# 位置パーソナル情報の保護と利活用に関する研究

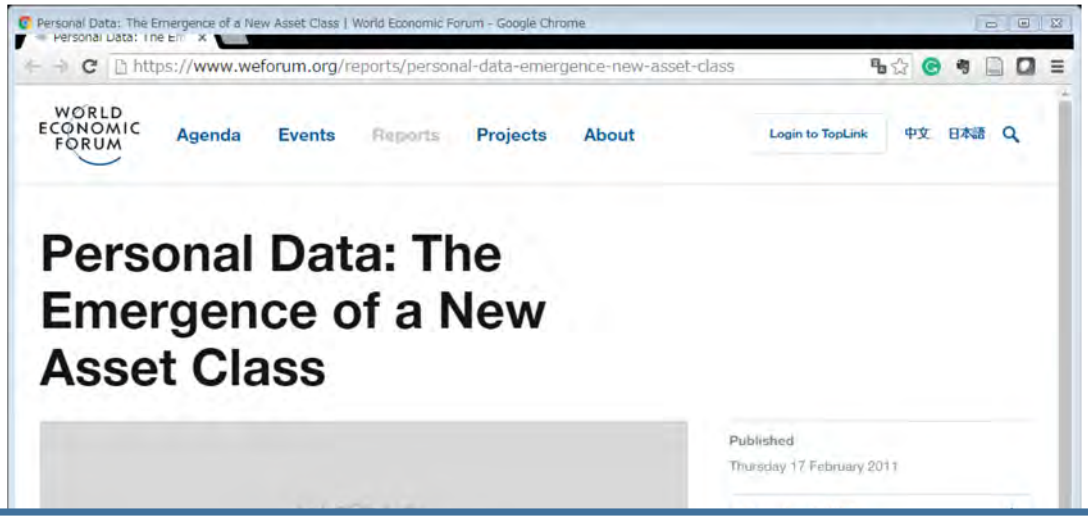
京都大学大学院情報学研究科
吉川 正俊

# “data as the new oil”

データメジャー(GAFA)4社の時価総額は
2010年代前半に石油メジャー（エクソンモービル，
ロイヤル・ダッチ・シェル，BP，シェブロン）を抜いた

- **Clive Humby**, UK Mathemetician and architect of Tesco’s Clubcard, 2006 (widely credited as the first to coin the phrase): *“Data is the new oil. It’s valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value.”*
- **Meglena Kuneva,** European Consumer Commissioner, 2009: *“Personal data is the new oil of the internet and the new currency of the digital world.”*

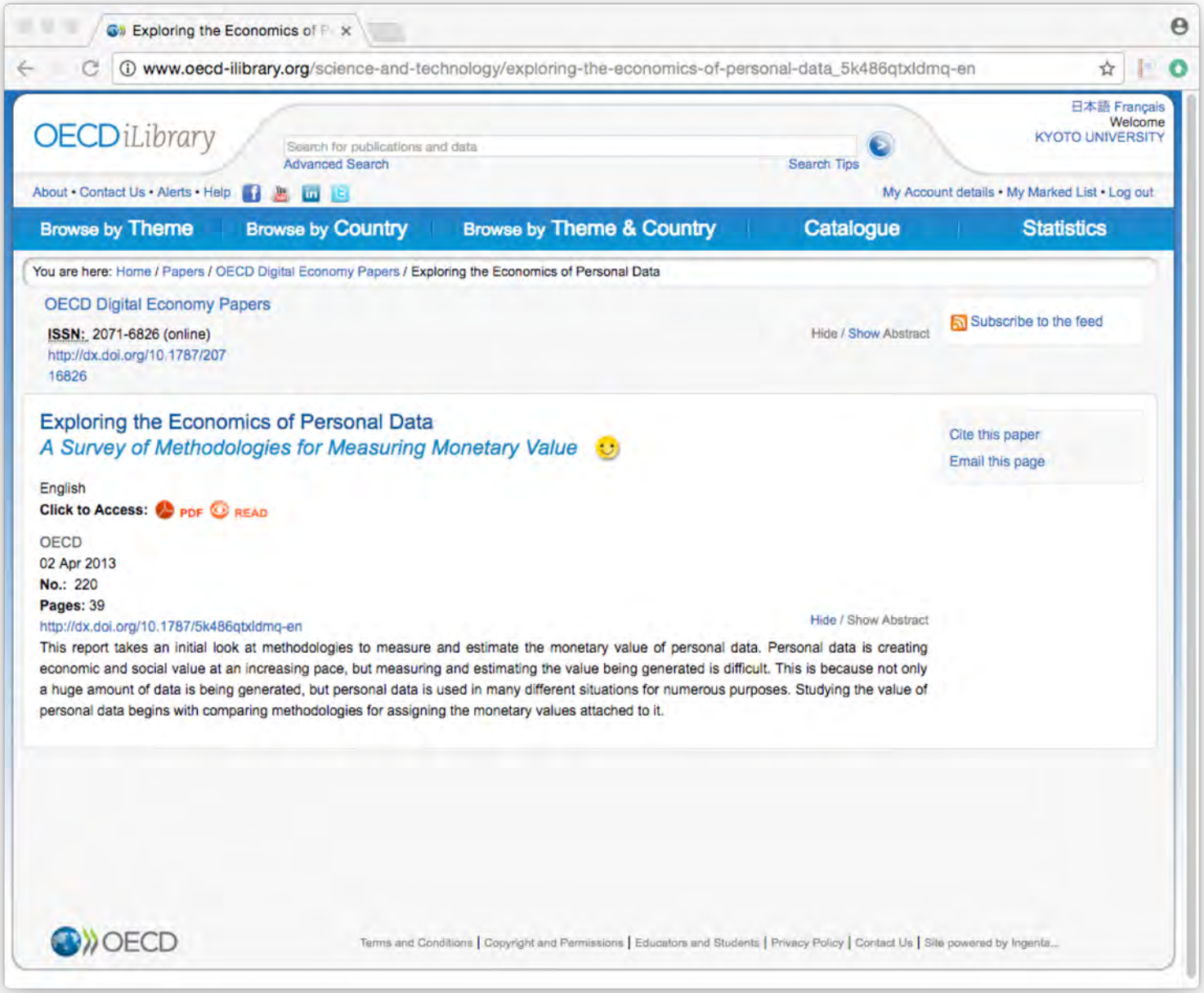
# World Economic Forum (2011)



