Purposeful permissions

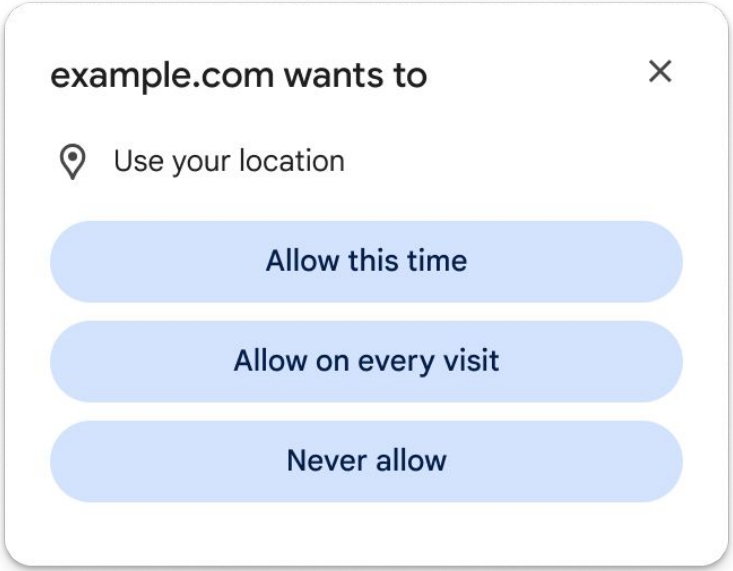
Adding data use and other contextual information to permission prompts

Alexandra Reimers (Google Chrome)
Serge Egelman (ICSI/UC Berkeley)
Nick Doty (CDT)

Agenda

- Intro & Why?
- Hypotheses about user behaviour and perception
- The role of different stakeholders
- How and what to declare?
- Challenges and opportunities

Current permission requests



User questions:

- What can I do if I allow?
- What else is my location data used for?
- How long is my location data retained?
- Who else gets my location data?
- Is my location data being sold?

Current state:

- Some answers are in the privacy policy.
- Some websites prepare the users before asking.
- Neither structured, nor consistent across websites.

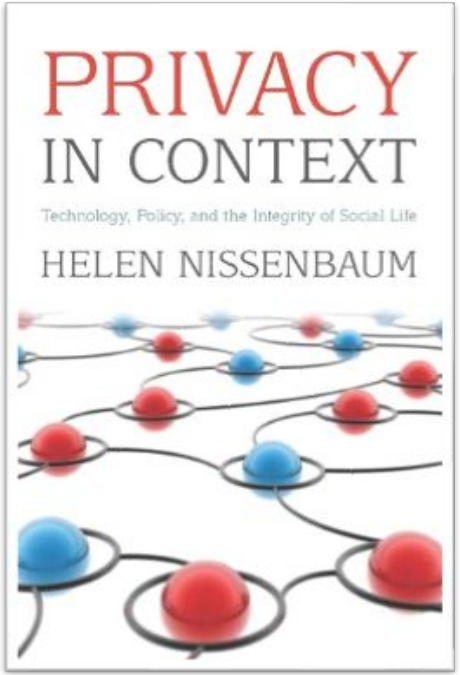
Key example: government credentials

Privacy as Contextual Integrity

- Philosophical theory of privacy:
  Privacy is the appropriate flow of information within contexts according to the norms of those contexts

- Explains why events turn into privacy outcries and why other simpler theories fail to explain these outcries

- Models data flows as:
  - Data subject
  - Sender
  - Receiver
  - Data type
  - Transmission principles (constraints)
    …all of which occur within a *context* (purpose)



Adding data use info: hypotheses about user perception

How we assume users currently engage with permission prompts:

- Users likely have automatisms/habituation and make fast decisions on permission prompts.
- Decisions based on site context or prior experiences.
- Some users currently don’t know where to find data use information.

What we assume users might value or expect:

- Data use info perceived as valuable.
- Data use info especially valuable when data is used for unexpected purposes.
- Reassurance when data is used only for site functionality.
- Expectation that data use info comes from website.
- Expectation that the browser verifies websites’ data use declarations.

Prior Example: The Platform for Privacy Preferences (P3P)

- Developed by the World Wide Web Consortium (W3C)
  http://www.w3.org/p3p/
  Final P3P1.0 Recommendation issued 16 April 2002
- Offers an easy way for web sites to communicate about their privacy policies in a standard machine-readable format
  - Can be deployed using existing web servers
- Enables the development of tools (built into browsers or separate applications) that
  - Summarize privacy policies
  - Compare policies with user preferences
  - Alert and advise users
- P3P support built into IE6 and Netscape 7
- Many WG meetings were spent arguing about the data collection purpose tags!

