Geolocation, privacy and the Web

Nick Doty

UC Berkeley, School of Information

Agenda:

Increasingly common geolocation technologies
Effects on privacy
World Wide Web Consortium's Geolocation API and our research on its use
Future of privacy on the Web
How we think academics can get involved

Google

SKYHOOK

Increasing availability of GPS chips in cellphones and even cameras

GPS: uses satellites and can be very precise, but only works outdoors and slowly
Cell-site triangulation: variable precision, but wide coverage for cellphones
WiFi triangulation: quite precise in some areas, requires large databases

Database providers just these three companies

Effects on privacy

Location information is:

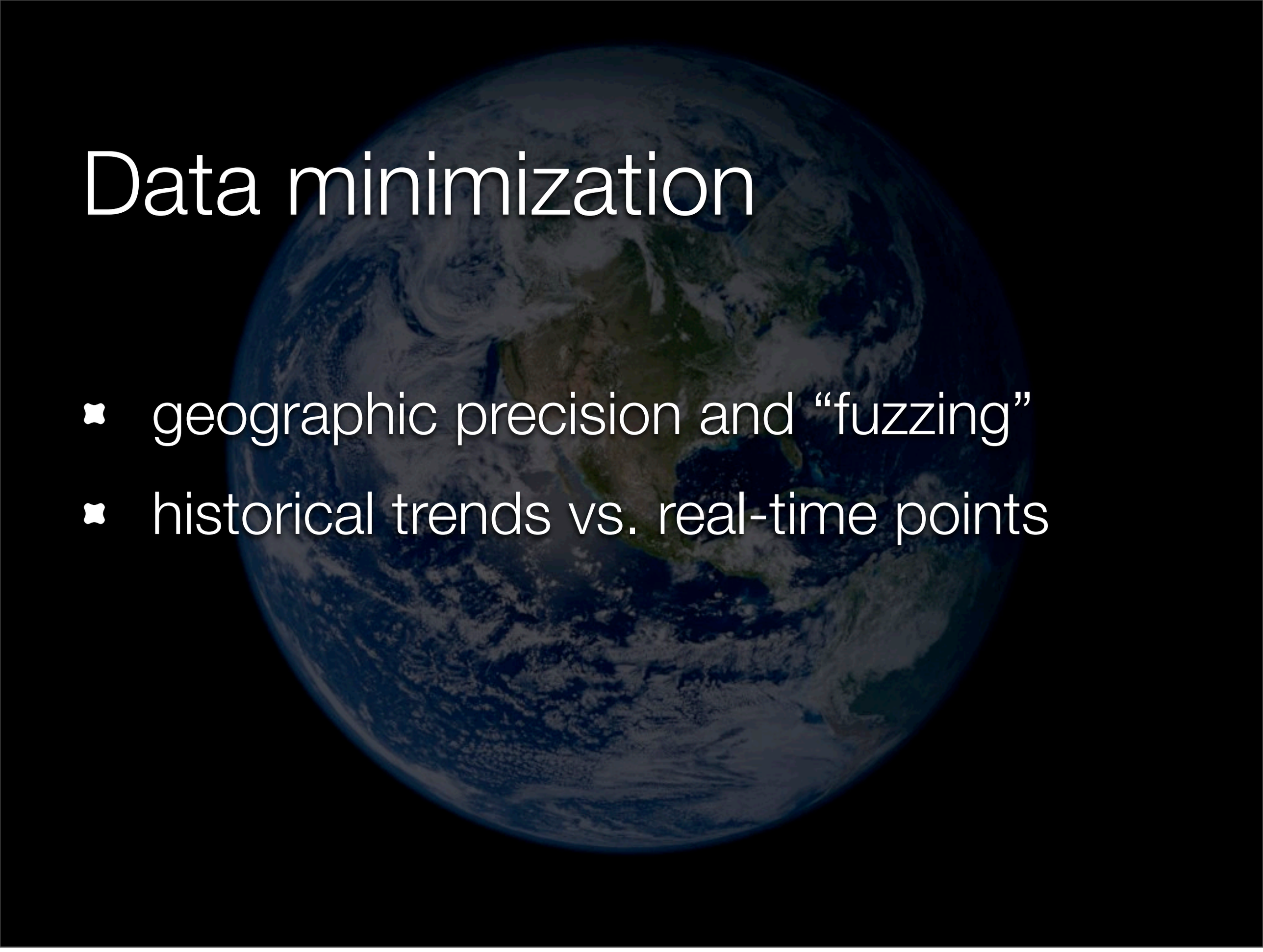
- informationally revealing
- personally identifying
- physically intrusive

