Personal infospheres

Jérôme Euzenat (INRIA & LIG) Philipp Cimiano (Bielefeld university)
John Domingue (The Open university) Siegfried Handschuh (DERI)
Hannes Werthner (TU Wien)

February 2, 2010

1 Data sharing and semantic web

Semantic web technologies are spreading to numerous applications: semantic desktop, semantic sensor networks, semantic web services, linked data, etc. The purpose of many of these applications is to collect data and to interpret them through interlinking.

On the side of users, the more bits of information are given away, the better the services they can have and the more the semantic web improves. Users want that the information provided to them is relevant to what they are doing or what they have to do. Hence, the bit of additional information that they give away (GPS coordinates, account information, etc.) should be used for providing contextual information.

However, users also want their data to be protected: they will accept to give more information, if they can control how and by whom this information can be accessed. This is a question of balance between services and control.

This, of course, is related to the raising concern of "privacy" which applies to the semantic web as well as the general web. In a system like the semantic web, allowing for connecting all the information, the lack of control is an obstacle to the adoption of the technology. However, semantic web technologies can also be used to tackle the issue.

The standpoint of this contribution is to consider how people could be encouraged to give away part of the information they hold so that it can be used to the benefit of a wider group of people. Only by providing more control to users, it will be possible to have a more positive and reasoned data sharing.

2 The continuum

The "privacy" problem is stated as if there were private data and public data. It is in fact ill-formulated. In reality, there are various kind of data with various degrees of privacy.

These degrees are related to:

- *what* data;
- *who* can access it;
- in what occasion (*when*): context;
- with what degree of detail (*how*).

Hence this problem is:

- multidimensional: it depends on the various dimensions mentioned above (what/who/when/how)
- gradual: instead of a simple private/public, the degree to which the data is exposable, e.g., I, family, family-and-friends, public.

3 Control to the people

Examples here are centered around individuals. However, agents in this model can be people as well as services (in the web service sense). Indeed, it can be understood that information can be communicated to such a service for providing its benefit, e.g., disclosing where someone is for delivery. Similarly, services may hold information that other users may want to be disclosed.

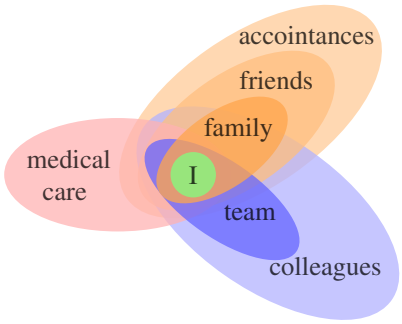

Figure 1: Typical spheres of relationships (who). Spheres are always centered on the individual.

Hence the important principle is that control over data must be decided by those who have data to disclose. This means that control over data is decided by individuals on their personal basis instead of by general policies, like social network software policies. Of course such policies may and certainly should play well with other policies such as corporate policies, but this is not the concern here. This may seem like access control. However, instead of rigid access control schemes, it is necessary to define access control in function of flexible concepts closer to the individuals (instead of dictated by operating systems and administrators).

The use of semantic technologies should provide more flexibility (because they can be extended) and more precision (because they can go to deep levels of details). So they are an ideal tool for giving people the control over their information.

4 The sphere model: approximating the continuum

There is a, from the user experience point of view, a continuum from data stored on one's disk, phone, social network accounts, the "cloud" and the web. There is also a seamless continuum between personal sensors, those in phones and computers, and mass sensors, CCTV cameras. It is very difficult to control this data: indeed, who has access to the MAC address of the WiFi card in a telephone when it is connected to a network?

It is difficult to prohibit or grant access to information in general (like distinguishing in the absolute that something is public or private), so we propose a general model for expressing the multidimensional space in which this data evolve. It is called the "sphere model" and is illustrated by Figure 1.

A sphere is a set of elements (person, event, context) identified by a name. It defines a compact area in a space centered at one point, e.g., the user. A system of spheres is a set of such spheres centered on the same point and partially ordered. So, basically these spheres are organised in a direct acyclic graph.

The spheres defined here are rather "personal spheres". They are in fact related to what is called "personal information". Of course, one could also consider corporate spheres or national spheres. Personal information on the semantic web is modelled by project such as Nepomuk. This means that starting from her semantic desktop, a user has already a lot of information available for defining spheres. Our goal is to use the information that matters to people (PIM data) to control the access of information that matters to people.

5 Who: spheres of relations

One can have various spheres: personal, colleague, people sharing a train compartment, people with whom one is currently writing a paper, etc. They are dynamic and can be subdivided into close family, work team, etc.

These kind of spheres are basically a group of agents which have the same rights with regard to (part of) one's data. Having these categories is convenient for granting or prohibiting access, based not on individual identities but on their belonging to a sphere.

