CM50175
Research Project Preparation
Project Proposal: A P3P Editor

Thierry Wouters
MSc Computer Science
University of Bath

6 May 2003

Contents

Introduction

The purpose of this project is to build a P3P editor for use by universities' Web Teams. P3P is a specification that enables web sites to issue their privacy statements in a machine-readable format.

In order to further describe this project, this introduction will argue why web privacy is an important topic for universities, why their privacy statements should be expressed in a machine-readable format, and how this can be achieved. The requirements and objectives of the project will then be presented.

Web privacy is often considered to act as a brake on the development and the popularisation of e-commerce, but why should universities be worried about it? The answer is twofold: Universities can be seen as an ISP by its students and staff and more and more online services are proposed by universities to their members via their web sites.

Staff and students generally use the universities' infrastructure to access the Internet, to send and receive e-mails, to print documents,etc. The amount of space required to report an exhaustive list of all the purposes this infrastructure is used for makes it unsuitable to do so here. From this point of view, the relation between students or staff and their universities is very similar to the relation customers have with their ISP. The main difference however, is that from a juridical point of view, there is nothing stopping the universities from collecting and exploiting individuals' data in relation to their usage of IT resources. Why should a university not try sharing this information across the whole organization, or with external parties, were it not for ethical reasons?

The second aspect of the interest universities may have in web privacy is that their web site is used more and more as a service provider for staff, students, and prospective staff and students. Prospective staff and students can apply online, answer surveys, register for e-learning session, whilst registered staff and students can in addition get lecture notes, browse the library catalogue,. . . To what extent can members be sure that, for example, all this data is not linked to their records or that this data is not used or communicated to third parties for marketing purposes?

There is a need for universities to communicate their privacy policies to their members. This project will concentrate on the optimization of the communication of privacy statements related to the second aspect mentioned above.

Publishing machine-readable statements is a major improvement compared to human-readable statements, in that the content of an actual privacy policy can be interpreted by a web browser, rather than simply being displayed to the user. The quality of the communication is thus determined by the browser and takes the user's preferences into consideration. Web site owners can therefore concentrate on the content of the statements rather than on their appearance, as the content will be rendered by the user's browser.

The communication can also be improved in that statements do not necessarily have to be displayed as a text content on the user interface. The browser can alternatively display icons or symbols, alerting the user of the currently visited web site's privacy practice. In such cases, the user is not required to have any particular knowledge of web privacy and is able to evaluate the degree of risk he is facing at any time.

The ultimate purpose of this project is to develop a tool that can be used by universities' Web Teams to assist them in the process of creation (or updating) and deployment of such P3P policies, so that no particular knowledge of the technologies involved in this process will be required.

The system will not only be tested to ensure that all the functionalities that are required have properly been implemented, but it will also be tested in collaboration with prospective users to ensure that it can be used by the target user group without any technical assistance, as stated in the user requirements.

The functionalities and the interaction of the users with the system will be designed and developed so as to optimize it for this community of users. It therefore involves the analysis of the universities' needs in terms of privacy statements communication, the analysis of existing editors, a usability analysis, and the testing of these user requirements. The system will also be implemented as a web-based application so as to free the users from having to install and configure it, and to provide them with the same look and feel of any other web interface.

This proposal will, in its first part, demonstrate the importance of web privacy by looking at the kind of data web sites can collect about their visitors and what this data can be used for. It will then be shown how human-readable statements can help in preventing Internet users'right from being scorned and why the P3P specification should be chosen to implement them. These arguments will be illustrated by showing how existing P3P-enabled web browsers interpret such statements and communicate them to the user.

The second part of this proposal will more formally define the above-mentioned motivation and goals of the project.

The third and last part will give a further description of the project by presenting how the requirements for the system will be collected, by specifying the functional and non-functional requirements, and by describing the system's architecture and the development decisions that have already been or will be taken.

Part I

Project Background

1 The Web Privacy Landscape

This section is intended to be a brief introduction to the web privacy domain. After having define what web privacy is, its importance will be stated and the reason why web privacy nowadays is causing controversy will be investigated. Various approaches to the solving of this problem will then be presented.

1.1 From privacy to web privacy

The recent advent of the Internet as a widespread information medium has implied quite an impressive number of commonly used neologisms as well as the use of already existing terms in this new context, including ″web privacy″. Providing a clear and unambiguous definition of this term before any further developments is thus required.

The Oxford Dictionary defines ″privacy″ as ″a person's right to the state of being private and undisturbed″. In order to protect the individuals' privacy, laws usually grant them ″the right. . . to control the collection, use, and dissemination of their personal information that is held by others″ (Electronic Privacy Information Center, 2000). Transposing these terms to the Internet world, web privacy can be interpreted as the right Internet users have to control the information they reveal while browsing a web site, as well as the right to be informed of the usage that that will be done of that information (Cranor, 2002).

From the above, it should be clear that web privacy differs from web security, though both are related. Web security aims to prevent a set of data from being accessed by unauthorized entities (Garfinkel, 1997). A privacy policy claiming that collected data will not be disclosed to third parties is meaningless if the data is not stored in a secure database. Conversely a secure database is somewhat absurd if the data it holds is not protected by privacy policies and/or commercial interests.

1.2 Web privacy matters

As the Internet is becoming a distribution channel for a wider and wider variety of industries, the amounts at stake when talking about e-commerce development are at the scale of billions of U.S. dollars.

According to Cranor (2002), web privacy policies have a high impact on Internet users' online behaviour and attitude towards online shopping. Surveys have shown that the lack of information on privacy policies makes Internet users reluctant to shop online.

Forrester Research (2001) asserts that 60% of online customers are concerned about the disclosure of their personal data and puts forward a 2001 estimated loss of profit of $15 billion as a direct outcome of this fear. Cyber Dialogue, a technology and service consultancy, has published the results of a survey carried out in October 2001 on a sample of 500 Internet users revealing that 27% of their interviewees had abandoned an order online because of privacy concerns, 21% of them preferring to switch to an offline order.

Though these figures are given by industry independent companies, they have to be considered with the usual cautiousness. It is indeed difficult to estimate to what extent the subjects of these surveys have voluntarily or unconsciously mixed up web privacy with web security. These two notions, as explained above, are closely related and it may be difficult for non-experts to perfectly distinguish them. Beyond the figures, it is the need for well-defined and accessible privacy policies that this surveys have revealed that had to be pointed out.

Knowing the web privacy financial and commercial stake, the chaotic situation depicted here may seem unrealistic. Why do web sites not take a greater care of making their privacy policies acceptable for customers? This question does not have a simple answer. An analysis of the dilemma companies are facing is required to understand the origin of the web privacy problem.

1.3 The web privacy dilemma

The dilemma companies are facing when taking decisions concerning their privacy policies will be exposed in this subsection by looking at the data companies can gather about their customers and the usage they can do of this data.