revealing: not just where I am, but what I'm probably doing; where I'm not, who I'm with, my financial associations, medical condition, even sexual orientation

identifying: on the otherwise anonymous Web, location histories can easily be tied to a real name (who else spends the night at my house and the day at my job?)

intrusive: the most basic form of privacy, stalker or ex-husband (WSJ) or government agents

Data minimization



- geographic precision and “fuzzing”
- historical trends vs. real-time points

Studies of privacy often consider the principle of DATA MINIMIZATION, which has a couple of interesting applications to location data in particular

geographic precision: in many cases, the usefulness of a location-based service doesn’t require precise location (weather, time-zone, even finding nearby movie theaters); unfortunately, reliance on latitude and longitude as a lingua franca of location-based services has held back this minimization; arguments over the best fuzzing algorithm have slowed progress

historical trends: many of the most revealing aspects of location depend on historical trend data, where a single point might not be as sensitive; this makes aggregation by a small number of providers all the more concerning;
this also has legal implications: an explosion of litigation shows intense debate, some courts have found that extensive, historical GPS data violates 4th Amendment rights even on public roads; others have found that historical cell-site data requires a lower standard as an electronic record (ECPA/SCA)

W3C Geolocation API

World Wide Web Consortium
Candidate Recommendation

- High-level, JavaScript API
- Agnostic to underlying geolocation technology
- Latitude and longitude only

W3C: technical standards body responsible for Web standards, things we all know and use in our browsers every day, HTML, JavaScript, etc. Founded by Tim Berners-Lee.

Recently promoted to Candidate Recommendation, over privacy objections.

DEMO

W3C Geolocation API

Security and privacy considerations

- Browser implementations require yes-or-no consent
- Web site implementations require “clear and conspicuous disclosure”

W3C Geolocation API

Security and privacy considerations

W3C Candidate Recommendation

4.2 Privacy considerations for recipients of location information

Recipients must only request location information when necessary. Recipients must only use the location information for the task for which it was provided to them. Recipients must dispose of location information once that task is completed, unless expressly permitted to retain it by the user. Recipients must also take measures to protect this information against unauthorized access. If location information is stored, users should be allowed to update and delete this information.

The recipient of location information must not retransmit the location information without the user's express permission. Care should be taken when retransmitting and use of encryption is encouraged.

