

## INSIDE

April 2008

### 1 WWW2008

Tom Heath looks at the importance of Linked Data, and outlines key issues for discussion in Beijing.

### 4 THE VALUE OF WEB 3.0

Mills Davis of Project10X asks how Web 3.0 differs from the previous phases of the Web.

### 6 PODCAST ROUNDUP

News of the latest Semantic Web podcasts from Talis.

### 7 TALIS ENGAGE

Nadeem Shabir reports on development work that delivered a commercially viable enterprise application with the help of the Talis Platform.

### 9 SIR TIM BERNERS-LEE TALKS WITH TALIS ABOUT THE SEMANTIC WEB

The full transcript of this recent interview with the Father of the Web.

### 19 THE SEMANTIC WEB GANG

Experienced Semantic Web practitioners debate the readiness of this technology for mainstream corporate adoption.

## Looking ahead to Linked Data on the Web

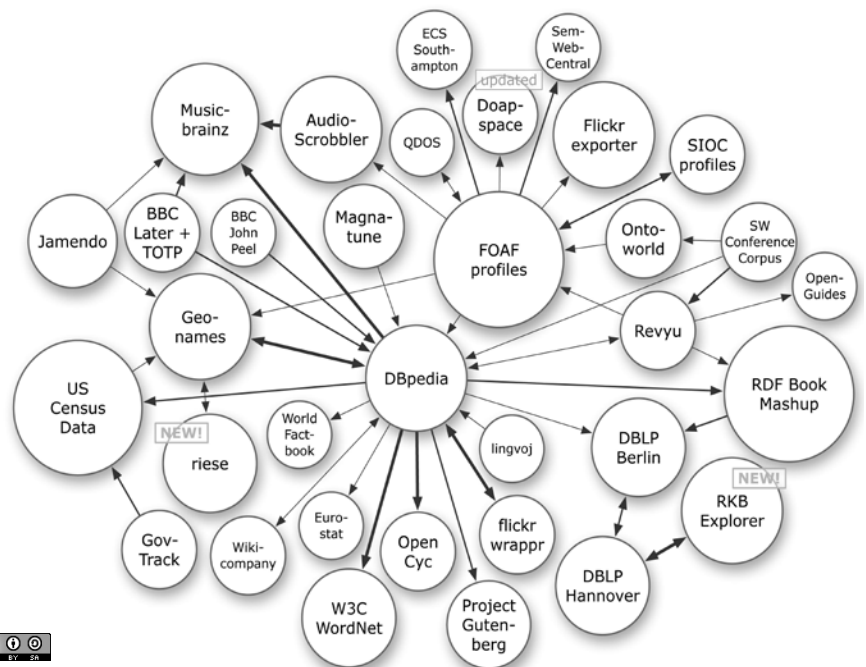


Image provided by Richard Cyganiak

**Ahead of the Linked Data on the Web workshop at this year's World Wide Web Conference in Beijing, Tom Heath introduces the concept of Linked Data and outlines some of the key issues that workshop participants will be discussing.**

Linked Data is a style of publishing data on the Web that emphasises data reuse and connections between

related data sources. To understand the value of this proposition, imagine a traditional Web site with no incoming or outgoing links. How much value would such a Web site generate for its owner, compared to another that was identical in content but richly interconnected with the Web at large?

*continued on page 3*

## SUBSCRIBE NOW

Nodalities Magazine is made available, free of charge, in print and online. If you wish to subscribe, please visit [www.talis.com/nodalities](http://www.talis.com/nodalities) or email [nodalities-magazine@talis.com](mailto:nodalities-magazine@talis.com).

## MEET US AT

Talis staff will be presenting and attending several events over the next few months, including;

### WWW2008

Beijing, China 21-25 April 2008

### XTech 2008

Dublin, Ireland 6-9 May 2008

### 2008 Semantic Technology Conference

San Jose, California 18-22 May 2008

### 5th European Semantic Web Conference

Tenerife, Spain 1-5 June 2008

### LinkedData Planet Conference 2008

New York City 17-18 June 2008

For further information visit: [www.talis.com/platform/events](http://www.talis.com/platform/events)



Nodalities Magazine is a Talis Publication

ISSN 1757-2592

Talis Group Limited  
Knights Court, Solihull Parkway  
Birmingham Business Park B37 7YB  
United Kingdom  
Telephone: +44 (0)870 400 5000

[www.talis.com/platform](http://www.talis.com/platform)



The views expressed in this magazine are those of the contributors for which Talis Group Limited accepts no responsibility. Readers should take appropriate professional advice before acting on any issue raised.

©2008 Talis Group Limited. All rights reserved. Talis, the Talis logo, Nodalities and the Nodalities logo where shown are either registered trademarks or trademarks of Talis Group Limited or its licensors in the United Kingdom and/or other countries. Other companies and products mentioned may be the trademarks of their respective owners.

## EDITOR'S NOTES



Welcome to this first issue of Nodalities Magazine, which I hope you will enjoy. Don't forget to visit [www.talis.com/nodalities](http://www.talis.com/nodalities) and sign up to receive subsequent issues by post.

Today's Web of Documents is only the first step in a vision of far greater richness and utility; a vision that anticipates meaningful links between discrete pieces of information currently locked up within the body of those web documents or hidden away from view altogether in a database. In this long-held vision of the Semantic Web, the meaning (or semantics) of the link between resources becomes explicit, as does much of the structure implicit to those documents and databases. On p.4, Mills Davis of Project 10X picks some key trends from his recent Semantic Wave report, and uses these to illustrate the opportunities created by application of Semantic Web ideas.

As Sir Tim Berners-Lee noted in a recent conversation (see the full transcript from p.9), this vision of the Semantic Web has reached a point at which it really is ready for commercial use. In the first episode of the new Semantic Web Gang monthly podcast (see p.19 for a transcript of this episode), some of those at the forefront of efforts to build businesses around the Semantic Web share their interpretations of Berners-Lee's assertion. On p.7, Nadeem Shabir illustrates the ease with which one particular application was built from scratch on top of Talis' Semantic Web Platform; an application that paying customers are using to deliver some of their core services.

With emphasis shifting from developing underlying standards and specifications toward using them, Linking Open Data project offers a valuable illustration of that which is possible. On p. 3 Tom Heath reports on the project's progress over its first year, ahead of the Linking Open Data workshop that he will be co-chairing at this year's World Wide Web Conference in Beijing later this month.

With Nodalities Magazine we intend to provide a regular and accessible round-up of key trends and issues affecting those who are putting the Semantic Web to work in the mainstream. We will draw upon work within Talis, as well as industry trends surfaced by the monthly Semantic Web Gang podcasts, Danny Ayers' weekly Semantic Web round-ups, and business news and analysis on ZDNet's new Semantic Web blog. I also welcome suggestions of content and contributors for future issues.

Thank you for reading Nodalities Magazine. I hope that you will subscribe in order to receive subsequent issues by post, and I hope that you will engage in conversation to make this magazine yours, and to help nurture the Semantic Web community that we are here to report, inform, and grow.

[paul.miller@talis.com](mailto:paul.miller@talis.com)



*Continued from front page.*

As a publishing medium, the Web enables linking to related content rather than duplication. Not only does

this allow site owners to add value to content without incurring additional data management costs, but these links also provide a means for discovery, such that quality content adequately linked can generate value for its publisher. Linked Data takes these ideas and applies them not to “content” in the traditional Web sense, but to data. By following the principles of Linked Data, anyone can publish their data in the Web of Data, or Semantic Web, and stand to benefit from the value that connections add.

The core principles of Linked Data were first introduced by Sir Tim Berners-Lee in 2006, and have since been adopted by a growing number of data publishers who recognise that the value of their data is not just in the data itself, but in the connections it has to other data sources. The focal point for these efforts has been the grassroots Linking Open Data project, that has succeeded in publishing billions of pieces of interlinked data, covering domains as broad as music, films, locations, hotels, books and reviews.

In a previous issue of the Talis Platform Newsletter I wrote about the significance of the Linking Open Data project in bootstrapping the Web of Data, and took a historical look at what the project has achieved in its first year of existence ([tinyurl.com/4b2fu2](http://tinyurl.com/4b2fu2)). As this first issue of Nodalities Magazine goes to press, the Linked Data community is gearing up for the first formal event dedicated to the subject - the Linked Data on the Web workshop (LDOW2008). In the remainder of this piece I'll look ahead to the workshop, provide a brief overview of the state of the art work being presented at the event, and situate this in the broader context of the ongoing Linked Data research and deployment agenda.

The Linked Data on the Web workshop will take place on April 22nd in Beijing,

as part the 17th International World Wide Web Conference (WWW2008). As one of the organisers I'm delighted with the range and quality of the papers that will be presented at the workshop and published in the accompanying proceedings. In all, 23 papers were accepted from 37 submissions. These are made up of a mix of long, short and demo papers, which together balance the more theoretical side of Linked Data research with demonstrations of practical applications emerging in the field. The programme of the workshop is organised around the themes of publishing and consuming Linked Data, supported by an additional session related to architectural and conceptual issues.

Publishing Linked Data is the largest of these three themes, in terms of number of papers. This reflects the priorities of the community over the last twelve months, during which time initiatives such as the Linking Open Data project have emphasised demonstrable bootstrapping of the Web of Data. Grouped under this theme are a number of papers that highlight how existing data sets held within social media and photo sharing sites, library catalogues, enterprise and administrative databases, institutional repositories and even plain HTML documents can be exposed on the Web according to Linked Data principles. A number of authors use this opportunity to share and demonstrate their experiences of Linked Data generation and publishing - valuable knowledge which will no doubt benefit those who follow. These practical experiences are complemented by papers that address publishing issues such as provenance and licensing that are more theoretical in nature but no less important as the community moves forward.

Searching and browsing are often seen as the two primary modes of interacting with the traditional Web of Documents, and this distinction can be seen in the papers that address consumption of Linked Data. The Linked Data browsers that will be presented take a range of forms, from add-in widgets for Web pages, to location-enabled applications

for mobile phones, to desktop browsers that can also publish back to the Web of Data.

Supporting the practically-oriented aspects of the workshop programme are a number of papers addressing more architectural and conceptual issues of Linked Data. As is to be expected given the importance of the issue, the process of assigning identifiers to items described in data, and the nature of these identifiers themselves, is discussed in a number of papers. Methods for interlinking music data sets are reported, as are techniques for linking stable URIs to potentially ambiguous keyword tags. Lastly the question of whether social processes can be harnessed to aid data publishers in selecting commonly used, and therefore desirable, elements for describing their data is raised, and a solution proposed by which this may be achieved.

As I hope this account demonstrates, the world of Linked Data has developed significantly in the last twelve months. In addition to the increased availability of data, we're also seeing ever growing numbers of applications being developed and deployed that are predicated on the existence of this Web of Data, and specifically designed to exploit it. I expect these trends to continue apace in the coming year, and gain significantly in momentum. The primary question for data publishers and consumers is no longer *whether* to get involved in this next generation of the Web, but simply *when*.

*Tom Heath is a Researcher in the Talis Platform Division, and one of the organisers of the LDOW2008 workshop.*

**[tom.heath@talis.com](mailto:tom.heath@talis.com)**

[www2008.org](http://www2008.org)  
[events.linkedata.org/ldow2008/](http://events.linkedata.org/ldow2008/)  
[tinyurl.com/2vsdnr](http://tinyurl.com/2vsdnr)

Photo of Tom Heath,  
 courtesy of Gregory Todd Williams, [kasei.us](http://kasei.us)

# The Value of Web 3.0

*Mills Davis recently published **Semantic Wave 2008**, a comprehensive introduction to semantic technologies and the business opportunities that they represent. In this article, Mills asks how Web 3.0 differs from previous phases of the Web.*

**By Mills Davis**



How is Web 3.0 different from previous stages of internet evolution? It employs "knowledge computing" to drive new value creation and solve problems of scale and complexity.

The basic shift occurring in Web 3.0 is from information-centric to knowledge-centric patterns of computing. Web 3.0 will enable people and machines to connect, evolve, share, and use knowledge on an unprecedented scale and in new ways that make our experience of the internet better.

Web growth continues to accelerate. Dimensions of net expansion include communications bandwidth, numbers of people connected, numbers and kinds of devices that are IP-aware, numbers of systems and applications, quantities of information, and types of media. As the internet expands, needs world-wide are outstripping the capacities and capabilities of current information and communications technologies (ICT) and architectures. Information-centric patterns of computing have reached the limit of what they can provide to cope with problems of scale, complexity, security, mobility, rich media interaction, and autonomic behavior.

Web 3.0 will solve these problems and lay a foundation for the coming ubiquitous Web of connected intelligences. The Web 3.0 solution, simply put, is to give the internet a knowledge space. In the following topics we identify key characteristics of this knowledge space, sketch out how its semantic computing works, and examine how Web 3.0 knowledge-centric patterns of computing drive new value creation.

The key notion of semantic technology is to represent meanings and knowledge (e.g., knowledge of something, knowledge about something, and knowledge how to do something, etc.) separately from content or behavior artifacts, in a digital form that both people and machines can access and interpret. As a platform, Web 3.0 will embrace all semantic technologies and open standards that can be applied on top of the current Web. It is not restricted just to current Semantic Web standards.

Knowledge exists in many forms in today's Web. All computing processes represent some type of knowledge in some way in order to process information, for example: knowledge about how information is organized in order to search it; rules that tell a computer program how to make a decision; or action steps to take to complete a task.

