Unreasonable Expectations:

An Examination of the Semantic Web in the Light of the Original Vision for the Project

Dónal Deery

A research paper submitted to the University of Dublin,

in partial fulfillment of the requirements for the degree of

Master of Science Interactive Digital Media

2014

Declaration

I declare that the work described in this research paper is, except where otherwise stated, entirely my own work and has not been submitted as an exercise for a degree at this or any other university.

Signed: ___________________

Dónal Deery

28th February 2014

Permission to lend and/or copy

I agree that Trinity College Library may lend or copy this research Paper upon request.

Signed: ____________________

Dónal Deery

28th February 2014

Acknowledgements

I would like to thank my supervisor Rachel O’Dwyer for her advice, patience, and grammatical assistance during the preparation of this paper. I would also like to thank my parents for their continuing support and positivity.

Summary

This paper is concerned with the relationship between the Semantic Web as it was originally envisioned and the present status of the endeavour. The Semantic Web is an enhanced version of the existing World Wide Web in which data that can be processed by computers is added to web pages in order to make it easier for users to locate and exchange information. It was proposed by Sir Tim Berners-Lee, the invention of the original Web. The paper begins with a consideration of the original vision for the Semantic Web outlined by Berners-Lee and others around the turn of the millennium. The new generation of web technologies which were designed based upon this vision are then explored. The second chapter is centred upon the issue of which organisations are currently making use of Semantic Web technologies and principles, with particular attention being paid to major companies such as Facebook and Google. The third chapter takes as its focus the successes and difficulties experienced within the field of Semantic Web research in recent times. Chief among the successes is the Linked Open Data initiative which allows for related sets of structured data to be exposed in a uniform fashion, permitting like elements from distinct datasets to be related directly to each other, which enables new insights to be made at unprecedented scales. This paper suggests that some of the difficulties currently being encountered in the Semantic Web community are related to the engagement with artificial intelligence research, specifically the area of knowledge representation (KR). KR employs techniques based on traditional formal logic to encode meaning in a manner that allows computers to perform reasoning operations. This paper identifies a conflict between KR methodologies and the heterogeneity of data on the Web. The paper concludes that the field of Semantic Web research has become overly focused on KR problems and needs to return to the user-centred original vision if it is to generate worthwhile applications.

Table of Contents

Table of Figures

“There may be aspects of love, poetry and jokes that are too elusive to state clearly.”

John F. Sowa

Introduction

This paper is concerned with the relationship between the Semantic Web as it was originally envisioned and the present state of the endeavour. Simply put, the Semantic Web is an enhanced version of the existing World Wide Web in which data that can be processed by computers is added to web pages in order to make it easier for users to locate and exchange information. Contrary to what one might imagine, the idea did not originate from a working group within the World Wide Web Consortium (W3C), the international organisation which develops and maintains the software technologies from which the Web is assembled. Instead, the creation of a Semantic Web was proposed by Sir Tim Berners-Lee, the English software engineer who was responsible for the invention of the original Web. As the years have passed, more and more people have become involved in the effort, and Semantic Web research is now a considerable field whose members are drawn from both academia and commercial bodies.

However, it would be unreasonable to deny that Semantic Web research has a low public profile, and that even within computer science and related disciplines, understanding of and engagement with the objectives of the Semantic Web community remains limited. This would seem to be rather surprising for a number of reasons. Firstly, the impact that Semantic Web technologies are intended by their proponents to have on the manner in which knowledge is organised and exchanged online is a profound one. Secondly, the Semantic Web intersects with certain activities within artificial intelligence research, perhaps the most high-profile domain within computer science. Thirdly, as was noted above, the Semantic Web is the brainchild of the individual who conceived of and delivered the World Wide Web, which has been widely hailed as one of the most significant inventions in human history, and as such this successor proposal is worthy of serious analysis.

This paper is an attempt to explore these issues. It will begin with a consideration of the original vision for the Semantic Web outlined by Berners-Lee and others (perhaps tellingly) around the turn of the millennium. It will then be necessary to consider the new generation of web technologies which were designed based upon this vision and the debate which surrounded it; these are the tangible means developed to realise the Semantic Web.

Once the relatively brief history of the Semantic Web has been explored, the remainder of the paper will be broadly concerned with the present status of the endeavour.

The second chapter proceeds from the premise that large corporations are presently some of the most powerful actors on the Web. As such it will be centred upon the issue of which entities are currently making use of Semantic Web technologies and principles, with particular attention being paid to major companies such as Facebook and Google.

The third chapter takes as its focus the successes and difficulties experienced within the field of Semantic Web research in recent times. Chief among the successes is the Linked Open Data initiative, personally spearheaded by Berners-Lee, which has been the most high-profile Semantic Web project to date. It is the combination of two main elements, Linked Data and Open Data. Linked Data allows for related sets of structured data to be exposed in a uniform fashion, permitting like elements from distinct datasets to be related directly to each other. This enables new insights to be made at unprecedented scales. The datasets can be crowd-sourced or centrally sourced, or both. Linked Data could not be realised without Semantic Web technology. Open Data is an associated campaign to encourage the release of large datasets into the public domain, with a particular focus on government-held data that was funded by taxation. Provided that such datasets are made available under an open license, they can then form the basis of further Linked Data projects, or be integrated into existing ones.

Some of the difficulties currently being encountered in the Semantic Web community are related to the engagement with artificial intelligence research alluded to above, specifically a constituent area called knowledge representation (KR). KR employs techniques based on traditional formal logic to encode meaning in a manner that allows computers to perform reasoning operations. The apparent incongruity of KR methodologies with the real heterogeneity of data on the Web will be examined. Finally, the conclusion will assess the current state of the Semantic Web endeavour within the context of the original vision.

Chapter One - Image and Reality

1.1 - Visions

There are two rather distinct sources which can be said to outline the original Semantic Web vision. The first of these is Tim Berners-Lee's book, *Weaving the Web*, published in 1999 to coincide with the first decade of the Web's existence. The other is an article published in *Scientific American* in 2001 entitled "The Semantic Web" which Berners-Lee co-wrote with Jim Hendler and Ora Lassila, both artificial intelligence researchers. *Weaving the Web* was written when the Semantic Web project was in the very early stages of formalisation, and as such the vision outlined within it is, at certain points, ambitious enough to be classed as utopian. In a justifiably oft-cited passage, he states:

> I have a dream for the Web . . . and it has two parts. In the first part, the Web becomes a much more powerful means for collaboration between people. I have always imagined the information space as something to which everyone has immediate and intuitive access, and not just to browse, but to create. …the dream of people-to-people communication through shared knowledge must be possible for groups of all sizes, interacting electronically with as much ease as they do now in person.
>
> In the second part of the dream, collaborations extend to computers. Machines become capable of analyzing all the data on the Web - the content, links, and transactions between people and computers. A "Semantic Web," which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy, and our daily lives will be handled by machines talking to machines, leaving humans to provide the inspiration and inuition. The intelligent "agents" people have touted for ages will finally materialize. This machine-understandable Web will come about through the implementation of a series of technical advancements and social agreements that are now beginning… (Berners-Lee & Fischetti 1999, pp.169–170)

By contrast, the scenario envisaged in the *Scientific American* article is considerably more prosaic in its pragmatism, being concerned with the automated solution of a tricky scheduling task (Berners-Lee et al. 2001). The article is regarded as disclosing the canonical vision of the Semantic Web, and it attained a very high profile, being the first reference to the Semantic Web encountered by the majority of people; it has also been cited a great deal in the years since (Hendler 2011).

1.2 - Origins

For Tim Berners-Lee, the idea which would come to be called the Semantic Web is connected inextricably to the version of the World Wide Web which he brought into being in 1989, and which has become a ubiquitous feature of life in much of the world in the subsequent years. Hendler has said that for Berners-Lee, it was a concession to even apply the prefix 'semantic', which he felt was implied (ibid). As early as 1994, Berners-Lee was outlining the affordances which a Web containing machine-readable data would provide above and beyond the extant system of interlinked documents legible only to people (Berners-Lee 1994). In a summary of his plenary address to the inaugural International World Wide Web Conference, he stated that: "To a computer …the web is a flat, boring world devoid of meaning. This is a pity, as in fact documents on the web describe real objects and imaginary concepts, and give particular relationships between them" (ibid). The latter statement is telling as it expresses the conception of the relationship between form and content which the Semantic Web project seeks to inscribe – a movement by which the Web will come to consist not merely of referents, but pure meanings.

In his book, *Weaving the Web*, published in 1999, Berners-Lee situates the invention and expansion of the Web over its initial decade within the context of his vision for its future development; this is reflected in the book's subtitle: *The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. Its latter section in particular functions as something of a manifesto and inchoate blueprint for the endeavour. Yet the centrality of a machine-comprehensible Web to Berners-Lee's conception of the true value of the network is continually evident throughout the text.

This is never more the case than at the outset of the tale when Berners-Lee describes a program he wrote while working as a software consultant at the European Particle Physics Laboratory, better known as CERN, in 1980. The program, which was called ENQUIRE, was a type of database created in order that Berners-Lee might keep track of the "connections" between the thousands of CERN staff, their various projects, and the different computer systems at the lab upon which these projects were being run (Berners-Lee & Fischetti 1999, pp.4, 12). Berners-Lee describes ENQUIRE's design thus:

> The program was such that I could enter a new piece of information only if I linked it to an existing one. For every link, I had to describe what the relationship was. For example, if a page about Joe was linked to a page about a program, I had to state whether Joe made the

> program, used it, or whatever. Once told that Joe used a program, Enquire would also know, when displaying information about the program, that it was used by Joe. (ibid, p.11)

As can be seen, within this program lay not solely the germ of the Web, but also that of its prophesied successor; the concept of a machine-interpretable semantic relationship between the items in the network was as much a part of Berners-Lee's thinking as were the links themselves.

