



# Privacy Risk and Preservation for **COVID-19** Contact Tracing and Information Sharing

Jan 29, 2021

**Fang Liu**

Applied and Computational Mathematics and statistics

University of Notre Dame


# Data Privacy (Protection) Day

- Jan 28
- First time: 2007
- Observed in United States, Canada, Israel, and 47 European countries.
- Educational and Raises awareness among individuals, families, and businesses, about the importance of protecting privacy
- Encourages compliance with privacy laws and regulations
- Promotes events and activities that stimulate the development of technology tools for privacy protection
- Creates dialogues among stakeholders interested in advancing data protection and privacy.

# Sponsor



Committee on Privacy and Confidentiality

- <https://community.amstat.org/cpc/home>
-  @ASACPC1
- 2020: Differential Privacy and the 2020 Census: Modernizing Disclosure Avoidance at Scale to Mitigate Growing Privacy Threats by Michael Hawes
- 2019: Toward Protecting the Privacy of Individuals When Disseminating Data: Challenges in Disclosure Risk Assessment by Tom Krenzke and Jianzhu Li
- 2018: What is a Privacy Loss Budget and How Is It used to Design Privacy Protection for a Confidential Database? by John M. Abowd.

# CHANCE

Using Data to Advance Science, Education, and Society

Articles

Columns

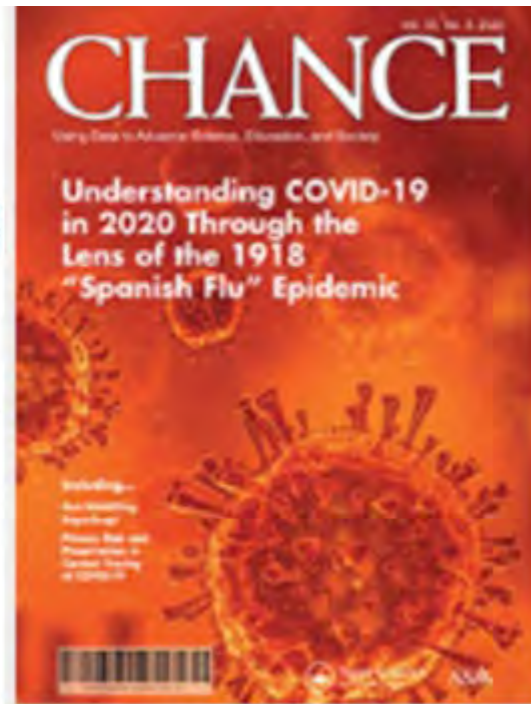
Editor's Letter

Letters to the Editor

## Privacy Risk and Preservation in Contact Tracing of COVID-19

• [Columns: O Privacy, Where Art Thou](#)

*Dong Wang and Fang Liu*



Talk will be posted @ <https://community.amstat.org/cpc/home>

# Outline

- I. Public awareness and perception of privacy during the pandemic
- II. Data collected during the pandemic and the associated privacy risk and privacy protection
- III. Application of formal privacy concepts for data collection and sharing during the pandemic
- IV. Conclusions

# I. Public Awareness and Perception of Privacy during the Pandemic

# Survey Results

## Pew Research Center (April, 2020)

- 60% say it wouldn't make much of a difference in limiting the spread of COVID if the government tracked people's locations through their cellphones
- 52% say it would be at least somewhat acceptable for the government to use people's cellphones to track the location of people who have tested positive for COVID
- 45% says it is acceptable for the government to use cellphones to track the location of people who may have had contact with someone who tested positive for COVID.

# Survey Results

Zhang et al on public's attitudes toward surveillance measures (Jun, 2020)

- 62% support enforcing temperature checks
- 57% support expanding traditional contact tracing
- 49% support carrying out centralized quarantine
- 44% support deploying electronic device monitoring
- 44% implementing immunity passes
- 42% support the government encouraging everyone to download and use contact tracing apps
- Decentralized data architecture increases the likelihood of people downloading apps by 5.4%



# Survey Results

## CISCO 2020 Consumer Privacy Survey (June 2020)

- Top privacy concerns:
  - 31% say the data will be used for unrelated purposes
  - 25% say the data will be shared too broadly with third parties
  - 24% say the data will not be deleted or anonymized when no longer needed to combat COVID
- Permissible data sharing during the pandemic
  - 57%: Medical checks and questionnaires for workplace safety
  - 49%: Location/contact tracking
  - 37%: Information about infected people

# Survey Results

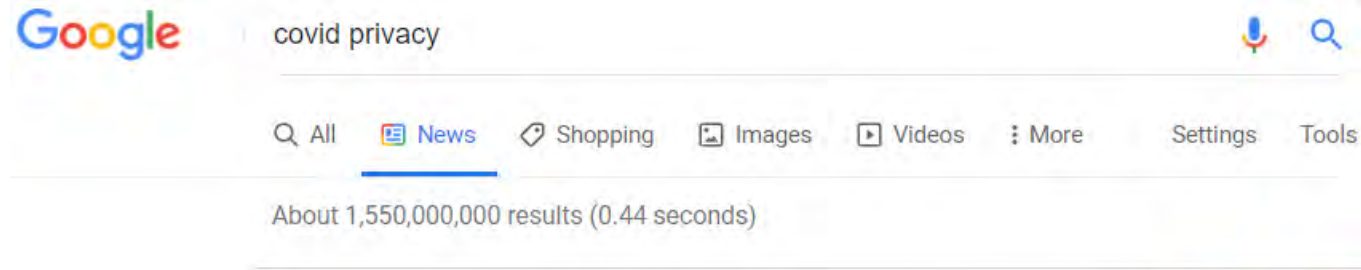
## Cisco 2021 Data Privacy Benchmark Study

- 93% organizations turned to their privacy teams to help navigate these challenges and 90% report privacy metrics to their C-suites and boards
- 87% consumers expressed concerns about the privacy protections of the tools they needed to use to work, interact and connect remotely
- More than 140 jurisdictions have now passed omnibus privacy laws, and ~80% of respondents found these laws to have positive impacts.
- 57% support employers using data to help make workplaces safe.
- <50% support location tracking, contact tracing, disclosing info about infected individuals, or using individual information for research

# In Summary

- There are big privacy concerns regarding personal data collection and sharing during the pandemic
- Limited trust in government or businesses in handling privacy in collected data during the pandemic
- Some willingness to share data to help control the spread of COVID, should the privacy rights be respected.


