The Social Contract Core

James H. Kaufman, Stefan Edlund, Daniel A. Ford
IBM Almaden Research Center
650 Harry Rd.
San Jose, CA 95120-6099
{kaufman, edlund, daford}@almaden.ibm.com

Calvin Powers
IBM Raleigh Software Lab
3901 S. Miami Blvd.
Durham, NC 27703
cspowers@us.ibm.com

ABSTRACT
The information age has brought with it the promise of unprecedented economic growth based on the efficiencies made possible by new technology. This same greater efficiency has left society with less and less time to adapt to technological progress. Perhaps the greatest cost of this progress is the threat to privacy we all face from unconstrained exchange of our personal information. In response to this threat, the World Wide Web Consortium has introduced the "Platform for Privacy Preferences" (P3P) to allow sites to express policies in machine-readable form and to expose these policies to site visitors [1]. However, today P3P does not protect the privacy of individuals, nor does its implementation empower communities or groups to negotiate and establish standards of behavior. We propose a privacy architecture we call the Social Contract Core (SCC), designed to speed the establishment of new "Social Contracts" needed to protect private data. The goal of SCC is to empower communities, speed the "socialization" of new technology, and encourage the rapid access to, and exchange of, information. Addressing these issues is essential, we feel, to both liberty and economic prosperity in the information age[2].

Categories and Subject Descriptors
D.2.11 [**Software Engineering**]: Software Architectures – *domain-specific architectures*

General Terms
Design, Experimentation, Standardization.

Keywords
Privacy, P3P, Social Contract.

1. INTRODUCTION
The information revolution, like the industrial revolution before it, promises to create greater economic prosperity as it speeds access to goods, services, and information. Just as the advent of the production line increased the rate of wealth generation, the creation of the Internet has accelerated information flow, enhanced the efficiency of payment systems, and increased the speed of financial transactions [3, 4]. The increased efficiency with which companies operate (from supply chain management to new product definition) depends upon rapid access to and interchange of information. The information itself often increases in value as the velocity of information flow increases [3]. Market data, for example, is most valuable when a company can rapidly enhance or modify product lines to fill a new market niche and is less valuable as it becomes outdated. The information also rises in value as suppliers can identify individuals and groups to target new products and offerings. In the information age, enterprises are driven to learn the needs of their customers quickly; also, they want to group their customers in order to target individuals according to their needs, ability to pay, price willing to pay, related interests, how they can be contacted etc. As the ability to deliver "location-based services" is enabled by the development and deployment of new cellular phone technology enterprises will add to this list "where are customers of type 'A' located right now?"

From the individual's point of view, increased prosperity may come at an enormous cost to their privacy. This is not the first time in history that a technological revolution has led to a tradeoff between economic progress and individual rights, rather that choice has been a characteristic of every technological or social revolution experienced by humankind. When people moved from hunter-gatherer societies to agricultural communities, they sacrificed some of their individual freedoms for a more stable food supply and better shelter. Similarly, in the industrial revolution, farmers gave up considerable privacy and freedom to define working hours as they left their rural homes and adopted the crowded lifestyle of the metropolis and the regimented work schedule of manufacturing lines. What distinguishes the current information revolution from previous technology driven transformations of society is the speed with which change is occurring. In prior revolutions, individuals had much more time to invent the social structures that make life in a transformed society not only tolerable, but on balance better. As Hobbs, Locke, and Rousseau concluded, people will sacrifice some individual freedoms for a greater good [5-8]. However, this sacrifice is based on a Social Contract that establishes norms of behavior required in the transformed society. These contracts may manifest themselves as new government regulations, but more often they emerge as accepted conventions between people that evolve as individuals adapt to new technologies.

In the past, Social Contracts had many decades to develop. Consider, for example, the introduction of the telephone. Alexander Graham Bell invented the Telephone on March 10, 1876. However, rapid expansion of the long-distance phone network did not really begin until the 1950's with the introduction of automated digital switching technology [9]. According to the U.S. Census [10], the number of households with telephones did not exceed 80% until the 1960s. By comparison, in the year 2000, 9-in-10 school-aged American children had computer access, and ninety four million people used the Internet from home, up from 57 million in 1998 [11]. Through the early part of the twentieth century, society could adapt to new technology on a timescale of generations. Since the latter half of the twentieth century, in just

WWW 2002, May 7-11, 2002, Honolulu, Hawaii, USA.
ACM 1-58113-449-5/02/0005

one generation, society has had to adapt to an explosion of information technology innovation and change. The Internet, e-mail, cell phones, plastic credit cards all were introduced in the past 50 years. With the increased rate of introduction of new technology, there is a pressing need for tools to help individuals protect their privacy, and for societies and groups to establish and define new norms of behavior.

2. THE NEED FOR A NEW SOCIAL CONTRACT

A "Social Contract" defines collective rules that constrain the behavior of individuals and groups living in a society in such a way as to protect the individual, while also benefiting the society as a whole. For example, acceptance of telephone technology requires that individuals respect certain norms of behavior (e.g., don't call me at 4:00 in the morning). Once such a Social Contract is broadly accepted, most people are willing to make their personal phone numbers available for the individual and collective good. With the advent of the Internet and the information age, we are all asked to provide a variety of personally identifiable information (PII data), often with little or no control over how that information will be used, stored, or distributed. There is no technology available today to either assist a person in protecting their privacy or to assist society to establish accepted norms of behavior and constraints on the use of PII data. The "Platform for Privacy Preferences" (P3P) standard developed by the World Wide Web Consortium, does not yet address this problem [1,12,13]. P3P is a mechanism to describe "policies", that is, statements by recipients of individuals' PII data that express how the recipient intends to deal with that data. However, P3P does not protect the privacy of individuals, nor today does its implementation empower communities or groups to negotiate and establish standards of behavior. Sites will often express "privacy policies" because they intend to distribute or sell private information.