1.3 - The Real Structure

In March 1989, Berners-Lee prepared a proposal to develop not just the network of hypertext links suggested by the logic of ENQUIRE, but also the machine-readable markup which was to him an equally integral property of the system (ibid, p.23). Part of the reason that this aspect of the proposal held such appeal for Berners-Lee rested with his conviction that machine-readable markup of such a network, if analysed in an appropriate fashion, could yield profound insights into the structure of relationships in a given environment (ibid, p.24). Citing this original proposal almost a decade later, Berners-Lee appears somewhat sanguine about the scope of his latter-day ambition, presumably due to the challenge of conveying the potential affordances of semantic markup in the intervening decade:

> I was brash enough to look forward to having a web of data that could be processed by a machine. I said: "An intriguing possibility, given a large database with typed links, is that it allows some degree of automatic analysis. Imagine making a large, three-dimensional model, with people represented by little spheres, and strings between people who have something in common at work…perhaps you see tightly knit groups in some places, and in some places weak areas of communication spanned by only a few people. Perhaps a linked information system will allow us to see the real structure of the organisation in which we work." (ibid)

This contention is reiterated later on in the text in relation to the Semantic Web (ibid, p.175). In this manner, a semantic web is conceptualised as serving as a mirror of reality, provided that the necessary tools are available to render it legible. This remarkable capability can, in Berners-Lee's conception, be furnished by semantic markup of the nodes in the network. In this conceptualisation, the network infrastructure plays no part in the structuring of the social relations in the situation under observation; they instead reveal a truth previously obscured in the same manner as a no longer-novel technology like the microscope once did.

The methodological presuppositions that Berners-Lee displays here suggest an epistemological paradigm imported almost wholesale from the sciences. In this positivistic approach, a new technology allows nature to be glimpsed as it truly is, but no consideration is given to the possibility that the same novel technology plays a role in *constructing* the natural object (in this case an organisation) under observation. Another example of this would be John Johnston's critique of the Human Genome project, wherein "Molecular biology remains metaphysical…insofar as it disavows the conditions of its own possibility, namely, its complete dependence on information technology" (Johnston 2008, p.5).

1.4 - The Virtue of Simplicity

Berners-Lee's perception of this matter is further underlined in a passage later in the *Weaving the Web* where the logic of the burgeoning Web is compared to the fundamentals of physics: "One of the beautiful things about physics is its ongoing quest to find simple rules that describe the behaviour of very small, simple objects. Once found, these rules can often be scaled up to describe the behaviour of monumental systems in the real world" (ibid, p.38-9). This analogy is also an expression of Berners-Lee's conviction that the efficacy of a network like the Web lies in its putative resemblance to an organic system, with all the robustness and potency associated with same. Thus, a major design imperative for the endeavour was that it be constructed from as few simple components as was feasible in order to ensure that the technology be adopted as widely and rapidly as possible, allowing the network to expand exponentially (ibid, p.39).

The Hypertext Transfer Protocol (HTTP), operating in conjunction with Universal Resource Locators (URIs), were the fundamentals of the system that Berners-Lee designed and implemented. This configuration has been widely praised for facilitating the success of Web by virtue of its simplicity in precisely the manner he intended, as a system with: "…as close as possible to no rules at all" (ibid, p.17). The ingenuity of the design has of course drawn Berners-Lee a great deal of praise, even from those critical of his other proposals (Quitney Anderson & Rainie 2010, pp.15, 17). The acclaim for the architecture of the extant Web is extremely pertinent in relation to the Semantic Web because Berners-Lee contends that in order for the latter to be realised, the unprecedented software architecture devised for it must adhere to those same principles of simplicity:

> A reason for the success of the Web is that hypertext is so flexible a medium that the Web does not try to constrain the knowledge it tries to represent. The same must be true for the web of meaning [tellingly, this term refers to the Semantic Web]. […] The trick …is to make sure that each limited mechanical part of the Web, each application, is within itself composed of simple parts that will never get too powerful. […] The mechanisms for metadata, privacy, payment and so on will all work in a well-defined way. The art of designing applications in the future will be to fit them into the new Web in all its complexity, yet make them individually simple enough to work reliably each time. (Berners-Lee & Fischetti 1999, pp.197–8)

This statement is rather crucial to the issues which will be dealt with in the third chapter, as it articulates Berners-Lee's commitment to a robust Semantic Web which does not demand standardised input data in order to function effectively.

1.5 - Bottom of the Stack

As Berners-Lee was writing *Weaving the Web* in the late 1990s, he was also deeply involved in the early stages of the design of the Semantic Web's architecture, which taking place under the aegis of the World Wide Web Consortium (W3C), the organisation that he established in 1994 to develop and maintain standards on the Web. At this point in time, the Extensible Markup Language (XML) had just been formalised and adopted by the W3C, permitting it to be employed as the language in which many of the Semantic Web's component layers would be written. Like the Hypertext Markup Language (HTML) developed by Berners-Lee for use on the Web, XML was derived from the Standard Generalised Markup Language (SGML), which itself was designed for use as a printer control language (Schmidt 2010, p.4).

It should be noted that while HTML documents must broadly conform to a specified configuration with respect to the tags or elements which they contain, the tags contained in an XML document are defined entirely by its creator, leading to a plethora of potential usage scenarios, and also massive degrees of variation and idiosyncrasy. Although Berners-Lee was conscious of the potential challenges for interoperability presented by the protean properties of XML, he was confident that they could be overcome (Berners-Lee & Fischetti 1999, p.174). These issues were raised during the initial design of the Semantic Web architecture, or "stack", particularly in relation to the component layers that were expressed in XML, the first of which to be defined was the Resource Description Framework (RDF).

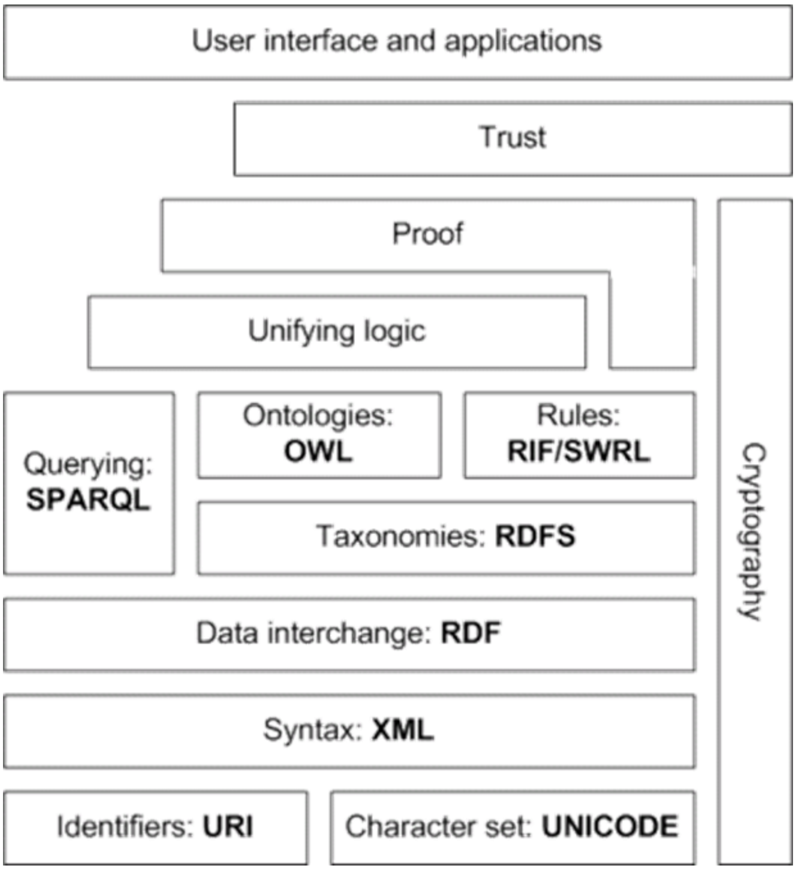

Fig.1.1: Diagram of Semantic Web Stack derived from slides by Berners-Lee (Wikipedia Editors 2014)

RDF is in a sense the most prominent component of the Semantic Web, and to an extent, it embodies the underlying concept of the project in microcosm. It has been suggested by members of the Semantic Web research community that it would be more suitable if RDF stood for "Rich Data Format", which is appropriate given that it effectively represents an attempt to create the most efficacious metadata format possible (Antoniou et al. 2012, p.4). One might consider the function of RDF to be a development of the data which has traditionally been included in the `<meta>` tags within the `<head>` elements of a web page. Even today, this descriptive information, sometimes based on the widely used Dublin Core metadata format, generally represents the sum total of data on a given page that is intended to be processed by machines. It should be borne in mind however that this data is not truly “semantic” in the sense of the term employed by the Semantic Web community, as it is intended to be indexed by search engine web crawler programs using traditional string-based keyword matching techniques, with no deduction or reasoning occurring on the part of the machine – no processing of “meaning”.

The solution to this presented by RDF is to permit not only the description of the contents of a page, as is the case with conventional metadata, but also the capacity to represent on the Web, by means of markup, “…any identifiable thing, including things that may not be directly retrievable on the Web”, such as a person (Miller & Manola 2004). This is achieved by associating unique URIs with relevant items of information, and decomposing them small groups of three statements called triples, each having a defined relationship with the other items which itself is associated with a URI. Triples adhere to a standard “subject: predicate: object” format, thus a person might be identified by a given URI, which might be linked via an attribute URI `#fullName` to a string containing that full name (ibid). RDF is characterised as having a “graph-centric” data model, and is perhaps most easily comprehended when represented visually; of course, in order that it be machine-readable, it must then transferred into a one-dimensional form to be expressed in any of the set of syntaxes which have been developed for the purpose, some in XML, some not (Antoniou et al. 2012, p.28).

The layer directly above RDF is RDF Schema, a taxonomy system the function of which is to provide a structure of generalised terms within which the particulars of the data stored in RDF can be usefully contextualised. Whereas RDF is concerned largely with individual objects or instances, RDFS can be said to be concerned with classification and defining the types of objects (ibid, p.40). In a sense, RDFS defines the relationships that can exist between the individual subject and object items in RDF triples. The other key aspect of RDFS is properties, which permit “inheritance” between classes. For example, given that a particular apartment A is a member of the class of apartment buildings in general, then it will inherit the ability to be rented from its parent class (ibid, p.43). Antoniou et al. observe that the capacity to enforce an intended meaning of the subclass relationship in this manner, so that it will be interpreted identically by all parsing software – to fix its semantic definition – qualifies RDFS as “a primitive ontology language” (ibid).

