Towards a Visual Vocabulary for Privacy Concepts

Vivian Genaro Motti[1] and Kelly Caine[2]
[1] George Mason University, Fairfax - VA, USA
[2] Clemson University, Clemson - SC, USA
vmotti@gmu.edu

Designing privacy controls that are intuitive for end users involves several challenges. Graphic representations could certainly contribute to tackle such challenges, however, a visual vocabulary for privacy solutions is currently lacking. To contribute to that end, in this paper we report a review of online images related to privacy. We identify graphic representations that illustrate privacy objects, mechanisms and actions, and categorize them according to the concepts covered. Seeking to develop a set of icons to illustrate privacy concepts, we analyze online contents publicly available from social media and extract representations that commonly refer to privacy concepts. The contributions of this paper include: (1) a set of graphics that represent privacy mechanisms, objects and actions; and (2) a vocabulary for such graphic representation. The evidence is drawn from the analysis and discussion of 241 images selected by the authors from ten online image repositories.

INTRODUCTION

Emerging technologies, such as: mobile and wearable computers enable pervasive data collection and instant information access. Despite the many benefits that emerging technologies provide for end users in their daily lives, given the widespread data collection and content sharing online, these technologies also require higher levels of privacy control.

Despite the significant efforts dedicated to study intuitive privacy controls, privacy is a multidisciplinary, complex concept without a universal solution. Intuitive privacy-enhancing solutions are complex to design and challenge all stakeholders involved on it (e.g., developers, designers, researchers).

Graphic representations though have already been proven valuable solutions to inform users in information systems. Visualization tools have been effectively employed to enhance privacy protection in online communities (Balebako et al., 2012), as an alternative for long, verbose and complex privacy policies. Their easy recall, quick recognition, and steep learning curve, make them suitable to provide privacy controls and notices for users in resource-limited settings (Schaub et al., 2015), including wearable devices with small graphic displays.

This work focuses on understanding the users' perceptions on privacy, aiming to identify a shared graphical representation of this concept. To gain a deeper insight in to users' mental models about visual privacy, we analyze an extensive set of online images (n=241) published by social media users, UI designers and content producers. We coded and categorized the images retrieved to identify the concepts that are often associated with privacy illustrations.

The privacy images analyzed are mostly related to users' objects, actions and control mechanisms. We hypothesize that daily objects used in real world tasks to ensure users' privacy can inspire design requirements for privacy icons and lead to more intuitive privacy solutions. Graphic representations provide intuitive solutions for users by immediately grabbing their attention and fostering their understanding, besides allowing to nudge them towards careful privacy decisions.

VISUAL PRIVACY

Visual notices for privacy consist of text, images, icons, or a combination thereof (Schaub et al., 2015). In visual privacy, both presentation and layout are important. Colors, fonts, and white space all impact users' attention and comprehension.

Previous work on visual privacy has explored different graphic solutions to enable privacy controls.

To analyze the influence of users' ages in their privacy understanding, Lorrie Cranor collected and tagged privacy images uploaded by users around the globe in an online repository (Cranor et al., 2015), (Balebako et al., 2012). The most frequent tags associated with the images are: control, computer, camera, smartphone, social media and surveillance. Most images uploaded are complex representations that include various objects and textual descriptions, being thus unsuitable for small graphic displays. This online repository focuses on data collection, and provides a word cloud (created with the images' tags) for content analysis.

To represent privacy visually, Matt McKeon created an interactive tool that represents the evolution of the default privacy settings of Facebook (McKeon, 2015). This tool shows significant changes on default privacy settings along 5 years (from 2010 to 2015). Although the graphic is limited to one social media channel, it revealed a strong interest of the community. By explicitly representing privacy settings to end users, his tool fostered discussion about privacy, gathering more than 500 online comments about the representation.

To investigate privacy visualizations for data sharing, Caine et al. (2011) employed concentric circles, icons, numbers and text, and studied how the users' perspectives could be influenced by alternative representations of disclosure control. Despite also focusing on privacy illustrations, this work covers mainly privacy requirements for data sharing.