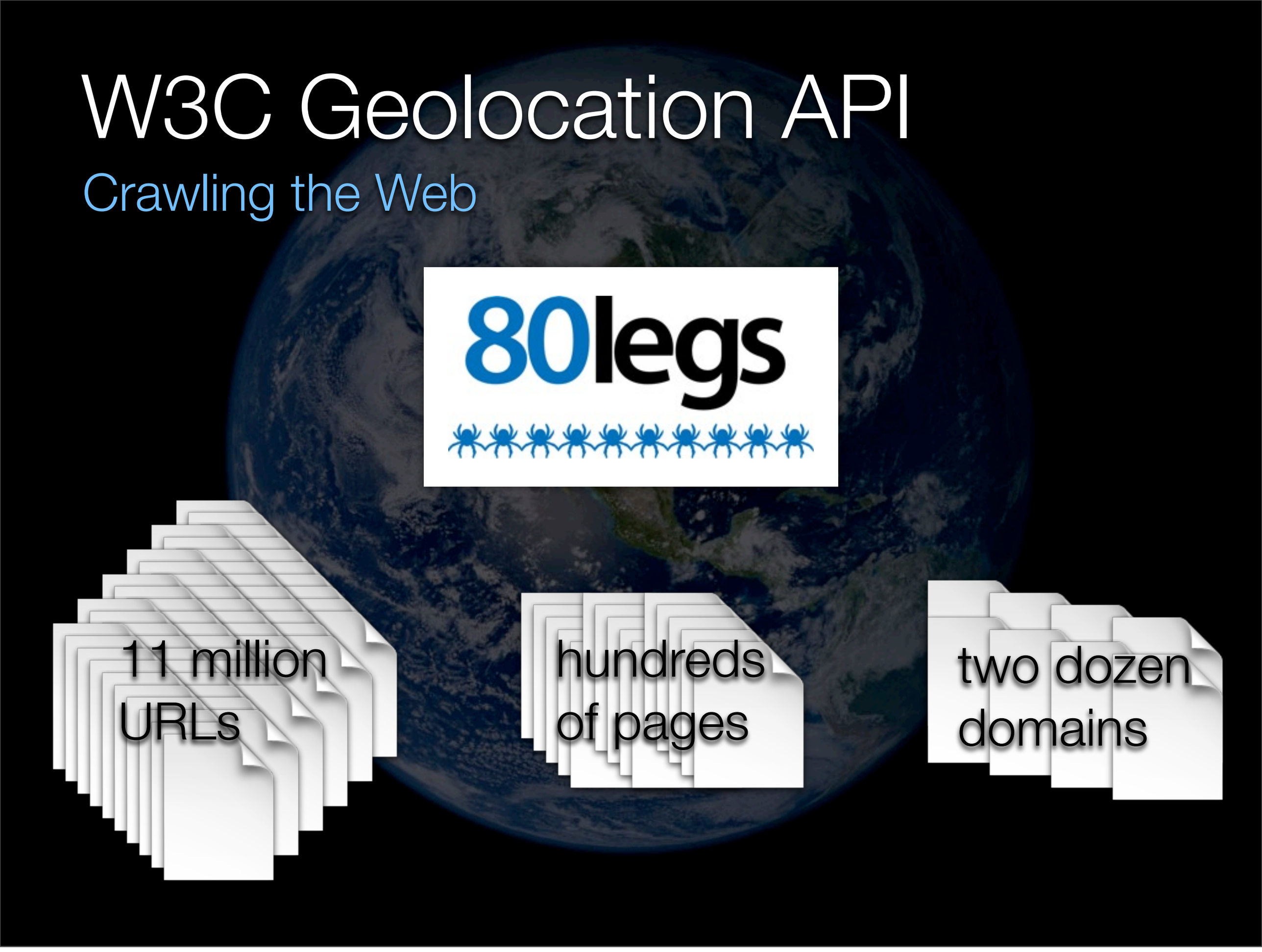
Recipients must clearly and conspicuously disclose the fact that they are collecting location data, the purpose for the collection, how long the data is retained, how the data is secured, how the data is shared if it is shared, how users may access, update and delete the data, and any other choices that users have with respect to the data. This disclosure must include an explanation of any exceptions to the guidelines listed above.

This was mostly written by privacy advocates, and so is actually a good, strict set of rules.
This is normative (MUST, SHOULD, etc.) but not a technical part of the API.
Also, the W3C has far less influence over sites than it does over browser makers.

As a result, we wanted to look at existing web sites that use the API and see whether they followed these strict requirements...