1.6 – Top of the Heap

In the terminology of information science, an ontology is a document, file or language that defines the relationship between terms (Berners-Lee et al. 2001, p.40). As was noted above, RDFS can be considered to be a simple ontology language, but the primary ontology function in the Semantic Web is performed by the Web Ontology Language (OWL), which represents perhaps the most involved layer of the stack. OWL is written in the RDF syntax, but its purpose is to facilitate the expression of "…more advanced, more 'expressive', knowledge" than is permitted by the intentionally limited structure provided by RDF and RDFS (Antoniou et al. 2012, p.91). The term "OWL" has become a general name that now denotes a set of distinct sublanguages. The W3C's original specification was originally formalised in 2004 at the same time as RDF and RDFS, but was subsequently revised, with a replacement version containing additional functionality, called OWL 2, becoming a Recommendation in 2009 (Dean & Schreiber 2009).

The original OWL Recommendation is divided into three sublanguages that have an ascending scale of descriptive power, and therefore complexity. OWL Lite is the smallest, most restricted variant, intended for users primarily needing a classification hierarchy and simple constraints (McGuinness & van Harmelen 2009). OWL Description Logic (OWL DL) is a superset of OWL Lite; it is intended to supply maximum expressiveness while still remaining capable of providing an answer in a reasoning program within an acceptable time period; to employ the language of mathematics, it is computationally decidable (ibid; Antoniou et al. 2012, p.98). OWL Full is the superset of OWL DL by virtue of the fact that it has fewer constraints. It gains this expressive power at the cost of computational completeness; Antoniou et al. assert that this "…[dashes] any hope of …efficient reasoning support" (ibid). Even the W3C specification editors concede that because Description Logic is derived from traditional predicate logic: "It is unlikely that any reasoning software will be able to support complete reasoning for every feature of OWL Full" (McGuinness & van Harmelen 2009).

Not all of the layers of the Semantic Web stack have been realised in the decade and a half in which work has been ongoing on the project. Among the other layers that have been realised is the SPARQL Protocol and RDF Query Language (SPARQL), a means to query data stored in RDF triples, developed from the ubiquitous Structured Query Language (SQL) used in databases. There is also the Rule Interchange Format

(RIF), a W3C Recommendation the title of which indicates the fact that the objective of the working group which developed it was not to attempt to create a new rule language which would suit all purposes, but instead focussed on the interchange among the various rule systems in use on the Web (Antoniou et al. 2012, p.148). Other systems for rule interchange, such as the Semantic Web Rule Language (SWRL) are also currently under consideration (ibid, p.155). The development of rule languages can be a problematic area of activity due to some of the issues regarding consensus and universality which also hinder the broad deployment of ontology languages, a matter which will be examined in the third chapter.

The complexity of the Semantic Web stack is unmistakably evident, even in the much simplified form in which it has been presented here.

1.7 - Conclusion

It has been shown that ‘semantics’ – or machine-readable markup – has been integral to Tim Berners-Lee’s vision of the original Web since prior to its inception, due in part to his successful experiments with ENQUIRE. Berners-Lee identifies such semantic markup on the Web as a means to obtain a clearer picture of the manner in which complex phenomena like social organisations operate on a grand scale, but without considering the role that the network plays in constituting the structure of such phenomena. He also expresses the view that the success of the Web is due to the fundamental simplicity of its component parts, and that the technologies upon which the Semantic Web will be based must be equally simple in order for it to flourish in a similarly exponential fashion. The actual stack architecture which was developed over the better part of a decade by many W3C working groups was then considered. It could certainly be argued that the Semantic Web stack as delivered appears, to an observer, to be considerably more complex than the system which Berners-Lee outlines in *Weaving the Web*. The area of ontologies appears to be particularly involved.

With an understanding of the purpose of the Semantic Web intended by its creator, in addition to an awareness of the software technologies developed to support it, the next priority should be to ascertain how these ideas and systems are faring out in the wild, beyond the confines of the W3C.

Chapter Two - Adoption or Adaption

2.1 - A Changed Landscape

To assert that the world has changed immeasurably in the years since Tim Berners-Lee first outlined his vision for the Semantic Web in *Weaving the Web* would be something of an understatement. The Web of the 1990s, which formed the backdrop to the W3C's initial efforts to formalise the Semantic Web, was certainly less saturated by online commerce than is the case today, as many companies were slow to identify the vast potential market which could be reached via the Web.

The intervening period has seen the ascent of Google, which was one search engine provider among many when it was launched in 1998, to its present position as one of the most prominent, influential, and powerful companies in the world, with an effective monopoly in its original sector, and diverse additional activities. The past decade or so has also seen the emergence of a great many sites designed to facilitate the generation and publication of content by non-expert, non-enthusiast web users, with social media applications constituting a large part of this. These developments are often collectively referred to under the rubric: "Web 2.0" (DiNucci 1999, p.32). The nature of this appellation can be read as an indication and perhaps also as an unwitting indictment of the enduringly low public profile of Semantic Web activities; after all, according to Berners-Lee's vision it is the latter which would be the new and improved second iteration of the original Web.

The major developments outlined above are merely a very brief summary, and much else has happened to the Web in the past fifteen years or so. It would however, be a mistake to assume that such matters have passed the field of Semantic Web research by, or indeed rendered its objectives largely obsolete or invalid. In order to ascertain the degree to which the Semantic Web has remained relevant to the contemporary Web, it is necessary to examine the level of engagement with its methodologies by these newly ascendant companies. Facebook and Google are especially appropriate here as their popularity, combined with their modes of operation mean that a great many users increasingly interact with the Web through the filter of these sites.

2.2 - Absent Social Machines?

Considering the original Semantic Web vision in relation to the advent of user-generated content and associated social media, it could be argued that Berners-Lee was successful in identifying areas in which the original Web was lacking. Although ultimately, the area of social networking, for lack of a better term, appears to have also developed in a manner somewhat different from what he anticipated, certainly with regard to scale and sophistication.

However, according to Berners-Lee himself, the "abstract social machines" he speaks of in *Weaving the Web* have been realised, albeit with some caveats. In that text 'social machines' were glossed in vague fashion as "…processes in which the people do the creative work and the machine does the administration" (Berners-Lee & Fischetti 1999, p.186). Berners-Lee believes this definition describes the unseen operation of Wikipedia, specifically the interaction between the mediawiki content management system and the social layer of super-users who oversee quality control on the site (Berners-Lee & Hendler 2010, p.156). Further examples cited include the "trackbacks" and commenting systems employed by blogging sites in order to foster interaction between users, and also include prominent social networking sites such as Facebook, and the now-marginalised MySpace. Berners-Lee and Hendler assert that while these are all examples of 'social machines', they are merely "early versions", impoverished by virtue of the fact that they are not yet true semantic applications because "…they function largely isolated from one another" (ibid, p.157).

In other words, they constitute discrete datasets which remain sealed off from each other, often referred to as "siloing" – a state of affairs lamented by those within the Semantic Web research community and related fields (Walpole 2013, p.100). Berners-Lee and Hendler reaffirm that the design imperative of the Semantic Web is the representation on the Web of that which was previously unrepresented:

> Much as the architecture of the current Web allows a virtually unlimited scaling of the Web of documents, the architecture of the future Web must be designed to allow the virtually unlimited interaction of the Web of people. (Berners-Lee & Hendler 2010, p.157)

According to the authors, this new horizon – which they feel compelled to dub "Web 3.0" – is urgently required: "…because the magnitude of the problems that our society faces today are such that only the concerted effort of groups of people operating with a joint power much greater than that of the individual can hope to provide solutions” (ibid). The attainment of this goal demands a considerable increase in the sophistication of the ‘social machines’ on the Web. The authors aver that the Semantic Web technology developed by the W3C, which is, after all, the outcome of years of work by domain experts, is the optimal choice to serve as the basis for such an undertaking (ibid).

This is rather a crucial matter as far as the current status of the Semantic Web is concerned. In large part, the health of the ongoing endeavour can be judged by the extent to which the actors with real agency on the Web, namely major companies, are adopting Semantic Web methodologies. The present state of affairs could be perceived as a sliding scale, with a scenario wherein the W3C-derived Semantic Web architecture is employed in full as designed at one end, and at the other, a scenario in which the core concept of semantic linking is realised through entirely distinct technical means. The degree to which cases near the latter end of the scale can be considered evidence of the gradual realisation of the Semantic Web vision is a matter of continuing debate within the community (Hendler 2011).

The utilisation or otherwise of open standards is a key issue here, as the deployment by corporate actors of systems which are proprietary in nature could serve to exacerbate the ‘siloing’ problem. In her response to the Pew Research Center’s 2010 survey concerning the future of the Semantic Web, Susan Crawford, Internet law professor at the University of Michigan, envisions a scenario in which “[t]here will be more and better meta-information, but it will continue to be opportunistic, siloed, and ad hoc” (Quitney Anderson & Rainie 2010, p.10).

2.3 - User Friendliness and the Facebook OGP

A key locus of this debate is the Open Graph Protocol (OGP), a component of the Facebook Platform software environment. Introduced in 2010, the OGP permits Facebook users to “like” content on many different websites, where previously the facility was available only on the Facebook site itself (Allemang & Hendler 2011, p.203). The OGP was thus designed to be used primarily by third-party sites, who could embed it within their own pages in order in the hope of potentially increasing the visibility of their product by leveraging Facebook's vast user base. As a consequence of this, OGP has a very simple data model of some two dozen types; user testing by the company showed that external content managers were unwilling to learn the pre-existing metadata properties (for example, the different names used to refer to email addresses) upon which the system could otherwise have been built (ibid, p.204). Hendler and Allemang argue that this disjunction is easily solved by the application of an overarching layer of RDF-Schema which could explain that the OGP property `og:email` is equivalent to the `foaf:mbox` property from Friend of a Friend (FOAF), a common metadata format for describing relationships between people (ibid). Although this method resolves a technical issue raised by multiple formats in this instance, the reluctance of semi-skilled users – the content managers on third-party sites – to engage with the pre-existing formats, including those developed by the W3C, should be a matter of some concern to proponents of the Semantic Web.

