A "Nutrition Label" for Privacy

Patrick Gage Kelley,* Joanna Bresee,* Lorrie Faith Cranor,* Robert W. Reeder**

* Carnegie Mellon University
School of Computer Science
pkelley, jbresee, lorrie@cs.cmu.edu

** Microsoft
Trust User Experience (TUX)
roreeder@microsoft.com

ABSTRACT

We used an iterative design process to develop a privacy label that presents to consumers the ways organizations collect, use, and share personal information. Many surveys have shown that consumers are concerned about online privacy, yet current mechanisms to present website privacy policies have not been successful. This research addresses the present gap in the communication and understanding of privacy policies, by creating an information design that improves the visual presentation and comprehensibility of privacy policies. Drawing from nutrition, warning, and energy labeling, as well as from the effort towards creating a standardized banking privacy notification, we present our process for constructing and refining a label tuned to privacy. This paper describes our design methodology; findings from two focus groups; and accuracy, timing, and likeability results from a laboratory study with 24 participants. Our study results demonstrate that compared to existing natural language privacy policies, the proposed privacy label allows participants to find information more quickly and accurately, and provides a more enjoyable information seeking experience.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces; K.4.1 [Computers and Society]: Public Policy Issues–Privacy

General Terms

Design, Experimentation, Human Factors, Standardization

Keywords

privacy, P3P, policy, user interface, information design, labeling, nutrition label.

1. INTRODUCTION

Website privacy policies are intended to *assist* consumers. By notifying them of what information will be collected, how it will be used, and with whom it will be shared, consumers are, in theory, able to make informed decisions. These policies are also meant to inform consumers of the choices they have in managing their information: whether use of their information or sharing with third parties can be limited, and if it is possible to request modification or removal of their information.

However, Internet privacy is largely unregulated in the United States (except for children's privacy and some sector-specific regulations) and the privacy policies created by companies are frequently difficult for consumers to understand. Online privacy policies are confusing due to the use of specific terms that many people do not understand, descriptions of activities that people have difficulty relating to their own use of websites, a readability level that is congruent with a college education, and a non-committal attitude towards specifics [14]. These issues are complicated by companies creating policies that are tested by their lawyers, not their customers. It has further been established through numerous studies that people do not read privacy policies [21] and make mistaken assumptions based upon seeing that a site has a link to a privacy policy [26]. A recent study estimated that if consumers were somehow convinced to read the policies of all the companies they interact with, it would cost an estimated 365 billion dollars per year in lost productivity [20].

In addition, research has shown that consumers do not actually *believe* they have choices when it comes to their privacy. Based solely on expectations, they believe there are no options for limiting or controlling companies' use of their personal information [16]. This is a finding that we again validated in our work.

In short, today's online privacy policies are failing consumers because finding information in them is difficult, consumers do not understand that there are differences between privacy policies, and policies take too long to read. We set out to design a clear, uniform, single-page summary of a company's privacy policy that would help to remedy each of these three concerns.

This paper first presents related work describing standardization efforts in other domains in which companies present information to consumers to aid in their decision making, as well as early standardization efforts for privacy policies. Our approach comes from a broad survey of work that provides consumers with information: nutrition labeling, drug facts, energy information, and most recently work commissioned by the Federal Trade Commission to create a standard financial privacy notice. We discuss our iterative design approach, including focus group testing, as we developed and refined our information design over several months. Finally, we describe our 24-participant laboratory study and discuss the results of our initial evaluation.

2. RELATED WORK

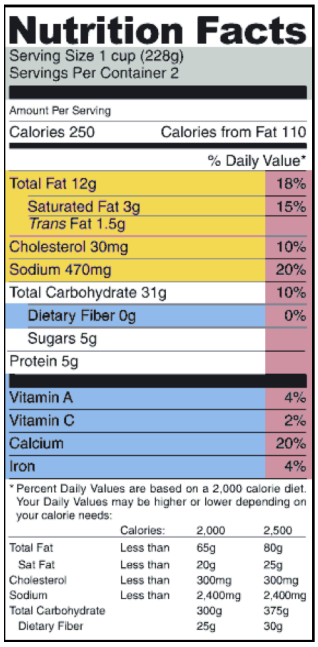
To better inform our design process we surveyed the literature surrounding other consumer labeling efforts: the "Nutrition Facts" panel, energy and drug labeling, and recent work on creating a standardized financial privacy notice. Additionally, we summarize our previous work on a standardized privacy policy format.

2.1 The "Nutrition Facts" Panel

In the United States, the nutrition label seen in Figure 1, has become iconic after being mandated by the Nutrition Labeling and Education Act of 1990 (NLEA) [28]. In the last nineteen years, its increasing ubiquity has led to a number of studies examining the

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee.

Symposium On Usable Privacy and Security (SOUPS) 2009, July 15-17, 2009, Mountain View, CA, USA.



Nutrition Facts

Serving Size 1 cup (228g)
Servings Per Container 2

Amount Per Serving

Calories 250	Calories from Fat 110
	% Daily Value*
Total Fat 12g	18%
Saturated Fat 3g	15%
Trans Fat 1.5g	
Cholesterol 30mg	10%
Sodium 470mg	20%
Total Carbohydrate 31g	10%
Dietary Fiber 0g	0%
Sugars 5g	
Protein 5g	
Vitamin A	4%
Vitamin C	2%
Calcium	20%
Iron	4%

*Percent Daily Values are based on a 2,000 calorie diet. Your Daily Values may be higher or lower depending on your calorie needs:

	Calories:	2,000	2,500
Total Fat	Less than	65g	80g
Sat Fat	Less than	20g	25g
Cholesterol	Less than	300mg	300mg
Sodium	Less than	2,400mg	2,400mg
Total Carbohydrate		300g	375g
Dietary Fiber		25g	30g

Figure 1. The Food and Drug Administration's Nutrition Facts panel as regulated by the NLEA. Source: www.fda.gov

costs of adoption and the ability to inform and change consumer purchasing decisions.

The sparse literature around the design of the nutrition label [3] focuses on the decisions made to simplify the information as much as possible for consumers. These decisions were made in part to address low literacy rates and the needs of older Americans. These guidelines include defining a zone of authority, providing quantitative information about nutrients, defining minimum font sizes, and equalizing labels across products by providing defined serving sizes and calculating percentages based on standardized daily amounts.

Surveys indicate that consumers would prefer that nutrition labels include more information. However, studies have shown that including more information would not actually be beneficial [10]. Studies conducted to examine the impact of the NLEA have found that it is the populations of people who are educated and already motivated to investigate nutritional information who benefit the most from nutrition labels [2][10]. Another study found that nutrition information had the greatest impact when there was a limited number of items from which to make a selection [24]. This result implies that the nutrition label made it easier to compare between a small set of items, allowing consumers to benefit, through informed decision making. Studies have demonstrated that nutrition labels have an impact on consumer decision making, with some user-reported effect sizes up to 48% after the initiation of NLEA [10]. For most studies, however, the effect of the nutrition label is small and most studies focus on specific nutrients such as fat intake or specific products such as salad dressings. We are not aware of controlled studies that measure the impact of nutrition labels on consumer behavior over an extended period of time.