The problem is that existing knowledge on the Web is fragmented and difficult to connect. It is locked in data silos and operating system file system formats. Knowledge is hidden in object-oriented black boxes and layers of stack architecture. It is embedded in program code and squirreled away in proprietary algorithms.

Web 3.0 changes this. The convergence of pattern discovery, deep linguistics, and ontological symbolic reasoning technologies make it feasible to automatically extract embedded and intrinsic knowledge from today's Web. Evolution of semantic social computing will enable communities to create, curate, and share knowledge in human readable and machine executable forms.

Web 3.0 taps new sources of value by modeling knowledge, adding intelligence, and enabling learning. Semantic technologies have the potential to drive 2-3 order of magnitude improvements in capabilities and life cycle economics through cost reductions, improved efficiencies, gains in effectiveness, and new functionalities that were not possible or economically feasible before now. New sources of value include:

## Value from knowledge modeling

Semantic models are sharable, recombinant, and executable. To model first, then execute the knowledge reduces time, risk, and cost to develop and evolve services and capabilities. Semantic model-based approaches achieve added development economies through use of (a) shared knowledge models as building blocks, (b) autonomic software techniques (goal-oriented software with self-diagnostic and self-management capabilities such as self-configuration, self-adaptation, self-optimization, etc.), and (c) end-user and do-it-yourself life-cycle development methodologies (rather than requiring intervention by IT professionals). Knowledge that is sharable, revisable, and executable is key to applications where facts, concepts, circumstances, and context are changing and dynamic.

## Value from adding intelligence

A working definition of intelligence is the ability to acquire, through experience, knowledge and models of the world (including other entities and self), and use them productively to solve novel problems and deal successfully with unanticipated circumstances. A key new source of value is adding intelligence to the user interface, to applications, and

to infrastructure. An intelligent system or agent is a software program that learns, cooperates, and acts autonomously. It is autonomic and capable of flexible, purposeful reasoning action in pursuit of one of more goals. An intelligent user interface (UI) knows about a variety of things such as system functionality, tasks users might want to do, ways information might be presented or provisioned. Intelligent UIs know about the user (via user models), which enables tailoring system behavior and communications. Adding intelligence helps users perform tasks, while making working with the computer more helpful, and as invisible as possible. As a result, systems do more for the user, yield more relevant results with less effort, provide more helpful information and interaction, and deliver a more enjoyable user experience. Adding intelligence can produce ten-fold gains in communication effectiveness, service delivery, user productivity, and user satisfaction.

## Value from learning

Machine learning is the ability of computers to acquire new knowledge from past cases, experience, exploration, and user input. Systems that learn increase in value during their lifetime. Their performance improves. They get better with use, and with scale. In addition to new or improved capabilities, systems that learn during operation may improve system life cycle economics by (a) requiring less frequent upgrading or replacement of core software components, and (b) enabling new incremental extensions to revenue models through add-on knowledgeware and software-as-a-service.

## Value from semantic ecosystem

An ecosystem is a self-sustaining system whose members benefit from each other's participation via symbiotic relationships (positive sum relationships). Principle drivers for semantic infrastructure and ecosystem include the economics of mobility, scale, complexity, security, interoperability, and dynamic change across networks, systems, and information sources. These

problems are intractable at Web scale without semantics. The corollary is the need to minimize human labor needed to build, configure, and maintain ultra-scale, dynamic infrastructure.

Semantic ecosystems that emerge in Web 3.0 will consist of dynamic, evolvable systems consisting of ensembles (societies) of smart artifacts. This means a shift in design focus from static, performance-driven design to: (a) design for robustness & resilience; (b) design for uncertainties; (c) design for distributed, autonomous pervasive adaptation; (d) design for organically growing systems; and (e) design for creating self-evolving services.

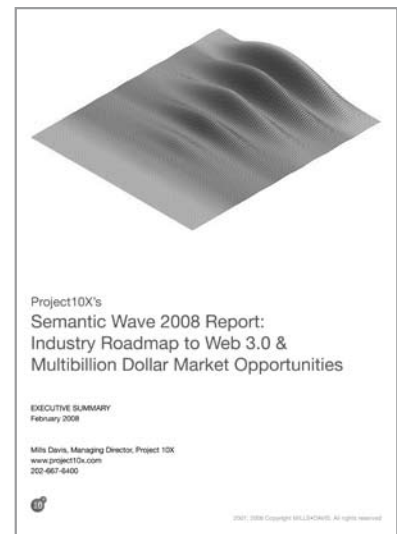
Current systems including the internet are designed to operate with predefined parameters. Change spells trouble. Mobility is a problem. Semantic ecosystems, however, will be future-proof, able to grow dynamically, evolve, adapt, self-organize, and self-protect. Web 3.0 will lay the foundations for an ubiquitous Web that includes autonomic intellectual property, Web-scale security and identity management, and global micro-commerce in knowledge-based assets. The value vector for semantic infrastructure is 2-4 orders of magnitude gains in capability, performance, and life cycle economics at Web scale.

**mdavis@project10x.com**

*Mills Davis is Founder and Managing Director of Project10X — a research consultancy specialising in next wave semantic technologies and solutions.*

[www.project10x.com/](http://www.project10x.com/)  
[www.semanticexchange.com/](http://www.semanticexchange.com/)

A summary of the Semantic Wave report can be downloaded from  
[www.project10x.com/](http://www.project10x.com/)





# Talis and the Semantic Web

*Talis CTO Ian Davis explains why Talis see the Semantic Web as a significant growth opportunity, deserving of their ongoing investment.*

By Ian Davis



We think the Semantic Web has a huge potential to radically change our society and economy and we believe that those applications and

services that understand the nature of data and its relationships will make the greatest difference. The World Wide Web demonstrated the power of simple links between documents, creating the most successful information system in history. The Semantic Web builds on this success, reusing the same notions of linking but greatly enhancing its utility by adding meaning to the links. This means that the Web can move beyond free text to structured data and the potential for data integration at an unprecedented scale.

The Semantic Web technologies are reducing the costs of providing, consuming and processing data and unlock business models and activities that would not otherwise be viable. At the same time the networking of this data, using the ubiquitous technologies of the World Wide Web, is generating new value. The value of data now comes not only from its intrinsic nature but also from its relation to other data

in the network. This extrinsic value could potentially be far greater than the existing intrinsic value. We believe that this change in the economics of data will create a new disruptive wave of innovation and growth.

Talis is creating key enabling infrastructure for the Semantic Web to help unlock this wave. The Talis Platform is an application platform delivered Software as a Service and enables any software developer to create data rich applications with minimal upfront investment, reducing risk and time to market. Developers using the Platform can spend more of their time building extraordinary applications and less of their time worrying about how they will scale their data storage and processing.

While the barriers to participating in the Semantic Web are low, there are still difficulties in dealing with the scale and variety of data that is available. Our objective is to dramatically lower the complexity and costs of storing, indexing, searching and augmenting large quantities of data so that every application can participate in the Semantic Web. Our Platform provides a range of services to suit all levels

of developer needs, from simple updates and free text searching to deep structured querying using standards like SPARQL. In the spirit of shared innovation, our developer community is open and free to anyone who wishes to develop with the Talis Platform.

Talis is already using the Platform to transform our own existing products and to enable rapid development of our new generation of semantically rich applications and services. The first of these applications are already achieving commercial success, as well as generating structured data for other Semantic Web applications to reuse, increasing the value of the data for all.

*Ian Davis is Chief Technology Officer at Talis, responsible for the ongoing development of the Talis Platform.*

**ian.davis@talis.com**

Find out more about the Talis Platform:

**Businesses:**

[www.talis.com/platform](http://www.talis.com/platform)  
[blogs.talis.com/nodalities](http://blogs.talis.com/nodalities)

**Developers:**

[n2.talis.com/](http://n2.talis.com/)  
[blogs.talis.com/n2](http://blogs.talis.com/n2)

## Podcast roundup

A round up of the latest podcasts available at [www.talis.com/platform/podcasts](http://www.talis.com/platform/podcasts)

Talis routinely records conversations with leading proponents of the Semantic Web, and makes them available online in podcast form. These include Danny Ayers' technical conversations with those at the coalface, as well as Paul Miller's executive briefings with the leaders of companies putting the Semantic Web to work. Recently, we also began monthly round-table discussions with the Semantic Web Gang, and a transcript from the first of these conversations can be found on p. 19.

Recent conversations that may be of interest include Sir Tim Berners-Lee (see p. 9 for a full transcript), Jim Hendler on the Semantic Web and Artificial Intelligence, Barak Pridor on Reuters' work around OpenCalais, and Richard Cyganiak on Sindice and Linked Data.

# Talis Engage: Delivering Community Information with a little help from the Semantic Web

*Nadeem Shabir describes the way in which a Semantic Web application platform and Agile development methods enabled a small team at Talis to rapidly design, build and deliver a new application which is now selling to enterprise customers.*

**By Nadeem Shabir**



*'I am of the opinion that my life belongs to the whole community and as long as I live, it is my privilege to do for it whatever I can.*

*I want to be thoroughly used up when I die, for the harder I work the more I live.' - George Bernard Shaw.*

A little over eighteen months ago we embarked on what we felt was an interesting challenge - to build an online community information solution that would allow communities to organise, create and publish details about local events, history, organisations, and groups.

Having already built solutions in this space we recognised an opportunity to build a next-generation solution. This would be sensitive to changes within the business domain with the growing emergence of community engagement and empowerment agendas.

This leveraged ideas and requirements gleaned from actively following local and national initiatives aimed at strengthening communities that attempted to find ways to facilitate wider community engagement.

In an attempt to validate some of our assumptions and ideas analysts from Talis took part in community engagement courses and workshops. Our analysts also worked with our existing customers and other advisory groups to get an appreciation of what it was that made communities engage with each other, as well as what it was that made it difficult for communities to engage with each other and their regional governments

through intermediaries such as their local library. We felt that understanding this would help us to build a community information solution that truly supported the communities in which it was deployed.

Once we identified the need and scope for what we wanted to deliver, we embarked on an equally innovative approach to developing this solution. At this point we could, like many software companies, have taken the requirements we had identified and then disappeared off into the ether for twelve months, to return with a community information product that we 'thought' satisfied the needs of local authorities and their users. However, in my experience, that rarely works and what you often develop might meet some of the business needs but it is rarely a solution that compels or delights users.

The approach we took was to develop Talis Engage one iteration at a time. We adopted and embraced the Agile Development philosophy and specifically the SCRUM process - we then chose a number of local authorities to work with from around the United Kingdom.

Representatives from local libraries from those areas formed a group of Advisory Customers to whom we demonstrated the product at the end of every four week iteration. The individuals were given time to use the software, comment on it, tell us what they liked as well as what they disliked.

This continuous feedback meant that over each iteration the product evolved into an extremely rich and compelling

community information solution which in real terms represents not only a product of the engineering talent at Talis but it also represents and embodies the ideas and vision of a number of local libraries around the country which the solution would be deployed into.

With Talis Engage the community owns its information. All of it. Library staff, local authorities and citizens are all empowered to suggest changes to information and contribute to its maintenance. The net effect of which is that this encourages communities to manage their own resources and promote the visibility of local resources.

*'The sign of success in social software is when your community does something you didn't expect.'* - Anil Dash.

Instead of describing Talis Engage in terms of a detailed technology stack a better way to understand what the product is might be by describing some of its key features or attributes.

Firstly as a solution Talis Engage is hosted at our data centres and is deployed and delivered using the Software as a Service model. This immediately lowers the cost of ownership for our customers - they don't incur the installation, hardware maintenance or licensing costs that traditionally come with purchasing solutions which customers then have to install themselves.

This also means that our customers pay a subscription to use this service and so they receive continuous product updates and improvements at no additional cost.

We can continue to receive feedback from them and continue to improve the product without forcing them to purchase a new version of the product as is the case in more traditional models of deployment.

Talis Engage is designed to be easily managed. It allows communities to define their own record types and relationships which are suited to their needs and the way in which they want to use the system. By building this level of flexibility into the system we allow users and communities to find ways to use the system that we ourselves could not have predicted.

Community is also about identity and as such we ensured that our customers could customise the look and feel of the application to reflect their community. Also by adopting industry standards like XHTML and CSS coupled with open API's we have given customers the ability to integrate Talis Engage with their existing web sites and applications.

Talis Engage is built upon the Talis Platform, a Semantic Web application platform. This means that data created by communities within Talis Engage can be easily interlinked to other data on the web. Talis Engage takes advantage of the ontology and themes of the W3C Submission for Semantically Interlinked Online Communities (SIOC) which uses Friend of a Friend (FOAF) to describe contacts and Simple Knowledge Organisation System (SKOS) to describe taxonomies such as IPSV (Integrated Public Sector Vocabulary).

The information in Talis Engage is modelled using industry standard structures, rather than proprietary schemas developed by Talis, this means any data created in Talis Engage is portable to other applications and sites, and again this is facilitated and encouraged through API's. In many ways Talis Engage is a thin layer built upon the Platform which is doing all the underlying searching, indexing and semantic relationships between records.

Consequently Talis Engage truly is one of the first commercial applications utilising Semantic Web technologies at its core. It solves a real problem, but by utilising Semantic Web technologies it enables communities to link and share data in ways that would not have been possible otherwise. When you consider that it is then possible to aggregate the linked data generated by each local community into a more holistic regional or national view we begin to see that applications like Talis Engage bring us one step closer to realising the dream of a Web of Linked Data.