Hendler is somewhat fond of dismissing a straw-man argument against the Semantic Web advanced by those ignorant of the genuine subtleties of the undertaking. This simplistic view, associated with the media commentator Clay Shirky, holds that Semantic Web research is an inherently misjudged attempt to construct a monolithic schema capable of describing all knowledge, and is *ipso facto* an impossible task (Shirky 2003; Hendler 2011). Hendler quite reasonably points out that the inevitability of dealing with multiple languages has always been an aspect of the Semantic Web's design, indicating the "Non-Unique Naming Assumption", which permits the same resource to be addressed by different names using distinct URIs (Allemang & Hendler 2011, p.9). However, this state of affairs does not negate more nuanced critiques, and the fact remains that every time a new format like the OGP is added to the anticipated myriad, someone must be allocated to the task of integrating them into the W3C's Semantic Web architecture. If this labour-intensive translation activity is not scrupulously pursued, two unpleasant consequences result: first, the prized diversity of the Web as embodied by these distinct formats may be undermined, and second, the deplored 'siloing' problem may endure in a renewed form, as Susan Crawford suggested.

2.4 - Semantic Web Principles or Semantic Web Technology?

With respect to the sliding scale detailed above, Jim Hendler can be placed among those who believe that it is the *principles* underpinning the Semantic Web that truly matter, over and above the W3C's technologies. For him, the Facebook OGP is "…arguably the most successful Semantic Web model ever" (by virtue of the scale and rapidity of its uptake) (Allemang & Hendler 2011, p.205). At the 2011 European Semantic Web Conference Hendler's contribution was facetiously entitled "Why the Semantic Web will never work" and directly addressed the current state of the art with specific reference to the famous *Scientific American* article of 2001 that he co-wrote with Berners-Lee and Ora Lassila, among other Semantic Web vision pieces (Hendler 2011). During the course of the lecture, he avers that the OGP is the only evidential example necessary in order to justify the assertion that the Semantic Web is being realised; that, contrary to what some might say, it is already "here" (ibid).

There is pragmatism in Hendler's position, insofar as the fate of the Semantic Web, like so many other endeavours, will likely be determined by the corporate entities which are the actors with the most agency on the contemporary Web. Facebook's OGP may be the most prominent example of a major company employing technology that draws upon Semantic Web principles, but it is not the only one. Good Relations (GR) is an ontology developed to allow retailers to provide detailed descriptions of their offerings. Like OGP, it is included in the `<meta>` tag of a page using RDFa[1], which allows it to be indexed by web crawlers, ultimately allowing the search engine to relay these detailed descriptions in their results lists, ultimately benefiting both parties (Allemang & Hendler 2011, p.280).

2.5 - Finding Answers with Google

It would be reasonable to assert that for the most part, Semantic Web research was focussed on the realisation of an alternative paradigm of information retrieval distinct from the string-matching based keyword search system which has, sometimes in combination with other factors, served the various search engine companies so well. Jim Hendler has remarked that the hundreds of thousands of page results returned by a typical query to a

[1] Resource Description Framework in Attributes, a type of RDF designed to be included in the Meta element of a HTML page along with existing metadata types.

contemporary search engine – none of which will necessarily contain all the terms from the query – is radically different from the system envisioned by himself and other Semantic Web researchers (Hendler 2011). In their scenario, the user would receive a list of results which were wholly relevant; the list might be very brief or even empty, depending on the query (ibid). This is indicative of a belief that web users would rapidly tire of the crude nature of keyword search, whereupon Semantic Web researchers would be able to introduce their preferred approach, wherein finesse, in the form of reasoning programs operating over a rich set of ontologies, would prevail over reasoning-free 'brute force' methods.

The most powerful company on the Web ultimately has a greater influence over the future of the Semantic Web than any other factor. Unfortunately, Google's intentions in this area remain typically obscured. For a considerable period, there was little indication that the company was engaging in any way with the ideas emerging from the Semantic Web community. Perhaps this was unsurprising, given that it had had so much success with the keyword-based searching that Semantic Web advocates like Hendler aim to obsolesce. Then, in 2010, Google purchased Metaweb, the company that created Freebase, a collaborative knowledge base which anyone can contribute to, which functions by storing facts as triples in accordance with the Semantic Web design methodology. The language used in the attendant press release is strikingly reminiscent of Tim Berners-Lee's original vision of a semantic web:

> The web isn't merely words—it's information about things in the real world, and understanding the relationships between real-world entities can help us deliver relevant information more quickly. …we've acquired Metaweb, a company that maintains an open database of things in the world. Working together we want to improve search and make the web richer and more meaningful for everyone. (Menzel 2010)

Approximately two years after their acquisition of Metaweb, Google announced the Knowledge Graph, a knowledge base designed to augment its existing search functionality with semantic data. Freebase has been characterised as the core of the Knowledge Graph; as of November 2012 it held some 23 million entities in its database, almost twice as many as it contained when Metaweb was purchased (Filloux 2012). Other sources upon which the Knowledge Graph relies include the CIA World Factbook, and Wikipedia (presumably DBpedia, a knowledge base containing information from Wikipedia expressed as RDF triples), and as of June 2012 it contained more than 500 million objects, as well as more than 3.5 billion facts about and relationships between these different objects (Singhal 2012).

The outward evidence of this major upgrade to Google's search system was typically subtle. If a user searches for a term with a corresponding entity in the knowledge base they will be presented by a fact box at top of the first page of their results. This will contain information, primarily drawn from Wikipedia, including associated data items determined by the type of search subject; thus the fact box for a band would also list some of their songs and albums, the entry for a city lists area, weather, points of interest. Each such item is itself a link to its corresponding Knowledge Graph entry and attendant search results. Where possible, Graph fact boxes also contain a "People also search for" list, which is intended to potentially anticipate the user's next query as well as facilitating serendipitous discovery (ibid). When a search term has no corresponding entry in the Graph, the search results appear just as they always have. In this manner, the Knowledge Graph supplements keyword-based search, which would appear to be in accord with Google's strategy, in so far as that can be ascertained.

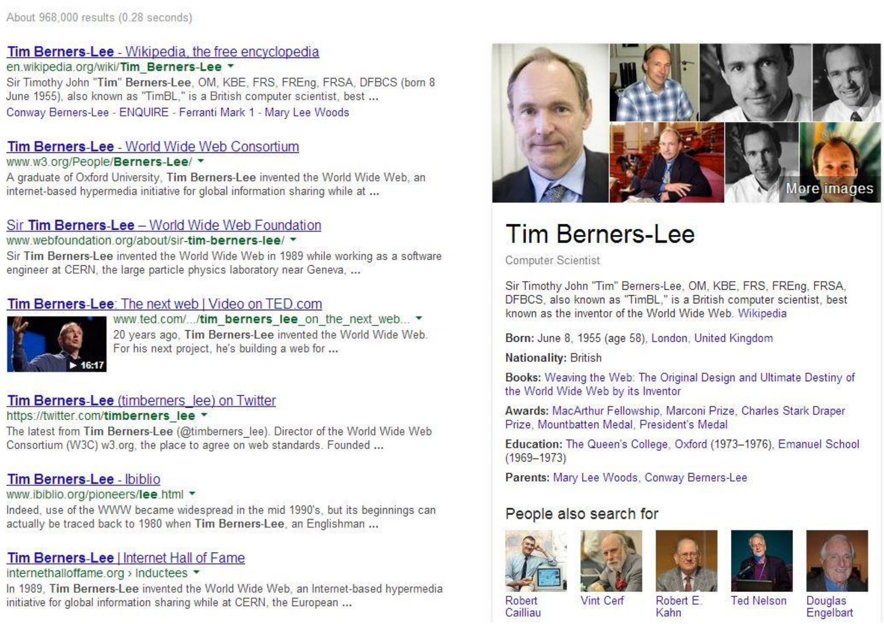

Fig.2.1: Screen capture of Google results page with Knowledge Graph entry

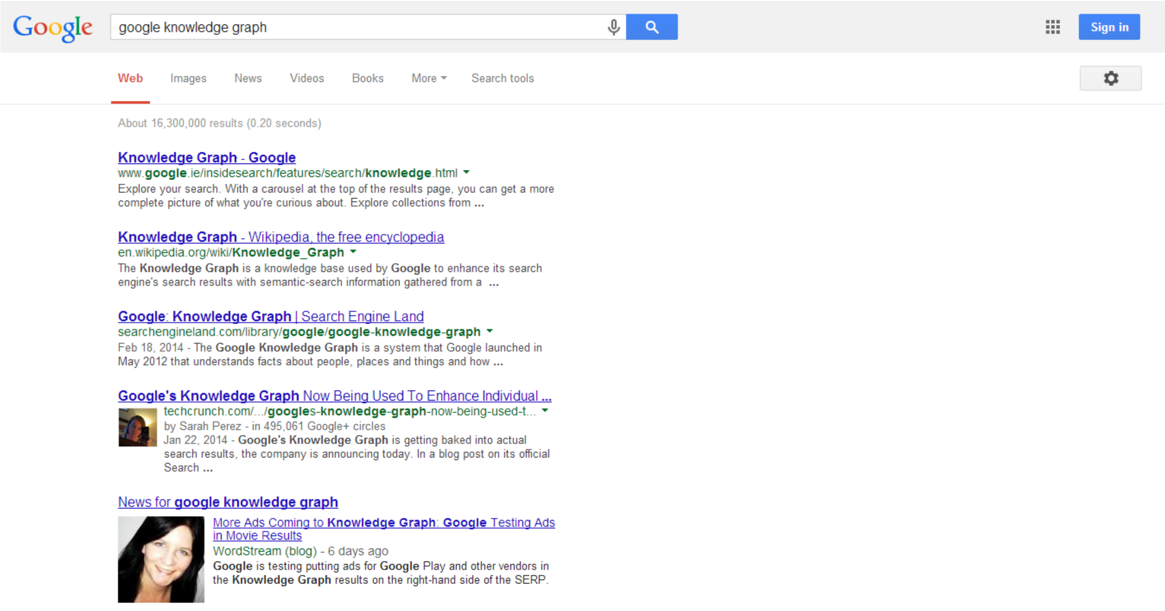

Fig.2.2: Screen capture of Google results page without Knowledge Graph entry

The company appears to be focussing on processing natural language queries in order to better facilitate what it calls "conversational search", and the largest major update to its search algorithm in three years, entitled "Hummingbird", was introduced in either August or September of 2013 in support of this policy (Gibbs 2013). Hummingbird was designed to leverage the semantic data held in the Knowledge Graph to process and understand *all* of the terms within a given query, in a manner akin to that advocated by Hendler (ibid; Hendler 2011). As a consequence, the Google search engine now has simple reasoning functionality, such that users can receive an answer to factual queries expressed in natural language, for example: "how tall is Barack Obama's wife?" will yield a Knowledge Graph fact box containing the answer in addition to the Graph entry for Michelle Obama.