Other studies have found that the effects of providing calorie information (not a complete nutrition facts label) in restaurant menus are often very small and the effects may vary depending on the population studied. In a study of meal choices at a sandwich shop, Downs et al. found that if participants were given menus that included calorie information, they ordered meals with about 50 fewer total calories than participants who did not receive calorie information. However, the authors stated that this was "an effect smaller than this study was powered to test." Nonetheless, they pointed out that if the finding proved reliable, it could be significant if it caused people to reduce their caloric intake by a similar amount for multiple meals each day. In a related study of food purchases at three New York City restaurants before and after a law went into effect mandating the posting of calorie information on menu boards, the authors found no effects of the legislation at two of the three restaurants. At the third restaurant they found a small effect. They noted that the effect was larger for dieters than for non-dieters, suggesting that the availability of label information may again be most useful to people who are already interested in the information provided by the label [9].

2.2 Other Privacy Notices

Layered Privacy Policies, a policy display format popularized by the law firm Hunton & Williams [25], involve a short form or summarized version of a privacy policy created using a step by step process. This summary has standardized headings for the policy information, but the information itself is provided by the company, in free-form natural language text.

The US Federal Trade Commission (FTC) is currently leading an effort to develop a standardized financial privacy notice. The Kleimann Group used an iterative design process to develop a prototype notice for the FTC, focusing on user comprehension, allowing users to "identify differences in sharing practices," and compliance with the regulations surrounding financial privacy notices specified in the Gramm-Leach-Bliely Act. Over a 12-month period the Kleimann Group iterated on several design prototypes, conducting focus groups and diagnostic usability testing [16]. Our iterative design approach followed a similar process of testing labels for comprehension and then overall design through focus groups.

The Kleimann Group final prototype consists of four parts: the title, the frame, the disclosure table, and the opt-out form. The disclosure table, which actually displays the company's privacy practices, makes up the majority of our label. The rest of the Kleimann Group prototype was educational information to build a foundation of terms and understanding for the user [16].

More recently, the Levy-Hastak report was released, detailing the results of a 1032-participant mail/interview study [17]. The authors conclude that the table format performed the best "on a diverse set of … measures." Additionally, this success is attributed to the table providing a more holistic context for the particular sharing of each financial institution.

2.3 Other Labeling Programs

We also explored energy labeling programs from the European Union [12] and Australia [11], the US Consumer Products Safety Commission's toy and game warnings [8], and the US FDA Drug Facts label [29], to gain a broader understanding of practices used in designing and defining labeling requirements.

In general, the standards documents [7][12][28] are occupied with defining precise guidelines to describe compliance with the various labeling requirements. This includes point sizes of rules and text, allowable typefaces, allowable colors, and minimum sizes. In some instances, such as choking warnings on children's games, standards also include placement requirements.

Recently, a number of labels have been introduced to provide ratings to consumers on a fixed scale, focusing on a single metric or small number of metrics. The Australian Water Efficiency Labeling System (WELS) [32] and the British Food Standards Agency's Signposting (or Traffic Light) [13] use very small indicators with accompanying ratings. The WELS program uses an indicator with a possible score out of six blue stars. The Signposting initiative rates the quantities of fat, saturates, sugar, and salt in foods using a red, amber, green traffic light coloring system. Early research [2][18] has shown that Signposting enhances consumers' ability to evaluate products more accurately and surveys show that ninety percent of consumers find this type of label useful.

2.4 The Platform for Privacy Preferences

Due to the difficulties surrounding the use of text privacy policies, the World Wide Web Consortium created the Platform for Privacy Preferences (P3P) [30]. P3P is a standard machine-readable format for encoding the online privacy policy of a company or organization. Once this P3P policy has been provided, consumers must use a user agent to interpret it into something understandable. Unfortunately, widely available P3P user agents have limited functionality. These include the P3P policy processing elements of common web browsers and a few privacy specific browser add-ons [6].

To provide consumers with an active tool where they can investigate and explore the privacy policy of a website, earlier work from the CyLab Usable Privacy and Security Lab (CUPS) produced the P3P Expandable Grid. This user agent was based on one of the central Expandable Grid objectives of displaying a holistic policy view [22]. The interface was created to use the entire P3P specification, broken down by categories. An example of the grid is shown in Figure 2.

The P3P Expandable Grid has two main parts: the header and the information display. In the header, there is a title, a legend that explains the 10 possible symbols (8 pictured) that may appear in the body of the grid, as well as expandable column headers that explain how that company uses data, and who they will share it with. Finally, in the top-right corner of the header is a button that toggles between showing and hiding information that isn't collected (i.e., hide rows that would be blank).

In the body, information is displayed in blocks that correspond to P3P Statements. Each block starts with a title and a short textual description (if available) and is followed by a hierarchy of expandable rows, which list what information this company collects. The symbols in each row show how that specific piece of information could be used or shared according to the policy. In this way we were able to show the entire depth of the P3P specification in a two-dimensional grid.

Based on an online survey of over 800 people in the summer of 2007, we found further evidence that people generally do not understand the information presented in privacy policies and also do not enjoy reading them. When comparing three formats: a standard natural language policy; PrivacyFinder, which is a simplified human-readable version based on a P3P policy and consisting mostly of bulleted lists; and the above version of the

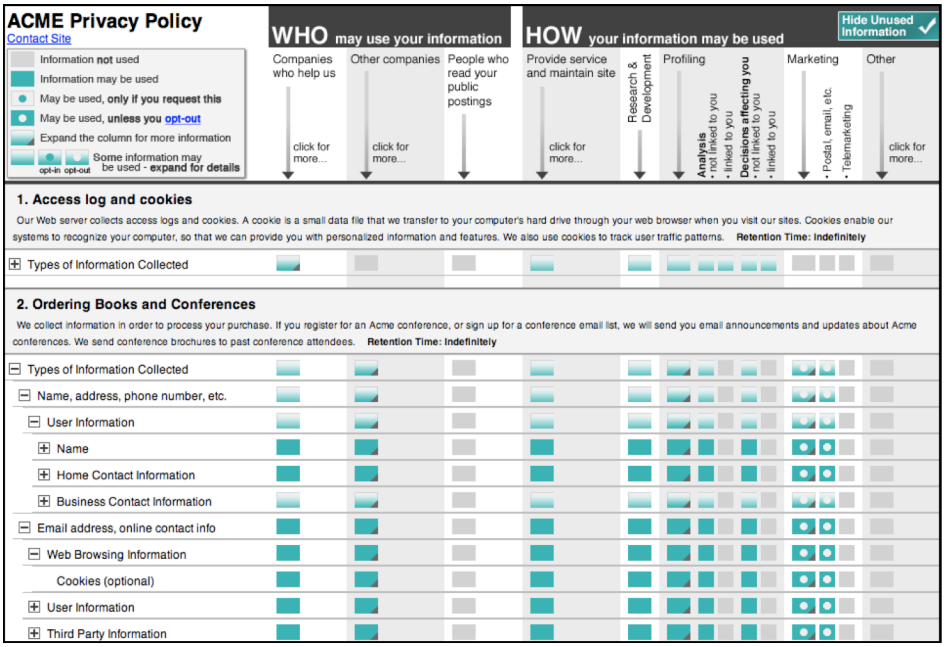


Figure 2. Our *P3P Expandable Grid,* an early attempt at a standardized information design for privacy policies. Due to its implementation of the entire P3P specification its complexity prevented large performance gains.

