Emergent Capabilities for Collaborative Teams in the Evolving Web Environment

Dave Braines
Emerging Technology Services
IBM United Kingdom Ltd
Hursley Park, Winchester, UK
Email: dave_braines@uk.ibm.com

Paul Smart
School of Electronics & Computer Science
University of Southampton
Southampton, UK
Email: ps02v@ecs.soton.ac.uk

Jie Bao
Dept. of Computer Science
Rensselaer Polytechnic Institute
Troy, NY, USA
Email: baojie@cs.rpi.edu

***Abstract*— This paper[1] reports on our investigation of the latest advances for the Social Web, Web 2.0 and the Linked Data Web. These advances are discussed in terms of the latest capabilities that are available (or being made available) on the Web at the time of writing this paper. Such capabilities can be of significant benefit to teams, especially those comprised of multi-national, geographically-dispersed team members. The specific context of coalition members in a rapidly formed diverse military context such as disaster relief or humanitarian aid is considered, where close working between non-government organisations and non-military teams will help to achieve results as quickly and efficiently as possible. The heterogeneity one finds in such teams, coupled with a lack of dedicated private network infrastructure, poses a number of challenges for collaboration, and the current paper represents an attempt to assess whether nascent Web-based capabilities can support such teams in terms of both their collaborative activities and their access to (and sharing of) information resources.**

I. Introduction

Research sponsored by the International Technology Alliance[2] (ITA) is investigating the interface between humans and machines in terms of information representation and visualisation, and it is specifically looking at techniques which could be used to support shared understanding [11] between human users in a networked team environment. A particular focus of attention in the ITA programme are situations where the team members are drawn from a diverse multi-national coalition team and are operating in a potentially unstable network environment, for example, one provided by a MANET (Mobile Ad-hoc Network) infrastructure. A key objective of our work in the ITA is to understand the way in which these kind of network environments enhance, augment, and (perhaps) undermine cognitive processing at both the individual and collective levels. [10]. This paper discusses the latest advances in the Social capabilities of the World Wide Web (WWW), some of which could foster positive collaborative team interactions and potentially improve team cognition. Our ultimate vision is one where the underlying electronic networks play a role in shaping and augmenting the individual and collective cognitive capabilities of agents (see [9], in press).

[1]This is an abridged version of the earlier ITA technical report of the same name, see http://www.usukita.org/papers/5390/details.html

[2]More information available from: http://usukita.org/

In our research, we also consider the potential effects on cognitive performance that may be caused by an over-reliance on unreliable or poor-quality network resources. Within this paper we consider the WWW as the underlying transport and facilitation mechanism upon which various applications and communities of users operate to create higher-level social interactions. These higher-level social capabilities (sometimes referred to as constituting the "Social machine"[8], [6]) are enabled through the universal adoption of various standards and protocols upon which the WWW is built, and we therefore consider the WWW as the standardised "glue" which can be used to create and foster web-enabled social interactions.

The majority of this paper is intended as a review of current capabilities rather than an attempt to propose new ideas. The paper provides an overview of a number of Web-based technologies and capabilities in various subsections of the "State-of-the-Web" section (see Section II). Each of these subsections has a specific focus on one particular area of interest. These areas, when considered as a whole, provide the basis for much of the evolving ecosystem that is the current WWW, and it is these areas which we believe can be built upon in evolutionary stages to further foster social interactions and to support the construction of the future web of data (rather than the current web of documents). The potential military relevance for these technologies and capabilities is summarised in Section III.

II. State-of-the-Web

Our discussion of the current "state-of-the-web" is split into various discrete subsections, each offering commentary and typical examples of the capabilities and services that exist or are emerging in particular areas. There is a degree of overlap between the various areas that we have identified; however, we find it useful to discuss the areas in separate sections as a means of presenting and compartmentalising our thoughts. Areas in which we see particular overlap are the Social Web and Web 2.0. However, for the purposes of this paper, we chose to differentiate these areas on the basis of the Web 2.0 *solutions* provided by underlying technical capability or specific user behaviours (See Section II-B), versus the higher level Social Web *requirements* or capabilities which support social interaction or a particular socially-related attribute (See

Section II-A). It should be noted that this characterisation of requirements and solutions does not imply a rigid one-way flow of influence, but instead characterises a more symbiotic relationship where technical solutions can foster unanticipated new social behaviours in addition to the evolution of social behaviour placing new requirements on the underlying technology.

