

PDBANK: Utilizing Personal Data for Social and Commercial Benefits While Protecting Privacy

Li, Babu, Rahadiano, Katsushika, Zheng

January 25, 2019

FBL/PBL: Personal Data as "New Oil" and its Market

TEAM MEMBERS

Huiyong Li - D1, Graduate School of Informatics

Sridhar Babu M - M2, Graduate School of Informatics

Haris Rahadiano - M1, Graduate School of Informatics

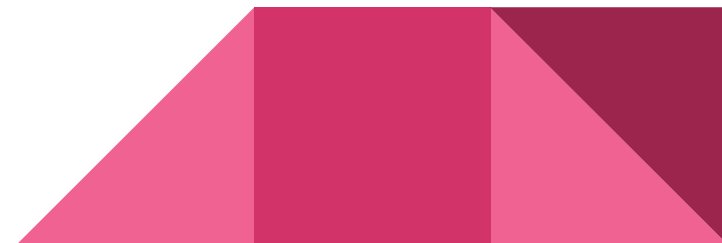
Daisuke Katsushika - M1, Graduate School of Government

Shuyuan Zheng - M1, Graduate School of Informatics

SUPERVISORS

Professor Masatoshi Yoshikawa

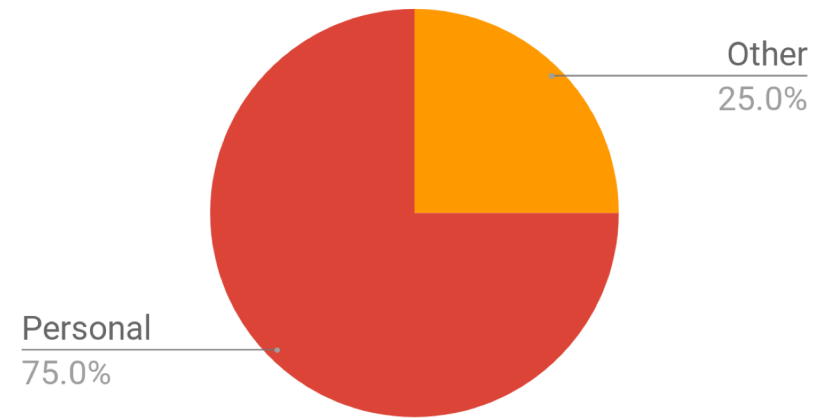
Assistant Professor Yang Cao



Background

- **Vast amounts** of personal data available
 - Purchase histories
 - Trajectories
 - Heartbeat records
- **Technology and business** transformation are at speed
 - Smartphones and Internet of Things (IOT)
 - Personal data economy: € 739 billion by 2020 in the EU

Digital data (175 zettabytes by 2025)



Sources: IDC, Accenture

Opportunity =>

Personal data as “New Oil”



Background

- Mining **benefits** for business and beyond
 - What value could be derived from personal data
 - Not only for business but also for the whole society
- Personal data **privacy**
 - The EU General Data Protection Regulation (GDPR)
 - New technical approaches to protect personal data and privacy

Challenge =>

New infrastructure for social benefits and privacy



Motivation of team and individual

How to utilize personal data for **social and commercial benefits** while **protecting privacy**?