3. DEFINING A SOCIAL CONTRACT

In order to understand how a social contract protects privacy it is first useful to consider the differences between privacy protection, access control, and security. Privacy protection becomes important after an individual has voluntarily transferred private information or PII data to a "user" of that information. Once the PII data is in the user's possession, the user has three separate responsibilities:

1. Controlling Access to the data. Access control restricts the types of users who may gain access to data based on an expressed agreement between the PII data collector and the PII data owner.
2. Honoring obligations expressed by the policy and implementing a process for resolving disputes.
3. Ensuring security of their IT infrastructure (making sure no one steals the data).

Security and Access control are both important and are readily addressed by state of the art commercial middleware [14]. The key to ensuring privacy is the establishment of expressed agreements or "contracts" between the users of PII data and the owners of PII data. The P3P standard makes it possible "for Web sites to publish their privacy policies in a machine-readable syntax". Using P3P along with other middleware, sites may:

1. Express a Policy that describes how PII data will be used.
2. Attach that policy to PII data instances.
3. Establish and maintain audit records to ensure that they have complied with the policies they declared with respect to PII data at the time the data was acquired.

Today, if a user is dissatisfied with an expressed policy, their only options are to avoid visiting the website or to put their dissatisfaction aside. Choosing to completely abstain from transactions on the Internet does protect privacy, but opting out diminishes the value of new technology. If a majority of people makes this choice, the technologists have failed and individuals and society as a whole suffer from missed opportunities.

Designers of the P3P framework have made it extensible and considered various implementations to alert users when they come across policies that fail to match their preferences. For example, P3P supports the creation of user agents that can be implemented as browser plug-ins or proxies [1].

P3P user agents look for references to a P3P policy... A P3P user agent ... would retrieve P3P policies, compare them with user's preferences, and authorize the release of data only if a) the policy is consistent with the user's preferences and b) the requested data transfer is consistent with the policy. If one of these conditions is not met, the user might be informed of the discrepancy and given an opportunity to authorize the data release themselves.

Faced with a policy discrepancy, a user's choices are limited. P3P allows users to understand a site's policy and, if that policy is unacceptable, to choose not to access the site. However, there is no mechanism in place today for individual users to express their own preferences, provide feedback to a site, or otherwise affect a change to a site's policy. In this regard, P3P does not, today, completely fulfill its goal to provide "*a simple, automated way for users to gain more control over the use of personal information on Web sites they visit.*" Furthermore, in order to indemnify them against legal action and to simplify audit compliance, there is every incentive for site owners to establish and express policies that explicitly fail to protect privacy (e.g. "if you give us your data we can do with it anything we like"). Some authors have suggested that the name "Platform for Privacy Preferences" is misleading as it fails to allow users to enforce their preferences [15].

P3P is designed not to protect data privacy but to facilitate the gathering of data…

There is nothing about P3P that would enforce or even aid the enforcement of the "deals" that are struck through its algorithms. In this sense, P3P embraces the technical while ignoring the entire social context that such a technical solution should exist within.

Many people will not understand that "privacy practices" are not the same as "privacy." P3P therefore allows sites to create an air of privacy while they gather personal data.

Privacy advocates involved in the development in the P3P standard understand these criticisms and point out that [16]:

P3P does not protect privacy, in and of itself… P3P needs a regulatory or policy context to help protect privacy, it cannot do this by itself… It does, however, help create a framework for informed choice on the part of consumers… bringing Web site privacy policies to the foreground.

Indeed, several governments have passed legislation to protect individuals and to regulate how institutions may treat PII data. Title V of the Gramm-Leach-Bliley Act deals applies such regulations to U.S. financial institutions [17]. Among other things, it requires "clear disclosure by all financial institutions of their privacy policy regarding the sharing of non-public personal information with both affiliates and third parties", and "a notice to consumers and an opportunity to "opt-out" of sharing of non-public personal information with nonaffiliated third parties subject to certain limited exceptions". Similarly, the "Health Insurance Portability and Accountability Act" (HIPAA) of 1996 empowers the U.S. Department of Health and Human Services to set regulations governing (among other things) the use of medical PII data (if congress does not explicitly set such regulations by law) [18]. In 1998, Congress passed the Children's Online Privacy Protection Act (COPPA), to protect the privacy of children using the Internet [19]. Great Britain passed, in 1998, The Data Protection Act declares that anyone processing personal data must comply with the eight enforceable principles of good practice [20].

The European Union Organization For Economic Co-operation and Development has established voluntary guidelines for Information Security and Privacy [21]. Legislation can play an important role in protecting privacy. However, given the fast pace of technology, regulatory mechanisms may act too slowly to effectively moderate the policies expressed by sites. A far better mechanism would allow individuals and groups to gain control and influence sites to adopt privacy preserving policies through collective social pressure. Clearly there is a pressing need for the definition of genuine social contracts which will not only "bring Web site's policies to the foreground", but will also empower groups to assert norms of behavior to govern and restrict the abuse of privacy. This can be done with a combination of market forces and technologies designed to speed the flow of information between individuals regarding the way competing Websites handle PII data. What better solution than to ask technology to assist in this process?

4. THE SOCIAL CONTRACT CORE

We propose an architecture, which we call the Social Contract Core, which extends P3P in such a way as to allow groups of users to define privacy norms and preferences. There is no unique privacy policy that defines a social contract appropriate for all groups or communities. Rather, individual groups and communities should be able to express their preferences. The Social Contract Core is composed of three functional units; the Client unit, the Site Owner unit, and the SCC Conventions Site (Figure 1). These units share several common SCC components

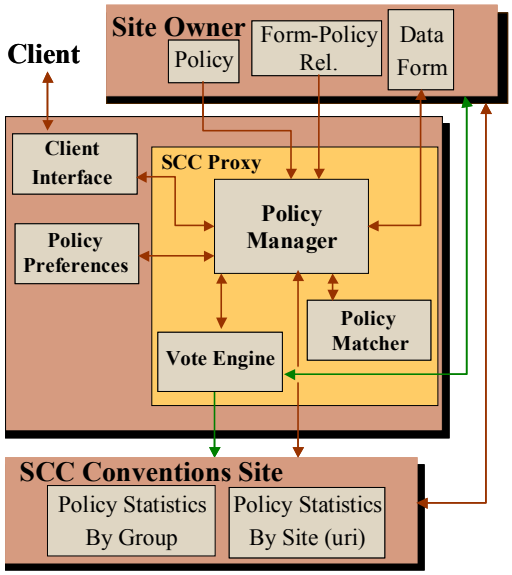

Figure 1. The social Contract Core had three functional components: A client component, a Site Owner, and a Conventions Site.

including "policy" documents, a policy matcher, a statistics gatherer, etc.