A. Social Web

The Social Web is the manifestation of collective user behaviour on top of the underlying WWW infrastructure. This takes many forms, and current examples tend to be disparate, or at best loosely linked, usually concentrated around particular centralised sites or services. At the time of writing this paper the most visible examples of the Social Web tend to be the particular sites that attract great volumes of social activity from large user bases, i.e. Facebook, Twitter, MySpace, Wikipedia, Flickr, YouTube etc. The key differentiating factor that separates the Social Web from the underlying technologies is that the Social Web is defined by the people whose collective actions and contributions constitute it, rather than the underlying computer infrastructure and technology which supports it. That is not to say that the underlying technical infrastructure is not important: it is critical, and small technological innovations can enable large cultural or social phenomena as a result. The advent of Web 2.0 technologies (See Section II-B) played a very large role in enabling the Social Web to emerge; however, there are examples such as early blogging, early wikis and particular sites, such as Amazon, which exhibit clear Social Web attributes prior to the advent of the current set of Web 2.0 technologies. Some of the key characteristics of the Social Web include:

1) Openness: Much of the infrastructure of the WWW and the Social Web is built around the notion of openness, both in terms of user behaviour and social etiquette, as well as in terms of software and standards. There are also ongoing efforts to extend the openness further, into domains where commercial dominance is currently present. These emerging standards aim to open various components which support the Social Web, with a desire to create an "Open Social Web" which is more generic, more pervasive, more interlinked, portable, and ultimately easier for end users to work with from a data perspective. Whilst good progress is being made on the definition and agreement of technical specifications, it is not yet clear whether these will be widely adopted, and if so how this adoption could be facilitated. There are clear commercial reasons why the popular Social Web organisations might wish to retain their current control (and even sometimes ownership) of the underlying data, but the commercial case for adopting open distributed standards in order to achieve the next iteration of the Social Web is less clear, even though the ultimate benefits to the community at large are potentially significant. For emerging organisations in this environment a good value proposition to the potential future user base is one of portability, convenience and centrality (based on open standards), which should foster increased transition between social networks and allow social network services to compete more on real benefits to their users rather than simply trying to achieve critical mass and then hold on to it. Some potentially valuable new initiatives such as The Mine! Project[3], which is described as "an Open Source project for online data and relationship logistics" are starting to investigate issues and potential approaches in this area. It is also critical that the user interaction with the underlying data represented in these open standards is at least as rewarding as the current incumbent applications, otherwise the user community will not be motivated to move towards this potentially more valuable, open, and portable future.

2) Identity: There are also a number of recent proposals around the topic of identity, and, in particular, managed identity, which is delivered through distributed identity providers. Popular examples of identity management solutions which are gaining widespread adoption are discussed in Section II-B4. Many of these are also examples of Web 2.0 technologies, which are emerging to support identity management functions on the Social Web. One aspiration of distributed identity management technologies is that they give the individual users back their control of the data that they create. Rather than being forced to conform to the default decisions of the applications individual users are using, these capabilities can allow users to grant limited access to a specific subset of their data for a temporary period rather than being forced to grant access to all of their data in order for any one part to be accessed. Solutions in this space will also help to reduce or remove the "Password anti-pattern"[4] which has become an all-too-common occurrence in the Social Web, along with the opportunities for phishing and fake sites whose intention is simply to farm user login and password details for nefarious purposes. Also, an open, distributed identity management standard will help to reduce the large number of separate userids and passwords which each user has to manage for their portfolio of online presences, in addition to providing a potential basis from which a centralised reputation solution can be built.

3) Intellectual property: Another key factor in the Social Web is that of intellectual property ownership and rights. In an environment where material can be quickly re-purposed, mashed-up and generally used (and maybe abused) it is important that the members of the community understand any ownership issues that relate to the content they are using. The traditional world works with patents, trademarks and copyrights and the specific legal interpretations of these can vary by country which makes for an unclear and potentially unresponsive environment within which to operate. Such an environment does not lend itself to the kinds of behaviour (or speed) that is desired by the Social Web community. Also, since much of the original material in the Social Web is created by individual users, and these users may wish to actively promote reuse of their material, there is a need for a more

[3]See http://themineproject.org/

[4]Providing the userid and password for one site to another in order to allow the second site to access secure resources in the first.

consumable and user friendly rights system. Creative Commons[5] is a widely adopted recent standard in this area, which defines various levels of usage and starts where traditional copyright ends. Using a creative commons license, users can define for each of their electronic assets the specific terms under which that asset can be reused, and it is common for users to simply state a requirement for attribution of their work. Prominent Social Web sites such as Flickr have adopted the creative commons license at the core of their offering, thereby allowing users to search for electronic assets (in this case typically photographs) not only by their distinguishing features, but also by their creative commons license type. The use of this more granular approach to licensing, and the alignment of this with the desires of the Social Web community, helps to reduce the friction associated with the re-mixing or interlinking of assets, and the community itself often helps with the enforcement and policing of violations.

