Lecture 12: (i) Anonymization; (ii) Privacy

CMSC 25900 / DATA 25900 Spring 2021 The University of Chicago



"...a tension that shakes a foundational belief about data privacy: Data can be either useful or perfectly anonymous but never both." – Paul Ohm

Historical Conceptualizations of Anonymization and Personal Data

Personally Identifiable Information (PII)

- Also termed "personal data"
- 2010 NIST Special Publication 800-122 Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)
- General Data Protection Regulation (GDPR)

NIST 800-122 Definitions

"PII is —any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information."

NIST 800-122 PII Examples

- Name, such as full name, maiden name, mother's maiden name, or alias
- **Personal identification number**, such as social security number (SSN), passport number, driver's license number, taxpayer identification number, patient identification number, and financial account or credit card number
- Address information, such as street address or email address
- Asset information, such as Internet Protocol (IP) or Media Access Control (MAC) address or other hostspecific persistent static identifier that consistently links to a particular person or small, well-defined group of people
- **Telephone numbers**, including mobile, business, and personal numbers
- Personal characteristics, including photographic image (especially of face or other distinguishing characteristic), x-rays, fingerprints, or other biometric image or template data (e.g., retina scan, voice signature, facial geometry)
- Information identifying personally owned property, such as vehicle registration number or title number and related information
- Information about an individual that is linked or linkable to one of the above (e.g., date of birth, place of birth, race, religion, weight, activities, geographical indicators, employment information, medical information, education information, financial information).

NIST 800-122 Definitions

- "To distinguish an individual is to identify an individual. Some examples of information that could identify an individual include, but are not limited to, name, passport number, social security number, or biometric data. In contrast, a list containing only credit scores without any additional information concerning the individuals to whom they relate does not provide sufficient information to distinguish a specific individual."
- "To trace an individual is to process sufficient information to make a determination about a specific aspect of an individual's activities or status. For example, an audit log containing records of user actions could be used to trace an individual's activities."

NIST 800-122 Definitions

 Linked information is information about or related to an individual that is logically associated with other information about the individual. In contrast, linkable information is information about or related to an individual for which there is a possibility of logical association with other information about the individual. For example, if two databases contain different PII elements, then someone with access to both databases may be able to link the information from the two databases and identify individuals, as well as access additional information about or relating to the individuals. If the secondary information source is present on the same system or a closely-related system and does not have security controls that effectively segregate the information sources, then the data is considered linked. If the secondary information source is maintained more remotely, such as in an unrelated system within the organization, available in public records, or otherwise readily obtainable (e.g., internet search engine), then the data is considered linkable.

GDPR Definitions (Article 4)

 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

GDPR Definitions (Article 4)

- 'processing' means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;
- 'restriction of processing' means the marking of stored personal data with the aim of limiting their processing in the future;

GDPR Definitions (Article 4)

 'pseudonymisation' means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person;

Example from UChicago IRB

 4. * Do you anticipate that the research data will be transferred or transported at any time? Yes No Clear
 5. * Do you plan to store data on a server or cloud service? Yes O No <u>Clear</u>
a. * Which server or cloud service do you plan to use?
UChicago Box
6. * Will you collect any identifiers from the research participants (including names, addresses, Social Security Numbers, email and phone contact information etc.)?
O Yes No <u>Clear</u>
7. * What identifying information about research participants will be linked to the data?
Data/specimens will be directly labeled with personal identifying information
Data/specimens will be labeled with a code that the research team can link to personal identifying information through a crosswalk to the coding system
Data/specimens will be labeled with a code but the research team will not have access to the crosswalk that connects the codes to participant identifiers
Data/specimens will not be labeled with any identifying information and a coding system will not be used
Other
8. If you will be using a coding system, who will have access to the crosswalk that links participant identifiers to the data/specimens and where will you store the crosswalk?
Not applicable.

Models of Data-Release Stewardship

Scope of Releasing Data

- Release to third parties
- Release to the public
- Release to others within your organization

Models of Data-Release Stewardship

- (Note that we just made up most terms on this page)
- A Release-and-Forget Model: Try to remove PII and otherwise "deidentify" data, but then provide unrestricted access (e.g., through publicly posting a dataset)
- A Release-Under-Conditions Model: Try to remove PII and otherwise "deidentify" data, but then provide restricted access to them (e.g., through data processing covered under contractual obligations and an approval process) and sometimes conditions upon the processing or the release of aggregate data
- A Managed-Processing Model: The data steward never releases the data, but will run computation for others and provide aggregate answers

General Techniques for Anonymization

Nonanonymized Data

Name	Race	Birth Date	Sex	ZIP Code	Complaint
Sean	Black	9/20/1965	Male	02141	Short of breath
Daniel	Black	2/14/1965	Male	02141	Chest pain
Kate	Black	10/23/1965	Female	02138	Painful eye
Marion	Black	8/24/1965	Female	02138	Wheezing
Helen	Black	11/7/1964	Female	02138	Aching joints
Reese	Black	12/1/1964	Female	02138	Chest pain
Forest	White	10/23/1964	Male	02138	Short of breath
Hilary	White	3/15/1965	Female	02139	Hypertension
Philip	White	8/13/1964	Male	02139	Aching joints
Jamie	White	5/5/1964	Male	02139	Fever
Sean	White	2/13/1967	Male	02138	Vomiting
Adrien	White	3/21/1967	Male	02138	Back pain

TABLE 1: Original (Nonanonymized) Data

Suppressing Data

• Suppression: Deleting or omitting data

 TABLE 2: Suppressing Four Identifier Fields

Race	Complaint
Black	Short of breath
Black	Chest pain
Black	Painful eye
Black	Wheezing
Black	Aching joints
Black	Chest pain
White	Short of breath
White	Hypertension
White	Aching joints
White	Fever
White	Vomiting
White	Back pain

Generalizing Data

• Generalization: Re-code data to be less granular

Race	Birth Year	Sex	ZIP Code*	Complaint
Black	1965	Male	021*	Short of breath
Black	1965	Male	021*	Chest pain
Black	1965	Female	021*	Painful eye
Black	1965	Female	021*	Wheezing
Black	1964	Female	021*	Aching joints
Black	1964	Female	021*	Chest pain
White	1964	Male	021*	Short of breath
White	1965	Female	021*	Hypertension
White	1964	Male	021*	Aching joints
White	1964	Male	021*	Fever
White	1967	Male	021*	Vomiting
White	1967	Male	021*	Back pain

TABLE 3: Generalized

From Paul Ohm. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. UCLA Law Review Vol. 57, p. 1701, 2010.

Aggregating Data

• Aggregation: Release summary data rather than raw data

 TABLE 4: Aggregate Statistic

Men Short of Breath 2

[Broken] K-anonymity

 "A release of data is said to have the k-anonymity property if the information for each person contained in the release cannot be distinguished from at least k – 1 individuals whose information also appear in the release."

[Broken] K-anonymity

Name	Age	Gender	State of domicile	Religion	Disease
Ramsha	30	Female	Tamil Nadu	Hindu	Cancer
Yadu	24	Female	Kerala	Hindu	Viral infection
Salima	28	Female	Tamil Nadu	Muslim	ТВ
Sunny	27	Male	Karnataka	Parsi	No illness
Joan	24	Female	Kerala	Christian	Heart-related
Bahuksana	23	Male	Karnataka	Buddhist	ТВ
Rambha	19	Male	Kerala	Hindu	Cancer
Kishor	29	Male	Karnataka	Hindu	Heart-related
Johnson	17	Male	Kerala	Christian	Heart-related
John	19	Male	Kerala	Christian	Viral infection

[Broken] K-anonymity

Name	Age	Gender	State of domicile	Religion	Disease
*	20 < Age ≤ 30	Female	Tamil Nadu	*	Cancer
*	20 < Age ≤ 30	Female	Kerala	*	Viral infection
*	20 < Age ≤ 30	Female	Tamil Nadu	*	ТВ
*	20 < Age ≤ 30	Male	Karnataka	*	No illness
*	20 < Age ≤ 30	Female	Kerala	*	Heart-related
*	20 < Age ≤ 30	Male	Karnataka	*	ТВ
*	Age ≤ 20	Male	Kerala	*	Cancer
*	20 < Age ≤ 30	Male	Karnataka	*	Heart-related
*	Age ≤ 20	Male	Kerala	*	Heart-related
*	Age ≤ 20	Male	Kerala	*	Viral infection

This data has 2-anonymity with respect to the attributes 'Age', 'Gender' and 'State of domicile' since for any combination of these attributes found in any row of the table there are always at least 2 rows with those exact attributes. The attributes available to an adversary are called quasi-identifiers. Each quasi-identifier tuple occurs in at least *k* records for a dataset with *k*-anonymity.^[14]

How Do You Find Personal Data?

- Example: Google's Cloud Data Loss Prevention (DLP) API
 - <u>https://cloud.google.com/dlp/docs/infotypes-reference</u>
- Uses heuristics to try to identify particular information types

InfoType	Description
ADVERTISING_ID	Identifiers used by developers to track users for <i>advertising purposes</i> . These include Google Play Advertising IDs, Amazon Advertising IDs, Apple's identifierForAdvertising (IDFA), and Apple's identifierForVendor (IDFV).
AGE	An age measured in months or years.
CREDIT_CARD_NUMBER	A credit card number is 12 to 19 digits long. They are used for payment transactions globally.
CREDIT_CARD_TRACK_NUMBER	A credit card track number is a variable length alphanumeric string. It is used to store key cardholder information.
DATE	A <i>date</i> . This infoType includes most date formats, including the names of common world holidays. Note: Not recommended for use during latency sensitive operations.
DATE_OF_BIRTH	A date of birth. Note: Not recommended for use during latency sensitive operations.
DOMAIN_NAME	A domain name as defined by the DNS standard.
EMAIL_ADDRESS	An <i>email address</i> identifies the mailbox that emails are sent to or from. The maximum length of the domain name is 255 characters, and the maximum length of the local-part is 64 characters.
ETHNIC_GROUP	A person's ethnic group.

MALE_NAME	A common <i>male name.</i> Note: Not recommended for use during latency sensitive operations.
MEDICAL_TERM	Terms that commonly refer to a person's <i>medical condition or health</i> . Note: Not recommended for use during latency sensitive operations.
ORGANIZATION_NAME	A name of a <i>chain store, business or organization.</i> Note: Not recommended for use during latency sensitive operations.
PASSPORT	A <i>passport number</i> that matches passport numbers for the following countries: Australia, Canada, China, France, Germany, Japan, Korea, Mexico, The Netherlands, Poland, Singapore, Spain, Sweden, Taiwan, United Kingdom, and the United States.
PERSON_NAME	A full <i>person name</i> , which can include first names, middle names or initials, and last names. Note: Not recommended for use during latency sensitive operations.
PHONE_NUMBER	A telephone number.
STREET_ADDRESS	A street address. Note: Not recommended for use during latency sensitive operations.
SWIFT_CODE	A SWIFT code is the same as a Bank Identifier Code (BIC). It's a unique identification code for a particular bank. These codes are used when transferring money between banks, particularly for international wire transfers. Banks also use the codes for exchanging other messages.
TIME	A timestamp of a specific time of day.
URL	A Uniform Resource Locator (URL).

