Expressive Privacy Promises — How to Improve the Platform for Privacy Preferences (P3P)

Matthias Schunter, Els Van Herreweghen, Michael Waidner (eds.)
IBM Zurich Research Laboratory, Switzerland
{mts,evh,wmi}@zurich.ibm.com

September 30, 2002

Abstract

P3P can be a viable tool for organizations to advertise their privacy promises. This position paper summarizes our experiences and some problems we have encountered when using P3P. The main criticisms are amboguities in the specification and missing guidelines for user agents. The main suggested improvement is a extended but simplified syntax and a revised consent model that groups opt-in and opt-out to multiple statements into one 'consent block'.

1 The Scope of P3P

The W3C Platform for Privacy Preferences (P3P) [4] aims at an organization-independent way to formalize privacy statements. It constitutes a formal language that aims at enabling organizations to advertise *privacy promises* in a well-defined way. How these promises are actually enforced within an organization is not addressed by P3P.

In order to identify potential improvements, we first need to clarify our understanding of the scope of P3P. From our point of view, there are at least three different scenarios for privacy policies. P3P only adresses the one we describe last.

Data Subject Policies The preferences of a particular data subject who's data may be collected by an organization. Preferences can be formalized using APPEL [5]. Another example are the IBM Privacy Services that use data subject policies to protect the release of personal data that is stored in a personal wallet [1].

Privacy Promises The privacy promises that an organization advertises. Privacy promises enable a user to determine whether his data subject preferences or policy matches and whether data shall be released or submitted. Privacy promises can be formalized using P3P [4]. Since P3P promises can be interpreted as a legally binding statement, such statements are usually reviewed or created by the legal department and describe only basic guarantees that the organization is willing to give.

Privacy Practices The access- or privacy-control policy that governs the actual usage of the data by one or more organization. An example formalization is E-P3P [3]. Such policies can be used to formalize the privacy policies associated with data that is handled inside an enterprise or exchange between two organizations in one or two enterprises. Privacy practices are more detailed and restrictive than promises. They are usually defined and/or implemented by non-legal players.

In this position paper, we only describe observations and improvements that strenghten P3Ps ability to formalizing privacy promises. We do not try to twist P3P into applicability for the other areas.

2 Perceived Problems of P3P

This section surveys some problems of P3P 1.0. We structured them into syntactical and semantical issues of the language and issues related to P3P user agents.

2.1 Syntax of P3P Privacy Statements

The core of a P3P policy are a data schema and (privacy) statements. A data schema defines data elements that can be used in privacy statements. Each privacy statement defines what data elements can be used by what recipients for what purpose as well as the retention policy for the data elements. The meaning of a statement with multiple elements of each type is that *all* listed data elements can be used by *all* listed recipients for *all* listed purposes (i.e., corresponding to a cross-product). If multiple statements contain the same data element, both can be applied (i.e., implementing a union).[1]

In general, P3P allows statements to be ambiguous and redundant without associating any meaning to such statements. Recipient, retention. and purpose are not clearly separated. We now elaborate some details.

Consent and Choice Single elements (data-users, purposes, data-elements) inside a P3P statement can be declared opt-in or opt-out. Opting in for, e.g., a purpose allows this purpose for all other data-users, and data-elements. This makes statements with multiple elements of each type and some of them optional very hard to understand. In addition, if all recipients and purposes are optional but its collection is required, it is unclear whether the data may be collected (but not used) or whether it should not be collected in the first place.

Retention Statement P3P uses pre-defined labels to advertise how long data can be retained. An actual maximum time-span is only required in the human-readable text. The main label is `stated-purpose` that signals that the data will be retained for the stated purposes. With multiple purposes in a statement, this means that the data can be retained as long as *any* of the stated purposes is still active. The consequence are a variety of retention times for the same data element.

Pre-defined Recipients, Purposes The pre-defined values for recipients and purposes are fuzzy and mix notions of business relationship and policy. The recipient `delivery`, for example, mixes purpose and recipient. This can easily mislead user agents. The names are sometimes misleading. The purpose `contact`, e.g., conceals that the data will be used for marketing.

Categories Categories are elements inside data elements that provide hints to users and agents as to the intended **uses** of the data. The multitude of categories give an impression of a lot of choice/granularity, but definitions are unintuitive and data cannot but overlap between categories, or is at least link-able. The privacy-relevant question "what can be derived from all that collected data" is never even addressed. E.g., collecting click-stream data is usually harmless unless it is used to link other data.

2.2 Unclear Meaning of a P3P Policy

A P3P policy should make clear what recipient may perform what purpose on which data element. In addition, it should define what data can be collected, whether it needs to be anonymized at collection, and how long can it be retained.