P3P Expandable Grid, we found that none of the three formats were found to be pleasurable to read or easy to comprehend. Notably, we found the P3P Expandable Grid to be slightly worse than the other formats, both in enjoyment and comprehension [23].

3. DESIGN METHODOLOGY

This section elaborates on our iterative design process, presenting several prototype labels with benefits and criticisms, and highlighting where knowledge from other label designs was applied. Throughout this process we leveraged informal user feedback as well as focus groups, which are discussed in detail in Section 4.

3.1 Problems with the P3P Expandable Grid

Based on the analysis of the previously mentioned P3P Expandable Grid study results and a subsequent lab evaluation, we identified five major problems with the Expandable Grid [15]:

- Many of the P3P labels are not clear to users. For example, "Profiling" and "Miscellaneous Data" are not terms that users encounter in the context of their use of websites.
- The legend has a large number of symbols including multiple symbols for expansion (depending on directionality), which the user may not understand.
- Multiple statements that may be related to the same types of information in a P3P policy are displayed separately, possibly requiring the user to check multiple rows to answer a single question.
- The Hide Used Information button in the top right only condenses unused rows, not columns.
- Rows with a plus symbol may be expanded; however, many users (40.7%) never expanded any data types. By not expanding data types, users never saw some important parts of the policy [23].

With these initial five problems in mind we abstracted several general principles from the nutrition labeling literature [3][4][27][28].

- Putting a box around the label identifies the boundaries of the information, and, importantly, defines the areas that are "regulated" or should be trusted. This is a common issue when the label is placed in close proximity to other information, but may not be as significant an issue online.
- Using bold rules to separate sets of information gives the reader an easy roadmap through the label and clearly designates sections that can be grouped by similarity.
- Providing a clear and boldfaced title, e.g., Privacy Facts, communicates the content and purpose of the label specifically and assists in recognition.

While much of the labeling literature also focuses on quantifiable properties, such as amounts of fats or fiber or percentages of active ingredients or calories from a standardized expected daily value, privacy policies typically do not include quantifiable measures, and the P3P specification includes no quantifiable fields. The Kleimann Group dealt with this lack of quantifiable information by moving to binary Yes/No statements, which they found to be readily understood by focus group participants.

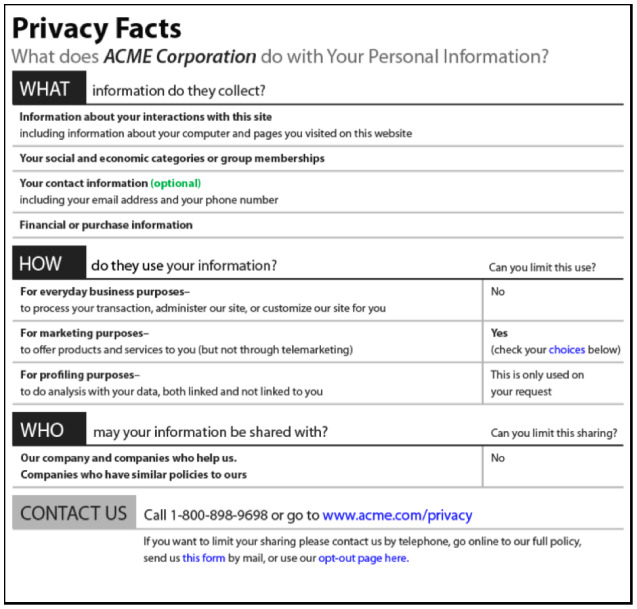


Figure 3. Our *Simplified Label*, an early attempt at a privacy label.

3.2 The Simplified Label

Our next design, following the P3P Expandable Grid, was the *Simplified Label.* In creating the Simplified Label, we used Yes/No statements and applied the three general principles discussed above. The Simplified Label is shown in Figure 3. (Note: as with each of the screenshots shown below, this is one of many variants of a similar vein. We show only one of each that we believe is representative of the entire series.)

While we made visual changes including adding a title and sub-head, adding bold lines, and simplifying the table view, the most significant change is a reduction in complexity. Two changes contributed most to simplifying the label: eliminating P3P statement groupings and eliminating the use of P3P data hierarchies. These changes are detailed below.

3.2.1 P3P Statements

P3P specifies data groupings called STATEMENT elements [31]:

> The STATEMENT element is a container that groups together a PURPOSE element, a RECIPIENT element, a RETENTION element, a DATA-GROUP element, and optionally a CONSEQUENCE element and one or more extensions. All of the data referenced by the DATA-GROUP is handled according to the disclosures made in the other elements contained by the statement.

This means that all of the collected information in a statement can be used for certain purposes, and can be shared in the same way. A useful model is to think of P3P as consisting of multiple triplets of information, {data, purpose, recipient}. We do not include retention because our analysis of over 5000 unique P3P policies collected by the Privacy Finder search engine [6] shows that the majority of P3P policies state that data is retained indefinitely. In cases where a website has a different data retention policy we include a note at the bottom of the label.

Due to P3P information naturally falling into these triplets, a display such as the list in Figure 3 suffers some information loss. For example, it is possible contact information is used for

marketing exclusively and purchase information is used for profiling purposes exclusively. Or it is possible that both contact and purchase information could be used for either purpose. By removing the triplets and only displaying a list, we lose that distinction. This tends to make privacy policies appear more permissive than they actually are.

A P3P policy may also have multiple statements. In the P3P Expandable Grid, statements were displayed in a numbered list. In the Simplified Label we have merged multiple statements into a single list. For example, consider a policy where the first statement of a policy was about cookies and the second dealt with web activity. In the P3P Expandable Grid we would list the categories twice. The first time only cookies would be highlighted; the second, web activity. With the Simplified Label we show the information from all of the statements in a single list.