4) Social graph: One of the core capabilities of a traditional social networking website is the social graph: the identification of relationships between people (usually users registered to that website) and sometimes between people and other electronic entities (such as photos, meetings, events, groups etc). Sites such as Facebook, Twitter and MySpace implement specific user interfaces to support the creation and management of these relationships separately within each of the different websites. An early open standard proposal for the representation of this data is known as FOAF (Friend Of A Friend), although this has gained only limited adoption in certain sections of the academic community. FOAF is based on an RDF representation (See Section II-C for more details on RDF) and could form a key dataset for the Linked Data Web and Semantic Web communities. Other more recent standards have been proposed, such as OpenSocial[6] and Facebook Connect[7], although it is worth noting that FOAF differs from proposals such as these in that it proposes a fundamental representation of the social networking data itself, rather than simply providing a mechanism for sharing data between existing social networking sites.

5) Collaboration: Perhaps the most important capability for members of the Social Web is that of collaboration: the ability to create, edit, link to and share material between users in an efficient and useful manner (this is something that is clearly an extension of the initial publisher-centric and static incarnation of the WWW). The collaborative environment that has been fostered by the Social Web has created new interaction models between individuals and between organisations and individuals, for example, the new roles of "Prosumer" (Producer/Consumer) as introduced in [12]. Collaboration takes many forms, some of which are more obvious than others. The explicit collaborative construction of information resources such as Wikipedia represents a clear concerted effort from many users over an extended period, the end result of which is a single high-quality reference resource. However, there is also the collective creation of material and collaborative tagging of pictures (in Flickr) or videos (in YouTube), the user-generated comments on (e.g.) blogs (or any other type of content), the Collective Intelligence gained from user reviews or rating of products, or even the implicit collective buying (e.g. Amazon recommendation) and music listening (e.g. iTunes Genius) habits of users. These and other examples are equally valuable but perhaps less obvious forms of collaborative behaviour. In addition to these, there are more specific task oriented examples such as sharing calendars, to-do lists, communication via instant messaging, IRC (Internet Relay Chat), discussion forum, or the accomplishment of complex goals in online virtual worlds and game environments. This wide variety of collaborative activities, and the artifacts that result from this collaboration, has prompted the need for aggregation, syndication and filtering services, which attempt to provide centralised perspectives (usually focused around a particular user) on multiple sources of online material. For example FriendFeed provides a capability for aggregating conversations across multiple Social Web applications, enabling contextual responses to be made, and emerging developments such as Mozilla Raindrop[8] focus on the subtly different task of ranking and fragmenting this stream of disparate information into separate contextual pieces.

6) Social search: The Social Web offers great potential in terms of personalised (and therefore hopefully more relevant) search results. The main search service providers are looking at techniques concerning the integration of social network information into search results, for example, by prioritising results which have been commented on, linked to, or created by people close to you within your social network. Whilst this capability does not yet exist in any of the main stream search services, it is a tantalising prospect[9], and it could be a good demonstration of the potential benefits of semantic enrichment for online data. It should also be noted that the rapid dominance achieved by the Google search service has been heavily attributed to the success of the PageRank algorithm in terms of returning results which users found to be significantly more relevant than competing offerings. PageRank uses the hyperlinks created by human users between documents on the WWW to better rank the results, and it was one of the first examples of an algorithm that tapped into the existing social structure of the WWW, rather than simply relying on the textual content of each page.

B. Web 2.0

Web 2.0 is a widely used term and means different things to different people. Web 2.0 does not mean a distinct second version of the WWW, but is instead a more collaborative and fundamentally user-centric evolution of the current WWW environment: one within which social interaction, publication,

[5]See http://creativecommons.org/

[6]See http://code.google.com/apis/opensocial/

[7]http://developers.facebook.com/connect.php

[8]See http://labs.mozilla.com/raindrop/

[9]e.g. a very recent announcement at: http://googleblog.blogspot.com/2009/10/introducing-google-social-search-i.html

co-creation, community building and sharing of information become easier and gain greater focus. Sometimes Web 2.0 is referred to as the Social Web, or the Participatory Web, but for the purposes of this paper we define Web 2.0 in terms of the infrastructural components which specifically facilitate social interactions in a web environment[10]. In this Web 2.0 subsection we look at various technologies and techniques on which such Social Web capabilities can be built.

