Semantic Web: a personal privacy perspective

S. Atkinson[1], P. Jagodzinski[2], C. Johnson[2] & A. Phippen[1]
[1]*Network Research Group, University of Plymouth, Plymouth, UK*
[2]*University of Plymouth, Plymouth, UK*

Abstract

This paper describes an ongoing investigation into the potential impact that the Semantic Web could have on the personal privacy of individuals. The argument presented is that personal privacy should become part of the underlying architecture and design of the Semantic Web in order to limit the vulnerability of individuals.

The current structure of the Semantic Web does not explicitly address the issue of the control of personal data, which could in turn lead to individuals being placed in positions of vulnerability. Personal privacy is seen as a major element of vulnerability and has caused some dilemmas from the legal perspective, balancing those who would hide their wrongdoing behind privacy against those who would be exposed to harm if there were no privacy.

Issues of privacy are explored from the perspective of three groups of individuals: survivors of domestic violence; people who are not IT specialists; and teenagers. The impact of current web technologies on these groups is explored in order to gain insight into how the Semantic Web might be adapted.

The paper concludes with a proposal for further research into the development of a Semantic Web tool which will aim to support individuals in identifying the potential threats to their privacy that arise from each sortie into cyberspace. Particular emphasis will be given to the issues faced by vulnerable groups and individuals.

Keywords: Semantic Web, privacy, vulnerability, domestic violence, teenagers.

1 Introduction

The Semantic Web purports to be a solution to the frustrations of end-users when searching for information. Data is to be marked in such a way that computers are able to make use of it in a more intelligent fashion.

WIT Transactions on Information and Communication Technologies, Vol 36, © 2006 WIT Press
www.witpress.com, ISSN 1743-3517 (on-line)
doi:10.2495/IS060411

Today, more and more personal data is being collected and made available through the Internet. This is occurring through traditional means (for example the availability of public records online) but also as a result of the increasing pervasiveness of the Internet. For example, mobile phones, fridges and other household devices are now being designed so that they can be controlled over the Internet from a computer in a remote location. The ability to gather and aggregate data from a wide range of sources allows the creation of personal profiles which in turn raises issues of personal privacy.

Solove [1] has described the general impact of technology as an "Architecture of Vulnerability", and proposed that technology places people in situations of risk that they are powerless to mitigate. Whilst Solove highlights the problem of identity theft, other issues such as stalking and harassment have also been noted [2].

The purpose of this research is to explore how the Semantic Web might address these issues of vulnerability. The control of personal information is seen as an important part of mitigating the risks to privacy and our research seeks to embed such controls within the design and architecture of the Semantic Web.

The paper is structured as follows. Section 2 examines the structure and capabilities of the Semantic Web, together with the possible implications for personal privacy. Section 3 then reviews the current body of thought on personal privacy, including the relationship with vulnerability and the current legal dilemmas. Section 4 describes a pilot study which attempts to identify the vulnerabilities experienced by three groups, namely survivors of domestic abuse; everyday individuals not IT professionals; and teenagers. Finally Section 5 outlines our direction for further research suggested by our pilot study.

2 Semantic Web

The Semantic Web is the term given to a vision proposed by Tim Berners-Lee [3] where computers are able to process information found on the web, irrespective of its format. The Semantic Web thus has the benefit of making data interoperable, allowing the more widespread sharing and automated aggregation of information. However, this then has the potential to overcome the historical difficulties in creating personal profiles – i.e., the high cost in gathering together pieces of personal information held in different places. As yet, no consideration has been given to the implications for privacy.

The Semantic Web framework is created from a layering of protocols as illustrated in figure 1.

The standards of XML, RDF and OWL have been issued by the World Wide Web Consortium (W3C) as recommendations [5], and work is in progress into how logic, proof and trust are to be implemented.

Trust is defined by the Oxford English Dictionary (OED) as a belief in the reliability, or truth of a matter. Within the e-commerce context Guerra et al. [6] describes three components to trust: privacy, identity and security. However, the current focus for research [7] within the context of the semantic web echoes the OED definition, i.e., that of trustworthiness and credence given to data. In

particular, privacy together with the use and control of personal information is not an integral part of the semantic web infrastructure, but perhaps a sideline to trust. This leads to the possibility that privacy issues may be overlooked, or not properly implemented.