1.3.1 Customer data collection techniques and exploitation

Let us first identify the data a web site owner is technically able to extract from an online customer. This can be classified in two categories, depending on its origin:

- **Online Forms:** Obviously, all the information provided by the filling of online forms and sent to the web server can be exploited. It is usually referred to as the *personal customer details*.
- **HTTP request headers:** More subtly and often ignored by Internet users, a fundamental requirement for the HTTP protocol to work properly is to reveal personal user information. The Internet is built on a client-server architecture, which means that each page displayed in a web browser (the client) is a response to an HTTP request sent to the web server (Cho *et al.*, 1997). This request contains, along with the requested page URL, the client's IP address, the previous requested URL (known as the *referrer*) and information allowing the identification of the client's operating system and language preferences. These HTTP features can be used by web sites to track their customers' browsing behaviour

(Cranor, 2002). The data which is collected in the HTTP request headers is commonly called *tracking data* and is usually saved in log files by the web server.

The former tells the web site's owner who the customer is whilst the latter allows him to track his browsing.

During the client-server interaction, third parties may also access parts of the tracking data by means of a simple trick, as illustrated on figure 1. A web page often contains third party contents such as pictures or adverts. In reality, third party contents are, in the web page, nothing but references to their web sources. The web page is thus first downloaded from its hosting domain without the third party content. The client's browser then discovers that they have to be downloaded and sends HTTP request to the mentioned URLs, which are not required to belong to the same domain as the web page that contains them. A third party such as an advertising company is thus sent the HTTP request header, amongst which the referrer, enabling it to track customers across all the sites where their adverts appear.

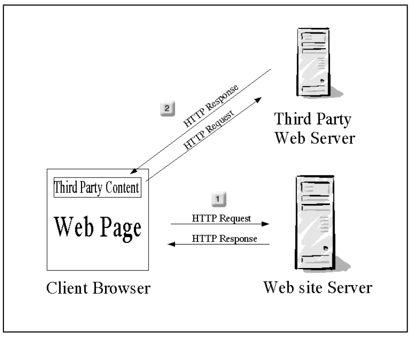


Figure 1: Third parties data collection

It is worth mentioning that third party contents do not necessarily have to be visible banner adverts or images. For example, a 1x1 transparent GIF image that is invisible to the end-user triggers exactly the same HTTP interactions between a browser and a third party web server.

Having identified the data customers reveal whilst browsing web sites, an analysis of how this information can be used and the extent to which this use threatens the customers' privacy can be carried out. Figure 2 illustrates the complex set of combination of data exploitation practices web site owners and third parties can achieve. Figure 2 clearly shows the possible outcomes of these exploitation practices. Each will be discussed from the customer's privacy perspective.

1.3.2 Commercial use of personal details

The data collected by means of online forms can be used to classify users in so-called *clusters*. Companies typically use this information to target their marketing campaign on specific clusters of clients. Such practices are nothing but the transposition of widespread practices to the online market. Customers can also have their personal data sold to third parties but this is not, regarding offline transaction practices, a really new privacy issue either.

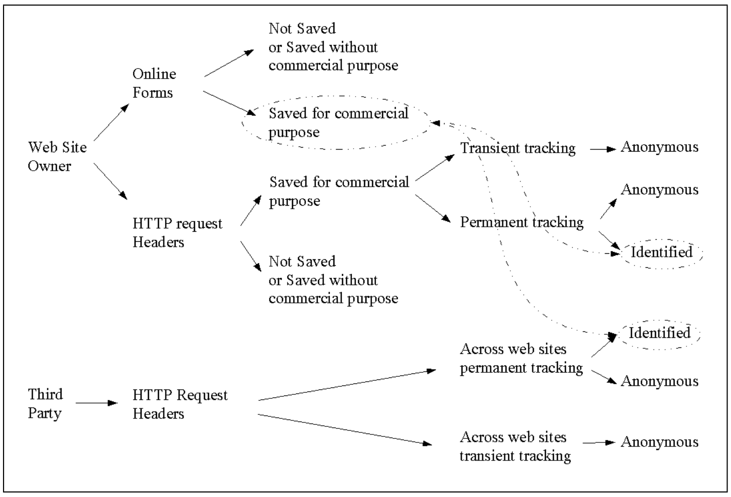


Figure 2: Customer data exploitation practices

The customer can relatively easily be aware of the web site's data collection and management practices.

1.3.3 Transient anonymous tracking

The link between personal details collected via online forms and tracking data can so far only be achieved by the web site owner by comparing the form submission IP address with the IP address contained in further HTTP request headers.

IP addresses being most of the time dynamically allocated on a per session basis by the customer's ISP, the collected information is transient and cannot be linked to information gathered in other sessions, unless the customer accepts to fill in the form for each session, which he is unlikely to. Considering that in addition third parties may gather some tracking data as well, the customer may at worse be shown customized adverts, which is not deemed to be a major privacy issue (Electronic Privacy Information Center, 2000).

Gavray (2002) points out that, from an economical point of view, web sites have a high interest in collecting as much information as possible on their customers. Most commercial companies are indeed putting a lot of efforts to exploit this data as far as statistical analysis and data mining techniques allow it. Customers receive in return a personalized browsing experience and service. Gavray identifies this interaction as a "win-win" relationship, as long as the data collection and exploitation does not end up in customers' privacy violation.

1.3.4 Permanent tracking

Web site owners can use a number of different techniques to achieve a permanent tracking of their customers. They can require users to log themselves in whenever the customer visits the web site, or they can use so-called *cookies*. Logins allow web site owners to link each HTTP request to a unique arbitrary key, hence the permanent tracking. The tracking may not be linked to the customer's personal data and can therefore remain anonymous. The login alerting the customer that he is identified and the information revealed to third parties remaining as seen before, the privacy issues are still quite limited. Cookies, on the other hand, cause much more controversy.

Roughly said, a cookie is a bit of text used by a web site server to tag the customer's browsers. Each time a "tagged" browser will request a page hosted on the web site, it will return this cookie. Cookies therefore work across multiple sessions. The permanent tracking is achieved by giving the cookie a unique identifier.

As cookies are silently installed on the client's browser, a customer profile can now be drawn without the customer's knowledge and/or consent. In this situation, one could wonder if the relationship between the web site and the customer is still well balanced. However, the collected information remains within the web site company and privacy laws may protect the customer.

1.3.5 Third parties' permanent tracking and identification

Major privacy issues arise when third parties attempt to identify Internet users on a permanent basis. In theory, a cookie is only returned to a web site belonging to the same domain than the web site that set it. The tracking of the customer's behaviour is thus limited to that particular domain. In practice, third parties are able, by means of third parties adverts or pictures inserted in the visited web page, to extend the tracking area to multiple domains, thereby increasing drastically the variety and accuracy of the user profiles (Cranor, 2002; Kristol, 2001). Third parties can set a cookie together with the adverts sent to the customer and, using this cookie and the referrer in the HTTP request header, track customers across all the sites where their adverts appear on a permanent basis. Advertising networks claim that it allows them to present adverts that are more likely to interest the customers, but, as pointed out by Kristol (2001),

> they can also charge more money for them, or they can hope for more revenue through a higher "click-trough rate", or they can hope to sell the profiles, perhaps after linking them with personally identifying information.

Here, the privacy issues are much more important and the legality of such practices is now questionable. It has to be mentioned that the data collected by third parties remains anonymous unless one and only one web site reveals the customer's personal details to a third party. A dotted line in figure 2 illustrates this disclosure requirement.