Li

Healthcare and **Education** marketplace

Healthcare and Geosocial marketplace

Babu

Haris

Government use and protection

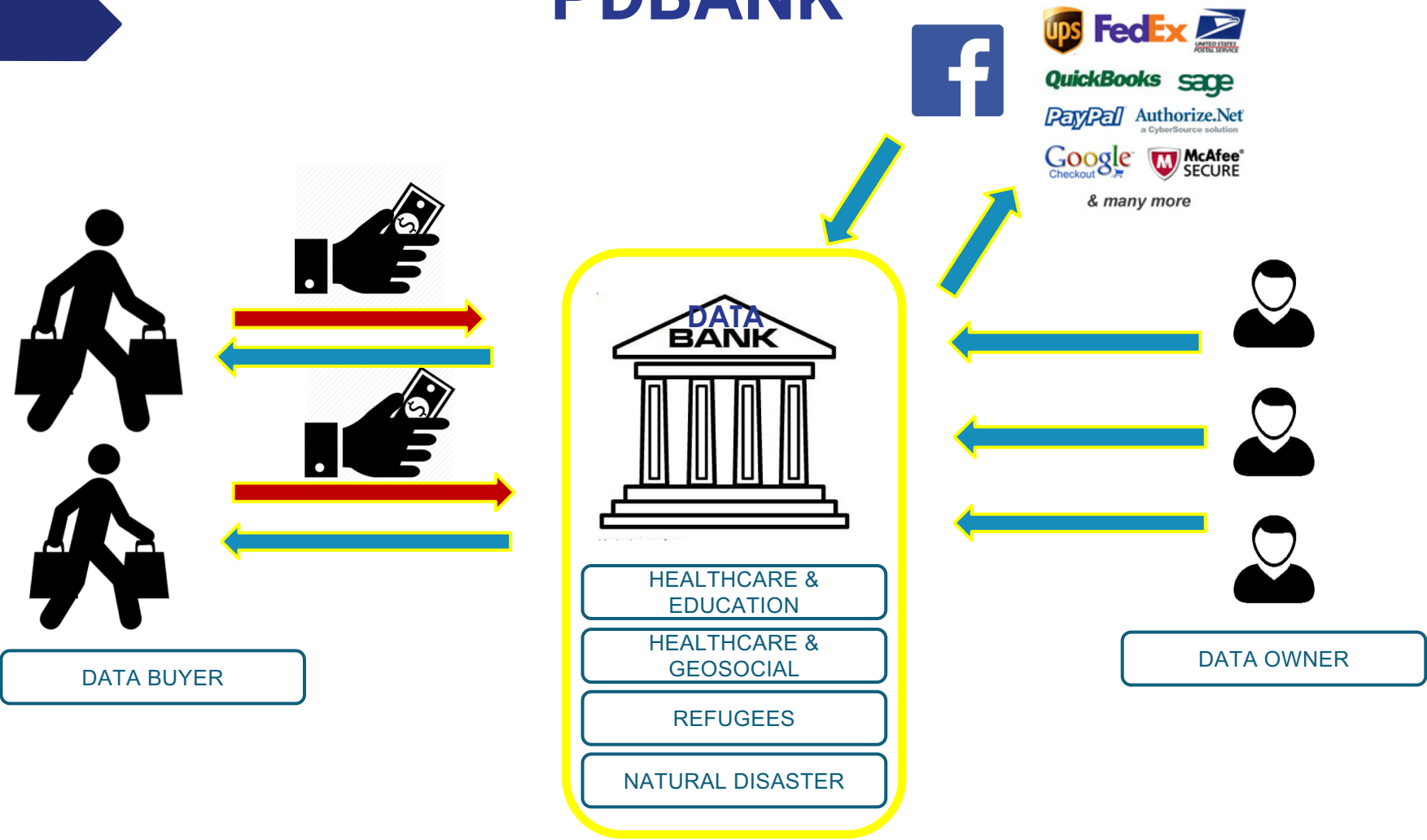
Unsolved **Humanitarian** issues

Katsushika

Zheng

Privacy protection and technology

PDBANK



Multi-Domain Personal Data Bank

PDBANK

- The Person is provided control over monetization of information that is related to the person
- Information may be received and stored by a storage entity, such as **PDBANK**
- Upon confirmation of purchase agreement between a person (**Data Seller**) and a purchaser (**Data Buyer**), the information is made available to the purchaser (**Data Buyer**)
- The person receives value in exchange for the information (**Monetization of Data**)
- The information may be collected and stored for later bulk or incremental monetization at the control of the person (**Data Owner**)

ECONOMY

Japan takes step toward enormous bank of personal data

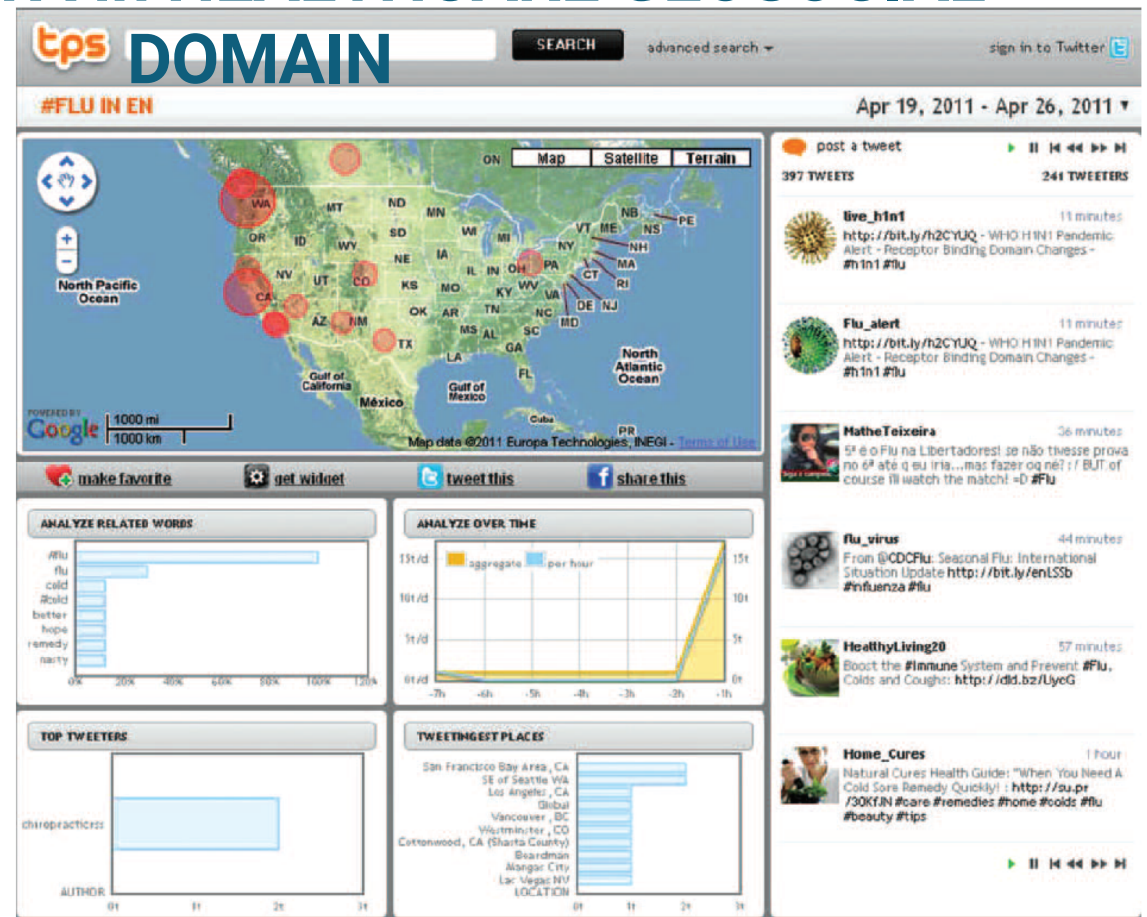
Information would help businesses tailor products, services

