Semi-Automated Seeding of Personal Privacy Policies in E-Services

Yee, George; Korba, Larry

Publisher's version / Version de l'éditeur:

Encyclopedia of E-Commerce, E-Government and Mobile Commerce [Proceedings], 2006

National Research Council Canada
Conseil national de recherches Canada

Institute for Information Technology
Institut de technologie de l'information

NRC · CNRC

*Semi-Automated Seeding of Personal Privacy Policies in E-Services **

Yee, G., Korba, L.
2006

* published in the Encyclopedia of E-Commerce, E-Government and Mobile Commerce. 2006. Publisher: Idea Group, Inc.. NRC 48235.

Canada

Routing Number: Manuscript 230
Submission Title: Semi-Automated Seeding of Personal Privacy Policies in E-Services

George Yee
Institute for Information Technology
National Research Council Canada
1200 Montreal Road, Building M-50
Ottawa, Ontario, Canada K1A 0R6
george.yee@nrc-cnrc.gc.ca
Phone: 613-990-4284
Fax: 613-952-7151

Larry Korba
Institute for Information Technology
National Research Council Canada
1200 Montreal Road, Building M-50
Ottawa, Ontario, Canada K1A 0R6
larry.korba@nrc-cnrc.gc.ca
Phone: 613-998-3967
Fax: 613-952-7151

Semi-Automated Seeding of Personal Privacy Policies in E-Services[1]

George Yee and Larry Korba

INTRODUCTION

The rapid growth of the Internet has been accompanied by a proliferation of e-services targeting consumers. E-services are available for banking, shopping, learning, government online, and healthcare. However, each of these services requires a consumer's personally identifiable information (PII) in one form or another. This leads to concerns over privacy.

In order for e-services to be successful, privacy must be protected (Ackerman, Cranor, and Reagle, 1999). An effective and flexible way of handling privacy is management via privacy policies. In this approach, a consumer of an e-service has a personal privacy policy that describes what private information the consumer is willing to give up to the e-service, with which parties the provider of the e-service may share the private information, and how long the private information may be kept by the provider. The provider likewise has a provider privacy policy describing similar privacy constraints as in the consumer's policy, but from the viewpoint of the provider, i.e. the nature of the private information and the disclosure/retention requirements that are needed by the e-service. Before the consumer engages the e-service, the provider's privacy policy must match with the consumer's privacy policy. In this way, the consumer's privacy is protected, assuming that the provider complies with the consumer's privacy policy. Note

that policy compliance is outside the scope of this work but see Yee and Korba (July 2004).

Initial attempts at conserving consumer privacy for e-services over the last few years have focused on the use of web site privacy policies that state the privacy rules or preferences of the web site or service provider. Some of these policies are merely statements in plain English and it is up to the consumer to read it. This has the drawback that very few consumers take the trouble to read it. Even when they do take the time to look at it, online privacy policies have been far too complicated for consumers to understand and suffer from other deficiencies (Lichtenstein, Swatman, and Babu, 1999; Jensen and Potts, 2004)). Still other privacy policies are specified using P3P (W3C) that allows a consumer's browser to automatically check the privacy policy via a browser plug-in. This, of course, is better than plain English policies but a major drawback is that it is a "take-it-or-leave-it" approach. There is no recourse for the consumer who has a conflict with the web site's P3P policy, except to try another web site. In this case, we have advocated a negotiations approach to resolve the conflict (Yee and Korba, Jan., May, 2003). However, this requires a machine-processable personal privacy policy for the consumer.

We assume that providers in general have sufficient resources to generate their privacy policies. Certainly, the literature is full of works relating to enterprise privacy policies and models (e.g. Karjoth and Schunter (2002), Barth and Mitchell (2005)). Consumers, on the other hand, need help in formulating machine-processable privacy policies. In

addition, the creation of such policies needs to be as easy as possible or consumers would simply avoid using them. Existing privacy specification languages such as P3P, APPEL (W3C; W3C, 2002), and EPAL (IBM) are far too complicated for the average Internet user to understand. Understanding or changing a privacy policy expressed in these languages effectively requires knowing how to program. Moreover, most of these languages suffer from inadequate expressiveness (Stufflebeam et al, 2004). What is needed is an easy, semi-automated way of seeding a personal privacy policy with a consumer's privacy preferences. In this work, we present two semi-automated approaches for obtaining consumer personal privacy policies for e-services through seeding. This paper is based on our work in Yee and Korba (2004).

