USABLEPRIVACY.ORG
the usable privacy policy project

project newsletter june 2016

the usable privacy policy project

Natural language privacy policies are the de facto standard for "notice and choice" on the Web. Yet, few users read these policies and those who do struggle to understand them. Initiatives aimed at the development of machine implementable standards or other solutions that require service providers to adhere to more stringent requirements have run into obstacles, with many website operators showing reluctance to commit to anything more than what they currently do.

Our project builds on recent advances in natural language processing, privacy preference modeling, crowdsourcing, formal methods, and privacy interfaces to overcome this situation. It combines fundamental research with the development of scalable technologies to:

1. Semi-automatically extract key data practices and privacy policy features from natural language website privacy policies, and
2. Present these features to users in easy-to-digest formats that enable them to make more informed privacy decisions as they interact with websites, mobile apps and online services.

Through our work we hope to overcome the limitations of current natural language privacy policies without imposing new requirements on service providers. Our work also involves the systematic collection and analysis of website privacy policies, looking for trends and deficiencies both in the wording and content of these policies both for a given organization as well as across different sectors. This analysis can in turn help organizations improve their policies, help regulators assess policies, and inform ongoing public policy debates.

our first newsletter!

With the project now in its third year, this newsletter is intended to highlight some of our progress and activities over the past year. For a more complete list of publications and activities, we encourage you to visit our project's website and subscribe to our mailing list. Our goal is to grow our project into a broader community of organizations and individuals interested in collaborating in this area. Drop me a line if you would like to discuss how you might be able to get involved.

Norman Sadeh
Lead Principal Investigator
Professor
Carnegie Mellon University
sadeh@cs.cmu.edu

www.usableprivacy.org

Join our mailing list!
tinyurl.com/uppp-mail

project team

principal investigators
Norman Sadeh (CMU) - lead
Alessandro Acquisti (CMU)
Travis Breaux (CMU)
Lorrie Faith Cranor (CMU)
Joel Reidenberg (Fordham)
Barbara van Schewick (Stanford)
Noah Smith (U. Washington)

affiliated faculty, post-docs & senior researchers
Eduard Hovy (CMU)
Alessandro Oltramari (CMU)
Pedro Giovanni Leon (Stanford)
N. Cameron Russell (Fordham)
Mads Schaarup Andersen (CMU)
Florian Schaub (CMU)
Shomir Wilson (CMU)

research staff
Linda Moreci (CMU)

graduate students
Hazim Almuhimedi (CMU)
Jaspreet Bhatia (CMU)
Aswarth Dara (CMU)
Harishma Dayanidhi (CMU)
Vlad Herta (Fordham)
Bin Liu (CMU)
Frederick Liu (CMU)
Thomas B. Norton (Fordham)
Dhivya Piraviperumal (CMU)
Ashwini Rao (CMU)
Kanthashree Sathyendra (CMU)
Sebastian Zimmeck (Columbia)
Ziqi Wang (CMU)

undergraduate students
Sushain Cherivirala (CMU)
Roger Iyengar (WUStL)
William H. Matchen (CMU)

This project is funded by the National Science Foundation's Secure & Trustworthy Cyberspace initiative under a Frontier grant.

Carnegie Mellon University

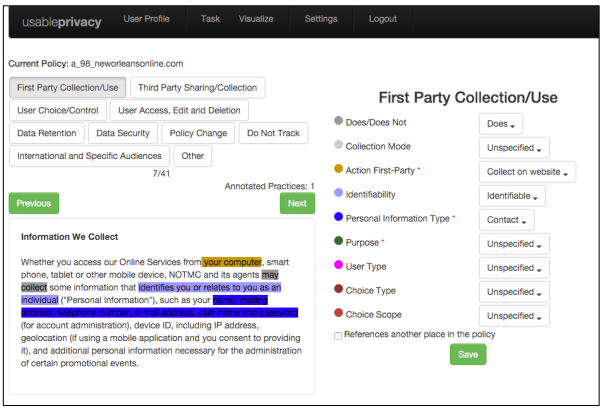
CLIP Center on Law and Information Policy AT FORDHAM LAW SCHOOL

CIS The Center for Internet and Society

NSF

corpus of annotated privacy policies

In July 2016 we will release our OPP-115 Corpus. The corpus covers 115 privacy policies and includes a total of 23,000 fine-grained data practice annotations extracted from these policies. The corpus is the product of hundreds of hours of annotation work by law students at Fordham University and the University of Pittsburgh, using an annotation tool developed at CMU. This corpus already provides a basis for our effort to analyze privacy policies and to scale up our annotation process using machine learning. Our ultimate goal is to eventually cover a substantial subset of the privacy policies on the Web.



