

# Ethics

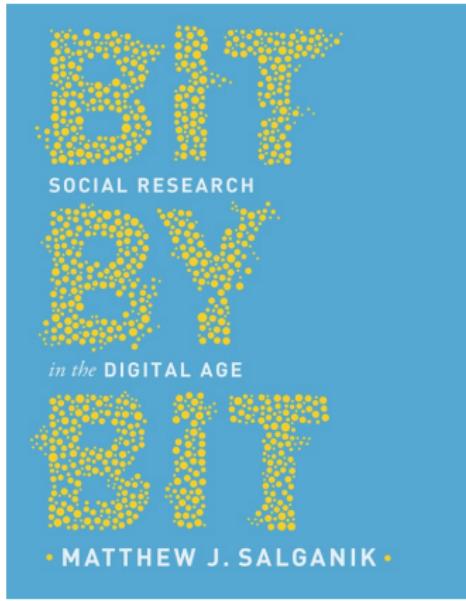
## Part 1, Part 2, Additions and extensions

Matthew J. Salganik  
Department of Sociology  
Princeton University

Summer Institutes in Computational Social Science  
Day 1  
2020

The Summer Institutes in Computational Social Science is supported by grants from the Russell Sage Foundation, Alfred P. Sloan Foundation, Facebook, and the Social Science Research Council.





- 1) Introduction
- 2) Observing behavior
- 3) Asking questions
- 4) Running experiments
- 5) Mass collaboration
- 6) Ethics
- Part 1, Part 2, [Additions and extensions](#)
- 7) The future

## Ethics, part 1 and part 2

- ▶ Introduction
- ▶ Three examples
- ▶ Digital is different
- ▶ Four principles
- ▶ Two ethics frameworks
- ▶ Four areas of difficulty
- ▶ Practical tips

## Addition and extension 1 of 5: Respect for Law and Public Interest

Four principles:

- ▶ Respect for persons
- ▶ Beneficence
- ▶ Justice
- ▶ Respect for Law and Public Interest

## Terms-of-service agreements

# **MapWatch: Detecting and Monitoring International Border Personalization on Online Maps**

Gary Soeller  
Northeastern University  
soelgary@ccs.neu.edu

Karrie Karahalios  
University of Illinois at Urbana-Champaign  
kkarahal@illinois.edu

Christo Wilson  
Northeastern University  
cbw@ccs.neu.edu

Christian Sandvig  
University of Michigan  
csandvig@umich.edu

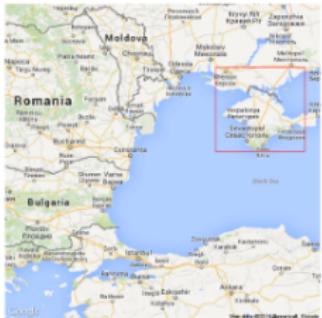
<http://dx.doi.org/10.1145/2872427.2883016>

## Abstract:

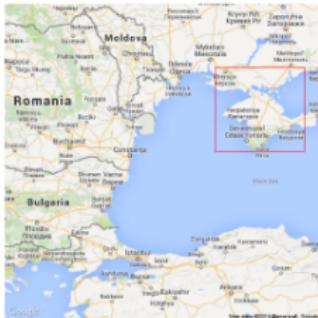
"Maps have long played a crucial role in enabling people to conceptualize and navigate the world around them. However, maps also encode the world-views of their creators. Disputed international borders are one example of this: governments may mandate that cartographers produce maps that conform to their view of a territorial dispute. Today, online maps maintained by private corporations have become the norm. However, these new maps are still subject to old debates. Companies like Google and Bing resolve these disputes by localizing their maps to meet government requirements and user preferences, i.e., users in different locations are shown maps with different international boundaries. We argue that this non-transparent personalization of maps may exacerbate nationalistic disputes by promoting divergent views of geopolitical realities."