1.4 Tackling the web privacy problem

Swire (1997), in a paper written for and published by the U.S. Department of Commerce, identifies and examines three theoretical models that could solve the web privacy problem. In practice, the application of each of these models is restricted by hypothesis that makes the model irrelevant if not met. The purpose here is to briefly describe these models so that the different approaches to the solving of the web privacy problem can be illustrated.

1.4.1 The pure market model

The pure market model rests entirely on the financial incentive companies have to respect their customers' privacy, as seen in the previous section. It is here assumed that the customers have access to the companies' privacy statements and have the ability to understand the implications these statements have for their privacy.

Studies have proven that this is currently not the case on the electronic market. More than the half of Internet users never or rarely read web site's web privacy policies, for the majority of them by lack of time and interest or because they are difficult to understand. Internet users mainly complain about the lack of standardization of the layout of privacy policies, when a conventional summary or a checklist could be used (Privacy Leadership Initiative, 2001; Culnan & Milne, 2001).

Consequently, there is an asymmetry of information between the customers and the companies in that companies do not divulge their privacy statement in a convenient way for the customer. On top of that, most customer lack some proficiency in web privacy and are reluctant to do the investment in terms of time and effort to acquire such a proficiency, preventing them from being aware of the implications companies' statements may have.

As long as such an asymmetry remains, the pure market model will not yield a fair equilibrium for both sides. It is therefore advocated by Swire that the market alone will not solve the web privacy problem, though any opportunity to near a pure market equilibrium should be considered.

1.4.2 The pure enforcement model

The pure enforcement model reckons the market limitations and suggests a governmental regulation of the customers' data collection, tracking, and use. Though the application of such a model can rapidly yield an acceptable level of customers privacy protection, it has a considerable cost in that the enactment and the enforcement of the legal framework have to be administered and industries have to make the required investments to ensure their compliance with these rules.

It also has to be assumed that, as stated by Swire, ”the people drafting and enforcing the rules are competent, well-informed, and wish to achieve the public good in the area of privacy protection”.

This is however the model that has been chosen by the European Commission. The EU now has a comprehensive and harmonized legal regime for privacy protection[1],

[1] The Directive on Protection of Personal Data (95/46/EC)

stating amongst others the circumstances of data collection, the restrictions to third parties transfers of data and their use for ”secondary” purpose.

1.4.3 Self-regulation

Self-regulation is presented by Swire as an alternative to the pure models which might create an acceptable protection of customers’ privacy without excessive costs for companies, customers, or governments. Self-regulation relies on industries to create and issue their own set of guidelines for good privacy practice. This generally requires the setting up of industry organizations that will be charged for that.

It is important here to mention that these guidelines do not necessarily have to be legally enforced. They simply may rather be made available to industry members, government agencies, and the general public. As a substitute to a proper legal enforcement system, industry organizations may publish a list of companies that do not comply with these industry guidelines, or cancel them the right to use a seal or logo, or disqualify them from membership in the association. Customers can therefore more easily be aware of the rules and the companies they can trust.

The intrinsic advantage of self-regulation is that it is built on the expertise of industry members, who are those who know the best how customer data are collected and used and those that are likely to be able to reckon whether the cost for the companies to comply with these guidelines will be affordable. Swire however tempers these arguments by mentioning that self-regulation may lead to the creation of cartels and can be used to wield market power.

Self-regulation is the approach that the U.S. Administration encourages industry actors to follow (National Telecommunications and Information Administration, 1998) to extricate the U.S. e-market from its current confusing situation, where web privacy is regulated by a patchwork of federal and sector-specific laws that apply to narrow industry sectors (Cranor, 2002).

1.5 Privacy tools

A totally different and much more technocratic approach to the web privacy problem is to try stopping customers from revealing data about their browsing activity. A variety of free or commercial tools claiming that they allow Internet users to anonymously surf the web are available[2]. Before closing this first section on web privacy, the principles on which such tools are based have to be explained and their efficiency needs to be discussed.

Two kind of tools can be identified. On the one hand, ”Web privacy-aware” browser may offer the possibility to Internet users to set up some cookie handling parameters in order to adapt their browser to the level of privacy they want. The major weaknesses of these tools are that the set up in itself usually requires a minimum of knowledge in the web privacy domain and that it is not unusual to have web sites not func-

[2] A list of tools can be found on the EPIC.org web site. *http://www.epic.org/privacy/tools.html*

tioning if cookies are not allowed to work in the way the were supposed to. Browser manufacturers therefore set low privacy level default values, which a customer rarely modifies.

On the other hand, there are dedicated privacy tool aiming to hide the information a browser automatically reveals when sending HTTP request headers. The idea is to use proxies between the client and the server. Proxies are originally used to optimize users' Internet surfing by caching previously requested web page. Proxies can also be set up so that they modify the HTTP request they receive and forward. When a client sends an HTTP request for a page in a domain, the proxy will act as an intermediary that will request the web page without revealing the origin of that request. The web site server will receive a request from the proxy, not from the client. When the server sends back the requested page to the proxy, it is up to the proxy to remember which one of its clients had requested that page in order to pass back the server response.

These tools, despite what their advertisements may sometimes suggest, do not prevent a browser from sending HTTP request headers information. As explained above, this information is compulsory for the client-server architecture on which the Internet is built to work. This information is only kept by the proxy. As emphasized by Cranor (2002), "users of 'anonymizing' proxy services are not anonymous to the proxy or to their own ISPs, who may log their users' web activities."

Consequently, the owner of the proxy knows everything about his customers. One could wonder what the usage free 'anonymizing' proxy owners can do with these data. Commercial tools may provide a better guarantee that the service owner will not misuse the data they are able to collect about their clients, but here one should rather put a question mark on the fact the customers have to pay to exercise one of their fundamental rights.

1.6 Conclusion

It is now obvious that customers personal and tracking data have a high commercial value that can be bought and sold as any commercial resource.

Some companies have as unique business the collection and exploitation of customer data. Their business is based on the commercial relationship they have with web site owners that ask a financial reward to make up for the appearance of third party content on their web pages.

However, these companies somehow have access to the customers' data through the intermediary of commercial companies. The source of the problem is thus the trade off commercial companies have to achieve between the extra revenue they earn by exploiting, selling, or letting third parties access their customers' information and the loss of profit they bear whilst discouraging customers to buy their products because of these data privacy issues.

The web privacy problem can be tackled by applying different models that all have their own strengths and weaknesses. It is however fair to say that the international nature of the Internet and the lack of harmonization of the decision that are taken (if any) to set about the problem make it difficult to efficiently protect customers from having their privacy right scorned while they are online. As a consequence, the web privacy problem is not expected to be solved in the short term.

2 The Platform for Privacy Preferences Project

Having identified the web privacy problem and the various approaches that can be used to tackle it, this section will introduce the Platform for Privacy Preferences (P3P) standard by defining and commenting its objectives and its implementation principles.

A brief assessment of P3P will be achieved by discussing how P3P is expected to contribute towards the solving of the web privacy problem and its expected level of acceptance by industries.

2.1 P3P: Origin, definition, and objectives

This first subsection aims to provide an understanding of how P3P has been developed, what it does, and what its objectives are.

2.1.1 Origin and development of P3P

P3P has been developed by the World Wide Web Consortium (W3C) and its current version (P3P 1.0) has reached the status of W3C Recommendation in April 2002 [3].