Credentials and secrets

The infoType detectors in this section detect credentials and other secret data.

InfoType	Description
AUTH_TOKEN	An <i>authentication token</i> is a machine-readable way of determining whether a particular request has been authorized for a user. This detector currently identifies tokens that comply with OAuth or Bearer authentication.
AWS_CREDENTIALS	Amazon Web Services account access keys.
AZURE_AUTH_TOKEN	Microsoft Azure certificate credentials for application authentication.
BASIC_AUTH_HEADER	A <i>basic authentication header</i> is an HTTP header used to identify a user to a server. It is part of the HTTP specification in RFC 1945, section 11.
ENCRYPTION_KEY	An encryption key within configuration, code, or log text.
GCP_API_KEY	Google Cloud API key. An encrypted string that is used when calling Google Cloud APIs that don't need to access private user data.
GCP_CREDENTIALS	Google Cloud service account credentials. Credentials that can be used to authenticate with Google API client libraries and service accounts.
JSON_WEB_TOKEN	JSON Web Token. JSON Web Token in compact form. Represents a set of claims as a JSON object that is digitally signed using JSON Web Signature.
HTTP_COOKIE	An <i>HTTP cookie</i> is a standard way of storing data on a per website basis. This detector will find headers containing these cookies.
PASSWORD	Clear text <i>passwords</i> in configs, code, and other text.

US_PASSPORT		A United States passport number.
US_PREPARER_TAXPAYER_IDENTIF	ICATION_NUMBER	A United States Preparer Taxpayer Identification Number (PTIN) is an identification number that all paid tax return preparers must use on US federal tax returns or claims for refund submitted to the US Internal Revenue Service (IRS).
US_SOCIAL_SECURITY_NUMBER		A United States Social Security number (SSN) is a 9-digit number issued to US citizens, permanent residents, and temporary residents. This detector will not match against numbers with all zeroes in any digit group (that is, 000-##-####, ###-00-####, or ###-##-0000), against numbers with 666 in the first digit group, or against numbers whose first digit is 9.
US_STATE		A United States state name.
US_TOLLFREE_PHONE_NUMBER		A US toll-free telephone number.
US_VEHICLE_IDENTIFICATION_NUMBER		A vehicle identification number (VIN) is a unique 17-digit code assigned to every on-road motor vehicle.
Uruguay		
InfoType Descriptio	'n	
URUGUAY_CDI_NUMBER A Uruguay	an Cédula de Identidad	(CDI), or identity card, is used as the main identity document for citizens.

Important: Built-in infoType detectors are not a 100% accurate detection method. For example, they can't guarantee compliance with regulatory requirements. You must decide what data is sensitive and how to best protect it. Google recommends that you test your settings to make sure your configuration meets your requirements.

Global

InfoType	Description
ADVERTISING_ID	Identifiers used by developers to track users for <i>advertising purposes</i> . These include Google Play Advertising IDs, Amazon Advertising IDs, Apple's identifierForAdvertising (IDFA), and Apple's identifierForVendor (IDFV).

Can You Screw Up Data Suppression?

• Yes!

Case Study 1: ZIP Code, DOB, Sex

Massachusetts Health Data

- Mid 1990s: Group Insurance Commission (GIC)
- Upon request, GIC will release records with 100 attributes for every state employee's hospital visits
- Latanya Sweeney, "Uniqueness of Simple Demographics in the U.S. Population":
 - 87%: ZIP code + full Date of Birth + Sex is uniquely identifying
 - 53%: *City* + full Date of Birth + Sex is uniquely identifying
 - 18%: County + full Date of Birth + Sex is uniquely identifying
- William Weld (Governor of Massachusetts) deanonymized when Sweeney purchased voter rolls from the city of Cambridge
 - Sweeney sent the governor's records (diagnoses/prescriptions) to him



Case Study 2: AOL Search Data

AOL Search Data Release

- AOL Research released 20,000,000 search queries for 650,000 users of AOL's search engine (3 months)
- Suppressed AOL username and IP address
 - Replaced them with unique, pseudonymous identifiers



AOL Search Data Release (Aftermath)

The New York Times A Face Is Exposed for AOL Searcher No. 4417749

By Michael Barbaro and Tom Zeller Jr.

Aug. 9, 2006

f y 🛛 🄶 🗌

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

AOL Search Data Release (Aftermath)

• "...User 4417749's identity in queries such as "landscapers in Lilburn, Ga,' several people with the last name Arnold and 'homes sold in shadow lake subdivision gwinnett county georgia." They quickly tracked down Thelma Arnold, a sixty-two-year-old widow from Lilburn, Georgia who acknowledged that she had authored the searches, including some mildly embarrassing queries such as "numb fingers," "60 single men," and "dog that urinates on everything."



Case Study 3: Netflix Prize

- Netflix released 100,000,000 records from 500,000 users
 - December 1999 to December 2005
 - Assigned a unique pseudonymous identifier to each user
- Each record included the pseudonymous identifier, the movie watched, the rating (1-5 stars), and rating's date



- Narayanan and Shmatikov correlated with IMDb
- Ratings on IMDb are public
- Databases are not perfect subsets of each other
- What can be leaked from knowing which movies an identified user watched?



$$\mathsf{Sim}(r_1, r_2) = \frac{\sum \mathsf{Sim}(r_{1i}, r_{2i})}{|\mathsf{supp}(r_1) \cup \mathsf{supp}(r_2)|}$$

Definition 3 (De-anonymization) An arbitrary subset \hat{D} of a database D can be (θ, ω) -deanonymized w.r.t. auxiliary information Aux if there exists an algorithm A which, on inputs \hat{D} and Aux(r) where $r \leftarrow D$

- If $r \in \hat{D}$, outputs r' s.t. $\Pr[Sim(r, r') \ge \theta] \ge \omega$
- *if* $r \notin \hat{D}$ *, outputs* \perp *with probability at least* ω

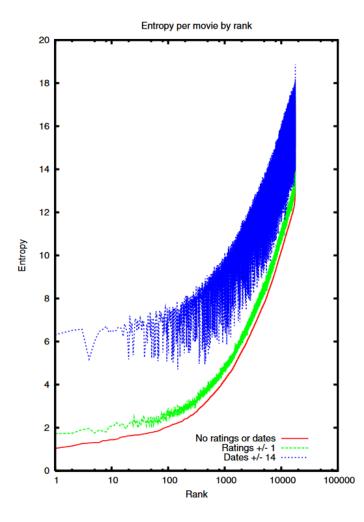


Figure 7. Entropy of movie by rank

Recap

The Surprising Success of Deanonymization

• The use of auxiliary information

- Extremely hard to control
- Errors suppressing data
- Personal data showing up in unexpected places
- It's hard to reason about what is/is not identifiable
- Thinking only about personal data / PII is not sufficient

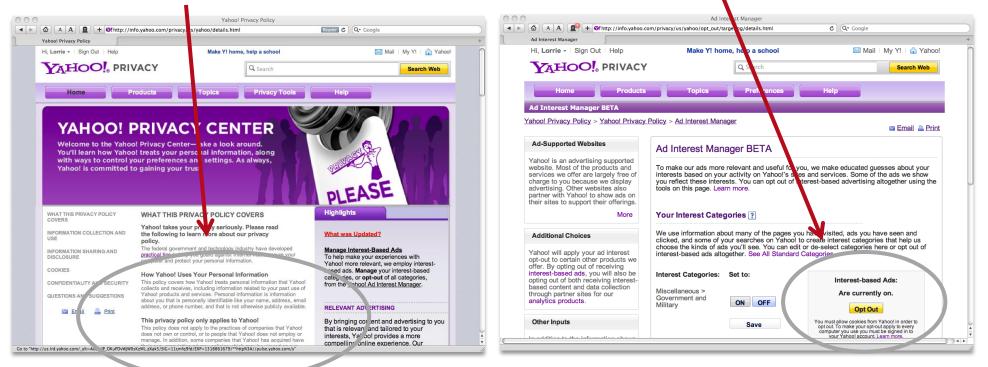
Privacy Communication

Notice and choice

Protect privacy by giving people control over their information

Notice about data collection and use

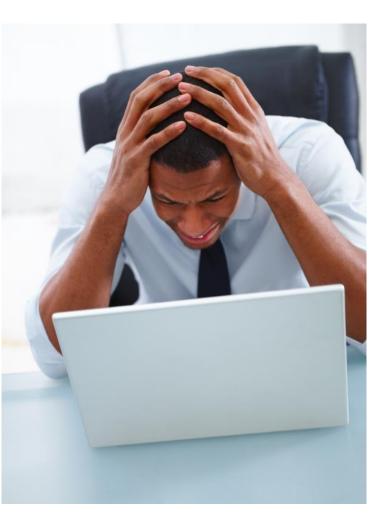
Choices about allowing their data to be collected and used in that way



Nobody wants to read privacy policies

"the notice-and-choice model, as implemented, has led to long, incomprehensible privacy policies that consumers typically do not read, let alone understand"

 Protecting Consumer Privacy in an Era of Rapid Change. Preliminary FTC Staff Report. December 2010.



- until it is no longer needed for business, tax or legal purposes. • We will keep your information safe using encryption or other appropriate security controls.

- We don't sell your Personal Information to anyone for any purpose. Period. • We keep your Personal Information in our business records while you are a customer, or

Our privacy commitments

Back to Top

your personal information.