## Abstract, part 2:

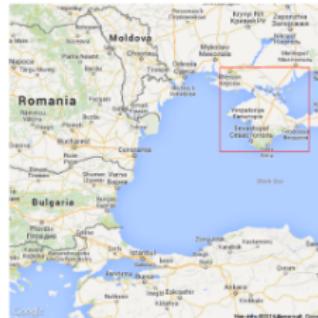
“To address this problem, we present MapWatch, our system for detecting and cataloging personalization of international borders in online maps. Our system continuously crawls all map tiles from Google and Bing maps, and leverages crowdworkers to identify border personalization. In this paper, we present the architecture of MapWatch, and analyze the instances of border personalization on Google and Bing, including one border change that MapWatch identified live, as Google was rolling out the update.”



(a) Russia



(b) Ukraine



(c) All countries except Russia and Ukraine

Figure 5: Border changes separating the Crimean Peninsula from Ukraine on Google Maps. Map Data ©2014 AutoNavi Google.

<http://dx.doi.org/10.1145/2872427.2883016>

**Ethics.** Many Web platforms discourage all automated crawling outside the API, and even within the API a platform may prohibit “research” in its online Terms of Service document (usually for competitive reasons). We agree with [57] that non-commercial research for the public good that deals with issues of societal importance must be able to access public Web resources for research purposes as long as automated processes do not produce an unreasonable load. This was our guiding philosophy in this research design.

<http://dx.doi.org/10.1145/2872427.2883016>

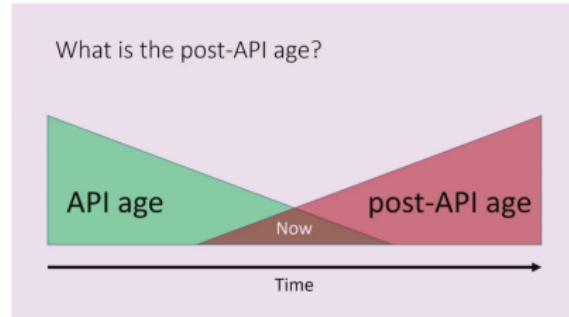
Researchers (with the support of the ACLU) have filed a case challenging the CFAA, Sandvig v Lynch:

[https://www.aclu.org/cases/  
sandvig-v-lynch-challenge-cfaa-prohibition-uncovering-racial-discrimination-online](https://www.aclu.org/cases/sandvig-v-lynch-challenge-cfaa-prohibition-uncovering-racial-discrimination-online)

Even if this is legal should we do it?

Even if this is legal should we do it?

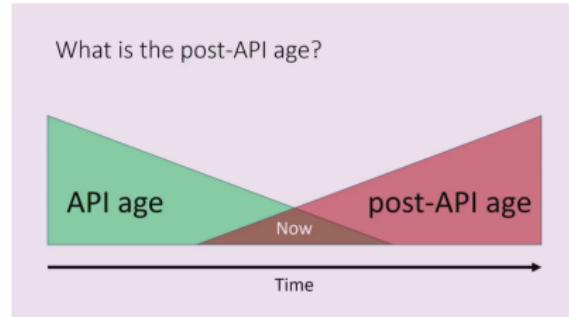
Deen Freelon at SICSS 2018: “[Surviving the post-API age](#)”



If you go “off the grid”:

Even if this is legal should we do it?

Deen Freelon at SICSS 2018: “[Surviving the post-API age](#)”

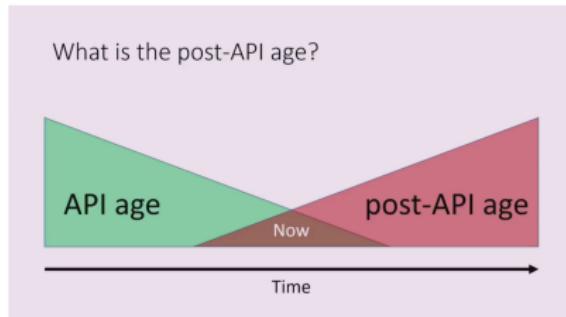


If you go “off the grid”:

- ▶ you might lose access during your research

Even if this is legal should we do it?