3.2.2 P3P Data Hierarchies

P3P allows for two interchangeable and different hierarchies of data (collectable information). The more commonly used is categories: a list of 17 types of information that companies can collect. When a category is specified a company reserves the right to collect any information that falls under that category (i.e. "Physical Contact Information" includes name and telephone number). The other data hierarchy, the base data schema, includes every data element that can be specified using P3P, hierarchically arranged (e.g., NAME is a child of USER and includes GIVEN[name], MIDDLE[name], and FAMILY[name]). Further complicating the situation, every element belongs to one or more category (NAME is a member of both demographic data and physical contact information because one's GIVEN name is part of their contact information while one's FAMILY name provides demographic information).

In the original P3P Expandable Grid, each category was displayed in its entirety in each statement, with each element of the base data schema hierarchically arranged as children. This led to nearly 800 elements per category (if fully expanded). To simplify, we decided to display only data categories. While this affords us a list of possible information that can fit on a page, it suffers when companies state they will only collect specific items. For example Contact Information would be displayed similarly if a company collected a consumer's name, their postal address, their telephone number, or all of the above information. One way of preserving some of this detail would be to display the specific data elements a company collects when a user clicks on the name of a category.

3.2.3 Design Notes

To further reduce complexity, information that is not collected or purposes that are not mentioned in a particular policy are not shown. The Show/Hide information button has also been removed; thus, there is no way to see uncollected information.

Finally, we have defined a maximum width of 760px for this label and all following designs in this paper. One important consideration was that the privacy label design be printable to a single page and viewable in the standard width of today's internet browsers.

3.3 The Simplified Grid

While the above label is extremely simple and closely follows a pattern established by the nutrition facts panel and the financial privacy notice, we felt that it sacrificed too much detail.

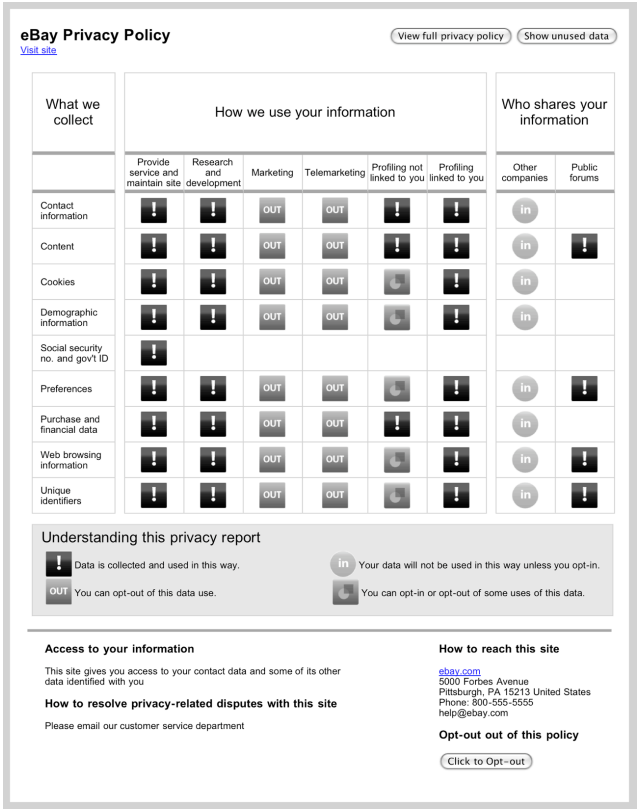


Figure 4. Our *Simplified Grid* in which the grid concept is reintroduced to the label.

The goal of our next design was to bring back more of the detailed information that privacy policies can provide without overwhelming users. To do this we decided to try to find a happy medium between our Simplified Label and the best aspects of the original P3P Expandable Grid. We adopted a two-dimensional grid layout, as shown in Figure 4. We call the resulting design the *Simplified Grid*.

3.3.1 Simplifying the P3P Expandable Grid

While the P3P Expandable Grid was not successful, this failure was not a result of the tabular display. Also, as discussed above, due to the nature of P3P Statements each reduction in dimensionality causes a loss of information and we wanted to minimize information loss to most benefit consumers. With the reintroduction of the two dimensional layout several changes were made. As mentioned in 3.2.2 we only used Data Categories to show what information companies collect, but we also simplified recipients and purposes .

Purposes, of which there are 12 specified[1] in the P3P specification, were grouped similarly to the categories in the P3P Expandable Grid. However the sub-categories were removed. Thus, Administration, Current Transaction, and Tailoring are all

[1] The P3P specification specifies 12 purpose elements: Current, Admin, Develop, Tailoring, Pseudo-analysis, Pseudo-decision, Individual-analysis, Individual-decision, Contact, Historical, Telemarketing, and Other-Purpose [31].

A bold title is used to set the context for the information.

Short labels are used for column and row headers, with longer definitions on our Useful Terms page.

Information that is not collected has a saturated label and a row full of the lightest symbol.

Four symbols on a scale from light to dark are used to indicate the severity of certain privacy practices.

Row and column locations are consistent so that two policies side-by-side can be easily visually compared.

A legend provides information about what each symbol means.

Figure 5. Our *proposed Privacy Nutrition Label.* This label is the one we tested in the second focus group and the laboratory study.

grouped under the title "Provide service and maintain site." We split the four P3P profiling-related purposes into two categories, based on whether that profiling is linked to the users' identity or performed anonymously. However, during our user testing, this distinction proved unclear to users.

Of the 6 recipients specified by P3P[2], Ours and Delivery are both never shown, as it is implied that the given company will always maintain the information. "Other Companies" merges the three remaining types of recipients, distinguished by their own privacy practices. We decided the importance of this column was to show whether any sharing with other companies was taking place. Public forums remained unchanged.

3.3.2 Symbols & Mixed Control

While you cannot opt-in or out to the trans-fat in your salad dressing, you might be able to have control over certain aspects of your information sharing on the internet. The Yes/No dichotomy advocated by participants in the Kleimann Group's studies works when there are only one, or maybe two, columns of information. Here we would have needed 8 columns and 10 rows of Yes/No information, which would have been visually difficult to parse.

[2] The P3P specification specifies 6 recipient elements: Ours, Same, Other-recipient, Delivery, Public, Unrelated [31].

Instead we again looked back to the P3P Expandable Grid and used symbols. However, while the P3P Expandable Grid had an array of 10 symbols, the Simplified Grid uses only four:

- **Exclamation Point**: Data is collected and used in this way.
- **OUT** (in a square): You can opt-out of this data use.
- **IN** (in a circle): Your data will not be used in this way unless you opt-in.
- **Square and circle**: You can opt-in or opt-out of some uses of this data.

Each of these four symbols was defined in a legend labeled "Understanding this privacy report" directly below the policy. The legend is another device borrowed from the P3P Expandable Grid; however, it has been moved below the policy.