[1] We were unable to determine whether P3P allows multiple statements for the same data element or not.

Unfortunately, the P3P specification only describes the meaning of a policy that restricts itself to the most primitive case. Complicated cases, like conflicts, are not sufficiently addressed. As a consequence, each reader that implements a user agent needs to perform his own educated guess how to resolve such issues.

Overlapping Statements The specification of P3P does not define whether the same data element can be included in multiple statements. If each data element can only be used once, the resulting policy would be very coarse grained. The following statement could no longer be expressed: " 'ours' can use the data for 'stated-purpose' and 'contact' while 'delivery' can use the data for 'stated-purpose' only.".

If we assume that the same data element can be used in different statements, P3P would need a notion of conflict resolution that defines what statements are considered as conflicting and how conflicts are resolved. This holds in particular if the same data element is collected as identifiable as well as non-identifiable, or as always as well as opt-in and/or opt-out.

Pseudonymous and Anonymous Use and Data The issue of data-anonymization is largely unclear. And what is P3P's definition of "anonymizing data"? It is unclear what pseudonymous use mean in combination with other purposes. Another open issue is what a `non-identifiable` label at a data element means if the same data element is also collected without this label.

2.3 Guidelines for User Agents

Interpretation by user agents is completely undefined. As a consequence, the meaning and the expressiveness of a policy mainly depends on the user agent interpreting it. There is no guidance or recommendations for writing 'agent-friendly' policies and how agents must interpret them.

3 Proposed Improvements

We believe that P3P can be a strong tool to advertise privacy promises to consumers. As a consequence, all considerations for the new P3P should take the user-agents and their user interfaces into account. A larger fraction of our criticism can be addressed by clarifying the P3P specification. We now sketch some more general improvements that go beyond clarifications of the specification.

3.1 Proposed New Features

Improved Consent Model Currently, data subjects opt-in or opt-out to elements within a statement. E.g., they can opt-out of a certain recipient for a given set of statements and retention policies. This implies that they automatically opt-in or opt-out to the resulting cross product with this recipient and all purposes and retentions. This is usually not what a user wants. In practice, a customer usually opts in for a abstract textual description that reflects many uses.

Since opt-in and opt-out usually corresponds to certain business processes in an organization that require multiple data elements for multiple purposes, it is advisable to introduce 'consent blocks' that enable to opt-in or opt-out to a set of statements. This can be formalized by named consent descriptors that can be opt-in or opt-out and describe (in text) what the consent means. Each statement can then specify a consent descriptor. If this particular consent has been given, the statement is applicable. Else, it is not.

Another advantage of consent blocks is that they can reflect the actual business process. This enables to collect consent one block at a time if and only if the process is actually

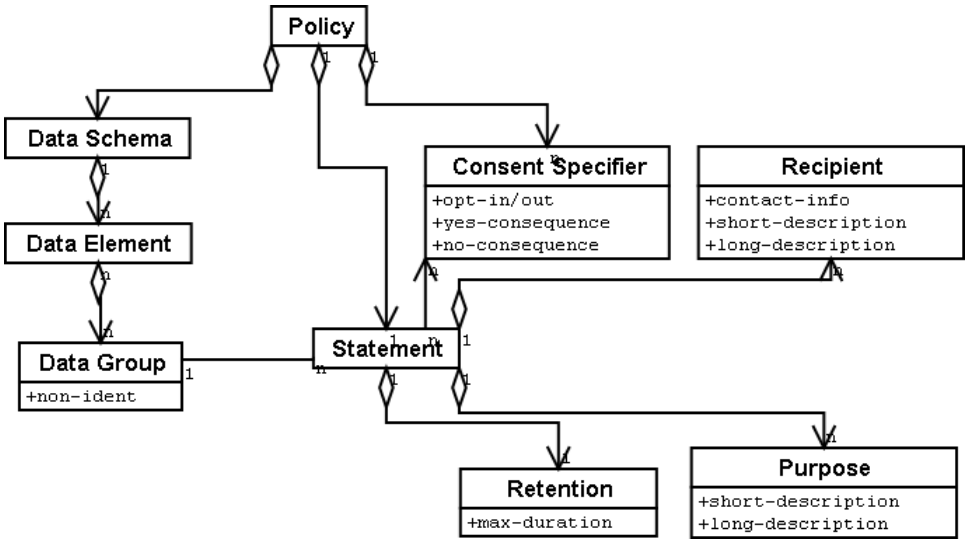

Figure 1: Outline of a Simplified XML Syntax for P3P in UML [2] (Lines denote references while lines with diamonds denotes XML sub-elements.)

started. Such real-time consent enables the organization to better illustrate why and how the data will be used. A user agent could then display a opt-in block only if the organization requires opt-in. An example is that the policy states a consent-block 'newsletter' that is connected to all statements that specify usage related to distributing the newsletter of the company.