FEBRUARY 24, 2017 05:00 JST



PERSONAL DATA in HEALTHCARE GEOSOCIAL

- **Personal Data in Healthcare related GeoSocial Media?**
- Merging the data generated from **GeoSocial application** and **health care domain**
- Eg. Twitter posts corresponding to user, location data, posts in social media by hospital and care centers



Twitter Positioning System (TPS)

Olson, John A. "Health-Related Geospatial Data and Social Media: Can You Harvest Geosocial Data?." *Journal of Map & Geography Libraries* 7, no. 3 (2011): 377-381.

PERSONAL DATA in HEALTHCARE GEOSOCIAL DOMAIN

What You Need to Know:

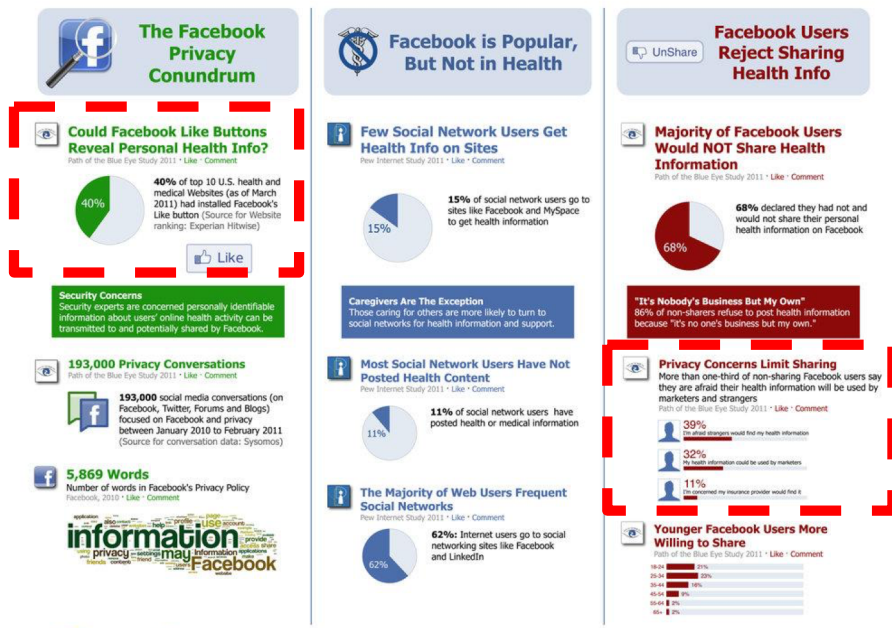
Facebook, Privacy and Health



Facebook and Health Don't Mix

Facebook is dominant, but not in health. Most don't use it to find health information. Fewer share personal health content on the site. Could privacy concerns be the reason?

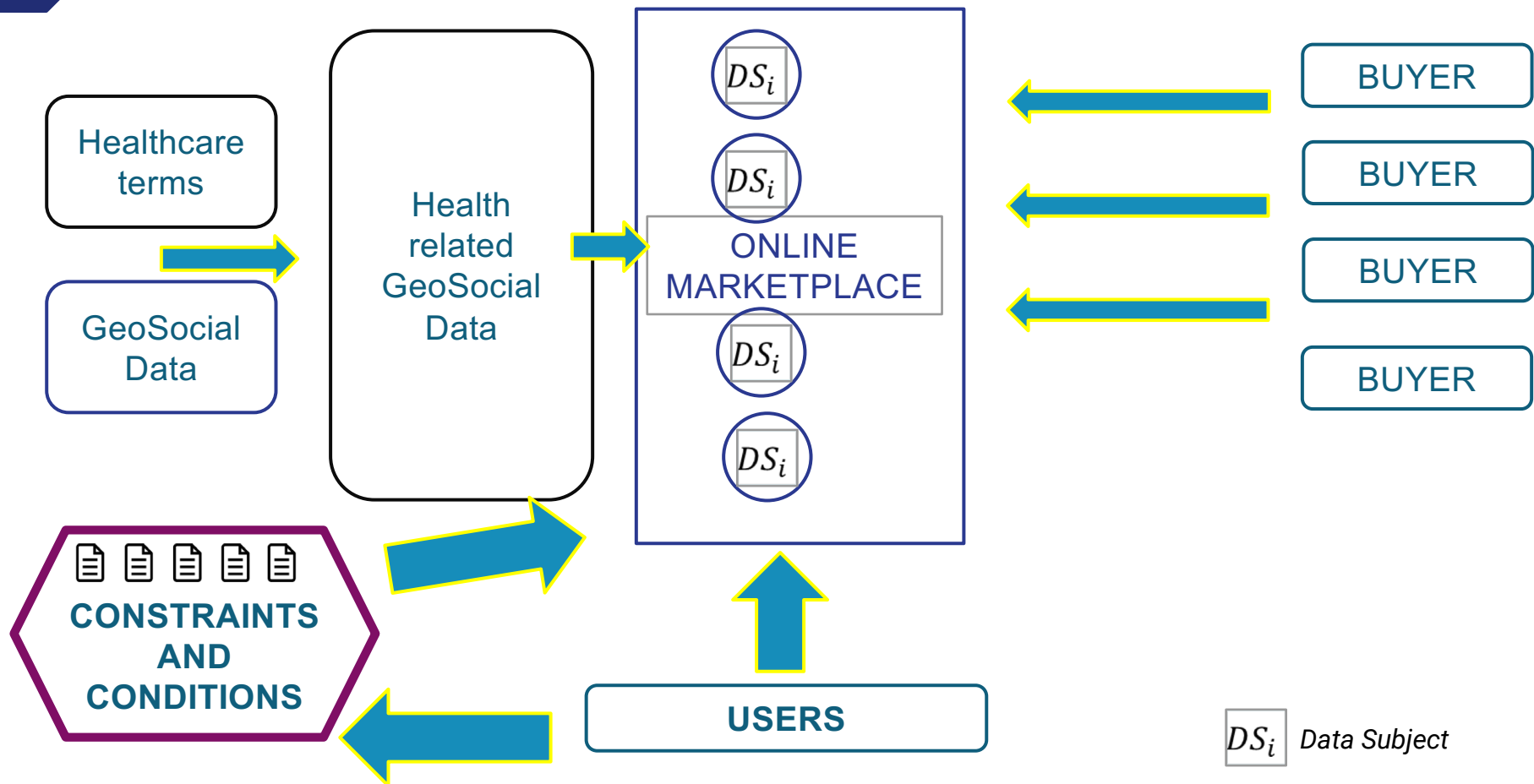
Infographic based on 2011 Path of the Blue Eye Project report. Learn how to access it here: www.pbeye.info/facebookhealth