Client: The Client unit reflects the goals and priorities of the individual. It contains several components (Figure 1) including a client interface or web browser, the client's personal "Policy Preferences", and a SCC Proxy that allows the client to express preferences to others. The SCC Proxy is configurable and the client may use it to select from a variety of SCC convention sites and to identify his/her self with one or more client groups. The Policy Preferences component is a file that contains the client's P3P preferences regarding data types, purposes, and recipients. Observe that a P3P preference file is a valid P3P document, but the interpretation of attributes such as purpose and recipient of data schemas reflect the preferred settings of the data owner rather than statements of a data user. This preferences document may be implemented in the P3P preference exchange language (APPEL, [27]).

The SCC proxy sits between a web site (site owner) and the client interface. It is materialized as a web browser extension that can take actions based on policy preferences stored on the clients file system. The proxy opens a second window or view that allows the client to invoke several possible functions with regards to the data obtained from the site owner.

1) Set opt-in opt-out selections allowed under the site owner's policy to the client's default Policy Preferences. In order to accomplish this using the current P3P data type definition (DTD), the SCC proxy must access not only the site's policy document using a GET request for the resource /w3c/p3p.xml, it must also access a "Form-Policy" relationship document which relates form fields and options (checkboxes) in a Web Site's html form to the policies expressed in the governing P3P policy for that web form and the choices allowed to the user within that policy. This relationship document is an extension to P3P. It can be constructed most easily by assigning ID's to the html form and relating the ID's to P3P data types and expressions in the policy document. In Appendix I we discuss how this might be

accomplished more simply using a Resource Description Framework (RDF) binding of P3P.

2) Update the default Policy Preferences based on the current policy settings defined in the client interface window. This is particularly useful if the policy obtained from a site owner contains a policy option that applies to a data type, purpose, or recipient not yet defined in the client's own Policy Preferences.

3) Invoke a policy matcher function to compare and contrast the Site Owner's Policy with the Client's Policy

4) Retrieve Preferences or Policy Statistics (by group or site) obtained from one or more SCC convention sites. This enables the PII Owner to adopt the "best practices" or social conventions of groups he or she identifies with. The client may wish to identify with multiple groups and to change groups while moving from site to site. For example, the PII Owner could download the policy preferences of a "Parent-Teacher Association" group to pre-configure their SCC proxy with the standard acceptable policies for how data collected from Children are used.

5) Vote. Voting sends an xml document with the client's Policy Preferences to both the Site Owner and to the SCC collection Sites. To upload a P3P Policy Preference the client can for instance use HTTP post requests to a URL that is predefined in the web site's P3P policy file (see Table 1). Note that neither the site owner nor the SCC collection site store profiles on individual users. These profiles are only used (on the server side) to generate statistics that reflect the privacy preferences of collections or groups of individuals in order to help set the emerging accepted standards (i.e., the "social contract") for a privacy policy.

Note that there is no a-priori reason to place an arbitrary limit to client votes or to the number of SCC sites they may join and send votes to. There is, of course, a need to prevent hackers from gaming the system and for this reason individual SCC sites would necessarily adopt mechanisms to detect and prevent mischievous behavior. There are several ways this can be accomplished, including the use of click stream data. The detailed solution is beyond the scope of this paper. Suffice it to say that whatever vote validation mechanisms are adopted, the mechanisms themselves must not provide a threat to privacy or disincentive to participation.

6) Along with this document the client sends an entity data group (Table 2) that identifies the social group (or groups) that the client identifies with. The client also passes the URL of the site owner to the SCC collection site. The SCC site and the Site Owner process this information and use it in several ways.

```
<?xml version="1.0" encoding="UTF-8" ?>
<META xmlns =
   "http://www.w3.org/2000/12/P3Pv1">
 <POLICY-REFERENCES vote-uri = "/p3p-
vote/submit-vote">
  <POLICY-REF about =
"http://www.site.com/w3c/all_p3p.xml">
   <INCLUDE>/*</INCLUDE>
  </POLICY-REF>
 </POLICY-REFERENCES>
</META>
```

Table 1. URI specifying a location to express preferences by voting.

Site Owner: The Site Owner (Figure 2) unit contains one or more policies defined in accordance with the P3P standard. Obviously the Site Owner also has other data relevant to the current transaction or purpose of the site. The SCC also provides the Site Owner with additional "Customer Policy Preferences and Statistics" which are collected when visitors to their site "vote" their preferences. A site might use these statistics to alter their P3P policy for the mutual benefit of the site owner and the site's customer groups (see below). When a client chooses to express their views to a Site Owner by "voting", the site owner receives from the client an XML document containing the client's policy preferences. This document is compared with the site's current policy by a vote monitor that uses a policy matcher to extract two types of data from the preferences document.

Within the P3P specification, a policy may apply to a data type or a data-group (collections of data types). To avoid contradictions, the Policy Preference document describes preferences with respect to data types only. The policy matcher looks for policy preferences which apply to data types or data groups within a site's existing policy and generates statistics of the number of clients who prefer to opt-in or opt-out of various policy purposes, recipients, etc. by data type. For example, consider a site owner collecting contact information as grouped in Table 3. This site might apply one policy to this entire group of data types (#business.contact-info.*) as shown in Table 4.