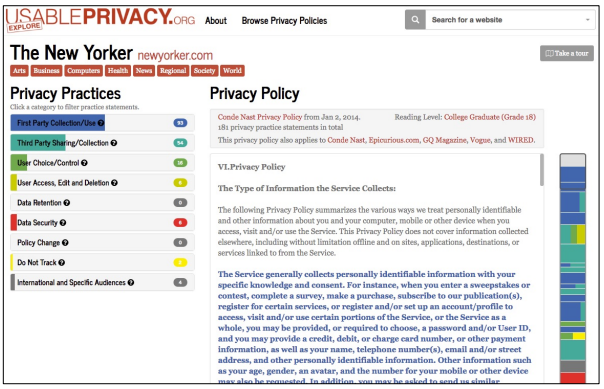
Learn more from our publications and get the corpus at:
www.usableprivacy.org/data

Wilson et al. **The Creation and Analysis of a Website Privacy Policy Corpus**. Proceedings ACL 2016

Wilson et al. **Demystifying Privacy Policies Using Language Technologies: Progress and Challenges**. Proceedings TA-COS '16

explore.usableprivacy.org

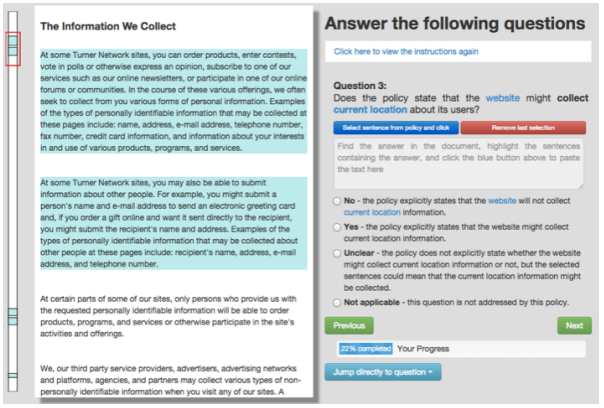
You can also interactively explore the OPP-115 corpus using our explore.usableprivacy.org website, which we launched earlier this year. This includes browsing policy annotations based on categories of websites, as well as visualizing and comparing the composition of individual policies. You can selectively visualize different types of annotations organized by the types of data practices they disclose (e.g. first party collection/use, third party collection/sharing, user choice/control, user access/editing/deletion, Do Not Track and much more). The site also includes a number of other relevant statistics such as readability levels and much more.



Visit **explore.usableprivacy.org**

privacy policy analysis with crowdsourcing

Crowdsourcing offers the potential to scale the analysis of privacy policies with micro-tasks aimed at labeling fragments of privacy policies and/or extracting different types of data practice statements. Our research results suggest that with careful task design many useful data practices can be reliably extracted by crowdworkers. Machine learning and natural language processing techniques can help improve the productivity of crowdworkers and reliability of results.



Some of our recent publications in this area include:

Wilson et al. **Crowdsourcing Annotations for Websites' Privacy Policies: Can It Really Work?** Proceedings WWW 2016

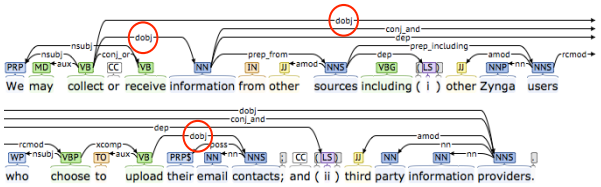
Schaub et al. **Crowdsourcing Privacy Policy Analysis: Potential, Challenges and Best Practices**. it – Information Technology 2016