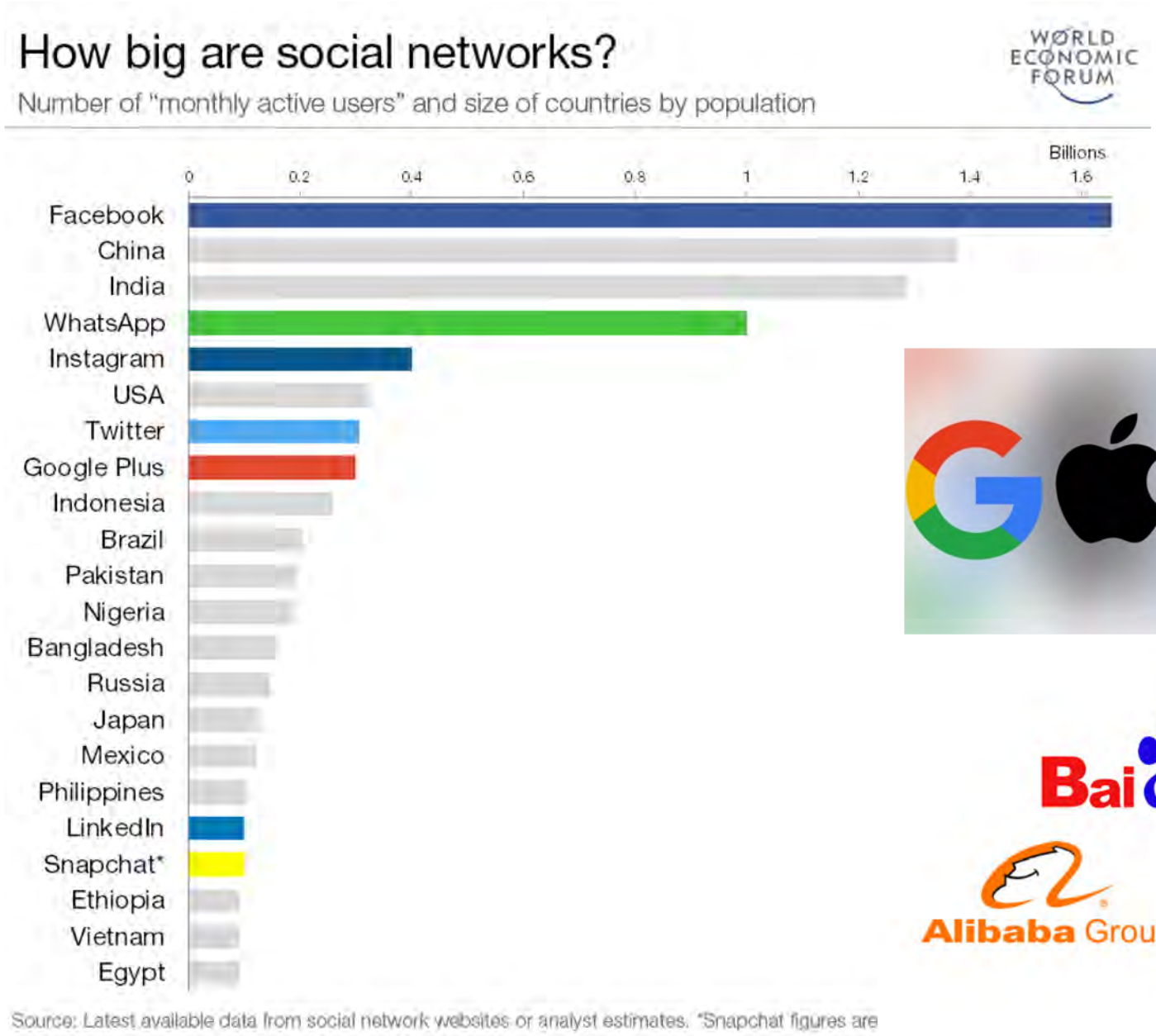

Building the legal, cultural, technological and economic infrastructure to enable the development of a balanced personal data ecosystem is vitally important to improving the state of the world.

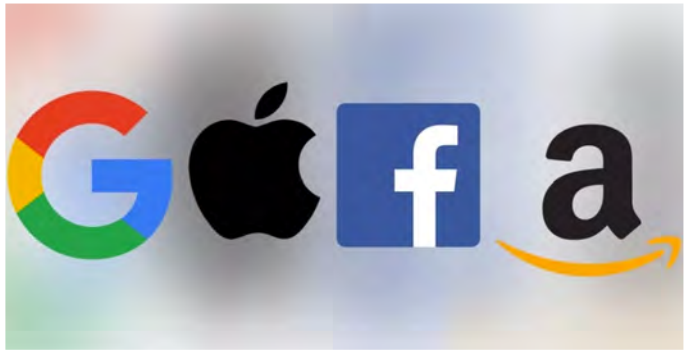
http://www.weforum.org/reports/personal-data-emergence-new-asset-class

# “Exploring the Economics of Personal Data,” OECD Digital Economy Papers (2013)



# GAFA and BAT



Baidu 百度

Alibaba Group

Tencent 腾讯

# A datacenter of facebook

VLDB 2013 Keynotes
“Data Infrastructure at Web Scale”
Jay Parikh, VP of Infrastructure Engineering, Facebook

# 保護と利活用のための法および政策

# EU一般データ保護規則

- 2016年4月14日、欧州議会本会議においてEU一般データ保護規則（General Data Protection Regulation、以下GDPR）が正式に可決
- データ主体に与えられる新たな権利としては、ウェブの検索結果に示される不適切な個人データに対する本人の異議申立を円滑化する忘れられる権利（消去する権利）
- 「データポータビリティ」の権利や「プロファイリング」に関わる権利は、グローバルなインターネットのエコノミクスに根本的な変革をもたらしうる概念である。
- 適用範囲はEU市民をターゲットとしたサービスを提供する全世界の企業にまで及ぶものであり、EUで事業を展開する日本企業はもちろん、インターネット上でサービスを提供するすべての主体に直接的な影響を与えうる。
- 違反者に課せられる全世界連結売上高の4%あるいは2,000万ユーロのいずれか高い方を上限とした莫大な制裁金は、グローバル企業の経営にすら大打撃をもたらす強大なエンフォースメント装置

EU一般データ保護規則可決、そして情報社会の民主主義について
生貝直人 | 東京大学大学院情報学環客員准教授、情報通信総合研究所研究員
2016年4月15日 20時55分配信
http://bylines.news.yahoo.co.jp/ikegai/20160415-00056652/

未来投資戦略2017－Society 5.0
実現に向けた改革－
内閣府 日本経済再生本部

## 産業界・個人におけるデータ流通・利活用の加速化

（残された課題）

・プライバシー保護に関して国民が抱く漠然とした不安、データ連携や活用によるメリットが分かりにくい、データの利用権限が明確でない等により、企業や業種の枠を超えたデータ連携・活用が十分進んでいない。

（主な取組）

・企業間での適切な契約締結を通じたデータ利用権限の明確化と共有を促すべく、本年５月に策定したデータ利用権限に関する契約ガイドライン等の活用を進める。同時に、本年度中を目途に産業界等との対話を通じて分野ごとに留意すべき点の整理を行い、個別分野への展開を進める。