When a client chooses to vote, the vote monitor will extract statistics on every relevant data type contained in the site's policy. For example, consider the purpose, "contact", in the policy code above. The site will gather statistics separately for "#business.contact-info.postal.*" data types and the "#business.contact-info.telecom.telephone.number" data types even though the site owner has chosen to define a single policy for the entire group. This is necessary for two reasons. First, the client may not have the same data group defined. Second, it is in the interest of the site owner to gather the data by data type. Imagine a hypothetical client group wherein 99% of clients prefer not to be contacted by phone under any circumstances. These individuals would opt out of any and all contact methods based on this site's data group. On the other hand, suppose 30% of clients in the same group would be willing to accept postal contacts. Gathering statistics by data type for a significant sample of clients would reveal to the site that they were missing a marketing opportunity for 30% of individuals. In this way the Social Contract Core is an aid to Site owners who wish to use client PII data. This example suggests part of a first business model for the Social Contract Core and a reason for Site Owners to support

```
<ENTITY>
 <DATA-GROUP>
  <DATA  ref="#scc.group.name">
   NewYorker
  </DATA>
 </DATA-GROUP>
</ENTITY>
```

Table 2. Policy Preferences and statistics may be expressed for a data group.

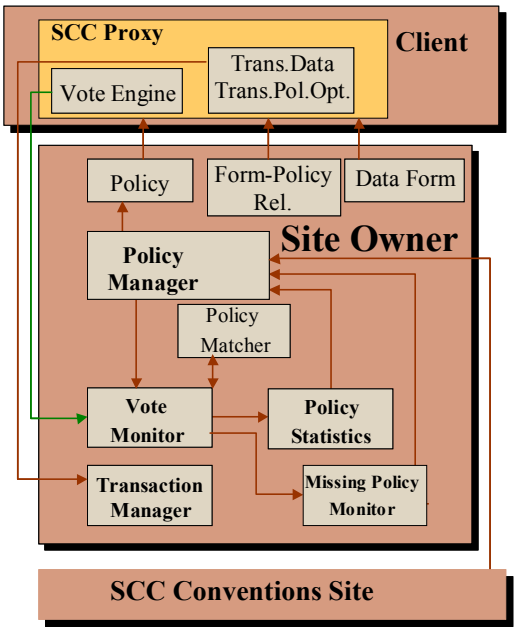

Figure 2. Site Owners gather statistics about their policies and discover missing policies from the vote process.

voting. By enabling a site-specific statistics gatherer, the site owner can discover market forces and at the same time gain a marketing opportunity by (in this example) providing separate contact policies for telephone and postal contact information.

The vote monitor gathers statistics (Table 5) about client's satisfaction or dissatisfaction regarding the existing default policies for data types and purposes defined by the site. The site can then use this information to change the default policy settings ("Most of my customers prefer not to be contacted by email.") or to split or combine data groups according to the needs of their client population. However, the vote monitor cannot gather meaningful statistics about policies or policy choices NOT defined by the site. The vote monitor also uses a policy matcher method (not shown) to discover these missing policies. A missing policy might correspond to a purpose not relevant to the site, or it might correspond to a missing "required" field for an existing purpose (i.e. a use of the client's PII data for which the client is not given a choice but prefers to have one). The missing policy monitor collects these missing policies. Both policy statistics and missing policies can be used to update a site's policy. A policy manager may even update default opt-in opt-out settings automatically and periodically based on the site's expiry cycle.

SCC Convention Site(s): The SCC Convention Site(s), or SCC site, for short, defines a new class of web service. The SCC site accepts policy preferences, URLs and entity data groups from clients. It uses a "Vote Monitor" and a "Policy Matcher" component to extract statistics and missing policy data by site. From a web site URL, the SCC Convention site retrieves a web site policy. Web Site policies may also be cached. Since every P3P policy has an expiry field, cached policies may be automatically refreshed based on the expiry time and the workload of the convention site. Matching site policies with client preferences, the SCC Convention site can generate statistics.

The SCC site acts as a library of both Group Social Preferences (GSPs) statistics and of Privacy Policy Standard

```
<DATA-GROUP>
  <DATA ref="#business.contact-info.postal.street">
   4000 Lincoln Ave.
  </DATA>
  <DATA ref="#business.contact-info.postal.city">
   Birmingham
  </DATA>
  <DATA ref="#business.contact-info.postal.stateprov">
   MI
  </DATA>
  <DATA ref="#business.contact-info.postal.postalcode
   48009
  </DATA>
  <DATA ref="#business.contact-info.online.email">
   catalog@example.com
  </DATA>
  <DATA ref="#business.contact-
info.telecom.telephone.loccode">
   248
  </DATA>
  <DATA ref="#business.contact-
info.telecom.telephone.number">
   2341234
  </DATA>
 </DATA-GROUP>
```

Table 3. A typical Data Group.

(PPS) statistics for various classes of sites. GSP's reflect the preferences of groups of clients whereas PPS's reflect the policies of Site Owners by, for example, industry sector. Both GSP's and PPS's are expressed as P3P policies with regards to P3P data types, purposes, and recipients. The data in the SCC library is freely available to both individual clients and to Site Owners. Individuals may use the SCC library to select default values for their personal "Policy Preferences" (selecting the SCC group or groups with which they most closely identify). They may also use the library to compare particular Site Owner's policies with policy "social norms" for other owners of the same type. Similarly, Site Owners may compare their own policies with those of sites similar to their own.

Note that the voting process described above sends privacy preferences to both the SCC convention site as well as to the Site Owner (whose policy is the motivation for the vote). In principle it is unnecessary to send such votes to both locations. Our architecture supports these independent vote actions for two reasons. First, the existence of numerous SCC convention sites would put a burden on site owners wishing to consider their customers preferences (requiring the owners to periodically survey a multiplicity of convention sites). Second, requiring the site owners to retrieve votes from convention sites would slow down the social contract feedback process. In our view, websites would be driven by market forces to enable site-specific statistics gathering so they could more quickly respond to client vote actions. If a website chooses not to enable statistics gathering, vote actions would still register at the SCC convention sites where both customers and competitors could view collective opinions regarding the site's policies.