The W3C is a consortium of industry actors and experts aiming to lead the technical evolution of the Web with the long term goals of universal access and optimal use of resources of the web, considering the legal, commercial, and social issues it raises (Herman, 2003). To achieve these goals, the W3C develops and publishes specifications, the most famous one probably being HTML.

Depending on their level of maturity and consensus, the W3C specifications are given a status, ranging from "Working Draft" to "W3C Recommendation". W3C recommendations are considered to be appropriate for a widespread deployment (World Wide Web Consortium, 2001).

2.1.2 Definition and objectives

From a technical point of view, P3P is a standard for communicating privacy policies of web sites to clients in a machine-readable format, and a protocol that enables web browsers to read and process these P3P policies (Cranor, 2002).

[3] The Platform for Privacy Preferences 1.0 Specification, *http://www.w3.org/TR/P3P*

To enable web sites to communicate their privacy policies, the P3P specification defines a standard set of data a web site may wish to collect and categorizes the general use that may be made of this data. In addition, P3P defines a format in which this information has to be encoded to make it machine-readable, as well as a means to associate it with web pages or cookies, and a mechanism for transporting it over HTTP (World Wide Web Consortium, 2002).

From the above, it should be clear that P3P is a communication standard with a purely informational aim. The idea is to enable web browser to interpret web sites privacy policies so that they can accordingly alert the user. As emphasized by Cranor, "the P3P project seeks to enable the development of tools for making informed decisions about when and if personal information should be revealed". The ultimate aim of P3P is thus to make privacy policies more transparent and accessible, and to enable Internet developers to integrate a privacy dimension in their tools (e.g. P3P-enabled browsers, search engine, comparison shopping services, . . .).

P3P does not aim to provide anonymity to Internet users, nor does it guarantee the compliance of web sites with their privacy policies. Its role is to enable tools such as web browsers to inform Internet users of web sites' privacy policies in a convenient and easily understandable way.

It should be noted that P3P has a positive side-effect on companies' approach to privacy. By providing standard sets of data and potential uses of this data, it forces companies to execute a thorough and systematic analysis of their data collection practices and use, revealing sometimes unintentional and unsuspected consumers privacy rights violation (Cranor & Wenning, 2002).

2.2 P3P and the web privacy problem

P3P in itself does not resolve the web privacy problem. It rather contributes to the solving of the problem by allowing users to be informed of web sites' privacy practices without requiring from them proficiency in the web privacy area. The P3P approach to web privacy is an application of the pure market model exposed in section 1.4. Putting P3P in this theoretical framework, P3P aims to reduce the asymmetry of information between companies and customers for the benefit of a fairer market equilibrium.

Having reached the level of W3C recommendation, it is reasonable to estimate that the competition between internet tool developers will lead them to integrate P3P in their future tools. Microsoft's Internet Explorer 6 and Netscape 7, for example, already implement parts of the P3P specification by allowing control of the cookies according to their privacy statements.

P3P may also work in conjunction with other privacy, anonymity, and security technologies as well as with laws and self-regulatory programs, and thereby encourages privacy experts to foresee a solution to the web privacy problem (Cranor & Wenning, 2002).

2.3 P3P-enabling web sites

P3P-enabling a web site requires the creation of P3P policies and the referencing of these policies to their respective web resources. A mandatory requirement for P3P policies and references is that they have to be readable by any browser, whatever the operating system and platform of the client.

Such a level of interoperability can be reached by encoding the policies and their references in text format files conforming to the eXtensible Markup Language (XML) specification, yet another W3C recommendation[4].

2.3.1 P3P and the eXtensible Markup Language

A huge amount of literature is available to describe XML. The purpose here is to briefly present the concepts that are essential to the comprehension of the P3P policies and reference file development process.

As stated by Johnston (2001), marked-up documents are text formatted documents containing data and information about the data, or meta-data. HTML is a well-know example of markup language. The HTML specification defines the tags that can be inserted in a text file to indicate to web browsers the format in which the content of the file should be rendered. The logical structure of a document is thus, in HTML, communicated to a human reader by means of presentation conventions. Consequently, HTML does not allow the content of a document to be interpreted by softwares. In order to exchange information between independent systems, an agreement between the parties involved has to be reached on a descriptive markup language. This language then can be used to tag the content of the documents exchanged between the systems such that the logical structure of the document is included in the document itself in a machine-readable way. Parsers will then have to be developed to format and process documents that implement these markup languages, allowing systems to store and retrieve the information and meta-information they contain.

XML is a specification allowing the creation of such markup languages. It can therefore be considered as a meta-language. XML lets people create their set of tags. These tags are defined either in a DTD (Document Type Definition) or in an XML Schema. Schemas and DTDs allow, in addition to the definition of tags, the definition of structural constraints on their use. The DTD specification is included in the XML 1.0 specification, whilst XML Schemas have their own specification[5]

The P3P specification defines the XML Schemas to which P3P policy reference files and P3P policies have to be conformed to. Any valid P3P policy and reference file can thus be interpreted by any software implementing a parser in combination with the P3P XML Schemas submitted by the W3C.

[4] Extensible Markup Language (XML) 1.0 (Second Edition). *http://www.w3.org/TR/REC-xml*

[5] XML Schema Part 0: Primer, W3C Recommendation, 2 May 2001, *http://www.w3.org/TR/2001/REC-xmlschema-0-20010502/*

2.3.2 P3P policy files

The P3P policy XML Schema defines all the markups that can be used in a P3P policy file, and thereby defines the data that such files can contain. This data can be classified in two broad categories: General assertions and Data-specific assertions (Cranor, 2002).

The general assertions provide information about :

- The location of a human-readable equivalent of the policy.
- An explanation of how to opt-in or opt-out for a data practice. An opt-in mechanism means that the user's data will not be exploited for a specific purpose, unless explicitly requested by the user. Conversely, an opt-out mechanism requires the user to explicitly request his data not to be used for a particular purpose.
- The web site contact information.
- The way people will be allowed to access their personal records.
- How disputes will be resolved.

Data-specific policy content refers to information related to :

- The consequence of providing data.
- The indication whether no identifiable data is collected.
- How data will be used and with whom it may be shared.
- the opt-ind and/or opt-out mechanisms that are available.
- The data-retention policy.
- The kind of data that is collected.

2.3.3 Policy reference files

The P3P policy files need to be referenced so that whenever an HTTP request for the web site resource is received, the URL of the policy related to the requested web resource can be returned to and processed by the client browser.

The content of a P3P reference file is determined by the P3P specification. The reference file may inform the clients about the URL where a policy is found, the parts of the web site that are (or that are not) covered by the policy, the cookies that are (or that are not) covered by the policy, the HTTP access methods for which the policy applies, and the lifetime of a policy or the policy reference file itself.

A nice consequence of the above is that each site can have different policies for distinct parts of the site. Should it be required, even each URL of a site can be associated with its own policy. The definition of the URL or set of URL to which a policy applies is easily declared in the policy reference file by means of inclusion and/or exclusion mechanisms.