Bhatia et al. **Mining Privacy Goals from Privacy Policies using Hybridized Task Recomposition**, ACM TOSEM, 2016

Reidenberg et al. **Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding.** BTLJ, 2015

Norton. **Crowdsourcing Privacy Policy Interpretation.** Proceedings TPRC 2015

automating policy annotation and analysis



An important aspect of our research aims to leverage natural language processing and machine learning to automate the analysis of privacy policies. A significant challenge in this regard stems from the intentional vagueness often found in the text of privacy policies. We are developing techniques for relevance prediction and alignment of paragraphs across privacy policies, methods to automatically label fragments of policy text, methods to automatically extract data practices, as well as approaches to analyze and quantify the vagueness of privacy policies.

Some of our recent publications in this area include:

Bhatia et al. **A Theory of Vagueness and Privacy Risk Perception.** Proceedings IEEE RE 2016

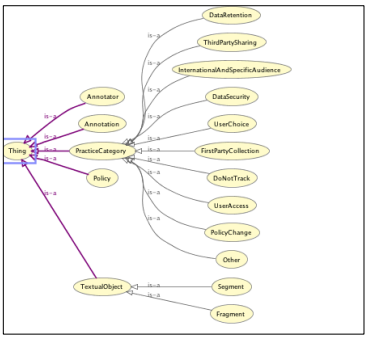
Wilson et al. **The Creation and Analysis of a Website Privacy Policy Corpus**. Proceedings ACL 2016

Bhatia et al. **Mining Privacy Goals from Privacy Policies using Hybridized Task Recomposition**. ACM TOSEM, 2016

Wilson et al. **Demystifying Privacy Policies Using Language Technologies: Progress and Challenges**. Proceedings TA-COS '16

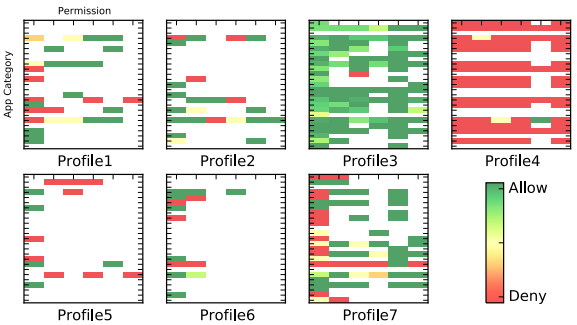
Reidenberg et al. **Automated Comparisons of Ambiguity in Privacy Policies and the Impact of Regulation.** J Legal Studies, 2016

semantic privacy policy analysis



Once data practices have been extracted from privacy policies, the natural next step is to use formal models to represent these annotations and analyze them. Are the policies consistent? How do policies in one sector compare to policies in another sector? Are the policies compliant with relevant regulations? Are the privacy policy's statements consistent with what the website or app actually does? To answer these and related questions, we have developed PrivOnto, a semantic framework for the analysis of privacy policies. PrivOnto relies on semantic web technologies to represent ontologies of concepts found in our privacy policy annotations and to support inference. An initial set of SPARQL queries has been developed to analyze the annotations extracted from a given policy and also support the comparison of statements made across multiple policies.

modeling privacy preferences & expectations



Yet another area of activity is the development of privacy notices that meet users' needs. This requires gaining a better understanding of people's privacy concerns, preferences, and expectations around the data practices of websites, mobile apps and services. We aim to identify data practices that are most critical to informing people's decisions when it comes to interacting with different websites and services, such as practices that are unexpected or surprising to a majority of people. We are also exploring the potential benefits of personalizing privacy notices and decision support based on profiles of individuals with similar privacy preferences or concerns.