Healthcare SocialMedia Privacy

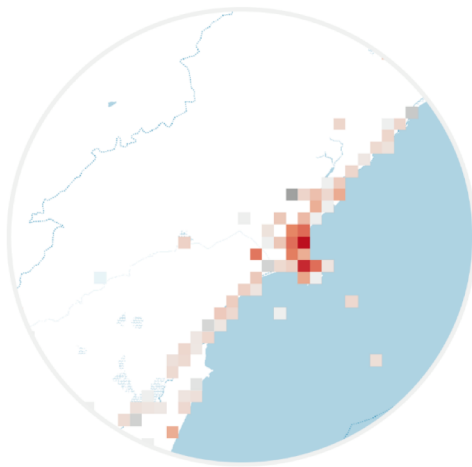
- **Personal Data in Healthcare related GeoSocial Media?**
- Facebook Like buttons reveal personal health information?
 - 40% of top US health and medical centers websites installed facebook like button
- Privacy concerns limit sharing
 - Afraid health information used by marketers and strangers

HEALTHCARE GEOSOCIAL DATA



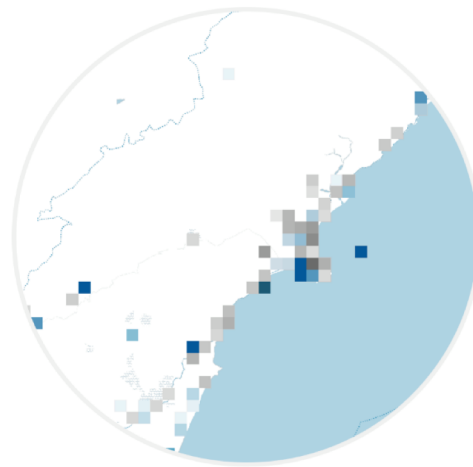
Online HEALTHCARE GEOSOCIAL MARKETPLACE

PERSONAL DATA IN NATURAL DISASTERS



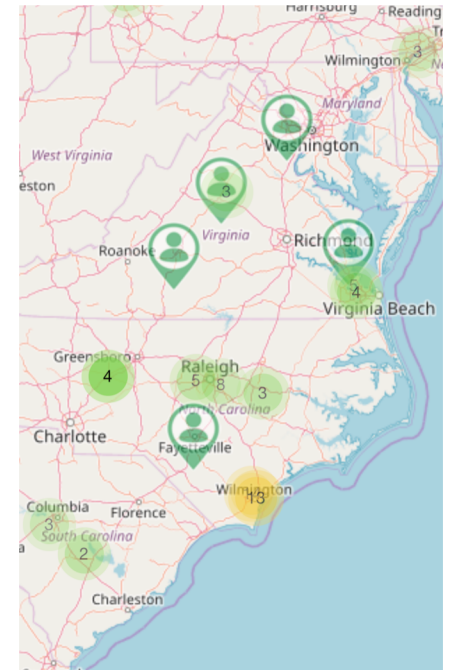
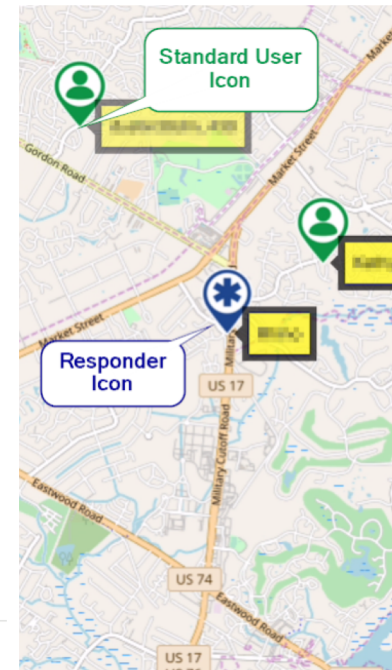
POPULATION DECREASES
(where people are leaving)