# Search on



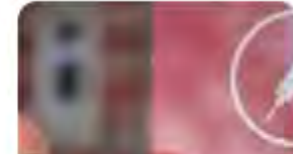
## The year we gave up on privacy

Our li

- 🔍 covid
- 🔍 covid
- 🔍 covid
- 🔍 covid
- 🔍 covid
- 🔍 covid
- 🔍 covid
- 🔍 covid
- 🔍 covid
- 🔍 covid

 The Guardian

German Covid vaccine officials forced to guess people's ages from names




... a C WSJ The Wall Street Journal  
as an

[WHO Plans Privacy, Security Rules for Covid-19 Vaccine](#)



...  
The World Health Organization is working on technical details and privacy


 Security Boulevard

Data Privacy Day: Understanding COVID-19's Impact

But we also know that COVID-19 has reshuffled the protocols and procedures for cybersecurity and data privacy, making compliance more ...

3 mins ago



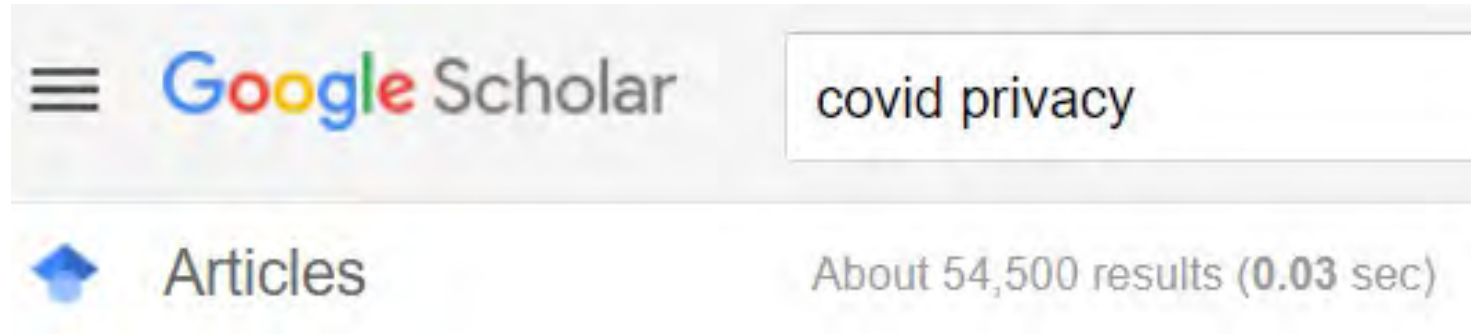
 Sixth Tone

Chinese Cities Vow to Better Protect COVID-19 Patients' Privacy



12 pm on Jan 25, 20  
About 6,400,000,000

# Search on Google Scholar



As of 2pm Jan 24, 2021,

Nature journals, The Lancet, IEEE journals, Informatics journals

CS, business and marketing, public policy, law, statistics, etc

## II. Data Collected and Released during the Pandemic and the Associated Privacy Risk and Privacy Protection

# What COVID-related data are collected by governments during the pandemic?

- Case surveillance data
- Public health surveillance data (screening, surveillance, and diagnostic testing data)
- Contact tracing data
- Pandemic-specific surveys (e.g. Household Pulse Survey by the census)
- Others (vaccination records...)