Web 2.0 technologies are fundamentally built around the needs of the Social Web (Section II-A), such as the desire to enable users to collaborate online, through direct interaction with other users, or in more passive activities from which subsequent social functions can be derived, such as tagging, recommendation, filtering or ranking. Another important capability supported by Web 2.0 technologies is the collaborative construction of information resources of various formats (often referred to as user-generated content), such as wiki content, blog and micro-blog entries, multi-media repositories (such as Flickr for photographs and YouTube for videos), user-generated "folksonomies" of tag data, mashups and visualisations. Examples of some of these technologies and the capabilities that they enable are listed below:

1) Collaborative tagging: The act of collaborative tagging was first made popular by the social bookmarking service del.icio.us and the photo sharing site Flickr. The result of a community of users collaboratively producing and tagging content is known as a "Folksonomy", and the collective agreement around particular tags of interest and their implied meaning within a community can create a valuable system for content classification. Folksonomies are created collaboratively in a bottom-up manner, rather than the more traditional top-down taxonomic approach, which is often centrally imposed. Folksonomies can evolve as circumstances change, and they may be more resilient to dynamic situations and unbounded domains compared to centrally managed taxonomies or ontologies. There can be issues with ambiguity, agreement and completeness of coverage with folksonomies, and they only truly flourish in environments that have gained a certain degree of critical mass amongst their user base. Collaborative tagging and the resulting folksonomies can be used to create, amongst other things, "tag clouds", which show the dominance of particular tags within a given subsection of material. Furthermore the user tagging activities can be used to identify things such as trending topics when the subsection is chosen to be recent material.

Tagging can be applied to any persistent online electronic entity, with the community collectively deciding which tags to accept and foster and which tags to overlook. Common entities which are tagged include photos, videos, blog posts, micro-blog entries, news articles and internet bookmarks, with popular communities emerging around the main applications in each of these areas of interest. Open questions remain regarding the mapping of different folksonomies or sets of tags across multiple sites, with a key issue being the retention of the simplicity of the folksonomic tagging approach whilst achieving a disambiguated relationship to other tagging systems[11].

2) Blogging and micro-blogging: A "blog" is a location on the WWW where a person, or group of people, create entries (blog posts) which offer some commentary on a topic of interest. The term blog is an abbreviation of Web-log. Blogs first became popular in the late 1990s and they mark the beginning of the Social Web. The desire for people to blog has led to the blogosphere, which is a rich ecosystem within the wider WWW. These include a variety of tools and software components to make the job of blogging easier (e.g. Wordpress or LiveJournal), dedicated blog search engines (such as Technorati), and various technology components to support cross referencing and commenting across blog posts within the blogosphere (such as Trackback). Whilst the original purpose of blogging was to create a largely textual resource, the act of blogging has also been extended to most other multi-media forms, such as photography, audio (e.g. podcasting) and video. The act of blogging has led to the rise of new social capabilities within the online community, such as "Citizen Journalism" and "Amateur Academia". These have the potential for widespread visibility of their material; however they are sometimes criticised for their lack of quality, ethics or credibility.

Micro-blogging is a more recent evolution, consisting of very short status updates rather than detailed blog posts. Micro-blogging has been made popular by sites such as Facebook, Twitter and others, and the act of micro-blogging is driving a new phenomenon that is known as the "real-time web". Whilst micro-blogging typically involves users saying what they are currently doing or thinking, this activity provides a high-volume near real-time information stream that can be a source of breaking news, user sentiment, or rumour, and the use of existing WWW components such as tags and hyperlinks provides a basis for micro-blogs to aggregate and link to other resources. At the time of writing, there is significant interest from the major search engine providers in harnessing this "firehose" of real-time activity[12] and making it available in search results, either as distinct real-time results in their own right, or as an additional component of the result ranking algorithm.

3) Wikis: In the mid 1990s wiki systems started to gain popularity, with perhaps the most widely known extant example of Wikipedia gaining widespread visibility around 2004. The premise of a wiki is that a community of users should have easy access to create and edit interlinked web pages using a WYSIWYG (What You See Is What You Get) or simple text-based interface, with little or no review and approval cycle (this reduces the friction associated with the context creation process). Additional key capabilities like a fully featured edit

[10]Although it should be noted that this paper does not focus in detail on the lowest level infrastructure components such as DHTML, Ajax and JSON.

[11]See http://tagcommons.org

[12]Both Google and Microsoft Bing have recently announced agreements with Twitter to access this real-time stream of information. See http://mashable.com/2009/10/21/google-twitter-search-deal/

history function have also evolved in response to the rapid adoption and uptake. These features have emerged partly to support identification in cases of vandalism or inappropriate changes.

Whilst there are a few high profile wiki systems on the WWW which focus on general topics, there are also many more wikis that cater for the more select interests of specific communities of users. There are a number of wiki software packages which can be used to quickly create a web-based environment within which a wiki can be hosted. A popular example is the Open Source MediaWiki[13] on which Wikipedia is based. In addition to the basic capability of allowing users to create and edit pages in a simple, user-friendly way, wikis also provide additional key capabilities which relate to the organisation and structuring of content, helping to ensure that the material entered in a wiki can be appropriately structured to support the subsequent reading of that information by future visitors.

