ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Online | Mobile | Global

Digital Epidemiology in the Age of Big Data

Specific Challenges around data analytics for Social Media Data

Prof. Marcel Salathé, Digital Epidemiology Lab, EPFL

@marcelsalathe

Journal of Infectious Diseases, 14. Nov 2016

Digital Pharmacovigilance and Disease Surveillance: Combining Traditional and Big-Data Systems for Better Public Health

Marcel Salathé

Digital Epidemiology Laboratory, School of Life Sciences and School of Computer and Communication Sciences, EPFL, Geneva, Switzerland

The digital revolution has contributed to very large data sets (ie, big data) relevant for public health. The two major data sources are electronic health records from traditional health systems and patient-generated data. As the two data sources have complementary strengths—high veracity in the data from traditional sources and high velocity and variety in patient-generated data—they can be combined to build more-robust public health systems. However, they also have unique challenges. Patient-generated data in particular are often completely unstructured and highly context dependent, posing essentially a machine-learning challenge. Some recent examples from infectious disease surveillance and adverse drug event monitoring demonstrate that the technical challenges can be solved. Despite these advances, the problem of verification remains, and unless traditional and digital epidemiologic approaches are combined, these data sources will be constrained by their intrinsic limits.

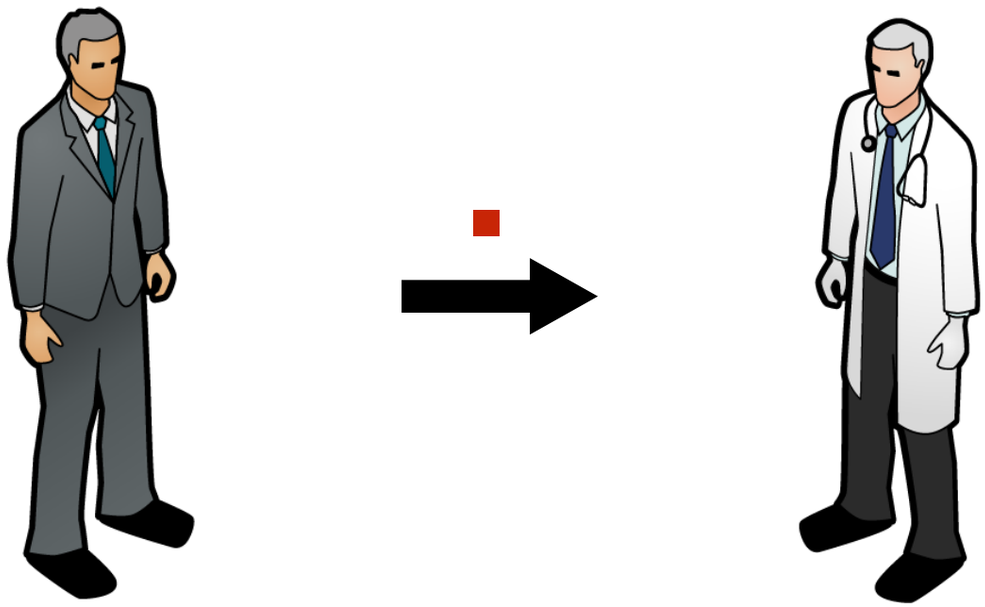
Keywords. digital epidemiology; disease surveillance; pharmagovigilance; Twitter.

Traditional Epidemiology

CDC et al.

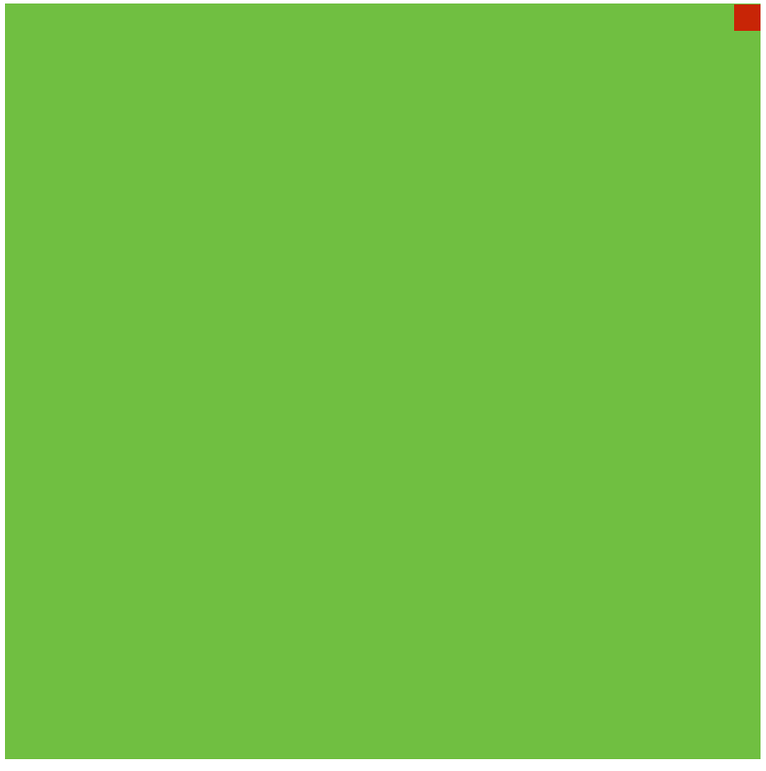
Academia

Traditional Epidemiology

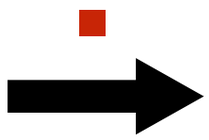


CDC et al.

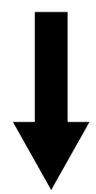
Academia



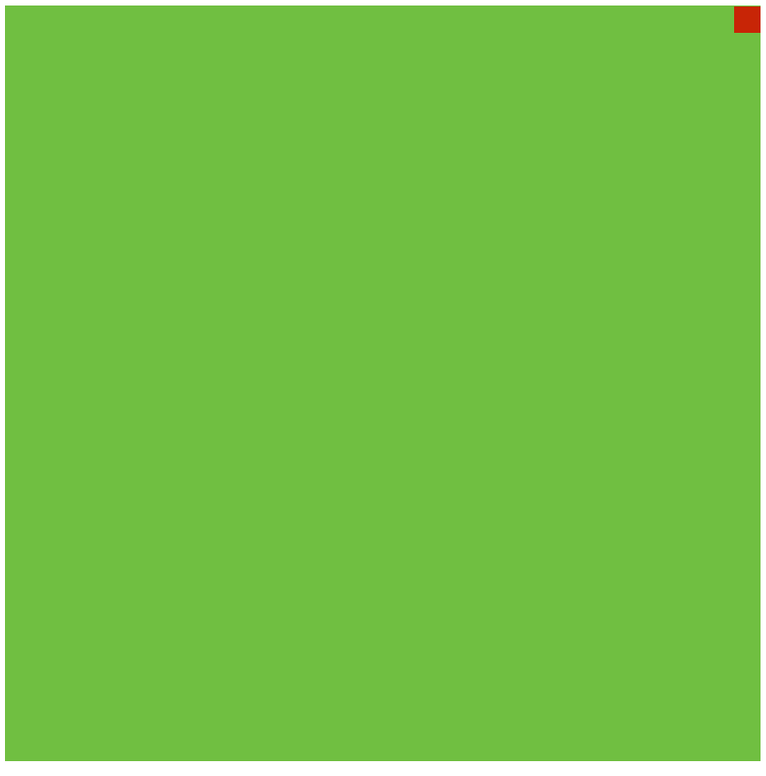
Traditional Epidemiology



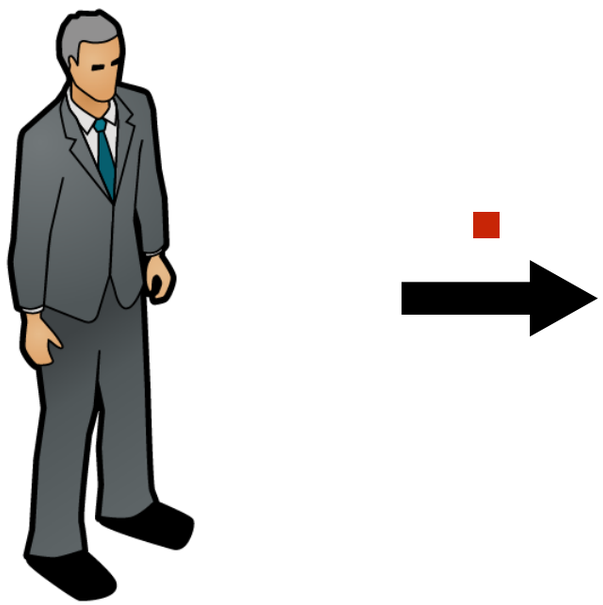
CDC et al.