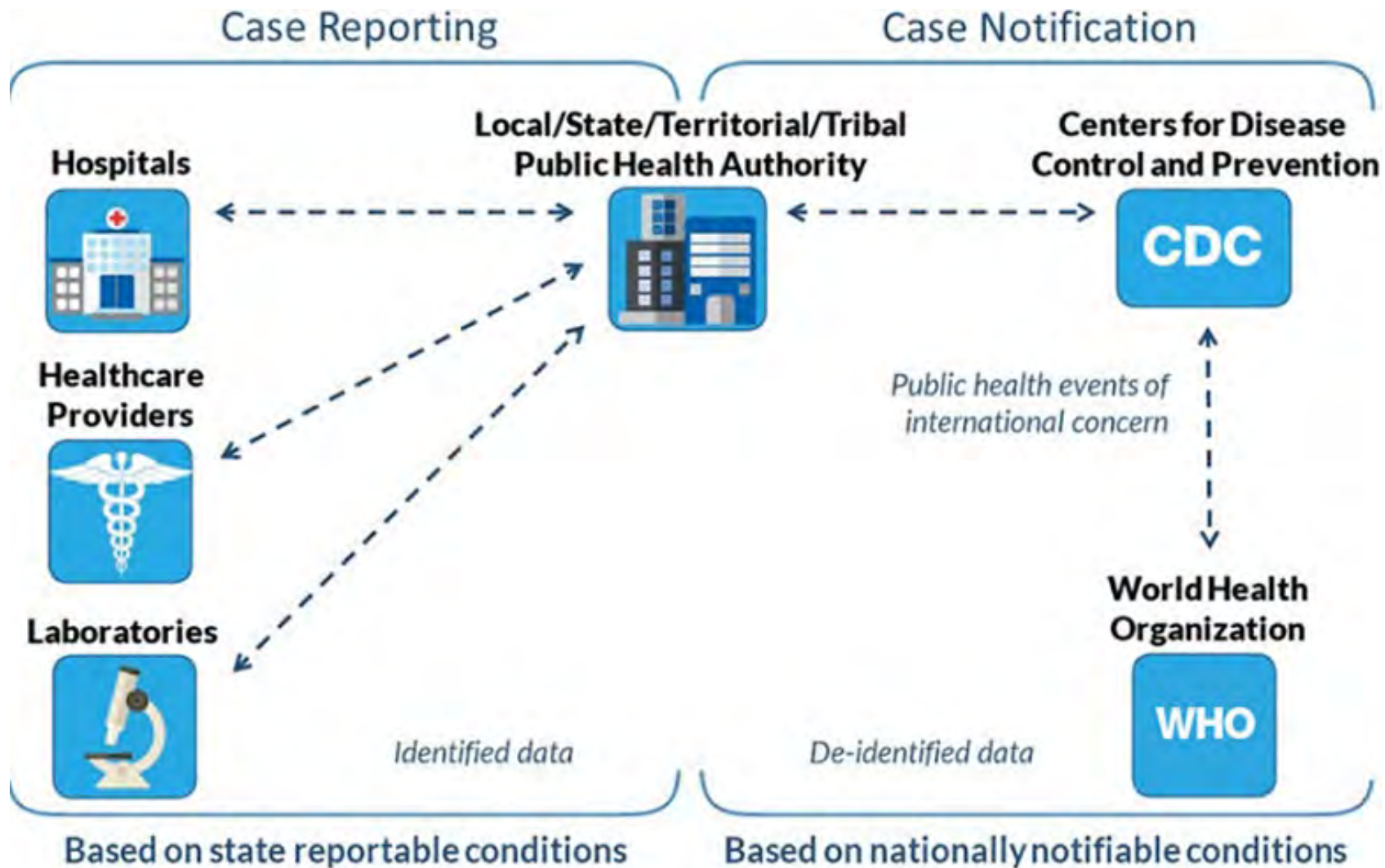
# Case Surveillance Data

- are important for understanding similarities and differences among cases in their demographic, clinical, and epidemiologic characteristics; exposure and contact history; care received, etc.
- help track the spread of COVID and identify areas of concern and groups at risk
- help develop guidance for the public, at-risk groups, and healthcare providers.

In the US, to **protect individuals' privacy**, case data are shared with CDC **without personal identifiers** (e.g., names or addresses) by the state and jurisdictional health departments **voluntarily**. But at the state/local levels, cases may be identifiable.



# Case Surveillance Data



# Leaking too much personal info?



Additionally, today (Monday, November 2), I was made aware that a student has tested positive for the COVID-19 virus. The student began having symptoms on [REDACTED] and has not been at school since [REDACTED]. The student was tested on Wednesday, [REDACTED] and received the positive results on Sunday, [REDACTED]. The student is in grade [REDACTED] and rides bus # [REDACTED].

# Leaking too much personal info?

- Early Dec 2020
- Chengdu, China
- Information of a confirmed case was released on the internet
- Last name, age, gender, occupation, coarsened residence info, clinical diagnosis, travel trajectory (beauty salon, movie theater, several restaurants, several bars on a single night) before diagnosis

患者1：赵某，女，20岁，无固定职业，居住于成华区崔家店华都云景台小区，系郫都区昨日确诊病例卢某、赵某的孙女。12月8日在对卢某密切接触者的隔离检测中检出核酸阳性。经省、市、区专家会诊，综合临床、影像学表现和实验室核酸检查结果，诊断为确诊病例（普通型），已转运至成都市公共卫生临床医疗中心医院隔离治疗。近14天内，患者主要停留的场所包括：中冶中央公园、嗨蓝调美甲店、小巷巷麻辣烫、海雾里小酒吧、playhouse酒吧、赫本酒吧等。

# Leaking too much personal info?

- The patient was doxed and cyberbullied regarding her social life and behaviors.
- (Fake) news, (fake) information, (fake) photos started to circulate about this patient on the internet and social media



# Public Health Surveillance Data

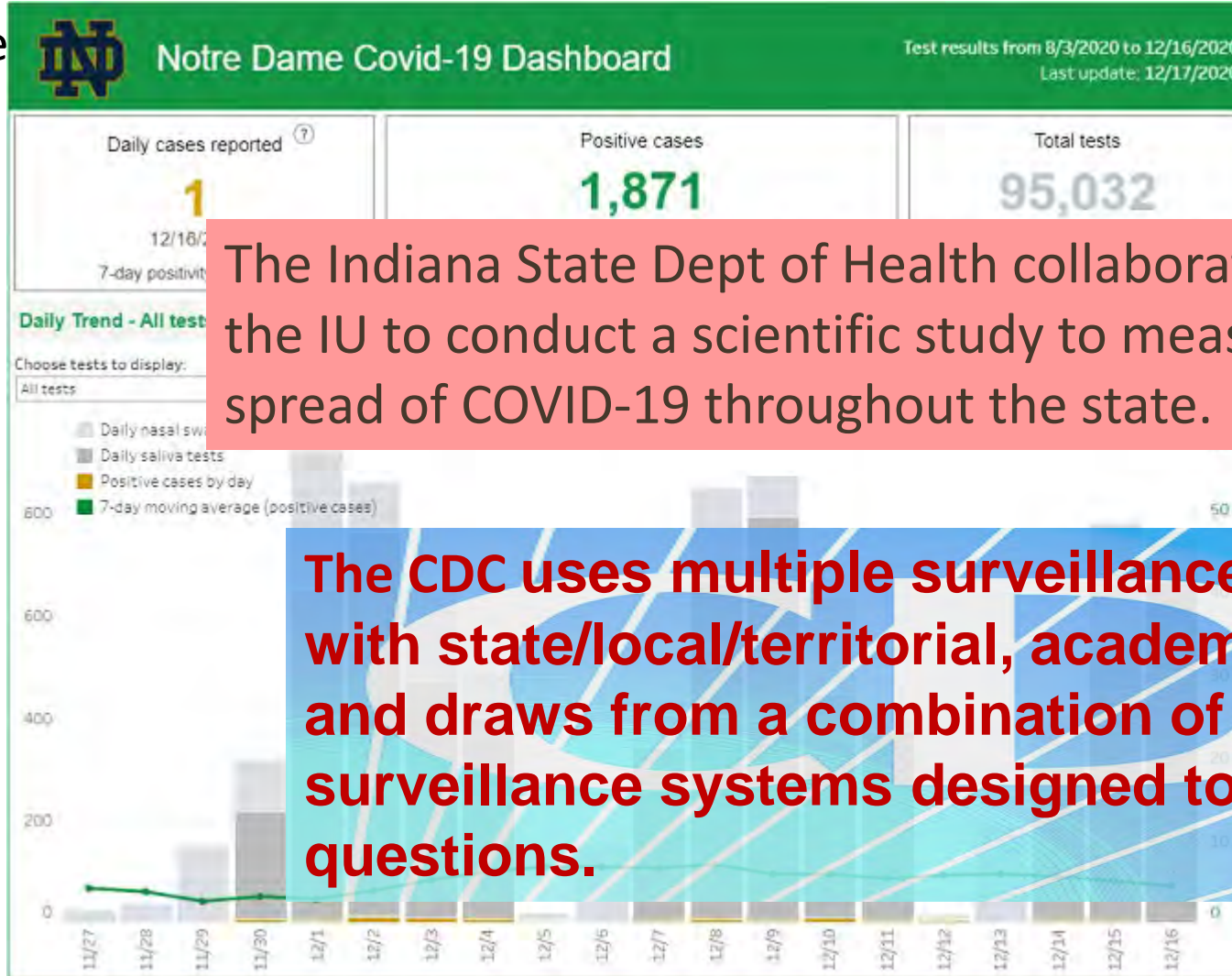
Surveillance is systematic collection, analysis, and interpretation of data to evaluate public health and plan and implement strategies.