5. SIMPLIFIED CLIENT SCENARIO

The previous sections describe a democratic embodiment of the Social Contract Core, which gives an end-user (the PII Owner) the ability to provide immediate feedback regarding a Web Sites policy by a voting mechanism. We recognize that a full implementation of a social contract core will require new infrastructure and security technology beyond the scope of this paper. There may be instances (for example in a peer to peer realization) where a complete voting architecture is too complex. In such cases, PII Owners may not want to invest much time in negotiating preferences with web sites or voting preferences with a Social Conventions site. While almost everyone expresses concern about maintaining control of the PII they submit online, there may be times individuals chose not to use such features while browsing the Internet. At such times, the end users may be willing to delegate the expression of their personal preferences to organizations or groups that they trust. This can be accommodated in the Social Contract Core with a simplified client component that lets the end user adopt privacy policy practices set by groups the end-user affiliates with.

In this simplified scenario, an end-user would configure their SCC proxy with the address of a Social Conventions site and then register affiliation with groups that the SCC site represents. The SCC client proxy will download the privacy preference standards set by these groups. Periodically the proxy will also check the SCC conventions site to look for updates to group's privacy standards. The end-user may also deregister an affiliation with the group.

As the end user browses the Internet, the SCC would record the end-user's preferences about which group policies to adopt for different sites, domains, or end user profiles. For example, while children are using the web browser, the SCC client proxy might use the preferences set by the National Parent Teacher Association. While the parents are browsing, they might adopt the standards published by their church for most web browsing, but use a more restrictive privacy policy published by their bank when buying from online stores.

The SCC client proxy would continue to be responsible for assisting the client's web browsing as described in earlier sections. It would assist in setting opt-in/opt-out choices based on the standards set by their current group affiliation. It would update the privacy preferences as the end-user moves from site to site based on the per site, per-domain, per profile the end-user expresses. It would continue to compare a site's privacy policy against the end-user's preferences and warn the end user if he or she is about to submit privacy sensitive to a site that has a policy that disagrees with his or her preferences.

```
<PURPOSE>
    <contact required="opt-in"/>
    <individual-decision required="opt-in"/>
    <tailoring required="opt-in"/>
</PURPOSE>
<RECIPIENT>
    <ours/>
    <same required="opt-in"/>
</RECIPIENT>
<RETENTION>
    <stated-purpose/>
</RETENTION>
```

Table 4. A single policy expressed for a data group.

In this simplified client scenario, the end-user would not have to create their own groups, would not have to explicitly make privacy preference choices while browsing the Internet, and would not have to explicitly vote their preferences with the SCC site. This would make privacy policy monitoring much less intrusive to the end-user's browsing experience. While not as fine-grained as the voting scheme described earlier, this simplified client scenario does provide a means of providing feedback or voting privacy policy preferences through affiliations. If the registration process of the Social conventions site has a basic authentication mechanism in place to ensure that each person can only register once, then the emerging social contract can be tracked by monitoring how many people affiliate with each group. In this system, Convention sites and their respective groups gain influence based on their membership and market forces. While this feed back mechanism is less granular (and perhaps slower) than the voting mechanisms in the full scenario, it nonetheless provides a way for the social contract regarding the usage of PII to evolve over time. The two approaches are not mutually exclusive. The simplified client scenario, "voting by affiliation," also creates another possible business model for the Social Contract Core. During the end user's registration with the social conventions site, the social conventions site might ask the end-user to respond to some basic demographic questions. These questions need not need to require that the end-user give identifying information. For example, the Social Conventions site might ask for the zip code that the end-user lives in, the household income level, hobbies, etc. The Social Conventions site would not ask for phone numbers, actual names, addresses, etc. Of course, the social conventions site should always enable the end-user to opt-out of supplying any demographic information. Demographic information collected by the Social Conventions site can be sold to businesses and organizations that want to tailor their privacy practices to their site's target demographic audience, thus creating a privacy preserving business model for supporting the social conventions site and creating a finer grained picture of the evolving social contract while still protecting the end-user's privacy.

6. EASE OF USE

The Social Contract Core extends P3P in such a way as to allow groups of users to define privacy norms and preferences. Perhaps the biggest obstacle to implementing a Social Contract Core is the user interface. Privacy policies expressed by sites using P3P are very complex and will only grow more complex as P3P evolves the capability to render a legal privacy contract in machine-readable form [16, 22]. Our proposed architecture was designed with this problem in mind. Users obtain default personal privacy preference files from a collection of neutral third party SCC Convention sites. Since no single default preference file will serve all social groups, the SCC Convention sites each contain a library of such files; each file represents the collective views of a particular community. Some of these library files may be static (e.g., representing a most restrictive default preference collection) and others may evolve. Furthermore, users need not learn a complex profile configuration process nor schedule recurring

```
<!DOCTYPE META SYSTEM "statistics.dtd">
<META>
    <total-votes>30</total-votes>
    <POLICIES>
        <POLICY>
            <STATEMENT>
                <PURPOSE>
                    <admin>
                        <opt-in>10</opt-in>
                        <opt-out>15</out-out>
                        <always>5</always>
                    </admin>
                </PURPOSE>
                <RECIPIENT>
                    <unrelated>
                        <opt-in>1</opt-in>
                        <opt-out>10</out-out>
                        <always>0</always>
                    </unrelated>
                </RECIPIENT>
                <DATA-GROUP>
                            <DATA ref =
                            "#dynamic.clickstream"/>
                </DATA-GROUP>
            </STATEMENT>
    </POLICY>
    </POLICIES>
</META>
```

Table 5. Example of possible vote statistics.

reconfiguration times to keep their default preferences up to date. The SCC proxy allows every client to add to or modify their default preferences in the course of visiting Internet sites. We are currently prototyping an SCC proxy using the WBI Development Kit for Java [23]. In one embodiment of the SCC, the proxy can be implemented as a browser plug-in that adds an SCC button to any web page. Pressing the button opens a dialogue or launches a new browser window containing the following components:

1) A policy match window that compares a site's policy to a user's default policy preferences

2) A pull down selection box allowing the user to choose from any number of SCC convention sites.