To analyze graphic representations for privacy policies, Kelley et al. (2009) drew inspiration from nutrition labels. The privacy table proposed by the authors employs different colors, however, no specific images or icons have been used.

Google material icons (2015) provide 794 icons an extensive number of graphic designs for designing user interfaces.

These representations are light weighted and intuitive being thus suitable small displays and limited interaction settings (e.g., slower Internet connections, small display sizes, limited power sources, black and white displays). Still, they cover generic UI features, and few icons are actually associated with privacy concerns (e.g., lock, person, group, data transmission). Regarding privacy specifically, Google material icons partially cover data collection, storage, transmission and sharing.

With a set of user studies, Egelman et al. (2015) investigated the most intuitive representations of privacy icons. In their method, designer solutions are combined with users' feedback to identify intuitive designs for icons. The applications covered focus on ubiquitous environments in general.

To the best of our knowledge, despite significant efforts to better understand users' perceptions on privacy, so far no work has investigated visual privacy based on online image repositories, in which users voluntarily contribute with contents and explicitly tag it with privacy.

METHOD

The methodology of this study is structured in three main steps: first, the image sources were selected; then, the images were collected, coded and categorized. Finally, we analyzed the results and defined a vocabulary for visual privacy.

Image Sources

To analyze privacy illustrations, we adopt an original approach, collecting images from online sources. The analysis of online contents has several benefits, e.g.: users' voluntarily upload the contents, participants are not recruited by convenience sample but are geographically distributed, and the study is anonymized. The images analyzed included photos, sketches, and illustrations from diverse online sources – social media channels, micro blogs, websites and search engines that allow users to upload media. To prevent bias from search results that are personalized based on previous search history, an incognito browser window was used for the search. The ten sources selected are: Instagram, Flickr, Pinterest, Lorrie Cranor's repository on illustrated privacy, Tumblr, Google Images, Shutterstock, Mozilla, Material Icons and Privicons. These sources involve three typical user profiles:

Crowd – end users: general public interested in graphic representations but in principle without commercial interests.

- Instagram: free online photo sharing and social network platform;
- Flickr: image hosting website, online community for photo researchers and bloggers;
- Lorrie Cranor's (LC) Repository: website featuring privacy illustrations from kindergartners through adults;
- Pinterest: web and mobile application that allows users to upload, save, sort and manage images;

UI designers – web designers: professionals, expert designers.

- Mozilla Icons: a small set of icons designed to illustrate privacy policies;
- Material Icons: an extensive set of icons to represent users' actions, as a command, a file, a device or a directory;
- Privicons: an approach to express and simplify privacy policies;

Professionals: journalists and content producers.

- Google Images: a comprehensive image search;
- Shutterstock: an online repository for photos, illustrations and vector art;
- Tumblr: a microblogging platform and social network website that allows users to post multimedia contents);

Tumblr is primarily an online repository for users to post personal images. As the images retrieved via this source were mainly re-posts for content producers and journalists, we classified its user profile as professionals instead of crowd.

The image sources were selected after an online search for image repositories and privacy icons. Large and commercial repositories (such as: Instagram, Flickr and Pinterest) were combined with private academic sources (Lorrie Cranor' repository) and corporative design solutions (Mozilla and Google). All contents were publicly available. We combined different image sources to reach a more diverse user sample.

Data Collection

Once the online sources were selected, we searched each source using the keyword 'privacy' as the search term. To collect the images, first three main inclusion and exclusion criteria were agreed among the study authors. The images had to be readable, clear and in a good resolution. Also, the images should not contain marketing and advertisement, quotes or only text. Finally, the images that are related to privacy were included. A broad definition of privacy was used.

All images that met the inclusion criteria were individually extracted, downloaded, and saved in a local collaborative repository. Table 1 provides an overview of the number of images that returned from the search, and the number of images that met the criteria, and were retrieved for further analysis. The search performed yielded more than 379,580 images, among which 241 (meeting the inclusion criteria) were extracted for analysis.

Table 1. Results retrieved from the search (more than 379,580) and number of images analyzed (n=241). The search was performed in August 2015. *NA* means *not applicable*, because Google images, Pinterest and Tumblr do not provide the total number of results retrieved.