In order to ensure that client browsers will find the reference file, its location has to be indicated by means of one of the following mechanism (Cranor, 2002; World Wide Web Consortium, 2002):

- **The ″well-known″ location.** The ″well-known″ location mechanism can be used by simply placing the reference file at the location `/w3c/p3p.xml`. A client requesting a resource from the site will first access this path to work out the policy related to the resource he is looking for. The main advantages of this mechanism are that it does not require any server configuration or modification to the site and that the P3P policies are processed by the browser prior to any resource transfer. Once the policy files and reference files are written, it is very easy to P3P-enable an existing web site using this mechanism. In cases where the web site has an important number of policy to reference, the time needed to transmit the reference file to the client may encourage web administrator to consider another referencing mechanism.
- **Embedded `link` tag.** In order to associate a single web resource with a policy, a `link` tag can be inserted directly in the resource. A policy reference can be referenced by a `link` tag placed in the `HEAD` area of the HTML document. Though this mechanism does not require to configure the server, it is quite tedious to implement for site with numerous resources.
- **HTTP header.** The server could be configured in such a way that the header of the HTTP response returned to the client indicates the location of the policy reference file.

2.4 P3P and human-readable statements

P3P has so far been advocated from the machine-readable statement redaction point of view. Using P3P may however have an impact on human-readable policies.

Any marked-up documents can be transformed to conform it to another markup language by means of so-called *stylesheets*. The W3C has published a stylesheet language specification named eXtensible Stylesheet Language Transformations[6] (XSLT) that allows the creation of such stylesheets.

P3P policies being, as explained above, XML files, their content can relatively easily be converted and interpreted as an HTML file by any web browser. The browser simply process the source XML file according to the presentation rules defined in the stylesheet associated to the document.

If different organizations are to use the same stylesheets on their own P3P policy files, the layout and presentation of their human-readable policy will be similar. It is not a utopia to foresee that sectorial associations may provide there members with such stylesheets, thereby indirectly bringing a certain level of standardization in the human-readable policies of their members.

[6] XSL Transformations (XSLT) Version 1.0, W3C Recommendation 16 November 1999, *http://www.w3.org/TR/xslt*

Another source for standardization can be foreseen from the fact that P3P-enabled browsers (see section 3 below) may offer as a built-in functionality such conversion of P3P policies into a human-readable format. Here, the user does not have to search for human-readable statements on web sites. At any time, he can simply ask his browser to fetch the current web site's privacy statement and it will display it according to its own P3P stylesheet. Provided that a user always uses the same browser, he will always be shown the same privacy statements format.

2.5 Conclusion

P3P allows web site to publish their privacy statements in a machine-readable format, so that web browsers can interpret and communicate their content to the user in an easily understandable way. By doing this, it is hoped that customers will be able to take informed decisions while surfing the Internet, thereby restoring a fairer equilibrium in their relation with web site owners.

By stating the P3P specification origin, it has been shown that, having reached the status of "W3C Recommendation", P3P is a vendor neutral specification that is ready to be implemented.

In addition, by relying on XML, it has been shown that the P3P specification guarantees enough flexibility in terms of portability and accessibility to be widely accepted in industries.

3 Existing P3P-enabled Tools

Different kind of tools that end-users can use to improve their control over their personal data have already been discussed (see subsection 1.5). This set of tools was rather aiming to limit the transfer of data between a client browser and a web server and, as explained, their efficiency is questionable.

This section will present some applications that implement the P3P specification. Should it not already be the case, these examples will definitely convince the reader of the appropriateness of the P3P approach to the solving of the web privacy problem . It will also further define how the privacy statements are ultimately communicated to end-users.

3.1 P3P-enabled web-browsers

Microsoft's Internet Explorer 6 and Netscape 7 both partially implement the P3P specification. These browsers check if the cookies that web sites are trying to set are related to a P3P policy.

The action undertaken by the browser in relation to these privacy statements (or their absence) depends on the browser's privacy settings. In the case where a cookie is not fully approved, Internet Explorer displays a small icon on the status bar to alert the user (see Figure 3). A cookie may for example be rejected, causing the web site not to work properly. Internet Explorer also offers the possibility to translate a web site's P3P policy into a human-readable format.

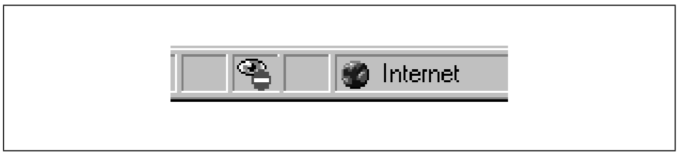


Figure 3: Internet Explorer 6 privacy alert

3.2 Browser plug-ins

Additional functionalities can be added to web browsers to improve their reaction to web sites privacy statements by means of a plug-in. An example of such a browser plug-in is the AT&T Privacy Bird[7]. In its principle, this tool provide functionalities that are very similar to the built-in features of Internet Explorer or Netscape. It however extends the checking of privacy statements to all the web pages, images, and third party contents that are downloaded by the client browser.

As it name suggests, this plug-in adds an icon representing a bird (see Figure 4) in the browser's title bar, and thereby making the end-user aware of the current threats he may be facing.

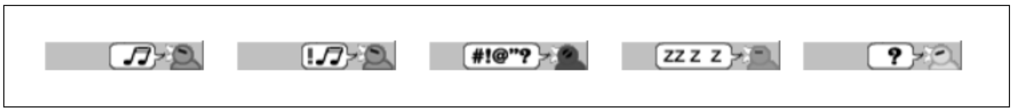


Figure 4: AT&T Privacy Bird

Efforts have been made by the Privacy Bird developers to make the tuning of the tool as easy as possible for a non-expert user. By clicking on the Privacy Bird icon, the user can also be presented a translation of the P3P policy of the web site in a human-readable format.

3.3 conclusion

Though the functionalities implemented by the tools presented in this section remain relatively basic, they have brought a first awareness of the privacy issues related to the Internet amongst parts of the Internet community. By blocking some cookies, they also made some web sites aware that they had to reconsider their privacy statement communication practices.

The simplicity of the privacy alerts that are given to the end-users having been illustrated, it should now be clear that an approach such as the one advocated by the P3P project designers can contribute to the solving of the web privacy problem.

[7] http://www.privacybird.com

4 Conclusion

This first part has introduced the web privacy problem and its complexity. Various approaches that are usually proposed to tackle it have been reported. It has been shown that a fair equilibrium could be reached between the interests of web site owners and their clients if these could be provided with an easy access to the web sites' privacy statement.

An emphasis has been made on the approach proposed by the W3C with the P3P specification. This approach has been assessed by analysing its origin, its technical implementation features, and by replacing it in the framework proposed by Swire (1997) to understand in what it could help in solving the web privacy problem.

An illustration of how the P3P specification can enhance the end-user awareness of web site data collection practices has been provided.

Part II

Project Motivation and Objectives

Having stated the web privacy problem, the extent to which P3P is intended to solve it, and some existing P3P-enabled tools, this second part will define the motivation for the development of a P3P editor tailored to universities' Web Teams needs. The objectives of the project will then be set.

5 Motivation for the project

Though most of the discussion has so far been illustrated by considering interactions between customers and online web sites, the web privacy problem is not restricted to commercial industries. As already sated in the introduction, the academic sector increasingly uses the Internet as a communication medium for an internal and external purpose.