Again, due to the way P3P uses data statements, it is possible that in some instances consumers might be able to opt-out of allowing their demographic information to be used for profiling, but in others it is required, or opt-in. The "square and circle" or "mixed choices" symbol attempts to convey this possibility; however, in our user testing it was found to be incomprehensible.

3.3.3 Visual Intensity

The Simplified Grid is the first iteration of our label to use visual intensity to provide a high-level indication of the quality of a given policy. Each of the four symbols has been colored such that darker symbols represent what could be more privacy-invasive practices. The use of intensity allows users to make quick visual comparisons that would not have been possible with text alone.

3.3.4 Testing

The most significant issue that arose in our testing was confusion over blank areas of the label. We thought that blank areas would clearly indicate information a company does not collect; after all, natural language policies typically leave out any mention of types of information the company does not collect. However, in testing, many participants were unclear on the meaning of the blank cells. Some inferred the accurate meaning that such information uses would not occur, but others thought it allowed the company free reign to do anything in those situations or that they simply had not yet decided their practices.

3.4 Final Proposed Privacy Nutrition Label

Our *Privacy Nutrition Label,* shown in Figure 5, is a direct descendent of the Simplified Grid. With the Privacy Nutrition Label, we sought to refine the strengths of the Simplified Grid by reducing clutter, introducing color, and simplifying symbols.

3.4.1 Types of Information Displayed

We made changes in the way we present data categories as rows in the table to better facilitate comparisons between policies and to reduce confusion about what data is being collected.

All of the P3P Data Categories are now represented in rows regardless of whether they are collected or not. For example, the label shown in Figure 5 indicates health & financial information are not collected (and thus not used or shared), but they have not been removed. Any policy displayed in this format will have exactly 10 rows, and the ordering will always be consistent. This allows two policies to be easily, visually compared side-by-side.

Participants in a focus group we conducted after making this change did not understand which information companies were not collecting. We indicated the information that was not collected with rows completely filled with minus symbols, but participants believed that companies collected every piece of information listed on the grid. One participant asked, "Why would they collect all that information if they're not going to do anything with it?" In the final prototype we grayed out the labels for data that companies did not collect, and we changed the minus symbol's description from "we will not use your information in this way" to "we will not collect or we will not use your information in this way." We also changed the row-heading label from "What we Collect" to "types of information." This change was made to highlight the fact that we now show even un-collected information and to reduce confusion about what was and was not being collected.

3.4.2 Symbol Changes & Color

In the Simplified Grid design, we marked types of information that companies collected and left other cells in the policy blank. However, half of the participants were afraid of the blank spaces; for instance, one said, "Nothing is mentioned. It is completely open-ended. These guys [the company] can modify these values." Therefore, in the final version we introduced a symbol to indicate that information was not collected or used.

Focus group participants found the mixed choices symbol confusing so we removed it. Instead we now display the symbol for the most invasive practice. For example, if in some circumstance one can opt-in and in another one can opt-out, we display the opt-out symbol.

We constrained our initial designs to grayscale to facilitate easy printing without loss of information and to reserve color for highlighting differences between a policy and a user's personal preferences (something we plan to implement later). However, feedback indicates that color seems to improve user enjoyment in reading the label, although we have not yet quantified this improvement. We selected the colors used in our label with care to accommodate viewers with color-blindness, allow for grayscale reproduction, and maintain the darker-is-worse high-level visual feedback discussed in Section 3.3.3.

3.5 Useful Terms

Even with the "understanding this privacy policy" legend in place there was still confusion over many of the terms used in the label. This was also a common issue during the development of the Kleimann Group's Financial Privacy Notice, and in response they developed what they call the "Secondary Frame." This portion of the prototype notice included both frequently asked questions and a series of extended definitions, which are: "[not] information as essential for consumers to have, but consumers often commented that they liked having it included." [16 p.27]

Our version of the "Secondary Frame" is a single page hand-out of useful terms. Our useful terms information was informed by the Human Readable definitions included in the P3P 1.1 Working Group Note [31] and consists of seventeen definitions, one for each of the row and column headers. Some are straightforward, others more detailed. For example, the definition of telemarketing states: "Contacting you by telephone to market services or products," while the profiling definition is:

> Collecting information about you in order to:
>
> - Do research and analysis
> - Make decisions that directly affect you, such as to display ads based on your activity on the site.

Information that the site collects about you may be linked to an anonymous ID code, or may be linked to your identity.

In future versions, clicking on or hovering over the headers could pop-up these definitions.

4. FOCUS GROUPS

We held two, hour-long focus group sessions to review the design and discuss participants' impressions and questions. We recruited focus group participants from the Carnegie Mellon University (CMU) Center for Behavioral Decision Research (CBDR) participant recruitment website. We paid participants $10 to participate in a 60 minute focus group.

The first focus group was composed of three female and seven male CMU students. The participants reacted positively to the Simplified Grid. For example, one participant stated, "This is more convenient than scrolling through reams and reams of paragraphs. I mean who reads them?" and another participant said, "I like the chart. [It's] better than long sentences." However, we found that some participants still had problems understanding privacy concepts. For example, one participant asked, "What is the difference between opt-in or opt-out?" and many others agreed that they did not understand this distinction. Additionally, many participants had trouble distinguishing different privacy concepts. Most participants were familiar with profiling, but did not understand the difference between "Profiling linked to you" and "Profiling not linked to you." Similarly, participants did not understand the different meanings of "cookies" and "unique identifiers." It was this vein of feedback that led to the inclusion of the useful terms definitions described in Section 3.5.

By asking participants to compare two policies, we found that participants could easily isolate and describe differences. Participants noticed that Policy A had more opt-in symbols and Policy B had more opt-out symbols. However, participants were not able to make accurate judgments about the policies. When we asked the participants to chose the company with whom they would prefer to do business, five of the ten participants chose Policy B: the company that collected and used more of their personal information.

Using the feedback from the first focus group, we initiated another series of rapid iteration and prototyping, which resulted in the final label prototype. Our second focus group compared the final Privacy Nutrition Label to the Simplified Label.

The second focus group was composed of four female and three male undergraduate students from CMU and the University of Pittsburgh. When reviewing the Privacy Nutrition Label vs. the Simplified Label we found that participants better understood the grid and were able to make more accurate side-by-side comparisons. Participants understood the significance of the red symbols, saying, "Red is for 'stop' or 'danger.'" We passed out two privacy policies, Policy A and Policy B, and asked the participants to raise their hands if they believed that Policy A is the better policy. Every participant raised his or her hand, correctly identifying Policy A as the more favorable policy. Participants demonstrated a detailed understanding of the differences between the policies with comments such as "It's very clear which site is best" and "You should pick a site with more opt-ins than opt-outs." Some participants even noted subtle differences between the two policies saying, "Policy A isn't perfect either, because they share your preferences, and this may include things like your religious or political preferences."

