

# small, n=me, data

Deborah Estrin

Professor, Computer Science, Cornell NYC Tech  
Professor, Public Health, Weill Cornell Medical College

[destrin@cs.cornell.edu](mailto:destrin@cs.cornell.edu)

work done with collaborators from  
Cornell, UCLA, openmhealth.org, ...

# Third IT pillar of personalized, precision, medicine

“Big data”  
(EHRs,  
Web mining) + Omics + “small data”  
(mHealth,  
digital traces)



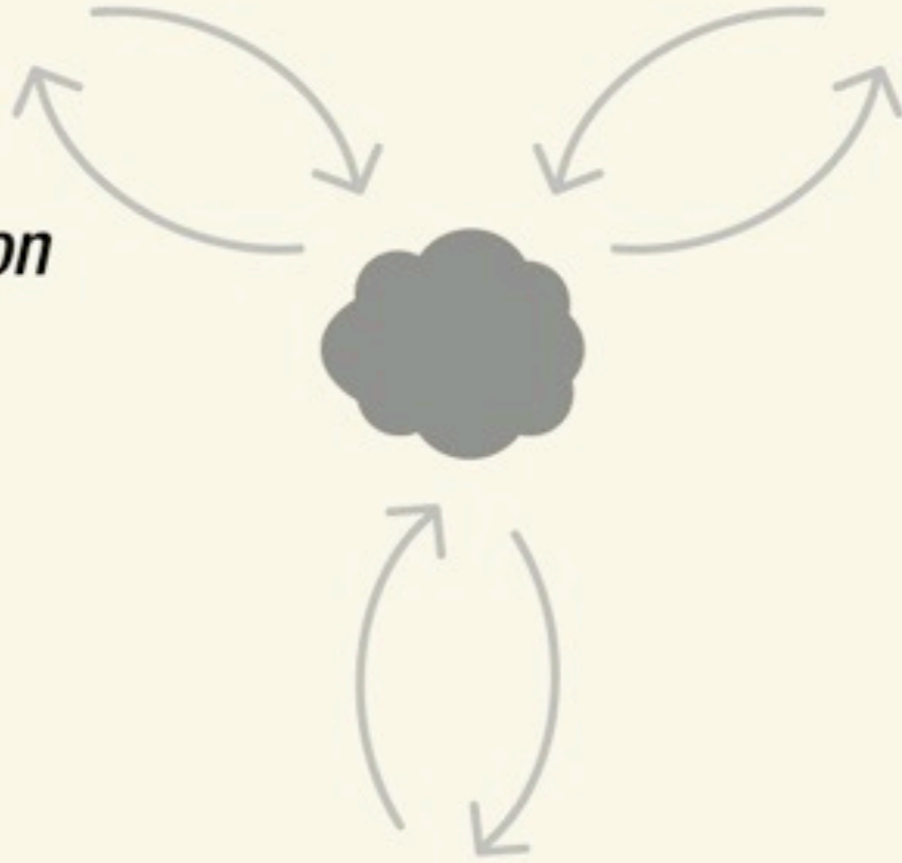
## **Participant self-care**

*How is this new medication working for me?*



## **Clinical care**

*How is the patient responding to new care plan?*



## **Research evidence**

*What works best in different contexts?*

# Profound potential to rephrase 'does it work?'

(Complexes of) Exposures



strength of association?

individual

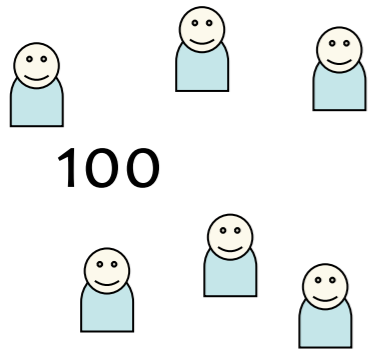


Outcome

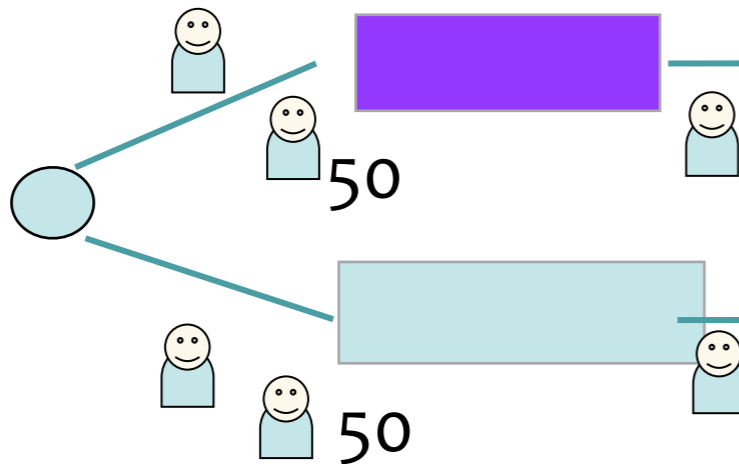


population

## 'does it work on average?' (RCT)



100



50

50

Outcome measure

Outcome measure

population

Outcome

Outcome

Sim,  
Kravitz

# Profound potential to rephrase 'does it work?'

(Complexes of) Exposures



strength of association?