*Nadeem Shabir is a Technical Lead at Talis, working on the development of several new applications that rely upon the Talis Platform.*

**nadeem.shabir@talis.com**

**[www.talis.com/engage](http://www.talis.com/engage)**



# Sir Tim Berners-Lee Talks with Talis about the Semantic Web

*Talis has been producing a series of podcasts exploring the Semantic Web for some time, more details of which can be found on p 6. In this instalment, Paul Miller talks with Sir Tim Berners-Lee of the World Wide Web Consortium. See [tinyurl.com/2k494w](http://tinyurl.com/2k494w) to hear the podcast itself and for links to resources mentioned during the conversation.*



Hello and welcome to this Talking with Talis Podcast with your host, Paul Miller. Today, I talk with Sir Tim Berners-Lee, inventor of the World Wide Web and

now Director of the World Wide Web Consortium. We talk about the Semantic Web, Linked Data and Tim's ambitions and vision for both.

**Tim:** thank you very much for joining me today for this podcast. Usually what I do with these podcasts is ask people to introduce themselves and talk a little bit about where they've come from. I guess with you, we probably don't need to bother. I will point people to your page at the Web Consortium and certainly to things like Weaving the Web for anyone who doesn't know who you are and what you have done to get to where we are today. So, I think, what we will do is, probably just move straight on to start looking at the questions.

**Tim:** OK, good to be with you.

**Paul:** Thank you; thanks for joining us. Over the past few years, there has been an awful lot of Semantic Web research going on in universities and inside research and development departments at - mainly - big corporations. What we are seeing now is really the results of that beginning to enter the mainstream. What do you think we have to do to move finally into the sort of mainstream deployment of some of these technologies and ideas that have been building for a very long time?

**Tim:** I think the Semantic Web is such a broad set of technologies and is going to do so many different things for different people. It is really difficult to put it on one

thing. What are the steps necessary right now for the life sciences community to be able to use it for their data about proteins is probably different from which steps do we need to be able to get interoperability between repositories of library data and museum data.

So, different communities have different faces, different communities always have different social considerations and often there is social steps, which when you finally get people to share data more, to be able to re-use data more; then, just like with interaction of the Web, there is a lot of echos of the same sort of social concerns.

People saying, "If I put up a Web server, then I will be out of the loop. Nobody will come and I won't get the credit." People don't have to come to my door and knock on it to get the information. And sort of misunderstandings like that. Or "If I give people my data, then they will be able to use it in ways which are better than the ways in which I have used it, and then, I will fade into the limelight." So, there are all these... we get these social things, but they tend to be different in different areas.

An important step we have just got over is bringing SPARQL out. So, SPARQL changed the landscape a lot because there is such a lot of opportunities to share, which was impractical to just load and people having then to do Linked Data, so, SPARQL gives access to those. So, I think, we will see a growing number of SPARQL endpoints and that is exciting.

**Paul:** OK. And we will touch on certainly Linked Data a little bit later on. You talked a little bit about people's concerns there with loss of control or loss of credibility,

or loss of visibility. Are those concerns justified or is it simply an outmoded way of looking at how you appear on the Web?

**Tim:** I think that both are true. In a way it is reasonable to worry in an organization, for example. Suppose are in a department in an organization and you own the data about a particular thing, whether it is when the machines are going to be maintained, fixed or what the temperature has been in each of your offices or something. You own that data, you are worried that if it is exposed, people will start criticizing your maintenance of heating systems or something.

So, there are some organizations where if you do just sort of naively expose data, society doesn't work very well and you have to be careful to watch your backside. But, on the other hand, if that is the case, there is a problem. So the Semantic Web is about integration, it is like getting power when you use the data, it is giving people in the company the ability to do queries across the huge amounts of data the company has. And if a company doesn't do that, then, it will be seriously disadvantaged competitively. If a company has got this feeling where people don't want other people in the company to know what is going on, then, it has already got a problem, this just exposes the problem. It is like what people say, "Well my data is actually a mess or a lot of my addresses are out of date, or inconsistent."

Well actually, it would expose... we see all these inconsistencies. Well, in a way, you got the inconsistencies already, if it exposes them then actually it helps you. So, I think, it is important for the leadership in the company, for example, to give kudos to the people who provided the data upon which a decision was made, even though they weren't the people who made the decision.

So, generally, to recognize the fact that people are providing access to their data is important. It's very important in Science, too. If you publish a paper in which you happen to have got a lot of the results by running a SPARQL query over existing cell line data, existing genomics data, existing clinical trials data, whatever it is, then obviously it is very important in the scientific ethos to credit the people who produced them.

If you produce the experiments and put those out there in RDF on the Web, then, the good news is you can expect credit back; and sometime in the future after you have retired, people may in fact... you may get credit from people who are using that.

**Paul:** OK, sounds good. Going a little bit broader than those questions then, back in 2001 in that Scientific American article, you and the other authors painted a very broad grand vision of where the Semantic Web could take us. Did you think we'd be closer to that seven years on?

**Tim:** Well, for one thing that article was, I think, too sci-fi. I think, that really what we have... the message has been... it was looking too far into the future. It imagined the Semantic Web was deployed, and then people had made all kinds of fairly AI-like systems which run on top of that.

In fact, the gain from the Semantic Web comes much before that. So maybe we should have written about enterprise and intra-enterprise data integration and scientific data integration. So, I think, data integration is the name of the game. That's happening, it's showing benefits. Public data as well; public data is

happening and it is providing the fodder for all kinds of mashups.

So, what we should realize is that the return on investment will come much earlier when we just have got this interoperable data that we can query over.

**Paul:** OK, and we are pretty close to that now with the Linked Data work that again we will probably dig into shortly. Weaving the Web, 1999 - is it time for another book that paints again sort of the picture given the experience of where we have gone in that time?

**Tim:** Yeah, I think, it has been time for another book for so long, but, when am I going to find the time to write it. I think, that the same things to be... it would be good to write a number of books.

The books I would like to write if I had time include... I would like to write a whole bunch of technical books about actually practically how to do Semantic Web things. I'd like to write a book about Semantic Web Architecture. And I'd like to write a book sort of painting the path for people in the industry, because I get a lot of questions along the lines of "OK, I read the specs, OK, but here I am, I am the CIO of a company, what does it mean for us now, what should we do?" So, there is a story about the answer to that one, typically "Well you should take an inventory of what you have got in the way of data and you should think about how valuable each piece of data in the company would be if it were available to other people across the company, or if it were available publicly, and if it were available to your partners."

And then, you should make a list of these things and tackle them in order. You should make sure you don't change the way any of your data is existing, is managed, so you don't mess up the existing systems and so on.

So, there's all this sort of advice, which is being repeated all of over the place. Semantic Web experts are being called up by CIOs and asked what they can

do. We need more books on that at the moment, I think, explaining how to put the Semantic Web as a win-only and win-win solution, and adding it to existing infrastructure in companies and things like that.

So, there's lots of books. But, when things get exciting, as they are now, come Monday morning, what should I do? It's just like back in the early days of the Web. Should I go and encourage a working group, participate in an open source project, should I go and give a keynote speech, should I go and do a podcast with Paul Miller? There's so many things to do, that I'm afraid writing another book just hasn't made it to the top of the pile yet.

**Paul:** OK, so there may be an opportunity there for someone else to write that book.

**Tim:** There's a lot of books out there to be written. Maybe also for people to interview the people who understand it, who understands things like Linked Data. Interview the people from the community and sort of write the books for them.

**Paul:** You mentioned SPARQL back at the beginning of our conversation. With SPARQL, I guess, a lot of the technical pieces are now in place. We've got RDF, we've got OWL, we've got GRDDL, we've got SPARQL, and we've got the rest. Are there any big gaps left in the puzzle, or do we have the bits we need now to stop using lack of standards as an excuse?

**Tim:** I think, really we've got all the pieces to be able to go ahead and do pretty much everything. I suppose, really you should be able to implement a huge amount of the dream, we should be able to get huge benefits from interoperability using what we've got. So, people are realizing it's time to just go do it.

If there's one thing which we've foreseen from the early stages, I think, it would be rule languages. So, the rule language is in the works. The first moment we started playing with the Semantic Web, the first thing I did was to write a rule engine. Because to me, that was way

of, for example, translating data from one ontology to another, for slimming it down, thickening it up, making general inferences, writing consistency checkers and things.

So, a rule engine is such a very general thing. Yeah, you see them effectively, you've got a rule engine when you filter your email. You set up email filtering rules through various smart mailboxes. Smart photo albums tend to be little rule engines. So, a lot of applications already are used to that. Users are used to that. Some users, I guess advanced users, to be fair, are used to using those rule systems to sort making their life run better, increasing the automation.

So, I think, the rule language will be really useful. The trouble is, there are so many different types of rule languages. And various other things as well. There's all kinds of things that are around. There's style sheet languages for Semantic Web forms. There are lots of things which I think, you'll see in the explosion of technology.

But the core, absolutely, is very solid now that we've got SPARQL.

**Paul:** OK. Another way that we've looked at the Semantic Web idea, is through the famous layer cake diagram. One of the obvious pieces on that that actually comes into play quite strongly with a lot of discussion around things like social networks and data portability and all the rest, is trust. What are we doing to address trust?

**Tim:** Well, yes, and trust is fairly high up the stack. Originally in the roadmap, I felt that we would want to... when we had rule languages, then, for example, we would have a more expressive system that would allow us to actually express the trust that we feel.

A lot of pre-Semantic Web trust-based systems do things like give people a numerical value of trust. "I trust this person to the level of 0.75," or something, or they say the guy is trustworthy, or he's not. And they're too simplistic.

So, I think, when you look at real systems in the world, you trust one person to give you a recommendation on a movie, and you trust a completely different person to give you a recommendation on whether a piece of code was good code. And so you trust different people for different things, and different agents for different things.

So, in fact, the code that I've been writing, the rule engine, very often it hasn't been for saying this is OK, it's a person with these particular properties, who meets this particular criterion, has said something. A message which comes from a certain source contains this information. And then you can do things with it.

So, one of the features of the N3 language, which extends RDF, is that it allows you to do that. It allows you to talk about what documents have said what, and allows you to write these rules, which say that sort of thing. You can argue about where the information is coming from.

Provenance, of course, is a really important word for almost anybody doing Semantic Web development; provenance somehow comes into their lives. If you're building a triple-store, often these things... we think of as triple-stores, storing the subject, the verb, and the object. Actually, for each of those little sentences they're also storing where it came from.

In the Tabulator, the data browser that we've got here at MIT, you can look at the aggregated view of data about all kinds of things. If you click on a cell, then, you can sharpen the list of sources and see, where did this particular data in that particular cell come from. People always need to go back to the source.

And I think, as we build trusted systems, we'll build them, which will not only go back to the source, they'll look at the metadata about the source. And then, they'll operate on that. So, for example, you'll be able to look at all the data you've got about something you're doing

for a school project. And you'll be able to say, "OK, now just show me the subset of that data, which is released under a Creative Commons license, so I know that I can use it in my school project. So, I won't get into trouble with the teacher."

So, that involves this awareness that a lot of Semantic Web systems are being built with now, which I think, is very important. It's the connection from the data to the provenance of the data, and not just for the name of the document that it came from, but the actual properties of that - the licensing, what it's supposed to be used for, what it's appropriate to use it for, whether I got it because I've gone through an authentication process, and actually whether it's private data, which I should not actually publish at all.

These are, I think... Building systems which track that sort of thing, where I got the data from, and what I'm allowed to use it for, is going to be more and more important. So, those are the ways we're going towards building trusted systems. I think, it's a very important part of the puzzle, but we'll build it using the things which mostly we have.

In the end, to strengthen it, we'll probably put cryptography in there. And N3 gives you a way of talking about what's in a document, you can relate it to the signature. So, you can boot-strap the whole trusted system using the things which we've really been playing with for several years.

**Paul:** Right. And it's about then making some of those things explicit, I guess. You mentioned, for example, that provenance is in the URI, but thinking far more carefully about what that means, rather than inferring meaning that perhaps isn't always there.

**Tim:** Yes. The really challenging thing about a lot of trust issues is the user interface. We're having this problem with the browser. In regular Web browser, at the moment, how do you let the person know that they're actually talking to the bank, their own bank, and not something which has got a URI which has got the

name of the bank somewhere in it, but not in a domain name?

How do you prevent these phishing attacks, for example. That's the great question. That when you look at the browser, you realize there's a certificate there, but the browser isn't actually showing you who the certificate is owned by. It's only showing the padlock to show it has a certificate. You're not checking it every moment to know that you're talking to the right person.

So, they obviously want to put in changes to the browser, so the user interface actually, instead of showing you the URI, it shows you the owner of the certificate and then you've moved up a level of trust. With Semantic Web applications, imagine you're looking at your calendar. On the calendar, you've got interoperability between different applications, suddenly on your calendar you've got bank statement transactions that show up. The photographs you've taken show up on the calendar because they've all got dates. And we've got interoperability.

So when you look at a calendar, suddenly it's got information from all the different places. It may have personal photographs from the family, bank statements from your company, financial statements from your company showing up there. You might use those for deciding where you want to meet with somebody at any particular point, figuring out what was happening on any particular day. You may not want to share that with the people you're in a meeting with.

So, to be able to see from the user interface that there's confidential stuff on the screen, and I should not share my screen. To be able to ask the user interface to filter it, so that I would only know, now I'm meeting with somebody else, can you just make sure that everything on the screen is the sort of thing which I'd be prepared to show somebody else. That also assists in understanding the policies.