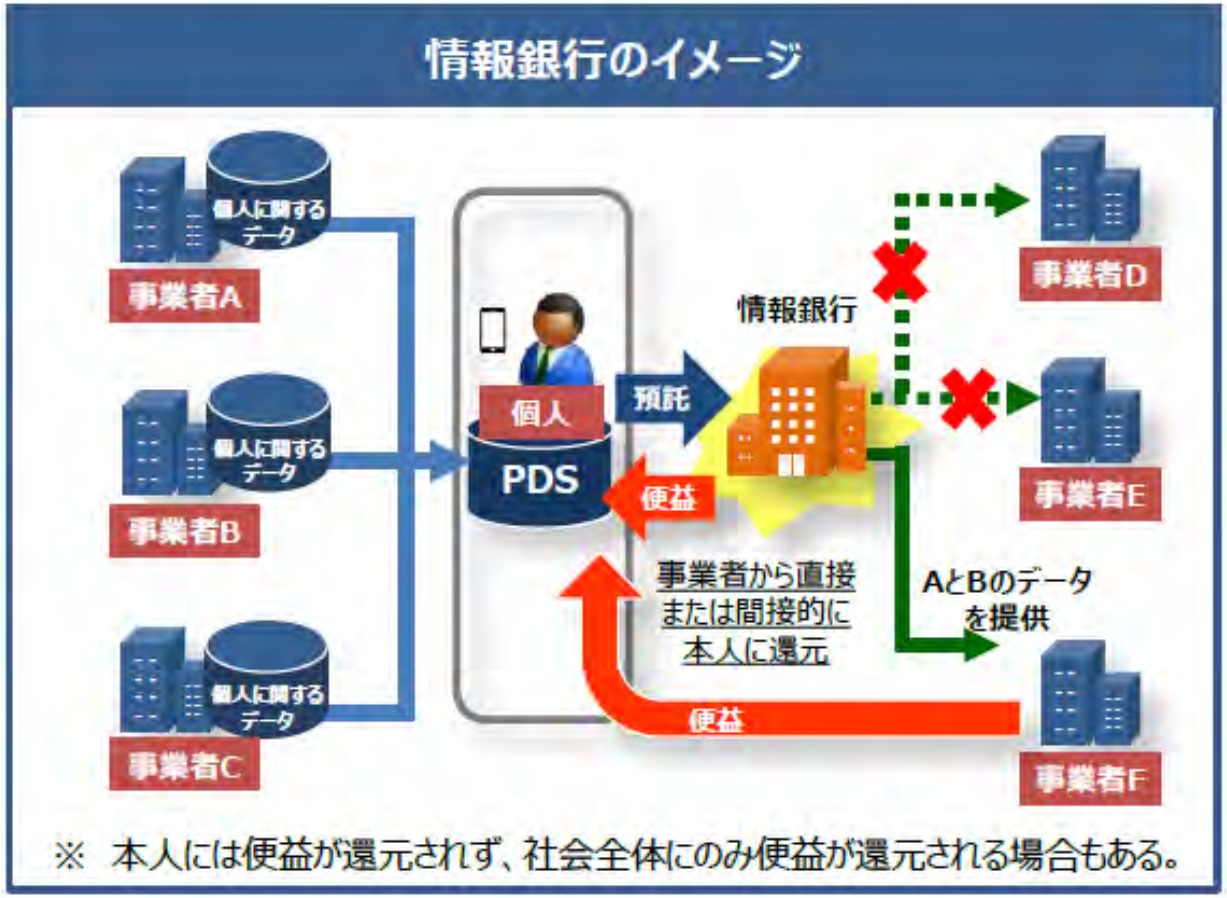
・データ利用者の利便性を高め、データ流通市場の拡大・活性化を促進するため、民間事業者間の自主ルールの策定及びその普及促進を図るための民主導の枠組みが本年度中に構築されるよう支援する。

・個人の関与の下でパーソナルデータの流通・活用を進める仕組みである PDS（Personal Data Store）や情報銀行、データ取引市場等について、官民連携実証事業を行う。あわせて、個人の関与の下で信頼性、公正性、透明性を確保するための制度の在り方等について検討し、本年中に結論を得る。

・個人情報及び匿名加工情報の取扱いに関する民間企業等からの相談対応や、これらを踏まえた事例集の公表等のデータ利活用促進に向けた情報発信等を本年度中に開始する。

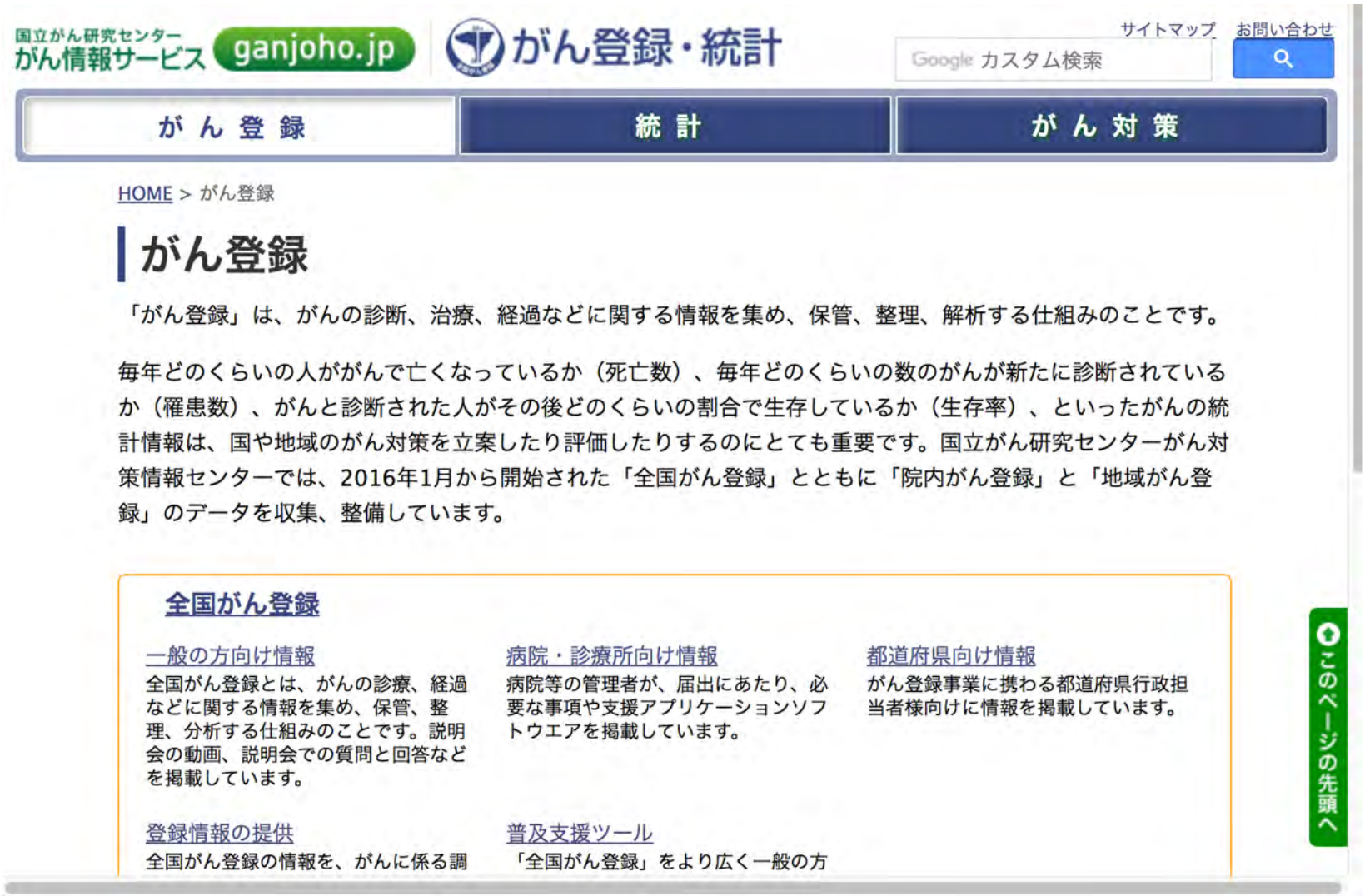
## 2．定義　（2）情報銀行

情報銀行（情報利用信用銀行）とは、個人とのデータ活用に関する契約等に基づき、PDS等のシステムを活用して個人のデータを管理するとともに、個人の指示又は予め指定した条件に基づき個人に代わり妥当性を判断の上、データを第三者（他の事業者）に提供する事業。




出典：AI、IoT時代におけるデータ活用ワーキンググループ. 中間とりまとめの概要. 内閣官房IT総合戦略室, 平成29年3月.