3) A pull down selection box allowing the user to choose from (identify with) a particular user group.

4) Three action buttons that allow a user to:

i. Preset current policy options from policy preferences (button inactive if no policy options exist on current website).

ii. Update policy preferences from policy options selected on current website (button inactive if no policy options exist on current website). This feature is very important as it allows users to add new default preferences in the course of browsing the web. As P3P is extended new purposes, and recipients will not doubt be added to the specification. Furthermore, policy defaults downloaded from SCC libraries may not always be complete. This feature simplifies the policy maintenance process.

iii. Vote as discussed in the client section above.

Users need not manage their privacy policy at all times. For instance, a user not concerned with site use of click stream data may choose to activate the SCC button when actively sending PII data to a website (e.g., when purchasing something over the web). Users concerned with click stream data could invoke the proxy and register their personal preferences with convention sites and with the sites they visit.

7. P3P EXTENSIONS

Within P3P, clients can specify whether they opt-in or opt-out of activities proposed by a web site. In consideration of user privacy, many groups may prefer to impose or attach additional constraints and rules to opt-in, opt-out, decisions. For instance, a client might decide that it is okay for a company to contact him/her every weekday between 6 and 8 pm, but would like to opt-out of being contacted at all other times. Another example is clients who would allow third-party access to their PII if they were properly compensated (e.g. a 5% discount on all purchases). Currently, no such rules are defined within the P3P framework.

If we use an RDF binding of P3P, we can provide additional statements to declare such rules in a P3P preference file. A constraint rule describing recurring time intervals where a condition applies (e.g. opting out of being contacted) can use the proposed RDF schema shown in Appendix II (defined in the "SCC" namespace). The schema is inspired by the recurrence rules used in the iCalendar [24] specification. As an example, a P3P preference file containing the following statement regarding a contact purpose:

```
<p3p:purpose rdf:ID="data-purposes">
  <p3p:contact/>
</p3p:purpose>
```

is attached to additional time constraints using the following RDF statements:

```
<rdf:Description about="#data-purposes">
  <SCC:time-constraint>
   <SCC:Recurrence>
    <SCC:start-dt>2001-11-07T0200Z</SCC:start-dt>
    <SCC:end-dt>2001-11-07T0300Z</SCC:end-dt>
    <SCC:frequency>DAILY</SCC:frequency>
    <SCC:count>300</SCC:count>
    <SCC:by-day>
      <rdf:Bag>
            <li>MO</li>
            <li>TU</li>
```

```
            <li>WE</li>
          </rdf:Bag>
        </SCC:by-day>
      </SCC:Recurrence>
    <SCC:time-constraint>
  </rdf:Description>
```

The statements specify that the client opts-in of the data purposes (i.e. being contacted) every Monday, Tuesday and Wednesday of the week between 6 and 7 pm (PST), starting November 11th, 2001 and continuing for 300 occurrences. The "time-constraint" RDF property is an extension of the traditional P3P properties for data purposes, and the domain of the property is an SCC Recurrence resource.

Other extensions to the P3P RDF schema, e.g. constraints allowing Web sites to specify financial incentives to persuade clients to give up their PII will be discussed in a forthcoming publication.

8. CONCLUSION

P3P as it exists today neither protects privacy nor empowers society to establish the social contracts needed to protect PII data and restrict the use of such data by various recipients. While government legislation can play a significant role in regulation the use of PII data, it is desirable to also invoke market forces. A suitable Social Contract Core is possible which will allow individuals to rapidly share information regarding privacy policies choices and the implementation of policies at competing websites. Such a three-component architecture will empower communities, encourage the acceptance of new technology, and encourage rapid access to and exchange of information essential to economic prosperity in the information age.

9. APPENDIX I

The Resource Description Framework (RDF) provides a mechanism for describing META-data, or data about data [25]. P3P policies are encoded in XML, but they may also be represented using RDF [26] as shown in Table 6A. In RDF, entities such as P3P data types are resources, and can be referenced from other documents. In the Table 6B, the header of an HTML page contains RDF data describing the relationship between form fields on the page and P3P data types. Such linking information can be used by an agent to provide more accurate feedback on policies behind data gathered on a particular Web page, and to preset fields according to local preferences. For instance, suppose field1 on a page represents a web-form field in which data of type name is collected. The data collected from this field will then be governed by the rules for type name in the governing P3P policy. The checkbox on the page ("checkbox1") is associated with the P3P contact purpose and is used to express the end-user's opt-in/opt-out choices for the contact purpose, where the governing P3P policy allows these choices to be made. This field can be preset by the SCC proxy to the opt-in or opt-out preference as specified by the client. The example also shows how P3P policy files can be associated with Web pages using the HTML <link> tag. Alternatively, this association can be made by pointing to a P3P policy using an RDF attribute in the header of the page, e.g.

```
<rdf:Description about="http://www.site.org/index.html">
  <p3p:policy rdf:resource =
"http://www.site.org/w3c/p3pv1.rdf"/>
</rdf:Description>
```

RDF P3P (at http://www.site.org/w3c/p3pv1.rdf):

```
<rdf:RDF>
 <xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:p3p =
"http://www.w3.org/2000/07/p3pmodel/p3prdfschema.txt">
 <p3p:Policy rdf:about =
"http://www.site.org/w3c/p3pv1.rdf">
  <p3p:data-group>
   <rdf:Bag rdf:ID="DG01" >
    <rdf:li>
     <p3p:name/>
     <p3p:jobtitle/>
    </rdf:li>
   </rdf:Bag>
  </p3p:data-group>
 </p3p:Policy>
 <rdf:Description aboutEach="#DG01"
rdf:bagID="stmt1">
  <p3p:purpose><p3p:develop/></p3p:purpose>
 </rdf:Description>
</rdf:RDF>
```

Table 6A. A Web Site may have a P3P policy encoded in RDF.