And we've got systems at MIT where you get in there and play around with things like mixing OpenID authentication with friend-of-a-friend, for example. Currently, if you want to comment on my blog, then you have to be related through the social network to somebody in the group, the Decentralized Information Group. You have to be a friend of a friend of a friend to some level of somebody in the group; just so we know that you're not a spammer. It's not that we want to cut down the people who can, we just want to cut out the spammers.

So, we see a lot of spammers who are using the social network, which is part of the Semantic Web, to produce the traffic, which actually shores up on the Semantic Web stack.

**Paul:** Well, that's clearly an area which will require a lot of activity moving forwards. Another area that will require a huge amount of effort moving forward is around data for the Semantic Web. We're going to need an awful lot of it. Where are we going to get it from?

**Tim:** There's an awful lot of data out there. And I think, one of the huge misunderstandings about the Semantic Web is, "oh, the Semantic Web is going to involve us all going to out HTML pages and marking them up to put semantics in them." Now, there's an important thread there, but to my mind, it's actually a very minor part of it. Because I'm not going to hold my breath while other people put semantics in by hand.

I'm not going to wait for other people to do it, and I don't want to do it either, to sort of add the semantics to HTML pages. So, where is the data going to come from? It's already there. It's in databases. So, most of this data is in databases. Often the data is already available through some kind of a Web interface.

So, if you take a government department, which is interested in defense data; you take a company's products. You take a printer manufacturer, it sells all these printers, it sells all these ink cartridges, it

can sometimes be an afternoon's work to try and figure out which ink cartridge is actually compatible with which printer. Because they're on different parts of the website and they haven't published that information in RDF.

Suppose they publish the information in RDF, then you could just look up the printer and find all the ink cartridges which are compatible with it. And you could write programs that automatically go and buy all the ink cartridges at the appropriate prices and from the appropriate stores, and it's all getting very automatable.

But, the thing that's holding us up is that, there's data which the companies have got on this, sitting and going round and round on its disks. Or it's in their SQL systems and needs to be exported in a way that we can get at it in linked RDF as a SPARQL. And then, that could be reused. And all the people, all the resellers, all the stationers who sell ink cartridges, for example, will be able to make much better websites because they'll be able to pull compatibility information from the user.

The company will find that its users are happier, and they'll end up selling more printers and more cartridges. The whole world will run more smoothly and we'll have more time to get on with more important questions.

**Paul:** How does a company that has one of these databases take the step to make it available to the Semantic Web? What do they have to do? You said they don't have to go away and rewrite all their pages, but presumably there is a step they have to go through.

**Tim:** Well, there's a couple of ways of doing it. Say that you've got a database-type website. One way to do it is to look at it... let's stay with the printers, for example. When you look at the website you notice there's a page on the printer, which has got the specifications, and it's got a little table of the properties of the printer. And there's a PHP script somewhere, which produces that.

So, you get somebody who understands these things to write another PHP script which is totally parallel, which just expresses the same information in RDF. That's all. Expressing it in RDF is actually kind of simpler than expressing it in HTML, because when you express it in HTML, then you have to make sure that the CSS is pretty, and you've got the icons in the right places, and it meets the organization's guidelines for being part of the website, and it's got a consistent style and everything, and it's got the navigation buttons.

When you do an RDF one, to a certain extent you need navigation buttons in the sense that when you output the data about the printer, you have to make sure that when you mention the compatible ink cartridge, you use the URI for the compatible ink cartridge, which will cause the RDF machine which is interested in that to pull up the RDF page about the ink cartridge.

But this RDF page is meant for machines or people using RDF browsers. So people using things like the Tabulator will be able to pull that up and follow that and look at the links, and then also, they're going to make tables of all vendors and tables of all the cartridges, and use the interface to concoct queries effectively, for all the printers that take ink cartridges that cost less than \$20 or something.

So, one way to do it is to parallel the website, parallel each Web page, which has got interesting data about a particular thing, such as a product, with the same thing in RDF. And then, you try it out using an RDF browser to see if it works and see if it's got the data. And then, you pass it to some of your friends or your colleagues or your peers in other companies who would be interested in the data and see whether they can use it and match it up.

The other way is that you start at the database. You just look at the database, you do that inventory of all the tables in the database, hopefully, you sit down with some of the people who designed the database tables, because often within

a company, it's really a bit of a black art knowing exactly what some of the columns in a database are actually meant to mean. Sometimes, you have to go and have coffee with the people who actually designed the database schema. But, then you sit down and look at the tables, and you can point a piece of code, such as for example, the D2R Server code from Berlin. We've got Python code, dbview, you can point it at MySQL database. And it will make a sort of default ontology and say, "That's OK." Let's assume that for each table you've got... each table is about a set of things. We'll make a class, which corresponds to the table. And we'll give a URI to each of the things, which is described by a row in the table. And it will generate for you a default ontology.

Now, the more you tell it about the database, then the better it will be, because it realizes, "Oh, yeah, this is a product ID that crops up here," and it knows where the product ID crops up in other tables. Then, when you do a look-up for the RDF URIs for a given product, it will not only give you the data in the product table, it'll give you the other links, the links pointing into it, that say that this is the product, or this company has ordered the product, it's in this invoice, and it's got this compatible product and so on.

So, as you tell the system more and more about the database, it starts to produce more and more a reasonable RDF view of the world. And you can wander around it with Disco or one of the other RDF browsers out there, and you can wander around it, and as you slowly add pieces or add labels to the ontology, then your mapping file is now becoming part ontology and part mapping. It's mapping from and explaining the internal database schemas, how they map to this ontology that you're making.

As this grows, you get a more and more useful system. So, you may in fact go to that process to a point where you actually start using the data. At a certain point, you may just decide that you've done enough for now and just export that

data and let other people clean it up.

One of the things they're likely to come back and say, "Well, you've got a start and end time there, could you just export those in an iCal ontology?" Because a lot of people use start and end times, and if you do that, then, we could all put it on our calendars, and we can all put it on our timeline views and things.

So, after you've done this initial export raw from your database, then you can look around for terms that you're using where there are actually ontologies out there, go to the Semantic Web Interest Group, go to your friends. If you're in Cambridge, come to the gatherings we have every second Tuesday at MIT, with people who are doing this sort of thing with the Semantic Web. Ask around, "Is this ontology useful? Has anybody got an ontology for this?" Use tools like Swoogle to search for ontologies.

You don't just use internal terms, but where you're sharing terms, you use terms that other people use as well.

**Paul:** Is this the kind of thing that the Linked Data Project are doing then?

**Tim:** The Linked Data Project as a whole has got lots of projects within it. It's a sort of linked open data movement, I suppose, in which there are different projects. And the greatest thing is, with this linked RDF, you get interoperability between projects which are really quite different. For example, DBpedia is one of the more famous pieces of it. DBpedia is an extraction from Wikipedia of all the little data boxes. So, you have a box, for example, in Wikipedia for cities, that gives its latitude and its longitude, which county it's in, how many people, the population, and so on. And so, they construct relationships between geographical entities and so on. And so, it produces a really interesting graph, all by scraping the rather formalized HTML which is in the forms of a page to mark up or mark down, which is in the Wikipedia.

On other systems like MusicBrainz, they have a database and there was an

ontology created and mapping done very deliberately. Of course, there are lots of people dealing with tracks and singers and albums, and so, there's a lot of interest in interoperability for music players and so on, and music look-up services.

So MusicBrainz could piggyback on lots of other ontologies, but also, the singers are often in Wikipedia, so, they could connect songs, in some cases, which are in Wikipedia and certainly artists which are in Wikipedia into Wikipedia entries. So, that took a certain amount of negotiation between the parties to actually go through and do a data cleanup, in which they made sure they identified appropriately the corresponding nodes and linked them together.

So, some data is scraped from HTML pages, some of it is pulled out of databases, some of it comes from projects which have been in XML. So, things come in many different ways. And once they're exported, as you browse around the RDF graph, as you write mash-ups to reuse that data, you really don't have to be aware of how it was produced.

**Paul:** That sounds what you need, more or less, isn't it? It's about making this data visible and seamless and available to other people and their applications. And how big is this activity at the moment? I do remember some figure suggesting very large numbers of records now being available.

**Tim:** There are. I'm never good at large numbers. But, there are various linked open data projects. There's a [linkeddata.org](http://linkeddata.org) that I see somebody has put together, there's Linked Data in Wikipedia. If you Google for 'linked open data', then you'll find pointers to various things. Where you'll find a completely up-to-date list, I'm not sure. Maybe, we should send attached comments at the end of the blog, pointing to things like Richard's Venn Diagram of how the different pieces overlap in the large Linked Data projects. There's a

lot of Linked Data which is not publicly advertised at DBpedia levels because it's more of a niche interest, which benefits from connecting into these people and maybe the Internet and those sort of things. I think, it's difficult to measure it all.

I think, one of the large contributions to it has been the friend-of-a-friend, which is exported by a bunch of social networking sites. So, the FOAF data. FOAF data used to be the largest contribution to the Semantic Web. I'm not sure whether it still is.

There's the Web conference in China, and there's going to be a Linked Open Data Workshop associated. That's in April. And there's also going to be some sort of linked open data conference in New York in June, I think. So, there's a lot of stuff happening with people interested in this. Go to the Wikipage and contact people in some of the interest groups, there's an IRC channel to find the people who are involved in that.

**Paul:** Yeah. And I'll include link to all of these in the show notes and people can follow along. There certainly is an awful lot happening. From our perspective here, we're certainly watching this as being the thing that begins to demonstrate the real value of the Semantic Web outside of the research community. It's what you can actually start to do with this data as you link it up.

**Tim:** Yeah. Well, I think, there's a certain role that's played by public data, so that if you say in your FOAF file... I say that I work in Boston. It's kind of nice to say, "Well, I live in Boston," and get a DBpedia reference to it, so, somebody's immediately got access to a lot of data about it.

I did a demonstration for Fidelity the other day, where I pulled some comma separated value files, some CSV files, from their website about their mutual funds. So, I could then put that in RDF of course, and in that case you could look at it in tables and you could look at other people's mutual funds, which you had

pulled from other websites. To distinguish them, of course, I could put a link to who's actually offering the funds, and I could use the DBpedia URIs for Fidelity. And that suddenly makes it much richer. And suddenly you could say, "OK, I'm looking at all the funds which are based in cities on the East Coast." Suddenly, they see this connection.

**Paul:** Yeah, definitely. I think, as we move from the original reasonably straightforward idea of the Web towards a more Semantic Web, are we having difficulty bringing understanding with us. The original idea of the Web was pretty easy for almost anyone to get.

It was also actually very easy to do. You could simply open a text editor and write HTML. I can remember doing that, very early on, before the other tools came along.

As we move towards a more complicated Semantic Web, does it become harder to understand, and also harder to actually engage with?

**Tim:** I think, it depends really, it depends so much on how you look at it. So, the Web, and the Semantic Web, the existing Web... maybe we should say, the hypertext Web, the document Web, and a Data Web, are... In some ways we are not leaving the document Web behind. It is not as though, when you say we are moving towards a Semantic Web... Yes we are moving towards and we are implementing a Semantic Web, but it very much complements the Web of documents. The Web of documents will continue to exist, and as we are adding more and more video, and all forms of interactivity, that is going to be very exciting too.

So, we got these two things that are maturing, in a way. If you use something like the N3 Syntax, the N-Triple Syntax, or the Turtle syntax for data it's very simple, it is a very simple language. You can write things down very easily. It is not really more complicated than HTML.



So, being able to write a FOAF file for yourself in N3 is easy, and then you can convert it into RDF/XML for output. So, in a way, it's got the same, more or less, structure as in N3 syntax. It has got that same sort of items, you can convert your RDF into it and look at it. I can fix it and put it out there just like I would have done it with HTML. That's a lot that can be said for it.

On the other hand, data is different from documents. When you write a document, if you write a blog, you write a poem, it is the power of the spoken word. And even if the website adds a lot of decoration, the really important thing is the spoken words. And it is one brain to another through these words.

And as a person is expressing himself to another human being, in such a way that the machines, they will try to understand it. But, they will really not be able to catch up with the poetry.

When we are creating data; when, for example, I am creating some information about an event. If I am creating information about an event, then, I am putting in the time and place in such a way that I can latitude and longitude out of it. And putting in the people who are invited to the event, using their email addresses, for example, in such a way that those people can be identified and linked in. So, I am constructing something which fits in and will be reused in all kinds of different ways. Just as the poem will be reused. In the future people will read the poem, and get all kinds of different things out of it.

But, the data, is in a way... the whole is more powerful mathematically. The fact that people will be able to do inference and they will be able to conclude from the fact that a person is at that event in that location, that they are not in a city 100 miles away, and therefore that they cannot attend something else.

They will be able to do things like conclude from the fact that a different person at the event, took the photograph on their camera, during the event;

therefore, the photograph was of the event. So, therefore, it would be reasonable to share the photograph with anybody else who was at the event. Those people would have been identified.

So, in principle, a whole lot of work can happen. Our life can be made much easier, because we put this in. And thinking about how that all works, I suppose it is more complicated, because it is more powerful, with building systems which is going to do a whole lot more.

There we are moving, I suppose, from the horse to the motorcar. In a way, the motorcar is a more... the internal combustion engine is a more complicated thing, but it will enable things the horse can't do. It can just go a whole lot faster and it can go on for a whole lot longer. Even though we still like horses.