Because some apps, including some AT&T and DTV information, or use information in different ways, they may hav and/or terms and conditions. These apps may also offer you a Back In Top

Ef

A

	<text><section-header><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></section-header></text>	<text><section-header><section-header><section-header><section-header><section-header><section-header><section-header></section-header></section-header></section-header></section-header></section-header></section-header></section-header></text>	<section-header><section-header><section-header><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></section-header></section-header></section-header>	<text><text><text><text><text><text><text><text><text><text></text></text></text></text></text></text></text></text></text></text>	<text><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></text>	<text><text><text><text><text><section-header><section-header><section-header><section-header><section-header><section-header></section-header></section-header></section-header></section-header></section-header></section-header></text></text></text></text></text>	<text><text><text><text><text><text><text><text><text><list-item><list-item><list-item></list-item></list-item></list-item></text></text></text></text></text></text></text></text></text>
About Our Privacy Policy	extension while program the second se	A stress stressmith A stress stressmith and the stress stressmith and the stress	product is using a (30)-C11 region single dimension and a group on a single dimension of the single dimension of the single dimension of the single dimension of the single dimension. The single dimension of the single dime	He may use the provide access the factories where an of the provide memory and the provide	E est d'architecture seu est avante avant to 41 Test. 14 E est d'architecture de la constante	We derive an an provint indelte in Virange stellung in der der ung VII werden der andere eine Stellung in der Stellung in der Stellung in der stellung in der andere eine Stellung in der Stellung in der Stellung in der stellung in der Stellung in der Stellung in der Stellung in der Stellung vorlichten der Stellung in der Stellung in der Stellung in der Stellung in der Stellung in der Stellung in der Stellun	(13) Such the strength, fairs ware exterime an encode 32 sugmets water that any encode of the strength of t
An Anter Ant	 Model And Landowscence comments¹⁰ Model and Landowscence comments¹⁰ Model and Landowscence and Land	<list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item>		<text><text><text><section-header><section-header><text><list-item></list-item></text></section-header></section-header></text></text></text>	<text><list-item><list-item><list-item><list-item></list-item></list-item></list-item></list-item></text>	<list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item>	<text><text><text><text><text><text><text></text></text></text></text></text></text></text>
(FAQ) format (/sites/privacy_policy/terms). We want as make informed choices about your privacy, and then spend the re products and services.	 Handreich er obergen ein mit gelf in 4 register beständet im sogaren eine Ander Sacht er der und Ander Sacht er der und	Herta's tradeministration Constraints and a second secon	A for such and your while in the other to such as a formation of the other to such as a suc	Manual exercised (and the effective and the effe	Sti filmstaat servergenergenergen het vergenergenergenergenergenergenergenerge	• Can get under detraders andere spotterers surf. • We deter detraders andere spotterers andere spotterers andere spotter. • On andere determine spotterers andere spotterers andere spotter. • On andere determine spotterers andere spotterers andere spotterers. • On andere determine spotterers andere spotterers. • On andere determine spotterers andere spotterers. • On andere determine spotterers.	No rest rais a processingletion to the processingletions (in a second se
Effective July 24, 2015	adoption considering and exclude of the intermediate intermediate of the intermediate	Hand a loader: If light the main an application of the large pulsability of the distribution of the system of t	Notice and a more formation than the result on a sub-product darge spaces and the relative backwards before any particular darge spaces and the result of the result of	Next service may be a contraction or material bits making and dis Allowing there are brack, service stress may be also approximately and approximately and approximately and approximately and approximately and approximately and approximately	Bernstein Witch erfehalt habe dage for a Landaux and the second of th	The second of the law is shown in the law is a law in the second of the law is a law in the second of the law is a law in the second of the law is a law	We say a Parallel Theodore Strengther, Bruchest Machades 19 and 19 an
A Quick Summary of Our	The at physical action and the second county of physical action action of physical action act	For all more and an analysis in the region problem is the short of the section	 Constraint infolder: (If the the profession include its Profession duration of the profession of the profes	 The provide comparison of the state of the s	Anothen manyther are able lossess of the sample straining them studied for advances of space of an space results or advances and the same logsess many time strain and straining and manufactures are as a straining and space of the souther of P base shows and advances are associated and provide any advances and the souther and the souther advances and the souther and P base shows and straining and the souther advances and the souther advances advances advances and the souther advances a	and the signature is consider as the Constanting states the signature double of Collader, the Constanting states the signature constanting states are constanting as a signature constanting as a signature constanting as a signature constanting states are constanting as a signature constanting as a signatu	 Benefit and an effective and the second of the phase and the second of the se
Policy	Information framework diversion-outry frame with plan spaceta point approximate of the spaceta	Northagene Plane is a strategin to approximate assuming? Security of the security of	standard (b) a transform and scale). Programs - remain control control on the standard biologic standard (b) and the standard (b)	411 Secure Measurement Measurement (more a manufacture measurement) and a manufacture of an and an an annual measurement (more a manufacture) and an annual measurement (more a manufacture) and an annual measurement (more a manufacture) and an annual measurement (more a manufacture) and an annual measurement (more a manufacture) and an annual measurement (more a manufacture) and an annual measurement (more a manufacture) and and an annual measurement (more a manufacture) and and an annual measurement (more a manufacture) and and an annual measurement (more a manufacture) and an annual measurement (more a manufacture) and and an annual measurement (more a manufacture) and and an annual measurement (more a manufacture) and an annual measurement (more a manufacture) and an annual measurement (more a manufacture) and and an annual measurement (more a manufacture) and and an annual measurement (more a manufacture) and an annual measurement (more a manufacture) and an annual measurement (more a manufacture) and and an annual measurement (more a manufacture) and an annual measurement (more a manufacture) and an annual measurement (more a ma	Now controllers a uniquelity to the definition of the definition	Buch for Customer Proprietary Network Information (CPNI) Maccore	This can be the SH-SH-SH-SH are not allow and fails into any equity, in . This can be a single series extracted on a field (SH-SH-SH) community on SH-SH-SH-SH-SH-SH-SH-SH-SH-SH-SH-SH-SH-S
Our Privacy Policy applies to your use of all products, services and our AT&T affiliates, such as DIRECTV, unless they have a	 Any operating that can be the layer the Wildow of the difference of the	Negl equipamente (Mu el parte par entre riske di relativo di antico di antic	Control bases for exception proceptions in detending time year factored for testing Subjection of type solitons requests Factors and the exception of the type in the type of	Honors III and Charles and Charles and Charles and Charles Registers are not with a 25 States of the Charles and Charles The States and Charles and Charles People Matt This have a calibration of the State Matter and Charles and State and Charles and Charles and Charles and Charles and Charles and Charles and Charles and Charles and Charl	Electric cap or particles, set spectra para reflect desse tepports and encode test and encode	phone may pair of home to bolismus the Turk "Memoria": A through the date that of a sectoring with any strain pair (b) of the phone may be the training of the and bolismus of the date and the training of the training of the training of the Base to Fig. Memory of the training of the sector of the training of the sector of the training of the sector of the training of the sector of the training of the sector of the training of	percentry bin us and norm in the surger certs. "In ref 174 reg (reg (reg (reg (reg (reg (reg (reg

E-Mail: Dway restaining event on and scheme individues and a mitheli of to she address meanings main to the protect or server type. You do no another to the AFST reprint galaxies have proprietions at contestance.

Best Wessagen, Oct-arty Treasagen,

web-sites that a pergendent of

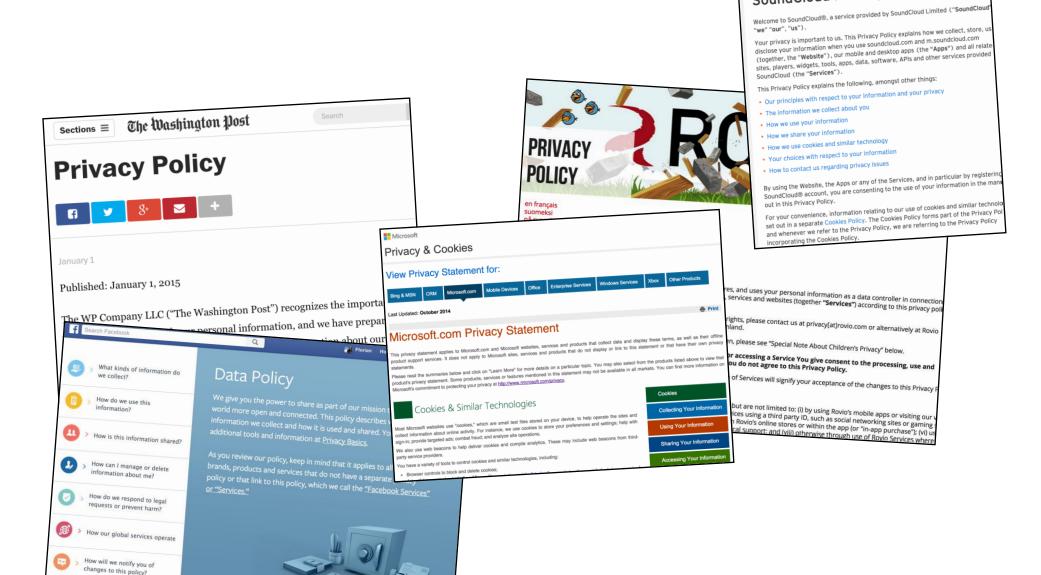
Medical - 1 400 107 4000 Manimes - 1 400 107 4000 Mender Str. - 1400 208 2000 Spender Str. - 1400 208 2000 Spender Language - 1 400 401 4005 Spikelin Fr Spinores - 1 400 401 507 V V

for assistance in other languages, please visit works

Legary #751 Schemer - 1-000-222-0000

Contensury of the following ATAT family of our the following



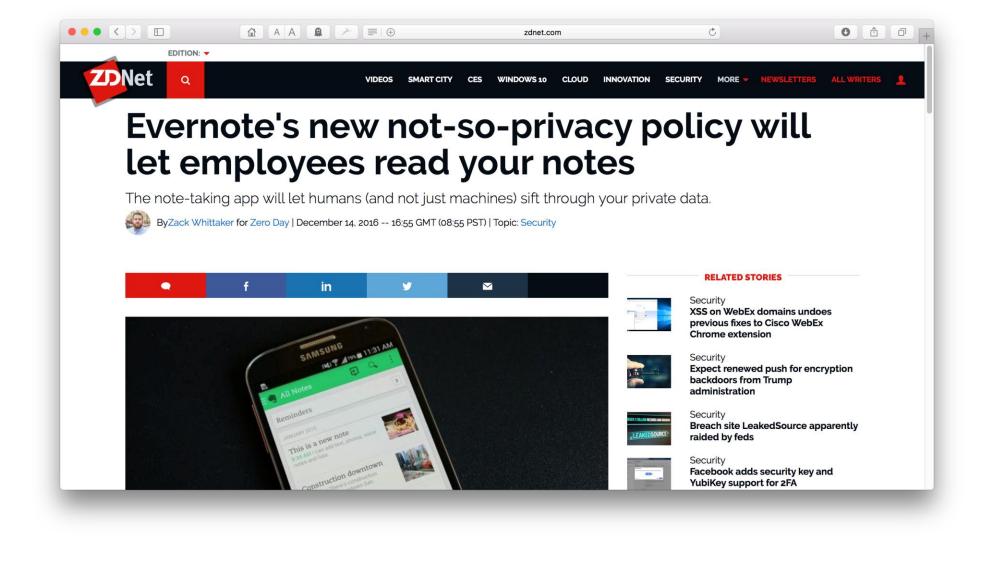


Search for artists, bands, tracks, pod

SOUNDCLOUD

We have updated our Privacy Policy as of 10 June 2015. SoundCloud Privacy Policy

CR Consumer Reports Privacy Policy				Q
Privacy Policy Highlights	-	📩 Like 2 Share	→Tweet G+1 Pinit 🖂 🚍	AA
Privacy Principles	Choice/Opt O			1
Privacy Policy FAQ		uı		
Cookies		il communications (pr	omotional, free email	
Protection of Information	newsletters, and R	- ,		
Surveys	ConsumerReports.org), our fr	ee email newsletters (for example,	nple, Consumer Reports magazine or those about cars, finance, drugs or safe hils about our other products and service	
Choice/Opt Out	ask you about your experience			3 01 10
			s by following the instructions provided a ving only that particular type of email me	
	2		ou can manage your email communication	



"Only in some fantasy world do users actually read these notices and understand their implications before clicking to indicate their consent"

UUUU

REPORT TO THE PRESIDENT BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE

Executive Office of the President President's Council of Advisors on Science and Technology