3.3.5 The *PURPOSE* element

Each *STATEMENT* element that does not include a *NON-IDENTIFIABLE* element MUST contain a *PURPOSE* element that contains one or more purposes of data collection or uses of data. Sites MUST classify their data practices into one or more of the purposes specified below.

<PURPOSE>
purposes for data processing relevant to the Web.

The *PURPOSE* element MUST contain one or more of the following:

<current/>
Completion and Support of Activity For Which Data Was Provided: Information may be used by the service provider to complete the activity for which it was provided, whether a one-time activity such as returning the results from a Web search, forwarding an email message, or placing an order; or a recurring activity such as providing a subscription service, or allowing access to an online address book or electronic wallet.

<admin/>
Web Site and System Administration: Information may be used for the technical support of the Web site and its computer system. This would include processing computer account information, information used in the course of securing and maintaining the site, and verification of Web site activity by the site or its agents.

<develop/>
Research and Development: Information may be used to enhance, evaluate, or otherwise review the site, service, product, or market. This does not include personal information used to tailor or modify the content to the specific individual nor information used to evaluate, target, profile or contact the individual.

<tailoring/>
One-time Tailoring: Information may be used to tailor or modify content or design of the site where the information is used only for a single visit to the site and not used for any kind of future customization. For example, an online store might suggest other items a visitor may wish to purchase based on the items he has already placed in his shopping basket.

<pseudo-analysis/>
Pseudonymous Analysis: Information may be used to create or build a record of a particular individual or computer that is tied to a pseudonymous identifier, without tying identified data (such as name, address, phone number, or email address) to the record. This profile will be used to determine the habits, interests, or other characteristics of individuals *for purpose of research, analysis and reporting,* but it will not be used to attempt to identify specific individuals. For example, a marketer may wish to understand the interests of visitors to different portions of a Web site.

<pseudo-decision/>
Pseudonymous Decision: Information may be used to create or build a record of a particular individual or computer that is tied to a pseudonymous identifier, without tying identified data (such as name, address, phone number, or email address) to the record. This profile will be used to determine the habits, interests, or other characteristics of individuals *to make a decision that directly affects that individual*, but it will not be used to attempt to identify specific individuals. For example, a marketer may tailor or modify content displayed to the browser based on pages viewed during previous visits.

<individual-analysis/>
Individual Analysis: Information may be used to determine the habits, interests, or other characteristics of individuals and combine it with identified data *for the purpose of research, analysis and reporting.* For example, an online Web site for a physical store may wish to analyze how online shoppers make offline purchases.

<individual-decision/>
Individual Decision: Information may be used to determine the habits, interests, or other characteristics of individuals and combine it with identified data *to make a decision that directly affects that individual.* For example, an online store suggests items a visitor may wish to purchase based on items he has purchased during previous visits to the Web site.

<contact/>
Contacting Visitors for Marketing of Services or Products: Information may be used to contact the individual, through a communications channel other than voice telephone, for the promotion of a product or service. This includes notifying visitors about updates to the Web site. This does not include a direct reply to a question or comment or customer service for a single transaction -- in those cases, ***<current/>*** would be used. In addition, this does not include marketing via customized Web content or banner advertisements embedded in sites the user is visiting -- these cases would be covered by the ***<tailoring/>***, ***<pseudo-analysis/>*** and ***<pseudo-decision/>***, or ***<individual-analysis/>*** and ***<individual-decision/>*** purposes.

<historical/>
Historical Preservation: Information may be archived or stored for the purpose of preserving social history as governed by an existing law or policy. This law or policy MUST be referenced in the *<DISPUTES>* element and MUST include a specific definition of the type of qualified researcher who can access the information, where this information will be stored and specifically how this collection advances the preservation of history.

<telemarketing/>
Contacting Visitors for Marketing of Services or Products Via Telephone: Information may be used to contact the individual via a voice telephone call for promotion of a product or service. This does not include a direct reply to a question or comment or customer service for a single transaction -- in those cases, ***<current/>*** would be used.

<other-purpose> ***string*** ***</other-purpose>***
Other Uses: Information may be used in other ways not captured by the above definitions. (A human readable explanation MUST be provided in these instances).