individual

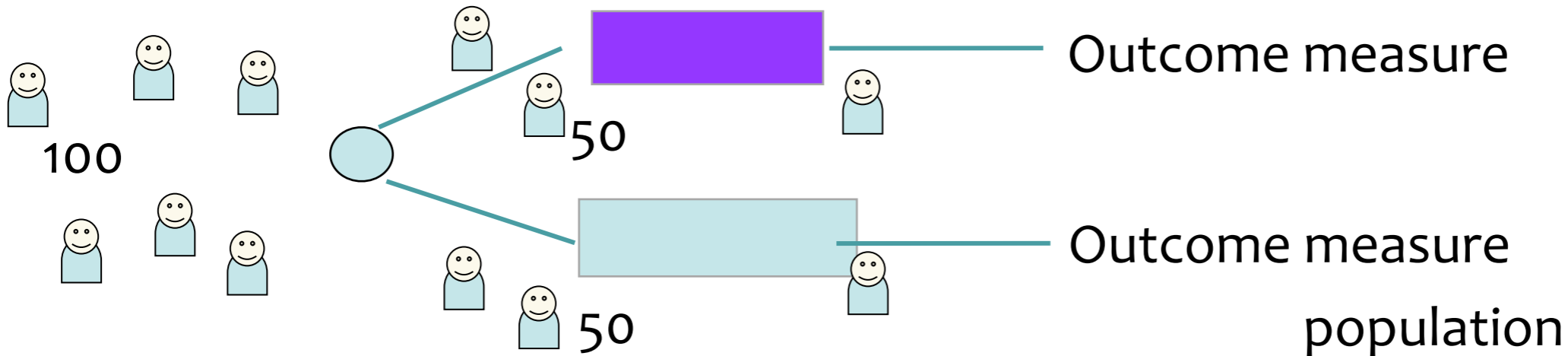


Outcome

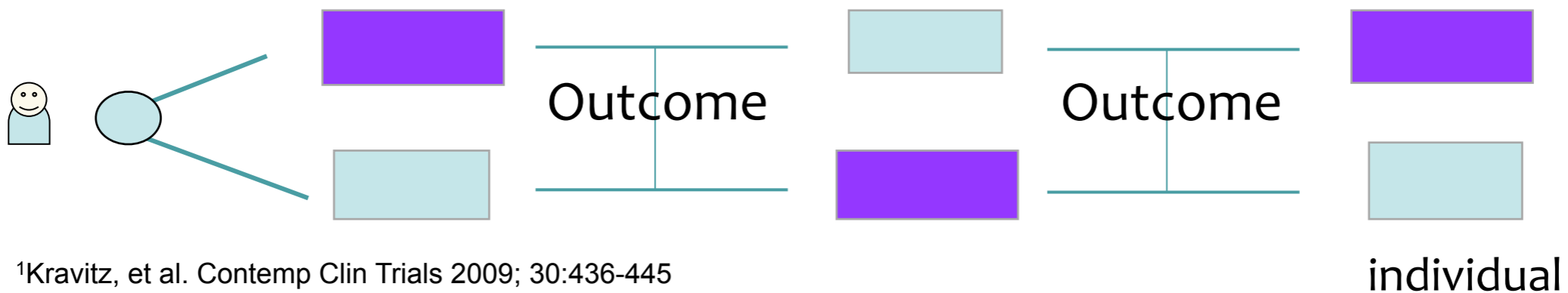


population

## 'does it work on average?' (RCT)



## N-of-1 study design: 'does it work for Mr. Jones?'



<sup>1</sup>Kravitz, et al. Contemp Clin Trials 2009; 30:436-445