One recent wiki-related advance of particular interest is exemplified by semantic wikis which aim to make explicit some of the semantic relationships between wiki content. Semantic wikis are important because they support the development of a collaborative, user-focused environment within which semantically rich user-generated content can be authored. A popular Semantic Wiki software package at the time of writing is Semantic MediaWiki[14], which is made available as an Open Source extension to the popular MediaWiki software described earlier.

The main difference between a normal wiki and a semantic wiki is that the latter has an underlying model of knowledge authored in the wiki, and it provides a mechanism for specifying distinct types for the hyperlinks that are expressed between articles in the wiki or located elsewhere on the WWW. Our research in the ITA has looked at various aspects of Semantic Wiki capabilities, specifically in terms of improving expressivity [3], providing capabilities for rule modelling [1], and providing a Controlled Natural Language (CNL) interface capability [2] in order to improve user interaction and understandability of the underlying semantic information.

A key usability consideration for semantic wikis is the hybrid approach, which allows for the entry of both unstructured textual data and structured factual information. This enables the users of the semantic wiki to define their knowledge in a form which is appropriate to their representational needs and potentially update or augment this unstructured information with increased structure over time. The semantic representation of the data in the wiki enables structured queries and exports to be performed which cannot be achieved with the normal unstructured data in a traditional wiki, and it supports a move towards a more data-centric vision of the online world (see Section II-C).

4) Identity management systems: OpenId[15] is a widely adopted open standard in the identity provider environment. It promotes a decentralised identity provider environment in which the user chooses their identity provider. Any organisation or website can become an identity provider, and no single organisation or entity owns the OpenId service. Users can login with a single userid to any website or service which supports the OpenId standard, and that website or service does not see their password details but instead receives a token from the corresponding identity provider to confirm that the user is authenticated. In simple terms, OpenId allows a user to login to multiple websites with the same userid and password. Related to OpenId is OAuth[16], which is also an emerging open web standard in the general domain of web identity. OAuth is a mechanism that allows different websites or services (which require authentication) to work together. For example, OAuth can be used to enable one service to access the resources of another service once the user is authenticated. This provides a powerful mechanism which avoids the need for websites to access the userid and password details of others. More importantly, this type of approach can increase the confidence of the user community, which should reduce the friction associated with the integration of related online services.

5) Content management systems: Whilst there is some overlap between blogs, wikis, and the more traditional Content Management Systems (CMS). CMSs are still distinguished as a source of structured information publishing within the WWW environment, albeit with wikis and blogs being particular specialisations which tend towards more fluid and unstructured information. CMS technology was very much used to drive the initial static content on the WWW, and, in many cases, the traditional CMS model has remained somewhat rigid and inflexible when compared to the more interactive advances of other components of the Social Web. The historic role of a CMS is to publish pages or product data which is largely static and intended for user consumption rather than user iteration. There are, however, some notable exceptions, and one in particular will be briefly covered here as it is of particular importance and relevance to the Social Web.

Drupal[17] is a popular Open Source content management system which is comprised of a core set of CMS capabilities and a module-based set of extensions, each of which provides specific capabilities. Both the core Drupal code base and the modules are created by the community and are made available as Open Source components. Drupal is used to power a number of high profile web sites such as the US Whitehouse[18] and Yahoo! Research[19], and Drupal is also used as the web based collaboration system[20] for the ITA research community. Drupal already supports a number of key Social Web functions (such as blogs, wikis, tagging, groups, integration with OpenId etc), and the planned version 7 release of Drupal is anticipated

[13]See http://www.mediawiki.org/
[14]See http://semantic-mediawiki.org/
[15]See http://openid.net/foundation/
[16]See http://oauth.net/
[17]See http://drupal.org
[18]See http://www.whitehouse.gov/
[19]See http://research.yahoo.com
[20]See http://www.usukitacs.com and http://www.usukita.org

to have significant support for key Semantic Web technologies such as RDF and SPARQL (See Section II-C for explanations of these terms). Drupal also offers a useful hybrid solution which provides an environment in which the flexibility of a Social Web application can be enabled, but it is still, fundamentally, a web-based CMS.