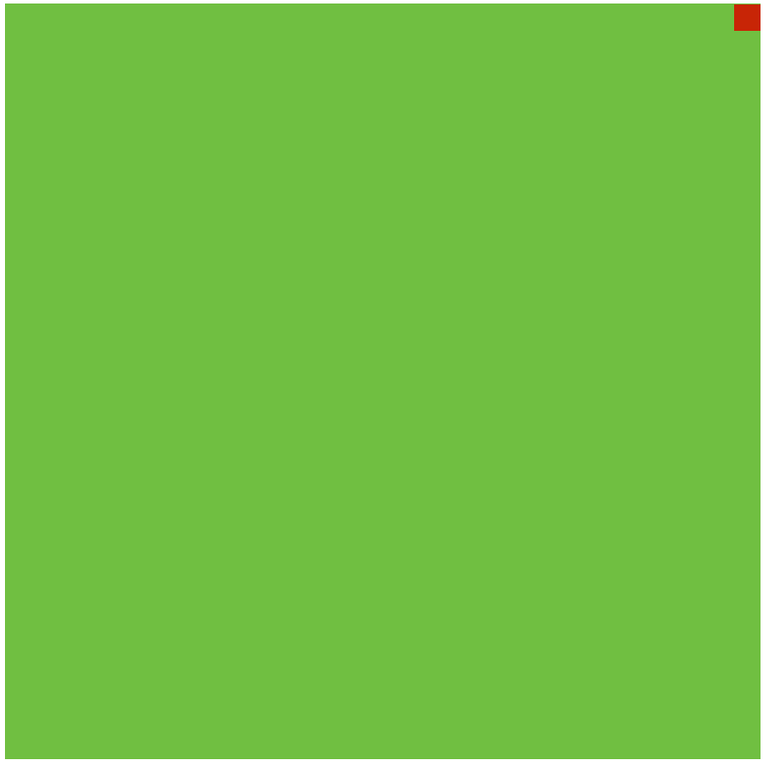


Academia

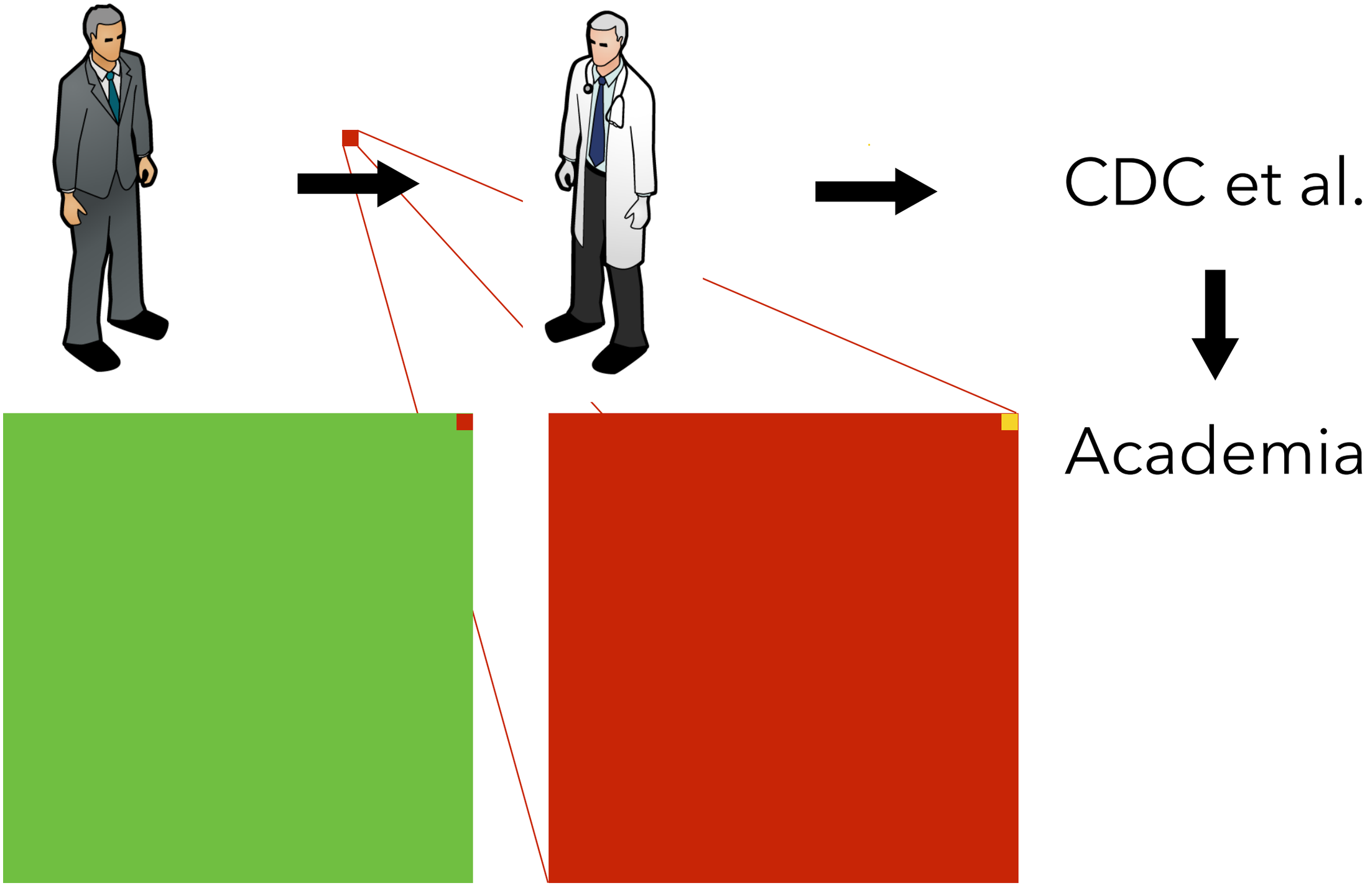


Traditional Epidemiology

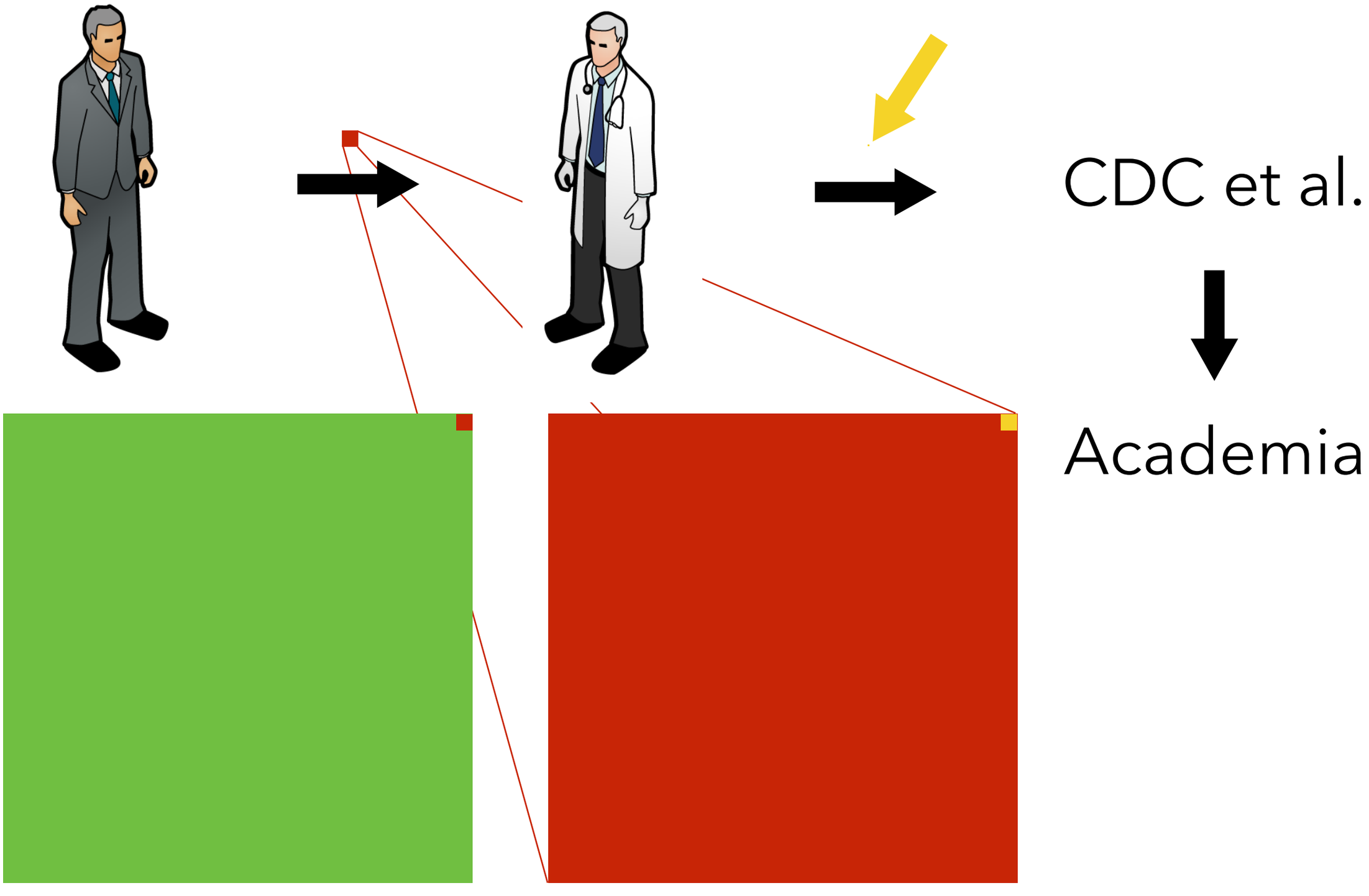




Traditional Epidemiology



Traditional Epidemiology



Digital Epidemiology: new data streams

"Got my flu shot this morning and now my throat is sore."

"Stomach flu & normal flu in the same month. I'm officially a germaphobe."

"Such an upset stomach today. I hope it's just a bug and not the Truvada."

Text
Images
Videos
Sounds
Location
Biological data
etc.

"My weight: 170.1 lb. 10.1 lb to go. #raceweight @Withings scale auto-tweets my weight once a week http://withings.com"

From Personalized Health…

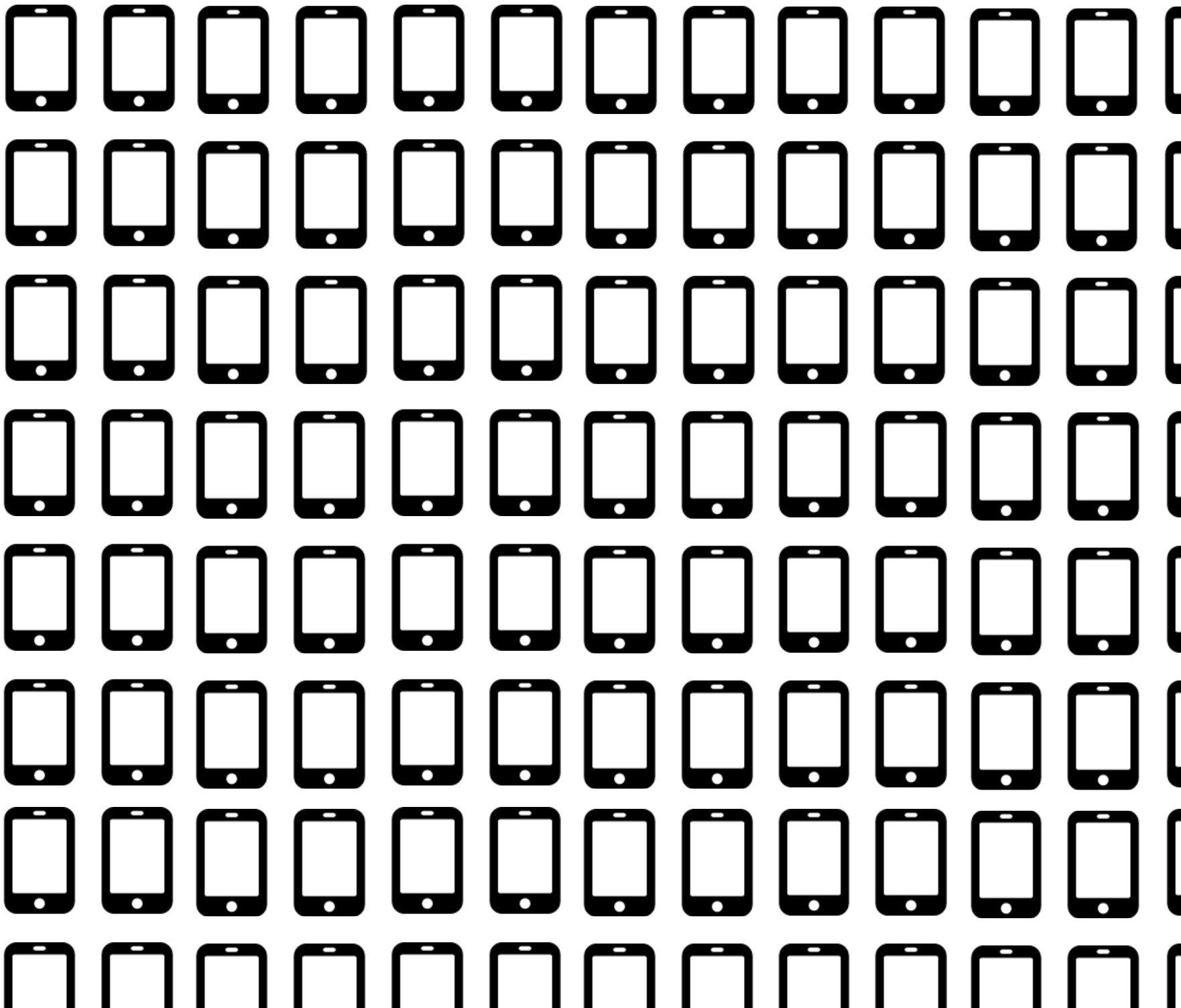
- my DNA
- my *omics data
- my location data
- my activity data (heart rate, etc.)
- my lab tests
- my drugs, side effects
- how I slept
- how I feel
- my health history
- etc.