# 公共の利益のために強制的に収集されるパーソナルデータもある



# データ資本管理学

# データ資本管理学

- 国家，企業，個人がデータを資本として適切に管理し，その健全な流通を実現する.

データ流通の形態：
- 税として徴収
- 保険料として収集
- 市場における自由取引
- オープンデータ化

×

重要データ：
- パーソナルデータ
- 産業データ
- 科学データ

# データ資本管理学（パーソナルデータの例）

- 情報の真正性の証明
  例）雑音により摂動化したデータ
  「四捨五入した年収額800万円」の真正性証明
- 価値付け
  - 相互に関連する多様なデータの譲渡によるプライバシ喪失の定量化
  - IPA（暗号）のようなプライバシ保護機構の公的推奨
- 価格付け
  - 無裁定価格 *Definition* 3.7 (*Arbitrage Free*). A pricing function $\pi(\mathbf{Q})$ is arbitrage free if, for every multiset $\mathbf{S} = \{\mathbf{Q}_1, \ldots, \mathbf{Q}_m\}$, if $\mathbf{S} \to \mathbf{Q}$ then

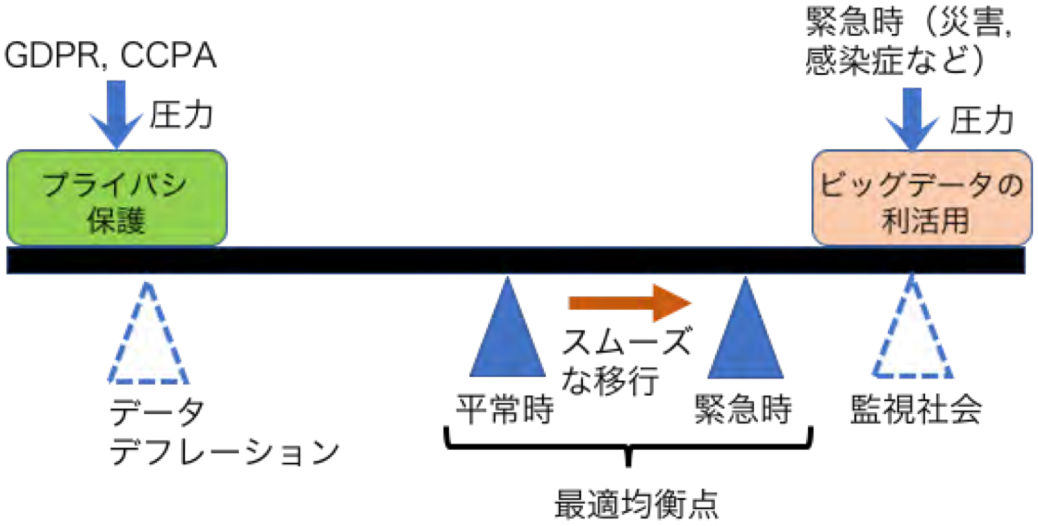
$$\pi(\mathbf{Q}) \leq \sum_{i=1}^{m} \pi(\mathbf{Q}_i).$$

- 貢献度評価
  例）多数の個人から収集したデータからの学習による生成モデル
- データ財布，データ家計簿
  - 過去のデータ取引履歴をすべて管理し，最適な家計戦略を提案
  - ブラウザ，携帯端末などに組み込む．

# プライバシ保護と利活用のバランスを取る最適均衡点を技術，社会の両面を考慮して見出すことは容易ではない

## ニーズ

- データの**真正性**確保
  ← EUのeIDAS規制
- プライバシ保護とビッグデータの利活用の最適**均衡点**探索（技術，社会の両面から）

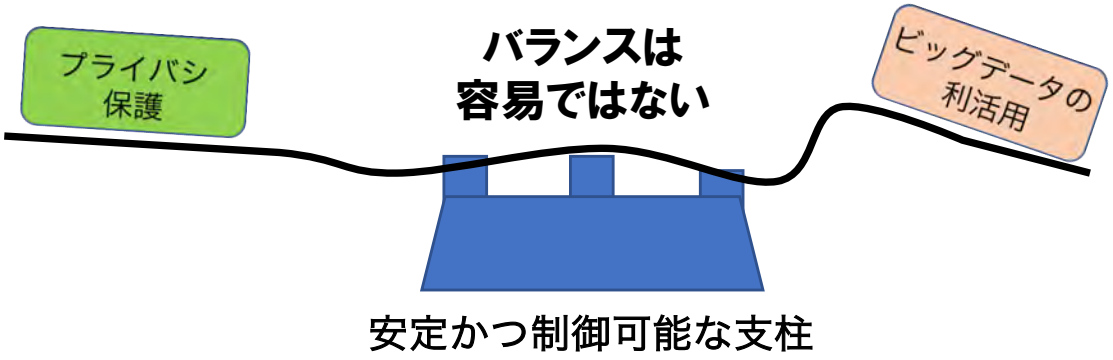



## 研究課題

- データの真正性 と セキュリティ，プライバシ，価値査定の**両立**
- 細粒度データの**価値査定**と貢献に応じた**報酬分配**（多数のセンサデータによる機械学習モデル構築への貢献など）
- **データの権利**（作成者の権利，パーソナルデータの自己コントロール権）保護

## 基盤技術

Secure processor, 暗号，認証，差分プライバシ，ブロックチェーン，provenance, Shapley値，形式的検証，メタデータ，…



# パーソナルデータとは？

**パーソナルデータとは何か？**（日本の個人情報保護法で規定される「個人情報」よりも広範囲）

any information relating to an identified or identifiable individual (data subject)

*2013 OECD Privacy Guidelines*

any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

*EU GDPR*

# Difficulties in Conceptualization of Personal Data

Distinguishing between personal and non-personal data is becoming increasingly difficult.

*2013 OECD Privacy Guidelines*

personal information rarely belongs to just one individual; it is often formed in relationships with others.