In addition to the publication of general interest information, universities' web sites may allow students to apply online, to get lecture notes, to browse the library catalogue, to answer a survey, or to register for e-learning sessions. This is of course not an exhaustive list of services that universities may wish to offer. Online services could for example be developed for staff.

The data concerning actual and prospective students and staff that is collected during these interactions can potentially be used in various and sometimes unwanted ways.

As the general public is expected to be more aware and to show more concerns about their privacy in a near future, universities are starting publishing human-readable privacy statements on their web site. These publications are deemed to be a necessary condition to ensure a smooth development and the effectiveness of the abovementioned online activities and services.

The first part of this proposal has shown the shortcomings of such human-readable statements. It has also been demonstrated how machine-readable privacy statements can leverage the impact of these publications. However, as stated by J. Hargreaves (Appendix A), only a handful of UK universities may nowadays have a machine-readable policy, if any. Moreover, J. Hargreaves estimates that the privacy problem UK universities are facing cannot be dealt with on a global basis, as privacy policy may greatly vary from one university to another.

To ease the process of creation and publication of their privacy statements, UK universities need to be provided with a flexible tool that would enable their current Web Teams to develop machine-readable privacy policies that are specific to the academic sector. In order to overcome the technical challenge that the P3P enabling of a web site may constitute for a Web Team, this tool should require from its users as little knowledge as possible of the P3P specification and the related deployment technologies.

This fundamental requirement will have, as explained below in section 7, a huge impact on the system architecture, as it implies the tool to be implemented as a web-based application.

From the point of view of the developer, this project is also a great opportunity to get accustomed to the most advanced web technologies and their related programming languages. The development of such a web-based application will indeed require a good level of proficiency in server-side programming technologies, as well as the mastering of XML document processing techniques. Such capabilities are nowadays highly valuable as they are the fundamentals for the so-called *Web Services* development, that is the establishment of automatized communications between heterogeneous applications or information systems.

6 Project Goals

Having stated the motivation for this project, its aims can be enumerated. These are:

- To analyse the needs UK universities have in terms of privacy policy development and publication. This will be done as part of the requirements gathering for the tool. Achieving this analysis will lead to the identification of the requirements and expectations the target user group has for the tool.
- To determine to what extent the P3P specification may be used to implement these needs in machine-readable statement. By comparing the scope of the P3P standard to the target user group requirements, parts of the specification that will have to be implement in priority will be identified.
- To analyse existing P3P editors and expose their strengths and inadequacies regarding the UK universities' needs identified in the first two points above. This will allow informed decisions to be taken regarding the specification of the software in terms of provided functionalities and user interaction.
- Design, develop, and implement that software.
- Design and implement a simple XSL stylesheet that converts the P3P policies created by the application in a human-readable HTML document that can be posted on any web site.

Part III

Project Description

Having identified the goals of the system, this section will expose how these are expected to be achieved. A further description of the system will also be provided. It has to be noted that this section is not intended to provide a definitive specification of the system. It rather aim to point out analysis and design decision that already have been or should be taken concerning the development and implementation of the system.

7 Requirements Analysis

7.1 User requirements

Identifying the requirements of the users is never an easy task. Failure to identify the user requirements at an early stage of the project is one of the most common reason for project being wrecked. This subsection will describe how the user requirements will be inferred for this project.

7.1.1 UK universities privacy needs analysis

Identifying the needs UK universities have in terms of web privacy communication is the first and probably the most important task of the requirement analysis of the system. The needs UK universities may have in terms of privacy statement will be derived from looking at a sample of existing human-readable policies issued by some universities.

By comparing the scope of the P3P standard with UK universities privacy needs, it is hoped to parts of the P3P specification the editor should implement in priority will be identified.

Attempt will be made at an early stage of the process to build a prototyped interface which could be tested by University of Bath's Web Team, and possibly some other UK universities' Web Teams. This would allow to assess not only the usability of the system but also to get feedback from prospective users regarding the functionalities that will be implemented.

7.1.2 User's expertise

The tool is intended to be almost exclusively used by universities' Web Team or PR Office members. In order to ensure an as broad as possible acceptance of the software in this community, no hypothesis will be made on the level of proficiency in the field of web privacy and P3P related technologies of the target user group. The system will have to be usable almost instantaneously by a user without any prior specific knowledge of the technologies involved in the creation and the deployment of P3P policies and reference files.

By narrowing the target user group to the one given above, it is expected that opportunities to simplify and optimize the interaction between the user and the system will arise.

As explained below (see subsection 8.1), this requirement has direct implications in the architecture that will be chosen to implement the system in that a web-based application is the most appropriate system architecture when such a minimization of user technical knowledge is required.

7.1.3 Existing P3P editors analysis

An analysis of the strengths and weaknesses of existing P3P editors will be conducted from the perspective of the UK universities' needs identified above and from the perspective of the stated target user group. This analysis is expected to indicate which aspects of the tools need to be dealt with in priority to guarantee that the above stated user requirements will be met.

7.2 Functional requirements

The system will have to provide the user with the possibility to create both policy files and reference files. In addition, the user has to be offered to choice to create these files from scratch or update existing versions of these files by uploading them into the system.

7.2.1 P3P policy file creation

The system has to allow its user to create a P3P privacy statement for a university web site. It has to be made possible for the user to create such policies without any prior paper version of the statement. In other words, the user will be presented with a set of multiple choice question and as little as possible open-ended questions. The order in which the questions will be asked to the user will be predetermined, but will take into account the answers provided by the user at previous stages of the process. Such restrictions on the navigation in the system will drastically enhance the ease of use and the efficiency of the system.

7.2.2 P3P policy reference file creation

The same reasoning applies to the policy reference file creation. When creating new privacy policies, the user should be offered the possibility to create a reference file that will indicate to the site's visitors' browser the location of these policy files. The gathering of the required information to create such a file will be accomplished in the same fashion as explained above for the policy files.

7.2.3 P3P policies and reference file updating

The system should also allow the user to upload an existing P3P policy or reference file so as to enable him to modify existing files rather than having to recreate new files whenever only minor updates are to be achieved.

This feature has wider implications in that it makes it possible to develop templates that any user could upload and use as a basis for their own privacy or reference files. It has to be noted that the design of such templates is not part of the project in itself, as an in-depth analysis of the fully implemented system usage has to be carried out.

The design of such templates can easily be achieved later on, when data about the usage of the system will be available. The creation the templates can be achieved by the maintenance team, without the assistance of the system developer. The software in itself can be used to create them, as a template is nothing but a P3P policy or reference file.

7.3 Non-functional requirements

In addition to the functional requirements, a few non-functional requirements for the system have already been identified and will be reported in this section.

7.3.1 System deployment and installation

The software will be delivered and deployed directly on a UKOLN hosting machine. Attempt will be made to render the installation and configuration of the application as easy as possible, should the application be subsequently re-installed. The eventuality of a concise installation guide will be discussed with the project supervisor.

7.3.2 Security

No security issues have been identified so far concerning the access to the software. The system will not save the data provided by the users longer than it is required to fulfill the above-mentioned functional requirements. Any data collected during the execution of the system will be lost once the required files will have been created and delivered to the user.

7.3.3 System documentation

As stated above, an installation guide will be provided to enable the maintenance team to re-install the software, should it be needed.