May 2014

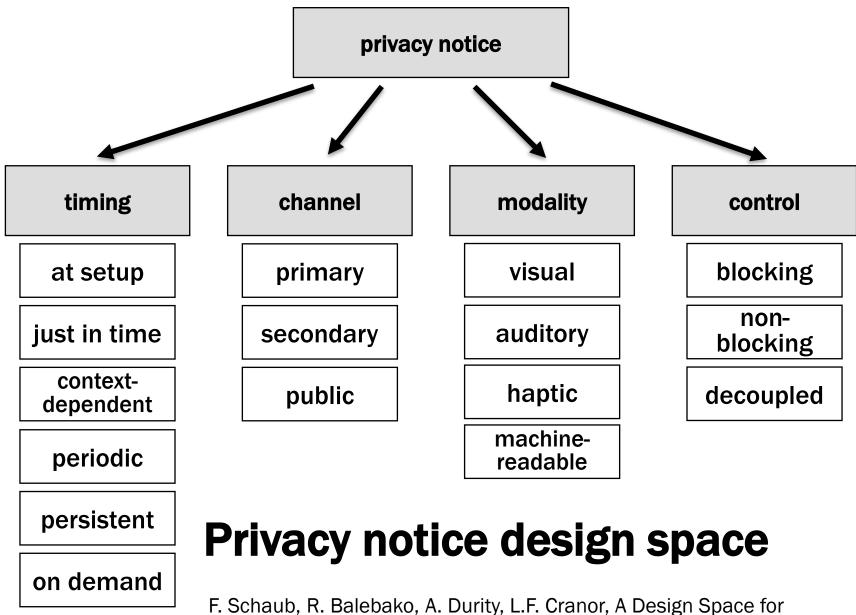


Cost of reading privacy policies

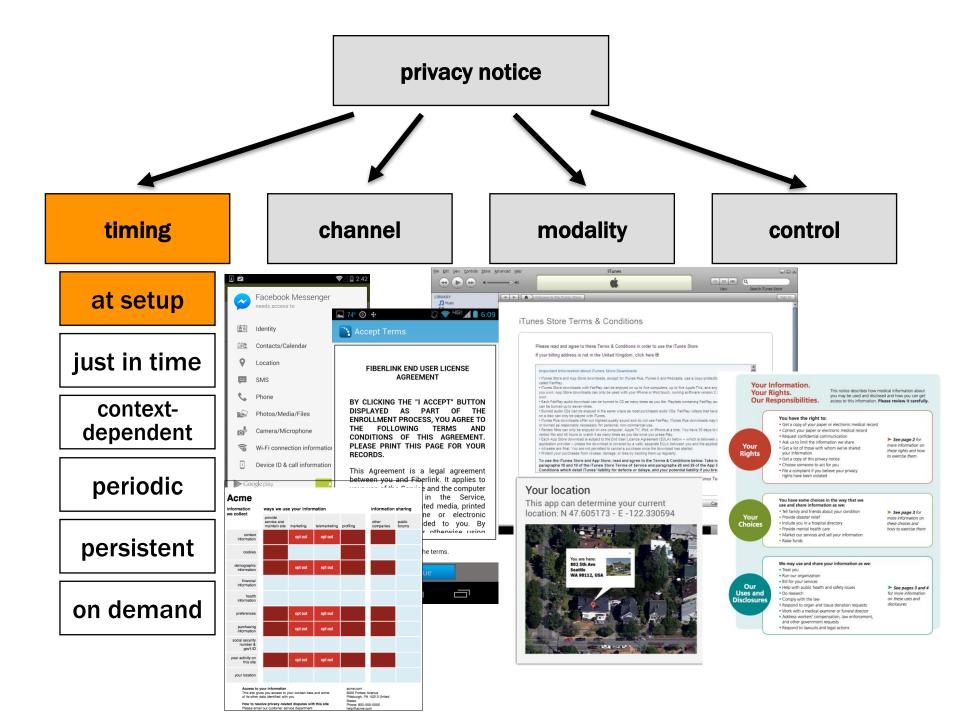
- What would happen if everyone read the privacy policy for each site they visited once per year?
- Time = 244/hours year
- Cost = \$3,534/year
- National opportunity cost for time to read policies: \$781 billion

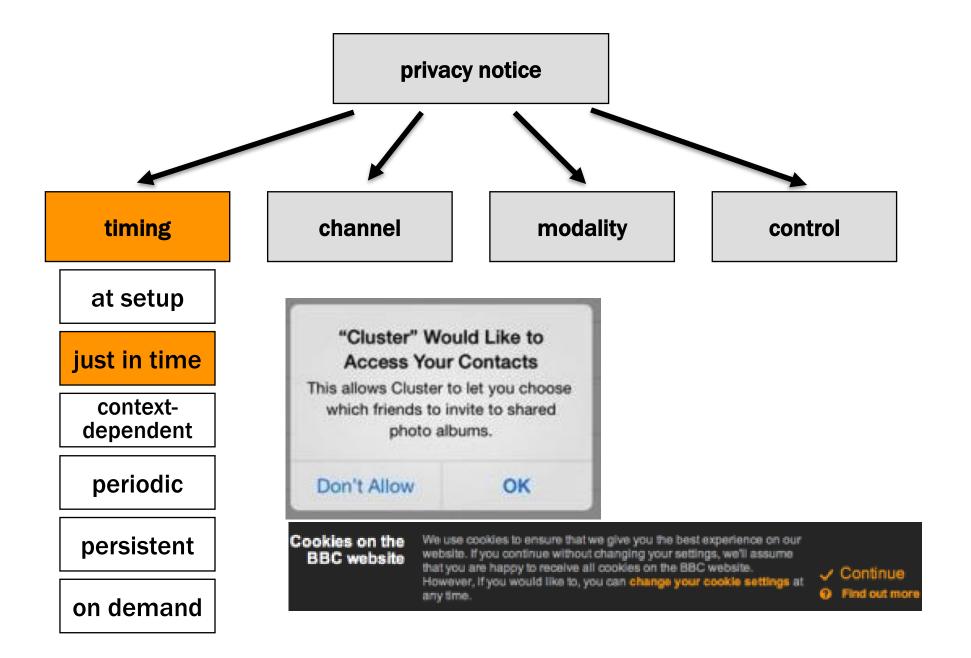


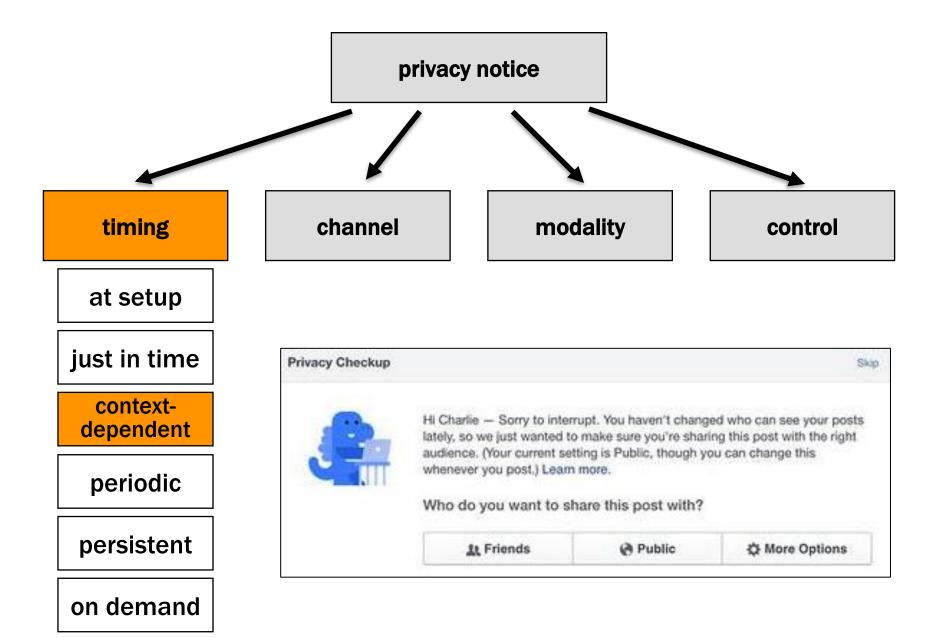
Aleecia McDonald and Lorrie Faith Cranor. The Cost of Reading Privacy Policies. I/S: A Journal of Law and Policy for the Information Society. 2008 Privacy Year in Review Issue. **Privacy Notice Design Space**