		Total	Retrieved
Crowd (n>270,925)	Instagram	118,702	21
	Flickr	152,053	20
	LC Repository	170	21
	Pinterest	*NA*	21
UI Designers (n=813)	Mozilla	13	13
	Material Icons	794	75
	Privicons	6	6
Professional (n>107,842)	Google Images	*NA*	21
	Shutterstock	107,842	21
	Tumblr	*NA*	22
Total		**> 379,580**	**241**

As an exhaustive analysis of all images would not be feasible, we retrieved the first search results (some repositories sort them per date, so that most recent content is presented, others sort it per relevance, so that only images linked with privacy tags are retrieved). The images analyzed reflect the most relevant results retrieved, based on the criteria used by the search engines, online repositories, and social media channels. Instagram and Pinterest for instance sort the search results based on the date when the content was posted. Our analysis was not exhaustive, as manually coding all images returned from online searches would be unpractical; instead we focus on a limited image sample (24 images on average) to illustrate updated results from 10 different sources.

Image Analysis and Coding

We used themes to organize the images, and connect the emerging concepts. In a pre-analysis, a sub-set of images was used to identify specific codes. Analogously to Schaub's design space to define privacy (Schaub et al., 2015), the images retrieved were broadly classified in four main themes:

- **Who:** people, institutions or organizations involved in discussing, providing or threatening users' privacy;
- **How:** objects, actions, behaviors, attitudes and mechanisms that enable privacy control;
- **Why:** users' goals to obtain privacy, feelings, intents, and emotions involved;
- **Where:** places and locations, real world scenarios where users perceive to have their privacy warrantied.

The qualitative content analysis consisted in observing each image, empirically analyzing its key features, similarities and differences, and coding it. The codes emerged ad-hoc, in a bottom up approach of individual analysis of each image.

Some images were straightforward to code (e.g. a door and a camera are both objects), others represent multiple concepts simultaneously. For images that evoked two or more descriptive codes, we used their key message for coding, or we collaborative discussed and agreed upon one unique code. Through coding and affinity diagrams, each image was manually and individually coded. Images that could not be classified were discarded from the analysis (n=6). The codes were analyzed and lead to a visual vocabulary to aid in graphic representations of privacy-enhancing solutions.

RESULTS

The search performed yielded more than 379,580 images, among which 241 were extracted for further analysis (Table 1). The images selected were categorized in four high level themes: who, how, why and where and seven general codes: people, institutions, objects, actions, mechanisms, concepts, and places, which were then subdivided in 15 descriptive terms: roles, public persons, circles, regulatory, social media & IT, control, storage, sensor, blocker, attitudes & behaviors, regulations, policy, feelings, indoor and outdoor. In the initial coding phase, with a bottom up approach 97 different descriptions emerged. Table 2 shows 5 codes, 9 terms and their respective descriptors. Figure 2 provides an overview of the images selected for analysis, per user profile and source.

Table 2. Five codes and their descriptors for visual privacy

Action	Analyzing, Authenticating, Blindfolding, Blocking, Blurring, Covering, Connecting, Closing, Dimming, Disclosing, Dimming, Erasing, Forwarding, Hiding, Localizing, Locking / Unlocking, Looking, Observing / being observed, Packing, Protecting, Protesting, Revealing, Sharing, Shredding, Spying, Surveilling, Synchronizing, Uploading, Uncovering
Objects	
Blockers	Blinders, Curtains, Diary, Door, Fence, Gate Key, Message, Padlock, Wall, Windows
Control	Semaphore
Sensors	Camera, Camcorder, Microphone
Storage	Memory Card, Cloud
Organizations	
Regulatory	NSA, HIPPAA
Social Media & I.T.	Ashley Madison, Bitcoin, Facebook, Google+, Instagram, Pinterest, RSS, Twitter, Whatsapp
People	
Role	Politicians, Legislators
Public Persons	Edward Snowden, George Orwell
Circles	Group, Individual
Abstract Concepts	Betrayal, Confidentiality, Creepiness, Exclusivity, Fear, Intimacy, Isolation, Loneliness, Public vs. Private, Safety, Secrecy, Shame

Codes and Descriptions

The seven high-level codes identified include:

1. People: specific roles, as politicians and legislators, public persons, i.e. individuals related to revelation, such as, Edward Snowden or "big brother", such as, George Orwell, or granular classifications (individual or group).