On the other hand, a proper user guide is not required for this system. It is up to the developer to guarantee that the system will be intuitive enough and provide sufficient guidance to the user to ensure a correct and efficient usage. The fact that the system will be implemented over the Internet corroborates that point in that the user interface is embedded in an environment (a web browser) any user can reasonably be assumed to be familiar with.

The maximum that can be done in terms of instructions given to the user is a brief web page that would be displayed right before the execution of the tool.

No programmer guide will be provided as such. However, the dissertation will detail the entire process that will have led to the delivered system. An attempt will also be made to document this process in such a way that any programmer could apply it again for the development of other specific XML document editors.

8 System Description

This section will briefly describe from a high-level point of view the decisions that have already been taken regarding the system development and implementation

8.1 System architecture

As stated in the user requirements, the system should be usable by users having only a limited knowledge of the technologies related to P3P policies and their deployment. The amount of time required to create and deploy the generated files should also be minimized. Under these conditions, the most adequate architecture for the system is to implement it as a web-based application, which means that the application in itself will be deployed behind a web server, whilst the client will access and pilot it from any web interface.

According to Maruyama *et al.* (2002), the advantage of using a web-based architecture is twofold. It first implies that the user is provided with a web interface, which means that he is offered the same look and feel as he can experience on any other web interface. It also implies that the user is freed from having to manage the application (installation, configuration, updates, . . .), which is a major advantage compared to more traditional architectures.

8.2 System development and implementation

Having specified the architecture of the system, this subsection will identify the programming languages, the APIs, and the server-side configuration that are the most suited for the system.

8.2.1 Server-side technology

Server-side application development means that the pages that are delivered by a web server to a client browser are at least partially dynamically generated by a program, rather than being static pages retrieved from the server's file system.

Broadly speaking, there are two different ways to achieve such a dynamic generation of web content:

- **Server-side scripting.** Originally done by coupling Perl or Visual Basic scripts to a Common Gateway Interface (CGI), this techniques involves the insertion of fragments of code, called *scripts*, inside a content that otherwise would have been static. The scripts are executed by the web server when such content is requested by a client. The result is a dynamically created static content sent to the client. That kind of implementation is today rather achieved by using PHP or Java Server Pages (JSP), for the convenience they bring.
- **Server-side programming.** Proper server-side programming involves the development of an independent application that is run on the server and that is piloted by the user's browser. The whole system can therefore be written in one consistent programming language. It also allows a better separation of the presentation layer from the application layer, for the benefit of a better manageability of the application. Two technologies are competing in the field of server-side programming: Microsoft's Active Server Pages (ASP) and Java Servlets.

Choosing the right approach is seen by Van Der Linden (2002) as a trade-off that is mainly determined by the size of the application. Scripts can be quickly written and deployed and are therefore best suited for smaller application. The deployment of server-side programs is a much more daunting task, as it requires the configuration of a dedicated server that communicates with traditional web servers. But the advantages it brings in terms of manageability and maintainability of the application make it the best choice for larger applications.

The P3P editor requiring as main functionalities the creation and modification of XML files, its size can be easily estimated to be above the critical size where the advantages of proper server-side programming outpace their disadvantages.

Regarding the determination of which of the ASP or the Servlet technology should be chosen, the key argument is again given by Van Der Linden: "Servlets free you from from being tied to one kind of hardware or one software vendor". As the Java programming language has been designed to be portable, any servlet code can be run on any servlet server, making it easier to add new functionalities that require more processing capacity without having to modify the existing code.

8.2.2 XML files processing

Accessing XML files' content from an application can be achieved by means of two APIs: The Simple API for XML (SAX) and the Document Object Model (DOM). Each of these API has a radically different approach to the parsing of an XML file and therefore have their own advantages and disadvantages (McLaughlin, 2001).

DOM is a tree-based API in that it reads in an XML document and creates an in-memory hierarchical representation of the content of the document. The API then provides methods to access and modify this internal structure, and ultimately allows its serialization to an XML document. A DOM-based parser is thus best suited to situations where the content of an XML document needs to be randomly accessed or sorted, or when it first has to be ensured that the given XML document is well-formed and/or valid[8].

SAX is rather an event-driven API in that it scans an XML document and communicates with the application by sending it an event each time an XML element is found in the document. The performance of a SAX-based parser is thus enhanced compared to a DOM-based parser because it can directly start dispatching events without having to load the entire document in memory. A SAX-based parser also requires less memory.

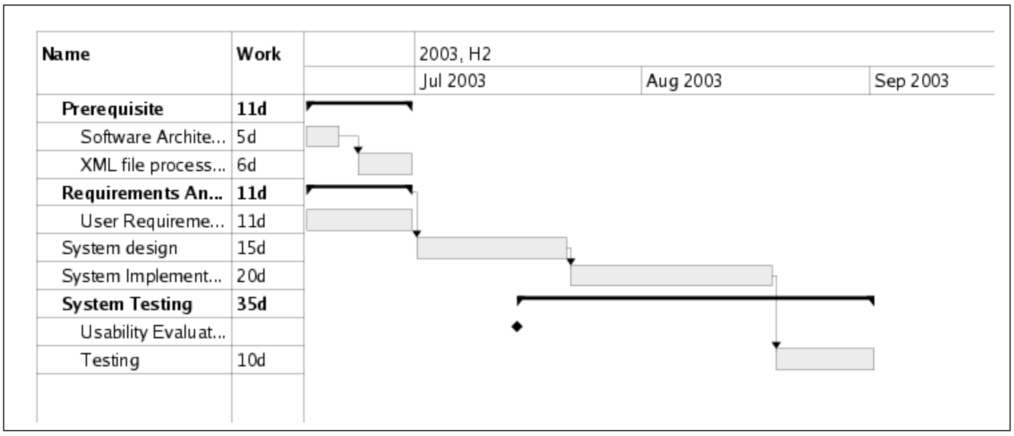
In addition, a DOM tree can be serialized to revert the data to an XML document, whilst SAX only provides methods for the parsing of a document. The building of P3P policies and reference files could have been achieved by means of scripts embedded in web pages. These scripts would have written in an output file the data entered by the user on the interface by means of the println() function. Maruyama *et al.* (2002) do not recommend to follow such an approach due to the complexity that the creation of well-formed and valid XML files involves.

The approach that will be followed to implement the system will consist in a servlet that is given all the required information from the user interface. The servlet will then build an internal DOM tree with this data, perform, if required, modifications to this tree, and serialize it in an XML file.

9 Project Evaluation and Timelines

In order to ensure that the delivered system meets its requirements, it will be tested at various stages of the development process. The proposed timelines for the realization of the project are depicted in the Gantt chart of the figure 5 and are further defined below. The project starting date has been fixed to the 16th of June 2003.

[8] An XML document is said to be *well-formed* if it conforms to the W3C XML specification and *valid* if its content is in accordance with its associated Schema



Name	Work
Prerequisite	**11d**
Software Archite...	5d
XML file process...	6d
Requirements An...	**11d**
User Requireme...	11d
System design	15d
System Implement...	20d
System Testing	**35d**
Usability Evaluat...	
Testing	10d

Figure 5: Project timelines

9.1 User Interface and user requirements

As explained above, it has to be ensured that the functionalities provided by the system meet users' expectations. Prototyped interfaces of the system will be used to simulate the environment in which the users will later use the fully implemented system. This evaluation will have to be achieved at an early stage of the process.