During the pandemic, surveillance data are used to

- understand disease severity
- identify risk factors
- monitor and forecast the spread of COVID
- understand how COVID impacts the capacity of healthcare systems
- etc.

# Public Health Surveillance Data

Surveillance



different levels

The Indiana State Dept of Health collaborates with the IU to conduct a scientific study to measure the spread of COVID-19 throughout the state.

The CDC uses multiple surveillance systems in collaboration with state/local/territorial, academic, commercial partners; and draws from a combination of data sources and new surveillance systems designed to answer specific questions.

# Public Health Surveillance Data

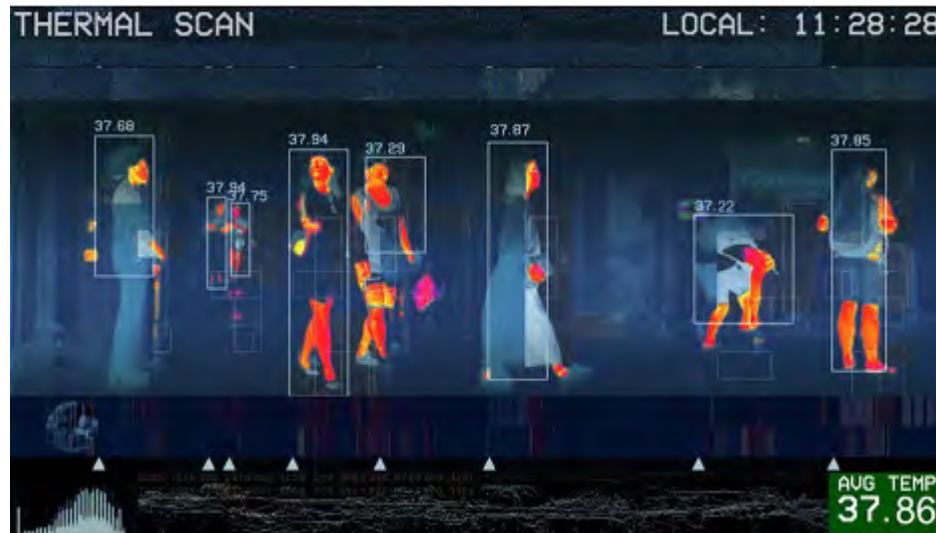
## China: Alipay Health Code

- Users enter personal info (eg, ID, phone numbers, addresses, places of work) and where, when and how they entered the region, etc.
- A series of health-related questions
- Based on the responses, a color-code (red, yellow or green) is generated – corresponding to their health status and level of risk for COVID-19.



# Public Health Surveillance Data

- **Taiwan and Singapore** use sensors (e.g. thermal imaging cameras and infrared sensors) in public spaces (e.g., airports, bus shelters, train stations) to identify potential cases.
- **Liechtenstein** uses bracelets that send data on vital biometrics (e.g., skin temperature, breathing rate, heart rate) to a Swiss laboratory for analysis.



courtesy of  
govtech.com





# Public health surveillance privacy concerns

- Often contain personal identifiable and sensitive info
- may reveal details about a person's lifestyle, behaviors, and health.
- concerns the data being used for other purposes and the surveillance system might be here forever.

Authorities in Shanghai consider integrating **a personal health index into** an app that ranks citizens on indicators such as how much they sleep, smoke, and drink and how many daily steps they take.

In the UK, **Serco, SITEL, and Amazon Web Services** have access to users data, which, reportedly, will be **held for 20 years**.

Some governments (e.g., Israel, Kenya, Mexico, Turkey) have reportedly used the pandemic as an opportunity to **analyze telecommunication** data

# Public health surveillance privacy protection

- **Aggregated statistics** from surveillance data are often shared with the public.
- The **International Covenant on Civil and Political Rights (ICCPR)** created a legal imperative to ensure that no one is “*subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to unlawful attacks on his honor and reputation.*”
- The **privacy-concerned Electronic Frontier Foundation** questions the accuracy of thermal cameras and argues that the surveillance infrastructure being built up now might become permanent.
- Seattle requires new technology operated by the city to go through a **rigorous vetting process** that includes public comments and participation from the city CTO and council.

# Contract Tracing

- Contract tracing can help identify potential transmission hotspots and measure effectiveness of public health interventions
- Contract tracing can be manual or digital. At least 40 countries are using contact tracing apps.