W3C Geolocation API

Crawling the Web



Last year around this time (when the API was still a draft but just beginning to get adoption)...

found a couple dozen web sites, whose policies and disclosures we inspected manually

	What does it do?	Informed up front?	In Privacy Policy?	Lets user inspect?
Google Maps	Zoom the map to your location.	✖	●	✖
Google Local	Nearby points-of-interest.	✖	✔	✖
Flickr	Show pictures taken nearby.	✖	✖	✖
Travelocity iPhone	Search for nearby hotels.	✖	✖	●
AskLaila	Search for businesses in India.	✖	✖	●
Search.ch	Find Swiss train schedules.	✖	✖	✖
Identi.ca	Attach your location to public microblog posts.	✖	✖	✖
Foreca Weather	Get the weather forecast.	✖	✖	✖
BooRah Restaurants	Show restaurants near you.	✖	✖	✖
GoThere	Singaporean points of interest.	✖	✖	✖
The Rocky Horror Picture Show	Find Rocky Horror showtimes nearby.	✖	✖	✖
GraffitiGeo	Show tagged locations nearby.	✖	✖	✖
GeoMail	Add your location to an email.	✖	✖	●
Our Airports (mobile)	Show nearby airports.	✖	✖	✔
Our Airports	Show nearby airports.	✖	✖	✔
Plemi	Find nearby concerts.	✖	✖	✖
AskAround.Me	Answer geotagged questions.	✖	✖	✖
gMapTip WordPress	Add a map to a blog post.	✖	✖	✖
Your Mapper	See map data for your location.	✖	●	✔
BackNoise	Semi-private conversations.	✖	✖	✖
BailBond.com	Find a nearby bail bondsman.	✖	✖	✔
Toupil.fr	Search for businesses in France.	✖	-	✖

Google web sites are commendable for actually explaining their use of location fairly well, in human language even, but multiple clicks away.

Identi.ca, a microblogging site, actually attaches your location to all public posts you make on the site, with no warning that it will do so.

OurAirports exemplifies one good practice that we didn't expect, not sending the location back to the site, just helping the user fill in a form field.

- in short, no one provides these disclosures up-front, and where disclosures are present, they're far out of the way
- though virtually everyone consented (in the sense of having a chance to say yes-or-no), virtually no one gave informed consent

W3C Geolocation API

Crawling the Web... again

	Fall 2009	Spring 2010
Crawled	11 million	5 billion
Pages	hundreds	hundreds of thousands
Domains	24	1000

we published these results in a report in February of this year, but are trying to improve upon them

- currently working on a larger crawl thanks to help from friends at a large technology company, crawling 5 billion URLs to find about a thousand domains
- we haven't gone through that data manually yet, but with help from the REU students this summer, we've built a survey that we hope to deploy through Mechanical Turk

- in short, though, things haven't changed

Alternative:
Machine-readable Policy

- Browsers can read and highlight important points
- User agents in interface-less devices can respect preferences
- Policy statements are written in an explicit, formal language

Machine-readable Policy

Platform for Privacy Preferences (P3P)

- XML encoding of full-site privacy policies
- 1996 — started
  2002 — recommendation
  2006 — abandoned
- Too complex?
  Disincentives?

