

# The social semantic desktop: Next generation collaboration infrastructure

Stefan Decker

*Digital Enterprise Research Institute (DERI), IDA Business Park, Lower Dangan, Galway, Ireland*

**Abstract.** This paper focuses on a next generation collaborative infrastructure. New developments in the computer science field, including semantic web, peer-to-peer, or natural language processing, will lead to new forms of collaboration and “social semantic information spaces”. These will have great impact on scientific communication.

## 1. Introduction

DERI is a research institute with about 60 researchers and is operating from Galway, Ireland. Sister institutes have been established in Innsbruck, Stanford and South-Korea. DERI is about to set up sister institutes in Canada, Sydney, Australia and China as well.

The main research interest of DERI is ‘web technology’ and ‘internet technology’ and the institute aims at establishing web semantics as the pillar of the internet infrastructure. In this paper we focus on issues in this context that we expect will be relevant in five to ten years.

## 2. Current problems in information management

Many of the present day problems (Table 1) in information management are problems that we are faced with on a daily basis. We travel around with laptops and in doing so, we experience insufficient information management and low level communication, because most communication is only through e-mail. Sending a document to 5–10 colleagues inside or outside an organisation multiplies the effort of management of this document times the number of people it is being sent to.

The information stored on a desktop is not interconnected with information stored on other desktops, so the desktop is a basic data silo. For example, information on the person that offers the document is not connected to an address book. And so the information is difficult to find and to maintain.

There is insufficient metadata support to data. For instance, everyone has to organise e-mail themselves. It would be much easier if a system would recognise that an e-mail is coming from a colleague of a certain project. Then, this e-mail can be automatically recognised and checked to give it priority compared to other e-mail, etc.

Also, there is an insufficient collaboration structure. If we want to exchange data stored on a mobile phone or laptop, we have to resolve to the time consuming process of exchanging the data physically. We hope that this problem goes away, so that we can establish an *ad hoc* collaboration infrastructure.

Table 1  
(Some of) today's problems . . .

- 
- Insufficient information management
  - Low level communication – everything is just e-mail
    - Sending a document to  $n$  people multiplies to overhead by  $n$
  - Information on the desktop is not interconnected (data silos)
    - Difficult to find and to maintain
  - Insufficient metadata support
  - Unsuufficient collaboration infrastructure
  - High cost of setting up/maintaining
    - Difficult to support ad-hoc collaboration
  - Specifically in European science
    - Large scale projects
    - Many project partners
    - Ad-hoc collaboration
- 

Specifically in European science over the last three years we have experienced that we have to deal with large scale collaboration projects with many partners. Often, we have to try to establish *ad hoc* collaboration. This is all without the luxury of being able to set up extra collaboration infrastructure in the future.

### 3. Today's problems solved and yesterday's ideas made possible

Many of today's problems described in Section 2 have been predicted by people in the past. Anyone who is familiar with the history of computer science is likely to know Vannevar Bush and his idea of the Memex. Or Doug Engelbart and his work in the late sixties on the mother of all demonstrating hypertexts and what could be done with organising information and his ideas of augmenting human intelligence.

Even Tim Berners Lee did not really envision the World Wide Web as a hypertext delivery tool, but as a tool to make people collaborate. We have to think ahead to try to come up with the next collaboration infrastructure. If the World Wide Web can transform the publishing industry, we can do it as well, and probably even more professionally.

What people like Vannevar Bush, Doug Engelbart or Tim Berners Lee tried to do, was to build a jet plane when the world had only just invented the parts to build a bicycle. In other words, it couldn't be done at that time. However, this has changed. What are we trying to come up with right, now is realising a "social semantic desktop" (Fig. 1).

So, why semantic? Firstly to make metadata and metadata items, ideas, entities, and topics all interconnected in a persons space. We could then share this information in social protocols; social not in a sense of a social party, but in the sense of individual to individual; or organisation to organisation; communication exchange. Now is the time to make this happen. In the past years, from the start of the World Wide Web until now, developments have taken place which make the dream that the before mentioned persons had, realisable. Now it can be done.