Deen Freelon at SICSS 2018: “[Surviving the post-API age](#)”

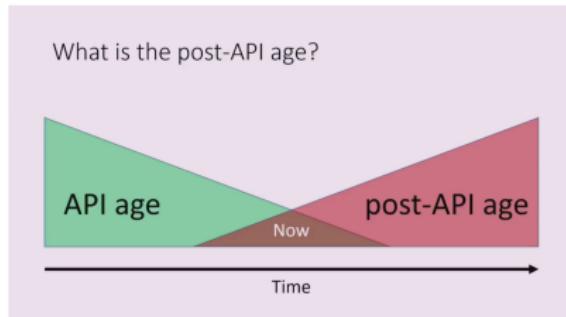


If you go “off the grid”:

- ▶ you might lose access during your research
- ▶ you might struggle to have your research funded, talk about it, and publish it

Even if this is legal should we do it?

Deen Freelon at SICSS 2018: “[Surviving the post-API age](#)”

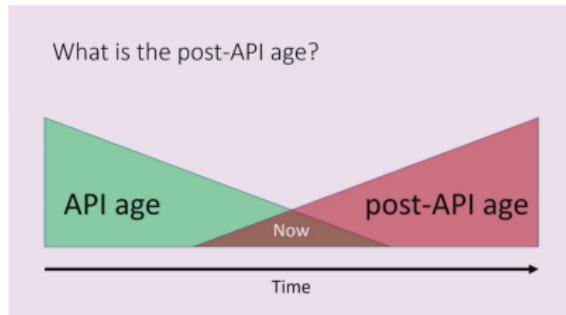


If you go “off the grid”:

- ▶ you might lose access during your research
- ▶ you might struggle to have your research funded, talk about it, and publish it
- ▶ you might not be able to share your data with other researchers

Even if this is legal should we do it?

Deen Freelon at SICSS 2018: “[Surviving the post-API age](#)”



If you go “off the grid”:

- ▶ you might lose access during your research
- ▶ you might struggle to have your research funded, talk about it, and publish it
- ▶ you might not be able to share your data with other researchers
- ▶ you might make it harder to other academics in the future

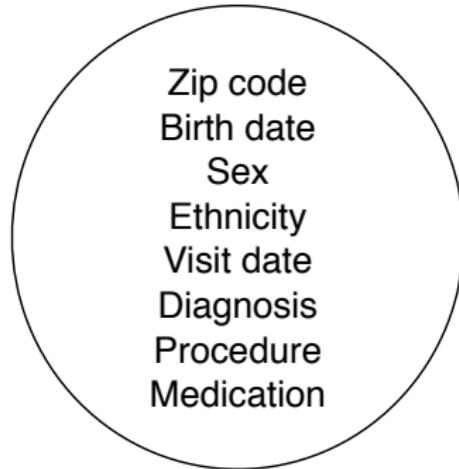
## Addition and extension 2 of 5: Informational risk

Areas of difficulty:

1. Informed consent
2. Informational risk
3. Privacy
4. Making decisions in the face of uncertainty

- ▶ Simple idea: Data can be made anonymous, and we can tell what data is sensitive

- ▶ Simple idea: Data can be made anonymous, and we can tell what data is sensitive
- ▶ Better idea: All data are potentially identifiable and all data are potentially sensitive



De-identified  
medical data

Sweeney (2002)