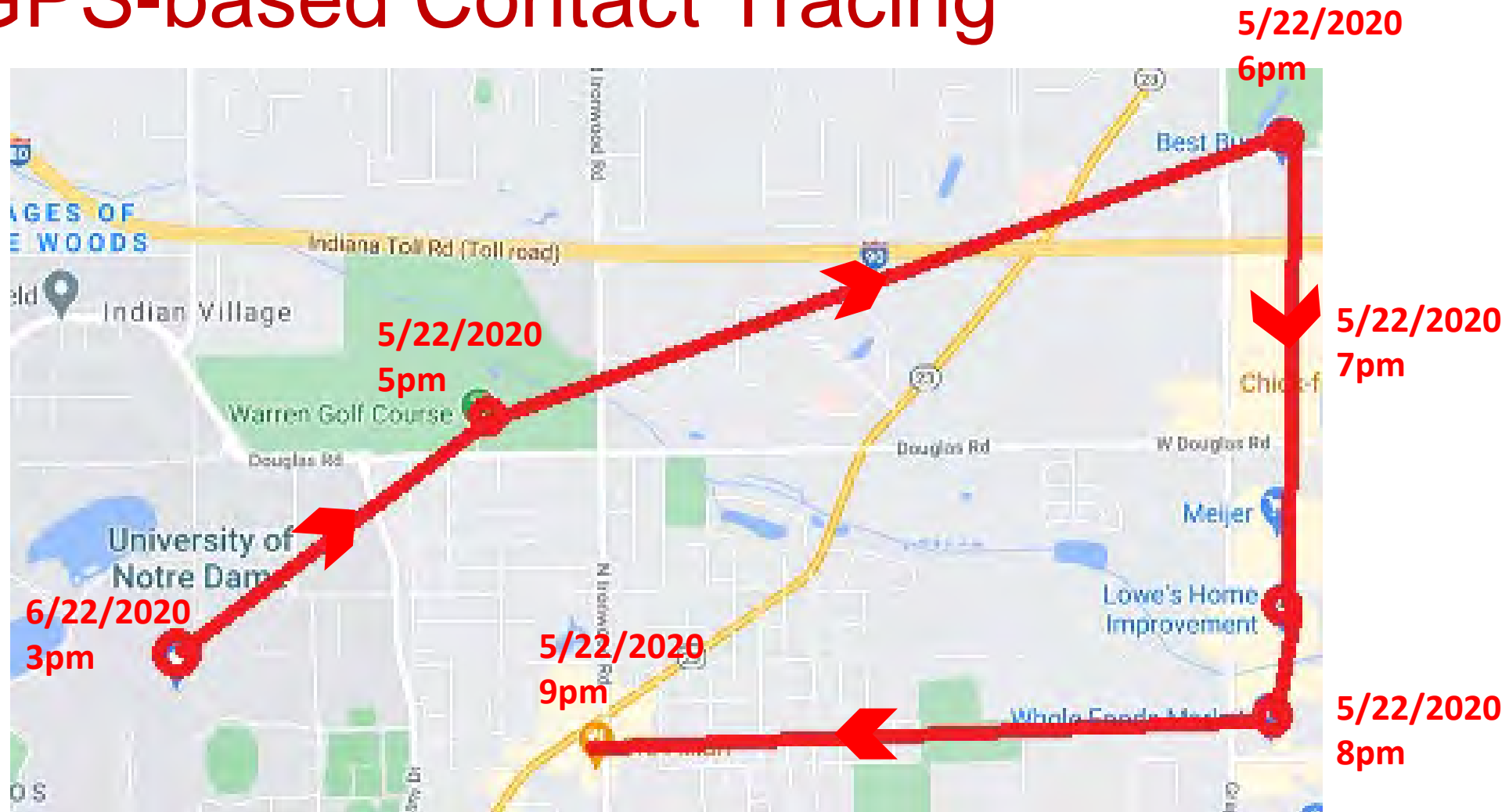
# Manual Contract Tracing



# Some Examples of Digital Contact Tracing

		technology	
		GPS (location based)	Bluetooth (proximity based)
<b>Model for data collection, storage, sharing</b>	<b>Centralized</b> (central server stores and process contact logs, and location info if GPS-based, send out notification)	<ul style="list-style-type: none"> <li>• Health Code (China)</li> <li>• Corona 100m (South Korea)</li> </ul>	<ul style="list-style-type: none"> <li>• TraceTogether (Singapore)</li> <li>• Pan-European Privacy-Preserving Proximity Tracing (PEPP-PT) (EU)</li> </ul>
	<b>Decentralized</b> (contact logs are processed locally; center server does not store or process location or contact info)	safe paths (US) (GPS+BT)	<ul style="list-style-type: none"> <li>• COVID watch (international)</li> <li>• DP-3T (EU)</li> <li>• Google-Apple Exposure Notification (US)</li> </ul>

# GPS-based Contact Tracing



# China Health Code (C+GPS)

- QR code-based quarantine apps during the early stages of the outbreak.
- Collects a vast amount of users' data — location data, self-reported medical history — before labeling each citizen with a risk score and issuing a colored health code

• 随申码/Shanghai QR Code



北京健康码/ Beijing Health kit



## 杭州健康码

Hangzhou Health Code



【绿码】  
凭码通行



【黄码】  
实施7天内隔离，连续  
(不超过)7天健康打卡正常  
转为绿码



【红码】  
实施14天隔离，连续14天  
健康打卡正常转为绿码

# South Korea Corona 100m (Co100) (C+GPS)

- Launched on February 11; had a million downloads in its first 10 days, according to SK government website Korea.net,
- Alerts users when they come within 100 meters of a location visited by an infected person.
- It is reported that health investigators will also have access to more data (e.g., surveillance camera footage and credit card transactions).





# Safe Paths (DC, GPS+BT)

An MIT-led, open source technology that aims to **maximize privacy** and **the effectiveness of contact tracing** in the case of a positive diagnosis.



Comprises and

- a web app **Safe Places**, where public health officials can broadcast location info of anonymized, redacted, and blurred location history of infected patients.
- a smartphone app **PrivateKit** that enables users to check if they have crossed paths with infected patients by comparing their personal location diary with the published redacted locations of patients

# Safe Paths: privacy protection strategy



Locally, GPS records are

- 1) mapped to a 3D discretized grid (point intervals/ PIs)
- 2) The PIs are further obfuscated with a hash function.

- Diagnosed patients share their **redacted, anonymized, hashed** PIs with a semi-trusted server (e.g. local health agency).
- App users periodically check the server to see if their hashed PIs match those shared by patients.
- Only the **app users know the matching results, not the server**; i.e., contact logs are never transmitted to or stored in the server.