1 Day After



POPULATION INCREASE
(where people are returning to)

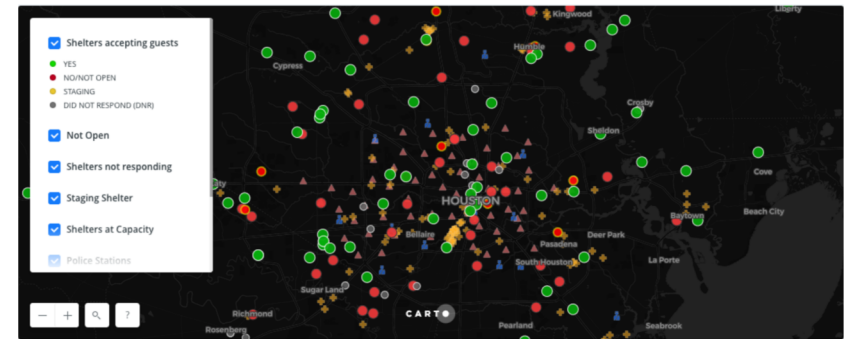
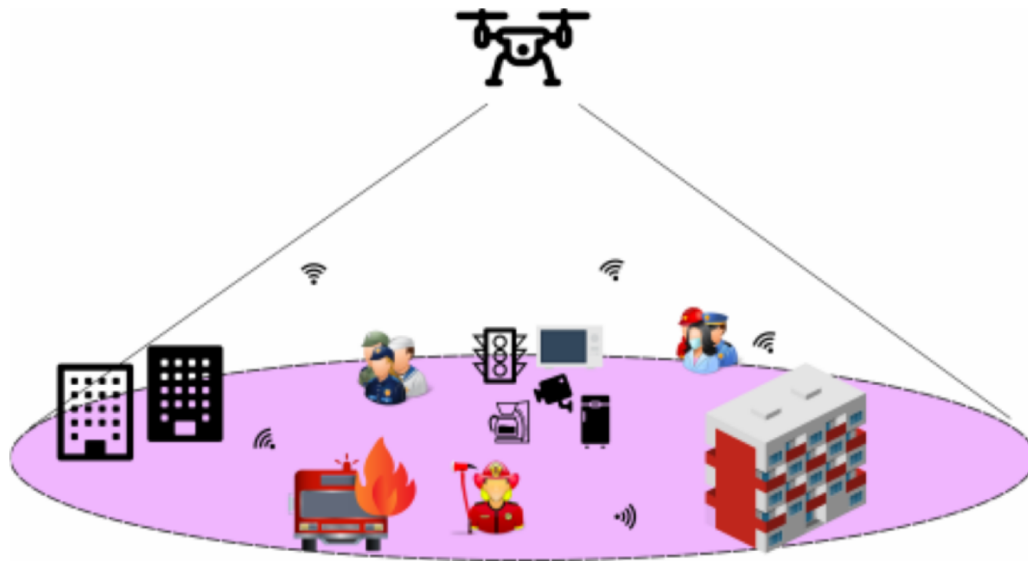
3 Days After



Left: <https://techcrunch.com/2017/06/07/facebook-will-share-anonymized-location-data-with-disaster-relief-organizations/>

Right: <https://aftermathdata.com/>

PERSONAL DATA IN NATURAL DISASTERS

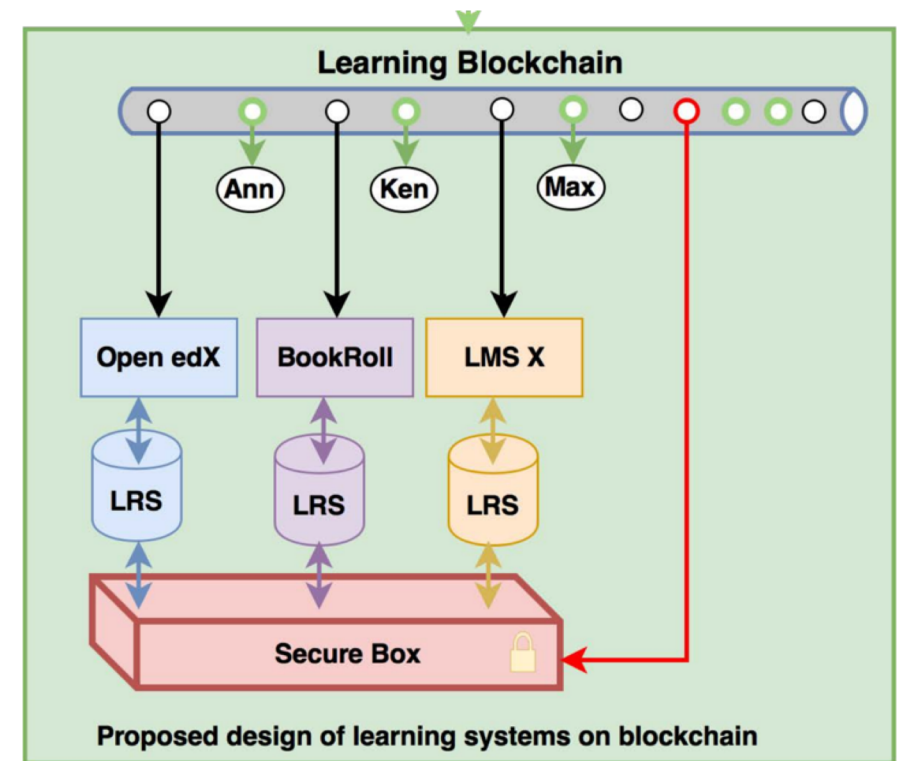
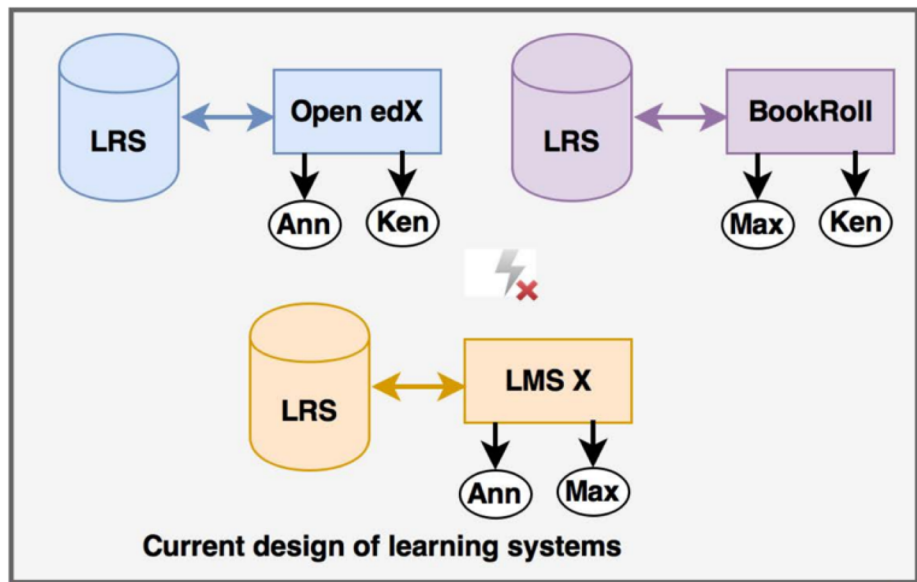