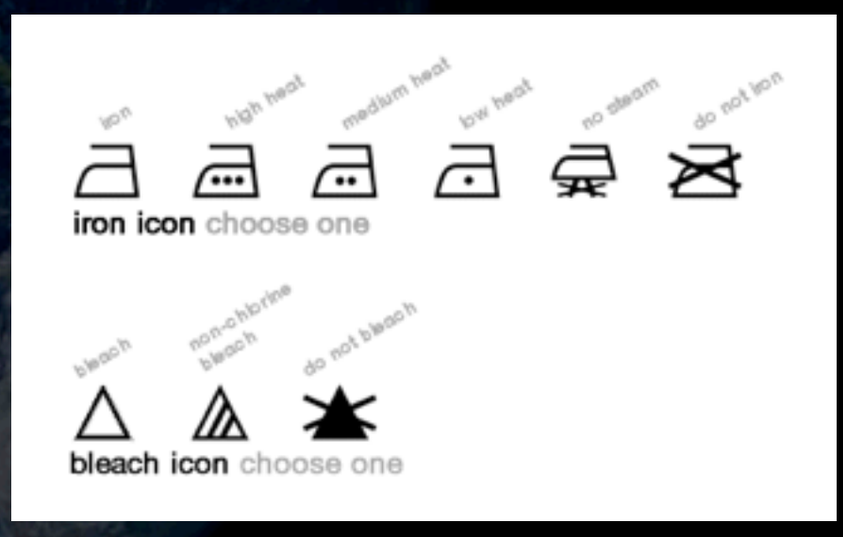
- P3P, 1996-2002-2006
- machine-readable encoding of complex privacy policies, so that browsers can read them automatically, make them more visible, make decisions on your behalf based on them
- suffered from being too complex and too rarely implemented, never took off
- lots of companies didn't want a simple, precise encoding of their privacy policies
- little variation in practices can make it hard to make any judgments based on these

- the ideas of P3P are frequently resurrected, most recently in Mozilla's Privacy Icons Project

Machine-readable Policy

Mozilla Privacy Icon Project

- XML encoding of full-site privacy policies, in laundry-tag style icons
- Large Firefox user base could drive adoption?
- Enough variation between sites?



- the ideas of P3P are frequently resurrected, most recently in Mozilla's Privacy Icons Project

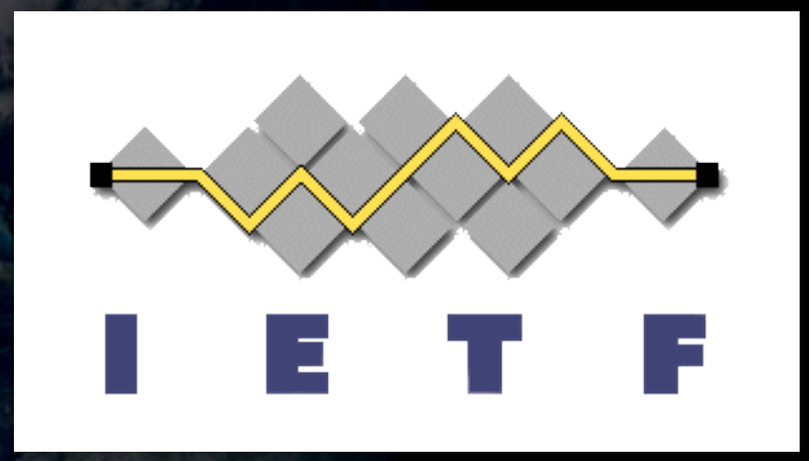
- GeoPriv (specific to location) takes a different angle, where policy is encoded in a machine-readable format, but policy is set by the user rather than the site
- user-specified rules are attached to data when it's returned (so that site's can't claim they didn't know: non-repudiation)
- W3C Geolocation WG thought this proposal was too complex, and didn't handle all use cases well

- we've recently proposed something in between these, allowing simple negotiation of policy information
  - sites provide a range of values that fit their use case, and users have the final say (screenshot)

Machine-readable Policy

GeoPriv — Internet Engineering Task Force

- User-specified XML encoding of personal privacy preferences
- Attached to location data
- Too complex? Flexible enough?