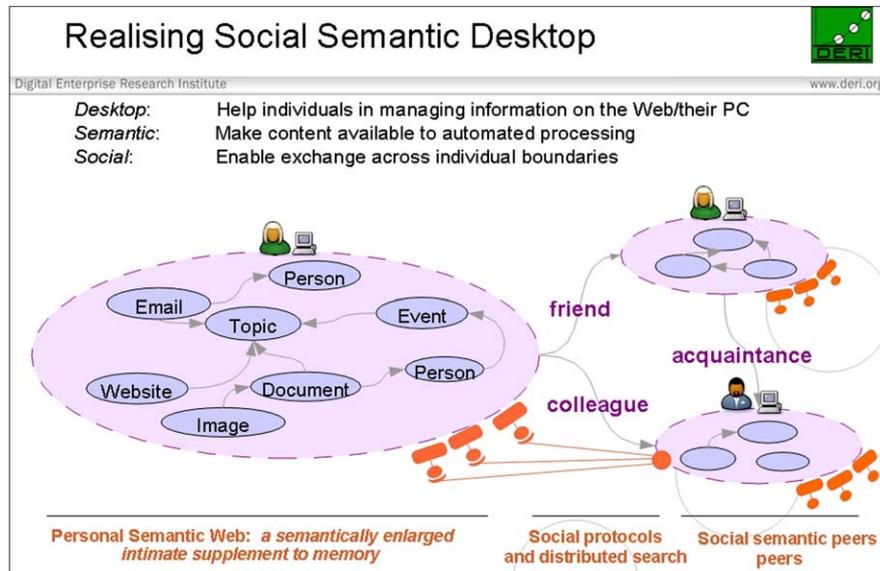


Fig. 1. Realising the Social Semantic Desktop.

#### 4. Social semantic information spaces

We have standardised metadata in the past few years, because these standards help us to formulate and to exchange data. We are now trying to develop a scalable distributed infrastructure. P2P-computing (which is now condemned because it is being used to download copyrighted material) enables us to interconnect without civilised infrastructures and to in-an-*ad-hoc*-way, communicate and exchange information.

We have knowledge articulation in the actual infrastructure, coming from desktop technology, but also from the web itself, for example from Wikis. Also, Natural Language Processing (NLP) is helping us deal with unstructured sources to a certain extent, so with decreased data. Last, but not least, there is a shift from pure technical matter and how to deal with sending information for NLP, to sending information about people to a community, or to an individual.

The Semantic Web was an effort that started about 60 years ago and was aimed to invent a standardised way to exchange metadata. This resulted in two standards. One is the Result Description Framework (RDF), originating from the library world, and the second is the Ontology Web Language (OWL).

Current desktop technology needs help in terms of metadata processing. Also from the web there are new entities dealing with metadata. Among these are Folksonomies, Wikis and Blogs, which are already well known.

Currently, there is also a level of research and development integrating the web and the desktop technologies, which is what we would like to call the "social semantic information space" (Fig. 2). With standards like ASS being interconnected and being put to use, a large space of references is created and the social semantic space becomes a living breezing entity on the web and on the Internet.

Specific examples that development is taking place would be semantic blogging, which is now being explored; integrating desktop data into a blog. Also, Wikis for personal information can be shared; which is a step further than what Wikis are able to do now.