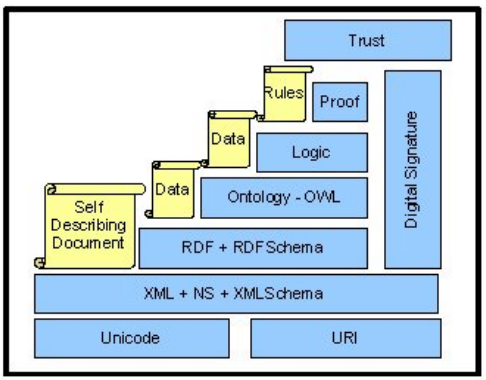

Figure 1: Suggested architecture for Semantic Web (Adapted from [4]).

This would appear to confirm the concerns raised by Solove [1] and Garfinkel [8] that technology not only places people in vulnerable positions but is also naturally privacy-invasive.

3 Personal privacy

Our focus is on the privacy of the individual (as opposed to that of organisations and businesses). The definition of privacy has adapted from originally being the "right to be left alone" [9], to explicit consideration of how easily information flows between entities [1, 8]. As noted in [10], it is clear that the amount of personal information released is linked to levels of vulnerability. Solove [1] has also observed that legal, social and technological elements of life combine to create the context where privacy problems can arise.

These problems are exacerbated because information is a highly valuable (business) commodity, and there are few controls on its dissemination or use. One example is the increasing number of public records now available through the Internet allowing many pieces of information to be gathered about people [11, 12, 13]. A second example would be the increasing trend for location tracking of mobile phones. This clearly puts at risk the phone-user's privacy, but significantly it is a risk that is difficult to mitigate. The user cannot gain information from their network provider to discover if their mobile phone is being tracked by a third party tracking service. They have to contact all the possible tracking services themselves to make this discovery.

European Law attempts to address the issue of personal privacy – specifically Articles 8 and 10 of the Convention of Human Rights enshrine the right to privacy of the individual and set out the principles for freedom of speech. However making use of the law to redress harm done relies upon individuals knowing who to sue, and being empowered enough to do that. The international nature of the Internet adds difficulties. The litigant must ascertain exactly where the infringement took place, where the defendant is located, and what law is relevant [14].

On the other hand, Articles 8 and 10 allow the state to intervene should there be a threat to the "economic well-being" of the country [15]. Invasions of privacy are becoming commonplace in the attempt to combat the fear of crime and terrorism. The purpose is to remove the protection of privacy for those who would harm others [16]. These invasions include the blanket DNA testing of those arrested for any offence [17], CCTV cameras observing public spaces [8], and the proposed introduction of identity cards [18].

4 Pilot study

A pilot study was conducted to explore fully the privacy context within which people find themselves. Special focus has been placed upon groups who may find themselves in vulnerable positions [16], specifically survivors of domestic abuse; everyday individuals who are not IT professionals; and teenagers. Once an understanding of vulnerability has been gained, steps can be taken to mitigate the risks.

4.1 Group one: survivors of domestic abuse

Women who flee abusive relationships are at most risk when they leave [20] and thus privacy (specifically remaining hidden) is paramount. Those who remain within abusive relationships will find technology increasingly used for power and control over them [21].

Semi-structured interviews were held with service providers to survivors of domestic abuse. These highlighted the threats from mobile phones; information given out from utility providers or government agencies to the wrong person; and websites that show postcodes gleaned from P.O. Box numbers.

4.2 Group two: everyday individuals who are not IT professionals

A short experiment carried out with ten people explored their perceptions of vulnerability and (with their permission) sought to gather together publicly accessible information about them. Most people were unsure about what information was held on them, and three individuals were very concerned about what might be held.

Date of birth was found to be the key to finding out more public information, for example the mother's maiden name and access to certificates. Some of the social networks and genealogical sites helped to find date of birth and mother's

maiden name, but on the whole this was unobtainable. Concern was raised about the ease with which the mother's maiden name could be found, the possibility of identity theft, and that the use of jargon throughout the Internet hid the potential to exploit vulnerability. Two people expressed concern that the Internet was creating a society that intruded on the personal privacy of individuals. One person felt it necessary to take action by ringing their bank and changing their identifying data from their mother's maiden name to something less likely to be discovered.

Public records caused concern in that the electoral roll could be combined with Land Registry information to infer that a woman is likely to live alone. This made one participant feel vulnerable to attack or burglary.

4.3 Group three: teenagers

Many young people wholeheartedly embrace technology and may not be completely aware of some of the risks. Social networks like www.bebo.com and www.faceparty.co.uk encourage the divulging of personal details, yet provide little control over who has access to that information. Magid [22] proposes that teenagers are most at risk from predatory behaviour.