The interaction between the target user group and the software as well as the software usability will be assessed in collaboration with Mrs J. Hargreaves, member of the University of Bath Web Team, Public Relation Office. This evaluation can occur at any stage of the development process, considering that the development of the tool can be achieved using the prototyped interfaces.

This usability evaluation is essential in that a positive outcome will ensure that the tool could be used by various UK universities's Web Team members. An attempt will be made to try and test the tool with one other university Web Team to ensure that the tool is usable for Web Teams that have not participated in the development process.

9.2 Software architecture

The software architecture will be tested early in the development process to definitively set the versions of the Java Runtime Environment and the Servlet container the software will run on. The UKOLN machine that will host the application will then be configured accordingly.

9.3 XML files parsing and creation

The next test will aim to acquire some experience in the parsing and creation of XML files using the DOM API. This will lead to an informed choice of the appropriate version of this API that will be used by the tool.

9.4 System testing

The testing of the first components of the final software are expected to be carried out by early August 2003, considering that the final hand-in date for the project is the 24th September 2003. A first version of the tool with limited functionality will be tested to ensure that the integration of the decisions taken above will occur smoothly for the entire system.

The software in itself will be tested according to the usual testing requirements taught to the MSc Computer Science students at the University of Bath.

Appendix

A Meeting report: University of Bath's PR Office Web team.

Present

Jacki Hargreaves	University of Bath, Public Relation Office, Web Team.
Andy Powell	UKOLN, Assistant Director Distributed Systems and Services, Project Supervisor.

University of Bath's web privacy policy. The University of Bath has a human-readable web-privacy policy that can be found on the University web site [1]. Jacki Hargreaves has composed this document on the basis of other Universities' web privacy policies that were available, in consultation with BUCS, and by paying attention to requests received from users. The aim of this policy is to make the users more comfortable whilst browsing the web site and, if applicable, whilst filling in forms

Machine-readable Universities' web privacy policies. Though a significant number of Universities probably have a human-readable policy, Jacki Haregreaves estimates that only a handful may have a machine-readable policy, if any. Andy Powell suggests an analysis of the existing Universities' web privacy policies to determine if they could be implemented in P3P. This would also allow the identification of parts of the P3P specification that academic institutions would use the most. Jacki Haregreaves mentions that Universities privacy policies may vary greatly.

User information. The web privacy policy should at least inform the users about

- What information is taken
- For what purpose
- Further uses of this information
- If the displayed policy is up to date

Number of privacy policies. A nice feature of P3P is that it allows the association of policies to only a subset of the domain. Multiple policies can thus be defined for the same domain, as long as policies do not overlap. The opportunity to implement multiple privacy policy has to be analysed for the case of Universities' web sites. This would allow online forms' policies to be considered on a more individual basis, so that each form's privacy policy could accurately be defined.

[1] http://www.bath.ac.uk/web/privacy.html

List of Figures

References

CHO EUN SOOK, KIM SOO DONG, RHEW SUNG YUL, LEE SANG DUCK & KIM CHANG GAP. 1997 (December). Object-oriented Web application architectures and development strategies. *In: Proceeding of the 4th Asia-Pacific Software Engineering and International Computer Science Conference (APSEC '97 / ICSC '97).* http://dlib.computer.org/conferen/apsec/8271/pdf/82710322.pdf.

CRANOR L.F. 2002. *Web Privacy with P3P*. Sebastopol (US): O'Reilly & Associates.

CRANOR L.F. & WENNING R. 2002 (April). *Why P3P is a Good Privacy Tool for Consumers and Companies*. [WWW]. http://www.gigalaw.com/articles/2002/cranor-2002-04.html.

CULNAN M.J. & MILNE G.R. 2001 (December). *The Culnan-Milne Survey on Consumers & Online Privacy Notices*. [WWW]. http://www.ftc.gov/bcp/workshops/glb/supporting/culnan-milne.pdf.

CYBER DIALOGUE. 2001 (November). *Cyber Dialogue Survey Data Reveals Lost Revenue for Retailers Due to Widespread Consumer Privacy Concerns*. [WWW]. http://www.cyberdialogue.com/news/releases/2001/11-07-uco-retail.pdf.

ELECTRONIC PRIVACY INFORMATION CENTER. 2000 (June). *Pretty Poor Privacy: An Assessment of P3P and Internet Privacy*. [WWW]. http://www.epic.org/reports/prettypoorprivacy.html.

FORRESTER RESEARCH. 2001 (September). *Privacy Concerns Cost eCommerce $15 Billion*. Consumer Technographics Brief, [WWW]. http://www.forrester.com/ER/Research/DataSnapshot/Excerpt/0,1317,13484,00.html.

GARFINKEL S. 1997. *Web Security & Commerce*. Sebastopol (US): O'Reilly & Associates.

GAVRAY G. 2002 (September). *Personnalisation des sites web: Elaboration d'une methodologie de mise en oeuvre et application au cas DGTRE*. MEng thesis, Universite Catholique de Louvain (Belgium).

HERMAN I. 2003 (February). *About the World Wide Web Consortium (W3C)*. [WWW]. http://www.w3.org/Consortium.

JOHNSTON P. 2001 (September). *XML : a brief introduction*. [WWW]. http://www.ukoln.ac.uk/metadata/presentations/nhs-2001/xmlintro/sld001.htm.

KRISTOL D.M. 2001. HTTP Cookies: Standards, privacy, and politics. *ACM Transactions on Internet Technology*, **Vol. 1**(Issue 2), pp. 151–198.

MARUYAMA H., TAMURA R. & URAMOTO N. 2002. *XML and Java, Developing Web Applications*. 2nd ed. Boston (US): Pearson Education.

MCLAUGHLIN B. 2001. *Java & XML*. 2nd ed. Sebastopol (US): O'Reilly & Associates.

NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION. 1998 (January). *Elements of Effective Self-Regulation for Protection of Privacy*. U.S. Department of Commerce [WWW]. http://www.ntia.doc.gov/reports/privacydraft/198dftprin.htm.

PRIVACY LEADERSHIP INITIATIVE. 2001 (November). *Privacy Notices Research Final Results*. (Conducted by Harris Interactive, Inc) [WWW]. http://www.fct.gov/bcp/workshops/glb/supporting/harris

SWIRE P. P. 1997 (June). *Markets, Self-Regulation, and Government Enforcement in the Protection of Personal Information*. In: Privacy And Self-Regulation In The Information Age (U.S. Department Of Commerce, Tech.Rept.) [WWW]. http://www.ntia.doc.gov/reports/privacy/selfreg1.htm#1A.

VAN DER LINDEN P. 2002. *Just Java2*. 6th ed. Palo Alto (US): Sun Microsystems Press.

WORLD WIDE WEB CONSORTIUM. 2001 (July). *W3C Process Document*. [WWW]. http://www.w3.org/Consortium/Process-200100719/tr.html.

WORLD WIDE WEB CONSORTIUM. 2002 (April). *The Platform for Privacy Preferences 1.0 Specification*. [WWW]. http://www.w3.org/TR/P3P.