2. Institutions: organizations that either promote privacy or threat it. Examples: NSA (National Security Agency), HIPPA (Health Insurance Portability and Accountability Act), Social Media and communication Channels (Facebook, WhatsApp), and Internet Providers (AT&T).

3. Objects: controls to manage privacy, storage to keep users' data, sensors to collect users' data and blockers to prevent data access by untrusted, unknown, or unwanted individuals. A semaphore exemplifies a control object. Storage items include memory cards and the cloud. Examples of blockers are plants, insufilm, notes, doors, tempered glass, gates, fences, walls, and windows. Sensors include webcams, camcorders, microphones, and mobile devices.

4. Actions: concern what users do to ensure their privacy, their common attitudes and behaviors in real world scenarios or information systems, such as: hiding, revealing, covering, blocking, sharing, or granting access.

5. Mechanisms: strategies used to ensure privacy control. Examples: a browser add on, terms of service, privacy policies, and privacy settings.

6. Abstract Concepts: abstract concepts related to users' concerns and feelings about privacy, e.g.: privacy vs. security, confidentiality, silence, secrecy, isolation, exclusivity.

7. Places: where users seek for and find privacy. Indoor locations, such as, a house, bedroom or bathroom; or outdoor locations, such as, a garden, or a deserted island.

Frequency of Codes

The number of codes related to the images varied per source, but certain trends emerged. The most common codes referred to actions (n=73) and objects (n=39), regardless of the image source. This pattern is common for all sources, but depending on the source, the images retrieved presented more divergent or convergent contents, i.e. the frequency of codes varied per source and user profile too.

As expected, for images authored by the UI designers (Privicons, Material Icons and Mozilla Icons) the most frequent code was action (n=45). The frequency of the action code was followed by object (n=8) and people (n=6). For professional images (generated or published by content producers), retrieved from Tumblr, Shutterstock, and Google Images, the codes varied, but still, action (n=20) was the most frequent, followed by concepts (n=9), mechanisms and object (both with n=7). For the crowd-sourced images, posted by end users, objects (n=23) and actions (n=12) were more frequent, followed by concepts (n=5). The remaining codes were employed just once or twice.

Most sources (n=5) had the images classified in three or less codes, and most images (n=112) were closely related to objects and actions, especially those produced by UI designers. The images uploaded by the crowd (UI users) were the most divergent concerning the codes employed, Instagram was the source with more variability and the only source whose images were classified with all codes. The most homogeneous results were found for Privicons, whose images received just action codes, and for Pinterest, whose images represented mainly objects (n=14), with one exception (the painting of a window glass to remove transparency, coded as action).

Nuances of Privacy

Due to the nature of the sources selected, and organic nature of this study, multiple nuances of privacy were noted. Instagram images focus on personal perspectives (Instagram, 2015), and Pinterest images target at physical controls for privacy (e.g., architectural and design solutions) (Pinterest, 2015). Google Images and Shutterstock, returned professional or presentation contents (e.g., diagrams, privacy settings). Flickr returned more professional photos (better quality, higher resolutions, larger dimensions, etc.) illustrating concrete examples of privacy in real world scenarios (e.g., surveillance cameras in the streets, privacy checks in online systems, shoulder surfing of a mobile phone in a metro).

Some online sources also suggested related terms and filters. Pinterest suggested window, plant, and Internet as related terms. When Internet was added to the search (Pinterest, 2015), most results showed privacy infographics (e.g., 'Is there such a thing as online privacy'), and information about privacy settings (e.g., 'How to set a VPN'). The contents were informative (illustrated instructions for online users to keep their information safe). Tumblr suggested privacy and technology, school of privacy, and privacy in libraries in the digital age, as related terms, and Google Images suggested fence, guard, and trees.

PRIVACY METAPHORS

The 109 terms that emerged from the images' analysis enabled us to define a 4-tier privacy vocabulary including themes, codes, descriptions, and examples. By referring to common objects and situations, we expect such vocabulary to inspire the design of privacy solutions that are closer to users' views.