6) Microformats: Microformats are small specifications of data structures which match a particular need when describing something. For example the hCard microformat is a simple specification for representing people, places and organisations, and XFN is a social network related microformat. Microformats build on existing browser-based technology which is already widely used (e.g. XHTML) rather than predicating themselves on richer, but not yet widely adopted, standards such as RDF. The design principles of microformats are that they should be closely linked with existing technologies and practices, should be incremental where possible and should use existing definitions and standards where possible. Microformats are therefore a very pragmatic and potentially useful approach to explicitly representing the details of a particular entity when it is described, for example, in a blog post. The Semantic Web community has mixed opinions regarding microformats. Some advocate that microformats should be abandoned in favour of more rigorous RDF-based representations, while others insist that some form of structured form of representation is at least preferable to the use of unstructured representations. The transition from microformat to RDF-based formats is likely to be a simpler proposition than that of extracting the underlying entities from the unstructured textual descriptions.

GRDDL (Gleaning Resource Descriptions from Dialects of Languages)[21] is a technique for translating XHTML based data into RDF, and it can be used to translate any microformat representation into corresponding RDF output. GRDDL is potentially a valuable resource which can be used to bootstrap data on the Semantic Web, but it is not yet being widely used.

Some of the popular search engines (such as Google, Yahoo and Microsoft Bing) are now consuming microformat data in order to enhance the relevance and accuracy of their search results (this is the case whenever microformat data has been specified in the pages they index).

C. Linked Data Web

Closely linked to the Social Web is the concept of the Linked Data Web, which is a proposal for a web of "linked data" rather than the current web of linked documents. From a simplistic perspective the idea behind the Linked Data Web[22] is that information entities (rather than pages) should be published and shared on the WWW through the use of publicly available Uniform Resource Indicators (URIs), which can be dereferenced (looked up) to allow navigation between entities thereby improving discovery opportunities. Using this technique, information publishers can provide these URIs for the data that they publish, but also, importantly, they can refer to other existing entities in the Linked Data Web through the use of existing URIs to cross reference their data. A key change in mind-set between the current WWW and the Linked Data Web vision concerns the move from a document-centric web to a data-centric web, and this is a significant change in terms of both the end result and the processes and techniques used to publish the information in the web environment. The notion of the Semantic Web can be seen as almost synonymous with the notion of the Linked Data Web[23]; however, the aim of the Semantic Web is to provide greater semantic enrichment of online data resources compared to the Linked Data Web initiative.

An important facilitating component of the Linked Data Web is RDF (Resource Description Framework), which serves as a common representation format for the information returned when URIs are resolved. The combination of resolvable URIs and RDF provides the fundamental basis for the technical implementation of a navigable network of linked data in a web environment. This improved ability to author, publish and inter-link data sources in a standard representation format also provides an opportunity for key de-facto data sources to be created, thereby acting as pivots in this wider web of data. This is already being observed for certain data sets such as Geonames[24] (for geographical data), MusicBrainz[25] (for data about the music business) and DbPedia[26] (for general entities in the world, extracted from Wikipedia) and others. The widespread adoption and refinement of such data sources will yield powerful points of common reference and disambiguation when referring to concepts and trying to manage co-references.

There are a number of defined standards and approaches within the Semantic Web community that are potentially useful in terms of the adoption of the Linked Data Web, primarily in terms of vocabulary definition. For example, the Semantically-Interlinked Online Communities (SIOC) initiative[27] aims to provide a semantic basis for the integration of online Social Web community data, and the Simple Knowledge Organisation System (SKOS)[28] is a useful taxonomic representation resource, as is the metadata-related work from the Dublin Core Metadata Initiative[29]. There is also SPARQL, which is a standard Semantic Query Language that can be used to query any data represented in RDF and located within a special semantic database known as a triple store. Finally, there is RDFa which is a recent derivative of RDF. RDFa can be used in a manner similar to microformats, with RDF-like statements being embedded directly within XHTML web pages.

Recent advances in Natural Language Processing (NLP) and

[21] See http://www.w3.org/2004/01/rdxh/spec

[22] See http://linkeddata.org/

[23] For the purposes of this paper we use the term Linked Data Web and Semantic Web loosely. We see them as the same core concept with a subtly different focus on semantic expressivity and prominence of data.

[24] See http://geonames.org/

[25] See http://musicbrainz.org/

[26] See http://dbpedia.org/

[27] See http://sioc-project.org

[28] See http://www.w3.org/2004/02/skos/

[29] See http://dublincore.org

entity extraction also offer the potential for creating linked data from traditional unstructured textual data, such as blog entries or web pages; however, the accuracy of these techniques will determine the degree to which they are widely used or trusted by the user community. A popular example of such a service, which is publicly available and which returns results with links into the wider Linked Data Web, is OpenCalais[30] from Thomson Reuters.