After reviewing the grid design, we passed out the simple text policy. Participants reacted negatively to the text policy because they felt that it did not provide enough information, saying, "This is an empty policy, it says nothing. I wouldn't trust it." Participants wanted to see how each piece of information was being used. For example, one participant stated, "With the grid it's easier to see things. What information is being shared? We don't know that anymore."

5. USER STUDY METHODOLOGY

Based on the feedback from our second focus group we performed a 24-participant laboratory user study comparing a standard natural language (NL) privacy policy with privacy policies presented in our Privacy Nutrition Label.

We used a within-subjects design where participants were randomly assigned to first use either the label or the natural language format. Each participant completed 24 questions relating to the policy format they were shown first and then the same 24 questions again with the other format. These tasks are detailed below. We recorded accuracy as well as time for each participant.

5.1 Participants

We recruited the 24 participants through the CBDR website. Our only requirement was that English be the participant's native language. We offered participants $10 to participate in a 45 minute study in our laboratory.

Our participants included 16 students and 8 non-students. Of the 16 students, 5 studied humanities, 5 economics or business, 2 science, and 4 information science. 16 of our participants were male, 8 were female.

5.2 Privacy Policy Selection

Our study used two NL privacy policies and two label formatted policies. We started with the current actual P3P policy of a popular online e-commerce website. We modified this policy in three ways to produce two different label policies for the mythical companies Acme and Button. The first change was to the data collected. Acme has preference information collected but not demographic information, whereas Button Co., collects demographic, not preference. This change is not incredibly significant but does distinguish the data collection. The second change was to the data uses. Acme does not do any profiling while Button Co. does. The third change was to information sharing practices. While Acme only shares information when consumers opt-in, Button Co. shares information unless consumers opt-out. These significant differences were introduced so that there would be a clear "correct" response for participant tasks that require them to determine which company better protects their privacy (see 5.3.3).

The two NL policies for the mythical companies ABC Group and Bell General represent the exact same policies as described above. The ABC Group policy is the natural language policy of the same company whose P3P policy was used to populate the grid, again with the three modifications above made to make it match Acme's. We could not however simply make the three modifications to the policy and also present it as the other natural language option because two different companies, no matter how

Table 1. Extended Text & Readability Comparison for NL

Policy Metric	ABC	Bell
Word Count	2287	2299
Sentence Count	136	130
Flesch Reading Ease	42.06	41.69
Flesch-Kincaid Grade	11.57	11.84

similar their practices, would not share the same text. The introduction, structure, and actual language used needed to be different. Thus, to create the Bell General policy we used the text of a different, yet comparable e-commerce website, and changed the practices so as to match that of Button Co.

In editing the natural language policies we removed any references to programs that would distinguish the companies (such as specially branded programs), removed lists of links from the beginning of the policies, removed references to Safe Harbor, and additionally modified the second policy so that both were approximately the same length. For a more complete comparison see Table 1.

We chose not to use layered policies. This decision was made because layered policy adoption is not consistent or widespread, most common layered policies would not be suitable for answering the questions we asked, and finally recent research has suggested layered policies are no better at helping consumers understand privacy than full natural language policies [19].

5.3 Task Structure

The task structure for each condition was exactly the same, with 24 tasks comprising a section. These sections can be split into four parts, each of which is detailed here:

5.3.1 Information Finding

The first 8 questions were all Yes/No questions asked of a single policy (ABC Group for NL, Acme for the label). Of these 8 questions, 6 were single-element questions, involving only one element of the P3P statement triplet. For example: "Does the policy allow the Acme website to use cookies?" to which the answer was Yes, or "Does the policy allow the Acme website to share your information on public bulletin boards?" to which the answer was also Yes.

The remaining two questions all required two parts of the triplet to answer the question, for example "By default, does the policy allow the Acme website to collect your email address and use it for marketing?"

5.3.2 Perceived Privacy Policy Understanding

Following the 8 information finding questions, participants were given 6 questions on a 5-point Likert scale, from Strongly Disagree (1) to Strongly Agree (5). Each of these is described below.

The first question: **L1**: "I feel secure about sharing my personal information with Acme after viewing their privacy practices" attempts to capture participants' reaction to the actual content of the privacy policy they read. **L2**: "I feel that Acme's privacy practices are explained thoroughly in the privacy policy I read" questions whether participants believe their practices are well displayed.

The next three questions deal with the experience of interacting with the privacy policy in the format we presented. **L3**: "Finding information in Acme's privacy policy was a pleasurable experience" has participants rate their enjoyment of finding information. **L4**: "I feel confident in my understanding of what I read of Acme's privacy policy" investigates participants' perceived accuracy in the earlier questions. **L5**: "It was hard to find information in Acme's policy" has participants rate the difficulty they had in finding information.

The final question, **L6**: "If all privacy policies looked just like this I would be more likely to read them" attempts to capture whether our proposed label would encourage more people to read privacy policies.

5.3.3 Policy Comparison Questions

The third section requires participants to compare two policies of the same format (ABC Group v. Bell General for NL or Acme v. Button Co. for the label). One of the policies in each comparison is the same policy from the initial 8 information-finding questions.

The first four questions in this section are True/False statements such as "By default, **Button Co.** can share information about your purchases with other companies, but **Acme** cannot."

The final two questions in this section are opinion questions, asking: "Which company will better protect your information online?" and "You're looking to buy a gift online. At which company would you prefer to shop?"

5.3.4 Policy Comparison Enjoyment & Ease

The final four questions are again on the 5-Likert scale presented earlier. They are in two pairs, the first pair asking if, "Looking at policies to find information was an enjoyable experience" and "Looking at policies to find information was easy to do." The second pair focuses specifically on the comparison task, "Comparing two policies was an enjoyable experience" and "Comparing two policies was easy to do."

6. RESULTS

The results from our laboratory study are presented below. First

Table 2. McNemar's p-values & Benjamini-Hochberg Correction p-values for information finding questions 1-8 (5.3.1), and policy comparison questions 15-18 (5.3.3).

	Label	NL	McNemar's	Benjamini-Hochberg Correction
1	96%	100%	NS	NS
2	**88%**	**29%**	**0.00024**	**0.0014**
3	100%	96%	NS	NS
4	92%	100%	NS	NS
5	54%	25%	0.12	0.21
6	**79%**	**21%**	**0.00012**	**0.0014**
7	75%	54%	0.3	0.45
8	88%	58%	0.09	0.18
15	96%	63%	0.06	0.14
16	92%	79%	NS	NS
17	**83%**	**38%**	**0.007**	**0.021**
18	**71%**	**25%**	**0.0009**	**0.0036**

we will address the issue of information finding through our quantifiable accuracy results. Next we describe the timing data on those questions, showing information finding is not only more accurate but also faster with label polices than with NL policies. To conclude this section we will present the "likeability" of the privacy label.