Left: Boukerche, Azzedine & Coutinho, Rodolfo. (2018). Smart Disaster Detection and Response System for Smart Cities. 10.1109/ISCC.2018.8538356.

Right-Top: <https://readwrite.com/2018/03/17/efforts-disaster-prediction-take-step-iot-sensors/>

Right-Bottom: <https://carto.com/blog/how-insurance-uses-location-data-prepare-natural-disasters/>

PERSONAL DATA IN EDUCATION



"Ocheja, P., Flanagan, B., & Ogata, H. (2018, March). Connecting decentralized learning records: a blockchain based learning analytics platform. In Proceedings of the 8th International Conference on Learning Analytics and Knowledge (pp. 265-269). ACM."

International problem

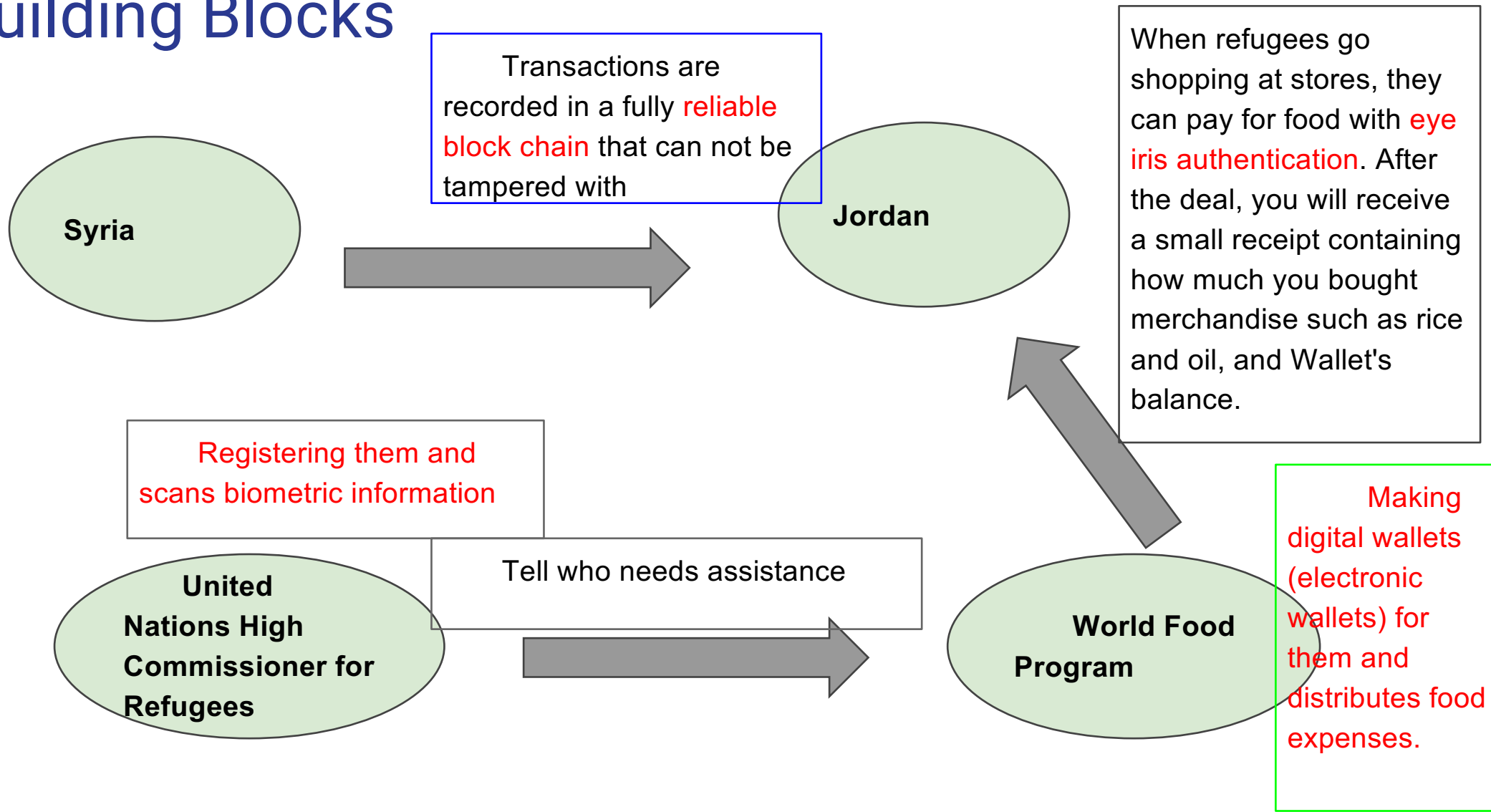
"Refugee Crisis"