2.6 - How Semantic is the Knowledge Graph?

It would seem that advocates of the Semantic Web should be relieved that Google has effectively offered a major validation of their principles by employing some of them in a major new undertaking. However, when the Knowledge Graph was first announced to the public in May 2012, Rafe Needleman, a reporter for CNET, observed that Jack Menzel, Product Management Director of Search at Google, didn't use the word 'semantic' at all in his presentation (Needleman 2012). When asked for his thoughts on the Graph, Ivan

Herman, Semantic Web Activity Lead at the W3C, asserted that it was closely akin to existing Semantic Web models, and as such welcomed it as "fundamentally…great stuff" (Zaino 2012b). As a consequence of this perceived inheritance, Herman expresses disappointment that Google failed to mention that: "…they benefitted from the work a lot of people have done. It's perfectly okay that they use it, and we're happy that they are, and it's okay if they use different terms that go down better for the average user, but [some acknowledgement] would have been nice" (ibid). According to Needleman, Menzel, for his part, argued that while the Semantic Web represented an ideal worth striving for, the Knowledge Graph is a more pragmatic endeavour, and "…not what people talk about when they discuss Semantic Web concepts. "We do continue to work on how to make search semantic," [Menzel] says, "but talking about it brings out the crazy people" (Needleman 2012).

It is somewhat difficult to accept Menzel's contention that the Knowledge Graph bears little conceptual relation to the Semantic Web when one is aware of the former's mode of operation and its origins. One must conclude that for Menzel, Google's focus on natural language processing, in which the machine is calibrated to make sense of human-readable documents, constitutes an entirely distinct approach to the Semantic Web paradigm, in which humans create machine-readable data in order to describe human-readable documents. It seems equally possible that he is of the view that Google's decision to develop a bespoke architecture for the Knowledge Graph instead of directly utilising the W3C's Semantic Web stack automatically places Google's semantic activities outside the field of Semantic Web research.

A perceived lack of gratitude toward the Semantic Web research community on Google's part, as broached by Herman, may also stem from the bespoke nature of the Graph's architecture – one can infer that the company considers their effort sufficiently original to negate any obligation to acknowledge any sources of influence. This conflicts somewhat with the position of those such as Hendler who consider the Semantic Web to be a set of principles rather than a given technological model, as corporations are not necessarily willing to share. This is a particular issue in relation to the Knowledge Graph, for although some of the component knowledge bases, such as Freebase, remain open to all, the mechanics of the Graph are wholly obscured as a consequence of their integration with Google's search algorithm, which remains one of the most valuable items of proprietary technology in the world.

As a result of the closed nature of Google's system, there is no comprehensive index of the entities contained within the Knowledge Graph available in the public domain (Zaino 2012a). Furthermore, there seems to be little prospect of this information being exposed, except by means of a laborious hack (ibid). The result is a situation in which perhaps the largest single system constructed on the principles of Linked Data cannot be connected to any of the other myriad Linked Datasets on the Web, violating the principles of openness and free exchange held dear by Berners-Lee and other Semantic Web proponents, - principles that underpin the operational logic of Linked Data itself. Similarly, if Google Search can answer queries directly by means of Knowledge Graph fact boxes, then there is no need for a user to leave Google's page, with Google's ads. Both these cases demonstrate the manner in which semantic technologies may be employed in order that Google maintain its competitive advantage.

2.7 – Conclusion

The position in which the Semantic Web community presently occupies is indeed a complex one. Berners-Lee can reasonably claim that because collaboration was, as we saw in the previous chapter, always a major component of his stated vision for the original Web, 'Web 2.0' is therefore something of a misnomer. His definition of 'social machines' is sufficiently vague that it encompasses a good deal of the web systems geared toward user-generated content, whether commercial like Wordpress and Blogger, or not, like Wikipedia. The question of commercial interest becomes more complex where the Facebook OGP is concerned. As was noted above, the Semantic Web community is presented with a double bind. On the one hand, refusing to accept the bespoke configuration of the OGP diminishes the diversity of input supposedly valued by the community. On the other, the any novel bespoke format, commercial or otherwise, must be translated into other formats, requiring considerable time and effort. If this is not attended to, then the much-maligned silo problem recurs. Thus, commercial practices can place Semantic Web researchers in awkward positions.

The situation with Google is also closely related to the issue of whether or not being a Semantic Web researcher involves a commitment to the technologies of the stack, or to a set of design principles. Whether or not the Google Knowledge Graph is a semantic project remains a matter of debate as long as the inner workings of their system remain confidential.

As was alluded to in the previous section, there are also further moral issues involved in the commercial exploitation of material, software or otherwise, created for the public domain. As Garton analyst Darin Stewart points out: “Linked Open Data is a public resource created by countless hours of effort from anonymous stewards. Acknowledging that contribution would not only be respectful, it would incentivise the creation of even more Linked Data” (Stewart 2012). Thus, the Knowledge Graph functions as a hoard-cum-silo, and the privatisation and exploitation of the collective, collaborative effort that is the lifeblood of the Web in the eyes of Berners-Lee and those like him, may ultimately harm the cause of the Semantic Web by discouraging the volunteer effort upon which its realisation depends. Thus, it can be seen that the interests of companies and those of the Semantic Web community cannot be expected to align in all cases, and that as such this presents a major source of friction given that companies are now the primary agents where the development of the Web is concerned.

Chapter Three - Successes and Challenges

3.1 - Uncertain Terrain

Continuing the examination of the present state of the Semantic Web project, this chapter is concerned with what is almost certainly the most prominent success from the field, namely the Linked Data initiative. It is also concerned with the most serious challenges facing Semantic Web research, which are the ontology situation and the potential neglect of end-users. Each of these issues will be examined in turn, and the manner in which each issue relates to the original vision of the project will be considered in turn, in addition to being contextualised with regard to the project as a whole.

Jim Hendler's "Why the Semantic Web will never work" lecture, delivered at the European Semantic Web Conference in 2011, is an appraisal of the status of the endeavour, a decade on from the landmark Berners Lee et al. *Scientific American* article. He sounds a note of caution regarding the original Semantic Web vision of Tim Berners-Lee and others, stating that while he believes that it is achievable in its entirety, many considerable barriers remain as far as the attainment of this objective is concerned (Hendler 2011).

Hendler expresses the conviction, discussed in the previous chapter, that commercial projects such as the Facebook Open Graph Protocol and the Good Relations ontology are incontrovertible evidence to sceptics and detractors that the Semantic Web is being realised (ibid). However, he does underline that these and other developments, such as the growth of Linked Data, collectively represent mere aspects of the original vision rather than its totality (ibid). Hendler considers Linked Data to be the preeminent example of a Semantic Web component that has been broadly embraced by those outside the Semantic Web research community (ibid).

3.2 - Linked Data: The Semantic Web's Successful Offspring

The term "Linked Data" was introduced by Berners-Lee around 2006 on his W3C *Design Issues* blog, although this did not coincide with the publication of any new technical standard by the W3C (Berners-Lee 2009a). This was due to the fact that the technology which underpins Linked Data had been a key component of the Semantic Web from the time of its first formalisation in the late 1990s; indeed one could argue, as Berners-Lee has, that it is the *sine qua non* of the Semantic Web, insofar as it is the RDF-based system by which URIs can be employed to represent concepts rather than merely documents (Berners-Lee 2009a). In the section of *Weaving the Web* pertaining to the Semantic Web, Berners-Lee advocated the linking of identical concepts which had distinct names in separate datasets in order to obtain insights at a greater scale than possible when related data remain in discrete silos (Berners-Lee & Fischetti 1999, p.201).

In the years following the 2006 *Design Issues* post, more effort was put into publicising Linked Data as an initiative in its own right. This culminated in an address by Berners-Lee to a 2009 TED conference, during which he called for major organisations, but primarily governments, to release their accumulated datasets into the public domain in order that their contents be processed and linked (Berners-Lee 2009b). The address could be perceived as a re-launch for the Semantic Web in a dilute, more pragmatic form; in an echo of *Weaving the Web*'s format from a decade prior, the Linked Data proposal is couched as the next step in the evolution of the World Wide Web, then almost twenty years into its existence (ibid). The attempt to synergise a re-publicising of the Semantic Web with movements by governments to be more open with the data they were gathering – soon to be called "Open Data" – was successful and Berners-Lee's appeal for "raw data now"[2] was largely heeded (Berners-Lee 2010). The decision to focus on state-held datasets, funded by taxation, was prudent insofar as the commercial enterprises to which Berners-Lee also addresses himself are under no tangible obligation to share their data – after all, it may well be their key asset. It remains to be seen whether the social networking sites he alludes to will end their policy of avoiding interoperability and maintaining distinct silos of user data, an issue dealt with in the previous chapter. At present it appears to be another instance of a Semantic Web proponent expressing an unreasonably optimistic attitude to the relationship between the interests of commerce and the interests of the public.

In a 2010 addition to the *Design Issues* post, Berners-Lee proposed a five-level grading system to reflect the degree to which a given dataset, released into the public domain could be said to reflect best practice for Linked Data. The minimum grade is

[2] Raw data is a term which describes data which are not yet in a form that can be processed by a Linked Data system. It refers primarily to data not yet in machine readable form, even a "photo of a scan of a fax of a table" (Berners-Lee 2009a). Structured data in proprietary formats, such as MS Excel, are preferable to totally raw data, open formats such as CSV are in turn preferable to proprietary ones (ibid).

attained when the raw dataset is placed on the Web in any form, provided that it is available under an open licence; the maximum denotes a dataset in non-proprietary RDF form, with entities linked to equivalents in related datasets to furnish context (Berners-Lee 2009a). To date the combined challenge of Open Data have been met most readily by the governments of the United States and Great Britain, and in 2011 half of the data in the Linked Data cloud had been drawn from national governments (Berners-Lee 2010; Hendler 2011).