One of the important things about the motorcar, of course, is that it has got actually, for most people when they use it, it has a very simple user interface.

So, I think, a crucial thing is that, most of the times we are using the Semantic Web, we are using it underneath the user interface, which will make it very, very easy in the way that people have got used to sorting their email, managing their calendars, and their contacts, and their address books, their appointment diaries.

These things have interfaces, which have matured over time. They are not perfect. They could all be improved upon. They will be improved upon. They get more complicated when they are linked together, but in a way they also become easier to use when they all have been linked together. When I can automatically go from my appointment diary, you know, through to the person, and then to the information about the place they work, and so on.

In a way, it will again put... The inline architecture of the Semantic Web is actually much simpler than that of

HTML. It is just these triples. So, in a way understanding it, and developing with it, is actually a whole lot simpler. It is not inherently more complicated.

However, when you look at the complexity of all the things on the Web, or the Semantic Web, when you look at the size of them, of data from all kinds of different places, and you think about the implications of that, then yes, it is complicated. The Semantic Web is already complicated. The public Semantic Web is a very interesting and complicated, very powerful, useful thing which is interesting to analyze.

So, from that point of view, the Web as a beast, gets more complicated and more intricate, in a way more exciting with the Semantic Web.

**Paul:** I guess one of the areas we have seen a lot of growth in easy to use tools on the Web recently, has been around the whole sort of participative movement that has loosely been labeled Web 2.0.

You wrote a blog post towards the end of last year, on the Giant Global Graph, which really made it quite clear how some of the Semantic Web ideas that had been around for a long time, applied to the sort of new kids on the block, in the social networking sphere.

Do you think developers of applications like, say, Facebook and LinkedIn and the rest, are ready to embrace the Semantic Web, or do you think they think they can do it themselves?

**Tim:** I think, there is two parts of that. There is whether they will need to give up the data, and whether they are willing to use the standards. You will find to start with a lot of places, like LiveJournal, for example, they expose FOAF. So standard RDF Friend of a Friend for your friend network.

If you look at MyOpera, not only do they expose a FOAF link but they allow you in your Opera profile to say, I am also this LiveJournal person. So, you can follow your links, you can follow the friend of a

friend, the social network, through one site and into other.

I think, it is a very grown-up thing to realize that you are not the only social networking site. When you do that, it is like a website that all of a sudden... otherwise it is like a website which doesn't have any links out. In the Semantic Web similarly, if you don't have any links out, well, that's boring.

In fact, a lot of the value of many websites is the links out.

So if you start off with one of these social networks that does have links out, then you will find out a huge amount. If you find one which doesn't, then you will be able to explore it using common tools, if they use the FOAF standards, but I bet you'll be limited; you will bump into the edges.

Now if you look at the social networking sites which, if you like, are traditional Web 2.0 social networking sites, they hoard this data. The business model appears to be, "We get the users to give us data and we reuse it to our benefit. We get the extra value." When one person has typed in who it is that's in a photo, then we can benefit. We give the other person extra benefit by being able to give them a list of photos that they are in. That's tremendously beneficial.

That's the power of the Semantic Web. And I think, the social networking sites, some of the ones that have become very popular have done it because captured the semantics. They haven't just allowed you to tag something with somebody's name, they've allowed you to capture the difference between somebody who took the photo and somebody who's in the photo, so that the power of the reuse of the data has been much greater.

So, first of all, are they going to let people use the data? I think, the push now, as we've seen during the last year, has been unbearable pressure from users to say, "Look, I have told you who my friends are. You are the third site I've told who my friends are. Now, I'm going to a travel

site and now I'm going to a photo site and now I'm going to a t-shirt site. Hello? You guys should all know who my friends are." Or, "You should all know who my colleagues are. I shouldn't have to tell you again."

So, the users are saying, "Give me my data back. That's my data." That was one of the cries originally behind XML, it was a desktop application. Don't store it in a format which I can't reuse. So, now it's, "Give it to me using the idea of standards. If you do that, then I can do things with it."

Now, there are two architectures which allow you to do this. The way some of the sites are working is that you'll go to, for example, a t-shirt site which is going to allow you to print a t-shirt or something. Or say you go to a photo site and say, "Now I want to see the photos of my friends. You don't know who my friends are. I am going to authorize you in some way, using something like OpenAuth, to go to another site. I'll open the gate with them, I'll tell them that it's OK to use the information about who my friends are."

So, just for the purpose of printing those t-shirts or just for the purpose sharing these photographs with my friends, I'll allow you to know who my friends are. So, we're getting this moving of user data between different sites. Now, we've got the user data stored in more than one place. Obviously, refreshing is important and we've got dangers of inconsistency and so on, and we've got all this third-party authentication going on.

There's another model, which is that I, the user, run an application in my browser, for example, or on my desktop. It could be an AJAX application. It could be an application which allows me to look at photos. But, what it does is, it pulls the photo information from many places, and I directly authenticate.

And when it pulls that information in, it pulls in all the information I rightfully have access to. It pulls my friends' information as well from different places. So, if I've got social networks, or for that matter, if

I've just got files in Web space. If I've got a friend-of-a-friend file, or even if I've got my local file on my desktop that now I can use. So, I can use my address book.

So, it now pulls all the information that I have access to about the social network, and it pulls all the information in that I have access to about photos, and then it allows me to browse the web of photographs of people using the full power of the integration of all those things. It allows me to look at photographs of friends, photographs of people that are friends of friends, but are not my friends, to see if I should be adopting them as friends and so on.

It can do all these powerful things, and it's happening actually in the user's browser, or it is happening on the user's machine. Both of these systems at some point allow people sharing data. The second system is much simpler. The second system involves people writing scripts which will operate across different data sources.

Web 2.0 is a stovepipe system. It's a set of stovepipes where each site has got its data and it's not sharing it. What people are sometimes calling a Web 3.0 vision where you've got lots of different data out there on the Web and you've got lots of different applications, but they're independent. A given application can use different data. An application can run on a desktop or in my browser, it's my agent. It can access all the data, which I can use and everything's much more seamless and much more powerful because you get this integration. The same application has access to data from all over the place. Does that make sense?

**Paul:** Absolutely, yes. And I think, creating and maintaining that split between the data and the interface or the application certainly has to be the way that we go. As you say, persuading some of the companies who have built a business model around holding the data is the task that we still have to get right. Although I suppose, as choice arises, people can choose not to go to those sites, can't they?

**Tim:** People can indeed choose not to go to that site. It reminds of the story of what happened when bookshops went onto the Web. Sometimes I have to remind people about this. When originally the Web came up and bookshop owners learned about it, they said, "OK, I was told at my dinner party that we have to have a website, so get us a website." The website would come up and it would say, "We recommend you go to see this wonderful bookshop. This is the address." And it would give you the directions.

But, they never put up a list of books that they sold. If people put that up, they wouldn't go to the bookstore. The important thing was people should go to the bookstore. And anyway, it was also commercially sensitive data. If you put up a list of books that you were carrying, if you put up a catalogue, then, your competition could immediately use that information to compete unfairly.

And then suddenly they would realize, they would be told, "Well, excuse me, sir, the competition already has its catalogue on the site. So, everybody is going to their website. Nobody is going to our website because it doesn't have any information. When they go to the competition, they go to the website first to check if they have the book, and then, they know if they go to the store they're going to be able to find it."

"Oh, OK. Well, I guess we'd better put our catalogue up." "Shall we put the prices up?" "Oh, no. Don't put the prices up, because that's commercially sensitive. They should see that when they go to the store." "Oh, the other people have put their prices up now? And so now they're taking our customers again?" "OK, I guess we'll have to put our prices up."

Stock levels? Of course you don't put the stock levels! "That's our backend information, you don't put the stock levels up. People can come to the store and when they order it they'll find out whether we've got it in stock or not. Oh, really, they don't like that?" Clearly then,

bookstores moved to putting their stock levels online because people got fed up with finding that they'd ordered a book and then they get a little email saying that it's been backordered for two months.

So, there's this syndrome of competitive disclosure. When actually business works better, when people have disclosed and are communication one to the other. Once it starts, then it can snowball. So, once we have people putting their catalogues up in RDF, it may be that there will be aggregators that look at products and they won't see your products if they're not up there using Semantic Web standards.

Which, you're giving a talk, and the person doesn't advertise it using the standards, it won't be streamed. It won't be up there, people won't have it on their calendars. People won't come and see you talk because the information wasn't made available publicly in interoperable fashion.

So, I think, the lesson of the bookstores on the Web is an important one. If you're working for a company and there's a sort of hesitation about sharing information with peers, that you know actually will make the company work better, tell them that story.

**Paul:** Will do. We will write it down and disseminate it widely. As we come to the end, I am conscious of the time, what do you think - and this is probably the hardest question of them all - what do you think the biggest challenges facing Semantic Web adoption and Semantic Web rollout are over the next couple of years?

**Tim:** Oh, that is a great question. I suppose, I think, the paradigm shift is the biggest hurdle. The fact that when you think in terms of Semantic Web, you think differently. It was actually a problem for the Web too. People look back and they say, "Well, the Web is so easy, you just download the Web browser and then you could just...." And the moment they use the Web browser, you had to write HTML, and then you could edit HTML pages

with editors and the whole world took off.

Well actually, before there was a significant amount of Web, it was really difficult to persuade people it would be a good idea. They just didn't understand how fundamentally essential it would be to be on the Web. They didn't understand what a kick they'd get out of finding that somebody had reused their information in a different way. They didn't understand how beneficial it would be to have more or less all information that they could think of available.

And imagining it, now imagine people write a SPARQL query as though the world, as though all the data to which you actually legally practically have access, actually is technically available to you as well - just anything which comes up into your mind as a scientist, as a businessman, just as a school kid wondering the answer to a science project question... There are obviously a set of people who get it.

They have a twinkle in their eye, they are incredibly fired up, because they understand it is going to be really really exciting when it all happens. To a certain extent, they are finding that these areas like life science, like social networking, like the Linked Open Data projects, where it is all starting to come together.

There are other areas where somebody who has worked in data systems doesn't get it. So, to understand, so explaining it, why you can't do it in an audio blog of 60 minutes. Because when you explain the new way of looking at the world, new way of looking at data, moving up a level from the database to the Web of things, you have to listen to where somebody is coming from, you have to understand what concepts they've got at the moment. All this is coming to this point of view of an object-oriented programmer or a database person, because the way you paint the Semantic Web is going to be very different.

And the misunderstandings they will have, they naturally get about Semantic Web will be very different. But, it is

happening more and more. So, I think, it is a question of how this meme can spread or how understanding about what this is. But, I hope that having Linked Data online, having user interfaces to it will help.

One of the crazy things, one of the big impediments we have had for the last few years, I guess was maybe a planning fault, is we didn't have user interfaces, we didn't have generic interfaces. When people asked me what the Semantic Web browser would be like. I'd say, well you don't understand, it is not really... documents are for browsers, data is for applications, so these applications will use it.

And in fact, I realized we need to get that feedback, "Oh look, Ma, at my Semantic Web data", just like "look, Ma, at my Web page." Hence the development of things like the Tabulator, which are very much in their infancy, but starting to be able to give people that instant gratification. I put my data up there and now I can see it up there, now I can show you, now I can immediately get kudos. Now, I can stop having to answer the phone. I can point people to the data. They can go and use a Semantic Web browser on it, they don't have to come and ask me.

So that, I think, is an important thing. We are only really at early stages of sort of the art and science of producing good Semantic Web, generic cross-domain Semantic Web browsers... and editors of course.

**Paul:** And that is an interesting point actually. Your original Web tool was a browser and an editor. Was it a mistake not to push harder to maintain that right back at the beginning?

**Tim:** I really wish we could have for a lot of reasons. To start with, people shouldn't have had the pain of having to write angle brackets, and that people were prepared to... that was a total shock to me. I had assumed that people wouldn't. Also if we'd have had... so we would have had I think, a much more collaborative space had all the browsers been editors.

And also, we wouldn't have all this terrible markup, because we would have had the markup where they are generated automatically, and it would have actually had matching tags. So, in that respect, in a number of respects, an ambiguous one being a collaborative space. We had to wait too long for blogs and wikis. Blogs and wikis would have happened sort of very much more easily if they had been editable things, if people had editors.

The problem of course was HTML got complicated. It had all kinds of things like sorting DIVs, which are difficult to edit. When you have nested lists, it is more difficult to edit than unnested lists. And I think, that is part of it. Also I think, the fear of actually being able to edit a page; then we would have had to develop some sort of templating. When you edit a blog, actually you edit only the very middle of the page, you edit a stream of text, then you have a very limited markup, and you don't get an option of editing all the stuff around it that is generated automatically.

So, I think, what we need are editors, so that is a limited. We should have a type of HTML form where you can just type and do bold, and strong, and emphasis and so on, very easily using these interfaces, which are supported by the browser instead of by a bunch of java scripts. And that will help us become more - that way we need to be more collaboratively creative.

And at the same time, for the Data Web, it is important that when people see data that is wrong, if they've got the rights to access it, they should be able to fix it. And if they see an address, see a wrong address up there or an email address and also it should be, actually this is not right.

It should be very easy for people to enter data and also things that we do like entering bug reports, entering agenda items for meetings, entering new events, all kinds of things that are really generating data and we should be able to do that really easily, and yet keeping both the Web of hypertext the Web of

Data, keeping it the Read/Write Web is a really big priority for me.