…

the problem with the current theories of privacy is the method of conceptualizing. The theories fail on their own terms --- they never achieves the goal of finding the common denominator, …

*Daniel J. Solove "Understanding Privacy", Harvard University Press, 2010.*

John Locke's conception of property as the fruit of labor and as an extension of the self has formed the backbone of intellectual-property law, …

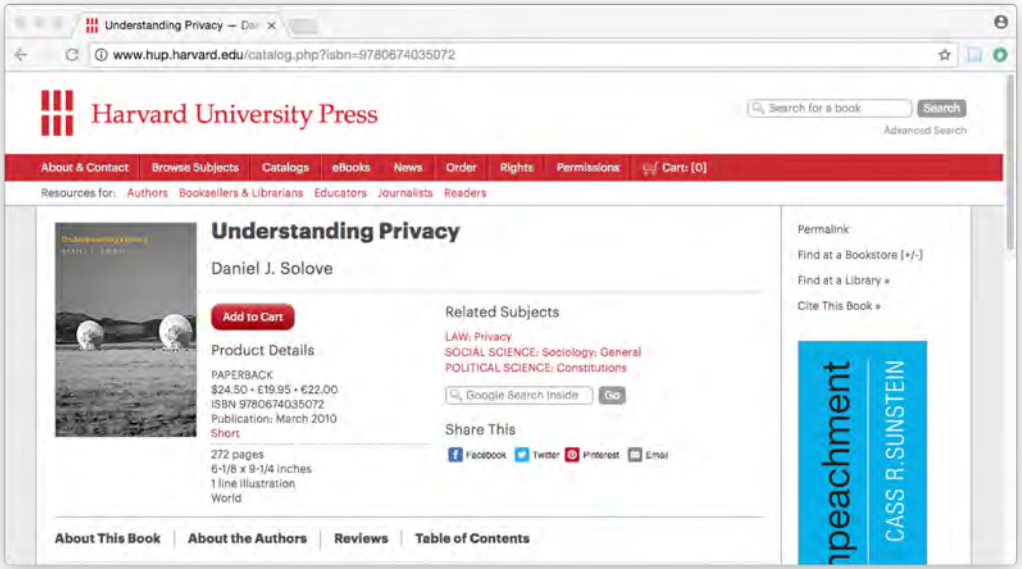
…

Extending property concepts to personal information, however, has difficulties.

…

A claim of privacy is not the same as a claim of ownership.

Daniel J. Solove
"Understanding Privacy",
Harvard University Press, 2010.



プライバシー・個人情報の財産権論は、通説的見解の問題をあぶり出すという役割は果たすものの、固有の意義を見いだすことは困難であるといわざるを得ない。

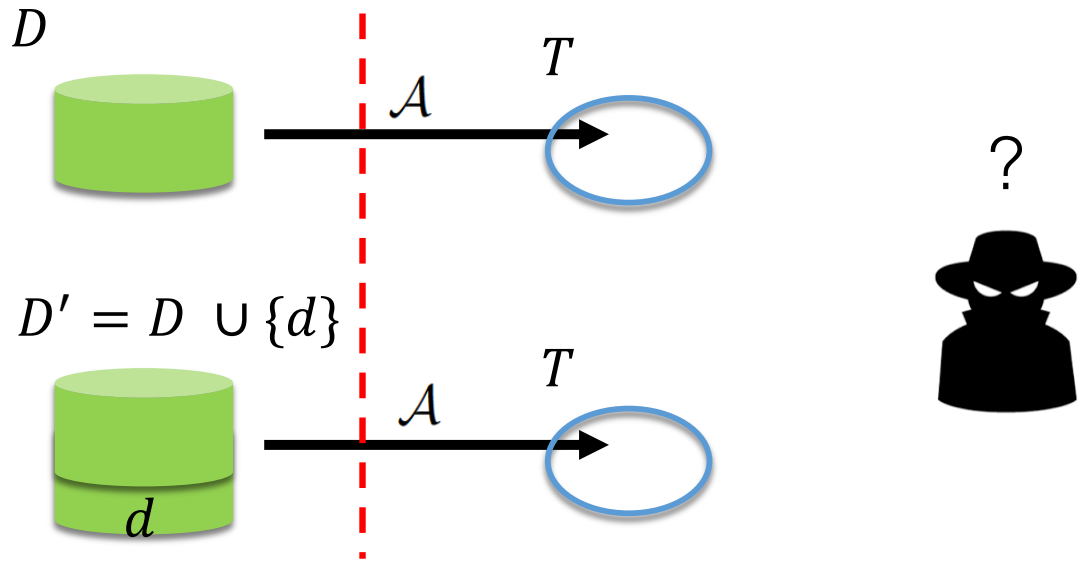
石井 夏生利. プライバシー・個人情報の「財産権論」-ライフログをめぐる問題状況を踏まえて. 情報通信政策レビュー, No. 4, 2012.

# 差分プライバシー

# Differential Privacy（差分プライバシー）


# Differential Privacy (DP)

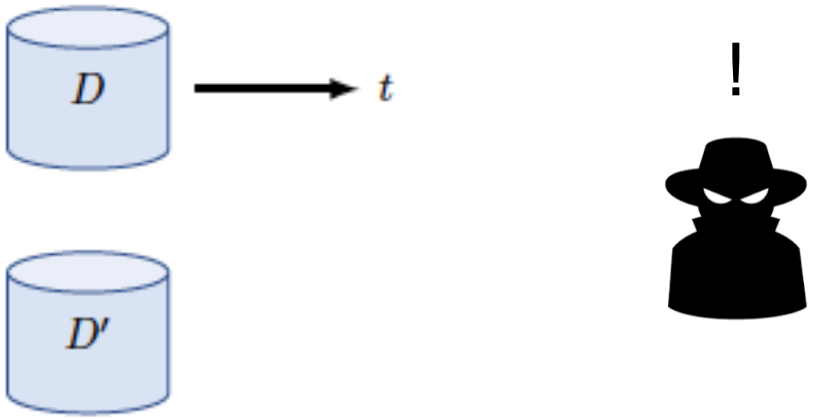
- Output of an algorithm should **not** be significantly affected by individual's data



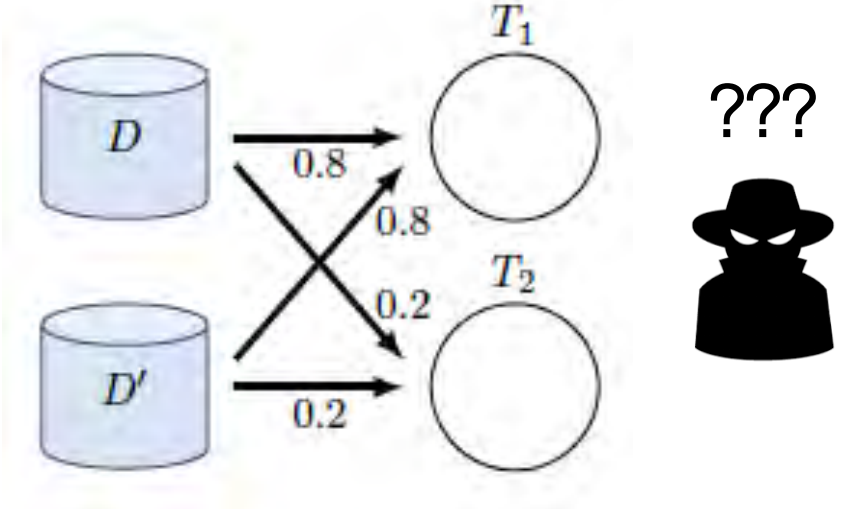

$D$ and $D'$ : neighboring databases
$\mathcal{A}$ : an algorithm
$T$ : the range of $\mathcal{A}$

