, USA, ACM Press (1999) 413-422
[2] Alani, H., Kalfoglou, Y., O'Hara, K., Shadbolt, N.: Towards a killer app for the semantic web. In: International Semantic Web Conference. (2005) 829-843
[3] Beckett, D.: Semantics Through the Tag. In: Proceedings on XTech 2006: "Building Web 2.0" (2006)
[4] Berner-Lee, T.: Putting the Web back in Semantic Web. In: Keynote Presentation, 4th International Semantic Web Conference, Available at: http://www.w3.org/2005/Talks/1110-iswc-tbl/ Accessed on May 22, 2006.
[5] Brickley, D., Guha, R.V.: RDF Vocabulary Description Language 1.0: RDF Schema. Available at: http://www.w3.org/TR/rdf-schema/ Accessed on: May 21, 2006.
[6] Brickley, D., Miller, L.: FOAF Vocabulary Specification. Available at: http://xmlns.com/foaf/0.1/ Accessed on May 22, 2006.
[7] Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. In: WWW7: Proceedings of the seventh international conference on World Wide Web 7, Amsterdam, The Netherlands, The Netherlands,
[8] British Broadcasting Corporation, BBC News Online. Available at: http://news.bbc.co.uk/ Accessed on May 22, 2006.
[9] Carr, L., Miles-Board, T., Woukeu, A., Wills, G., Hall, W.: The case for explicit knowledge in documents. In: DocEng '04: Proceedings of the 2004 ACM symposium on Document engineering, New York, NY, USA, ACM Press (2004) 90-98
[10] Cayzer, S.: Semantic blogging and decentralized knowledge management. Commun. ACM 47(12) (2004) 47-52
[11] Chapman, S., Dingli, A., Ciravegna, F.: Armadillo: harvesting information for the semantic web. In: SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, ACM Press (2004) 598-598
[12] Cunningham, H., Wilks, Y., Gaizauskas, R.J.: Gate: a general architecture for text engineering. In: Proceedings of the 16th conference on Computational linguistics, Morristown, NJ, USA, Association for Computational Linguistics (1996) 1057-1060
[13] De Roure, D. Jennings, N.R. Shadbolt, N.R. The Semantic Grid: Past, Present, and Future, Proceedings of the IEEE, Volume 93, Issue 3, March 2005, Pages 669-681.
[14] Del.icio.us. Available at: http://del.icio.us/ Accessed on May 21, 2006.
[15] Fensel, D. Ontologies: Their Glory and the new bottlenecks they create. OntoWeb Workshop.Semantic Web Project Proposal.,Vrije Universiteit Amsterdam (the Netherlands), 2000.
[16] Golbeck, J., Hendler, J.: Filmtrust: Movie recommendations using trust in web-based social networks. In: Proceedings of the IEEE Consumer Communications and Networking Conference, January 2006.
[17] Google Co-Op. Available At: http://google.com/coop Accessed on May 21, 2006.
[18] Google Local. Available At: http://local.google.com/ Accessed May 20, 2006
[19] Google Maps Mania. Listings of Google Maps-based mashups. Available at: http://googlemapsmania.blogspot.com/ Accessed May 20, 2006.
[20] Green, H. Yahoo Snaps Up Del.icio.us. In: BusinessWeek Online, Available At: http://www.businessweek.com/the_thread/blogspotting/archives/2005/12/yahoo_snaps_up.html Accessed on May 21, 2006.

[21] Gruhl, D., Chavet, L., Gibson, D., Meyer, J., Pattanayak, P., Tomkins, A., Zien, J.: How to build a webfountain: An architecture for very large-scale text analytics. IBM Syst. J. 43(1) (2004) 64-77
[22] Hey, T., Trefethen, A.E.: Cyberinfrastructure for e-science. Science 208(5723) (2005) 817-821
[23] Huynh, D., Mazzocchi, S., Karger, D.: Piggy Bank: Experience the Semantic Web Inside Your Web Browser, in Lecture Notes in Computer Science, Volume 3729, Oct 2005, Pages 413 - 430, DOI 10.1007/11574620_31, URL http://dx.doi.org/10.1007/11574620_31
[24] Kline, D., Burstein, D.: blog!. (2005) Squibnocket Partners LLC. ISBN 978-1-59315-141-1.
[25] Komatsu, E., Wolf, G., eds.: Saussure's first course of lectures on general linguistics. Elsevier (1907)
[26] Lerner, R.M.: At the forge: syndication with rss. Linux J. 2004(126) (2004) 8
[27] m. c. schraefel, Shadbolt, N.R., Gibbins, N., Harris, S., Glaser, H.: Cs aktive space: representing computer science in the semantic web. In: WWW '04: Proceedings of the 13th international conference on World Wide Web, New York, NY, USA, ACM Press (2004) 384-392
[28] Maslow, A. H. A Theory of Human Motivation. In: Psychological Review, 50, (1943) 370-396.
[29] McLaughlin, A.: Google In China, Google Blog. Available at: http://googleblog.blogspot.com/2006/01/google-in-china.html Accessed on May 22, 2006.
[30] Mishne, G., Carmel, D., Lempel, R.: Blocking Blog Spam with Language Model Disagreement. In: Proceedings of the 1st International Workshop on Adversarial Information Retrieval (2005).
[31] Programmable Web. Matrix of Mash-ups. Available at: http://www.programmableweb.com/matrix Accessed May 20, 2006.
[32] Quan, D.A., Karger, R.: How to make a semantic web browser. In: WWW '04: Proceedings of the 13th international conference on World Wide Web, New York, NY, USA, ACM Press (2004) 255-265
[33] Trott, B., Trott, M.: Trackback Technical Specification. Available At: http://www.sixapart.com/pronet/docs/trackback_spec Accessed May 21, 2006.
[34] Tuffield, M., Harris, S., Dupplaw, D. P., Chakravarthy, A., Brewster, C., Gibbins, N., O'Hara, K., Ciravegna, F., Sleeman, D., Wilks, Y. and Shadbolt, N. R. (2006) Image annotation with Photocopain. Submitted to First International Workshop on Semantic Web Annotations for Multimedia (SWAMM 2006), Edinburgh, United Kingdom.
[35] Weibel, S.: The Dublin Core: A Simple Conent Description Model for Electronic Resources. In: Bulletin of the American Society for Information Science (1997).
[36] Wikipedia, Folksonomies, Available At: http://en.wikipedia.org/wiki/Folksonomy Accessed on May 22, 2006.
[37] Wikipedia. Entry for "The Grey Album". Available at: http://en.wikipedia.org/wiki/The_Grey_Album Accessed May 20, 2006.
[38] Wikipedia. Wikipedia Dispute Resolution Policy. Available at: http://en.wikipedia.org/wiki/Wikipedia:Dispute_resolution Accessed on May 22, 2006.
[39] Wilson, M., Russell, A., m. c. schraefel, Smith, D.A.: mspace mobile: a ui gestalt to support on-the-go info-interaction. In: CHI '06: CHI '06 extended abstracts on Human factors in computing systems, New York, NY, USA, ACM Press (2006) 247-250