# Bluetooth Based Technology

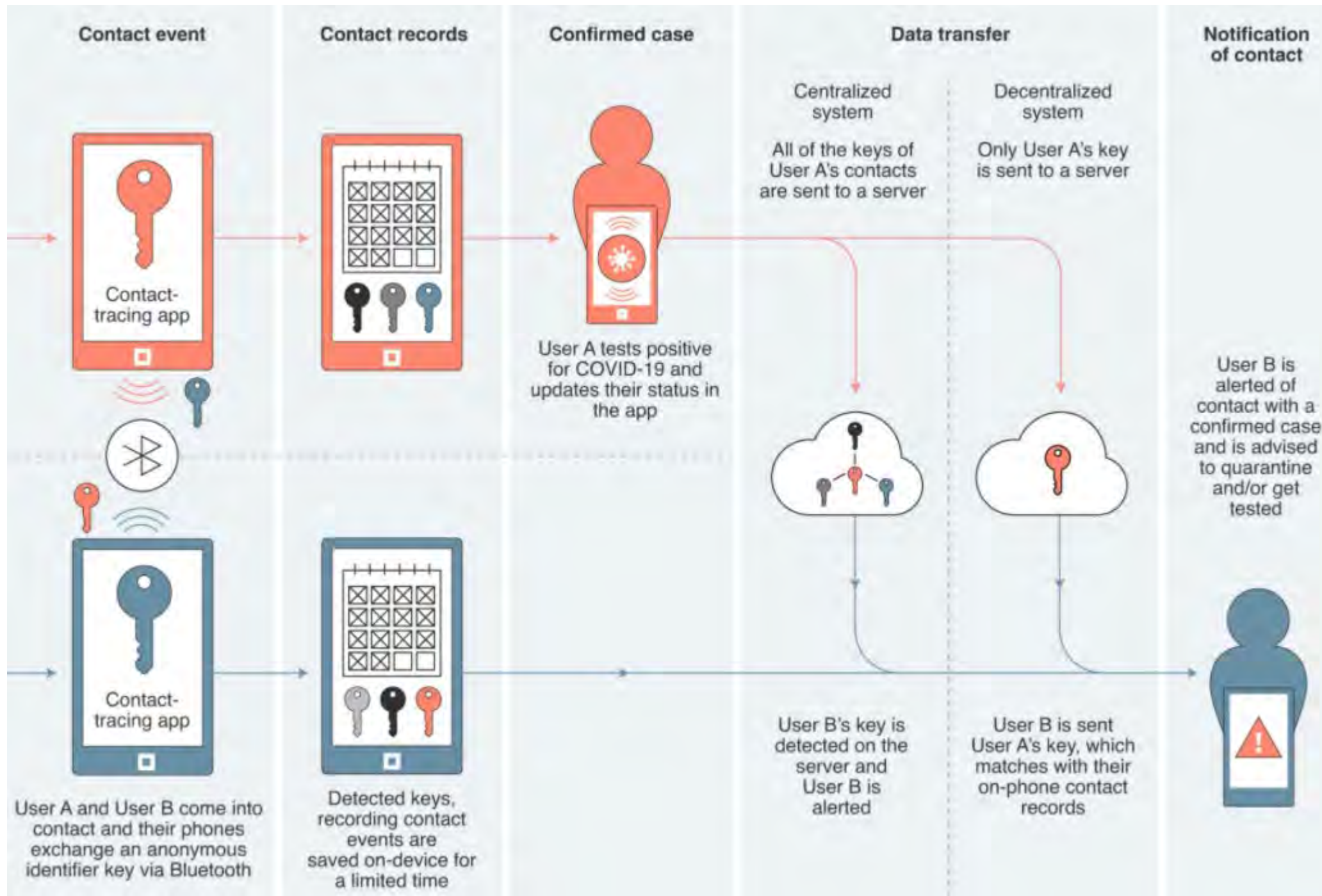
Bluetooth-based contact tracing apps do not collect location info from their users, but whether 2 people have appeared within 6 feet of each other at the same time (**proximity-based tracing**)

## Centralized model

- There is a central server that processes and stores contact log data, and notifies clients of potential contact with an infected patient.
- One of its advantages is that it's easier to audit the system and adapt it more quickly if needed.

## Decentralized model

- Better privacy protection compared to the centralized model
- Contact logs are processed and stored locally, but this comes with a higher cost for the computing power on the user side.



Adapted from Fig 2 In Budd et al. (2020)

# Examples of Centralized Bluetooth Model

Pan-European Privacy-Preserving  
Proximity Tracking



- Voluntary download and usage
- Data collected and stored in a secure server, and but never shown to the public
  - A random anonymized user ID
  - contact/mobile number
  - identification details
- With user's consent, the app exchanges encrypted and anonymized Bluetooth signals with nearby devices
- Data after 25 days is automatically deleted

# Examples of decentralized Bluetooth model

Decentralized Privacy-Preserving  
Proximity Tracing



华为接触卫士 (Huawei Contact Shield)

Google/Apple Exposure Notification (GAEN) system

In the US, as of 1/23/21, GAEN is implemented in 21 states, each has a different name (e.g., COVID Alert MI, COVIDaware MN, SlowCOVIDNC, GuideSafe, DC CAN, AlohaSafe Alert, etc) with a different logo

- 500K+ (2): NY, CA
- 100K+ (13): MI, WI, PA, MN, VA, MD, NC, AL, WA, NJ, CT, CO, NE
- 10K+ (5): DC, DE, ND/WY, HI, GUAM
- 1K+ (1): LA

Review score: 3 to 4.5 with a median of 3.6



# MI COVID Alert



## MI COVID Alert Exposures

### Exposure Notifications

Exposure Notifications are turned ON

### Past Exposures

No Exposures Detected

If this app detects an exposure that meets MDHHS's criteria, you will receive a push notification and additional information will be displayed here.

Read our [Privacy Policy](#).

MI COVID Alert version 1.2

- Exposures
- Notify Othe...
- Virtual MDHHS
- Share
- Stats

## Virtual MDHHS

Greetings! How can I assist?

I have a general inquiry

I received a positive COVID-19 test

I received a notification of possible exposure

### Share

The more people who join the fight against COVID-19, the more effective our Exposure Notification efforts will become!