Section "BACKGROUND" examines related work and the content of personal privacy policies. Section "SEMI-AUTOMATED SEEDING OF PERSONAL PRIVACY POLICIES" shows how personal privacy policies can be semi-automatically seeded or generated. Section "FUTURE TRENDS" identifies some of the developments we see in this area over the next few years. We end with "CONCLUSIONS".

BACKGROUND

We have been able to find only two other authors who have written on the derivation of personal privacy policies. Dreyer & Olivier (1998) describe a tool called the "Privacy Workbench" for creating and analyzing privacy policies. However, it is not clear from their paper how one comes up with the privacy policy in the first place, as it seems to just appear followed by a description of how the tool can perform conflict analysis. It is a

model-based rules inference approach for validating an existing privacy policy. More importantly, Privacy Workbench is a tool for a programmer, as it is far too complex for the average consumer to understand and use. Snekkenes (Snekkenes, 2001) wrote about the derivation of personal location privacy policies for use with a location-based service, e.g. E911 emergency location service in the United.States. Snekkenes' view is that "individuals should be equipped with tools to become in the position to formulate their own personal location privacy policies". This author provided concepts as well as fragments of a language for formulating personal location privacy policies. Unfortunately, the language presented can only be understood by programmers and not the average consumer. Our approaches for generating personal privacy policies are not model-driven or service specific and have been designed for ease-of-use by the average consumer.

Privacy Legislation and Directives

Before we can consider how to seed a personal privacy policy, we need to know what such a policy should contain in terms of privacy provisions. We use privacy legislation to obtain what must be specified in a personal privacy policy. Therefore, this gives a minimum policy in the sense that all elements required by law have been specified, but additional provisions can be included at the discretion of the consumer.

In Canada, privacy legislation is enacted in the *Personal Information Protection and Electronic Documents Act* (Department of Justice; Government of Canada) and is based on the Canadian Standards Association's *Model Code for the Protection of Personal*

Information (Canadian Standards Association) recognized as a national standard in 1996. This Code consists of ten Privacy Principles (Canadian Standards Association) that for convenience, we label as CSAPP. Data privacy in the European Union is governed by a comprehensive set of regulations called the Data Protection Directive (European Union). In the United States, privacy protection is achieved through a patchwork of legislation at the federal and state levels. However, privacy has been recognized as a constitutional right and there exists a highly developed system of privacy protection under tort law for the past century (Industry Canada).

We seek attributes of private information collection using CSAPP as a guide. We use CSAPP because it is representative of privacy legislation in other countries and has withstood the test of time, originating from 1996. We will then apply these attributes to the specification of privacy policy contents.

Personal Privacy Policy Content Based on Legislation

Based on an exploration of CSAPP (Yee and Korba (2004, 2005)), the contents of a privacy policy should, for each item of PII, identify a) *collector* - who wishes to collect the information (for consumer policies only), b) *what* - the nature of the information, c) *purposes* - the purposes for which the information is being collected, d) *retention time* – the amount of time for the provider to keep the information, and e) *disclose-to* – the parties to whom the information will be disclosed. Figure 1 gives 3 examples of consumer personal privacy policies for use with an e-learning provider, an online bookseller, and an online medical help clinic. The first item in a policy indicates the type of online service for which the policy will be used. Since a privacy policy may change

over time, we have a *valid* field to hold the time period during which the policy is valid. For a consumer policy, the proxy field holds the name of the proxy if a proxy is employed to provide the information. Otherwise, this field has the default value of "no". For a provider policy, the proxy field has a default value of "yes" indicating that the consumer can use a proxy to provide the information. Otherwise, this field has the value "no".