•Why data protection is needed?

- Refugees are often in a **vulnerable position** and their information is **highly sensitive**; they face far greater risks than most of us if their data is misused or shared with those who have no right to it.

- Growing number of actors in the humanitarian field
 - Increased processing of personal data, including data sharing among such actors.
 - UNHCR, as a main collector of personal data, therefore needs to lead efforts to ensure respect for data protection.
 - **Using the cloud** for personal data may save costs and increase flexibility, but could compromise confidentiality.

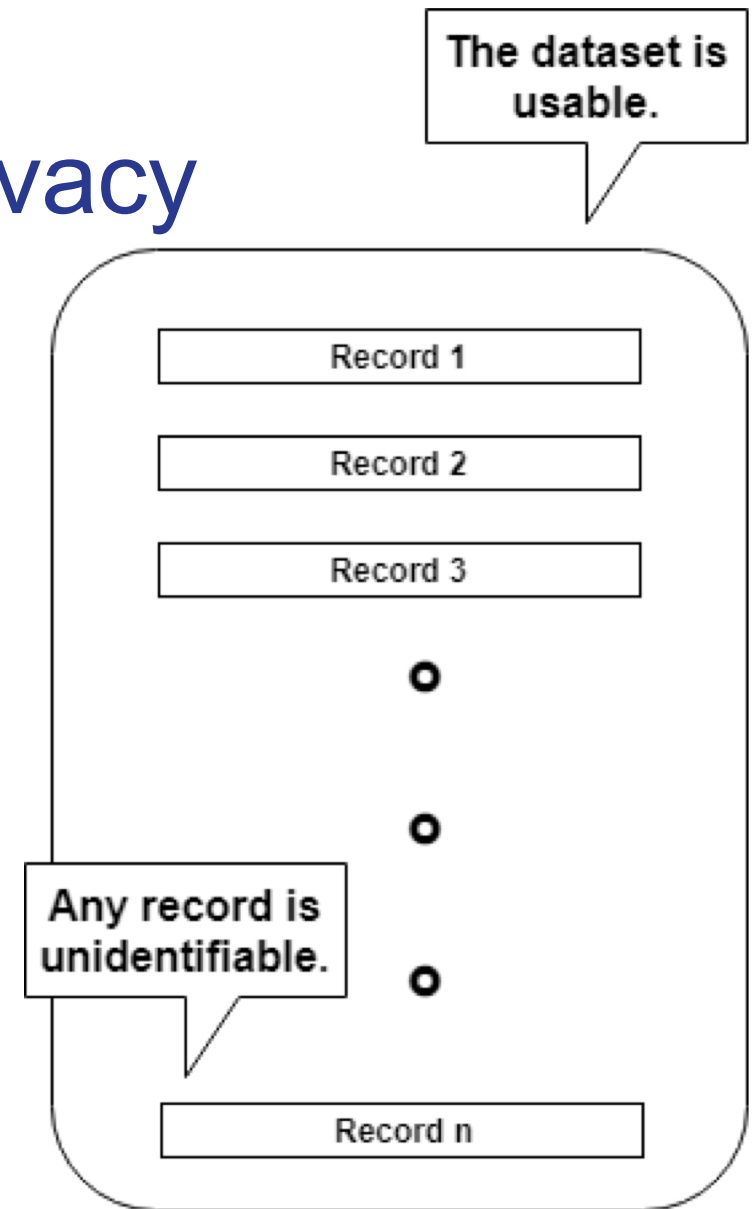
Building Blocks





Personal Data & Privacy

- Personally **identifiable** information
- Publishing personal data causes **privacy loss**.
- Our goal: to **utilize personal data** but still protect any record from **being identified**.