Share the app with others to help us reach our goal.

SHARE MI COVID ALERT

## Notify others

### Share A Positive Result

If you test positive for COVID-19, you can share your result anonymously.

To get a random PIN required to share a result, call your [local health department](#) first. You can also request a PIN through the MI COVID-19 hotline by dialing [2-1-1](#) or [888-535-6136](#).

Other app users who were near you in the past 10 days may be notified of exposure. Your identity will not be shared.

Please self-isolate if you tested positive. Learn more at [www.michigan.gov/ContainCOVID](http://www.michigan.gov/ContainCOVID).

SUBMIT POSITIVE RESULT

## Michigan Coronavirus Stats

Deaths  
14,951

Total Tested  
9,086,109

Testing Positivity  
6.57%

VACCINE TRACKER  
[Learn more about vaccines](#)

Doses Administered  
642,111

VACCINE TRACKER  
[Learn more about vaccines](#)

Doses Administered  
642,111

People Fully Vaccinated  
105,882

% Fully Vaccinated  
1.06%

## Black or African American

14% of population

14% of cases

24% of deaths

## Hispanic or Latino

5% of population

7% of cases

8% of deaths

## American Indian and Alaska Native

1% of population

1% of cases

0% of deaths

## HOSPITALIZATION

This tool contains the most updated information on hospital inpatient bed and ICU bed occupancy rates for last week and the week prior.

Beds (ICU inc.)

Beds Occupied Last Week 61%

2 Weeks Ago 60%

ICU

Beds Occupied Last Week 75%

7 Weeks Ago 76%

COVID-19 Non COVID-19 Empty

# Privacy Risk of Decentralized Model

- On April 16, 2020, the EU started the process of assessing the proposed decentralized systems for compatibility with privacy and data protection laws, including the GDPR.
- On April 17, 2020, the UK's Information Commissioner's Office published an opinion analyzing some decentralized CT protocols, stating that the systems are "*aligned with the principles of data protection by design and by default*".
- Vaudenay (2020) initiated 12 identification attacks of diagnosed people on DP3T. DP3T claims that out of the 12, 8 are also present in centralized systems, 3 do not work, and 1 works but involves physical access to the phone.
- Adversaries might
  - Place phones in fixed locations pick up the Bluetooth signals and reconstruct the movements of a person using the app who's later diagnosed
  - place a Bluetooth receive in places where there are a lot COVID patients and then a Bluetooth transmitter in a crowded place to create a lot of false alarms



# III Applications of Formal Privacy

# Applications of formal privacy concepts

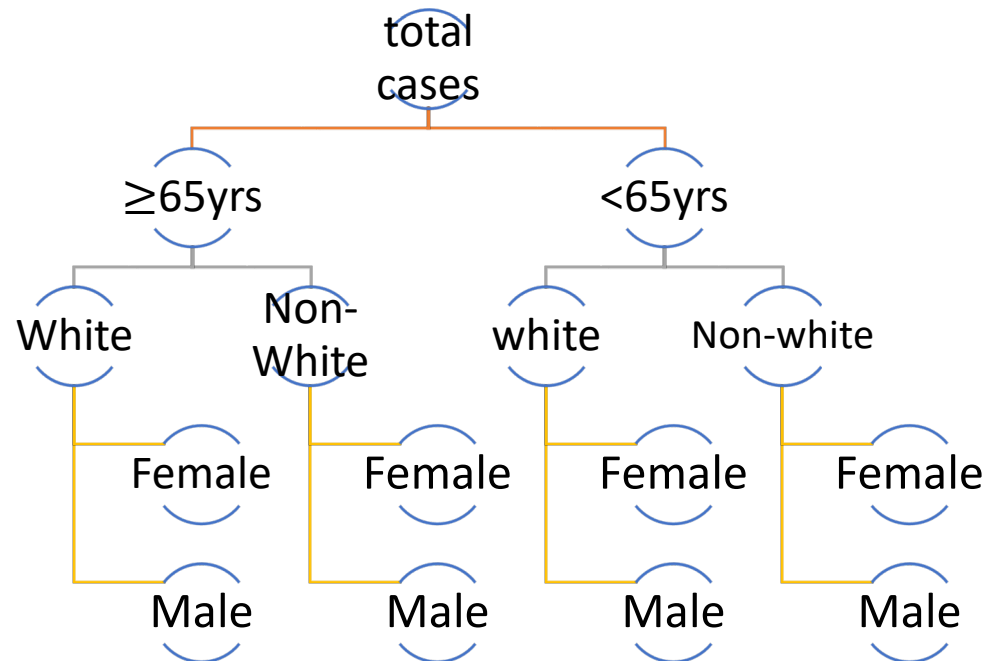
- Balance between privacy protection vs. information accuracy
- Some work exist, most leveraging existing formal privacy methods/techniques
  - Google community mobility report, symptom search trends report, foresting analysis
  - Location privacy (Cao et al.; Lyer et. al.; Vepakomma et al.; Xiong et. al. )
  - Immunity passport (Butler et. al.)
  - Privacy preserving diagnosis using imaging data (Müftüoğlu et al; Ulhaq and Burmeister)
  - Exposure notification (Hammoud and Yu)
- Examples:
  1. Integration of differential privacy (DP) to publish granular case numbers
  2. Usage of randomized response mechanisms or geo-indistinguishability for collecting or publishing contact tracing networks

# Application of DP for info publication

Differential Privacy (DP) (DMKS'06)

$$\frac{\Pr(\mathcal{M}(x))}{\Pr(\mathcal{M}(x'))} \leq e^\epsilon$$

Various relaxed and generalized versions.