Policy Use: *E-learning* Owner: *Alice Consumer* Proxy: *No* Valid: *unlimited*	Policy Use: *Bookseller* *Owner:* Alice Consumer Proxy: *No* Valid: *June 2003*	Policy Use: *Medical Help* *Owner:* Alice Consumer Proxy: *No* Valid: *July 2003*
Collector: Any *What:* name, address, tel *Purposes:* identification *Retention Time:* unlimited *Disclose-To*: none *Collector:* Any *What:* Course Marks *Purposes:* Records *Retention Time:* 2 years *Disclose-To*: none	*Collector:* Any *What:* name, address, tel *Purposes:* identification *Retention Time:* unlimited *Disclose-To*: none	*Collector:* Any *What:* name, address, tel *Purposes:* contact *Retention Time:* unlimited *Disclose-To*: pharmacy *Collector:* Dr. A. Smith *What:* medical condition *Purposes:* treatment *Retention Time:* unlimited *Disclose-To*: pharmacy

Figure 1. Example Consumer Personal Privacy Policies

A personal privacy policy thus consists of "header" information (policy use, owner, proxy, valid) together with 5-tuples or privacy rules

<collector, what, purposes, retention time, disclose-to>

where each 5-tuple or rule represents an item of private information and the conditions under which the information may be shared. A personal privacy policy therefore consists of a header plus one or more privacy rules.

SEMI-AUTOMATED SEEDING OF PERSONAL PRIVACY POLICIES

A semi-automated seeding (or derivation) of a personal privacy policy is the use of mechanisms (described below) that may be semi-automated to obtain a set of privacy rules for a particular use (see above). We present two approaches for such derivations. The first approach relies on third party surveys of user perceptions of data privacy. The second approach is based on retrieval from a community of peers.

Seeding Through Third Party Surveys

(a) A policy provider makes use of third party surveys performed on a regular basis as well as those published in research literature to obtain user perceptions of the level of privacy for various sets of PII separated according to their uses. This gives a sensitivity or range of privacy levels for different PII in different situations.

(b) Corresponding to a provider's privacy policy (which specifies what PII is required), the policy provider or a software application constructs and ranks the privacy rules for each use using the PII in (a), according to their sensitivity levels, such that the rules are selectable by a single value privacy level from a "privacy slider". The outcome of this process is a set of consumer privacy rules, ranked by PII sensitivity, for different providers. The policy provider would express the resulting privacy rules in a policy language such as APPEL. There are different ways to do this

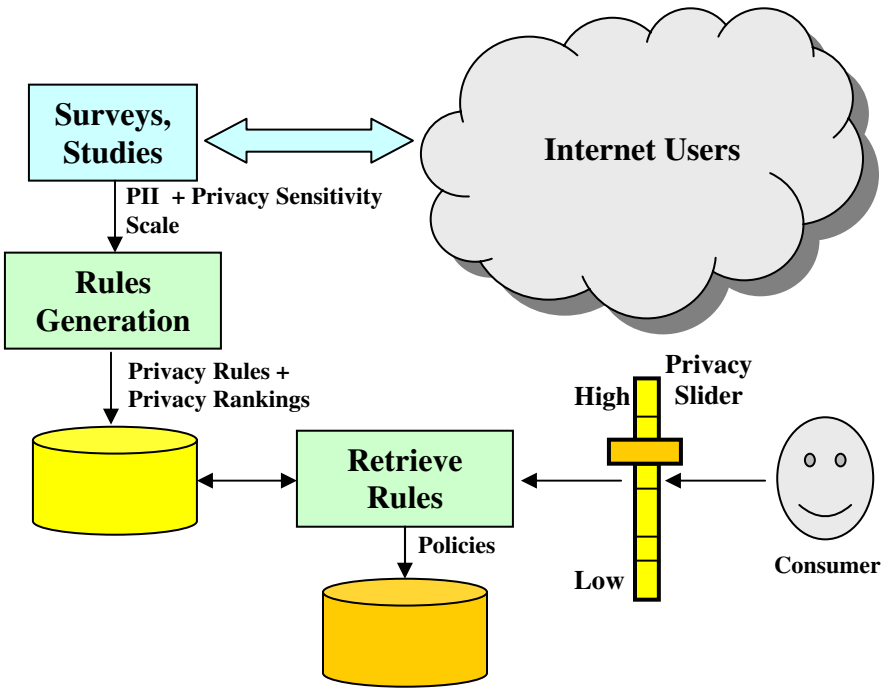
ranking. One way is to assign a privacy rule the median of its sensitivity range as its privacy level (illustrated below).

(c) Consumers obtain online from the policy provider the privacy rules that make up whole policies. They do this by specifying the use for the rules, the provider for which a consumer privacy policy is required, and the level of privacy required using the privacy slider. The consumer then completes each privacy policy by adding the rest of the header information. This can be done through a human computer interface that shelters the user from the complexity of the policy language. In this way, large populations of consumers may quickly obtain privacy policies for many service providers that reflect the privacy sensitivities of the communities surveyed.