Perhaps unsurprisingly the majority of the released data are in the 'raw' state, coming as they do from "…a wide variety of source formats and collection methodologies, [which result] in idiosyncratic data representations" (Allemang & Hendler 2011, p.188). Once made available, this raw data can then be converted from, for example, a proprietary spreadsheet format into RDF, allowing it to be used in conjunction with other applications such as plotting data on maps or timelines, among various other types of data visualisation (ibid, p.195). Berners-Lee and others were also emphatic about relating Linked Data to the familiar virtues of online collaboration at grassroots level (Berners-Lee 2010).

Linked Data can be seen as an aspect of the Semantic Web whose value has been demonstrated both by the strength of the response to the Linked Open Data initiative, in addition to the tacit endorsement represented by its employment within the infrastructure of commercial systems such as the Google Knowledge Graph. However, as the latter case makes clear, the technological principles of Linked Data may well endure in a form different than that originally intended by those in the W3C and Semantic Web community at large. Hendler asserts that while the eventual realisation of a Semantic Web along the lines of the original vision remains an achievable yet precarious prospect, Linked Data will almost certainly abide as the core of a poor substitute "JSON-based external URI Web"[3] (Hendler 2011). This impoverished postulate is extrapolated from the status quo wherein simple RDF-based data (primarily directly-embedded RDFa) are held in traditional databases, with no semantic operations like inferencing or reasoning being performed on them (ibid).

To summarise, after a decade and a half of development, a key figurehead of the Semantic Web community has suggested that only one element of the original vision is

[3] The JavaScript Object Notation is a system used to store data (such as RDF triples) on the web as attribute-value pairs.

assured of its future survival, and it may endure only in a circumscribed form. It is Hendler's thesis that the profoundly problematic area of ontologies continues to stymie the advancement of the Semantic Web endeavour (ibid). This is not surprising, as ever since Berners-Lee and others began to outline the concept of a semantic web, observers have identified ontologies as posing the gravest threat to its successful realisation. This is due to the fact that the entirety of the Semantic Web, but most particularly the ontologies which it requires, are based on principles drawn from the field of knowledge representation (KR).

3.3 - Knowledge representation and Unreasonable Expectations

Knowledge representation (sometimes with the appended suffix "and reasoning") is a discipline within artificial intelligence research, and is defined in simple terms by Stuart Russell and Peter Norvig as: "…the study of how to put knowledge into a form that a computer can reason with" (Russell & Norvig 2010, p.16). They assert that the essential principle underpinning KR since its inception some five decades ago is that: "…it is useful to have a formal, explicit representation of the world and its workings and to be able to manipulate that representation with deductive processes" (ibid, p.19).

Thus the integral relationship between this established field of enquiry and the initial conception of the Semantic Web is plain to see. However, it should be noted that the canonical implementation of same outlined in the 2001 *Scientific American* article (as distinct from the more diffuse version discussed in *Weaving the Web*) is the result of collaboration between Berners-Lee, a software engineer by training, and a group of researchers including Hendler, whose backgrounds were in artificial intelligence and robotics (Hendler 2011). As Antoniou et al. assert, "In AI there is a long history of developing and using ontology languages. It is a foundation that Semantic Web research can build on" (Antoniou et al. 2012, p.12). This proves to be rather a telling phrase, insofar as the negotiation with this complex heritage has proven to be a profoundly complex problem for the Semantic Web community. It should be borne in mind that this has not been a unidirectional exchange; in fact, it could be argued that the contrary is the case.

As Peter Patel-Schneider, a KR practitioner closely involved in the development of the OWL standard has pointed out, for the greater part of the field's existence, the KR systems which have been developed were merely "academic toy-things", with no practical application to any real-world scenarios (Patel-Schneider 2006). This state of affairs came to an end with the inception of Semantic Web research, and the work undertaken by Hendler and others on an "Agent Markup Language" (DAML, the original prototype web ontology language which would form the basis of OWL), under the aegis of the US Defence Advance Projects Agency (DARPA) (DAML Working Group 2006). Patel-Schneider argues that from the perspective of a KR researcher, the Web can be conceived of as a robot of a type more appealingly tractable than those with which they might usually work, insofar as its "sensors" and "effectors" are services and applications rather than the video and audio inputs and motor outputs of conventional robotics (Patel-Schneider 2006). It is his contention that this difference makes error-handling far more straightforward, and that such factors, combined with the scale of the endeavour make the Semantic Web an area of great interest to formal representation researchers (ibid).

However, Patel-Schneider's characterisation of the scale of the Web as a compelling property as far as KR is concerned belies the major difficulties that result when an attempt is made to describe complex phenomena in a precise and unambiguous fashion. Hendler contends that the major problem within Semantic Web research is the inflexibility and byzantine complexity of the ontology systems which have been developed – primarily by KR researchers – for use on the new Web (Hendler 2011).

The ultimate outcome of this desire to create a language with sufficient expressive power to encapsulate all the meaning that might appear on the Web was the diverse OWL family of languages delineated in the first chapter. As was noted there, OWL Full, the superset of all the sundry variants, is too richly descriptive, and therefore potentially ambiguous to be used in any reasoning programs; the necessary computing operations would almost certainly continue in perpetuity. OWL DL was then introduced as a version of OWL Full with constraints that allowed it to be decidable, at the cost of all loss of some expressivity. As was alluded to in the first chapter, the challenge of designing these languages lies in the greater part with the fact that they are entirely based on traditional logical formalism, and set theory in particular. These systems are inevitably accompanied by a set of unresolved problems pertaining to contradiction, not to mention a great deal of ideological baggage. Much of the effort that went into the development of the ontologies intended for deployment in the Semantic Web was expended upon the engineering of workarounds in OWL to circumvent the pitfalls presented by such august logical contradictions as the liar's paradox[4] and Russell's paradox[5] (Patel-Schneider 2006).

3.4 - Conceptual Baggage

It is Hendler's judgement that OWL has thus far been of far greater value to those in the KR field than to those in Semantic Web research, as it is the first *de facto* standard for formal representation, with all the associated benefits for research interchange which result from interoperable systems (Hendler 2011). While he does not begrudge the KR community their success, he argues that their determination to persevere in striving for the ideal monolithic ontology system reflects a failure to come to grips with the impossibly diverse environment that is the Web, in all its discord (ibid). Hendler has been active in attempting to encourage such an engagement. In a 2007 editorial in *IEEE Intelligent Systems*, he contends that:

> For many AI researchers, [the] social part of the Web really is like the dark side of the moon. We're so used to thinking that "knowledge is power" that we fall into a slippery slope, more-is-better fallacy. If some expressivity is good, lots must be great, and in some cases this is correct. What we forget, however, is something that's become sort of a catch phrase in Semantic Web circles: "A little semantics goes a long way." In fact, I'm just now beginning to understand exactly how little is needed to go a long way on something as mind-bogglingly huge and unorganized as the Web. (Hendler 2007, p.3)

It would appear that this advice was not very broadly heeded within the KR community, given that in 2011 Hendler felt it necessary to identify excessively powerful yet inflexible ontologies as the primary obstacle on the path to the realisation of the original vision of the Semantic Web (Hendler 2011). He went so far as to suggest that the myriad restrictions imposed by these ontologies were a significant motivation behind the establishment of many Linked Data projects which sought to operate with as little recourse to ontologies as possible (ibid). It could thus be said that the Semantic Web community as a whole was in some sense split as a result of the methodological paradigm embodied in the design of OWL.

[4] e.g., "This statement is false".
[5] Is the set of all sets who are not members of themselves also a member of itself?

An especially unfortunate aspect of this situation is that it validates to an extent the profound scepticism expressed from several quarters about the feasibility of the Semantic Web. This can be attributed in part to the involvement of researchers from artificial intelligence; a field a sometimes held in ignominy within computing and related disciplines for perceived tendency toward hubris (Quitney Anderson & Rainie 2010, p.19) (Russell & Norvig 2010, p.24). The NYU media commentator Clay Shirky has provoked the ire of James Hendler, among others, by characterising the Semantic Web community as "a witness protection program for AI researchers" (Ray 2011). The Pew Research Center's 2010 survey *The Fate of the Semantic Web* is a collection of hypotheses concerning the likely status of the endeavour in 2020. Several of the respondents espouse the anti-AI viewpoint; Jonathon Grudin of Microsoft Research proclaims the Semantic Web to be: "…the latest incarnation of a long line of futile AI endeavours that have not succeeded and wouldn't do much of what is hoped for them even if they did reach fruition. This one I would not expect to see by 2030 either" (Quitney Anderson & Rainie 2010, p.19). His opinion can be said to broadly typify the views of those grouped under the rubric for those unconvinced by the track record of AI proponents (ibid).

As far back as the period when the original vision of the Semantic Web was being outlined, Tim Berners-Lee was at pains to explicitly clarify that the project most assuredly was *not* an attempt to create an artificial intelligence (Berners-Lee 1998). Being a pragmatic software engineer, Berners-Lee goes further, dismissing any such enterprise as "magical"; elsewhere he has chided those who fear the unintended consequences of reasoning programs as imagining an absurd scenario in which "…library cards begin composing music…" (ibid) (Berners-Lee & Fischetti 1999, p.197). This determination to foreclose any attempt to associate the Semantic Web with strong AI could be read as a response to the 1990s "AI winter", a period when the reputation of the field was considerably diminished by the failure of a large number of ambitious projects (Russell & Norvig 2010, p.24).

3.5 - Wisdom of the Original Vision Lost

One of the major counter-productive effects of the ontology problem in Semantic Web research is the lack of use of ontologies on the Web, in contrast to the widespread employment of low-level stack components such as RDF (Hendler 2011) (ibid, p.469). Russell and Norvig assert that utilisation of Semantic Web markup technologies was "…inversely proportional to representational complexity…Usage of sophisticated RDF and Owl ontologies is not yet widespread" (ibid). This is very much in accord with Hendler's aphorism about the power of simple semantic markup when used sparingly. For Hendler, this state of affairs is regrettable yet unsurprising, as was indicated above, he contends that the ontologies developed for the Web are too similar to traditional KR models for closed systems, with an excessive focus on expressivity (Hendler 2011). Although he was involved in the design of the first OWL standard, it is his judgement that *all* OWL variants, even including OWL Lite, are too "heavy", in other words, they are too powerfully expressive to be of practical use, a problem not resolved by the revised OWL2 (Dean & Schreiber 2009) (Hendler 2011).