# DP: Two extreme cases

- No Protection of Privacy: Full Utility

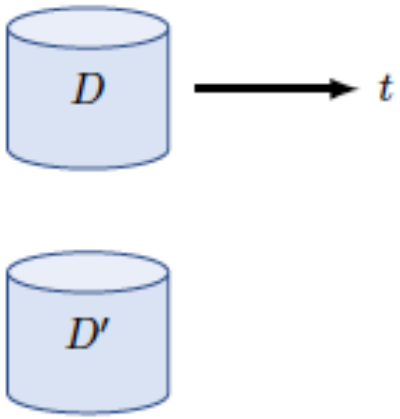



- Full Protection of Privacy: No Utility



# DP: Two extreme cases

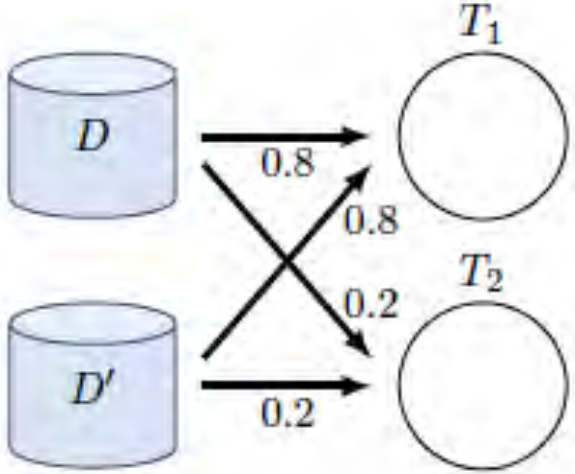
- No Protection of Privacy: Full Utility





$$\frac{\Pr[\mathcal{A}(D) = t]}{\Pr[\mathcal{A}(D') = t]} = \infty.$$

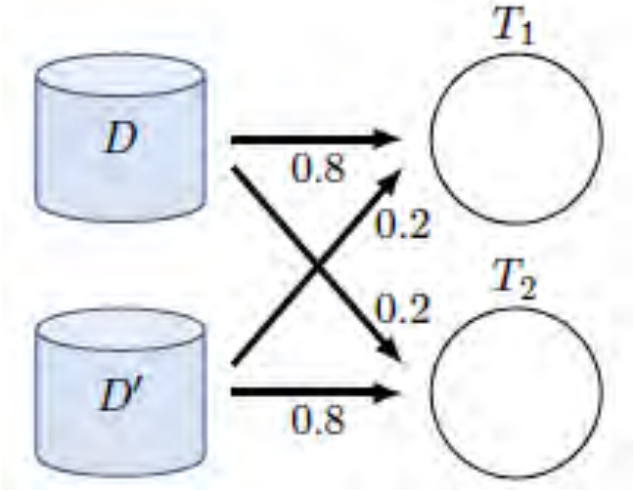
- Full Protection of Privacy: No Utility





$$\frac{\Pr[\mathcal{A}(D) \in T]}{\Pr[\mathcal{A}(D') \in T]} = 1$$

- **Modest** Protection of Privacy: **Modest** Utility

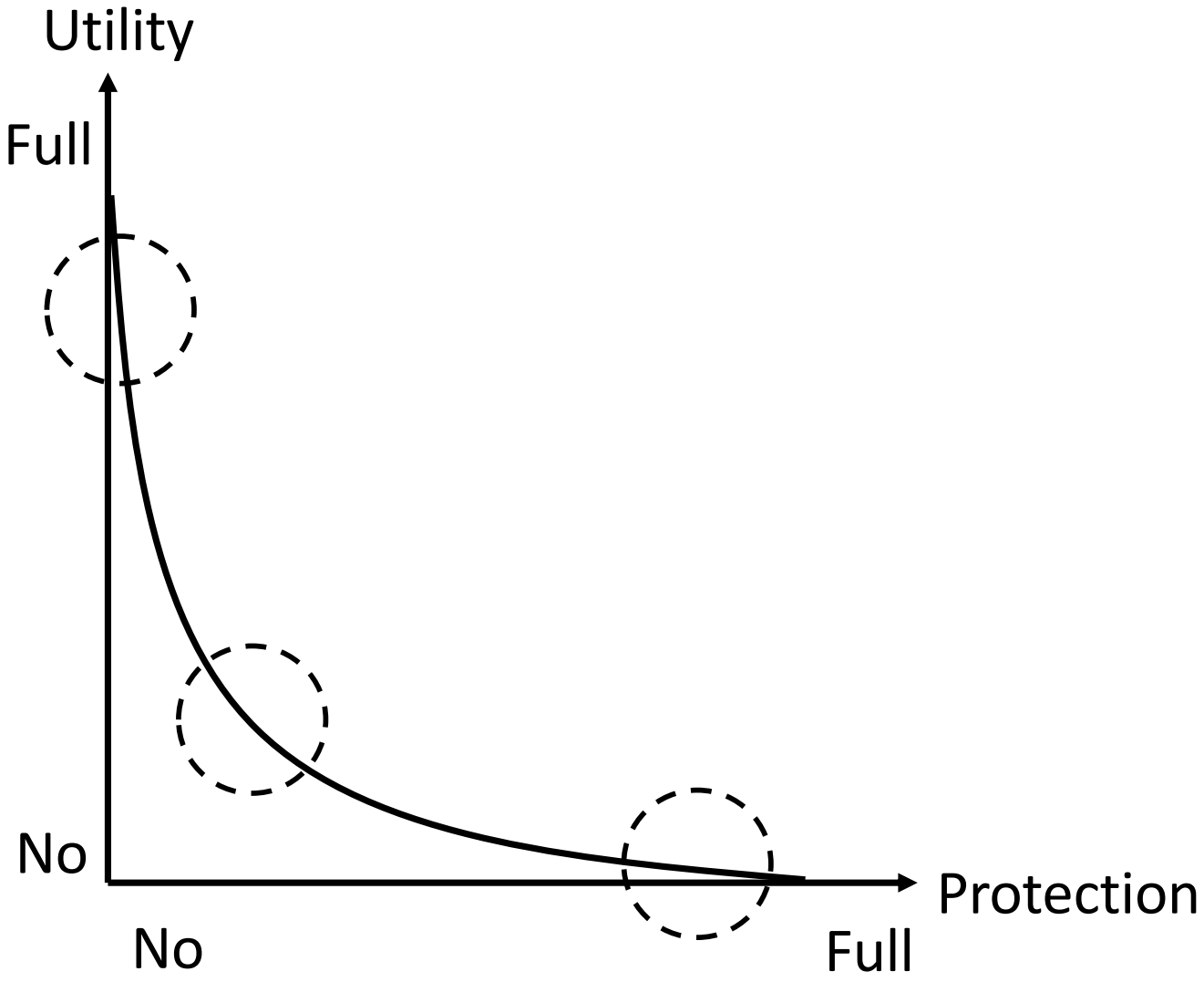



?!