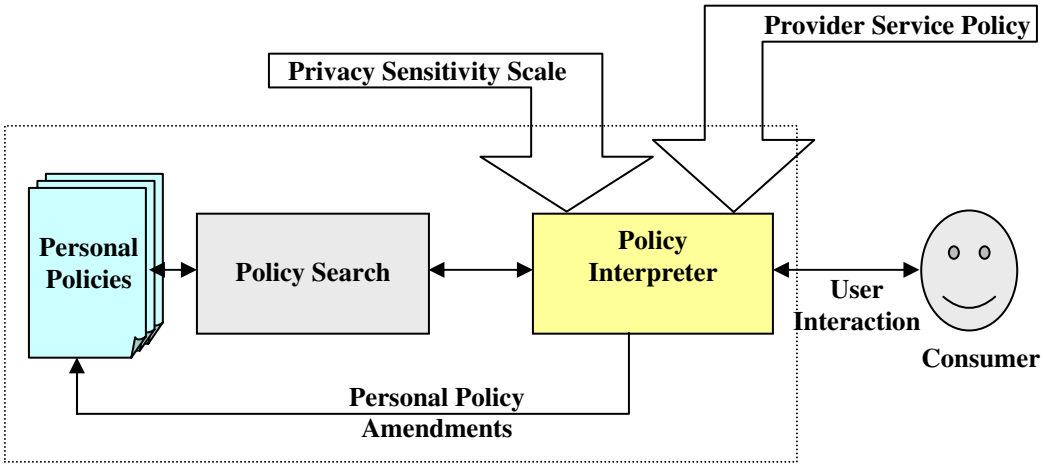
(d) Consumers may interactively adapt their privacy policies for different service providers based on their current policies, the sensitivities of the privacy rules, and the policy of the service provider. This assumes the availability of an easy to understand interface for the interaction as well as software to reflect the changes back into the policy language.

This approach requires trust in the policy provider. Effectively the policy provider becomes a trusted third party. A certification process for the policy provider is probably required. For instance, in Canada, the offices for the provincial and federal privacy commissioners could be this certification body. They could also provide this policy creation service.

A notification process should be used during the policy exchange phase between a consumer and a service provider to let the consumer know when “sensitive data” is exchanged. The degree of consumer sensitivity to different PII for different situations would also be available from the policy provider. This information could be updated regularly by the policy provider, or updated through a short online survey. The sensitivities would either modulate the base policy or set a trigger level for user warnings during policy exchange. During the warning, the user is presented with options that may allow the “degradation” or shoring up of the privacy policy. Figure 2 illustrates this approach.



a) Derivation of personal privacy policies from surveys



b) Adapting personal privacy policies to the service provider

Figure 2. Derivation of personal privacy policies through surveys

Example:

Suppose the item of PII for which we wish to derive a privacy rule is "course marks retention time" from the e-learning privacy policy in Figure 1.

Then the above steps are implemented as follows:

a) The third party survey generates the following results for course marks retention time (the higher the privacy level, the higher the privacy; the highest level is 5, the lowest level is 1). Note that CMRT stands for "Course Marks Retention Time".

PII	Privacy Level
CMRT: 6 months	3
CMRT: 6 months	4
CMRT: 6 months	4
CMRT: 6 months	5
CMRT: 12 months	1
CMRT: 12 months	1
CMRT: 12 months	2
CMRT: 12 months	3

CMRT = Course Marks Retention Time

Note that the other parameters in a privacy rule may change too, not just retention time. We change retention time only to keep the example simple. Actually, each different combination of parameters represents a different privacy level. The privacy level is inversely proportional to the retention time of the marks. The different privacy levels obtained for the same PII constitute one part of the privacy sensitivity scale.

(b) In this step, the policy provider constructs privacy rules from the PII in (a) and ranks them using the median value from the corresponding sensitivity range. Thus for the 4 course mark retention times of 6 months, the lowest value is 3, the highest value is 5, and the median is 4. Therefore the rule < any, course marks, records, 6 months, none > is ranked with privacy level 4. Similarly, the rule < any, course marks, records, 12 months, none > is ranked with privacy level 2.

(c) To obtain his/her privacy rules, the consumer specifies the use as e-learning and a privacy slider value of 4 (for example). He/she then obtains the rule

< any, course marks, records, 6 months, none >

and proceeds to complete the policy by adding header values for *owner*, *proxy*, and *valid*.