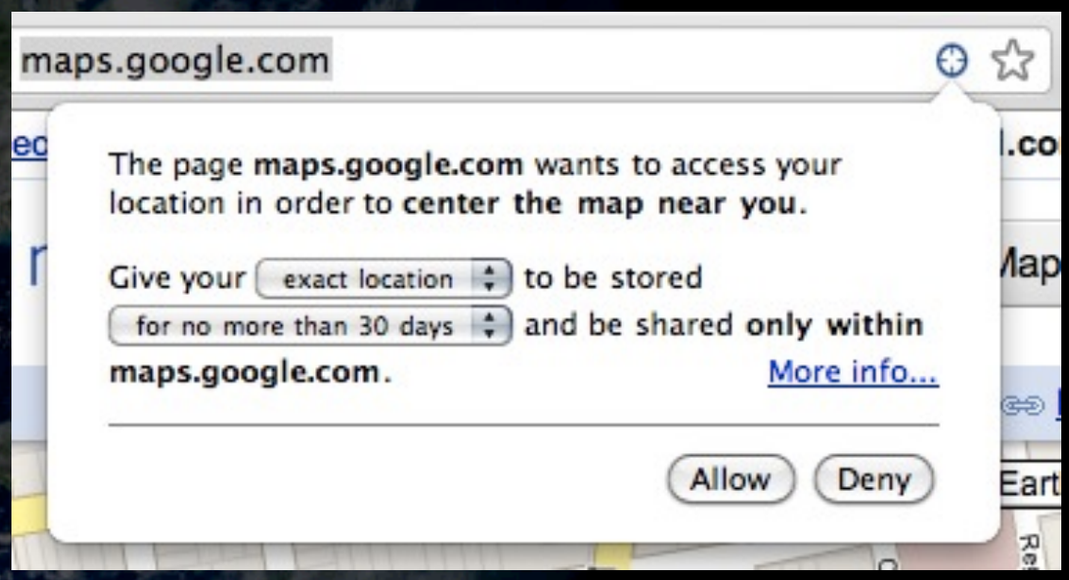


- GeoPriv (specific to location) takes a different angle, where policy is encoded in a machine-readable format, but policy is set by the user rather than the site
- user-specified rules are attached to data when it's returned (so that site's can't claim they didn't know: non-repudiation)
- W3C Geolocation WG thought this proposal was too complex, and didn't handle all use cases well

Machine-readable Policy

Simple Policy Negotiation for Location Disclosure

- Sites suggest a range of options (JavaScript, no XML)
- Users have the final say & attach their policies



- we've recently proposed something in between these, allowing simple negotiation of policy information
  - sites provide a range of values that fit their use case, and users have the final say

(screenshot)