6.1 Accuracy Results

At a high level, people were able to answer more questions correctly with the label. We compared the correct number of total questions, per participant, for the label vs. the natural language policy, M = 10.13 and M = 6.83 respectively, $t(23) = 7.41$, $p < 0.001$.

We explored each of the questions individually by testing the proportions of correctness for each question by condition, using McNemar's test. These results combine participants who saw the label first and with participants who saw the label second as accuracy differences were not significant between these two conditions. These comparisons show that the label is significantly more accurate in 2 of the 8 information-finding questions and 2 of the 4 policy-comparison questions. The accuracy rates for each question are shown in Table 2, with statistically significant comparisons shown in bold.

We performed a Benjamini-Hochberg correction to account for multiple testing across comparisons. Each of the paired proportions are shown in Table 2 along with the McNemar's p-values and the corrected p-values.

6.2 Timing Data

For each of the information-finding and policy-comparison questions we collected time-to-task completion data. As shown in Table 3, the label was significantly faster than the natural language policies for both the group of information-finding questions and the group of policy-comparison questions ($p<0.001$).

To test the mean task completion time for accurate answers we removed all timing results where the resulting answer was inaccurate and calculated means per question, per condition. Using a 2-sided t-test the label is significantly faster in 2 of the 8 information-finding questions and significantly faster in 3 of the 4 policy-comparison questions. In only one question was the average time faster for participants using the natural language policy, and this difference was not significant. The full results for this test can be found in Table 4.

6.3 Satisfaction Results

The satisfaction results were captured based on participants' responses on a Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree). We computed the mean response for the label and for natural language, both combined, and also separated by which format was viewed first. For each of these questions higher is better, including Question L5 "information was hard to find," which was reversed to be consistent with the remaining questions.

We performed t-tests for each of these questions, to compare the label to the natural language policies. All but 2 of these 10 questions resulted in significant results. The label was rated significantly more pleasurable, easier to find information in, and easier and more enjoyable to use when comparing two policies.

Table 3. Time-to-task comparisons between the label and natural language policies. Shorter times are better. Information Finding is questions 1-8 (5.3.1), Policy Comparison, questions 15-18 (5.3.3)

Times in seconds.	**Label**	**NL**
Information Finding	174.5	349.6
Policy Comparison	120.0	292.4
Average Total Time	339.9	692.0

Table 4. Time differences and p-values for average time per question comparing only correct answers. All times reported in seconds.

	Label	NL	Difference	**p-value**
1	37.58	61.27	23.69	0.07
2	**21.67**	**85.7**	**64.03**	**0.04**
3	**14.35**	**50.07**	**35.72**	**<.001**
4	18.89	23.09	4.2	0.4
5	34.51	29.95	-4.56	0.46
6	20.19	50.24	30.05	0.06
7	16.32	22.82	6.5	0.88
8	26.93	36.79	9.86	0.73
15	**46.58**	**132.69**	**86.11**	**0.0006**
16	**34.36**	**68.32**	**33.96**	**0.05**
17	21.91	35.48	13.57	0.28
18	**12.24**	**47.36**	**35.12**	**0.03**

The results from each of these questions are shown with means and p-values in Figure 8.

Additionally we performed 2-sample t-tests between conditions to exploring priming effects, where opinions have changed based on the policy format a participant viewed first. When looking at how participants answered the Likert scale questions about the label by condition, 3 questions had significant results. Participants felt significantly more secure when viewing the grid if they saw the NL policy first, (label first=2.92, NL policy first=3.92, p=0.03) reported they were significantly more likely to read policies more in the label format if they saw the NL policy first (label first=4, NL policy first=4.5, p=0.04), and found comparisons on the label significantly easier when viewing the NL policy first (label first=3.92, NL policy first=4.58, p=0.004). These results show significant priming to appreciate the grid more when the NL policy was viewed first.

6.4 Observations

The initial results we have presented above are very strong, however there is still much room for improvement. We observed that some participants still found elements of the label confusing. We began an additional round of iterative design and testing to address some of the issues we observed during the lab study.

Several participants were confused by the symbols we used to indicate opt-in and opt-out. For instance, one participant did not understand what "out" meant, saying, "I've been messing things up because I thought 'out' meant 'out of the question.'" To

improve users' comprehension, we will alter the symbol design to include the full phrases "opt-out" and "opt-in."

In addition, several participants in the lab study were completely unfamiliar with the terms opt-out and opt-in, and they assumed that the terms meant exactly the same thing. We will continue to refine our glossary definitions to help educate users about these concepts. The original definitions did not explain the terms opt-in and opt-out, with the legend reading "we will collect and use your information in this way unless you opt-out." The new definitions help explain the concepts, stating: "we will collect and user your information in this way unless you tell us not to by opting out." We plan to further test our design changes in focus groups, and believe that the design iterations will continue to improve the speed and comprehensibility of the Privacy Nutrition Label.

7. DISCUSSION

We began this paper with three factors in mind: the ability to find information, the understanding that there are differences between privacy policies and control over one's information, and the simple time-based costs of reading privacy policies. We strove to design a single page summary of a company's privacy policy that would help to remedy each of these three concerns and at the same time be enjoyable.

We believe that the results presented above clearly show that each of these areas was addressed. Accuracy results were better or similar for information finding and policy comparison. Task completion times were significantly lower when using the label than when using a natural language policy. And across the board, participants believed information was easier to find and had a more pleasurable time finding it using the label.

The final label design allows for information to be found in the same place every time. It removes wiggle room and complicated terminology by using four standard symbols that can be compared easily. It allows for quick high-level visual feedback by looking at the overall intensity of the page, can be printed, can fit in a browser window, and has a glossary of useful terms attached. People who have used it to find privacy information rated it as pleasurable. They not only rated it better than the natural language, but actually rated it enjoyable to use.

When using the label people far more consistently selected the company that had the stronger privacy policy. Participants also realized the benefits of the label for comparison: "This may actually be the biggest advantage of this system because you can put down two polices that are formatted the same and see the exact differences between them. It's really easy." Even more directly one participant said "I guess I'll look to see which policy has more blue," exactly capturing one of our intended design goals.

A number of open questions remain about how people will use the label in practice. Will people make more use of the label than they currently do of privacy policies? How will their use change as they become more familiar with the labels through continued use over time?

Our next step will be to iterate on a number of additional minor changes and then run a large online study, similar to Reeder et al.'s original test of the P3P Expandable Grid [23]. This will further confirm over a much larger and more diverse group of people that the label is in fact, more accurate, faster, and more pleasurable. Additionally as this study will be conducted online, people will be viewing privacy policies just as they normally would, at their computer, which is very different than performing these tasks in our laboratory on paper.

Finally, we plan to integrate a version of the privacy label into Privacy Finder, a privacy search engine maintained by the CUPS Laboratory. This will allow people to use the label outside of the context of a research study and will allow us to monitor frequency of use while collecting feedback on the label design. It is likely this public online deployment that will bring us closer to answering how much a standardized label design assists people over time as they become accustomed to using it.