Retrieval from a Community of Peers

This approach assumes an existing community of peers already possessing specific use privacy policies with rules according to desired levels of privacy. A new consumer joining the community searches for personal privacy rules or whole personal privacy policies (sets of rules). The existing personal privacy policies may have been derived using the third party surveys as above. The privacy policy rules are each stored along with its privacy level so that it may be selected according to this level. Where a rule has been adapted or modified by the owner, it is the owner's responsibility to ensure that the slider privacy value of the modified rule is consistent with the privacy sensitivity scale from surveys.

(a) All online users are peers and everyone has a privacy slider. The new consumer broadcasts a request for privacy rules to the community, specifying use and slider value. This is essentially a peer-to-peer search over all peers.

(b) The community responds by forwarding matching (in terms of use and slider value) rules to the consumer. This match may be a fuzzy match as well.

(c) The consumer compares the rules and selects them according to use, popularity (those that are from the greater number of peers), and best fit in terms of privacy. After obtaining the rules, the consumer completes the privacy policies by completing the headers as in the above derivation from surveys approach.

(d) Consumers may adapt their privacy policies for different service providers as in the derivation by surveys approach.

There is a challenge here regarding how to carry out this approach in a timely fashion. Efficient peer-to-peer search techniques will collect the policies in a timely manner, but the amount of information collected by the requester may be quite large. As well, since the various policies collected will probably differ from each other, the requestor will have to compare them to determine which one to select. Quick comparison so as to reduce the amount of data collected would be through a peer-to-peer policy search that employs a policy hash array, containing hashed values for different portions of the policy for more rapid comparison.

FUTURE TRENDS

We expect that over the next few years, consumers will become increasingly aware of their privacy rights. This is already happening as consumers are faced with the practical implications of privacy legislation. In Canada, this has meant that consumers are being asked for permission before their private data is collected every time they walk into a dentist's office or visit an optician for glasses. To ensure that their privacy rights are respected, consumers will need to express their privacy preferences in personal privacy policies. Hence the need for consumers to be able to create their personal privacy policies easily will grow. In response to this need, researchers will discover more ways for them to do so easily. In addition, there will be a need for technologies that are associated with personal privacy policies, such as policy negotiation, policy compliance, and trustable interfaces for interfacing the consumer to the provider for the purpose of privacy policy management (see "Introduction" for policy negotiation and compliance). Consumer privacy will only be truly protected once these technologies are available and used.

CONCLUSIONS

The protection of personal privacy is paramount if e-services are to be successful. A privacy policy approach to privacy protection seems best for e-services. However, for this approach to work, consumers must be able to seed their personal privacy policies easily. We have presented two semi-automated approaches for seeding the policies: one based on third party surveys of consumer perceptions of privacy, the other based on retrieval from a peer community. Both approaches reflect the privacy sensitivities of the community, giving the consumer confidence that his/her privacy preferences are interpreted with the best information available. As well, they might be effectively combined.

Clearly, the notion of a trusted third party as a personal policy provider may be controversial to some. Any error made by the policy provider could affect PII for many hundreds or thousands of people. Having privacy commissioners' offices take responsibility for this process seems to be a natural fit, given their mandate as privacy watchdog for the consumer. However, the process would have a cost. Costs might be recovered via micro-charges to the consumer, or the service provider for the policies provided. Aggregated information from the PII surveys might be sold to service providers.

REFERENCES

Ackerman, M., Cranor, L., and Reagle, J. (1999). Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences. Proceedings, *E-COMMERCE 99*, Denver, Colorado.

Barth, A. and Mitchell, J. (2005). Enterprise Privacy Promises and Enforcement. Proceedings, *WITS'05*, Long Beach, CA, USA, January 10.

Canadian Standards Association. Model Code for the Protection of Personal Information. Available as of Sept. 5, 2003 at: http://www.csa.ca/standards/privacy/code/Default.asp?articleID=5286&language=English

Department of Justice. Privacy Provisions Highlights. Available as of Feb. 28, 2005 at: http://canada.justice.gc.ca/en/news/nr/1998/attback2.html

Dreyer, L., and Olivier, M. (1998). A Workbench for Privacy Policies. Proceedings, *The Twenty-Second Annual International Computer Software and Applications Conference (COMPSAC '98)*, pp. 350-355, Aug. 19-21.