**Paul:** Absolutely. And it is fascinating to see how people do take to things like wikis and certainly conversations internally within Talis, where the development team uses Wikis all the time. And as you roll them out to other parts of the organization, there is an initial fear to get over, but once you get over it, people take to it like ducks to water. It is remarkable to see, yeah absolutely. Good, thank you very much Tim. Before we wrap up, do you have any final things that you wish I'd asked you?

**Tim:** No fundamental thing. Paul it is really good of you to do this series. I find it really useful to be able to delve into and it has been great listening. So, thanks for keeping on this tradition. I think, it is great for us now and maybe it is pretty interesting for posterity as well to track people's ideas in this space.

**Paul:** Thank you very much and thank you for taking the time to take part.

*See [tinyurl.com/2k494w](http://tinyurl.com/2k494w) to hear the podcast itself and for links to resources mentioned during the conversation.*

*This podcast is part of a series in which leading proponents of the Semantic Web share their thoughts.*

*See [www.talis.com/platform/podcasts/](http://www.talis.com/platform/podcasts/) for the full catalogue.*

# The Semantic Web Gang

*In March, we started the Semantic Web Gang series of podcasts, in which a group of Semantic Web practitioners share their views on the issues of the day. In this first installment, we meet some of the regular gang members and explore Tim Berners-Lee's assertion that the Semantic Web is ready. See [semanticgang.talis.com](http://semanticgang.talis.com) to hear this and subsequent episodes, and for links to resources mentioned during the conversation.*

**Paul Miller:** Hello and welcome to the Semantic Web Gang with your host, Paul Miller. In today's first episode of Semantic Web Gang I introduce some of our regular Gang members and we take a look at recent assertions that Semantic Web is ready for business.

Welcome to this first episode of the Semantic Web Gang. My name is Paul Miller and I am joined by some of our regular Gang members on this, our first call.

As you won't yet know their voices, I'll ask them to introduce themselves one at a time so that you can associate names with voices. And first of all, I'll go to Alex, Alex Iskold.

**Alex Iskold:** Hi everyone. My name is Alex Iskold. I am founder and CEO of a smart browser company called AdaptiveBlue. And I'm also a regular contributor to a Web blog called ReadWriteWeb.

**Paul:** Thank you, Alex. Good to have you. Next on my list I've got Thomas Tague. And you're all going to get very confused because there are two Toms on the call. Thomas.

**Thomas Tague:** Thanks Paul. I am Tom Tague. I'm the leader of the Calais initiative at Reuters and one of many evangelists for that initiative within Reuters.

**Paul:** Thank you, Tom. And next, so that you can identify their voices, I'll go to our other Tom, Tom Heath.

**Tom Heath:** Hi. Thanks Paul. My name is Tom Heath and I'm a researcher on the platform team at Talis in the UK. My primary interests are in recommender systems and linked data. Before I joined

Talis I was a Ph.D. student at the Open University in the UK, where I developed Revyu.com, which was a review and rating site for the Web of data, and winner of last year's Semantic Web Challenge.

**Paul:** Great. Thanks Tom. Next, Daniel.

**Daniel Lewis:** Hi, I'm Daniel Lewis and I'm a Technology Evangelist for OpenLink Software. And we develop Database and Semantic Web technology.

**Paul:** Thank you, Daniel. And last but not least, we have Greg.

**Greg Boutin:** Thank you, Paul. My name is Greg Boutin. I am a marketer at heart. What I do is turn technology into product. I came to Semantic Web probably more recently than everybody else here. I was in a different technology called Renewable Energy before.

I worked a bit on starting my own company and I am now working as a director of market strategy for a company I cannot really talk about because it is a stealth-mode in Canada.

**Paul:** OK. Thank you, Greg. And thank you all. With these calls each month, we are going to try and cover some of the hot issues that happen to be in the news in the Semantic Web space at that time.

As this is our first episode, we can take a slightly broader look. There will be some current affairs stuff, but we can also take a little bit of a step back and ask some slightly larger questions.

So I guess to start that off, I recently did a podcast interview with Tim Berners-Lee, the inventor of the World Wide Web and director of the World Wide Web Consortium. And we were talking about the readiness of the Semantic Web for

essentially enterprise use. The Semantic Web idea has been around for a very long time but really it's only in the last 12 months or so that we are beginning to see real financial engagement with the Semantic Web and some of its capabilities.

So in that interview, Tim was very strongly suggesting that the Semantic Web now ready for adoption. To align with that, we also have things like the recent news from Yahoo!, where they are going to be incorporating a lot of Semantic Web ideas, Semantic Web specifications and related things like microformats into their search systems. So clearly they are seeing some value here.

So I guess a question to all of you, and I'll start off by naming answerers until we get more of a discussion going, but a question for all of you: is the Semantic Web then ready for the big time or are there still things we need to address? And I guess first of all, I will go to Alex, who of course has a company that presumably believes the Semantic Web is ready for the big time because it's paying the bills.

**Alex:** Well, I think this is a pretty broad question. And I think there are different axes to it.

I think there is a technology axis and there is a sort of consumer axis and a marketing axis. I think we are seeing a lot of companies in the space, and so in terms of the technology I think there are definitely a lot of people who are recognizing that semantics can be a very powerful thing and you can do a lot of interesting stuff with it.

Now this treatment is fairly academic and leads to the specific question: what is it exactly that we can deliver to consumers

that you know would make them believe that Semantic Web is useful? And that's sort of, that's different, monetizing it and proving that there is a value for the end user.

And I think that while we are seeing kind of good growth in terms of the technologies, we are still sort of searching for killer apps and the applications of the technology that can demonstrate really, really interesting consumer value.

**Paul:** OK. Tom, what would your view be from Reuters? Again, you are investing and focusing quite a lot on things like Open Calais. So you are clearly seeing a value. Do you think it's ready for mainstream adoption? Or is this still quite a niche idea?

**Thomas:** It's the big question right now. I think, you know the interview that you had with Tim Berners-Lee, those are wonderful because it's great to see one of the big thinkers speaking in very simple terms.

There is so much complexity associated with discussions around anything with the term semantic in it, and he kind of cuts back to the chase.

I guess my view on it right now is the construction kit is now ready. We have the basic tools that we need to start building the Semantic Web and semantic applications there. But that's very different than ready for adoption. It's ready for construction. It's ready for us to start building great things.

But like Alex mentioned, right now I think we are all struggling to some extent on what are those great things. Now, my view is they will come out in two waves. There will be a set of capabilities and applications that are essentially augmentation of what we know today. You know, just like Web 1.0 was viewed as a new way to publish content and Web 2.0 has dramatically surpassed that.

I think our first round of things will be augmentation of what we do today. It'll be better search, better navigation. I'm certain advertising will creep its way

into there as well. And I think we could probably lay out several cases where semantic capabilities can dramatically improve things we do today.

I think where we all kind of hit the wall is what's next. You know, whether we call it the killer app or the new paradigm. And I think we are going to talk about it a lot for the next couple of years. But the only way we are going to learn what it is, is by having thousands of people out there trying to build, trying to build capabilities and start to learn what works.

**Paul:** And presumably that's why we need to be doing the kind of things you are doing, sort of opening up APIs and saying, "Come and try it."

**Thomas:** Yeah. In many ways our goal with Calais is pretty simple. It's to remove one of the barriers to people getting started. You know, people sitting down and handcrafting semantic metadata for news articles and blog postings and whatnot is probably not realistic, given the economics and time pressures in those industries.

So our hope is by putting a tool out there that simplifies that, and does it for free by the way, we just want to see thousands of people experimenting. And frankly, we want to watch and we want to see what looks compelling, what looks exciting, what looks interesting.

I think one of the big lessons we have learned over the last 10 years is it doesn't matter how many big brains you get in a conference room inside company X. There will be more innovation out of the tens of thousands of people trying new things out in the world than you can ever possibly have inside any one organization.

**Tom:** This is Tom Heath here rather than Tom Tague. But I think Tom Tague is absolutely right. There is a certain level of maturity now that has been reached, which means that developers can get hands on, they can use this stuff. There is enough data now to start playing with.

And we've really got to just put that out there and allow the innovation process to

take place out in the world and then try and learn the lessons that arise from that. I think that Tom is right when he says that the kinds of things that we will see in the first instance, in the first wave will really be more of the same kinds of things that we've seen now, better search engines, better browsers.

But I think as we learn more about interacting and using this data, then we'll start to understand a bit more and in more detail about the kinds of different functionality and different end-user benefits that can be brought by having a real Web of data that really is interconnected.

**Paul:** And presumably the really important point there is that none of that would have been possible without the investment in the generic specifications from organizations like the Web Consortium.

You know, if Reuters were putting Calais up and Alex was putting AdaptiveBlue up and we were putting up the platform and all these different things all on the basis of proprietary technologies and proprietary specifications, this linking that you're talking about wouldn't happen.

**Thomas:** Yeah, this is Tom Tague. You know, when you're reading about general Semantic Web topics out on the Web, particularly where people are contributing their opinions, there is clearly a camp that says "This is all academic. This is all very arcane. There are a lot of conferences. There are Ph.D.'s meeting in rooms writing specifications and it doesn't have anything to do with the real world."

I think that's unfair. I think there has been a group of people who have labored in the background to set the stage for themselves and other people to do some amazing stuff. And so things like the Consortium and the standards that have come out of that, frankly this is where we get to stand on the shoulders of giants and harvest some of what they have built for us. And they deserve credit.

**Tom:** Yeah, absolutely. I completely agree. I think it's also really important that we understand that the Semantic Web or



the Semantic Web of data isn't about a completely new technology stack.

Right there at the bottom there is HTTP and there are URIs. And it's those things, which really make this work as well as RDFs and OWL, etcetera, etcetera.

**Paul:** OK, agreed. Daniel, do you have any thoughts to share on this?

**Daniel:** Well, yeah. I think as he said, the development tools are definitely ready enough for people to start using this stuff. And we are even starting to media coverage as well like a lot of BBC articles are now talking about the Semantic Web and Web science and things like that.

And so we've really got to try to make simple things as well, simple explanations of what this stuff does.

**Paul:** Yeah, agreed. Greg, what is your perspective on this, sort of coming from slightly to one side, but certainly you've been looking quite closely at what a lot of people have been doing. So what would your perspective be on this readiness question?

**Greg:** Yeah, I think it's the billion dollar question really. I think, like as far as I see it, the Semantic Web really started by identifying the problem with the existing Web, which was the data is not linked enough. Or at least the links that are out there and the data, between the pools of data are not qualified enough.

And these problems were mostly identified by people like yourself and people who were using the Web or who were building the Web itself. Now the question is, with all this technology that has been created and that seems to be maturing really rapidly, what problem these are going to address for the end user. And is there a problem out there?

And I mean, I'm positive there are many problems that could be addressed by the technologies that are being created. But there isn't yet out there a valid proposition as far as the end user is concerned.

And I think when we ask the question, "Is the Semantic Web about to go

mainstream?" to me, when I think about the criteria, one of them is can the Semantic Web generate its own money beyond venture capital money that's going into the sector? Are there people paying real dollars for those applications that are actually useful?

And then you can split the problem into maybe two fields. Right now we are seeing a lot of efforts towards the B2B sector and that may be very well where the major problem lies that can be addressed by the Semantic Web because companies have a pressing need to really link data within and beyond their organization.

But beyond that, for me like the real criterion is going to be mainstream adoption by end users, by the actual consumer. And a bit like we saw with the Web, I mean, Google has really made it a mainstream, has made the - I mean Yahoo! and Google have made the entire mainstream specification just by making it so much easier for the average person to understand it and to find what they were looking for on it.

So it's a bit of a long-winded answer to that question. But at the end of the day, it's really is there money that is already being generated? At this stage there is some, clearly not enough so that the Semantic Web is mainstream. And then, who is going to really adopt that? And then there are other questions around that to adopt that.

**Paul:** I guess a question that springs to mind from that answer is to what extent does a consumer have to adopt anything Semantic Web, or will a consumer be going out and taking a product, like one of yours Alex, and simply using it because it does something they want to do? It solves a problem they have. And you just happen to have used some Semantic technologies behind-the-scenes to meet that purpose.

**Alex:** Yeah Paul, I absolutely agree. This was just on my mind as well as I was listening to this answer, which is, you know a lot of it is just about improving user experience.

And you know, a classic example of a computer automatically booking me a better vacation, you know, which I think is actually a bit of a stretch and were not going to be getting that with Semantic Web. But you know, just the whole point is if we can somehow leverage existing information or you know if we can re-annotate information in a different way to then improve the end-user experience, you know, people don't need to know the underlying technology is the Semantic Web.

There is an interesting question as to you know, how much of the classic technologies do you actually need? It's a fair question, given that if we are trying to sort of build it bottom up - meaning let's come up with a stack of technologies and then later come up with a value proposition for the end-user - that's pretty backward.

So if we were to look at the problem from the other angle, which is what is it that we are trying to improve? Are we trying to create a better search or are we trying to create a better browsing experience or a magic genie just solves all of our problems. So given the problem statement, there actually may be a solution that doesn't involve all the technologies about which we are talking.

I mean, I suspect that probably all of them are helpful. But my feeling is that you first have to define the problem and then look for the solution and not the other way around.

**Thomas:** I think Greg hit the nail on the head earlier on though when he said that a lot of the development of the Semantic Web stack or the Semantic Web idea conceptually would have been driven by people who had come from developing the Web of documents and really saw the limitation that this medium had in terms of linking specific bits of data together rather than just documents.