Fortunately, these sphere descriptions are everywhere in address books and social networks. Hence they can easily be expressed in RDF and other semantic technologies. General groups can be defined through classes either in extension (the list of people in one's team) or in intension (all the people in one's address book, all those working in the same company as oneself). One particular individual may be identified directly (through a URI) or indirectly through her role (like "my boss").

Identifying people can be achieved by specific technologies like foaf+ssl (but this is out of focus here).

6 What: characterising data

However, this is not all: data is also organised in such spheres. A calendar contains various types of events: personal, work, sport events. Not only access may be granted depending on the sphere someone belongs to, but it may depend on the sphere in which this information belongs to.

Of course, one's music band members will have access to the information about gigs, locations, etc. One's family may have it as well, but colleagues can only know that the person is unavailable. On the opposite, the band members will only know from business schedule that one is not available for practicing at some periods, while colleagues will have more precise information.

In terms of semantic technologies, these items can be characterised by classes and properties (sport events, family gatherings, work meetings) and filters about what is accessible can use designed in SPARQL or true rule languages.

7 How: granularity

Not only access may be granted depending on the sphere someone belongs to, but granularity of what is disclosed may be granted depending on the spheres people are in.

Granularity offers a gradual access to information over time and over spheres. Consider the following pieces of information describing the same event:

- "I am not available"
- "I am in Karlsruhe from October 26 to October 30"
- "I am working on a paper about S-match and algebra of relations in room 452 of Novotel with Paul, on Monday, October 28th, from 17h to 19h."

They are three representations of the same thing at three different granularities. Moreover, the granularity changes apply in three different dimensions:

- spatial,
- temporal,
- thematic.

The point with granularity is to be able to present information at different levels of detail so that only what is useful for the reader is available. This adaptation aims at efficiency by reasoning at the required level, and, may also be aimed at protecting privacy.

This can be defined through rules and/or views which alter information rather than simply filtering it, like: for sports events, colleagues can know that I am not available but in town.

8 When: characterising context

We call context information that characterises the situation, e.g., current task, location, time, purpose of disclosure, etc.

One may not want to disclose precise information about her localisation at anytime, but for a colleague in the surroundings, e.g., at the same conference, looking for her, this information may be useful to deliver. However, in particular contexts, e.g., an extraordinary event has happened, very important information (like medical data) may also need to be disclosed.

This information may of course be expressed with semantic web technologies. But it can also be expressed as a sphere: a travelling context is a narrower context than being simply awake and a wider context than driving to work.

In terms of semantic web technology, the disclosure can again be processed by filters which will consider if the context falls within a sphere in which the information can be communicated.

However, for characterising the context, one needs to ask the requester for information about his or her context, that he may not be willing to disclose. Hence, disclosure of information will be made on the basis of negotiation between two parties when one has to disclose the reason why some information is wanted, in order to have this information delivered. This negotiation will use exactly the same techniques as above. However, specific protocols may be needed for guaranteeing that the process converges to an acceptable result.

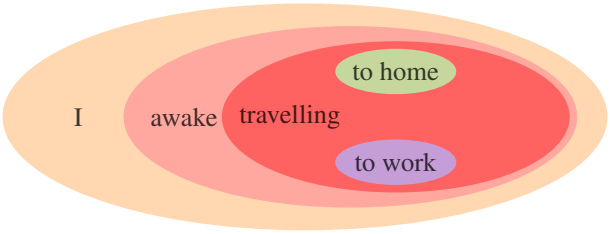

Figure 2: Context and granularity as spheres.

9 Conclusions

Semantic web technologies provide a very good basis for a more flexible and precise control over disclosed information. For this purpose, we introduced the abstract notion of spheres which symbolise the multiple and contextual aspects of situations.

We have identified five challenges for implementing this sphere approach:

- How to specify spheres and access policy;
- How to evolve spheres and create ad hoc spheres (like the train compartment);
- How to navigate between multiple context embeddings;
- Interaction between spheres and context.
- How to negotiate information access (on technical, social, or legal grounds);

This short presentation focusses on data access. However, spheres can be used for other purposes. For instance, users can use the same spheres or other ones for ascribing trust to information once they know their provenance (which was another important topic at the seminar).

10 Related projects

There are already several project that can be considered as providing some ground or early experimentation for the principles presented here:

Nepomuk by developing the concept of Semantik desktop and providing a semantic version of PIM categories has paved the way to the introduction of spheres[1].

Iyouit is a project of DoCoMo labs experimenting with the exploitation of automatically extracted context information from mobile phones. It does already use semanticised PIM information in order to implement access control to the data[2].

Persist has developed the notion of personal smart spaces which tries to provide services to users based on their contexts and preferences. When two users encounter, their smart spaces can interact and exchange information in a controled way[3].

[1] http://nepomuk.semanticdesktop.org/
[2] http://www.iyouit.eu/
[3] http://www.ict-persist.eu/