shared via mobile apps

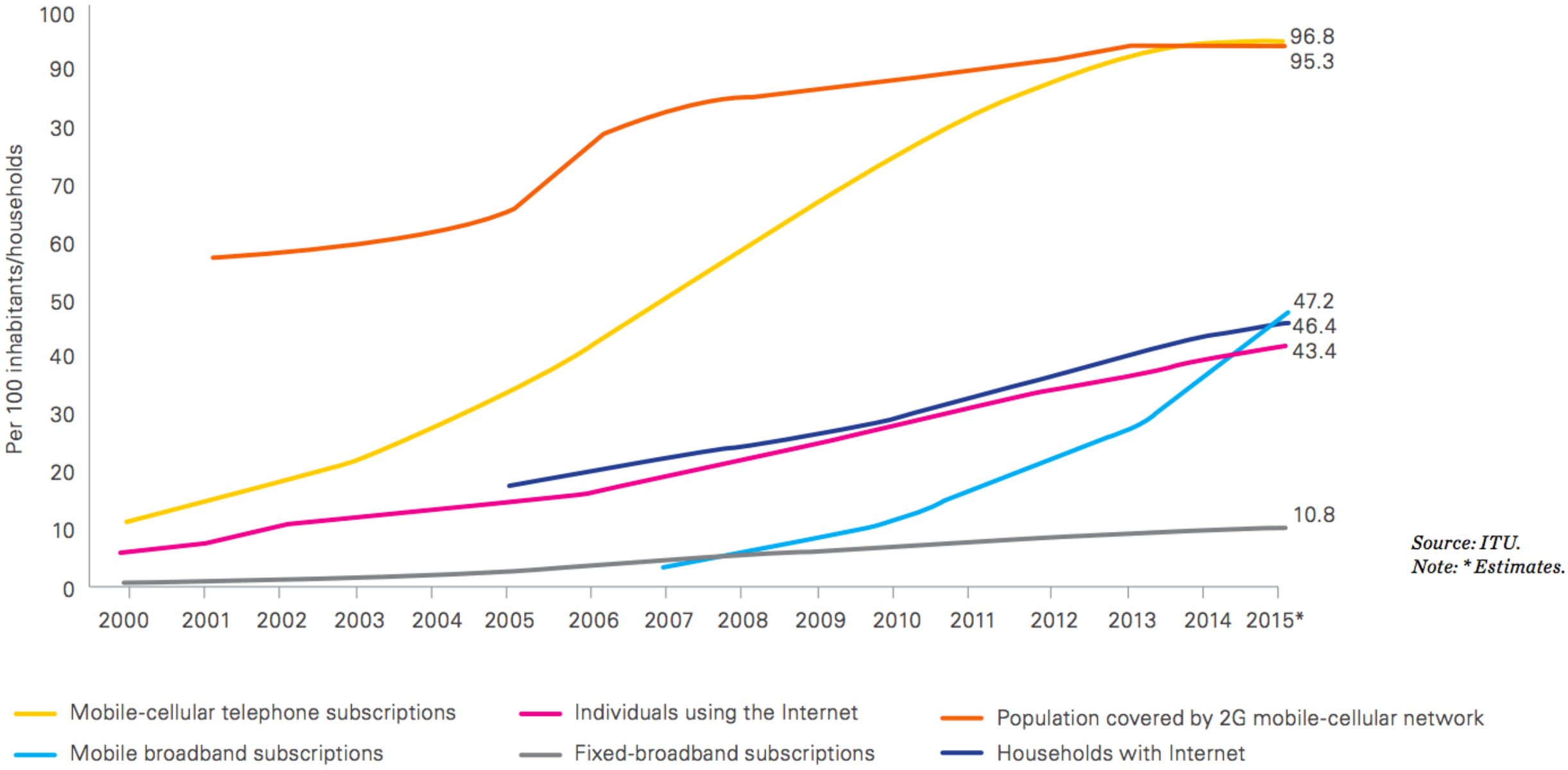
“The patient will see you now”

...To Truly Global Health

shared via mobile apps

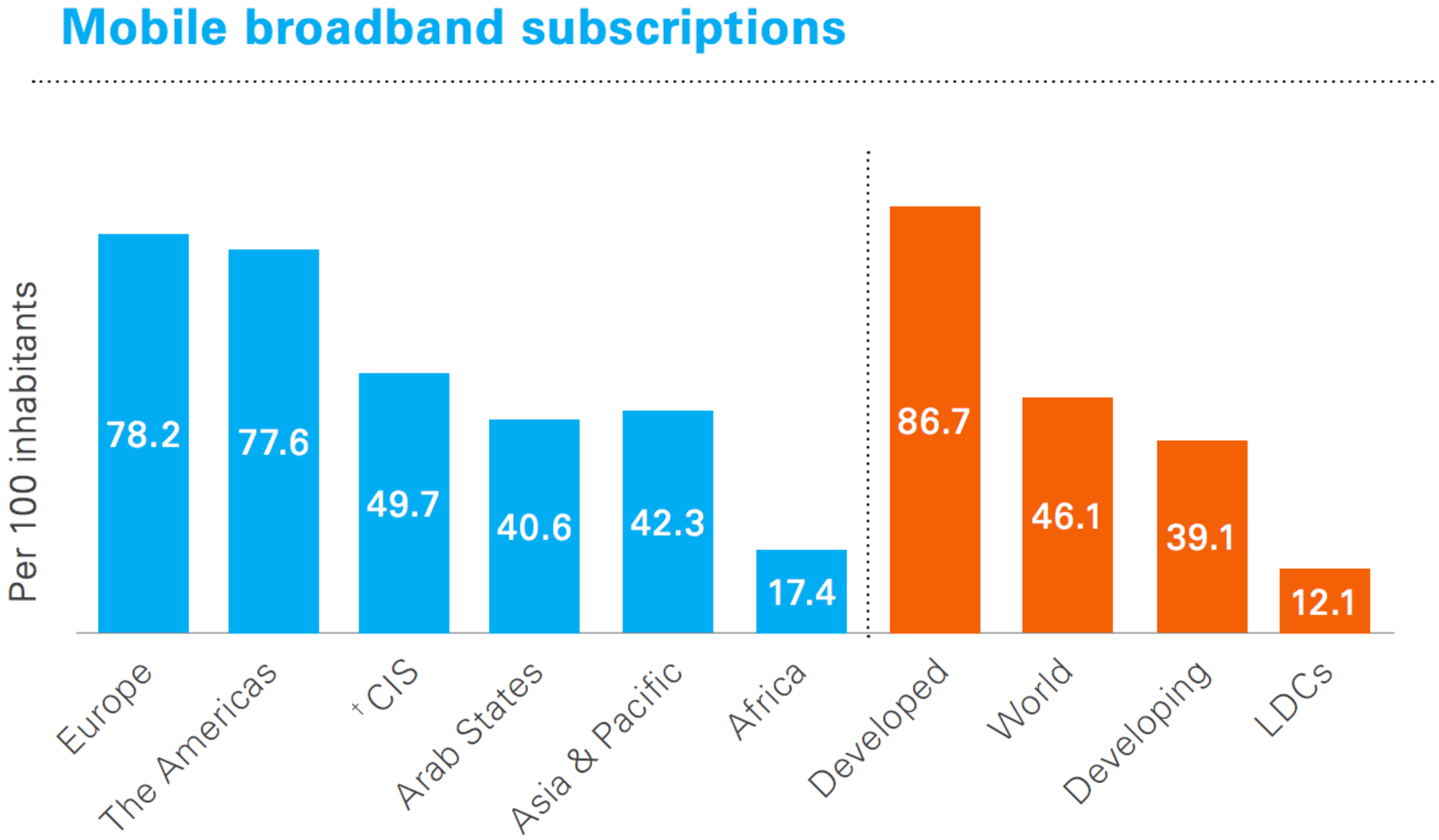


Mobile broadband



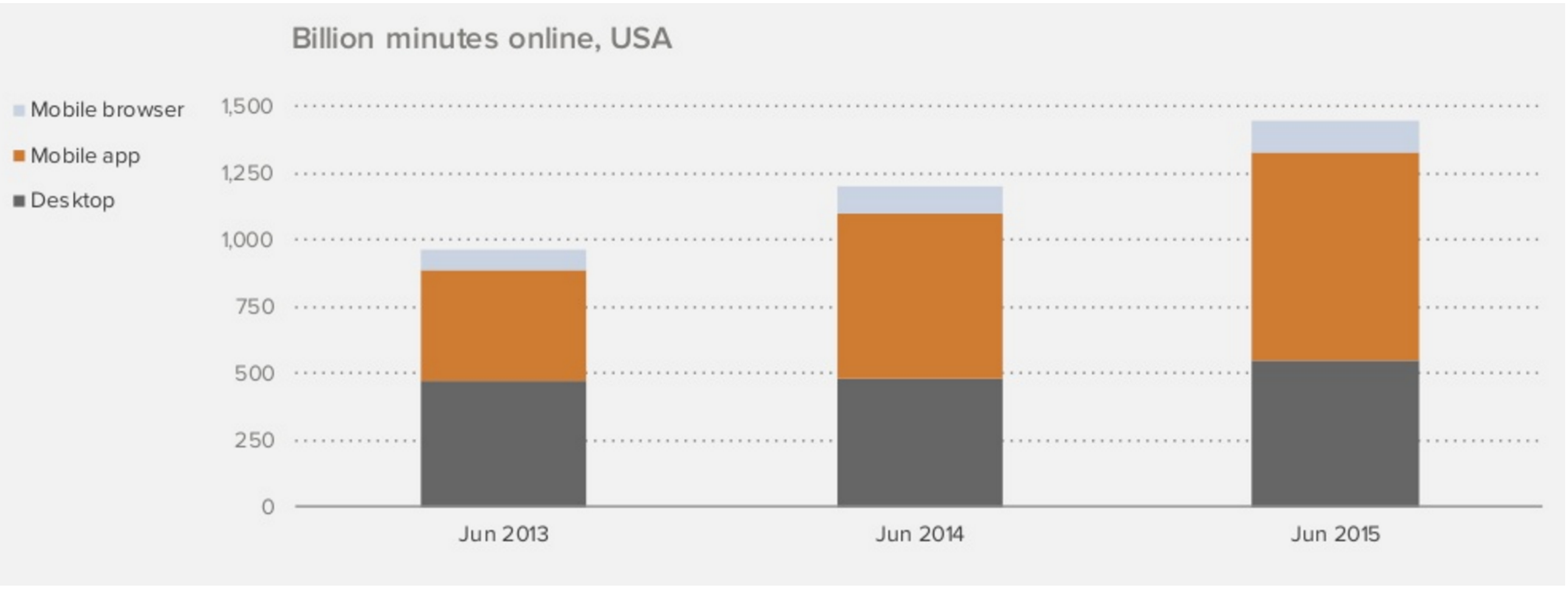
https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf

Mobile broadband



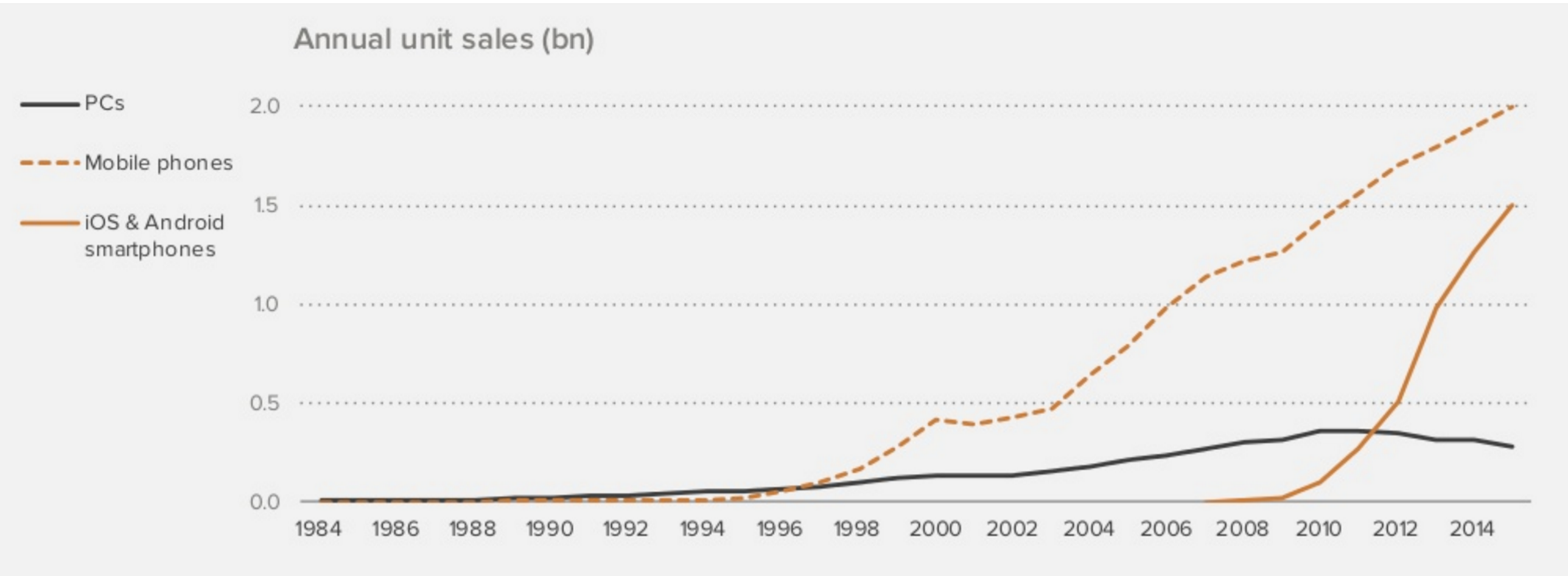
https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf

More time spent on mobile apps than all of web



http://ben-evans.com/benedictevans/?tag=Presentations

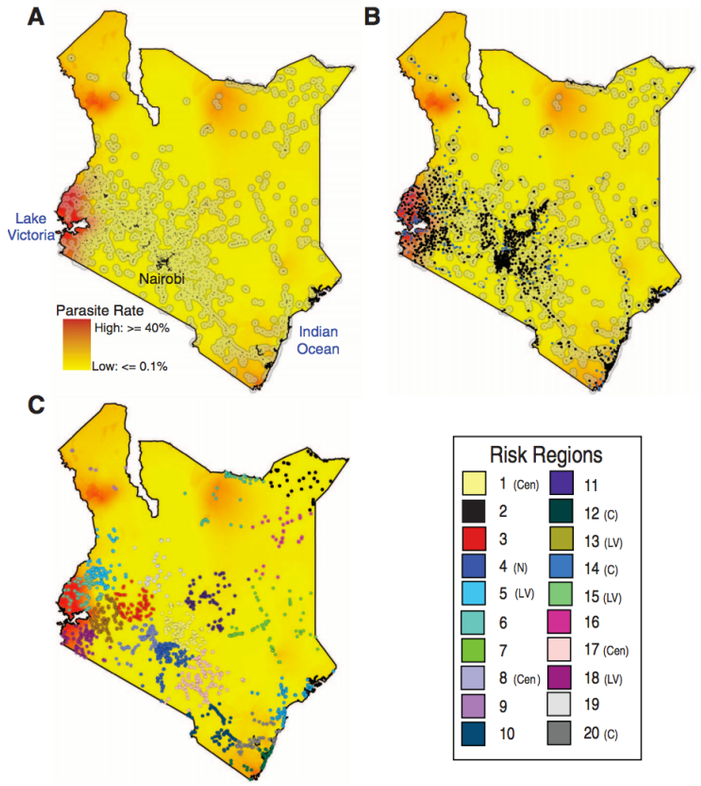
Mobile phone & smart phone revolutions dwarf the PC revolution