When we focus on how privacy is ensured (or threatened), we identify: (i) real world actions that users relate to privacy, (ii) physical objects that users employ to achieve privacy, and (iii) mechanisms that are useful for users to control their privacy. These actions, objects and mechanisms were the most frequent codes in our analysis. Their high occurrence suggests a potential for successfully employing them in the design of privacy solutions that are more intuitive and easier to use.

Drawing from the analysis and results of our study, frequency of the codes and descriptions, and focusing particularly on functional requirements for privacy-enhancing solutions (e.g., (Aquisti et al. 2015)), the images coded were analyzed under the light of four major privacy requirements.

For **data collection**: related icons can illustrate either the sensor (or object) that collects users' data (e.g., a microphone or a video camera), or a representation of the data itself (e.g., a location pin, a heart rate signal).

For **data transmission**: icons can illustrate the logo of the network protocol used for synchronization, communication, connectivity, and data transference (e.g., Bluetooth, NFC, Wi-Fi) or the physical connection used (e.g., USB connector).

For **data storage**: the representations include the physical object used (e.g., a memory card, a server) or a virtual metaphor (e.g., the cloud), representing where the data is stored.

For **data sharing and access control**: potential illustrations include who has access to the data, e.g. individual users represented through a contact photo or avatars, a granular representation of users, such as, one individual, or a closed group of individuals (e.g., family, friends), or even a general public (e.g., a globe showing global access). Thus far, metaphors, such as, a globe and an open (or closed) eye, have been used to distinguish access types (global, public, limited or none).

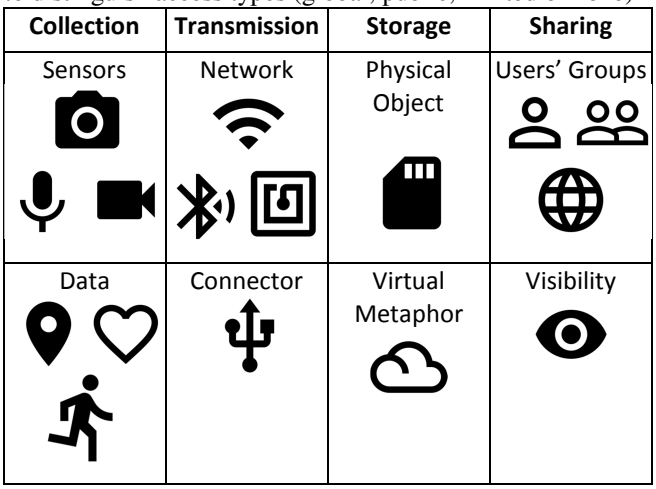

Figure 1. 16 icons illustrating privacy for: data collection (sensors or data), transmission (network and connectors), storage (physical object and virtual metaphor), and sharing (users' groups). Source: Google Material Icons.

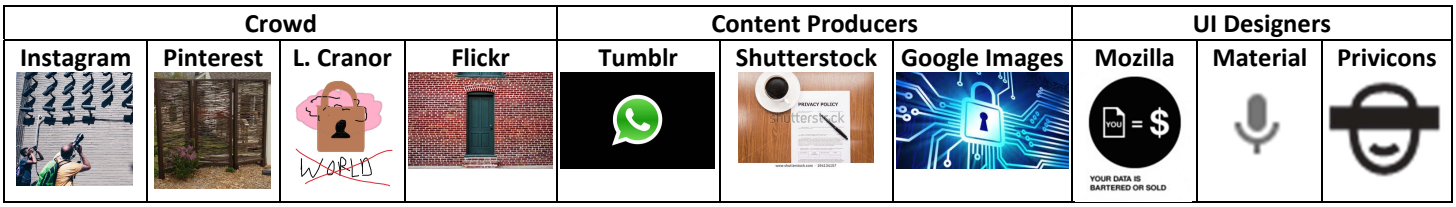

Figure 2. Examples of 10 privacy images retrieved for analysis, per user profile and source.