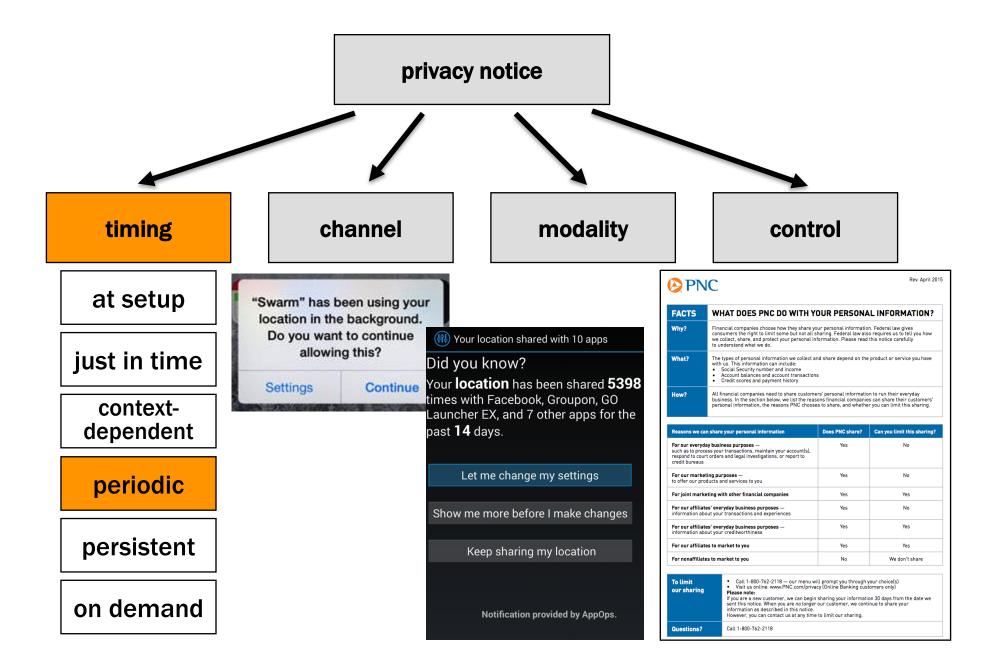


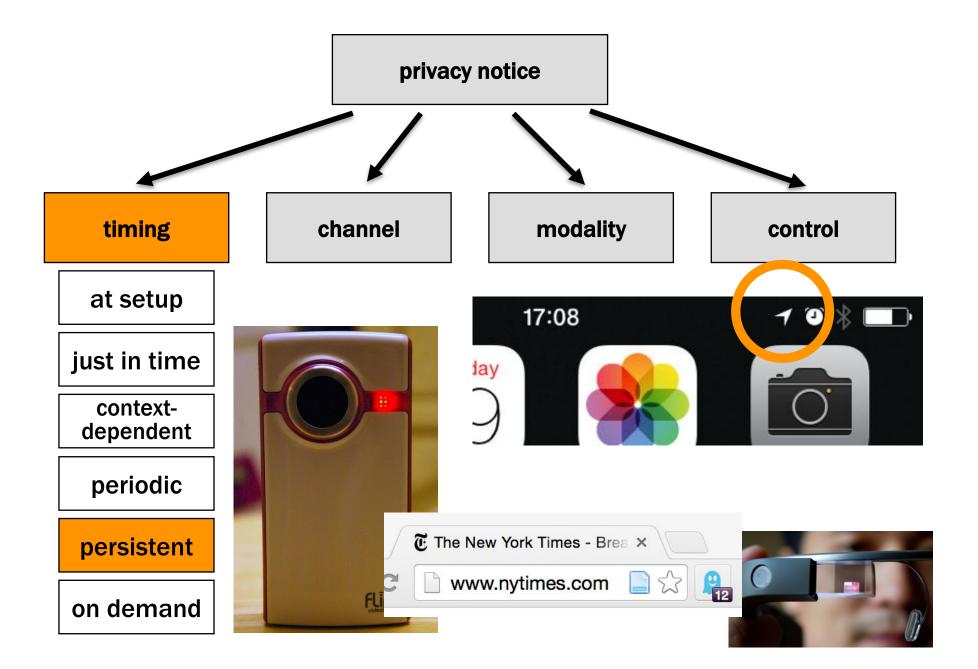
Effective Privacy Notices, SOUPS'15

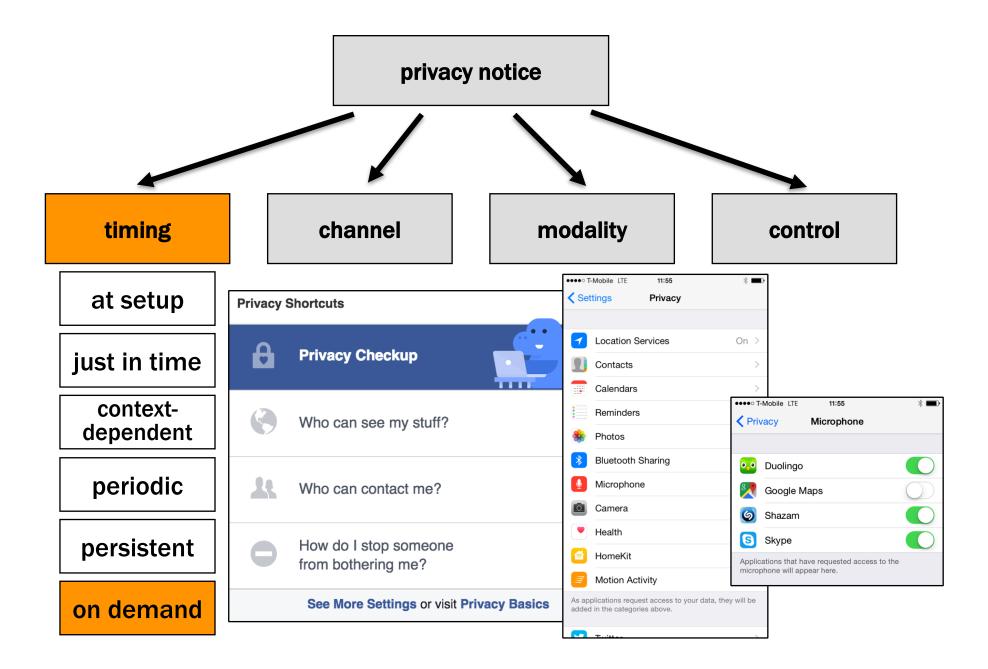


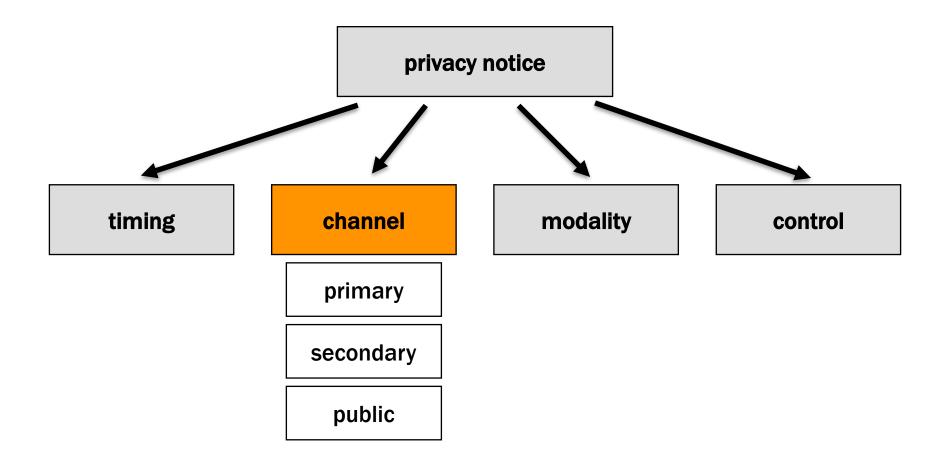


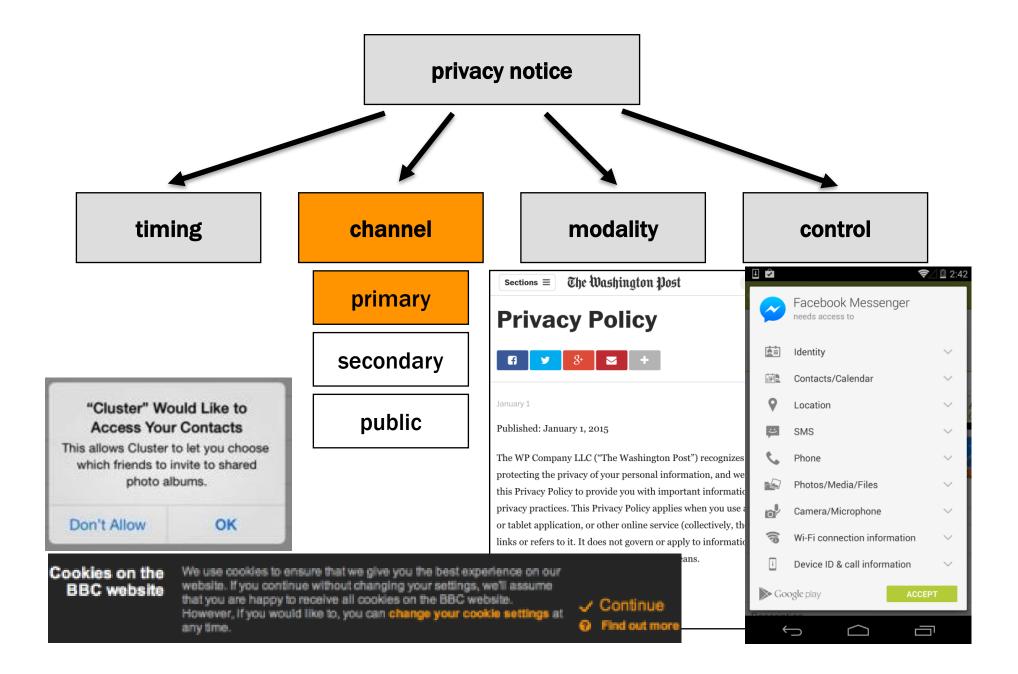


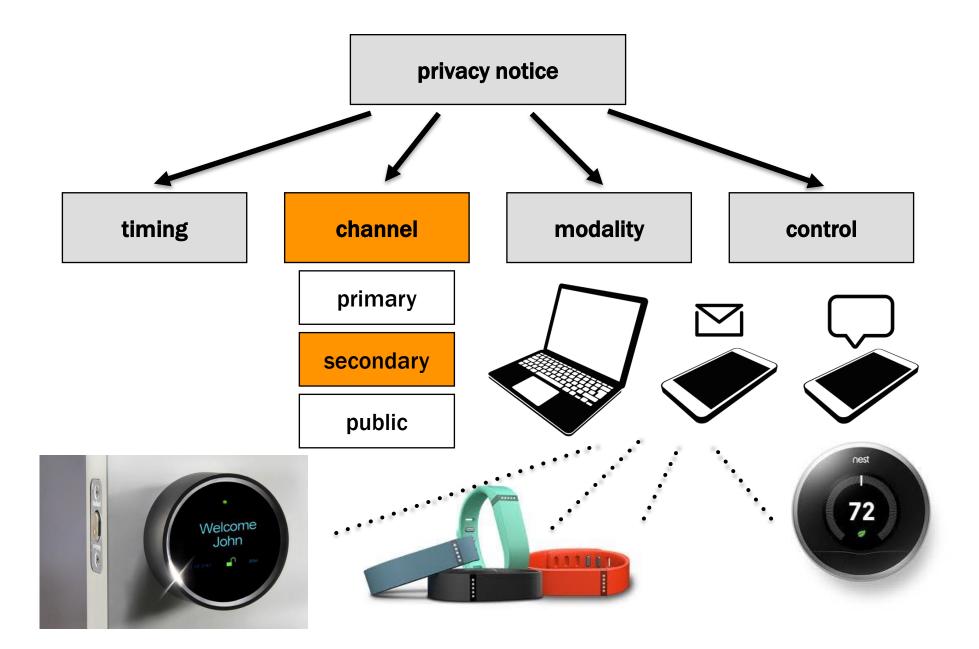


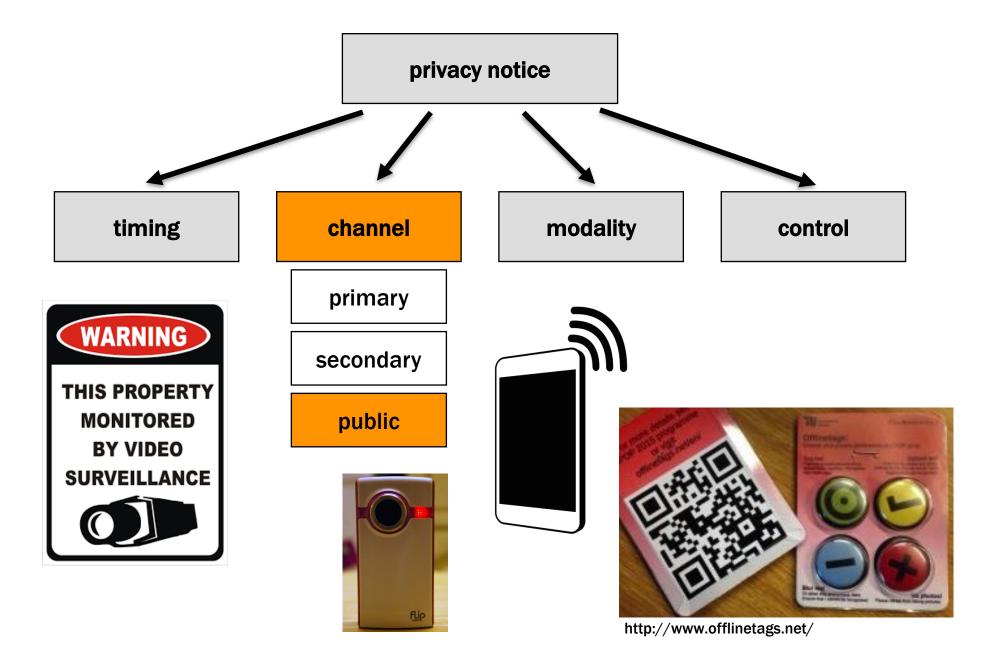


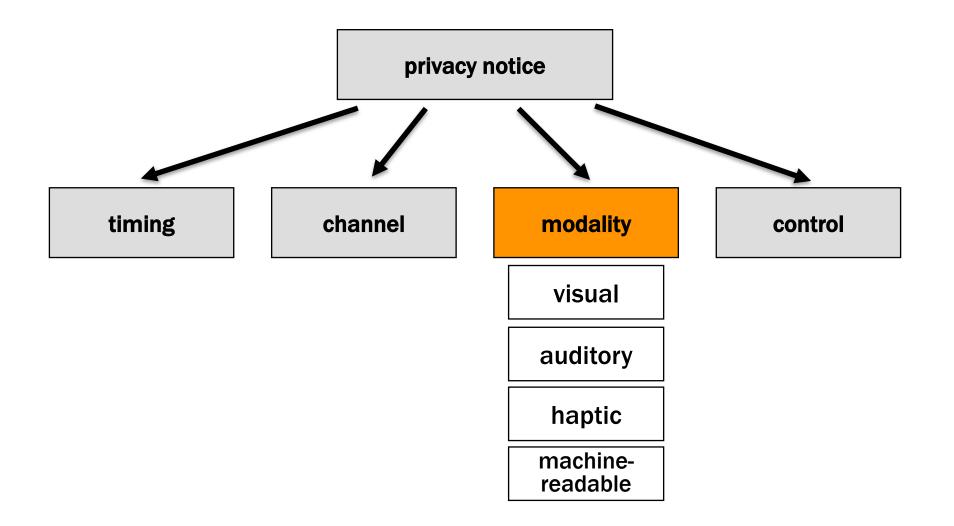


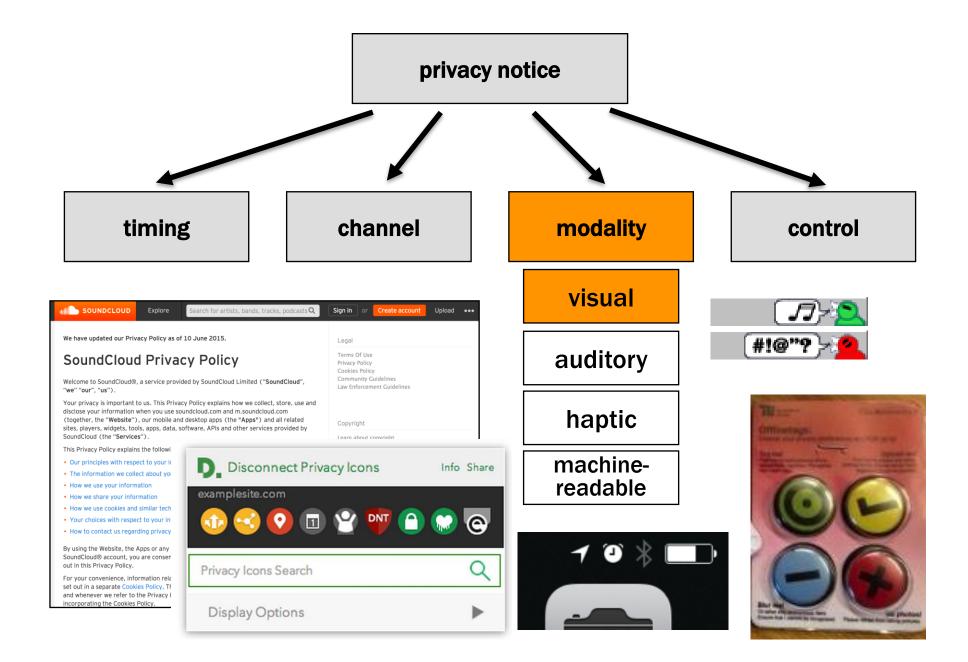


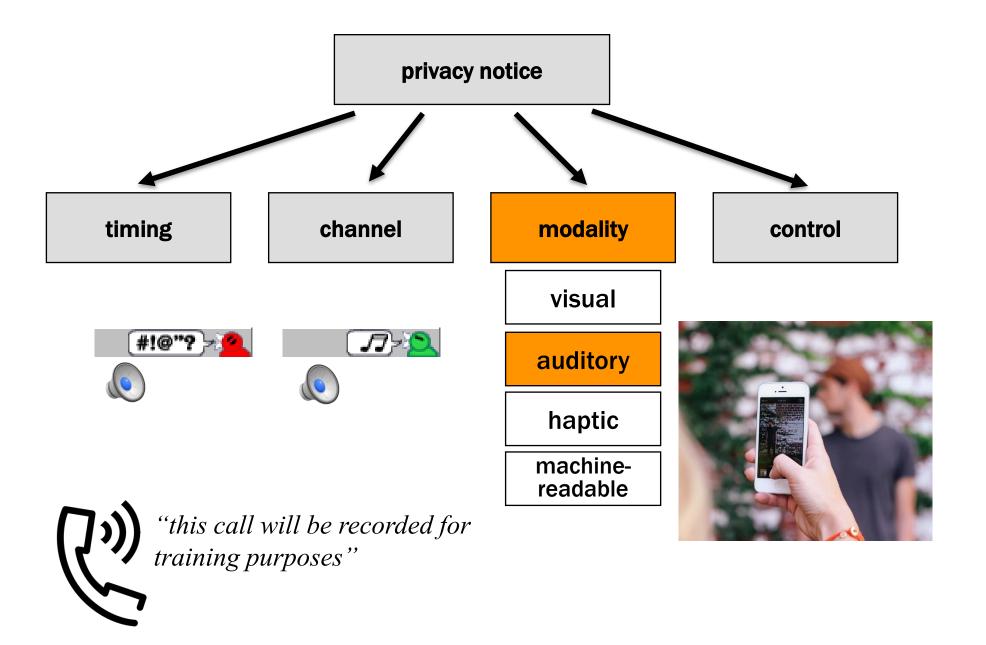


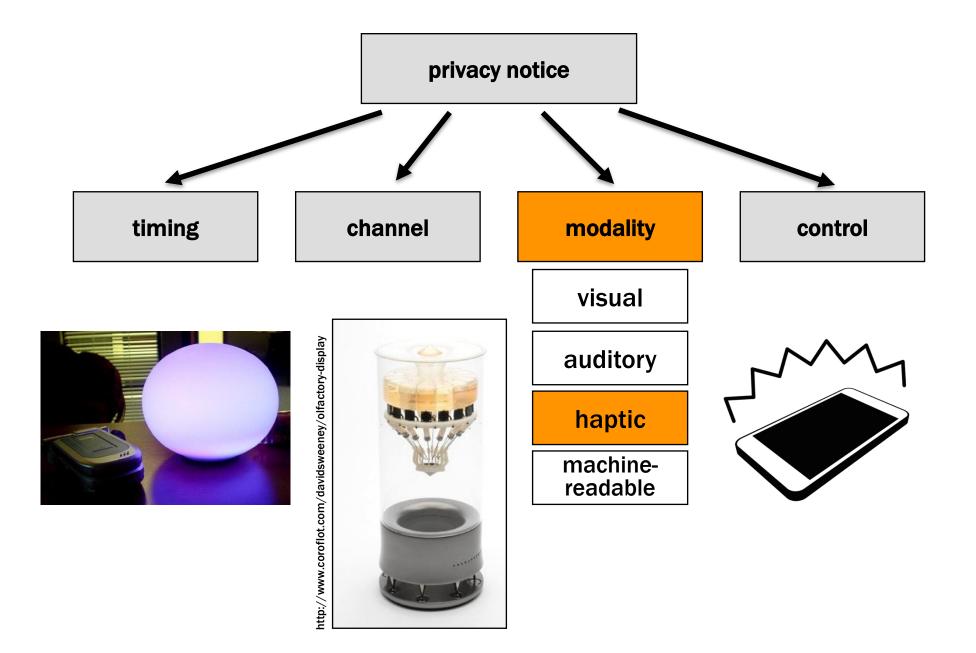


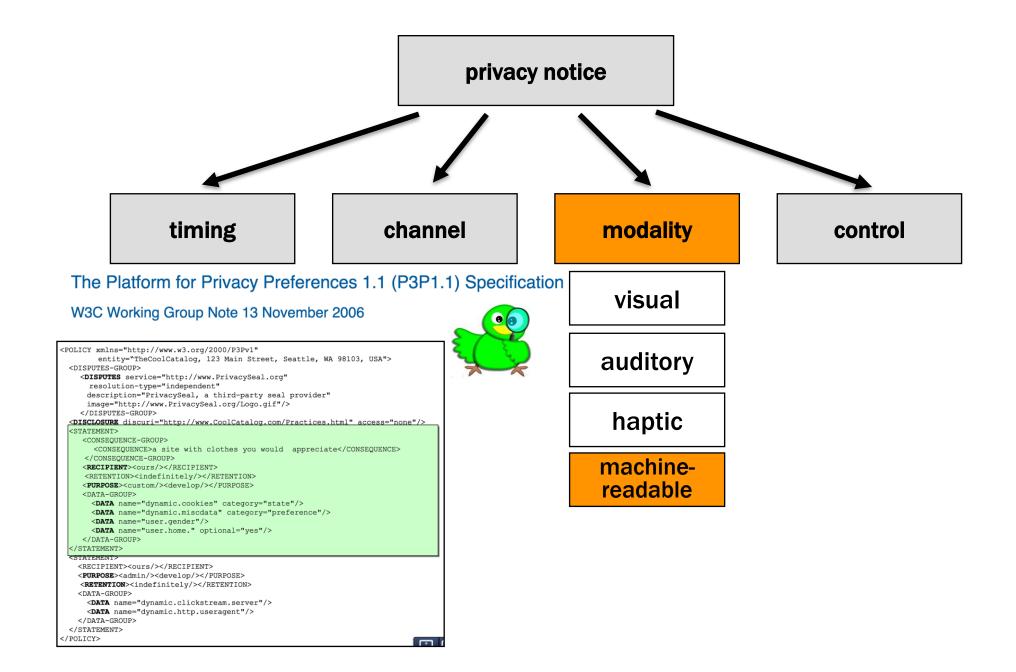


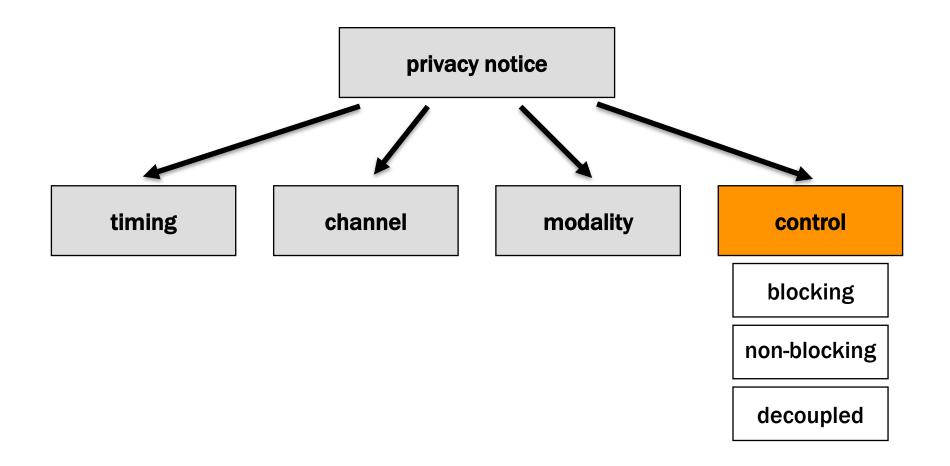


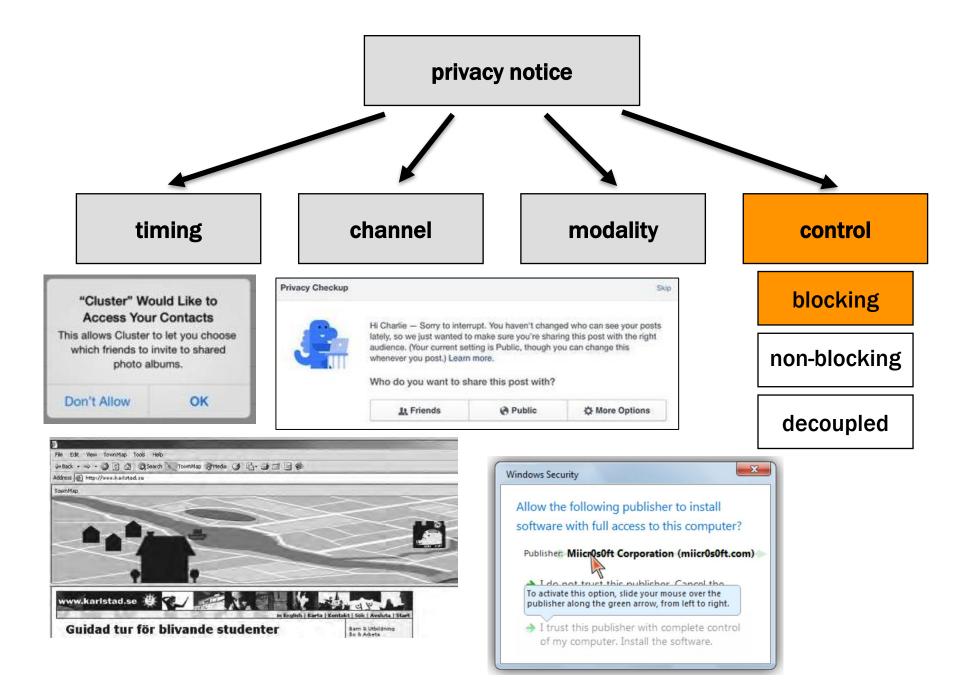


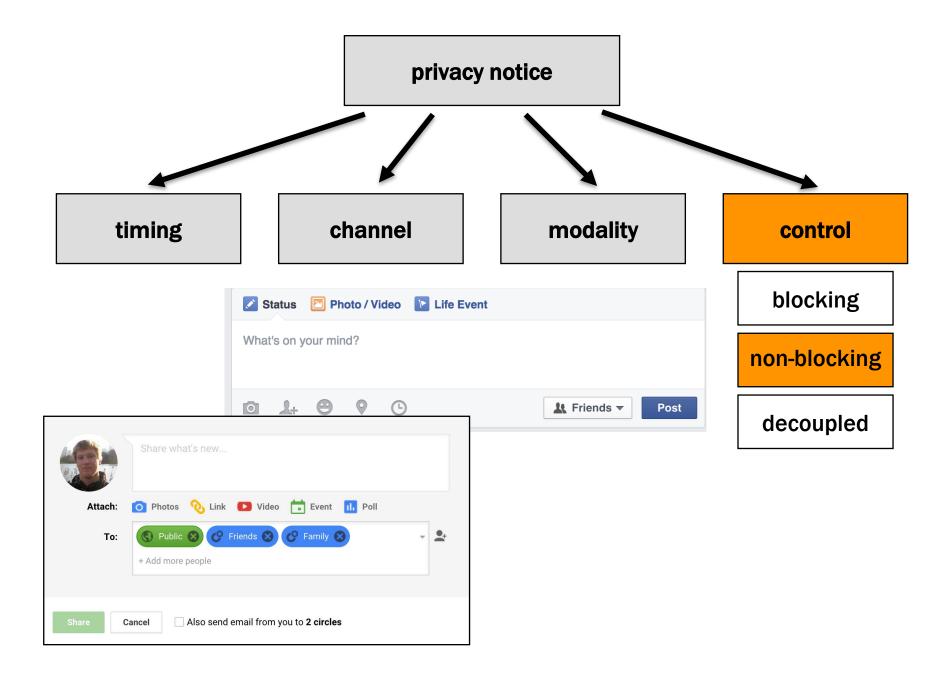


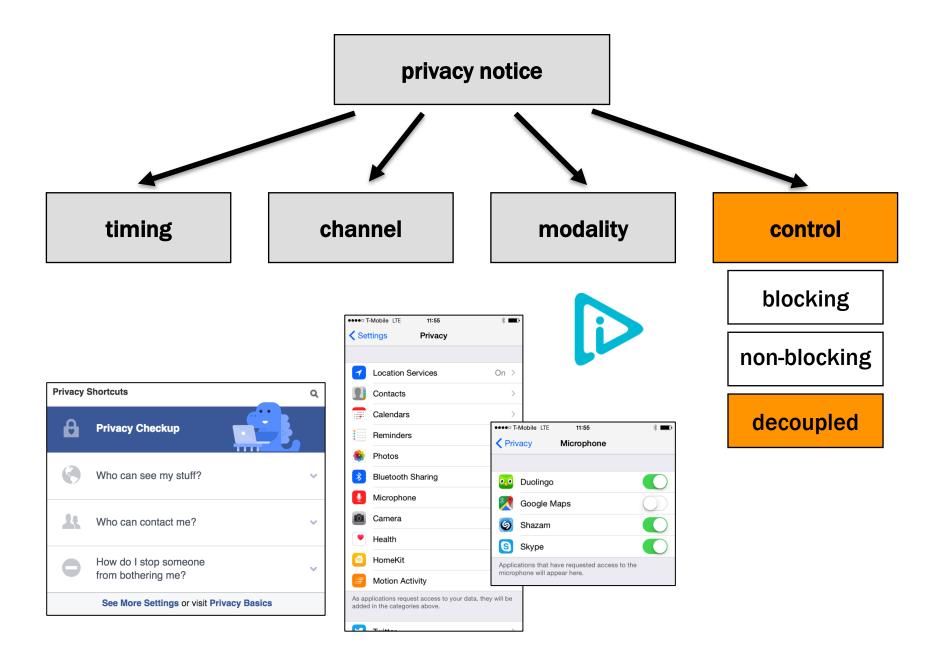




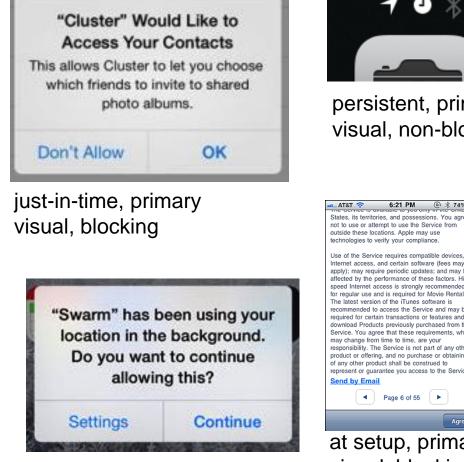








Layered examples from iOS



periodic, primary visual, blocking

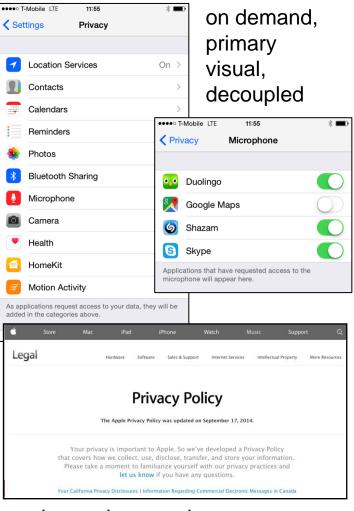


persistent, primary visual, non-blocking

6:21 PM @ 🖇 74% 🚍 States, its territories, and possessions. You agree not to use or attempt to use the Service from outside these locations. Apple may use technologies to verify your compliance