http://ben-evans.com/benedictevans/?tag=Presentations

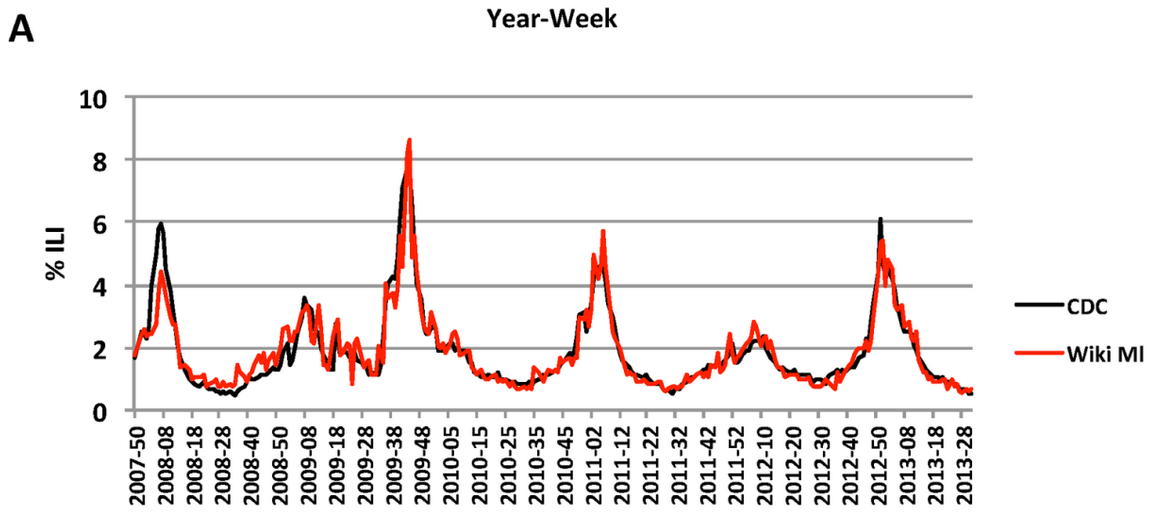
Mobile phones - Malaria elimination

Wesolowski et al 2012



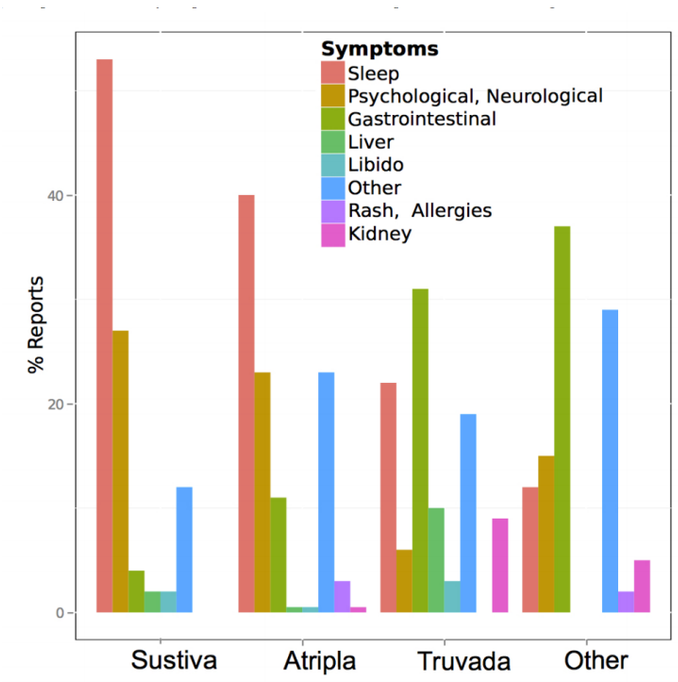
Wikipedia - Influenza forecasting

McIver & Brownstein 2014



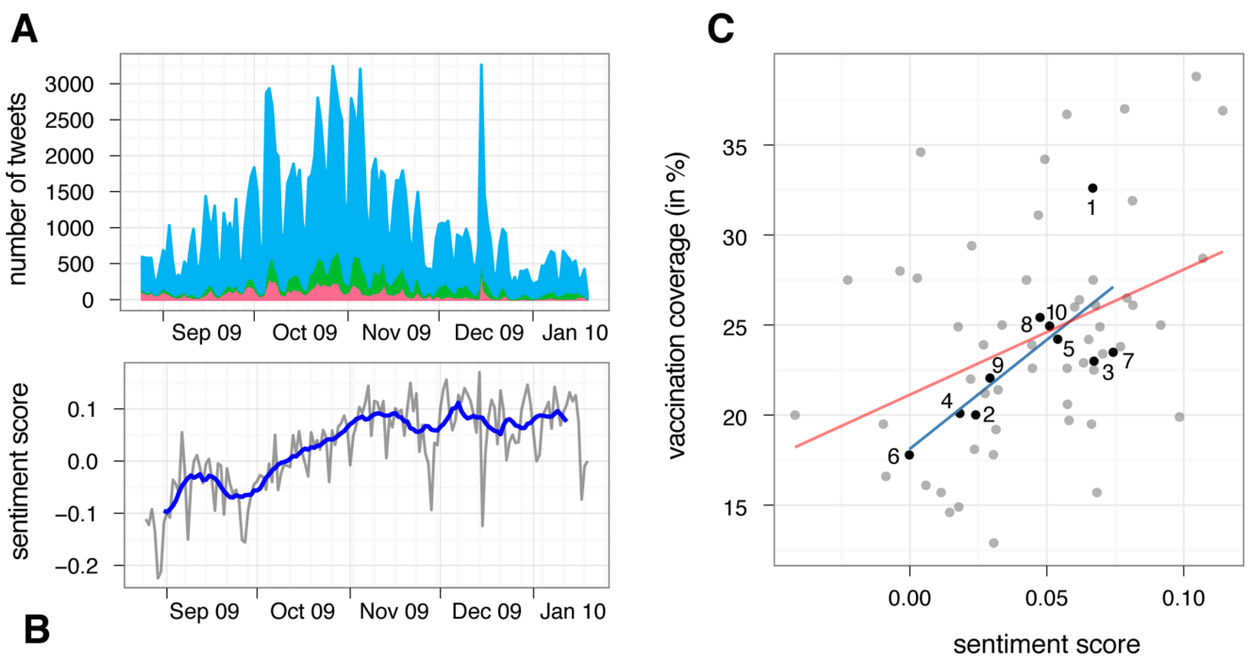
Twitter - Pharmacovigilance

Adrover et al 2015



Twitter - Vaccine uptake

Salathé & Khandelwal, 2011



Telegraph

By **The Telegraph** @Telegraph

Swiss drugmaker stresses that "no causal link" has been discovered between vaccine and deaths

View on web

1

Expand

Bird @suegrant54321 · 7h

Flu Shot Causing Brain Damage Court Awards Millions
experimentalvaccines.org/2014/06/12/flu…

1 2

Angelica Celaya @AngelicaCelaya · 13h

I'm guessing that wasn't a **flu shot** #Constantine

16 34

Active Minds, Inc. @Active_Minds · 22h

“Imagine what it would mean if people felt as comfortable saying they were going for counseling as they were going for a **flu shot**.”

42 35

CDCChronic @CDCChronic · Nov 28

All #cancer survivors should get a #flu shot every year & most should also get pneumonia shot. 1.usa.gov/1uM8LOp

24 8

Dr. Eric Hoskins @DrEricHoskins · Nov 27

It's not too late to get your free **flu shot**. To find the nearest location visit ow.ly/EZi5l

14 8

trutherbot @trutherbot · Nov 27

There is simply no reason to receive a **flu shot** when natural alternatives like vitamin D exist.

106 146

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control

Marcel Salathé*, Shashank Khandelwal

Center for Infectious Disease Dynamics, Department of Biology, Penn State University, University Park, Pennsylvania, United States of America

Abstract

There is great interest in the dynamics of health behaviors in social networks and how they affect collective public health outcomes, but measuring population health behaviors over time and space requires substantial resources. Here, we use publicly available data from 101,853 users of online social media collected over a time period of almost six months to measure the spatio-temporal sentiment towards a new vaccine. We validated our approach by identifying a strong correlation between sentiments expressed online and CDC-estimated vaccination rates by region. Analysis of the network of opinionated users showed that information flows more often between users who share the same sentiments - and less often between users who do not share the same sentiments - than expected by chance alone. We also found that most communities are dominated by either positive or negative sentiments towards the novel vaccine. Simulations of infectious disease transmission show that if clusters of negative vaccine sentiments lead to clusters of unprotected individuals, the likelihood of disease outbreaks is greatly increased. Online social media provide unprecedented access to data allowing for inexpensive and efficient tools to identify target areas for intervention efforts and to evaluate their effectiveness.

Citation: Salathé M, Khandelwal S (2011) Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control. PLoS Comput Biol 7(10): e1002199. doi:10.1371/journal.pcbi.1002199

Editor: Lauren Ancel Meyers, University of Texas at Austin, United States of America

Received May 10, 2011; **Accepted** July 30, 2011; **Published** October 13, 2011

Funding: MS acknowledges funding from Society in Science: the Branco Weiss fellowship. http://www.society-in-science.org/. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

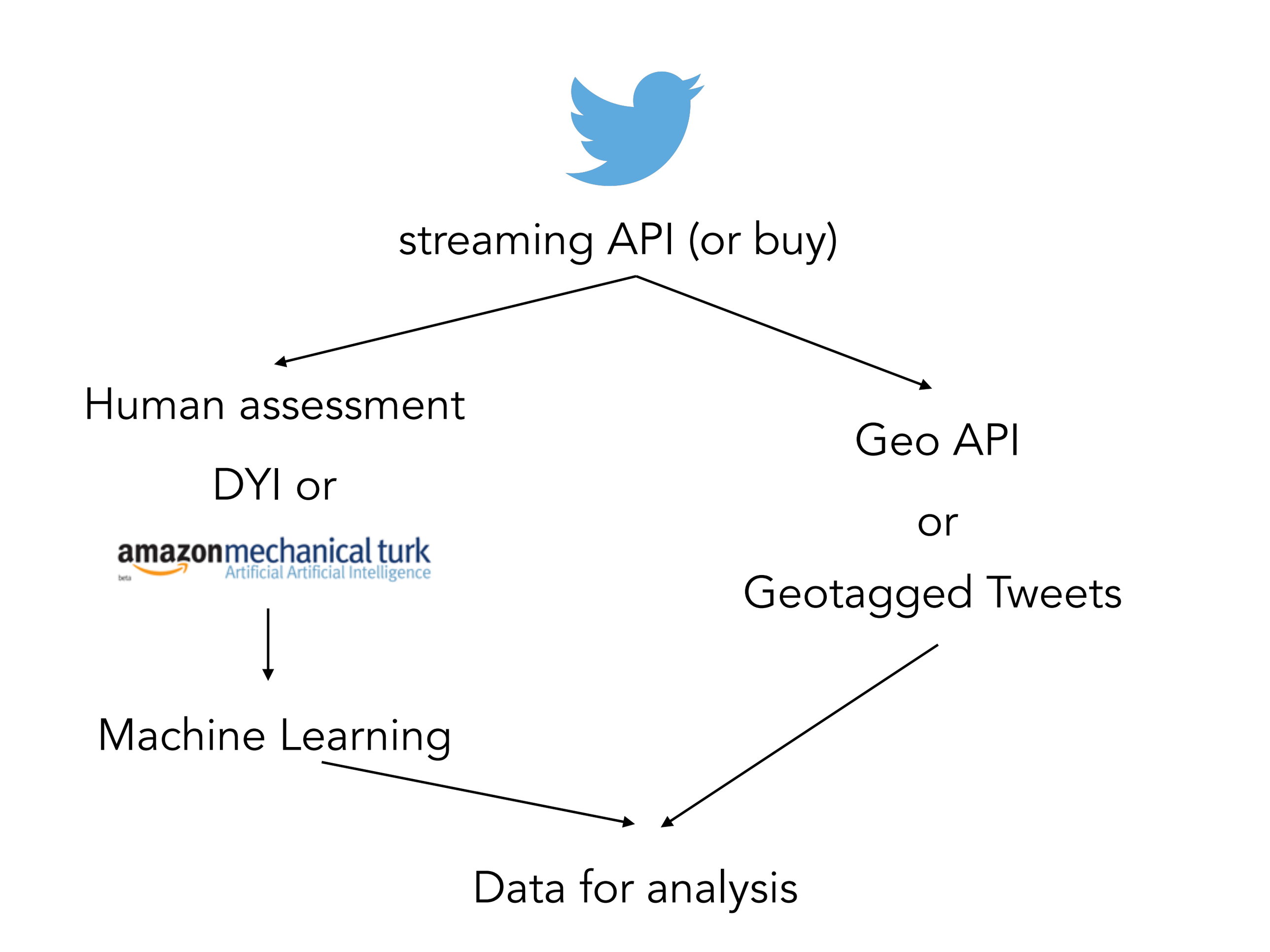
Competing Interests: The authors have declared that no competing interests exist.