Some of our recent publications in this area include:

Rao et al. **Expecting the Unexpected: Understanding Mismatched Privacy Expectations Online.** Proceedings SOUPS 2016

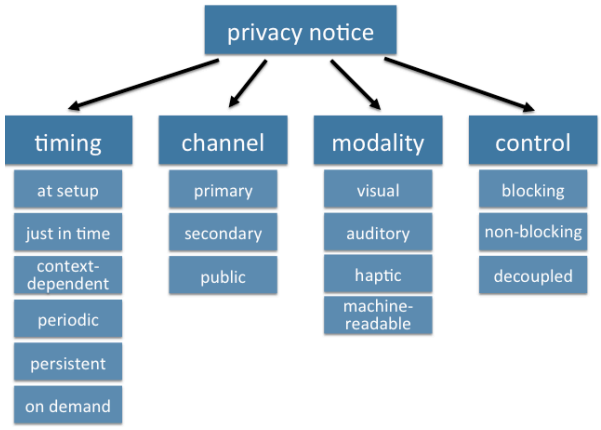
Liu et al. **Follow My Recommendations: A Personalized Assistant for Mobile App Permissions**. Proceedings SOUPS 2016

Leon et al. **Privacy and Behavioral Advertising: Towards Meeting Users' Preferences.** SOUPS PPS Workshop 2015

Reidenberg et al. **Privacy Harms and the Effectiveness of the Notice and Choice Framework.** I/S J. L. & P. Info. Soc., 2014

privacy notice design

Another significant set of activities revolves around the research, design, and development of more effective privacy notices. Our goal is to provide users with relevant and actionable information that enables them to make informed privacy decisions. Our research has focused on notice content and presentation, privacy indicators, and "privacy nudges." As part of this work, we have started to organize the design space of privacy notices, identifying a number of key dimensions that can be used to more systematically evaluate the effectiveness of different notice designs and privacy controls in different contexts.



Some of our recent publications in this area include:

Gluck et al. **How Short is Too Short? Implications of Length and Framing on the Effectiveness of Privacy Notices**. SOUPS 2016

Grannis. **Elements of Effective Notice in the Online Age.** Fordham Urban Law Journal, 2016

Reidenberg et al. **Rating Indicator Criteria for Privacy Policies.** SOUPS Workshop on Privacy Indicators 2016

Schaub et al. **A Design Space for Effective Privacy Notices.** Proceedings SOUPS 2015

Almuhimedi et al. **Your Location has been Shared 5,398 Times!: A Field Study on Mobile App Privacy Nudging.** CHI 2015

selected upcoming and recent events

We are committed to broadly disseminating our research across academia, industry and public policy circles, to identifying organizations interested in collaborating with us, and to more generally helping foster a vibrant research community in this area. Here are some recent and upcoming activities involving our project.

AAAI Symposium on Privacy and Language Technologies | November 2016

Shomir Wilson, Alessandro Oltramari and Fei Liu organize the AAAI Fall Symposium on Privacy and Language Technologies, to be held Nov. 17-19 in Arlington, Virginia. Call for papers: **sites.google.com/site/fsplt2016**

SOUPS Workshops on Privacy Indicators | June 2016

Florian Schaub co-organized the Workshop on Privacy Indicators and the Workshop on Future Privacy Indicators at the Symposium on Usable Privacy and Security in June 2016. More information: **www.usenix.org/conference/soups2016/workshop-on-privacy-indicators**

FTC PrivacyCon | January 2016

Ashwini Rao, Norman Sadeh and Florian Schaub each presented research results coming out of our project at the Federal Trade Commission's PrivacyCon on January 14, 2016. More information: **www.ftc.gov/news-events/events-calendar/2016/01/privacycon**

Data Privacy Day | January 2016

On January 28, Lorrie Cranor and Norman Sadeh organized CMU's celebration of Data Privacy Day. Ed Felten, Deputy US Chief Technology Officer, gave the event's keynote address. The event also featured presentations of our research. More information: **cups.cs.cmu.edu/privacy-day/2016/**

SOUPS Workshop on Privacy Personas & Segmentation | June 2015

Alessandro Acquisti and Norman Sadeh co-organized the Privacy Personas and Segmentation Workshop at SOUPS 2015. More information: **cups.cs.cmu.edu/soups/2015/pps.php**

CLIP Law and Information Society Symposium | May 2015

The Fordham Center on Law and Information Privacy organized the Ninth Law and Information Society Symposium on May 13, 2015 under the theme "Solving Privacy Around the World." More information: **www.fordham.edu/info/23833/past_conferences/5605/**