HTML document (at http://www.site.org/index.html)

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN"
 "http://www.w3.org/TR/REC-html40/loose.dtd">
<html xmlns:rdf = ”http://www.w3.org/1999/02/22-rdf-syntax-ns#”
  xmlns:p3p=”http://www.w3.org/2000/07/p3pmodel/p3prdfschema.txt#”>
  <head>
     <link rel=”P3Pv1” href=”http://www.site.org/w3c/p3pv1.rdf”/>
           <rdf:Description about=”http://www.site.org/index.html#field1”>
                    <p3p:data-type rdf:resource=” http://www.w3.org/2000/07/p3pmodel/p3prdfschema.txt#name”/>
          <scc:field-type type=”data-collection”/>
      </rdf:Description>
      <rdf:Description about=”http://www.site.org/index.html#field2”>
                    <p3p:data-type rdf:resource=” http://www.w3.org/2000/07/p3pmodel/p3prdfschema.txt#jobtitle”/>
          <scc:field-type type=”data-collection”/>
     </rdf:Description>
      <rdf:Description about=”http://www.site.org/index.html#checkbox1”>
           <p3p:purpose><p3p:contact/></p3p:purpose>
          <scc:field-type type=”option”>
                   <scc:optin value=”true”/>
                  <scc:optout value=”false”/>
          </scc:field-type>
      </rdf:Description>
  </head>
 <body>
   <H1>Please enter your name and job title:</H1>
             <form action=”…”>
           Name:<input type=”text” rdf:ID=”field1” />
          Title:<input type=”text” rdf:ID=”field2”/>
                    Contact? <input type=”checkbox” rdf:ID=”checkbox1”/>
          <input type=”submit”/>
      </form>
</body>
</html>
```

Table 6B. Forms in HTML pages are associated with P3P data types using RDF predicates.

Note that the RDF schema for P3P is still under development by the W3C consortium, and is subject to change [26]. This schema will need to be extended to be able to classify web form fields as collection points for a data type, or as places where preferences about the governing policy are expressed. Alternatively, this additional information could be expressed with tags in a separate namespace from the P3P RDF specification. These are shown in Figure 6B as elements in the “scc:” name space.

10. APPENDIX II

Table 7 shows a proposed extension to P3P, an RDF schema for recurring time intervals in the Social Contract Core.