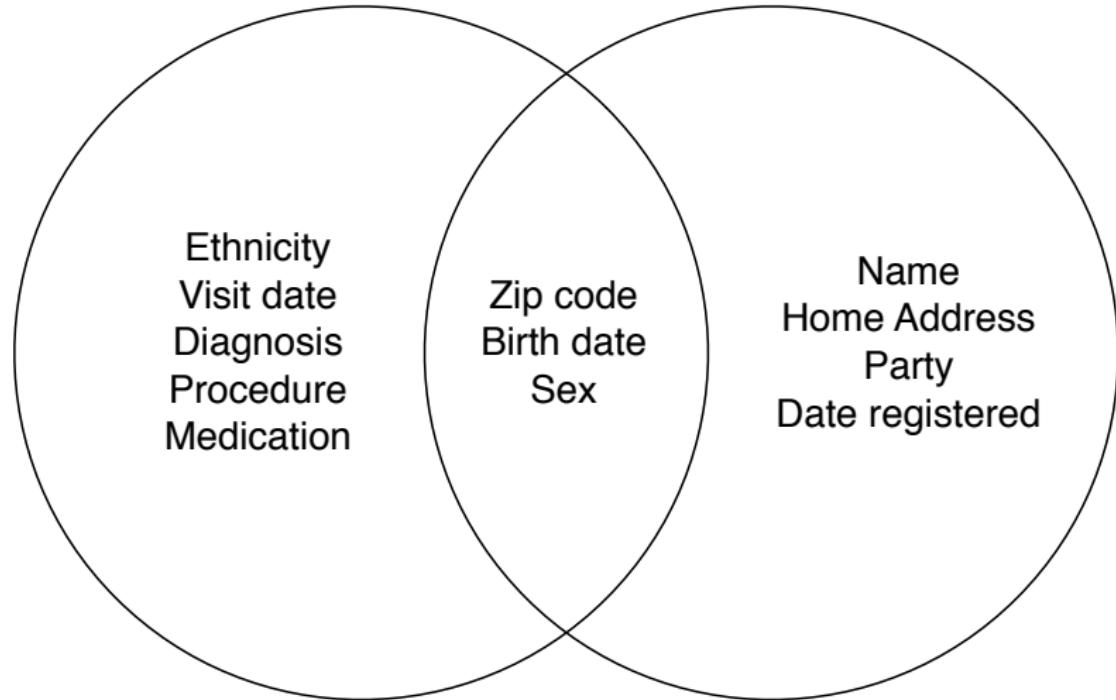
- Zip code
- Birth date
- Sex
- Ethnicity
- Visit date
- Diagnosis
- Procedure
- Medication

De-identified  
medical data

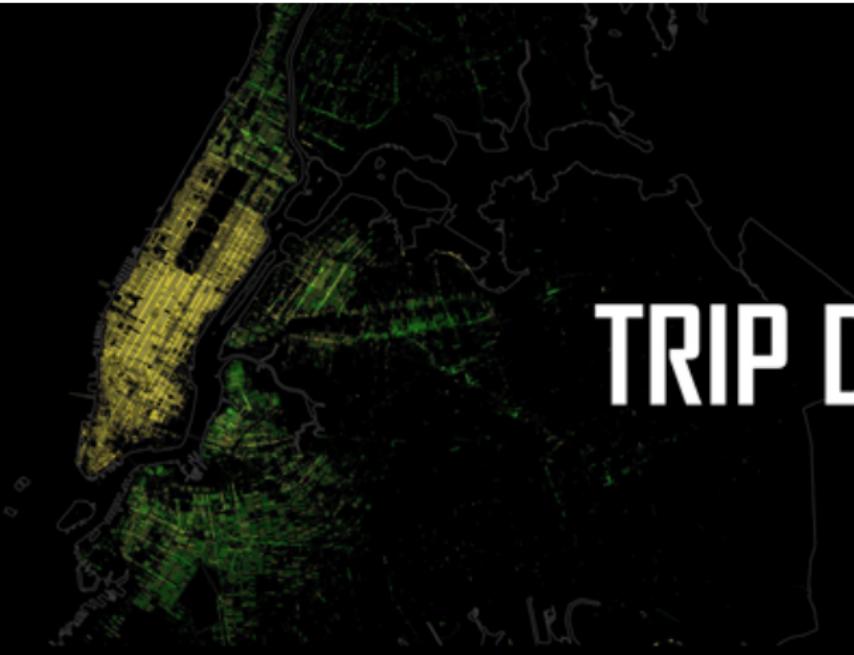
- Name
- Home Address
- Zip code
- Birth date
- Sex
- Party
- Date registered

Voter registration data

Sweeney (2002)