And I think we really need to understand that the problems have really come from the Web of documents and the way that the Semantic Web technology stack or the Web of data is really representing incremental improvements over that.

I think we have to be clear that we don't need to adopt this stack in one go, but there are small problems or small pieces that we can introduce one at a time. And then as we do that, then we will understand the limitations that presents. And then we might want to adopt something that's further up the chain.

I think that it would be wrong of us to see it as kind of an upside down process, which, Alex, I get the impression is what you are suggesting.

**Greg:** Yeah, I'm very aligned with all that actually. This is Greg again. I'm especially close to what, I mean to what all of you said, especially Alex in terms of what is going to be the killer application out there and what part of the Semantic Web is going to make it mainstream first.

I mean it's really like we are talking about the Semantic Web as a whole universe that everybody understands. I think this reflects very different definitions for different people. And I mean the basic thing would be the Semantic Web, as far as marketing might see it, is RDF. And then there are, I know, many other technologies around that. But that's really the tool that so far is getting the most traction in the market.

But it comes down to really what concrete application is going to use these linkages between information and make the best use of it.

And then, to Alex's point, what are the parts of the Semantic Web that have been, or the building blocks that have been created that are going to be re-utilized or utilized in order to improve the user experience? And I think there is still a lot of work that needs to be done in that area.

And we are seeing a lot of technologies maturing, but I think we are nowhere close to what the actual technology is going to be as far as the Semantic Web is concerned. I would say the Semantic Web is really an ecosystem in which we are going to evolve for many years going forward.

And I mean, it looks a bit like the prehistory and the technology has been really pushed forward. But the user experience - we have seen it with Twine for example, which has been the champion of trying to democratize the Semantic Web. They are still struggling to actually find what the actual value proposition is for the end user.

And there is still this mindset of looking at what Google did and what Yahoo! Did and what other applications did and try to do that better and I think Alex or Tom - I think probably Tom was saying. There is still this old mental model about what we can improve as far as the user experience is concerned.

And that is going to be huge. But I think what is going to be even bigger is a whole new set of value propositions around linking the data together and enabling new processes as far as the end user is concerned.

**Paul:** That linking of data together though, for a lot of traditional industries is going to require quite a significant mind shift, isn't it, because there is a recognition or a belief in a lot of traditional industries that the data, the information is their value proposition. And if you start exposing that for mixing and linking, you've undermined your existing business model. And unless you can see a new business model, you have undermined yourself.

**Mills Davis:** Paul, this is Mills.

**Paul:** Hi, Mills.

**Mills:** Hi. There are a couple of points. One, anybody who - I'm just going to put on my you know, as co-chair of the Semantic Interoperability Community of Practice in the Federal Sector hat - the government in the United States is spending between \$65 and \$70 billion a year in IT and has absolutely massive problems with information sharing, availability across you know boundaries.

And those boundaries can be within an agency, between agencies, across jurisdictions, between the U.S. and its allies and so on.

Anybody who is looking at exposing their data or trying to get out of silos, they know perfectly well what kinds of pain they are experiencing that makes them want to move in that direction. I don't think that the big mindset shift is that hard a proposition.

I mean, it was maybe three years ago because nobody understood the alternative, this idea of building a business ontology or some kind of model in the middle that could you know be itself an evolving structure.

But that idea, I think, has generally begun to take root as really just finding the other factors that de-risk making investment. You know, what buying category do we - what's the category of product for which we should budget? Who are the players? Where is the short list of people whose product solutions we want to buy? Who around here knows anything about this stuff that we can go to, to be the developer? I mean, the government outsources tons of stuff.

And you can kind of go on that way. But this is a classic problem of adoption of the critical mass of things that allow you to make a business decision. And what the Semantic Web scene has come to is that now these sorts of capabilities are starting to be on the road map of the larger IT companies.

I mean, you see Oracle you know pushing this stuff into their middleware. They are not doing it at the end-user product level. But they are doing it at kind of the plumbing level and for a particular reason, because that has the least visibility to the customer. They don't want to jeopardize an install base that can do lots of interesting linkups across databases by doing very complex hypercube SQL things that become a stack of paper that's an inch and a half high. It's terrible. Sparkle will let me do it so much easier. But there are hundreds of thousands of these people, and they already know how to do that.

So you have transitions, but there's nothing unique about semantic technology. What has to be educated and what has to be demonstrated and

what has to give the proof cases of adoption and so on.

**Thomas:** This is Tom Tague. I want to pick up on one thing Mills said, that there's nothing unique about the semantic web. You started off, Paul, with a question about organizations and concerns about exposing their data and essentially giving away the store. How do we make that happen? That's obviously a major concern for organizations whose lifeblood is information.

I thought that in your interview with Tim Berners-Lee, he had a great example: the bookstore example. The first website says "Here's our address, here's a pretty picture, come to our bookstore." Should we put up our product list? Should we put up our price list? Wherever we want to put it, Web 1.0 or Web 2.0, we're still in the middle of that argument and that discussion around content and rights to content and how much of it I'm willing to put out there.

We've had an explosion of content, but again, the bookstore still isn't publishing the books on the Web for free. We still haven't solved that problem, and I don't think the Semantic Web is going to magically address the issue of where we draw the line between publishing information about our products and capabilities, publishing snippets, and actually publishing the high value-added content. That's still a problem we're struggling with, and particularly for companies whose lifeblood is information.

**Alex:** This is Alex. Yeah, I completely agree with you on that, but I think there's an example of this already on the Web, which is Amazon. The Amazon e-commerce service has been around for five years and it's just an API into basically anything Amazon. Why they're doing this is because first of all, all the links point back to Amazon. They know they have a competitive advantage in terms of pricing and we know that people buy from Amazon, so they get a lot of good coming out of having an API.

A lot of the conversations around data interoperability are completely ancient

in the software industry. It's a Tower of Babel. We've lived through decades of this. The most critical issue between two computer problems is, how do we talk? And, you know, it almost doesn't matter what the language is, as long as we agree on the language, and the wide audience can agree on the language. So there are basically two distinct pieces. There's a piece of getting APIs in whatever format, and then there's a piece of getting adapters that can adapt these outputs to a common language. To me, as an engineer, when I look at this stuff, I see them as de-coupled. So more people on the Web look at interfaces. The adapters are not that hard to build. Even if the information is not an RDF, it can be transformed into RDF format. So that problem is solvable.

Beyond that, the bigger question is the incentives, like you were saying, for everyone to open. And I don't think that they are always there. I think for Amazon, there was a clear incentive, but I'm not certain that it's there for others.

**Tom:** This is Tom Heath. I've been thinking about this for the last few days trying to clarify my own views on the subject, and I've come back time and time again to the example of Amazon. I think, as well as thinking about the data and thinking about the API, I think we need to remember that there are identifiers in the question here and in the picture.

I think this is where something like the Semantic Web can have a more compelling story to tell over Web APIs, because there's the potential here for Amazon to expose identifiers for common concepts which many people want to refer to. If they do this with URI in a way that they don't currently do with, say, pieces of data which are embedded within the response from their e-commerce Web service, then there's the potential there for people to link into these URIs if they want to refer to a common author or a common book - Harry Potter or something like that. By doing that, then they potentially bring value into their applications at the data level as well as just the human interface HTML mashup level.

So there's potential there for them to say, "We are the authoritative source of identifiers for concepts in these whole range of domains." Then, if they move fast, then they've got a first-mover advantage and it makes it a lot harder for someone else to come along and create a bunch of identifiers for concepts in that area, because they won't necessarily get inward linking value.

**Paul:** Is that not what Freebase is trying to do?

**Tom:** Yeah. I guess maybe it is. I have an issue with the way they publish their identifiers. [laughs] But, that's a technicality.

But I think the value is there, the value of saying, "We are going to be around for a long time. We have great coverage in this area, we're reliable, we're authoritative." I think that's a great value proposition.

**Greg:** Yeah. I think I would just add one quick point. As far as an organization is concerned when it makes the decision of putting your data out there or keeping it to yourself, the main thing you're thinking about is, is this information more valuable to me as proprietary information, or is it more valuable to me as something that I'm sharing and where I'm going to reach people by putting it out there?

I think, again, the mental model is shifting progressively. It's not going to be black or white, and overnight everyone's going to put their information out there, obviously. But I think we see more and more organizations that are actually sharing more and more of what they actually do with other people out there. They see a return on their investment of their reputation through all the better information that they collect in turn. And then there's a question of timing as well. You can manage things by keeping information proprietary at first, and then releasing it.

So I think obviously there's a lot of legal work, as well, that needs to take place. That's lagging a bit behind. But things are being pushed in the right direction by the semantic web. The effects of the technologies that have been created are

creating changes in terms of how we look at the information. And I believe that in every field that we'd be early adopters and that we'd be organizations that would do it last, after everybody else has done it.

**Mills:** Paul, this is Mills again. I'd like to, for a second, come back to the question of mainstream. As I was listening to the comments when I tuned in, I was thinking to myself, "Well, who's mainstream, or what part of the mainstream are we talking about?" From my perspective, I tend to divide things out into several categories, four broad rubrics. One has to do with infrastructure issues of the Internet and so on. There, I see the recent announcement of Yahoo supporting the metadata and Semantic Web standards, and wanting to move in that direction for search.

Certainly letting you know one aspect of kind how the Net infrastructure is going to definitely start embracing representing knowledge about things separately from things, where they are expressions in particular languages.

But if you also look at what's going on with mobility, that's another kind of set of problems about the context switching for services, personalization of services and so on. All of the telecommunication companies and their associations and so on are all heavily committed now to semantic technology type research.

There are projects now too on kind of next-generation Internet. And why? Well, because kind of the fixed notions in the first round of the Internet require - it makes mobility tough. And to get to ubiquitous web, you fundamentally need a different class of technology that represents the knowledge separately in the way that different programs, different reasoners can get access to it.

Not like everybody for everything, but like applications that are needed. The second group, of course, is what we were talking about earlier - and Alex, I think, spoke well too - is consumer Internet. And the consumer Internet, from social to games to whatever, the quality of the user experience is going to be a very large

component of this.

And the question is, OK, what do semantic technologies ultimately add to this. One side of it is the SIOC kind of mashing up of things. Another thing might be architectures of learning, where we can get smarter UIs that can grow and learn and get new tricks and understand context better and so on. But that will be an area that will be driven by, does somebody see that this is valuable and interesting relative to something else?

And the third rubric, of course, was the enterprise. And that's a more - the shift there, I think, is already starting. In fact, at the Semantic Technology Conference, we're going to have a panel that's kind of just big players talking about where semantic technologies fit in their roadmaps. And certainly it's the interoperability, integration is one part of it.

It's also whether you view that from the standpoint of systems and processes or you view that from the standpoint of bringing together different kinds of information. These are huge problems. As well as the third dimension of that is, of course, the collaborative knowledge work, which wants to be able to relate knowledge, express it kind of at a document level with things that are schemas or structure levels, to behaviors.

So that is a very fertile area and one that, again, I think there's an increasing mind-share developing. Beyond that, there's the verticalization. The analogy I would draw here would be kind of to the computer graphics era 20 years or so ago, whenever it was.

But the verticals have particular classes of problems that can be solved with semantic technologies in ways that deliver capability or efficiencies or effectiveness that you can't get with other approaches. For example, one area would be governance, risk and compliance. Or anything that's event-driven, that requires you to mash together lots of different kinds of information, having kind of case workloads and so on.

There are tons of these, and there are sets of these applications. NASA can't go to Mars without robots that can figure out how to change their behaviors themselves because it takes too long to call home. So there's categories there, smart things, knowledge-centered engineering development disciplines versus document-centered ones.

There's just a whole encyclical of science, as opposed to just trying to search the documents. There will be a number of applications that will have the characteristics of putting knowledge to work, that take us beyond just the mashup stage, but actually to, "Let's put the theory and the information, the computer, and get more use out of it."

And also the kind of issue - I mean right now, for working at a certain scale, it's perfectly fine to talk about having some finite set of microformats that we can work with around a particular kind of problem. To do that at some scale, or to do that as a general proposition, then we'll probably find ourselves - of necessity - moving in the direction of semantic technologies no matter what.

**Paul:** That's an interesting point, which links back to the Yahoo! announcement as well. And I'm probably being silly asking this question this close to the end of the call, but the Yahoo! announcement supporting the regular pieces of the Web stack, things like RDF, it's also supporting a range of microformats. And some of the stuff that's come up in conversations outside this call, this feeling that the RDF and things like it are just too hard. Do you think that's true? Do you think maybe we should take the ideas but do it in microformat?

**Mills:** I'd love to weigh in on this one. Let me weigh in on this one. I think the analogy is, you go back to the computer graphics era, the first few years, it was all about graphics technology and then once the market started to happen it was about engineering graphics or publishing graphics or graphics for some particular kind of purpose.

And that's kind of the domain - the big win in the long run and big money is

going to be in the domains. Right? OK. Second point is that also, there were big debates at that time about which text format is the right one. And people went to standards committee meetings and they debated these things.

When PageMaker software came along, they said, "We'll take them all." And then eventually it was Word, because Microsoft won the game, or something. The point being that, we're at a stage now where we have lots of little work going on. Perfectly fine. They're structures, they're useable. They're little ontologies of useable expression, but by god, they're useful for something, for some people, for some purposes.

So we'll invest some energy in that. And I think we're at the stage now where, if you're at the scale of Yahoo, you're like the PageMaker of the Internet kind of thing, and you're just going to suck in what you need to suck in. And then later on it may sort out that RDF is, oh god, it turns out it's really cool.