Previous aspects of our ITA research have yielded various techniques to improve Semantic Web interoperability and accessibility. These include the GIDS (Global Interlinked Data Store) technique [4], which proposes a mechanism for distributing Linked Data across the network and treating the network as the database; the SWEDER (Semantic Wrapping of Existing Data Sources with Embedded Rules) technique [5] for making existing electronic data sources available in a readily consumable Semantic Web format with particular rules embedded to facilitate integration between datasets; and the POAF (Portable Ontology Aligned Fragments) technique [7] to address issues of efficient alignment between existing ontology resources. These techniques, and many others like them, are relevant to the ongoing Linked Data Web initiative and the broader desire for the evolution of the WWW into a future web of interlinked data sources. It is also acknowledged that Semantic Web environments are complex, so further extensions, standards and user interaction capabilities are required in order to make such environments sufficiently rewarding and user friendly.

III. Military Relevance

Throughout this paper we have identified a number of potential opportunities for various technologies and components to be used in support of rapidly formed coalition teams in a military environment. There are clearly components that can be reused directly or recreated within a managed intranet environment in order to provide basic collaborative capabilities. When collaborating with NGO and other non-military organisations it is possible that some information from the public internet will be referenced, so an ability to share and communicate into the public internet is likely to be valuable.

The challenges of the MANET network environment are (among other things) those of stability, availability and bandwidth constraints. These have a similar precedent in the WWW environment, specifically that of the mobile web environment, and specific new technologies to directly support disconnected working within a collaborative environment are now becoming increasingly common (e.g. Google Gears and HTML5 support for disconnected working). Other aspects of the mobile web, such as location-aware and direction-aware capabilities, are also directly relevant to the collaborative team-working scenario that we envisage, especially in cases where geographically-centered tasks and activities are being carried out, for example, disaster relief or humanitarian aid scenarios.

[30]http://www.opencalais.com/

There are militarily relevant opportunities for increased usage of core Web 2.0 technologies such as wikis and blogs, and the scenarios in which these can be used in a collaborative military team context are many. Perhaps a less obvious but potentially valuable capability which could also be harnessed is that of collaborative tagging, especially of resources such as photographs and videos or eyewitness reports etc. The advent of HTML5 and the opportunity for improved user interaction with data is also an emerging capability that could improve the general user experience, particularly if the web browser environment is appropriate for these kind of interactions.

In terms of linked data, the potential for using existing sources of structured information continues to grow; however, there are also opportunities for military (or government) organisations to contribute data into this open and publicly available network of interlinked data. This can either be in the form of collaborative incremental improvement or augmentation of existing resources, or through the publication of new datasets. Such actions would increase the value of the dataset which constitutes the Linked Data Web generally, and, would also have the effect of showing social participation in this vibrant environment rather than simply using the existing data resources. Some efforts are already under way in this area with both the US and UK government[31] beginning to make data available in various (usually non-semantic) online formats. This is a good start, and, can hopefully be improved upon over time, for example, by integrating data into the Linked Data Web[32].

The Linked Data Web can also be used as a template for the fusing of sensor-generated data, for example, through semantic representation of the sensor data and interlinking with other resources via resolvable URIs, as described by the Linked Data Web community. The potential for information fusion between automated sensor sources, user-generated content in collaborative systems, location-aware services and collaborative tagging techniques is significant. Also, the potential for Augmented Reality systems which fuse virtual data with real world objects is something that could be used to improve the capabilities of either individuals or teams, although Augmented Reality technology is still somewhat simplistic and immature at the time of writing.

There remain a number of issues related to security, identity and reputation in the WWW environment, and whilst positive steps are being taken in these areas, it is fair to say that all of the above potential opportunities for improving collaborative capabilities assume that this does not need to occur in a secure military environment. In the case of humanitarian aid or disaster relief, the need to collaborate rapidly and efficiently with diverse teams of people within and outside the military environment, it is likely that most collaborative exchanges will be undertaken in an insecure environment. In the case of the secure internal military environment, it is possible to foster collaborative techniques, perhaps based on the observed

[31]See http://www.data.gov/ and http://data.gov.uk/

[32]Efforts are under way at
http://data-gov.tw.rpi.edu/wiki/Generating_RDF_from_data.gov

behaviour of the Social Web, but it is unlikely that specific existing services (or even existing WWW technologies) could be used. It is also relevant to note the effect of the increased open and social nature of the WWW on military personnel, both in terms of their potential expectations for equivalent (or superior) capabilities in their military systems, and in terms of a potential increase in appreciation of the value of open collaborative systems. If this behaviour is manifest, it is likely to be somewhat inter-generational in nature, and as the generation in which these potential characteristics reach increasingly senior military roles it is possible that the organisational appetite for experiments or trial systems in these areas could increase, i.e. as the younger generation rise through the ranks, so they will serve as better champions for the adoption of such social technologies.