Re-identified medical data



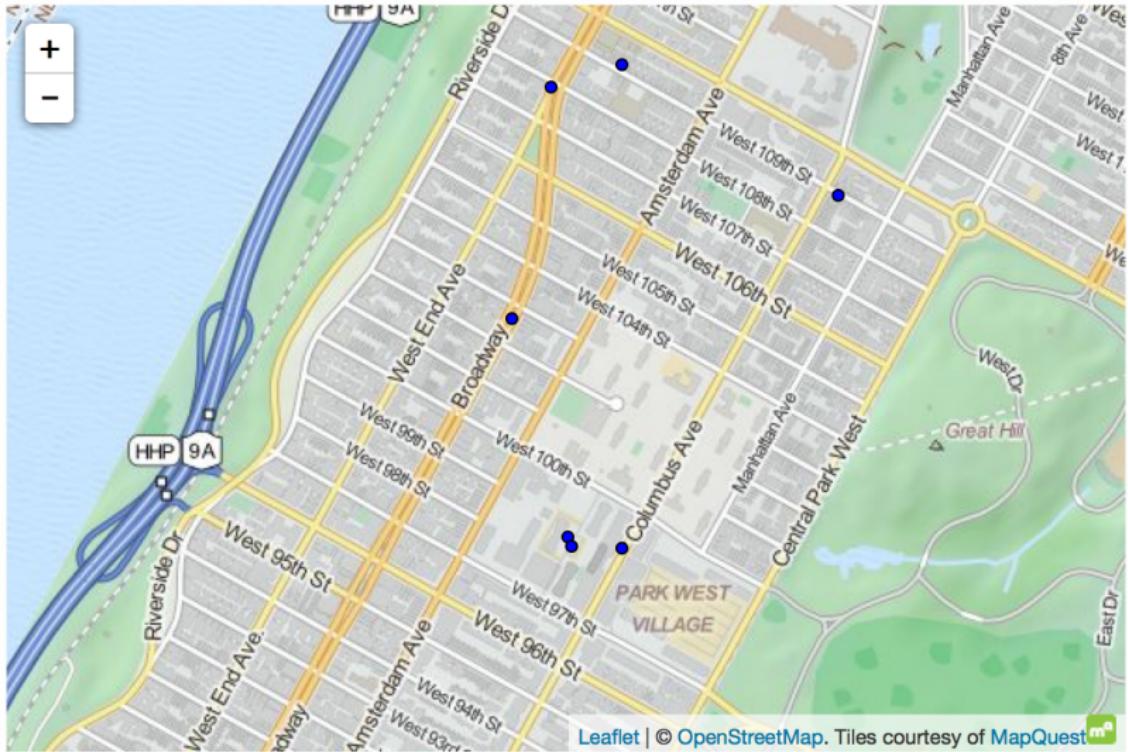
# TLC TRIP DATA





<http://research.neustar.biz/2014/09/15/>

[riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset](#)



[http://research.neustar.biz/2014/09/15/  
riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/](http://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/)

Data:

<https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

Data:

<https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

Code:

```
SELECT dropoff_latitude , dropoff_longitude
FROM tripData
WHERE pickup_lat > 40.767249 AND pickup_lat < 40.768
    AND pickup_lon > -73.996782 AND pickup_lon < -73.995538
    AND HOUR(pickup_datetime) < 6
    AND trip_distance > 5;
```

# Aol.



4417749, “numb fingers”

4417749, “numb fingers”

4417749, “60 single men”

4417749, “numb fingers”

4417749, “60 single men”

4417749, “dog that urinates on everything”

4417749, “numb fingers”

4417749, “60 single men”

4417749, “dog that urinates on everything”

4417749, “landscapers in Lilburn, Ga”

4417749, "numb fingers"

4417749, "60 single men"

4417749, "dog that urinates on everything"

4417749, "landscapers in Lilburn, Ga"

4417749, [several people with the last name Arnold]

4417749, "numb fingers"

4417749, "60 single men"

4417749, "dog that urinates on everything"

4417749, "landscapers in Lilburn, Ga"

4417749, [several people with the last name Arnold]

4417749, "homes sold in shadow lake subdivision gwinnett county  
georgia"

---

TECHNOLOGY

# *A Face Is Exposed for AOL Searcher No. 4417749*

---

By MICHAEL BARBARO and TOM ZELLER Jr. AUG. 9, 2006

<https://www.nytimes.com/2006/08/09/technology/09aol.html>

## Addition and extension 3 of 5: Fifth area of difficulty

Areas of difficulty:

1. informed consent
2. informational risk
3. privacy
4. making decisions in the face of uncertainty

## Areas of difficulty:

1. informed consent
2. informational risk
3. privacy
4. making decisions in the face of uncertainty
5. unanticipated secondary uses

eBird

[Submit](#)   [Explore](#)   [My eBird](#)   [Science](#)   [About](#)   [News](#)   [Help](#)

## Discover a new world of birding...

**► Learn more**

## Get started



Kirtland's Warbler *Setophaga kirtlandii* © Marky Mutchler Macaulay Library eBird

USERS

# Plants and Birds Need Privacy Online, Too

Our enthusiasm for sharing birds, plants, and superblooms has unintended consequences. But we can use the same tools that made the problem to fix it.

By APRIL GLASER

APRIL 04, 2019 • 3:56 PM

[https://slate.com/technology/2019/04/  
superbloom-california-nature-internet-collide-birds-poaching.html](https://slate.com/technology/2019/04/superbloom-california-nature-internet-collide-birds-poaching.html)

## Sensitive Species in eBird

eBird has recently been altered to better protect Sensitive Species, because **some bird species face risks including capture, targeted killing, or significant disturbance, and for these species eBird data output is restricted in some ways to protect them, while allowing important data to continue to come into eBird.**

<https://help.ebird.org/customer/en/portal/articles/2885265-sensitive-species-in-ebird>



- ▶ How would Lex Luthor use this?
- ▶ This requires adversarial thinking.

[https://en.wikipedia.org/wiki/Lex\\_Luthor#/media/File:LexLuthor1.png](https://en.wikipedia.org/wiki/Lex_Luthor#/media/File:LexLuthor1.png)

## Addition and extension 4 of 5: Practical tips

## Practical tips:

- ▶ IRB is a floor not a ceiling
- ▶ Put yourself in everyone else's shoes
- ▶ Think of research ethics as continuous not discrete

## Practical tips:

- ▶ IRB is a floor not a ceiling
- ▶ Put yourself in everyone else's shoes
- ▶ Think of research ethics as continuous not discrete
- ▶ Think of ethics as a research opportunity

# ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)

A computer science conference with a cross-disciplinary focus that brings together researchers and practitioners interested in fairness, accountability, and transparency in socio-technical systems.

<https://facctconference.org/>

Addition and extension 5 of 5: Think of ethics as a research opportunity, case study

# Privacy, ethics, and data access: A case study of the Fragile Families Challenge\*

Ian Lundberg<sup>a</sup>, Arvind Narayanan<sup>b</sup>, Karen Levy<sup>c</sup>, and Matthew J. Salganik<sup>a</sup>

In press at *Socius*, pre-print:

<https://arxiv.org/abs/1809.00103>

"Stewards of social science data face a fundamental tension. On one hand, they want to make their data accessible to as many researchers as possible to facilitate new discoveries. At the same time, they want to restrict access to their data as much as possible in order to protect the people represented in the data. In this paper, we provide a case study addressing this common tension in an uncommon setting: the Fragile Families Challenge, a scientific mass collaboration designed to yield insights that could improve the lives of disadvantaged children in the United States. We describe our process of threat modeling, threat mitigation, and third-party guidance. We also describe the ethical principles that formed the basis of our process. We are open about our process and the trade-offs that we made in the hopes that others can improve on what we have done."

In press at *Socius*, pre-print:

<https://arxiv.org/abs/1809.00103>

## Ethics, part 1 and part 2

- ▶ Introduction
- ▶ Three examples
- ▶ Digital is different
- ▶ Four principles
- ▶ Two ethics frameworks
- ▶ Four areas of difficulty
- ▶ Practical tips

## Additions and extensions

- ▶ 1 of 5: Respect for Law and Public Interest
- ▶ 2 of 5: Informational risk
- ▶ 3 of 5: Fifth area of difficulty
- ▶ 4 of 5: Practical tips
- ▶ 5 of 5: Think of ethics as a research opportunity, case study

Thank you