Machine-readable Policy

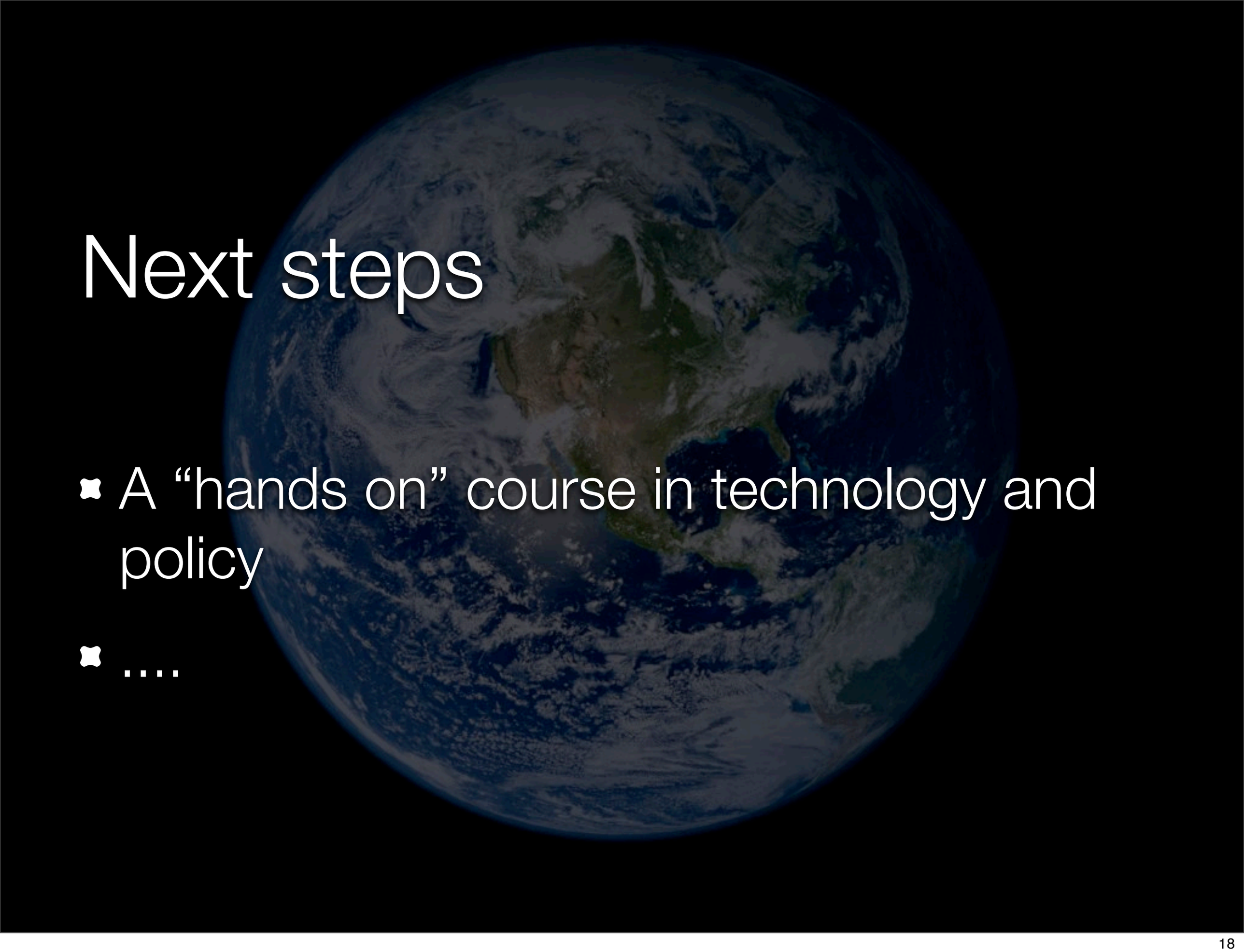
- "...but can't sites just lie?", the most common response from engineers
- yes, they can (and they do, in recently published research on P3P for cookies), so policy hooks in an API shouldn't be considered alone, or as an answer to security issues
- but for privacy, it's less a question of "good guys" and "bad guys" and more about making negotiations and disclosures of policy up-front
- up-front disclosures allow for enforcement through other means (FTC enforcement action, European Union, lawsuits, market-based measures)

I want to close by looking at the meta-question, by raising this issue not just from Geolocation privacy to privacy on the Web, but to academic involvement in these sorts of policy issues.

- technical standards bodies are made up almost exclusively of engineers from corporations; don't reflect the diversity of the population
- many of the decisions they make have profound policy implications, but never get put to a public vote and are largely considered just from this single engineering point of view