Internet access, and certain software (fees may apply); may require periodic updates; and may be affected by the performance of these factors. High speed Internet access is strongly recommended for regular use and is required for Movie Rentals The latest version of the iTunes software is recommended to access the Service and may be required for certain transactions or features and to download Products previously purchased from the Service. You agree that these requirements, which may change from time to time, are your responsibility. The Service is not part of any other product or offering, and no purchase or obtaining of any other product shall be construed to represent or guarantee you access to the Service Page 6 of 55 Agree

at setup, primary visual, blocking

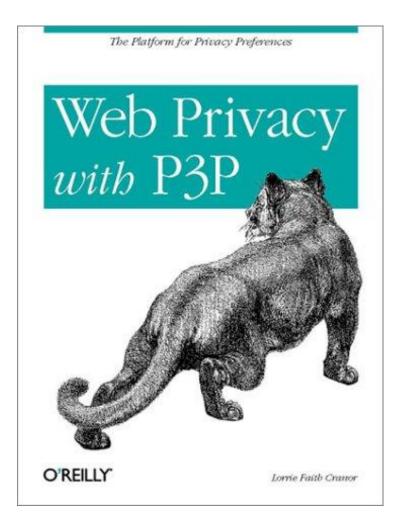


on demand, secondary visual, decoupled

Attempts at improving notice & choice

Attempt: Machine-readable privacy policies

Let your computer read for you



- Platform for Privacy Preferences (P3P)
- W3C specification for XML privacy policies
 - Proposed 1996
 - Adopted 2002
- Optional P3P compact policy HTTP headers to accompany cookies
- Lacks incentives for adoption

P3P in Internet Explorer

- P3P implemented in IE 6, 7, 8, 9, 10 ...
- Default privacy setting
 - Rejects third-party cookies without a CP
 - Rejects unsatisfactory thirdparty cookies

nternet Options 🛛 🖓 🔀
General Security Privacy Content Connections Programs Advanced
Settings Select a setting for the Internet zone.
Medium - - Blocks third-party cookies that do not have a compact privacy policy - - Blocks third-party cookies that save information that can be used to contact you without your explicit consent - - Blocks third-party cookies that save information that can be used to contact you without your explicit consent - - - Blocks third-party cookies that save information that can be used to contact you without your implicit consent - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Pop-up Blocker Prevent most pop-up windows from appearing. Turn on Pop-up Blocker InPrivate
Do not collect data for use by InPrivate Filtering Disable toolbars and extensions when InPrivate Browsing starts
OK Cancel Apply