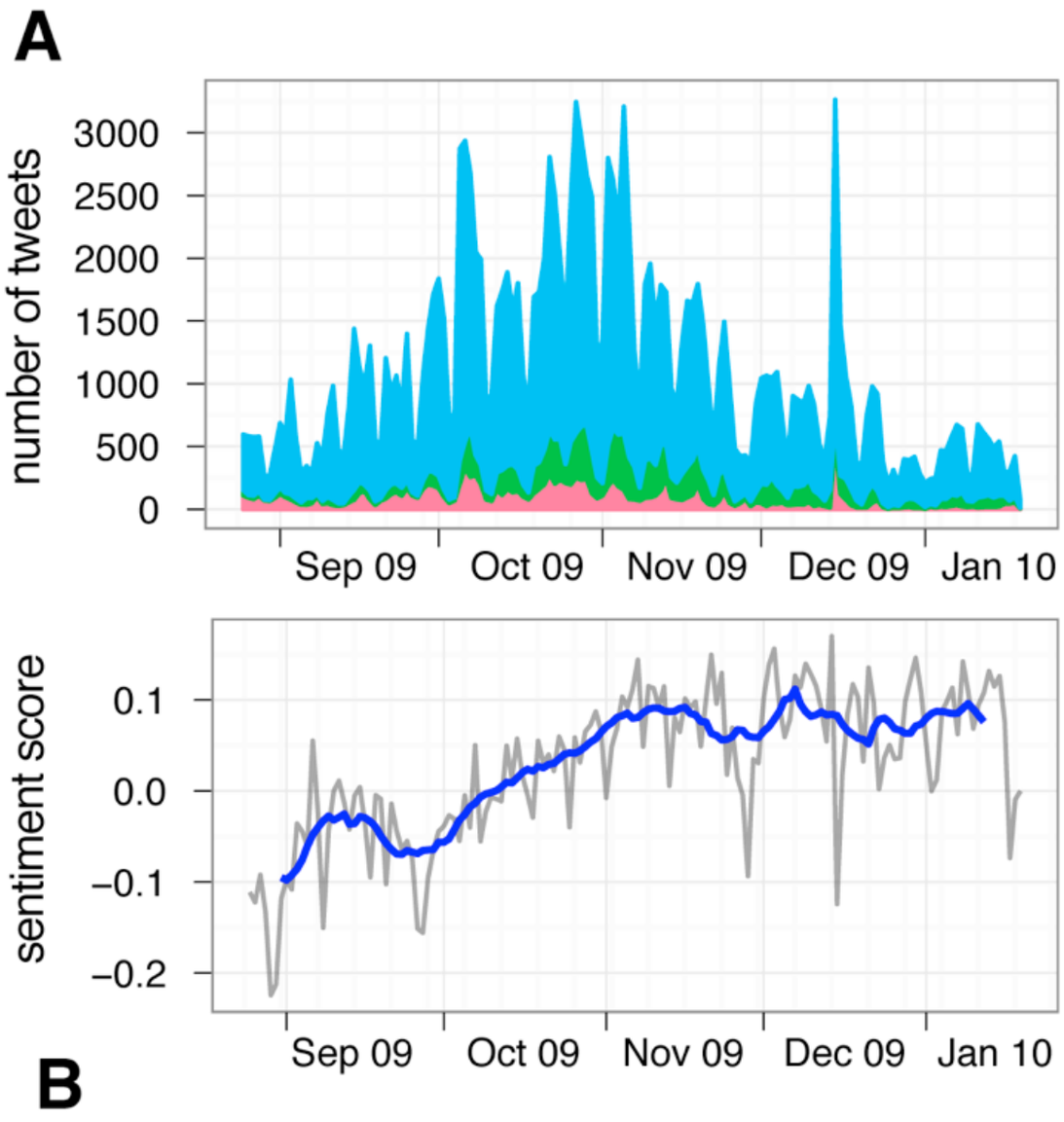
* E-mail: salathe@psu.edu

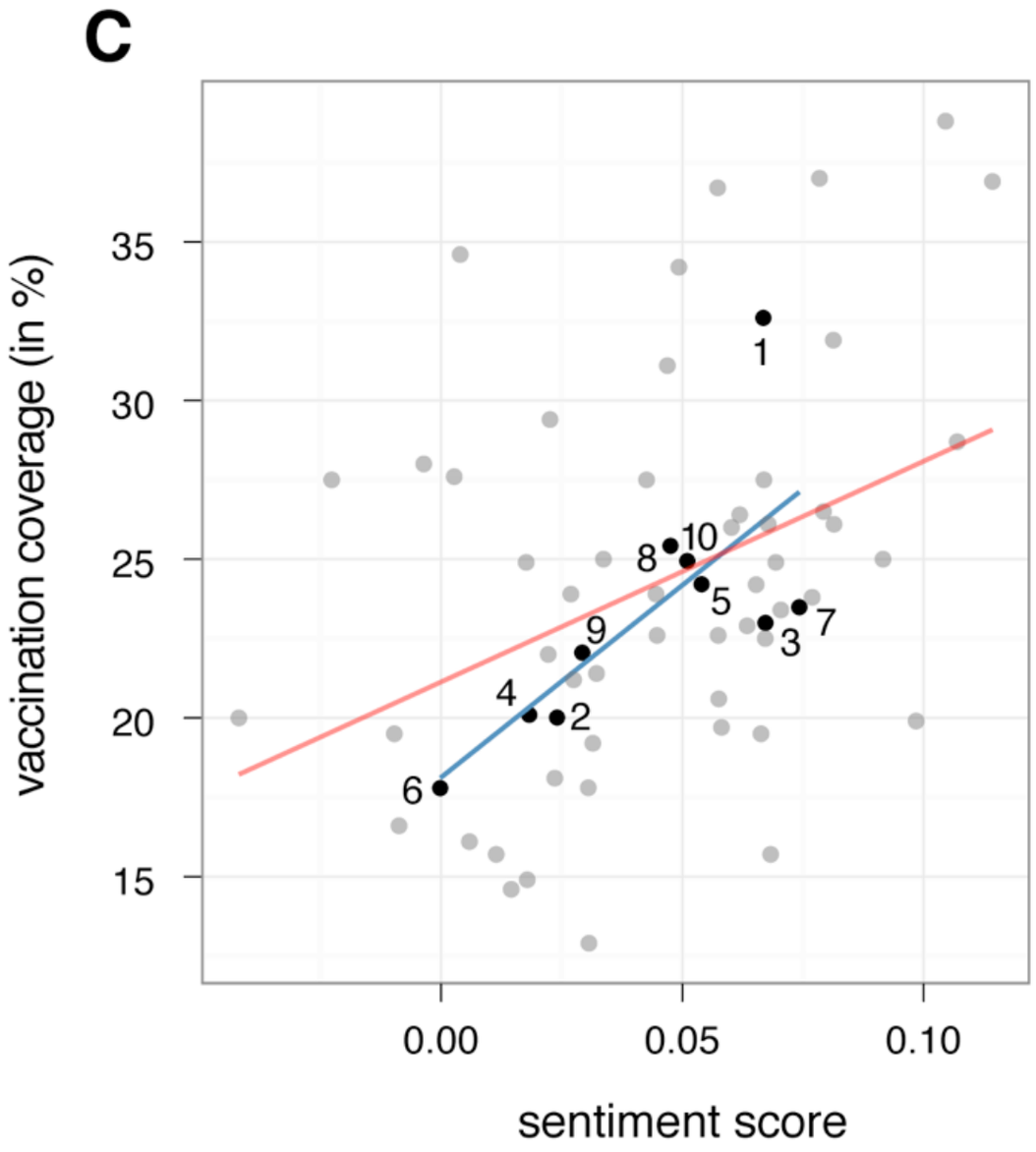
Introduction

Outbreaks of vaccine preventable diseases are a major public health issue. Outbreaks are more likely to occur if either overall vaccination rates decline [1], or if communities with very low

time, pandemic influenza A(H1N1) was spreading nationwide but a vaccine became widely available only very late in the year. We collected practically all publicly available text messages on Twitter (so called "tweets") containing English keywords relating to vaccination as well as location information provided by the


streaming API (or buy)
Human assessment
DYI or
amazonmechanical turk
Artificial Artificial Intelligence
Machine Learning
Geo API
or
Geotagged Tweets
Data for analysis





(~10% Tweets were assessed by humans, rest by machine learning algorithms)

Salathé & Khandelwal, PLOS Computational Biology, 2011

Salathé et al. *EPJ Data Science* 2013, **2**:4
http://www.epjdatascience.com/content/2/1/4

EPJ Data Science
a SpringerOpen Journal

EPJ.org

REGULAR ARTICLE Open Access

The dynamics of health behavior sentiments on a large online social network

Marcel Salathé[1,2,3*], Duy Q Vu[4], Shashank Khandelwal[1,2] and David R Hunter[1,4]

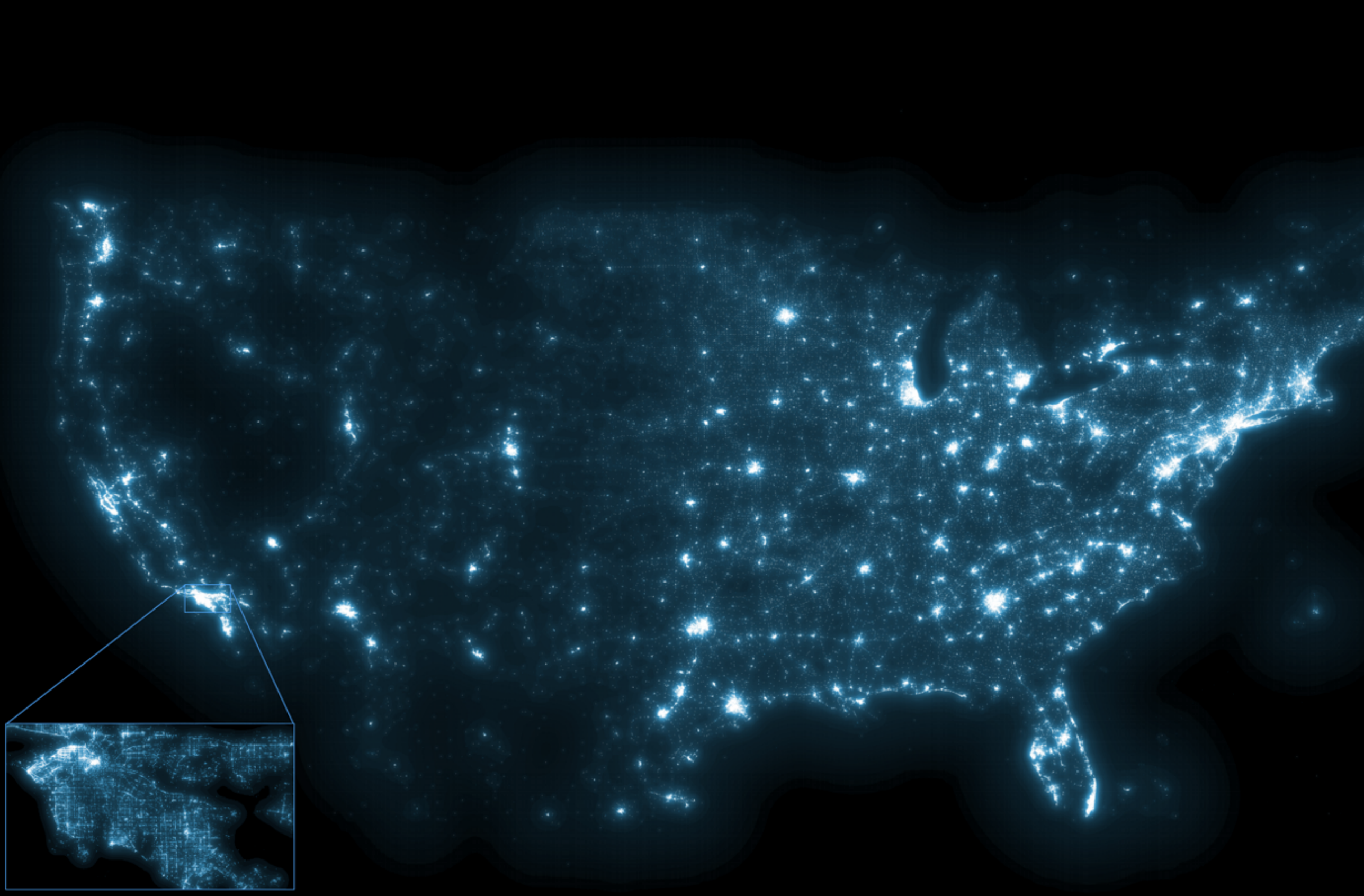
*Correspondence: salathe@psu.edu
[1]Center for Infectious Disease Dynamics, Penn State University, University Park, PA, USA
[2]Department of Biology, Penn State University, University Park, PA, USA
Full list of author information is available at the end of the article

Abstract

Modifiable health behaviors, a leading cause of illness and death in many countries, are often driven by individual beliefs and sentiments about health and disease. Individual behaviors affecting health outcomes are increasingly modulated by social networks, for example through the associations of like-minded individuals - homophily - or through peer influence effects. Using a statistical approach to measure the individual temporal effects of a large number of variables pertaining to social network statistics, we investigate the spread of a health sentiment towards a new vaccine on Twitter, a large online social network. We find that the effects of neighborhood size and exposure intensity are qualitatively very different depending on the type of sentiment. Generally, we find that larger numbers of opinionated neighbors inhibit the expression of sentiments. We also find that exposure to negative sentiment is contagious - by which we merely mean predictive of future negative sentiment expression - while exposure to positive sentiments is generally not. In fact, exposure to positive sentiments can even predict increased negative sentiment expression. Our results suggest that the effects of peer influence and social contagion on the dynamics of behavioral spread on social networks are strongly content-dependent.