Figure 1 illustrates a sample of 16 icons retrieved from Material Design that are related to: data collection (sensors and data), transmission (network and connector), storage (physical object and virtual metaphor) and sharing (users' group and visibility). Besides these four main requirements for privacy, the overall privacy control also needs a system representation. Potential icons that are associated with privacy controls include: shields, keys, engines, locks and semaphores.

The icons discussed above correspond directly to users' daily objects or situations, which can aid on the users' understanding and recall. However, those icons are an abstract representation of a concept, covering a static moment in the system status or a possible action for the user interaction. To add information or a dynamic behavior to icons, one potential strategy involves annotating the representation of a graphic icon to indicate a variation, e.g., by filling (or unfilling) the icon, crossing, double-crossing (or uncrossing) it, marking with specific signs (exclamation dot, question mark, etc.).

DISCUSSION

While privacy can be studied from multiple facets (Motti and Caine, 2015), only a user-centric approach can aid us to better understand users' perspectives and to design solutions that are more intuitive for their privacy control. This approach aids to bridge the gap between users' needs and privacy features, and to ensure that digital solutions (privacy icons, features and mechanisms) match with real world solutions that users already employ and are familiar with in their daily lives. Although we assume that users may find such solutions more intuitive and easy to use, further efforts are needed to cross-validate this hypothesis. At this stage, our findings are limited to one interaction modality (graphics) and further work is needed to translate and assess design implications to other modalities (e.g., audio, text, or widgets) and users' profiles.

CONCLUSION

Privacy is a multidisciplinary concept involving crosscutting concerns. Despite the fact that no universal solution can address privacy problems through intuitive solutions, there is a high demand for privacy-enhancing solutions that are intuitive and easy to use. By better understanding users' perspectives about privacy, and especially concerning its visual representations, with this study we shed light in to design opportunities for illustrating privacy. The key contributions of this work are twofold: (1) a better understanding of users' perspectives on visual representations for privacy and (2) a vocabulary for visual privacy. As future work, we plan to cross validate the findings of this study, conducting user studies to assess the effectiveness of the graphic representations identified.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1314342. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do notnecessarily reflect the views of the National Science Foundation.

REFERENCES

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. Science, 347(6221), 509-514.

Balebako, R., Leon, P., Shay, R., Ur, B., Wang, Y., & Cranor, L. (2012, May). Measuring the effectiveness of privacy tools for limiting behavioral advertising. In Web 2.0 Security and Privacy Workshop.

Caine, K., Kisselburgh, L. G., & Lareau, L. (2011, May). Audience visualization influences disclosures in online social networks. In CHI'11 Extended Abstracts on Human Factors in Computing Systems (pp. 1663-1668). ACM.

Lorrie Cranor – Privacy Illustrated Website. (2015) At: http://cups.cs.cmu.edu/privacyillustrated/

Egelman, S., Kannavara, R., & Chow, R. (2015, April). Is This Thing On?: Crowdsourcing Privacy Indicators for Ubiquitous Sensing Platforms. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 1669-1678). ACM.

Google Material Icons. Available at: https://www.google.com/design/icons/

Instagram. At: https://instagram.com/explore/tags/privacy/

Kelley, P. G., Bresee, J., Cranor, L. F., & Reeder, R. W. (2009, July). A nutrition label for privacy. In Proceedings of the 5th Symposium on Usable Privacy and Security (p. 4). ACM.

Matt McKeon – The Evolution of Privacy on Facebook. Interactive Visualization. At: http://mattmckeon.com/facebook-privacy/

Motti, V. G., & Caine, K. (2015). Users' Privacy Concerns About Wearables. In Financial Cryptography and Data Security (pp. 231-244). Springer Berlin Heidelberg.

Pinterest. URL queried: https://www.pinterest.com/search/pins/?q=internet%20privacy&term_meta%5B%5D=privacy%7Ctyped&term_meta%5B%5D=internet%7Cguide%7Cword%7C6&add_refine=internet%7Cguide%7Cword%7C6

Schaub, F., Balebako, R., Durity, A. L., & Cranor, L. F. (2015). A Design Space for Effective Privacy Notices. In Eleventh Symposium On Usable Privacy and Security (SOUPS 2015) (pp. 1-17).