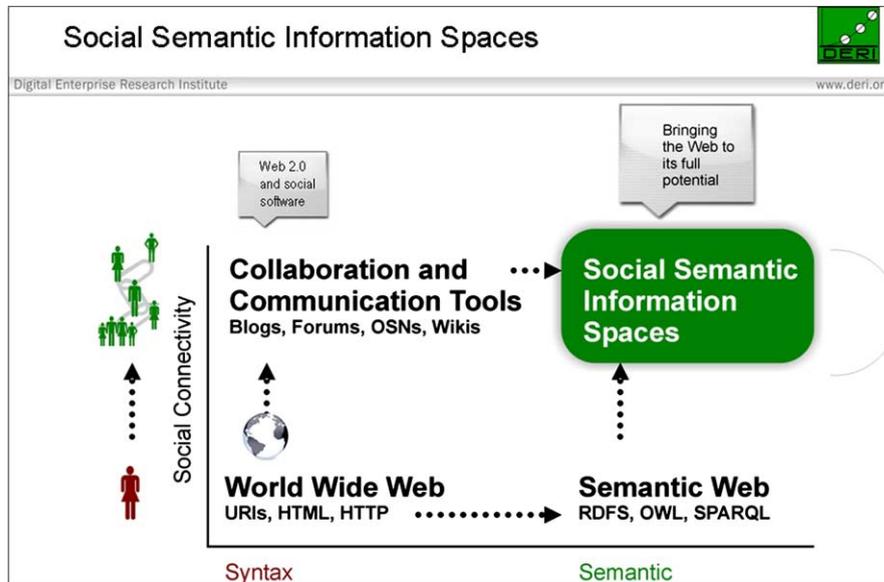


Fig. 2. The Social Semantic Information Space.

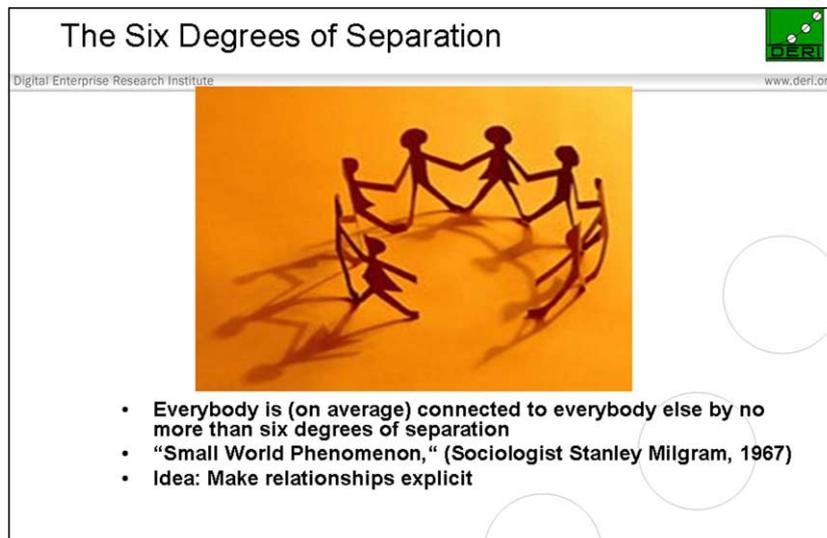


Fig. 3. The Six Degrees of Separation.

These semantic personal Wikis and the integration of Folksonomies and Desktop data are a very active research area. Folksonomies are being used to tag items from the web. Once this is done, it enables us to reuse tagged information. This function has been disregarded in the technical world in the last couple of years, but it is now reemerging. We are all collaborating, not only on our own island, but we are working with other people, on other islands. This is reflected in the work of the famous Stanley Milgram which states that on average everybody is connected to everybody else by no more than six degrees of separation (Fig. 3).

The idea of the six degrees of separation is now being used to make these relationships explicit and to see what can be done with them. An example of what can be done with this, are social networking sites. Millions of people worldwide are already members of sites that allow a person to enter personal information and through which you can see how close connected you are to any other person on the site, through social ties. There are networking sites not only for personal or social use, but also ones that are used in a commercial, businesslike, atmosphere. An example of the latter is linkedln.com.

#### 4.1. Business purposes/social purposes

The information entered on networking sites could be used by the person who entered the data to do useful things with, like spam filtering. The e-mail traffic to a person's inbox could be filtered and if the e-mail is not from this person's network, it's almost certainly spam. Microsoft has already developed a prototype of the MS Outlook software, which has a social network e-mail filter incorporated.

Another example of actual development the last couple of years, was P2P computing coming to the forefront. P2P computing is a very disruptive technology, because it takes out the middle person. It is very much in the spirit of the original internet, in the sense that there is no central control trying to enforce anything. The main feature of it is that everything has to be done based on social agreement and on protocols.