Keywords: social media; social network; diffusion; health behavior; contagion

Social networks play an important role in affecting the dynamics of health behaviors and the associated diseases [1–3], but identifying the main drivers of health behavior spread



Salathé et al. PLOS Computational Biology 2012

https://www.mapbox.com/blog/twitter-map-every-tweet/

https://www.mapbox.com/blog/twitter-map-every-tweet/

PlantVillage: Machine Learning for Disease Recognition

Image Data

Collecting 1M+ labelled images as training set for machine learning algorithm development

Machine Learning

Crowdsourced, open machine learning competitions based on open access images

[Collaboration with Penn State, Prof. David Hughes]

METHODS ARTICLE

Front. Plant Sci. | doi: 10.3389/fpls.2016.01419

Using Deep Learning for Image-Based Plant Disease Detection

 **Sharada P. Mohanty**[1],  **David P. Hughes**[2] and  **Marcel Salathé**[1*]

[1]EPFL, Switzerland

[2]Penn State University, USA

Crop diseases are a major threat to food security, but their rapid identification remains difficult in many parts of the world due to the lack of the necessary infrastructure. The combination of increasing global smartphone penetration and recent advances in computer vision made possible by deep learning has paved the way for smartphone-assisted disease diagnosis. Using a public dataset of 54,306 images of diseased and healthy plant leaves collected under controlled conditions, we train a deep convolutional neural network to identify 14 crop species and 26 diseases (or absence thereof). The trained model achieves an accuracy of 99.35% on a held-out test set, demonstrating the feasibility of this approach. Overall, the approach of training deep learning models on increasingly large and publicly available image datasets presents a clear path towards smartphone-assisted crop disease diagnosis on a massive global scale.

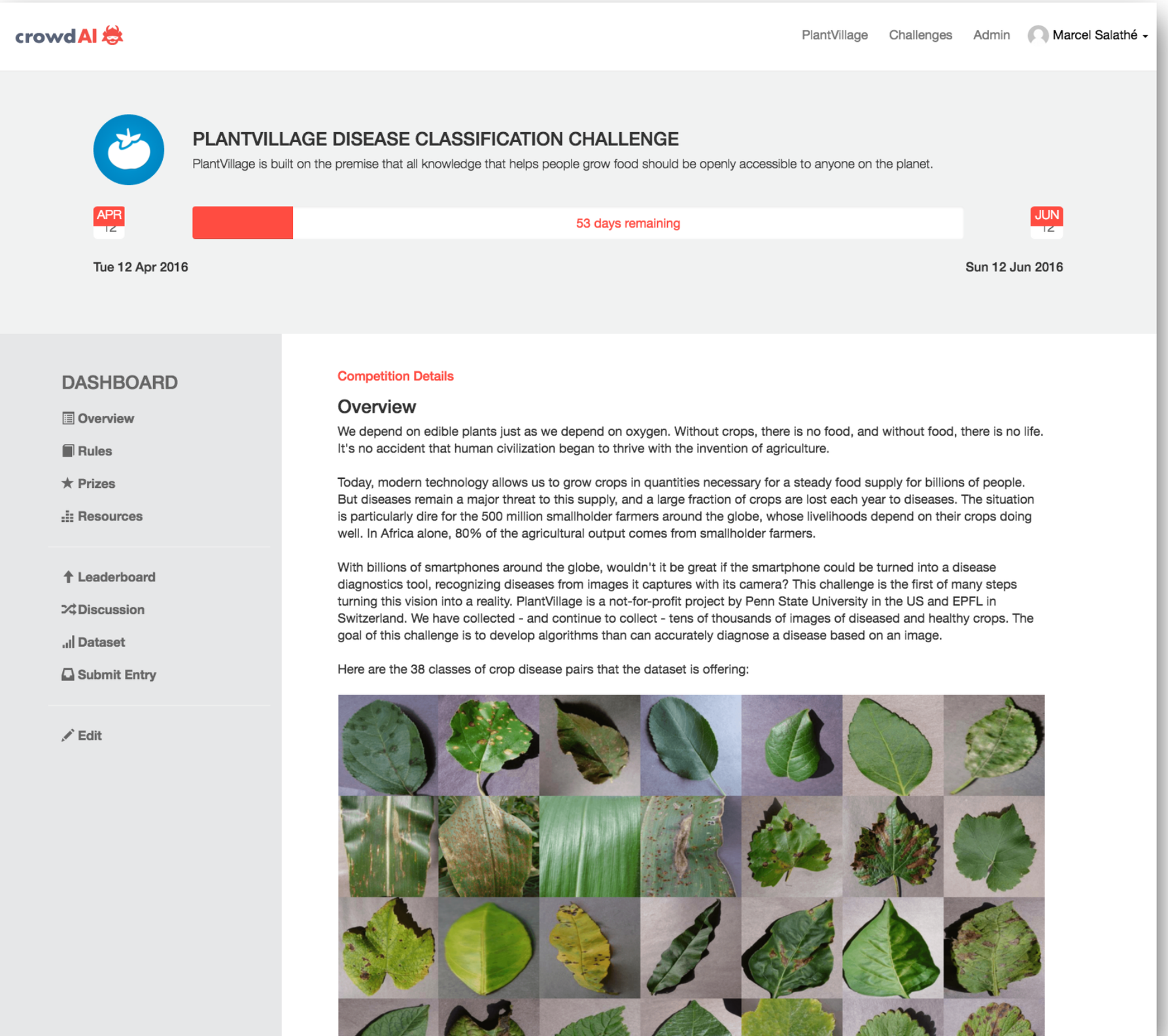
PlantVillage: Machine Learning for Disease Recognition

Image Recognition
(Machine Learning)

Diagnosis
Treatment Suggestions

crowdsourcing

Crowdsourcing Machine Learning



From Big Data to Knowledge Generation, *fast*?

In science, increasingly through crowdsourcing.

Netflix Challenge, Kaggle challenges, ImageNet Challenge, etc.

www.crowdai.org

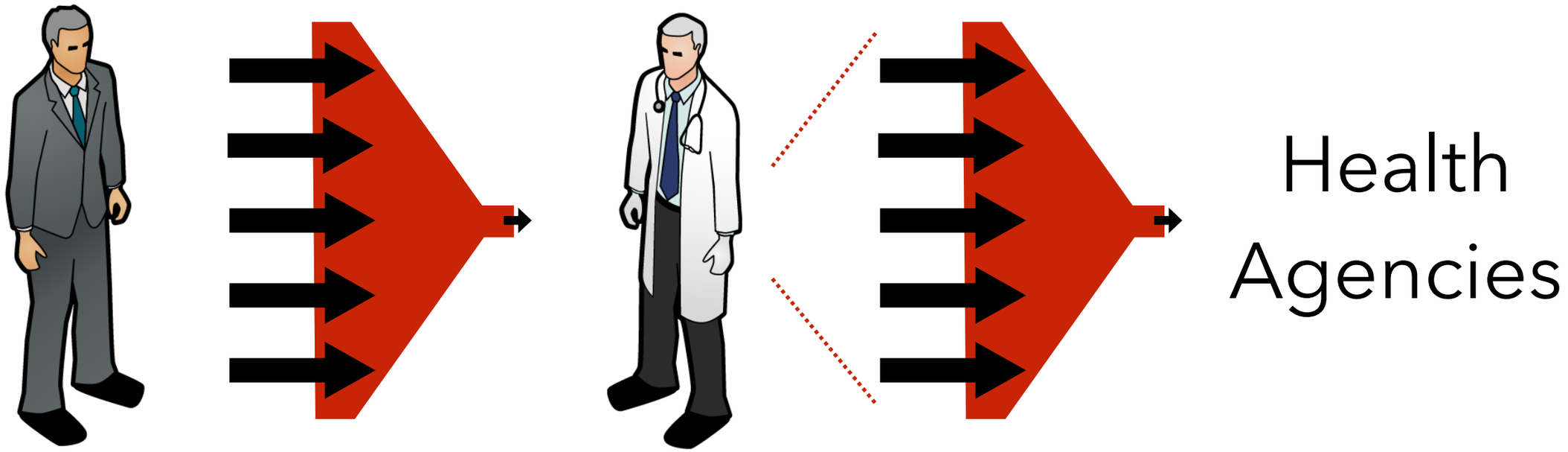
From Personalized Health…

- my DNA
- my *omics data
- my location data
- my activity data (heart rate, etc.)
- my lab tests
- my drugs, side effects
- how I slept
- how I feel
- my health history
- etc.