A number of the Web 2.0 and Social Web capabilities described in this paper have recently been applied to a recent UK ITA transition project, with particular emphasis on the support of advanced collaborative team working capabilities in a distributed network context. Early feedback suggests that the fusion of popular and successful collective intelligence capabilities with more advanced modes of semantic representation to support structured interaction are of great interest to our target user community.

IV. Conclusions

This paper summarises the latest advances in the areas of the Social Web, Web 2.0 and the Linked Data Web in the context of the capabilities that could be provided to support rapidly formed multi-national coalition teams. Examples are given of various leading edge websites and services which facilitate social interactions on the WWW, along with descriptions of key underlying attributes which are important to the ever-evolving Social Web environment. Specific technologies are outlined, particularly in the context of Web 2.0, with descriptions of how these might be applicable to supporting collaborative team behaviour in a military context. An outline of capabilities and technologies relating to the Linked Data Web is given, along with links to the Social Web and commentary on how future findings may be relevant to our particular area of interest. This paper finds that there are a number of technologies that can be used, and activities that could be undertaken to build environments within which these rapidly formed coalition teams could be more effectively supported. It is noted that the particular focus on MANET environments means that the need for disconnected modes of working are more critically important than for the usual Social Web, although mobile internet capabilities show similar characteristics and new standards and technologies are being created to better deal with disconnected modes of working.

Acknowledgement

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

[1] J. Bao, L. Ding, P. R. Smart, D. Braines, and G. Jones, "Rule modeling using semantic mediawiki," in *3rd Annual Conference of the International Technology Alliance (ACITA'09)*, September 2009. [Online]. Available: http://eprints.ecs.soton.ac.uk/17714/

[2] J. Bao, P. Smart, D. Braines, and N. Shadbolt, "A controlled natural language interface for semantic media wiki using the rabbit language," in *Workshop on Controlled Natural Language (CNL'09)*, June 2009. [Online]. Available: http://eprints.ecs.soton.ac.uk/17190/

[3] J. Bao, P. R. Smart, N. Shadbolt, D. Braines, and G. Jones, "A controlled natural language interface for semantic media wiki," in *3rd Annual Conference of the International Technology Alliance (ACITA'09)*, September 2009. [Online]. Available: http://eprints.ecs.soton.ac.uk/17713/

[4] D. Braines, G. Jones, P. R. Smart, J. Bao, and T. D. Huynh, "Gids: Global interlinked data store," in *3rd Annual Conference of the International Technology Alliance (ACITA'09)*, September 2009. [Online]. Available: http://eprints.ecs.soton.ac.uk/17705/

[5] D. Braines, Y. Kalfoglou, P. R. Smart, N. Shadbolt, and J. Bao, "A data-intensive lightweight semantic wrapper approach to aid information integration," in *4th International Workshop on Contexts and Ontologies (C&O 2008)*, July 2008. [Online]. Available: http://eprints.ecs.soton.ac.uk/15746/

[6] J. Hendler and T. Berners-Lee, "From the semantic web to social machines: A research challenge for ai on the world wide web." *Artif. Intell.*, vol. 174, no. 2, pp. 156–161, 2010. [Online]. Available: http://dblp.uni-trier.de/db/journals/ai/ai174.html#HendlerB10

[7] Y. Kalfoglou, P. R. Smart, D. Braines, and N. Shadbolt, "Poaf: Portable ontology aligned fragments," in *International Workshop on Ontologies: Reasoning and Modularity (WORM 2008)*, April 2008. [Online]. Available: http://eprints.ecs.soton.ac.uk/15749/

[8] T. B. Lee, *Weaving the Web*. Texere Publishing Ltd., November 1999. [Online]. Available: http://www.amazon.co.uk/exec/obidos/ASIN/0752820907/citeulike-21

[9] P. Smart, T. Huynh, D. Braines, K. Sycara, and N. Shadbolt, "Collective cognition: Exploring the dynamics of belief propagation and collective problem solving in multi-agent systems," July 2010. [Online]. Available: http://eprints.ecs.soton.ac.uk/17743/

[10] P. R. Smart, P. C. Engelbrecht, D. Braines, M. Strub, and C. Giammanco, "The network-extended mind," pp. 191–236, March 2010. [Online]. Available: http://eprints.ecs.soton.ac.uk/17739/

[11] P. R. Smart, T. D. Huynh, D. Mott, K. Sycara, D. Braines, M. Strub, W. Sieck, and N. R. Shadbolt, "Towards an understanding of shared understanding in military coalition contexts," *3rd Annual Conference of ITA (ACITA) 2009*, 2009. [Online]. Available: http://eprints.ecs.soton.ac.uk/17704/

[12] D. Tapscott and A. D. Williams, *Wikinomics: How Mass Collaboration Changes Everything*. Portfolio Hardcover, December 2006. [Online]. Available: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/1591841380