$$\forall T \subseteq Range(\mathcal{A}):$$

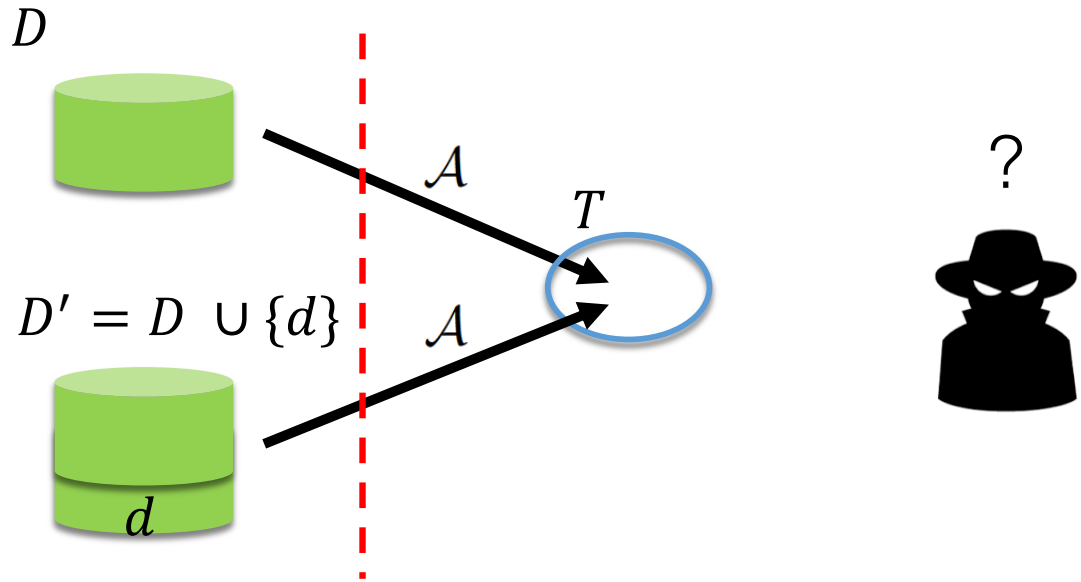
$$\frac{\Pr[\mathcal{A}(D) \in T]}{\Pr[\mathcal{A}(D') \in T]} \leq c$$

# Balance of Protection and Utility



# ε-Differential Privacy

- Output of an algorithm should **not** be significantly affected by individual's data

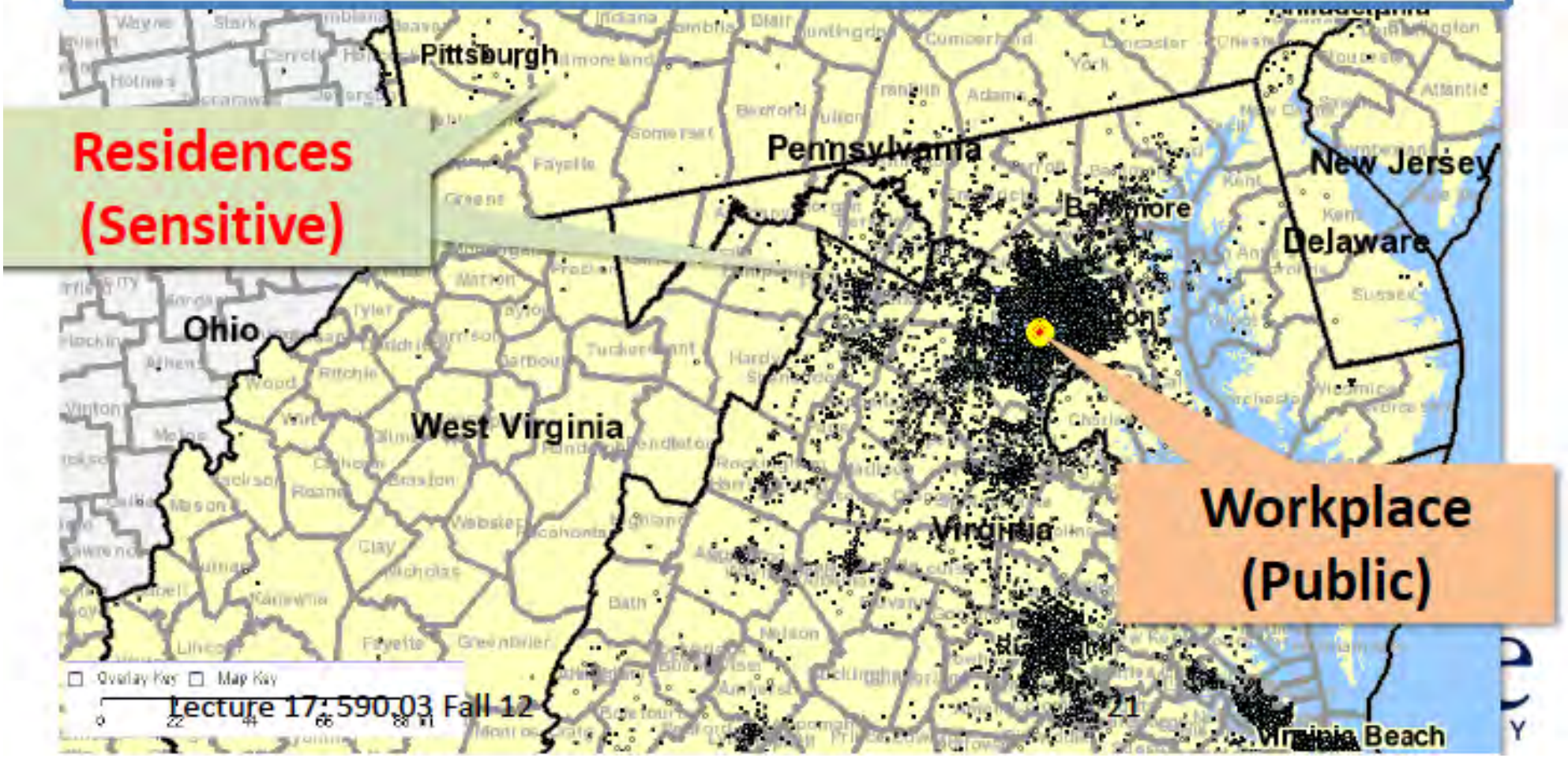



for any datasets $D$ and $D'$ that differ on one element, we have

$$\forall T \subseteq Range(\mathcal{A}): \frac{\Pr[\mathcal{A}(D) \in T]}{\Pr[\mathcal{A}(D') \in T]} \leq e^{\epsilon}$$

# OnTheMap: A Census application that plots commuting patterns of workers

http://onthemap.ces.census.gov/



# Bridging the Gap between Computer Science and Legal Approaches to Privacy


Alexandra Wood
Berkman Klein Center for Internet & Society at Harvard University

Privacy Semester Planning Workshop
May 24, 2017

# 位置パーソナル情報

# 位置情報

- (時刻, 場所)という対の集まり

- プライバシに関わる多様なデータを含む
  - 良く立ち寄る店の種類 → 年齢，性別
  - スポーツジム → 生活習慣
  - 教会 → 宗教
  - 政治集会開催会場 → 支持政党
  - 専門病院 → 疾患
  - 複数の個人が同時刻に同じ場所にいる
    → 会合，集会への同時参加
  - …

# 匿名化された位置情報から容易に個人を特定可能

- 150万人の15ヶ月に渡る匿名化された移動データ
- 携帯電話に最も近い基地局を記録
- 4個の(時刻, 場所)対データから95%の個人を特定できた．“If individual’s patterns are unique enough, outside information can be used to link the data back to an individual.”