The social networking aspect is becoming increasingly important and is being used and exploited as previously mentioned. There is an explicit meta data format and there is a convergence into one final goal. The train has left the station. Unattached to one individual technical vision these different steps are currently happening and culminating to the final result. All that can be done right now, is to steer the train in a certain direction that you want it to go, and facilitate and speed-control the train a bit. But based on technology developments these things are happening.

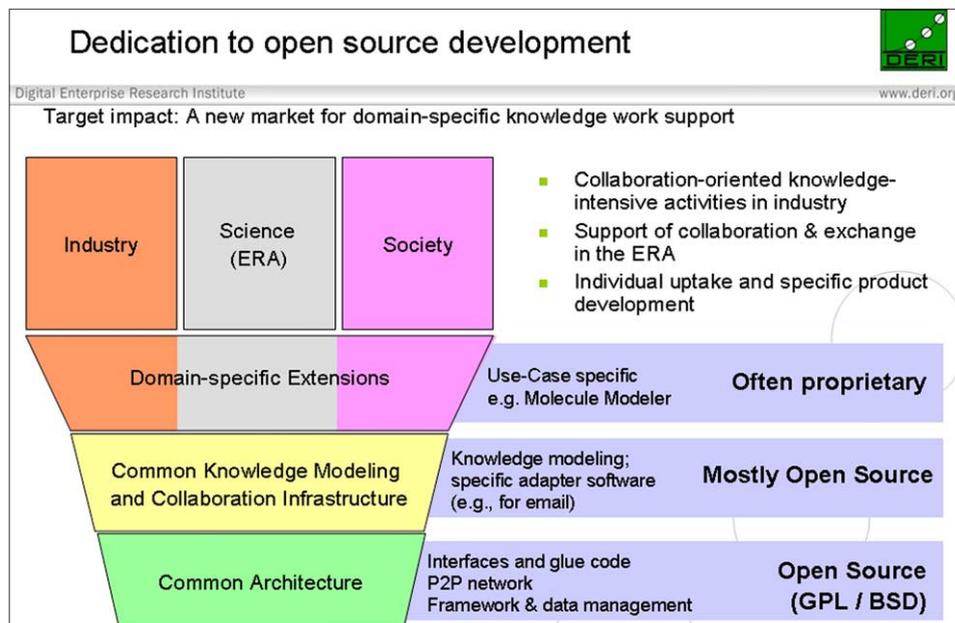


Fig. 4. Dedication to open source development.

One application that is being developed in a project of DERI is to interconnect scientists and networks and share information and publications directly with each other, based on the social ties, based on with whom scientists work on certain projects. So the community becomes part of a scientist's desktop. In fact, DERI has spent 17.5 million euro on projects in the last year, and on January 1st 2006 it started, in cooperation with companies like HP, IBM, SAP and a few research institutes, to develop crucial components of the infrastructure outline. Again, this is a development that is taking place anyway. We can only steer and try to facilitate certain developments.

Crucial and very much in the spirit of the original web is open source and open standard development (Fig. 4). Obviously, there is a business aspect to this, which is basically that we have a foundation, similar to the browser and the original server – which is open source-based on open standards, which we can just take and implement, but then we have propriety to domain-specific extensions.

One example that DERI is working on with the Institute Luis Pasteur, is the aim of Louis Pasteur to share information with fifty of their sister institutes, everywhere on the planet. This technology helps them to directly share information and work collaboratively on projects.

## **5. Conclusion**

The development of the social semantic desktop can be seen at [www.semanticdesktop.org](http://www.semanticdesktop.org). What we have tried to outline here will have an impact on scientific collaboration and publication. What exactly the effect will be, we still do not know. We are just building the technology and are thrown in it. What the outcome will be, what the business models will be, we can only guess. However, it is a step beyond e-mail, it enables a rethinking of scientific publications beyond text documents. Once we are able to point to other elements somewhere else, we can think of a publication as a collection of ideas, with just references to each other. Some of these ideas have been discussed at the APE 2006 conference. This “shaping” of the semantic web is taking place as we speak. The best way to predict the future is to invent it. It is very close to what we are all saying.