Potential roles of web ecosystem stakeholders

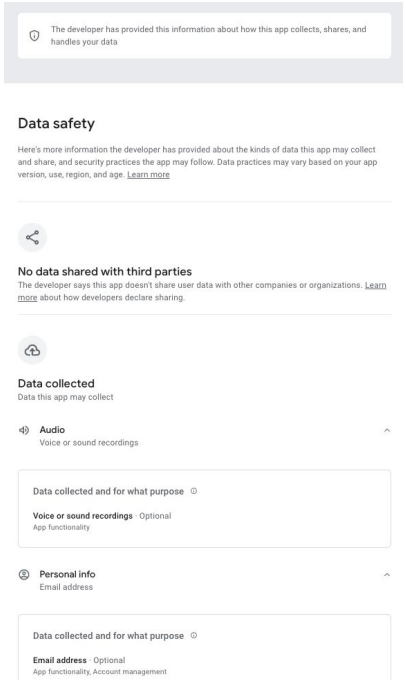
- **Websites:** declare how they use the data they’re asking for
- **Browsers:**
  - Show data use purposes (in the decision moment, as a learn more resource, in page info etc.)
  - Create aggregate reports that show how declarations evolve over time
- **Regulators / governments:** use to augment assessments of websites' data practices and disclosures
- **Standards community:** defines and evolves declaration format

Different granularities possible

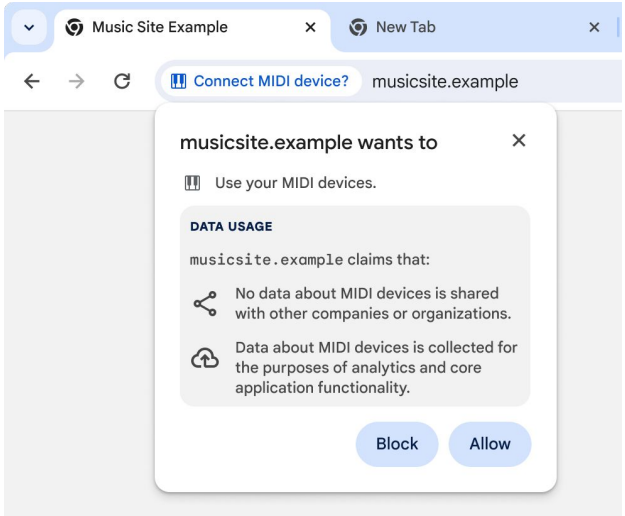
- Only link the website privacy policy in the prompt
- Free text
- Labels/Enums: Data type and purpose categories as in Play/App Store
- Combination of labels and free text
- per-origin well-known files, potentially with external verifications

Possible form factors

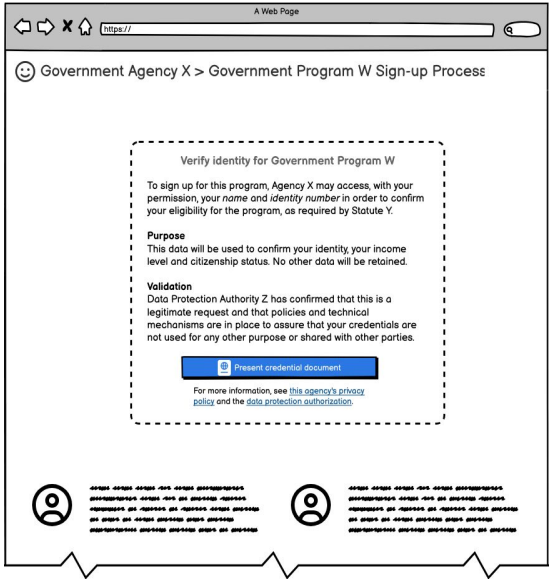
Pull on demand



In the prompt



In the content area



Challenges & Opportunities

Privacy policy link:

- Limited user value - privacy policies cover all of a site's data practices; not granular enough at the single permission level
- Easy for websites to adopt

Free text:

- Abusable
- Unstructured
- Inconsistent across websites
- Easy for websites to adopt
- Browsers unable to independently verify claim accuracy

Labels:

- Limited user understanding of broad categories
- Sub-categorizing to increase understanding → requires a full taxonomy of purposes
- Standardizable, structured and consistent across websites
- Scannable at a glance
- Comparable across websites and over time
- Browsers unable to independently verify claim accuracy
- Websites may take issue with generalized labels as applied to their specific practices

Adoption:

- Why should websites declare?
- What happens when the purposes change?

Open Research Questions

- Should the set of possible purposes be enumerated by a specification or open-ended text supplied by websites?
  - How can we best differentiate website-supplied content vs. browser-supplied content?
- What information do users need to make informed decisions?
  - Are there some purposes people don’t care about?
  - Are there some data types people don’t care about?
  - What does this say about defaults?
- Will users explore secondary UI (e.g., “more information” buttons)?
  - What information should be in primary UI?
- Which designs result in the narrowest gap between stated preferences and observed behaviors?
  - Are there better metrics that we should use?

Next Steps

Proposals to bring to incubation?

Volunteers to write up examples?

What further research is needed?