Case study 1: publish more granular case numbers [HRMS'10]

1. Sanitize the count in each node
2. Calculate weighted sanitized counts from bottom-up
3. Correct for inconsistency to obtain the final counts from top-down

# Applications of DP for info publication

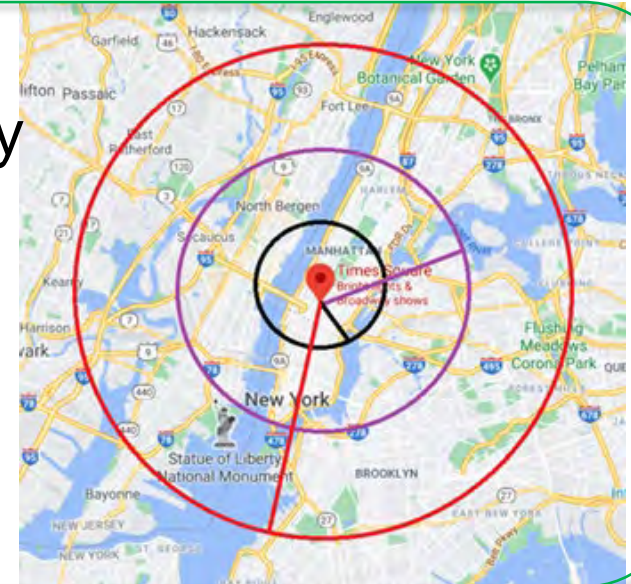
## Case study 2: contact tracing network sanitization



geo-indistinguishability  
[ABCP'13]

$$\frac{p_M(z|x)}{p_M(z|x')} \leq e^{l\epsilon},$$

where  $d(x, x') \leq l$



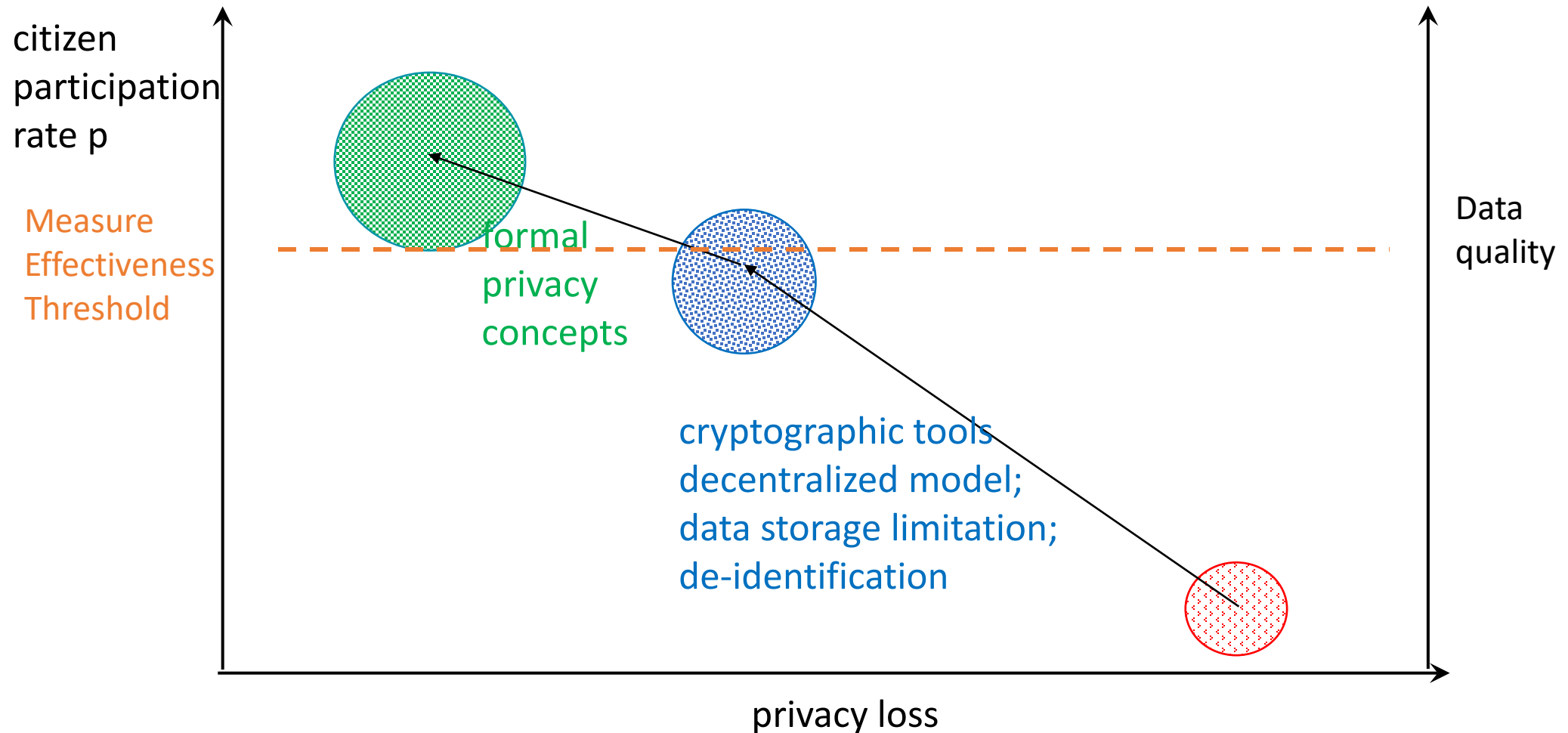
Randomized response to sanitize network edge [KKS'17]

$$\frac{p_M(y_{ij}^* = 1 | y_{ij} = 0)}{p_M(y_{ij}^* = 1 | y_{ij} = 1)} \leq e^{\epsilon_{ij}}$$

$$Pr_M(\text{retaining the original relation } y_{ij}) = \frac{e^{\epsilon_{ij}}}{1 + e^{\epsilon_{ij}}}$$

# IV Conclusions

1. Measures taken authorities to harness COVID is effective only when people trust those measures. Protection of privacy is key to building such trust.



## 2. Proper interpretation and application of the general data protection guidance

- Many privacy acts and regulations (e.g., HIPPA, GDPR) have allowed some flexibility during the pandemic, and does not prevent the processing and disclosure some personal data that is necessary to fight the pandemic.
- Meanwhile, it is important that general principles are respected (e.g., lawfulness, fairness, transparency, data minimization and purpose limitation).

### 3. Find the right balance

between allowing individual rights and freedom (privacy included) vs governments' responsibility managing a global/national crisis

Between individual privacy and freedom vs shared human values

between privacy loss vs information accuracy

limiting instantaneous vs. potential long-term negative impacts

.....



Thank You

