Preemption, Privacy, and People
Legal and Policy Foundations for the Long-term Success of Transportation Technologies

Jan Whittington

Urbanism Next
May 9, 2019
Oregon Convention Center
Portland, OR
Governing Transport Data

Trends in Transport Data Governance

- Public agencies want data, need questions answered
- Private firm: business competitiveness, public data requests
  "It's just analytics: I'll give this problem to another firm"
- Data standards address (some) issues

Trusted Data Problem

- Protecting the data (security)
- Protecting the data subjects (privacy)

Shifting Legal Landscape

- Is there a right to privacy in public space?
- Carpenter case: Data on our devices deserve privacy
- The future of privacy for location data
### Americans Hold Strong Views About Privacy in Everyday Life

In response to the following question: “Privacy means different things to different people today. In thinking about all of your daily interactions – both online and offline – please tell me how important each of the following are to you . . .

% of adults who say ... 

<table>
<thead>
<tr>
<th>Privacy Aspect</th>
<th>Very Important</th>
<th>Somewhat Important</th>
<th>Not Very Important</th>
<th>Not at All Important</th>
<th>Don’t Know</th>
<th>NET Important</th>
<th>NET Not Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>Being in control of who can get info about you</td>
<td></td>
<td></td>
<td>74%</td>
<td>19%</td>
<td>311</td>
<td>93%</td>
<td>4%</td>
</tr>
<tr>
<td>Being able to share confidential matters with someone you trust</td>
<td></td>
<td></td>
<td>72%</td>
<td>21%</td>
<td>211</td>
<td>93%</td>
<td>4%</td>
</tr>
<tr>
<td>Not having someone watch you or listen to you without your permission</td>
<td></td>
<td></td>
<td>67%</td>
<td>20%</td>
<td>8 12</td>
<td>88%</td>
<td>9%</td>
</tr>
<tr>
<td>Controlling what information is collected about you</td>
<td></td>
<td></td>
<td>65%</td>
<td>25%</td>
<td>5 11</td>
<td>90%</td>
<td>6%</td>
</tr>
<tr>
<td>Not being disturbed at home</td>
<td></td>
<td></td>
<td>56%</td>
<td>29%</td>
<td>9 22</td>
<td>85%</td>
<td>11%</td>
</tr>
<tr>
<td>Being able to have times when you are completely alone, away from anyone else</td>
<td></td>
<td></td>
<td>55%</td>
<td>30%</td>
<td>9 22</td>
<td>85%</td>
<td>10%</td>
</tr>
<tr>
<td>Having individuals in social/work situations not ask you things that are highly personal</td>
<td></td>
<td></td>
<td>44%</td>
<td>36%</td>
<td>13 2 4</td>
<td>79%</td>
<td>15%</td>
</tr>
<tr>
<td>Being able to go around in public without always being identified</td>
<td></td>
<td></td>
<td>34%</td>
<td>29%</td>
<td>25 6 4</td>
<td>63%</td>
<td>31%</td>
</tr>
<tr>
<td>Not being monitored at work</td>
<td></td>
<td></td>
<td>28%</td>
<td>28%</td>
<td>22 6 15</td>
<td>56%</td>
<td>27%</td>
</tr>
</tbody>
</table>


PEW RESEARCH CENTER
Unique in the Crowd: The privacy bounds of human mobility

Yves-Alexandre de Montjoye1,2, César A. Hidalgo1,2,4, Michel Verleysen2 & Vincent D. Blondel2,5

1Massachusetts Institute of Technology, Media Lab, 20 Ames Street, Cambridge, MA 02139 USA, 2Université catholique de Louvain, Institute for Information and Communication Technologies, Electronics and Applied Mathematics, Avenue Georges Lemaitre 4, B-1348 Louvain-la-Neuve, Belgium, 3Harvard University, Center for International Development, 79 JFK Street, Cambridge, MA 02138, USA, 4Instituto de Sistemas Complejos de Valparaíso, Paseo 21 de Mayo, Valparaíso, Chile, 5Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.