Attempt: Make your own machine-readable privacy policies

Use NLP to read policies

- Usableprivacy.org
- It's a hard problem
- Annotated corpus \rightarrow machine learning

Attempt: Crowdsourcing

Terms of Service; Didn't Read

SoundCloud Class B

🕐 You stay in control of your copyright

Collected personal data used for limited purposes

C 6 weeks to review changes

🖸 Indemnification from claims related to your content or your account

Personal information can be disclosed in case of business transfer or insolvency

More details

🖫 GitHub 🛛 🖓 🖫

🕐 You don't grant any copyright license to github

- 💽 Changes can happen any time, sometimes without notice
- You shall defend and indemnify GitHub
- C Your personal information is used for limited purposes
- Your account can be suspended and your data deleted any time for any reason

More details

Twitpic Class E

X Twitpic takes credit for your content

- Your content is for Twitpic and their partners
- Reduction of legal period for cause of action
- You indemnify Twitpic from any claim related to your content
- 🗙 Deleted images are not really deleted

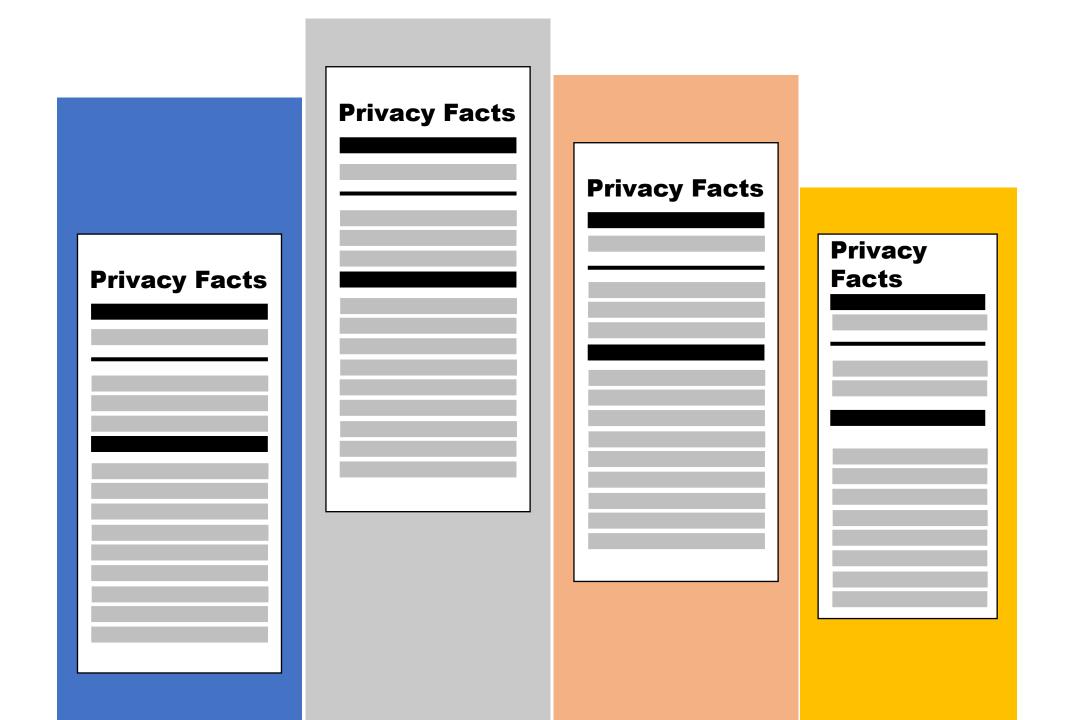
More details

Delicious Class D

- Very broad copyright license on your content, includes right for
- Delicious to distribute through any media
- X No Right to leave the service
- 💽 Only for personal and non-commercial use
- C [bad] delicious new terms 5. third party services get access to personal information
- Your personal information are an asset for business transfers

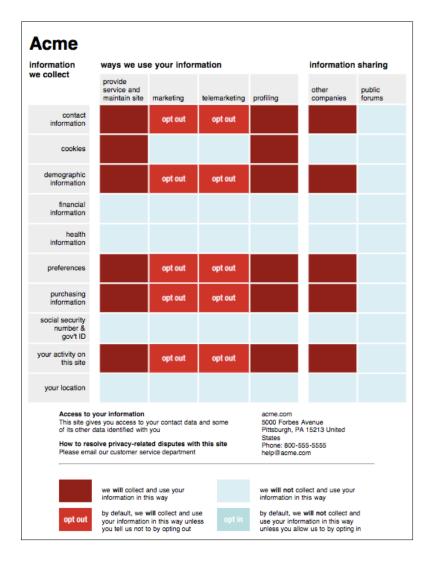
More details

Attempt: Standardized notices



Privacy nutrition labels

- Series of studies
 - Focus groups
 - Lab studies
 - Online studies
- Metrics
 - Reading-comprehension (accuracy)
 - Time to find information
 - Ease of policy comparison
 - Subjective opinions, ease, fun, trust
- P.G. Kelley, J. Bresee, L.F. Cranor, and R.W. Reeder. A "Nutrition Label" for Privacy. SOUPS 2009.
- P.G. Kelley, L.J. Cesca, J. Bresee, and L.F. Cranor. Standardizing Privacy Notices: An Online Study of the Nutrition Label Approach. CHI 2010.



- Gramm-Leach-Bliley Act mandated annual disclosures
- In 2009, regulators created a recommended form
 - High adoption because of safe harbor

Rev. [insert date]

FACTS	WHAT DOES [NAME OF FINANCIAL INSTITUTION] DO WITH YOUR PERSONAL INFORMATION?		
Why?	Financial companies choose how they share your personal information. Federal law gives consumers the right to limit some but not all sharing. Federal law also requires us to tell you how we collect, share, and protect your personal information. Please read this notice carefully to understand what we do.		
What?	The types of personal information we collect and share depend on the product or service you have with us. This information can include: Social Security number and [income] [account balances] and [payment history] [credit history] and [credit scores]		
	When you are <i>no longer</i> our customer, we continue to share your information as described in this notice.		
How?	All financial companies need to share customers' personal information to run their everyday business. In the section below, we list the reasons financial companies can share their customers' personal information; the reasons [name of financial institution] chooses to share; and whether you can limit this sharing.		

Reasons we can share your personal information	Does [name of financial institution] share?	Can you limit this sharing?
For our everyday business purposes – such as to process your transactions, maintain your account(s), respond to court orders and legal investigations, or report to credit bureaus		
For our marketing purposes – to offer our products and services to you		
For joint marketing with other financial companies		
For our affiliates' everyday business purposes – information about your transactions and experiences		
For our affiliates' everyday business purposes – information about your creditworthiness		
For our affiliates to market to you		
For nonaffiliates to market to you		

Questions? Call [phone number] or go to [website]

Who we are				
Who is providing this notice?	[insert]			
What we do				
How does [name of financial institution] protect my personal information?	To protect your personal information from unauthorized access and use, we use security measures that comply with federal law. These measures include computer safeguards and secured files and buildings.			
	[insert]			
How does [name of financial institution] collect my personal information?	 We collect your personal information, for example, when you [open an account] or [deposit money] [pay your bills] or [apply for a loan] [use your credit or debit card] 			
	[We also collect your personal information from other companies.] OR [We also collect your personal information from others, such as credit bureaus, affiliates, or other companies.]			
Why can't I limit all sharing?	 Federal law gives you the right to limit only sharing for affiliates' everyday business purposes—information about your creditworthiness affiliates from using your information to market to you sharing for nonaffiliates to market to you State laws and individual companies may give you additional rights to 			

Affiliates	Companies related by common ownership or control. They can be financial and nonfinancial companies. [affiliate information]
Nonaffiliates	Companies not related by common ownership or control. They can be financial and nonfinancial companies.
	[nonaffiliate information]
Joint marketing	A formal agreement between nonaffiliated financial companies that together market financial products or services to you.
	[joint marketing information]
Other important informat	tion
[insert other important informati	

- Built a parser and built an online database
- Lets people compare practices
- <u>https://cups.cs.cmu.edu/bankprivacy</u>

Attempt: Improve timing

Privacy label for Android



DESCRIPTION

Like word games? Like weasels? This is the game for you. Word Weasel is a fast word game where you find as many words as you can from 9 letters in 50 seconds. Compete with everyone else playing at the same time, a new game starts every minute!

"The most fun you can have on your own with 9 random letters. Brilliantly simple, devilishly addictive." --Kim, Android Market comment



comment

Ad supported, with ability to upgrade to adfree full version in game.

Uses the SOWPODS word list containing 100,000+ words. They may be obscure, but yes those are words!



n Credit card / financial Diet / nutrition Health / medical Photos

Analytics

REVIEWS

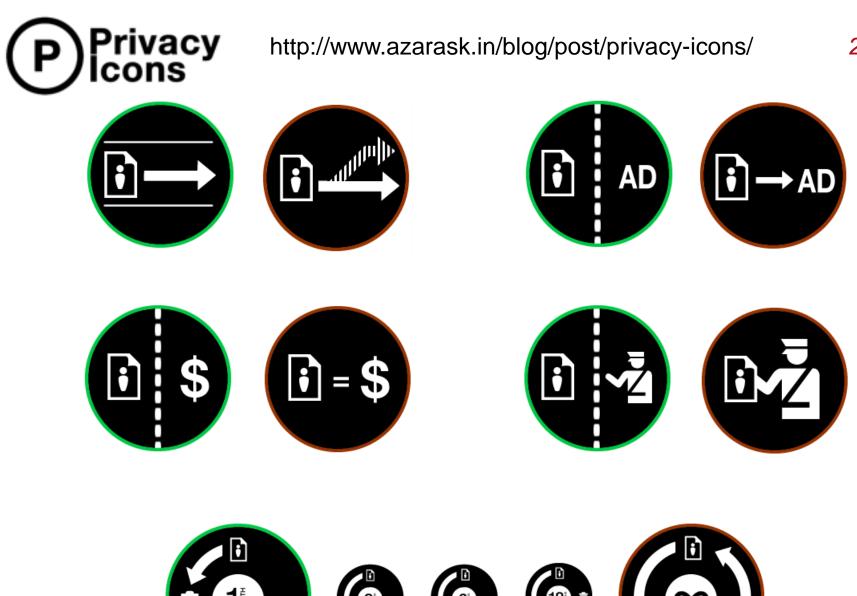
Average	5 stars	534
	4 stars	210
4.6	3 stars	37
*****	2 stars	5
793	1 star	7
	i star	/

Attempt: Personalized privacy assistants

Personal privacy assistants



Attempt: Icons





Attempt: Standardized disclosure icons