For people who are using it, fine, we'll work with that. But right now, it's "pull the structure in." The game is, do you have the structure or don't you?

**Greg:** This is Greg. I just was going to add, I'd like to have the perspective of Thomas at Reuters on this, because part of going mainstream is to create that ecosystem where developers are going to create applications. And we've seen Facebook being very successful at creating an ecosystem for social networks and making it extremely easy to create applications.

And part of what's implied in your question is, how easy is it to create applications using RDF and everything? And I know Reuters is moving, taking the lead on that, and I'd like to ask Thomas what response you've seen to that. Are developers rushed to create applications already, or how is it going?

**Thomas:** Well, one thing we've learned is, most bloggers don't create applications. We need to incent some developers to create applications for bloggers, which in four or five weeks,

we'll be announcing the first round of those easy-to-use tools that bloggers can incorporate that will not only aid in embedding semantic metadata into their blogs, but in some other basic blogging capabilities.

I'm actually really excited about the Yahoo! thing. I don't know how useful it's going to be; we'll have to see how they choose to deploy the data that they harvest. What I do know is that it will create a rush of interest on the part of a lot of people to make sure they've got metadata for Yahoo! to harvest.

And it will all of the sudden drive, I think, a lot of interest in, "How do I get this stuff on my web page? How do I make it accessible to Yahoo!? What is it?" And just drive a lot of thinking about it. It's not the killer app. But up until this point, I'm sitting down and writing a blog entry, I'm Joe Blogger, I don't have a lot of motivation to think about semantic metadata or standards or RDF or whatnot. I'm not too attached to that, and I'm certainly not going to bother to spend the time.

So I'm quite excited about the Yahoo thing, as just a pull to generate interest. I'm also personally very pleased because the timing is excellent for us, where we're going to be deploying many of the tools to allow publishers to easily and at low cost embed at least the starter kit of semantic data into their publication.

So I think my simple answer is, I'm pretty excited about it.

**Mills:** Cool.

**Tom:** Paul, you made a comment before about general conceptions about the ease of publications of these different kind of formats, whether it's microformats, RDF/A, or RDF/XML or whatever. And I think this kind of taps into a bigger issue, that there's a lot of education that needs to go right now.

I think Tim was pretty much right when he - and that's been kind of echoed in some of the things that have been said here today - that from a deployment point of view, the Semantic Web really is ready now. But we need to see

some applications, and to see those applications, we really need to have people building them.

And I think there's a huge gap now for people who are able to understand the way these technologies work, to really distill that down and communicate well to developers who've got the grounding in conventional web technologies and server-side scripting and basically that sort of stuff.

So I think that's a really, really key issue now. And whether they're being taught about microformats or RDF/A or any other serialization of RDF, then the general drift towards educating people about the value of semantics and about linking different pieces of data together as well as just documents, is really, really key.

**Paul:** OK. I'm conscious of the time. So I'm going to lob in one last small thing and get some quick feedback on that. There's clearly plenty of other stuff for us to continue to discuss in the future.

I don't know if you all saw the announcement yesterday of this semantic hacker competition. If you're in the U.S. and want to use the APIs of a company called TextWise, you can win a prize of up to a million dollars for suggesting a semantic application.

Does this mean that the Semantic Web is a really exciting place to develop, or does this mean a company needs to pay people to use its API?

**Mills:** Let me respond to that, Paul, because I know a little bit about that announcement. Also, let me comment that with the research that we're publishing - we have, like, a free summary of it - when Richard MacManus did a review of it, within the next three weeks or so we had 3,000 downloads. And when I ran some numbers on that, or some analytics, just to see - we'd asked some survey questions and so on.

Over half of the people who - or the title, the specialty or whatever, their discipline - over half identified themselves as software developers. So that, I thought, was kind of interesting. It indicated a

breadth and a range of interest in these kind of topics.

In the case of the TextWise Semantic Hacker, which is an API for semantic discovery which uses a technology which they call Semantic Signatures, which is looking at a lot of text to figure out patterns of meaning. And they had, like, I don't know, they had 4,000 people hit their site and 1,500 signed up for their challenge. I'd gotten an email from Connie Kenneally, their CEO, which kind of blew them away. That was just from the TechCrunch announcement. They haven't even done anything else yet. And they got two or three planned submissions which they thought were really cool.

So I don't know, do you have to pay to use your API? I think there's a lot of people with a lot of ideas that are looking for ways to engage. What I liked about what they're doing and what I like about Calais, is it's bringing this ability for people to have deep linguistic and semantic knowledge right to the fingertips, in the case of a developer or an end user, such that we can start to...

It's not just the tag that I make up, it's my ability to take advantage of some depth of knowledge about what language means and get me to a more stable kind of notion of tags in a name space.

**Paul:** Anyone else?

**Mills:** I think we'll see other examples of this as well, other developers. Say, there's your top 10 list of natural language companies, they see it's now the time to get to the Web.

**Thomas:** Obviously it's interesting, and I believe it will get some press. I guess I believe it's fundamentally not the right direction right now to offer people a million dollars, a billion dollars, whatever, to say, what's the killer app? I think we're in an evolutionary mode right now. And instead of spending all of our time and money trying to come up with that killer app, I think we should be encouraging exploration and experimentation, and watching very, very carefully and seeing what evolves out of usage of components

of the semantic toolkit by tens of thousands of people.

**Mills:** Who's speaking, by the way?

**Thomas:** It's Tom Tague.

**Mills:** OK, Tom.

**Thomas:** I cannot underestimate sort of my belief in the power of the crowd to distill, and more importantly to discard, poor ideas. So I think these contests are great, and we're sponsoring contests ourselves.

**Mills:** Yeah. Well, [indecipherable], for example, used that a year or two ago.

**Thomas:** Yeah. And it's great and it certainly gets you some attention and it gets some interest. And Calais will be sponsoring additional contest in the future. I don't want the contests and the hype to take precedence over our efforts, and I think it's everyone on this call's efforts, to make sure that components of this technology are available for thousands of people to play with. I think that's what's important right now: get it out there, let them play, and then watch incredibly carefully for exciting innovative stuff.

And maybe we'll learn a little faster than we did in the migration from Web 1.0 to Web 2.0. That's what I'm excited about.

**Mills:** Yeah. No, I think that's great.

**Tom:** Yeah. Hear, hear, Tom. Nice. Nicely said.

**Paul:** Anyone else? Alex, do you have any thoughts on this?

**Alex:** Yeah. I think competitions is actually a really good idea. I don't know if you guys have followed what Netflix has done with their million dollar prize to improve their recommendation algorithm. There was a whole bunch of article recently about this. It turned out to be nearly an impossible problem to solve, but there's a lot of innovation and a lot of people going at it, so it's pretty positive.

I was a little bit turned off in the announcement, when it was stated that, you know... It's one thing to ask to build

an application that's cool and interesting, it's a different thing to come up with a business model for Semantic Web, and yet it's another thing to say, basically, "You have to assign your intellectual property to us."

So, I can think some of the elements of the way this contest is set up I'm not necessarily supportive of, but I do think that, bottom-line, a competition environment is positive and it fosters innovation.

**Mills:** Yeah, with challenges. Look what happened to Creedance Clearwater Revival, they lost all their rights to their songs. So, I think the person entering that kind of relationship needs to have some awareness, right?

**Paul:** Yeah. And of course anyone entering will be listening to this call, so now they know to read the small print.

**Tom:** Paul, before you move on, I think the money is a separate question. But I think the take-home message that I'd like people to hear is that developing for the Semantic Web and developing with the Semantic Web's stack and all the data that's out there is really great fun. And whether you get a million dollars for your idea or whether you get one dollar, it's a lot of fun and I really encourage people to get involved and to try it out.

**Paul:** OK. We were doing some sums in the office earlier this morning, we reckon a million dollars is two pounds seventy-five at the moment.

**Alex:** That hurts.

**Paul:** Or three euros.

Not quite. Unfortunately, we are just about out of time. I could keep going for ages on this, so we'll have to save some of this for next month. Before we finish off, I'm just going to quickly run down the list and ask each of you if you have any closing remark to make and I'll go the other way up the list to the way we did it before.

So Greg, have you got any final closing comment?



**Greg:** Yeah, I would just build on what was said about discovering what the market is going to do with these tools now. I think it's an exciting time for technology developers. I think it's also an exciting time for marketers now, as well. To develop new tools to really understand what the market wants to do with it. Because as part of the innovation process, we can see that a lot of the marketing tools are out there that just don't work very well.

You can't really survey your customers to know what do you want to do with Semantic Web, that doesn't work this way. I think there are also a lot of tools that are being developed on that side that are going to help with that discovery. Contests are one way to do it and there are many other ways. Overall, I'm very optimistic that we are reaching a tipping point. Is that mainstream? I'm not sure yet, I don't know, I hope so and I am very optimistic.

**Paul:** Thank you, Greg. Daniel, do you have a closing comment?

**Daniel:** I guess we've already seen the development tools already. I think the idea behind RDF is a lot simpler than people have been saying. Two documents with a relationship between, that's been easy to understand, that's standard web development and just taking us down a notch to objects with a relationship between. It's quite simple, really. It's just a case of learning another syntax. So, yeah, I encourage developers to have a look.

**Paul:** Thank you, Daniel. Tom Heath?

**Tom:** No, I've got nothing further to add other than just echoing Daniel's comments about the simplicity of the underlying model and Greg's comments about being optimistic. So, nothing further to add other than I'm in sort of agreement there.

**Paul:** Good, that's a good way to end with everyone agreeing. Thomas, are you going to agree to?

**Thomas:** I, of course, agree with everything everyone said. I think in terms of a closing comment, we'll see

where our audience lands. I really want to encourage enterprises to not over-complicate things. The Semantic Web can look like a religion that you need to adopt, you need to learn a whole new vocabulary, people behave strangely, and you feel stupid, or if you dig a little bit deeper it's a great toolkit.

We have an incredible advantage right now of compute, bandwidth, and storage costs asymptotically approaching zero. It's time to experiment. You don't need to replace what you're doing. You don't need big committees. Start building little pieces and just move quickly on them. Use the pieces of the toolkit that you like and ignore the others. I'd just love to see a lot of experimentation.

**Paul:** Thank you, Thomas. I agree wholeheartedly with that as well. Alex?

**Alex:** Yeah. I think, first of all, I would definitely agree with what Tom said. I think it's very elegant and really great and I think enterprises would definitely benefit from that. My personal closing remark is that I think that this is a totally exciting opportunity for all of us to really distill the essence of what Semantic Web patterns are, especially from all of these different things that we've touched on.

Wouldn't it be amazing if we, over the course of the next months, were able to identify these bits and patterns ranging from consumer benefits to enterprise benefits to tools and just build the stack for people, a kind of blueprint to understand what this base is and what the opportunities are. I think if we could manage to do that that would be a great service for everyone.

**Mills:** I agree.

**Paul:** A nice easy job for us to do over the next couple of months. Thanks, Alex.

**Alex:** We'll leave that to you, Paul.

**Paul:** Oh, thank you. Last but not least, Mills, and please don't give me any more work to do.

**Mills:** Well, I guess the closing comment from me is more that the Semantic Web technologies, and more broadly,

semantic technologies are ready now. There's ample evidence in the market, there are plenty of proof cases. So, the major challenge going forward is basically educational. One of the things we're committed to is to work with other companies and interested parties to begin the process as broadly and as deeply as we can, to help educate

There's a program that we're talking over now with 20 different companies and we're hoping to do more by the Semantic Technology Conference in May. We're calling it Semantic Exchange. The idea there is to do outreach, to put articles in magazines and blogs, and basically communicate the case studies and things like this, and do this to different parts of the Internet community and to do it over a sustained period of time.

I think what we need now is kind of some consistent education and so on. My call to action is let's join, not only to do our conversations with Paul - which I think what Talis is doing here is absolutely superb - but, let's also look for opportunities to educate in the mainstream. Share what we've learned and also help other people pick up on these technologies and apply them and have fun with them. I think by the end of 2008, we'll see a big shift in mindset across the industry.

**Paul:** Great, thank you, Mills. Thank you all very much for your time. I look forward to doing it again next month, thank you.

*Episode 2 of the Semantic Web Gang will be recorded on 17 April, and published via [semanticgang.talis.com](http://semanticgang.talis.com).*



## WEB 3.0? IT'S WHAT HAPPENS WHEN YOU HAVE LOTS OF WEB, LOTS OF *SEMANTICS*, & LOTS OF *SOCIAL INTERACTION* WITH IT!

The semantic wave is here. Over the next decade semantic technologies will drive trillion dollar global economic expansions, transforming industries as well as our experience of the internet.

A new landscape is developing rapidly. Semantic technologies are being applied broadly across all industry sectors from consumer internet to mission-critical enterprise applications, and for new infrastructure from multi-core chips, to mobility, to web-scale cloud computing.

Along with investors, technology developers, and enterprises in the public and private sector, you want to know where the opportunities are, which to attack, what to build, what to buy, and why.

Project10X's new study provides answers. **Semantic Wave 2008 Report: Industry Roadmap to Web 3.0 and Multibillion Dollar Market Opportunities** charts the next stage of internet evolution and the role of semantic technologies in next generation applications and business models. This is business intelligence you can't afford to be without. To thrive in the coming business ecosystem — learn more at [project10x.com](http://project10x.com)!