We study fifteen months of human mobility data for one and a half million individuals and find that human mobility traces are highly unique. In fact, in a dataset where the location of an individual is specified hourly, and with a spatial resolution equal to that given by the carrier’s antennas, four spatio-temporal points are enough to uniquely identify 95% of the individuals. We coarsen the data spatially and temporally to find a formula for the uniqueness of human mobility traces given their resolution and the available outside information. This formula shows that the uniqueness of mobility traces decays approximately as the $1/10$ power of their resolution. Hence, even coarse datasets provide little anonymity. These findings represent fundamental constraints to an individual’s privacy and have important implications for the design of frameworks and institutions dedicated to protect the privacy of individuals.

Derived from the Latin Privatus, meaning “withdraw from public life,” the notion of privacy has been foundational to the development of our diverse societies, forming the basis for individuals’ rights such as free speech and religion, for example. Despite its importance, privacy has often relied on informal rules.
Municipalities and Data

Cities **PUSH** data to an open data portal

People **PULL** data out (e.g., public disclosure requests)

Cities **SPILL** data (e.g., external or internal breaches)

Challenges

Protecting privacy in data

+ Low bar of re-identifiability (mathematics)
+ Threat varies geographically (land use and density matter)
+ Query-based solutions (narrow mosaic ‘attack surface’)

Mitigating bias in data

+ What are the populations represented in the data?
+ The purpose of data collection & use, from public point of view
+ The private intent of its use

Need for PIAs or risk/benefit analyses

+ Should have a low tolerance for risk (merge technical & legal protection)
+ Inform of scenarios/trajectories for re-identifiability and harm
+ Recognize business decisions of public-facing firms
+ Recognize public need for evidence-based policy
Legal Landscape

Pre-emption v. Cities as laboratories of policy
- Cities as sites of experimentation in the public interest
- Cities are locations of concentrated market interest for firms
- Complex environments of cities elevate risk of conflict
- Opportunity to strike a balance at local level
- State and Fed backstop, elevate, provide expertise

Public Records Acts v. Right to Privacy
- Detailed data can answer policy questions (OD, routes, demo)
- Detailed data, better argument for injunction against release
- Public requests for release are perceived in the public interest
- Conflict: Publicly released data likely to be re-identifiable
Universities as Third Parties

Role for Universities

Repositories for private and public data research:
+ Research to advance methods of privacy protection (location data, genetics)
+ Data sharing and use agreements to specify mutual interests
+ Corporate affiliate agreements, enforceable research agreements
+ Products for public consumption that respond to policy questions
Focus, not FOMO

Public Agencies: What are the policy/compliance questions that need to be answered?

- Collecting data specific to the needs of public agencies
  + The needs of public agencies change over time
  + Standardized aggregates work for some questions (NACTO)
  + Query specific synthetic datasets if re-identifiable (OD, routes)
  + Some questions require merging private and public datasets

Firms: What are the allowable uses of the data?

- Research, expertise, governance of data repositories
  + Privacy protection, public reporting, permit compliance
  + Merged public and private datasets
  + Stopgap against push, pull, and spill
Roles for Universities

Protecting privacy in data
+ Low bar of re-identifiability (mathematics)
+ Threat varies geographically (land use and density matter)
+ Need query-based solutions (narrow mosaic ‘attack surface’)

Mitigating bias in data
+ What are the populations represented in the data?
+ The purpose of data collection & use, from public and subjects' POV

Need for analytics and governance
+ Merge technical & legal protection
+ Two levels of risk – to University, to society (if University does not act)
+ Inform of scenarios/trajectories for re-identifiability and harm
+ Develop new institutional rules of the game
A linked data repository with strong data governance.

https://www.uwtdc.org/

Center of Research in Privacy, Fairness, Accountability, Transparency Coordinated -- WA State AGO, IRB, CPO, CISO, Office of Public Records
Hosting Heterogeneous Data

Simple Architecture

Mobility data is intrinsically heterogeneous – multiple secure endpoints

- Triage hosts any files in any format for data that needs TDC protections but has not yet been parsed and processed.
- Lake hosts json for data that does not necessarily conform to a standard API, but can be represented in a semi-structured data model.
- Warehouse hosts structured data uploaded through one of several standard APIs.
For More Information

Jan Whittington
Associate Professor, Urban Design and Planning
University of Washington

janwhit@uw.edu

https://www.uwtdc.org/