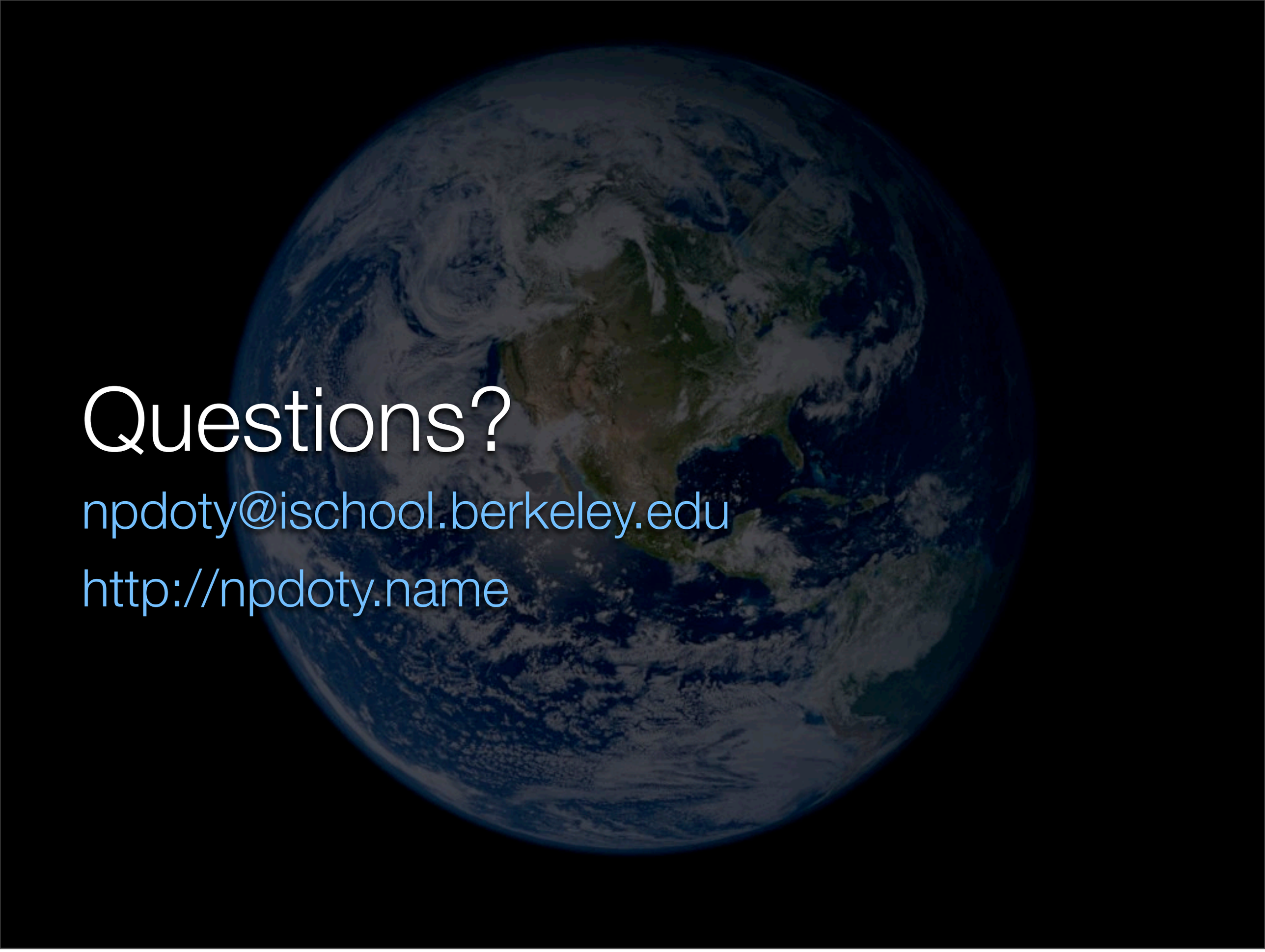
- not much better the other way around: lawmakers rarely understand the technical realities that their legislation affects

- solutions to these challenging technical/policy problems -- like privacy -- require a nuanced understanding of technology, policy, HCI, sociology, etc.; we think academics, particularly in interdisciplinary fields like the School of Information or like TRUST can make a needed, valuable contribution



- furthermore, this is a pedagogical opportunity: invaluable hands-on case studies for students to experience the variety of factors (technological, political, legal, etc.) characteristic of an information professional, issues they'll encounter once they're at Google, Microsoft, etc.
- this spring Professor Deirdre Mulligan and I will offer a "lab" course for technology and policy issues where we will:
  - continue research like what I presented to you today
  - get students participating in technical standards bodies
  - work on legislative analysis for relevant draft bills in Congress

- if we find this lab work to be successful, in the long term, we may consider a center at the iSchool to support this sort of research and participation from masters, PhDs and faculty throughout the campus



Questions?

npdoty@ischool.berkeley.edu

http://npdoty.name