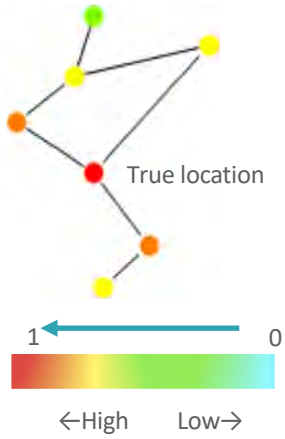

Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, “Unique in the Crowd: The privacy bounds of human mobility,” *Scientific Reports*, vol. 3, no. 1, p. 1376, Mar. 2013, doi: 10.1038/srep01376.

# The mechanism to satisfy geo-graph-indistinguishability : GEM

graph-exponential mechanism (GEM)

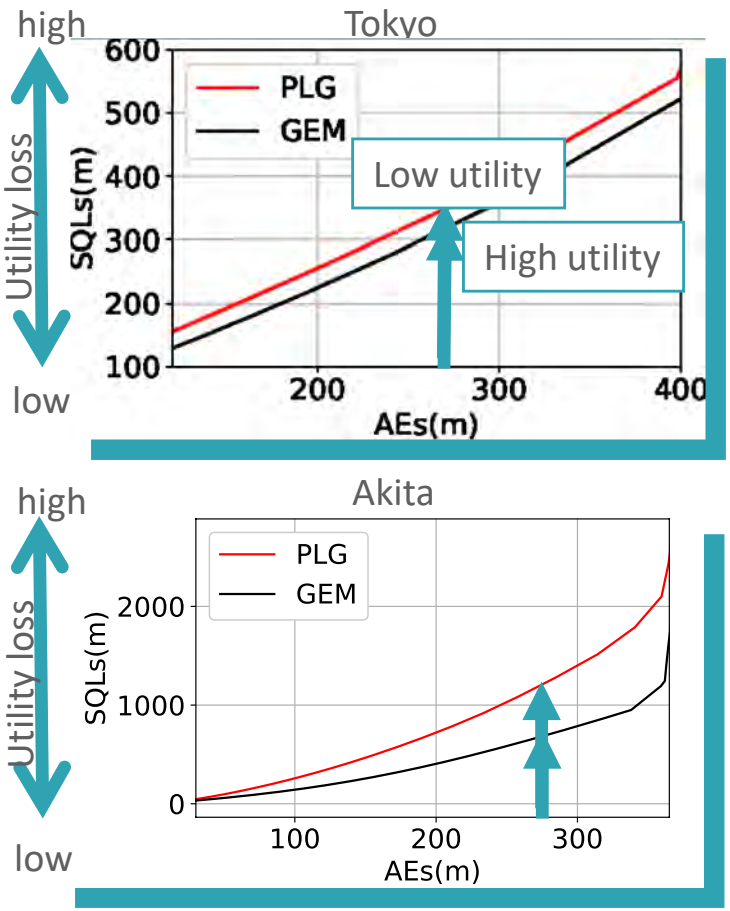
$$K(v)(w) = \alpha(v)\mathrm{e}^{-\frac{\epsilon}{2}d_s(v,w)}$$

- α is normalization factor

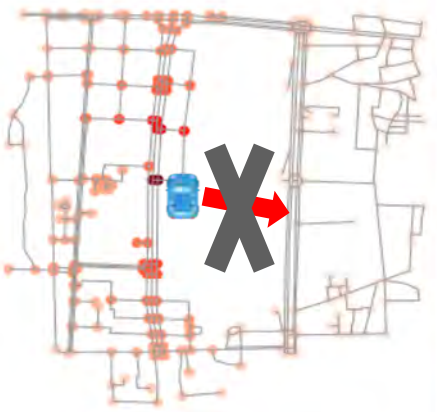



- Graph-exponential mechanism (GEM) satisfies $\varepsilon$-geo-graph-indistinguishability

# Experiments on Real Road Networks




- In the same privacy level,
  GE achieves lower SQL (high utility)
  because GE considers the graph shape and does not cause insufficiency 2
- The difference is larger in Akita graph
  because the difference of the shortest distance and Euclid distance is larger

# COVID-19感染拡大防止のための接触追跡アプリ

- 多くの国で開発されている

- プライバシ保護と公共的利益のバランスについても議論がある

[1] Q. Tang, “Privacy-Preserving Contact Tracing: current solutions and open questions,” *arXiv:2004.06818 [cs]*, Apr. 2020, Accessed: Apr. 30, 2020. [Online]. Available: http://arxiv.org/abs/2004.06818.

[2] R. Raskar *et al.*, “Apps Gone Rogue: Maintaining Personal Privacy in an Epidemic,” *arXiv:2003.08567 [cs]*, Mar. 2020, Accessed: Apr. 01, 2020. [Online]. Available: http://arxiv.org/abs/2003.08567.

[3] B. Mohanty, A. Chughtai, and F. Rabhi, “Use of Mobile Apps for epidemic surveillance and response – availability and gaps.,” *Global Biosecurity*, vol. 1, no. 2, pp. 37–49, Sep. 2019, doi: 10.31646/gbio.39.

[4] Y.-A. de Montjoye, F. Houssiau, and A. Gadotti, “Blogpost: Evaluating COVID-19 contact tracing apps? Here are 8 privacy questions we think you should ask.,” p. 7.

[5] H. Cho, D. Ippolito, and Y. W. Yu, “Contact Tracing Mobile Apps for COVID-19: Privacy Considerations and Related Trade-offs,” *arXiv:2003.11511 [cs]*, Mar. 2020, Accessed: Apr. 15, 2020. [Online]. Available: http://arxiv.org/abs/2003.11511.

[6] G. Arbia, “A Note on Early Epidemiological Analysis of Coronavirus Disease 2019 Outbreak using Crowdsourced Data,” *arXiv:2003.06207 [stat]*, Mar. 2020, Accessed: Mar. 24, 2020. [Online]. Available: http://arxiv.org/abs/2003.06207.

[7] M. Nanni *et al.*, “Give more data, awareness and control to individual citizens, and they will help COVID-19 containment,” p. 6, 2020.

# パーソナルデータ

- パーソナルデータは「新しい石油」
- パーソナルデータの取扱いは，計算機科学，経済学，法学などを含む学際的な課題
- 多くの問題が未解決のまま残されている