Hendler's solution is to utilise very low-level, primitive ontologies such as RDFS with its classes, subclasses, and properties, as well as the newly developed RDFS-Plus (Allemang & Hendler 2011). RDFS-Plus is, rather confusingly, yet another subset of OWL which is intended to provide a balance of expressivity and user-friendliness (ibid, p.24). RDFS-Plus is not a W3C standard, according to Allemang and Hendler it was formulated with direct input from commercial organisations in order to serve their identified needs (ibid) (Hendler 2011). It should be noted that Hendler considers Linked Data projects aiming to operate with little or no ontology structure to be as misguided as the KR community's pursuit of a monolithic ontology (ibid). The former are another group which he believes would derive great benefit from the implementation of an RDFS-Plus model (ibid).

The ontology problem and its consequences are particularly disheartening given the fact that the prudent rationale expressed in Hendler's "a little semantics…" aphorism is equally present in Berners-Lee's initial conception of the Semantic Web. In *Weaving the Web*, Berners-Lee asserts that the limited nature of HTML as a language has been integral to the success and growth of the Web (Berners-Lee & Fischetti 1999, p.196). As such this adherence to a "principle of least power" has also underpinned the design of RDF, as is evident in its fundamental structure of subject-predicate-object, with each linked to a URI (ibid, p.197). From Berners-Lee's perspective it was far more important to provide the tools to allow people to represent as much information as possible using as simple a structure as possible, and then to interrelate it in as many ways as possible, without ever requiring it to conform exactly to a single centralised set of definitions, as in classical KR (Berners-Lee 1998).

In other words, the present difficulties stemming from powerful, inflexible ontologies is a situation Berners-Lee and others had anticipated and planned to avoid, as is clear from the *Scientific American* article:

> Semantic Web researchers, in contrast [to formal representation practitioners], accept that paradoxes and unanswerable questions are a price that must be paid to achieve versatility. We make the language for the rules as expressive as needed to allow the Web to reason as widely as desired. (Berners-Lee et al. 2001, p.38)

Thus it can be seen that the core group of originators of the Semantic Web were always prepared to deal with the polyglot Web in all its undecidable complexity, and as such were aware that no monolithic ontology was possible. As Hendler and Allemang assert: "The Semantic Web isn't about getting everyone to agree, but rather about coping in a world where not everyone will agree, and achieving some degree of interoperability nevertheless" (Allemang & Hendler 2011, p.9). However, as we have seen, mission creep on the part of the KR community involved in the project has ultimately precipitated the division identified by Hendler in his address. Furthermore, it has validated the criticisms made by Shirky and others, which had been unjustified when they were made originally.

3.6 - "Less semantic, more Web"

David Karger of MIT is a highly respected member of the Semantic Web community who has for some time been been ploughing his own furrow with respect to some of the orthodoxies of the research area. Like Berners-Lee and Hendler, he is sceptical of the appropriateness of highly complex ontologies to the context of the Web (Ray 2011). Similarly, he has ascribed the prevalence of such ideas within Semantic Web research to the influence of classical KR practices, although he has been more forthright than Hendler in drawing attention to same (Karger 2013).

Karger made the keynote address to the 2013 European Semantic Web Conference, in which he observed that the vast majority of the high-profile papers being delivered at the conference were devoted to the underlying technologies of the Semantic Web (ibid). Further to this, he remarks that all of the major topics, such as ontologies and inferencing, were ones which had seen "…decades of study within the artificial intelligence community" (ibid). This being the case, Karger asks:

> Isn't such work on knowledge representation and reasoning still going on in the AI community? Given the fundamental nature of these problems, does the fact that we are doing our inferences over web data rather than (say) an expert system knowledge base change the problem at all? And if there is nothing specific to the Semantic Web about this work, what is the value of partitioning it from the AI community? (ibid)

Karger has effectively restated Hendler's reservations about the type of work being undertaken in Semantic Web research, but in a far more pointed fashion. Karger is being purposefully disingenuous when questions whether inferencing on the Web should be different from inferencing in a closed system, like Hendler, he considers the distinction to be a fundamental one too little recognised (ibid;) (Allemang & Hendler 2011, p.335). For Karger, however, the usurpation of Semantic Web research problems by KR research problems is not the source of the field's ailments but is instead a symptom of the community's persistent failure to address the needs of end-users (Karger 2013).

Karger's assessment of the situation is persuasive, in that the vast bulk of academic Semantic Web research continues to be concerned with the infrastructure for a Web of structured data at the expense of prototyping end-user applications which would leverage same. Hendler, in his 2011 state-of-the-art lecture, places the design of Semantic Web HCI (Human-Computer Interaction) for "real users" among the research challenges which should be addressed in the future (Hendler 2011). By contrast, it is within this area that Karger believes the success or failure of the W3C vision of the Semantic Web will ultimately be determined: "We have to describe specific end-user problems and demonstrate specific Semantic Web applications that will solve those problems. If we fail to do that …someone else will solve those problems without using Semantic Web tools, and the Semantic Web will be left behind" (Karger 2013). The low level of public awareness about the Semantic Web may be ascribed in part to the dearth of such prototype applications which might be used to demonstrate its usefulness. Like Berners-Lee, Karger believes that the flourishing of the World Wide Web was primarily due to the fact that it

"…[made] it easy for everyone to author, manage and share information. It wasn't really about novel systems …it was about a novel arrangement of those pieces that empowered end users" (ibid).

3.7 - Conclusion

The question of user empowerment is indeed a crucial one for the Semantic Web. Linked Data is premised in part on the idea that a multitude of small contributions can collectively assemble a remarkably useful resource, either from scratch or by transferring extant datasets into RDF form. The Open Data element of same is in a sense an ethical project, in that it is underpinned by the conviction that individual citizens are entitled to engage with the data collected on their behalf by state bodies. By contrast, an ontology problem has emerged within Semantic Web research due to the reluctance of knowledge representation practitioners to adjust their approach in order to engage with the new paradigm demanded by the plurality of decentralised, discordant data found on the Web. This state of affairs is all the more unfortunate because the original vision of Berners-Lee et al. had accounted for the challenges of a diverse open system. As was seen above, the status quo is symptomatic of the somewhat insular nature of the Semantic Web community, whose focus has been on unsuitable research problems for many years, to the point that, as Karger warns, their failure to attend to end-user application design may well result in much of their endeavours being ignored and going to waste.

Conclusion - Without Which Not

This research paper consisted of an examination of the present status of the Semantic Web, considered in the light of the original vision for the project as outlined by Tim Berners-Lee, and to a lesser extent Jim Hendler. The first phase of this process necessarily involved an analysis of the vision as it was expressed in unadulterated form by Berners-Lee in his book *Weaving the Web*, and also in the more restricted terms employed in the *Scientific American* articled he co-authored with Hendler and Ora Lassila.

While the aspirations expressed by Berners-Lee are at times so optimistic as to appear positively utopian, it is essential to note that a pragmatism underpins the design decisions he advocates in the effort to realise the Semantic Web. The remaining sections of the first chapter then delineated the structure and essential functions of the rather byzantine software architecture which has been agreed upon by the W3C, and which is ultimately the tangible, committee-approved result of Berners-Lee's vision.

With much, but not all of the Semantic Web 'stack' in place, the second chapter consisted of an assessment of the degree to which the field of Semantic Web research must negotiate a relationship with the various corporate entities which dominate the Web at the present time. Particular focus was given to the area of social networking, and more still to recent activities by Google, neither of which were significant forces when the Semantic Web vision was being outlined. It was determined that corporations such as Facebook were predisposed to creating semantic markup which was bespoke, and therefore of limited usefulness to other parties without expending the effort required to translate like terms across systems. For their part, Google seem to be very reticent about being seen to engage directly with Semantic Web research, even when it is implementing what appear to be ideas from the field for their opaque knowledge base, the Knowledge Graph.

The third chapter entails a survey of the Semantic Web's most visible success, in addition to a consideration of its most imposing challenges. The largest successful initiative to emerge from Semantic Web research to date has been Linked Data, which was explored at some length, particularly with respect to its semi-detached relationship to its parent field. It should be noted that Linked Data can be viewed as an attempt to strip away the most contentious aspects of the Semantic Web stack, in particular the system of ontologies inherited from knowledge representation. The difficulties which stem from the descriptive power of the OWL family were found to present the most significant obstacle to the comprehensive realisation of the original vision of the Semantic Web, according to Jim Hendler. The excessive expressive power and inflexibility of OWL mean that it is unsuited to the role it is intended to play in the Semantic Web stack and as such alternative solutions must be found. OWL would appear to be unfit for purpose because its design adheres too closely to the traditional ontology models used in KR, which are intended to function in closed systems, and never encounter terms from outside their vocabulary, or which have different definitions. Unfortunately, this is precisely the type of scenario a *Web* Ontology Language would be expected to handle.

This unsatisfactory state of affairs grows ever more egregious when one considers the fact that Berners-Lee was cognisant from the early 1990s onwards of the fact that a traditional KR approach to ontologies would be inadequate for dealing with the full polysemy of human discourse which runs through the Web. In his original vision, the optimal means by which to represent complexity was one involving a multitude of very simple components, a lower, RDF-level approach would be a the solution to the task of describing the world:

> The total Web of all the data from all the applications of RDF will make a very complex world in which it will be possible to ask unanswerable questions. That is how the world is. The existence of such questions will not stop the world from turning, or cause weird things to happen to traffic lights. But it will open the door to some very interesting new applications that do roam over the whole incalculable, intractable Web and, while not promising anything, deliver a lot. (Berners-Lee & Fischetti 1999, p.198)

This passage serves to expose the gulf in understanding between Berners-Lee's pragmatic view of the limits of formal logic, and the positivistic assumptions of those who conform to what George Lakoff has called the "objectivist paradigm", here in the form of the model theoretic semantics which form the basis of the knowledge representation program (Lakoff 1987, pp.196, 207; Patel-Schneider 2006). Berners-Lee's reference to unanswerable questions is particularly telling as it calls to mind the early Ludwig Wittgenstein's declaration that there can be no unanswerable questions because: "when the answer cannot be put into words, neither can the question be put into words" (Wittgenstein 1922, p.88). Invoking the grand claims advanced about the power of formal languages by august figures from the KR field such as John Sowa, Florian Cramer has cautioned that: "The history of computing is rich with confusions of formal with common human languages, and false hopes and promises that formal languages would become more like human languages" (Sowa 2000, p.420) (Cramer 2008, p.171). In this way it is possible to see the intrinsic limitations of formal logic as a system, but also gain a sense of how the pragmatic nature of Berners-Lee's original vision of low-complexity knowledge representation on the Semantic Web became subsumed by KR practitioners convinced of the indispensability of powerfully expressive languages in any given scenario.