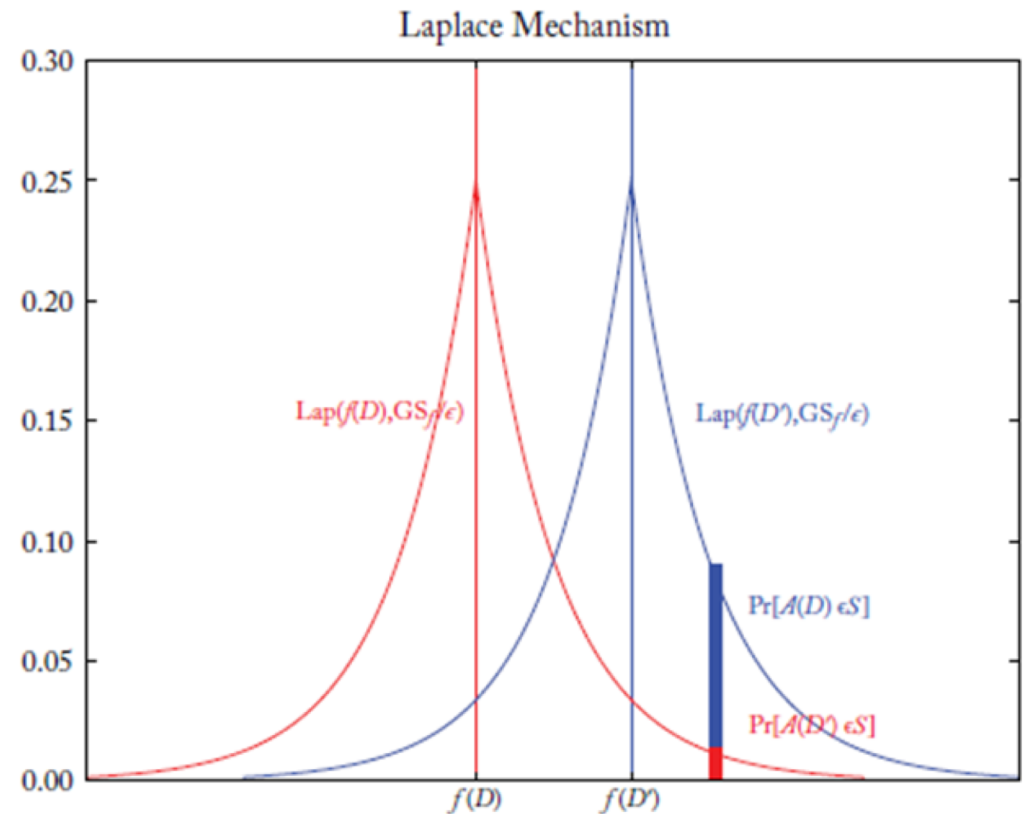


Privacy Protection

- Differential Privacy: any single record in a dataset has only a **limited impact** on the output.

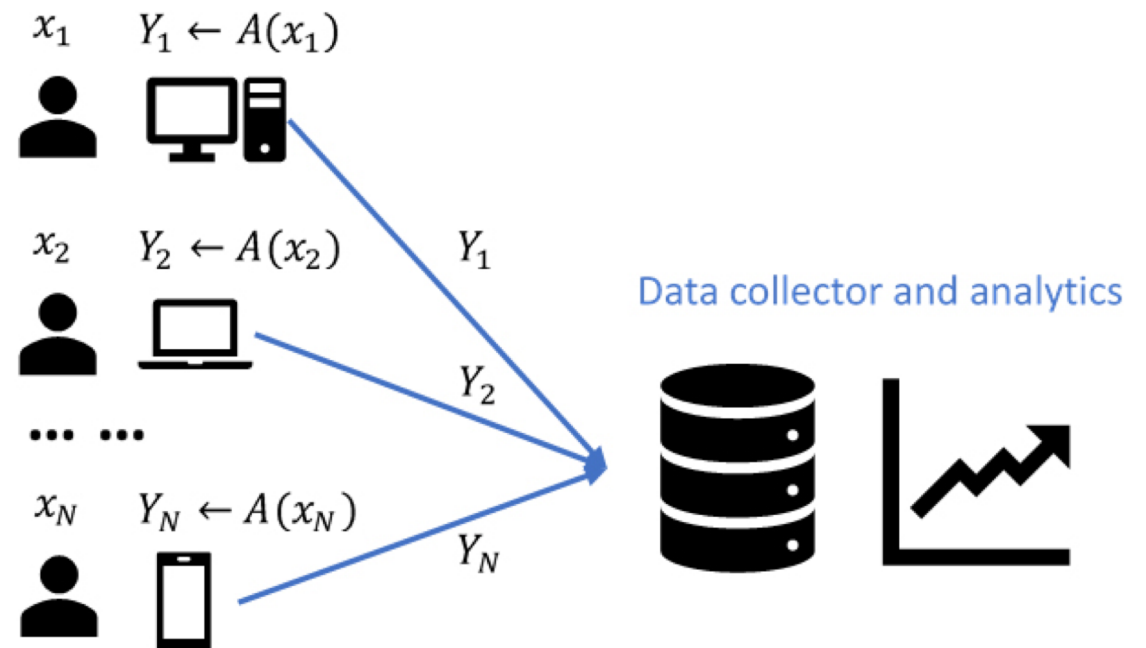
- We need to add random noise into the data. The higher the noise, the lower the privacy loss.

$$\frac{\Pr[A(D) \in S]}{\Pr[A(D') \in S]} \leq e^\epsilon$$



Local DP

- DP: the data bank is still able to access **raw data**.
- Local Differential Privacy: let users add noise **by themselves** before uploading their data to the bank.





Conclusion

PDBANK: Personal Data Bank

- Having potential **social benefits** for improving human life
 - Healthcare
 - Geo social
 - Natural disaster
 - Education
 - Refugees
- Balancing **economic value** and **privacy protection** of personal data
 - Mining potential value of personal data and boosting economic growth
 - Protecting personal data with bounded privacy loss and local DP