A small survey of teenage users of Microsoft Messenger was carried out, and to which 32 people responded. Of these, 17 had posted photographs of themselves on their profiles. Most people were happy to make their email address public, but were only comfortable in divulging address and phone number if they knew who it would be given out to. Overwhelmingly, 30 people were happy to have their gender known. 19 people did not regret putting personal information into their profile.

5 Further research

The research carried out to date has highlighted where vulnerability has been created. The ease with which information is obtained has created a situation where people are anxious about the risks created by technology and feel powerless to mitigate these risks. Examples are location based tracking of mobile phones; postcodes being shown on web sites; and inferences being made about single females.

Combining these vulnerabilities through shared data and the power of Semantic Web reasoning tools can of course lead to even greater vulnerabilities, and causes great concern. However, there is merit to be had in addressing this issue face on, specifically by handing back control to the individual in order to allow them to mitigate their risks for themselves.

To address this vulnerability, the next stage of the research will be to create, implement and evaluate a Semantic Web tool which will allow an individual more control over their personal details. This application will be embedded within a browser and will allow the user to identify where the threats to their privacy might arise. The tool will encompass the following requirements:

- Detection of when personal information is divulged.
- Recording of where personal information is sent.
- Monitoring of privacy policies where personal information has been divulged.
- Background checks on where the website is hosted.
- Monitoring of other personal information found on the Internet.

References

[1] Solove, D.J., *The Digital Person*, New York University Press: New York, pp 115 – 119, 2004.
[2] Spitzberg, B.H., Hoobler, G., Cyberstalking and the Technologies of Interpersonal Terrorism. *New Media & Society*, **4(1)**, pp71-92, 2002.
[3] Berners-Lee, T. with Fischetti, M., *Weaving the Web*, Texere Publishing: London, 2000.
[4] Berners-Lee, T, Semantic Web on XML, www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html.
[5] Semantic Web Activity Statement, W3C, http://www.w3.org/2001/sw/Activity.
[6] Guerra, A.G., Zizzo, D.J., Dutton, W.H and Peltu, M., Economics of Trust in the Information Economy: Issues of Identity, Privacy and Security, *Oxford Internet Institute, Research Report No 1*, April 2003, http://www.oii.ox.ac.uk/resources/publications/RR1.pdf.
[7] Golbeck, J., Parsia, B., Hendler, J., Trust Networks on the Semantic Web, in *Proc. of Cooperative Intelligent Agents, Helsinki, Finland,* eds. M. Klusch, S. Ossowski, A. Omicini, H. Laamanen, Springer-Verlag: Berlin, pp238-249, 2003.
[8] Garfinkel, S., *Database Nation*, O'Reilly Associates: Sebastopol, CA, 2000.
[9] Warren, S and Brandeis, L., The Right to Privacy, *Harvard Law Review*, **4**, pp193, 1890.
[10] Dinev, T. and Hart, P., Internet Privacy Concerns and their Antecedents - Measurement Validity and a Regression Model, *Behaviour and Information Technology*, **23**, pp413-422, 2004.
[11] Births, Marriages and Deaths Register, www.1837online.com.
[12] The UK Electoral Roll, www.electoralrolluk.co.uk.
[13] Companies House, www.companieshouse.gov.uk/index.shtml.
[14] Valongo, K., *Your Privacy on the Internet*, Internet Handbooks: Plymouth, 2000.
[15] Colvin, M., *Developing Key Privacy Rights*, Hart Publishing: Oregon, 2002.
[16] Schoeman, F., (ed), *Philosophical Dimensions of Privacy*, Cambridge University Press: Cambridge, 1984.
[17] Davies, S., *The Death of Privacy: A Personal View*, BBC: Video Cassette, 2000.

[18] ID card pilot scheme under way, 10 Downing Street, London, www.number-10.gov.uk/output/page5701.asp.
[19] Raab, C.D and Bennett, C.J., Distribution of Privacy Risks: Who Needs Protection, *Information Society*, **14**, pp263-274, 1998.
[20] Domestic Violence Statistical Factsheet 2002, Womens Aid Federation of England, www.womensaid.org.uk/dv/dvfactsh2002.htm.
[21] A High Tech Twist on Abuse: Technology, Intimate Partner Stalking and Advocacy, VAW Online Resources, Minesota, www.mincava.umn.edu/documents/commissioned/stalkingandtech/stalkingandtech.html.
[22] Teen Safety on the Information Highway, NCMEC, www.safeteens.com/safeteens.htm#Guidelines_for_Parents_0.