The other unfortunate effect of the rather fruitless focus on KR within Semantic Web research is that it has caused those within the community to neglect to consider the position of the end-users of the Semantic Web, and develop some prototype applications for them. This veritable dereliction of duty by Semantic Web academics, identified by David Karger, has, he argues led to a vacuum which will fill rapidly with inferior, presumably commercially-developed semantic applications (Karger 2013). Herein lies another instance where Berners-Lee's original vision, in this case of an empowered Web user, is rather at variance with the actuality of the present situation. As corporate control over the online space continues to grow and sector-based monopolies become further entrenched, commercial, quasi-semantic applications with all the limitations and consumer lock-in such systems typically involve may be the only options available to end-users in the future. In order to ameliorate such a scenario the Semantic Web research community must re-familiarise itself with Tim Berners-Lee's principled, pragmatic original vision and renew the field accordingly.

List of Abbreviations

AI - Artificial Intelligence

CIA - Central Intelligence Agency

DAML - DARPA Agent Markup Language

DARPA - Defence Advance Projects Agency

DL - Description Logic

FOAF - Friend of a Friend

GR - Good Relations

HCI - Human-Computer Interaction

HTML - Hypertext Markup Language

HTTP - Hypertext Transfer Protocol

JSON - JavaScript Object Notation

KR - Knowledge Representation

MIT - Massachusetts Institute of Technology

NYU - New York University

OGP - Open Graph Protocol

OWL - Web Ontology Language

RDF - Resource Description Framework

RDFa - Resource Description Framework in Attributes

RDFS - Resource Description Framework-Schema

RIF - Rule Interchange Format

SGML - Standard Generalised Markup Language

SPARQL - SPARQL Protocol and RDF Query Language

SQL - Structured Query Language

SWRL - Semantic Web Rule Language

TED - Technology Entertainment Design

URI - Universal Resource Locator

W3C - World Wide Web Consortium

XML - Extensible Markup Language

Bibliography

Allemang, D., Hendler, J.A. (2011) *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*, 2nd ed. ed, Morgan Kaufmann/Elsevier: Waltham, MA.

Antoniou, G., Groth, P., van Harmelen, F., Hoekstra, R. (2012) *A Semantic Web Primer*, Third Edition. ed, Cooperative Information Systems, MIT Press: Cambridge, MA.

Berners-Lee, T. (1994) Plenary at WWW Geneva 94 [online], *www.w3.org*, available: http://www.w3.org/Talks/WWW94Tim/ [accessed 9 Feb 2014].

Berners-Lee, T. (1998) 'What the Semantic Web can represent', *Design Issues*, available: http://www.w3.org/DesignIssues/RDFnot.html [accessed 26 Feb 2014].

Berners-Lee, T. (2009a) Linked Data [online], *Design Issues*, available: http://www.w3.org/DesignIssues/LinkedData.html [accessed 23 Feb 2014].

Berners-Lee, T. (2009b) *The next Web of Open, Linked Data* [online], available: http://www.youtube.com/watch?v=OM6XIICm_qo&feature=youtube_gdata_player [accessed 23 Feb 2014].

Berners-Lee, T. (2010) *The Year Open Data Went Worldwide* [online], available: http://www.youtube.com/watch?v=3YcZ3Zqk0a8&feature=youtube_gdata_player [accessed 24 Feb 2014].

Berners-Lee, T., Fischetti, M. (1999) *Weaving the Web: The Past, Present, and Future of the World Wide Web by Its Inventor*, Orion Business: London.

Berners-Lee, T., Hendler, J. (2010) 'From the Semantic Web to social machines: A research challenge for AI on the World Wide Web', *Artificial Intelligence*, 174(2), 156–161.

Berners-Lee, T., Hendler, J., Lassila, O. (2001) 'The Semantic Web', *Scientific American*, 284(5), 28–37.

Cramer, F. (2008) 'Language', in Fuller, M., ed., *Software Studies: A Lexicon*, Leonardo Books, MIT: Cambridge, Mass. ; London, 168–174.

DAML Working Group (2006) About DAML [online], *DAML.org*, available: http://www.daml.org/about.html [accessed 25 Feb 2014].

Dean, M., Schreiber, G. (Eds.) (2009) 'OWL Web Ontology Language Reference', available: http://www.w3.org/TR/owl-ref/ [accessed 13 Feb 2014].

DiNucci, D. (1999) 'Fragmented Future', *Print*, 53(4).

Filloux, F. (2012) 'Does Google really need news media content?', *The Guardian*, 5 Nov, available: http://www.theguardian.com/media/2012/nov/05/google-news-media-content [accessed 20 Feb 2014].

Gibbs, S. (2013) ‘Google introduces the biggest algorithm change in three years’, *The Guardian*, 27 Sep, available: http://www.theguardian.com/technology/2013/sep/27/google-biggest-algorithm-change-hummingbird [accessed 20 Feb 2014].

Hendler, J. (2007) ‘The Dark Side of the Semantic Web’, *IEEE Intelligent Systems*, 22(1).

Hendler, J. (2011) *Why the Semantic Web Will Never Work* [online], available: http://www.youtube.com/watch?v=oKiXpO2rbJM&feature=youtube_gdata_player [accessed 8 Feb 2014].

Johnston, J. (2008) *The Allure of Machinic Life: Cybernetics, Artificial Life, and the New AI*, MIT: Cambridge, Mass. ; London.

Karger, D. (2013) ‘Keynote at ESWC Part 3: What’s Wrong with Semantic Web Research, and Some Ideas to Fix it’, *Haystack Blog*, available: http://haystack.csail.mit.edu/blog/2013/06/10/keynote-at-eswc-part-3-whats-wrong-with-semantic-web-research-and-some-ideas-to-fix-it/ [accessed 26 Feb 2014].

Lakoff, G. (1987) *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, University of Chicago Press: Chicago ; London.

McGuinness, D.L., van Harmelen, F. (Eds.) (2009) ‘OWL Web Ontology Language Overview’, available: http://www.w3.org/TR/2004/REC-owl-features-20040210/ [accessed 13 Feb 2014].

Menzel, J. (2010) ‘Deeper understanding with Metaweb’, *Google Official Blog*, available: http://googleblog.blogspot.com/2010/07/deeper-understanding-with-metaweb.html [accessed 20 Feb 2014].

Miller, E., Manola, F. (Eds.) (2004) ‘RDF Primer’, available: http://www.w3.org/TR/rdf-primer/ [accessed 13 Feb 2014].

Needleman, R. (2012) Google Bringing New Smarts to Search with Knowledge Graph [online], *CNET*, available: http://news.cnet.com/8301-1023_3-57435114-93/google-bringing-new-smarts-to-search-with-knowledge-graph/ [accessed 20 Feb 2014].

Patel-Schneider, P. (2006) *Knowledge Representation and the Semantic Web* [online], available:

http://www.youtube.com/watch?v=rAkSY5Ha9vk&feature=youtube_gdata_player [accessed 25 Feb 2014].

Quitney Anderson, J., Rainie, L. (2010) *The Fate of the Semantic Web*, Pew Research Center's Internet & American Life Project, Pew Research Center, Washington, D.C.

Ray, K. (2011) *Web 3.0* [online], available: http://www.youtube.com/watch?v=bd8zR0v7Jts&feature=youtube_gdata_player [accessed 25 Feb 2014].

Russell, S.J., Norvig, P. (2010) *Artificial Intelligence: A Modern Approach*, 3rd ed. ed, Prentice Hall series in artificial intelligence, Prentice Hall: Upper Saddle River, N.J.

Schmidt, D. (2010) 'The inadequacy of embedded markup for cultural heritage texts', *Literary and Linguistic Computing*, 25(3), 337–356.

Shirky, C. (2003) The Semantic Web, Syllogism, and Worldview [online], *Clay Shirky's Writings About the Internet*, available: http://www.shirky.com/writings/semantic_syllogism.html [accessed 19 Feb 2014].

Singhal, A. (2012) 'Introducing the Knowledge Graph: things, not strings', *Inside Search: The Official Google Search Blog*, available: http://insidesearch.blogspot.ie/2012/05/introducing-knowledge-graph-things-not.html [accessed 20 Feb 2014].

Sowa, J.F. (2000) *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks/Cole: Pacific Grove ; London.

Stewart, D. (2012) Google's Knowledge Graph: Yeah, That's the Semantic Web (sort Of) [online], *Gartner.*, available: http://blogs.gartner.com/darin-stewart/2012/05/17/googles-knowledge-graph-yeah-thats-the-semantic-web-sort-of/ [accessed 21 Feb 2014].

Walpole, R. (2013) 'The National Archives Digital Records Infrastructure Catalogue: First Steps to Creating a Semantic Digital Archive', in *XML London 2013 – Conference Proceedings*, Presented at the XML London 2013, XML London: London, 87–101, available: http://xmllondon.com/2013/xmllondon-2013-proceedings.pdf#page=113 [accessed 19 Feb 2014].

Wikipedia Editors (2014) 'Semantic Web Stack', *Wikipedia, the free encyclopedia*, available:

https://en.wikipedia.org/w/index.php?title=Semantic_Web_Stack&oldid=578539977 [accessed 13 Feb 2014].

Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*, Routledge classics, Routledge: London.

Zaino, J. (2012a) Google Knowledge Graph Interview [online], *Semanticweb.com*, available: http://semanticweb.com/google-knowledge-graph-interview_b29172 [accessed 20 Feb 2014].

Zaino, J. (2012b) Get In On CrowdSourcing An Open Knowledge Graph API [online], *Semanticweb.com*, available: http://semanticweb.com/get-in-on-crowdsourcing-an-open-knowledge-graph-api_b31438 [accessed 21 Feb 2014].