Augmented Elements Much of the important contents of a policy is placed in the human-readable policy only. E.g, real retention policies and their destruction time tables. We feel that more information should be expressable inside optional elements of the policy.

All elements in privacy statements should be augmentable by additional qualifying information. Recipients should be augmentable by a contact address to enable the customer to find out to what organizations his data has been forwarded. The same holds for the purposes that should be augmentable with a concrete purpose and retention that should be augmentable by a concrete time-span.

An important advantage of such augmented P3P policies that they can be absolute. They can, e.g., contain an exact business purpose and exact recipients. This is a first step towards using P3P to store privacy promises given to a particular data subject. Additional changes that improve P3P's applicability as a promise storage format are consent blocks should that are augmentable by the actual choosen value of a particular data subject and an element with an opaque data subject identifier that can be used by organizations for promises management.

3.2 Simplified Syntax

The syntax can be clarified simplified and more structured. Our proposed updated syntax is depicted in Figure 1. The basic elements are

- a data schema that defines data elements,
- data groups that group data and define whether they are anonymized during collection,
- named consent specifiers that allow users to opt-in and opt-out to multiple statements at once,

- privacy statements that define retention, recipients and purposes, while identifying the required consent choices.

We will now describe some elements in more detail.

Simplified Statements The syntax should either prevent or else resolve conflicting or ambiguous statements. A simplified statement could identify a data group, a list of purposes for each recipient of the data group and an optional name of the required consent. For expressiveness, one should allow that the same elements can contain in multiple statements.

3.3 Clarified Specification

P3P should be augmented by documentation on how to resolve conflicts in a P3P policy. This documentation should associate a meaning to all P3P policies that are considered to be syntactically correct. If some P3P policies are excluded (i.e., are meaning-less by definition), this should be made explicit and such policies should not be allowed. Given a P3P policy, it needs to define answers to the following questions: "Can a given recipient use a given data element for a given purpose?" and "What is the retention policy for some data?".

3.4 User Agents

P3P is mainly built for user agents. Nevertheless, their behavior is completely unspecified by P3P. We would prefer if the P3P spec makes clearer statements and guidelines what or what should not be assumed by user agents. The main purpose of user agents is to unambiguously communicate the meaning of the P3P policy.[2] This is even more difficult without a semantics and guidelines how to communicate a P3P policy. In order to enable a wide range of agent implementations, such such guidelines should only describe requirements without defining the actual user-interface.

4 Conclusions

Even though P3P is an important step towards clear and machine-readable privacy statements, it is currently confusing. As a consequence, it is hard to understand for user-agents or humans. We have outlined some future directions that can help to simplify and improve P3P.

Many proposed changes (like a well-documented meaning and clearer user agent guidelines) can be resolved without changing the syntax. Some others (like a simplified syntax and a blocked consent model) will require changes to the P3P syntax.

A non-technical problem of P3P is the lack of a sufficient number of user-agents. Even though this is not a problem of the P3P language, it may be worth investigating how to facilitate the development and use of user agents.

Acknowledgements

We thank Martin Presler-Marshall and Calvin Powers for detailed comments and for detailed insight on the process that lead to P3P. We thank Xuan Liu for important inputs. Furthermore, we thank the many contributors at IBM that provided comments or feedback on drafts of this position paper.

[2] A first step could be to re-use some results that lead to APPEL [5].

References

[1] Kathy Bohrer, Xuan Liu, Dogan Kesdogan, Edith Schonberg, Moninder Singh, and Susan L. Spraragen. Personal information management and distribution. In *4th International Conference on Electronic Commerce Research (ICECR-4); Dallas, TX, USA*, 2001.

[2] Martin Fowler and Kendall Scott. *UML Distilled : Applying the standard object modeling language*. Addison-Welsey, Reading MA, 1997.

[3] G. Karjoth, M. Schunter, and M. Waidner. The Platform For Enterprise Privacy Practices – Privacy-enabled Management Of Customer Data. In *Proceedings of the Privacy Enhancing Technologies Conference*, San Francisco, CA, April 14-15 2002.

[4] W3C. Platform for Privacy Preferences. Available at `www.w3.org/P3P`.

[5] W3C. A P3P Preference Exchange Language 1.0 (APPEL1.0), 2002. Available at `www.w3.org/TR/P3P-preferences`.