European Union. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Unofficial text available as of Sept. 5, 2003 at: http://aspe.hhs.gov/datacncl/eudirect.htm

Government of Canada. *Personal Information Protection and Electronic Documents Act*. Available as of February 28, 2005 at: http://www.privcom.gc.ca/legislation/index_e.asp

IBM. Enterprise Privacy Architecture Language (EPAL). Available as of Feb. 28, 2005 at: http://www.zurich.ibm.com/security/enterprise-privacy/epal/

Industry Canada. Privacy and the Information Highway, Regulatory Options for Canada. Chapter 6, available as of Sept. 5, 2003 at: http://strategis.ic.gc.ca/SSG/ca00257e.html#6

Jensen, C., and Potts, C. (2004). Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices. Proceedings, *CHI 2004*, Vienna, Austria, April 24–29.

Karjoth, G. and Schunter, M. (2002). A Privacy Model for Enterprises. Proceedings of the *15th IEEE Computer Security Foundations Workshop (CSFW'02)*.

Lichtenstein, S., Swatman, P., and Babu, K. (2003). Adding Value to Online Privacy for Consumers: Remedying Deficiences in Online Privacy Policies with an Holistic Approach. Proceedings of the *36th Hawaii International Conference on System Sciences (HICSS'03).*

Snekkenes, E. (2001). Concepts for Personal Location Privacy Policies. Proceedings, *EC'01*, Tampa, Florida, USA, October 14-17.

Stufflebeam, W., Anton, A., He, Q., and Jain, N. (2004). Specifying Privacy Policies with P3P and EPAL: Lessons Learned. Proceedings, *WPES'04,* Washington, DC, USA, October 28.

W3C. The Platform for Privacy Preferences. Available as of Feb 28, 2005 at: http://www.w3.org/P3P/

W3C (2002). "A P3P Preference Exchange Language 1.0 (APPEL1.0)", W3C Working Draft 15 April 2002. Available as of Feb. 28, 2005 at: http://www.w3.org/TR/P3P-preferences/

Yee, G. and Korba, L. (Jan., 2003). Bilateral E-services Negotiation Under Uncertainty. Proceedings, *The 2003 International Symposium on Applications and the Internet (SAINT2003)*, Orlando, Florida.

Yee, G. and Korba, L. (May, 2003). The Negotiation of Privacy Policies in Distance Education. Proceedings, *14th IRMA International Conference*, Philadelphia, Pennsylvania.

Yee, G. and Korba, L. (2004). Semi-Automated Derivation of Personal Privacy Policies. Proceedings, *15th IRMA International Conference*, New Orleans, Louisiana, USA, May 23-26.

Yee, G. and Korba, L. (July 2004). Privacy Policy Compliance for Web Services. Proceedings, *IEEE International Conference on Web Services (ICWS 2004)*, San Diego, California, USA.

Yee, G. and Korba, L. (2005). Semi-Automatic Derivation and Use of Personal Privacy Policies in E-Business. *International Journal of E-Business Research*, *Vol. 1, No. 1*, Idea Group Publishing.

TERMS AND THEIR DFINITIONS

An **e-service** is an electronic service accessed via a network such as the Internet. Example e-services include online banking, online stock broker, online tax information, and online learning.

A **consumer** of an e-service is a user of the service, possibly by paying a fee.

A **provider** of an e-service is a business that operates the e-service and offers it to consumers, possibly earning fees for the use of the service by consumers.

Personal information or **personally identifiable information (PII)** is information that is personal about an individual that may be linked with the individual or identify the individual, eg. credit card number, birth date, home address, social security number.

Privacy is the right of an individual to determine when, how and to what extent his/her personal information is communicated to others.

A **personal privacy policy** is a description of personal privacy preferences, stating what personal information or PII may be communicated to others, to whom such information may be communicated, and under what conditions the communications may occur.

A **provider privacy policy** is a description of provider privacy preferences, stating what personal information or PII the provider requires from the consumer, and the conditions under which the information is required, in order for the provider to carry out its service.

A **peer** is another node in a network that is like every other node in the network.

A **community of peers** is a grouping of such nodes having something in common or considered grouped for a specific purpose, e.g. having particular types of privacy policies as discussed above.

---

[1] NRC Paper Number: NRC 48235