shared via mobile apps

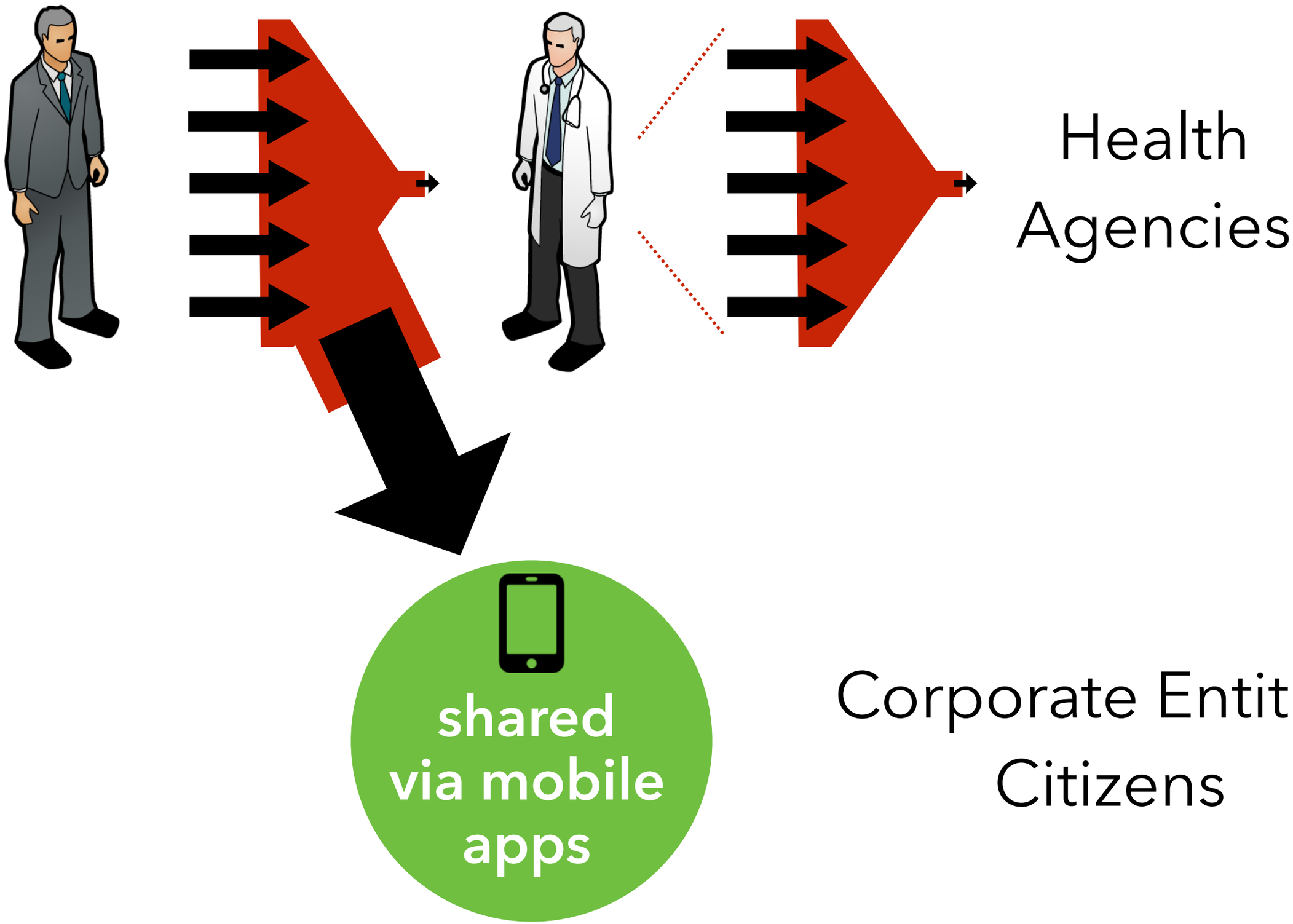
“The patient will see you now”

Follow the data…



Health Agencies

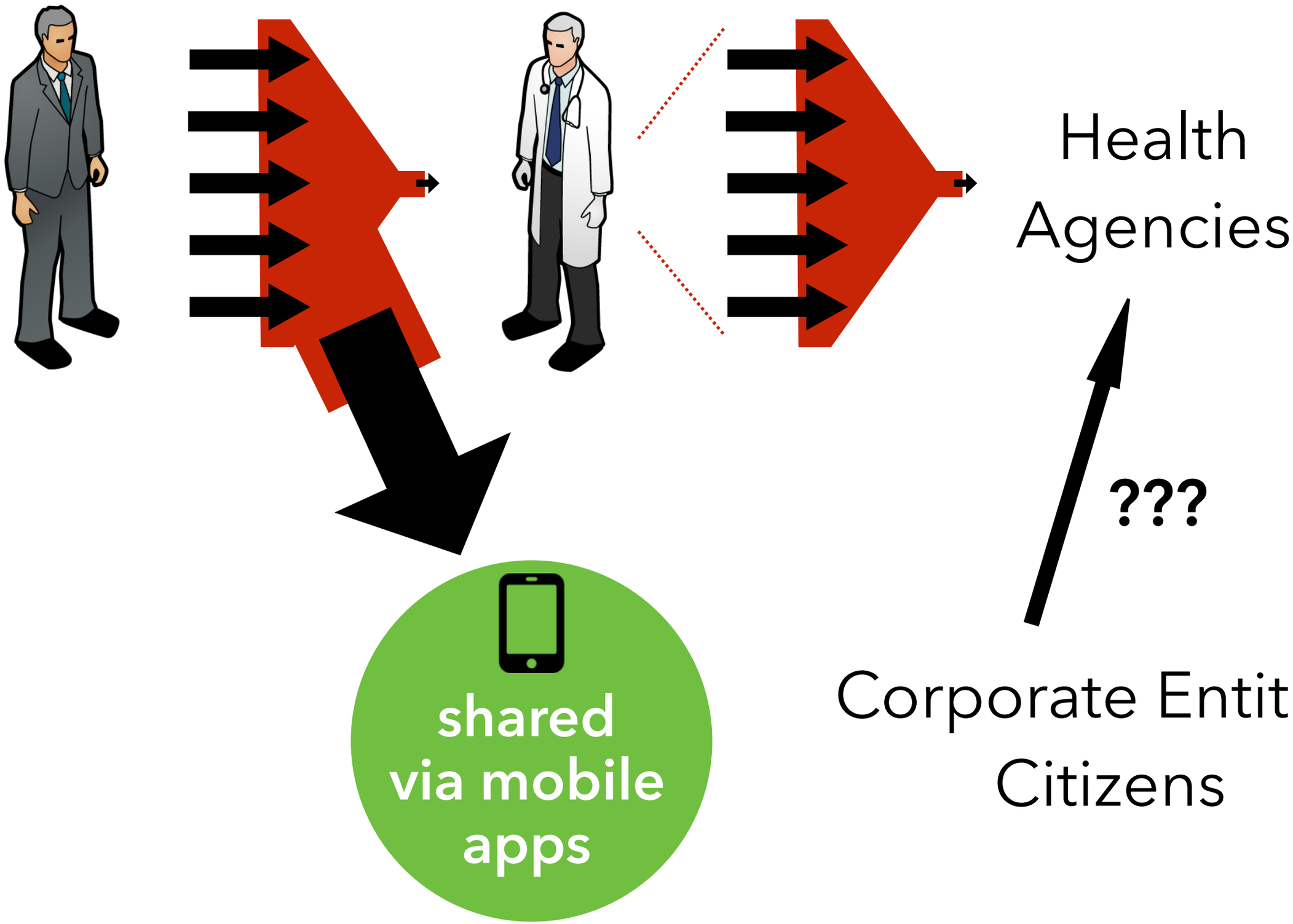
Follow the data…



Health Agencies

Corporate Entities, Citizens

Follow the data…



Health
Agencies

???

Corporate Entities,
Citizens

healthbank.coop

my data, my choice, my healthbank

We empower people across the globe to exchange their health data on our uniquely neutral and independent platform.

healthbank drives innovation in health sciences, from prevention to cure, at a better price with better quality for the benefit of both the individual and society.

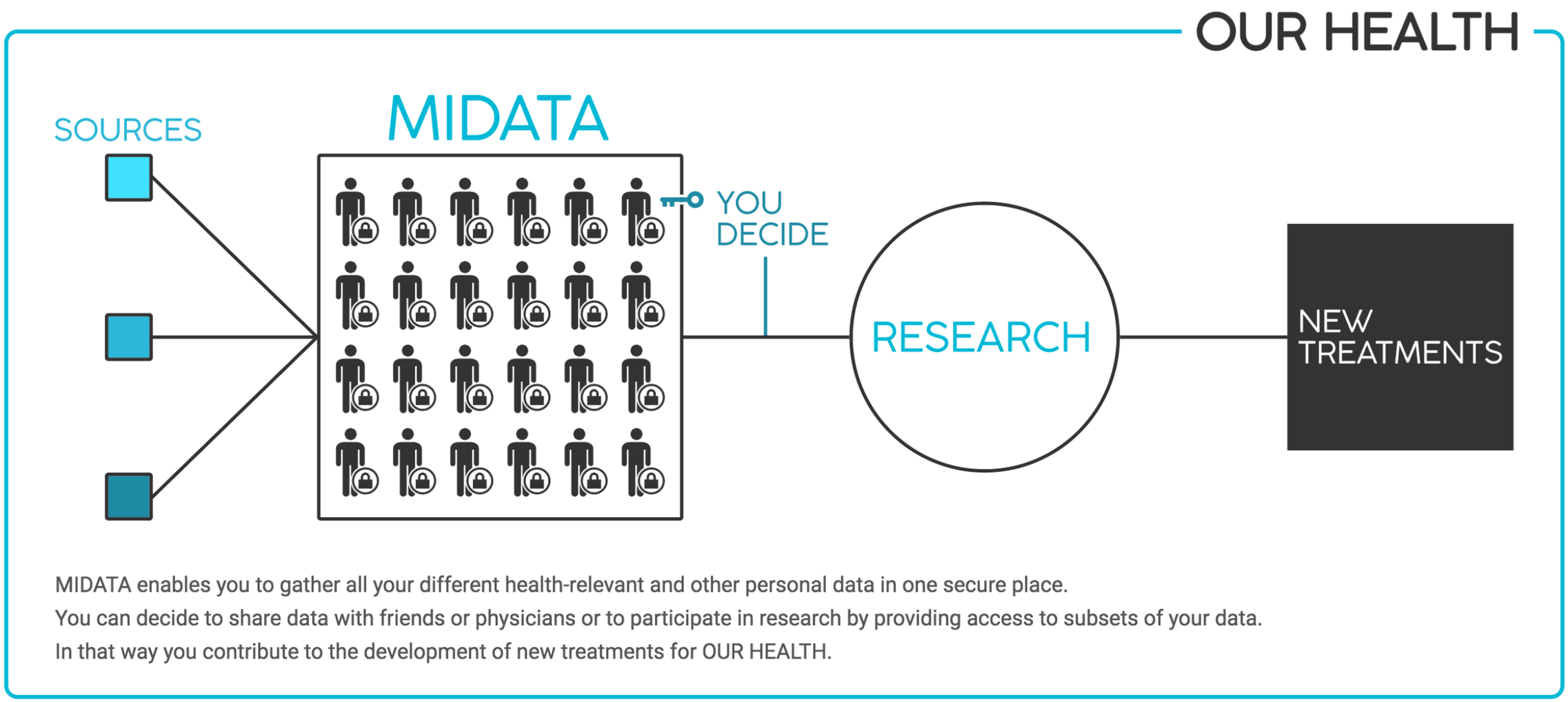
TELL ME MORE

midata.coop

MIDATA.COOP

DE FR IT EN

Home About Projects People Publications Contact



MIDATA enables you to gather all your different health-relevant and other personal data in one secure place.
You can decide to share data with friends or physicians or to participate in research by providing access to subsets of your data.
In that way you contribute to the development of new treatments for OUR HEALTH.

Challenges

- Digitial Epidemiology (data aquisition; data interpretation; machine learning) needs to become an integral part of public health institutions.

- The biggest challenge is data aquisition. We need to establish a "**WHO of data"** - health institutions are increasingly out of the loop, and no company alone has all the relevant data.

- Data can come from citizens, companies, institutions, etc.

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Online | Mobile | Global

Digital Epidemiology in the Age of Big Data

Specific Challenges around data analytics for Social Media Data

Prof. Marcel Salathé, Digital Epidemiology Lab, EPFL

@marcelsalathe