8. ACKNOWLEDGMENTS

The authors would like to acknowledge Sungjoon Steve Won for his early designs, including the simplified grid; Janice Tsai for her statistical expertise; Daniel Rhim, Robert McGuire, and Cristian Bravo-Lillo for their technical assistance and assistance in conducting user studies; Norman Sadeh and Aleecia McDonald for their guidance and advice; and everyone who provided input throughout the design process.

This work was supported in part by U.S. Army Research Office contract DAAD19-02-1-0389 ("Perpetually Available and Secure Information Systems") to Carnegie Mellon University's CyLab, by NSF Cyber Trust grant CNS-0627513, by Microsoft through the Carnegie Mellon Center for Computational Thinking, FCT through the CMU/Portugal Information and Communication Technologies Institute, and the IBM OCR project on Privacy and Security Policy Management.

9. REFERENCES

[1] Balasubramanian, S. and Cole, C. "Consumers' Search and Use of Nutrition Information: The Challenge and Promise of the Nutrition Labeling and Education Act." *Journal of Marketing.* 2002. Vol. 66, 112-127.

[2] Beard, T.C., Nowson, C.A., Riley, M.D. "Traffic-light food labels." *Med J Aust.* 2007;186:19.

[3] Belser, B. Designing the Food Label: Nutrition Facts. *AIGA Journal.* 1994.

[4] Buckley, P. and Shepherd, R. Ergonomic factors: The clarity of food labels. *British Food Journal.* 1993. 95

[5] Byrd-Bredbenner, C., Alfieri, L., Wong, A., and Cottee, P. The Inherent Educatiional Qualities of Nutrition Labels. Family and Consumer Sciences Research Journal, Vol 29, No 3, March 2001 265-280.

[6] Cranor, L., Egelman, S., Sheng, S., McDonald, A., and Chowdhury, A. P3P Deployment on Websites. *Electronic Commerce Research and Applications*, Volume 7, Issue 3, Autumn 2008, Pages 274-293.

[7] Consumer Product Safety Commission. "Labeling Requirements for Toy and Game Advertisements." 2008. http://cpsc.gov/library/foia/foia08/brief/toygameads.pdf

[8] DeJoy, D.M., Cameron, K.A., and Della, L.J. Post-exposure evaluation of warning effectiveness: A review of field studies and population-based research. *The Handbook of Warnings.* 2006. *(35-48)*.

[9] Downs J.S., Loewenstein G., and Wisdom J. Strategies for Promoting Healthier Food Choices. *American Economic Review.* 2009, vol. 99, issue 2, pages 159-64

[10] Drichoutis AC, Lazaridis P, Nayga RM.2006. Consumers' use of nutritional labels: a review of research studies and issues. Acad Marketing Sci Rev, no. 9.

[11] The Energy Label. 2007. www.energyrating.gov.au

[12] European Union Commission Directive 98/11/EC "Energy Labeling." 1998. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1998:071:0001:0008:EN:PDF

[13] Food Standards Agency. "Signpost Labeling Research." 2005 http://www.food.gov.uk/foodlabelling/signposting/siognpostlabelresearch/

[14] Jensen, C. and Potts, C. Privacy policies as decision-making tools: an evaluation of online privacy notices. SIGCHI. 2004.

[15] Kelley, P., A. McDonald, R. Reeder, and L. Cranor. P3P Expandable Grids. Poster at Privacy MindSwap Carnegie Mellon University. 2007. http://cups.cs.cmu.edu/soups/2008/posters/kelley.pdf

[16] Kleimann Communication Group, Inc. Evolution of a Prototype Financial Privacy Notice. February 2006. Available: http://www.ftc.gov/privacy/privacyinitiatives/ftcfinalreport060228.pdf

[17] Levy, A. and Hastak, M. Consumer Comprehension of Financial Privacy Notices. December 2008. Available: http://www.ftc.gov/privacy/privacyinitiatives/Levy-Hastak-Report.pdf

[18] Maubach, N., Hoek J. "The Effect of Alternative Nutrition Information Formats on Consumers' Evaluations of a Children's Breakfast Cereal" Proceedings of the EParternships, Proof and Practice – International Nonprofit and Social Marketing Conference 2008.

[19] McDonald, A., Reeder, R.W., Kelley, P.G., and Cranor, L.F. A Comparison of Online Privacy Policies and Formats. Privacy Enhancing Technologies 2009.

[20] McDonald, A, and Cranor, L. The Cost of Reading Privacy Policies. Telecommunications Policy Research Conference, 2008.

[21] Privacy Leadership Initiative. Privacy Notices Research Final Results, November 2001, Available at: http://www.understandingprivacy.org/ content/library/datasum.pdf.

[22] Reeder, R.W. *Expandable Grids: A user interface visualization technique and a policy semantics to support fast, accurate security and privacy policy authoring.* PhD thesis, Carnegie Mellon. 2008. http://www.robreeder.com/pubs/ReederThesis.pdf

[23] Reeder, R., Cranor, L., Kelley, P., and McDonald, A. A User Study of the Expandable Grid Applied to P3P Privacy Policy Visualization. *Workshop on Privacy in the Electronic Society.* 2008

[24] Seymore, J.D., Lazarus Yaroch, A., Serdula M., Blanck, H.M., and Khan, L.K. "Impact of nutrition environmental interventions on point-of-purchase behavior in adults a review." *Preventative Medicine* 2004. 29: S108-S136.

[25] The Center for Information Policy Leadership, H. . W. L. Multi-layered notices.

[26] Turow, J. Feldman, L., and Meltzer, K. Open to Exploitation: American Shoppers Online and Offline. The Annenberg Public Policy Center. 2005. http://www.annenbergpublicpolicycenter.org/NewsDetails.aspx?myId=31

[27] U.S. Food and Drug Administration. A Food Labeling Guide. Center for Food Safety & Applied Nutrition. 1999. http://vm.cfsan.fda.gov/%7Edms/flg-toc.html.

[28] U.S. Food and Drug Administration. "Guide to Nutrition Labeling and Education Act Requirements" 1994. http://www.fda.gov/ora/inspect_ref/igs/nleatxt.html

[29] U.S. Food and Drug Administration. "New OTC Drug Facts Label" *FDA Consumer Magazine.* 2002. http://www.fda.gov/FDAC/features/2002/402_otc.html

[30] W3C. The Platform for Privacy Preferences 1.0 (P3P1.0) Specification. http://www.w3.org/TR/P3P/

[31] W3C. The Platform for Privacy Preferences 1.1 (P3P1.1) Specification. http://www.w3.org/TR/P3P11/

[32] WELS Regulator (Australian Government). "WELS and Watermark." 2005. http://www.waterrating.gov.au/compliance.html