```
<rdfs:Class ID=”Recurrence”>
    <rdfs:comment>Recurrence class for specifying recurring time intervals</rdfs:Comment>
    <rdfs:subClassOf rdf:resource=”http://www.w3.org/2000/01/rdf-schema#Resource”/>
</rdfs:Class>
<rdf:Property ID=”start-dt”>
    <rdfs:comment>Start date/time (optional)</rdfs:comment>
    <rdfs:range rdf:resource=”#Recurrence”/> <rdfs:domain rdf:resource=”http://www.w3.org/2001/XMLSchema.xsd#dateTime”/>
</rdf:Property>
<rdf:Property ID=”end-dt”>
    <rdfs:comment>End date time (optional if duration is specified)</rdfs:comment>
    <rdfs:range rdf:resource=”#Recurrence”/> <rdfs:domain rdf:resource=”http://www.w3.org/2001/XMLSchema.xsd#dateTime”/>
</rdf:Property>
<rdf:Property ID=”duration”>
    <rdfs:comment>Duration of the interval, in milliseconds</rdfs:comment>
    <rdfs:range rdf:resource=”#Recurrence”/> <rdfs:domain rdf:resource=”http://www.w3.org/2001/XMLSchema.xsd#integer”/>
</rdf:Property>
<rdf:Property ID=”until”>
    <rdfs:comment>End date/time (optional). Do not use if a count property is specified!</rdfs:comment>
    <rdfs:range rdf:resource=”#Recurrence”/> <rdfs:domain rdf:resource=”http://www.w3.org/2001/XMLSchema.xsd#dateTime”/>
</rdf:Property>
<rdf:Property ID=”count”>
    <rdfs:comment>Max number of occurrences (optional)</rdfs:comment>
    <rdfs:range rdf:resource=”#Recurrence”/><rdfs:domain rdf:resource=”http://www.w3.org/2001/XMLSchema.xsd#integer”/>
</rdf:Property>
<rdf:Property ID=”frequency”>
    <rdfs:comment>Frequency. One of ‘SECONDLY’, ‘HOURLY’, ‘DAILY’, ‘WEEKLY’ or ‘YEARLY’</rdfs:comment>
    <rdfs:range rdf:resource=”#Recurrence”/><rdfs:domain rdf:resource=”http://www.w3.org/2000/01/rdf-schema#Literal”/>
</rdf:Property>
<rdf:Property ID=”by-hour”>
    <rdfs:comment>Hours within a day when recurrence occurs</rdfs:comment>
    <rdfs:range rdf:resource=”#Recurrence”/><rdfs:domain rdf:resource=”http://www.w3.org/2000/01/rdf-schema#Container”/>
</rdf:Property>
<rdf:Property ID=”by-day”>
    <rdfs:comment>Days within a week when recurrence occurs, e.g. “WE’, ‘TH’, ‘FR’</rdfs:comment>
    <rdfs:range rdf:resource=”#Recurrence”/><rdfs:domain rdf:resource=”http://www.w3.org/2000/01/rdf-schema#Container”/>
</rdf:Property>
<rdf:Property ID=”by-month-day”>
    <rdfs:comment>Days within a month when recurrence occurs, e.g. ‘1’, ‘2’, ‘3’</rdfs:comment>
    <rdfs:range rdf:resource=”#Recurrence”/><rdfs:domain rdf:resource=”http://www.w3.org/2000/01/rdf-schema#Container”/>
</rdf:Property>
<rdf:Property ID=”by-year-day”>
    <rdfs:comment>Days within a year when recurrence occurs, ‘1’ - ‘366’</rdfs:comment>
    <rdfs:range rdf:resource=”#Recurrence”/><rdfs:domain rdf:resource=”http://www.w3.org/2000/01/rdf-schema#Container”/>
</rdf:Property>
<rdf:Property ID=”by-week-no”>
    <rdfs:comment>Weeks within a year when recurrence occurs, ‘0’ – ‘53’</rdfs:comment>
    <rdfs:range rdf:resource=”#Recurrence”/><rdfs:domain rdf:resource=”http://www.w3.org/2000/01/rdf-schema#Container”/>
</rdf:Property>
<rdf:Property ID=”by-month”>
    <rdfs:comment>Months weeks within a year when recurrence occurs, ‘1’ – ‘12’</rdfs:comment>
    <rdfs:range rdf:resource=”#Recurrence”/><rdfs:domain rdf:resource=”http://www.w3.org/2000/01/rdf-schema#Container”/>
</rdf:Property>
```

Table 7. RDF Schema for Recurring Time Intervals in the Social Contract Core.

11. REFERENCES

[1] Cranor, L., Langheinrich, M., Marchiori, M., Presler-Marshall, M., Reagle, J., The Platform for Privacy Preferences 1.0 (P3P1.0) Specification, http://www.w3.org/TR/2001/WD-P3P-20010928/

[2] J. Kaufman, J. Ruvolo, and D. Ford, Tempus Fugit and the Need for an e-Social Contract, 5th International Conference on Autonomous Agents, Montreal, Canada. Full paper in Proceedings Workshop 11, pg 77, May, 2001 and extended abstract in Conference Proceedings, pg174, May, 2001.

[3] C.E. Unterberg, Tobin, ASP News, "ASPs Will Help Drive Our Global Economy" http://www.aspnews.com/analysis/analyst_cols/article/0,2350,4431_420701,00.html

[4] Macfarlane, R., Jacobs, G., and Asokan, N., The Role of Money & Internet in Social Development, Pacific Rim Allied Economic Organizations Conference Bangkok, Thailand (1998), http://www.icpd.org/development_theory/presentation_to_western_economic.htm

[5] Hobbes, T., The Leviathan (1650). http://www.orst.edu/instruct/phl302/texts/hobbes/leviathan-contents.html

[6] Hobbes, T., De Cive (The Citizen) Philosophical Rudiments Concerning Government and Society, (1651). http://www.constitution.org/th/decive.htm

[7] Locke, J. A Essay Concerning the true original, extent, and end of Civil Government" (1690). http://www.constitution.org/jl/2ndtreat.htm

[8] Rousseau, J.-J. Du contrat social (The Social Contract) 1762, Translated by G. D. H. Cole, public domain http://www.constitution.org/jjr/socon.htm

[9] Kahn, B., et al. History of Communications Infrastructures http://www.cs.berkeley.edu/~gribble/cs39c/Comm/telephone/telephone.html

[10] U.S. Census Bureau, Historical Census of Housing TablesTelephones, http://www.census.gov/hhes/www/housing/census/historic/phone.html

[11] U.S. Department of Commerce News, "9-in-10 School-Age Children Have Computer Access; Internet Use Pervasive", Sept 6, 2001. http://www.census.gov/Press-Release/www/2001/cb01-147.html

[12] Cranor, L., and Reagle, J., "Designing a Social Protocol: Lessons Learned from the Platform for Privacy Preferences Project" Telecommunications Policy Research Conference, Alexandria, VA., Sept. 27, 1997.

[13] Cranor, L., and Reagle, J., "Protocols for Automated Negotiations with Buyer Anonymity and Seller Reputations", Proceedings of the 1997 Telecommunications Policy Research Conference.

[14] Krishna, A., What Should You Be Asking About the Privacy Rights of Your Customers?, S.C. Magazine, October 2001. http://www.scmagazine.com/scmagazine/sc-online/2001/article/045/article.html

[15] Coyle, K. Protecting Privacy, NetConnect Library Journal, Winter, 2001, and P3P: Pretty Poor Privacy? A Social Analysis of the Platform for Privacy Preferences (P3P) http://www.kcoyle.net/p3p.html

[16] Mulligan, D., Schwartz, A., Caoukian, A., Gurski, M., "P3P and Privacy: An Update for the Privacy Community", Center for Democracy and Technology, March, 2000. http://www.cdt.org/privacy/pet/p3pprivacy.shtml

[17] Gramm-Leach-Bliley Act (1999). http://www.senate.gov/~banking/conf/

[18] Health Insurance Portability and Accountability Act, (1996), http://www.hcfa.gov/hipaa/hipaahm.htm

[19] Congress passed the Children's Online Privacy Protection Act (1998), http://www.ftc.gov/opa/1999/9910/childfinal.htm

[20] See e.g., Information Commissioner, Responsible for the Data Protection & Freedom of Information Acts. http://www.dataprotection.gov.uk/

[21] See: Organisation For Economic Co-operation and Development (OECD) http://www.oecd.org/ Guidelines for Information Security and Privacy. http://www1.oecd.org/dsti/sti/it/secur/prod/PRIV-EN.HTM

[22] W3C Platform for Privacy Preferences, Public Archive of Public Comments, http://lists.w3.org/Archives/Public/www-p3p-public-comments/

[23] WBI Development Kit for Java, IBM alphaWorks site, http://www.alphaworks.ibm.com/tech/wbidk

[24] Internet Calendaring and Scheduling Core Object Specification (iCalendar), http://www.imc.org/rfc2445

[25] Resource Description Framework (RDF) Model and Syntax Specification. http://www.w3.org/TR/REC-rdf-syntax/

[26] RDF Model schema for P3P http://www.w3.org/2000/07/p3pmodel/

[27] A P3P Preference Exchange Language 1.0 (APPEL 1.0) http://www.w3.org/TR/P3P-preferences.html