Sim,  
Kravitz

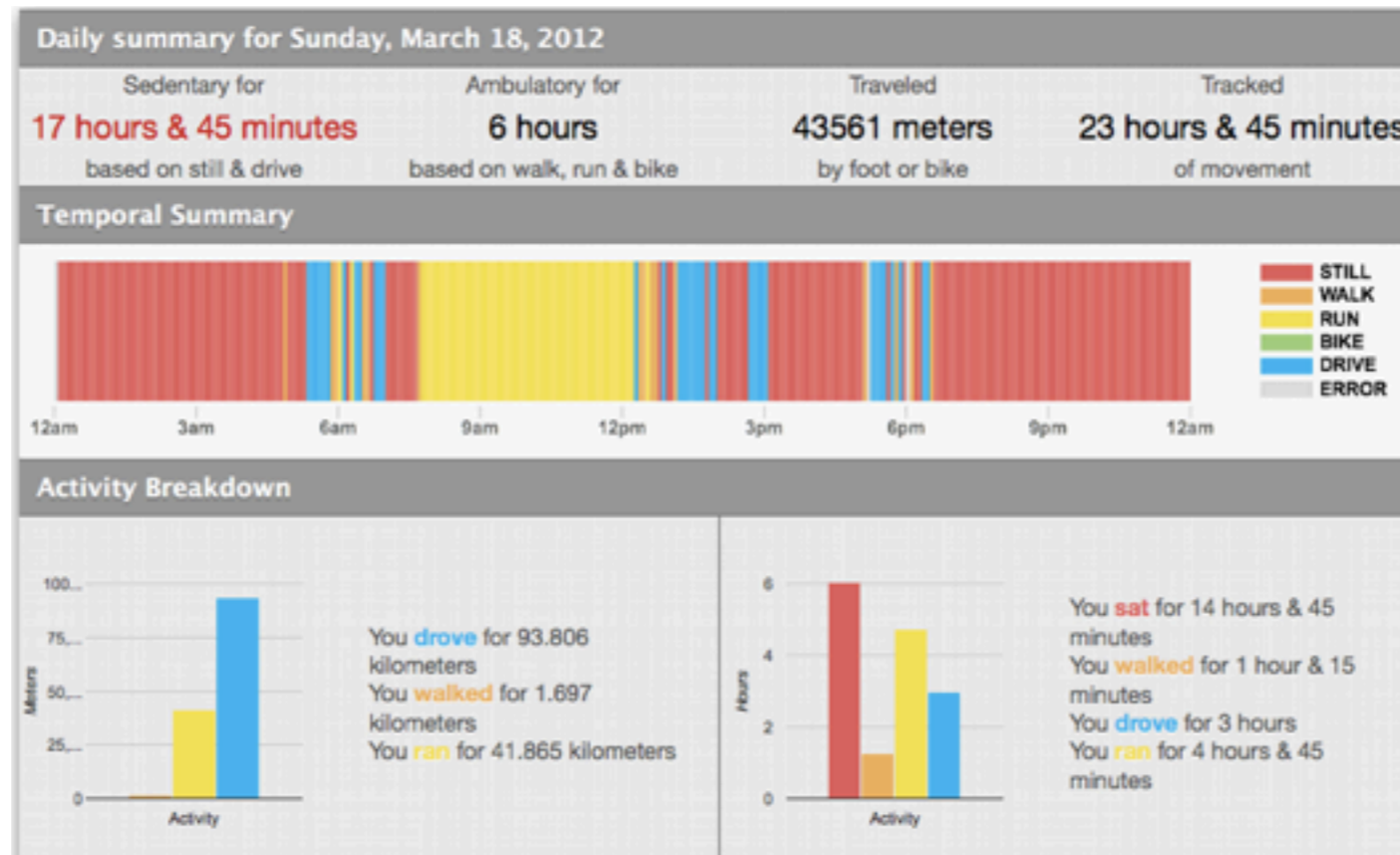
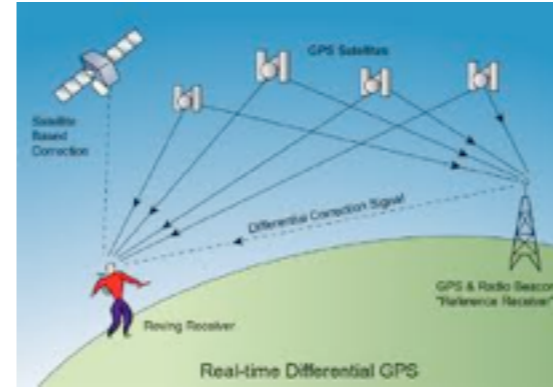
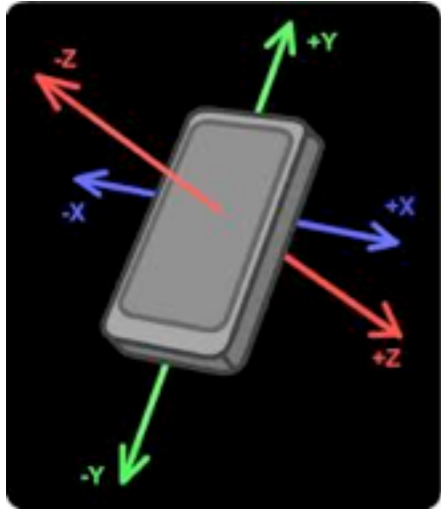
# “Real Sensor” streams



# mobile apps also generate relevant data

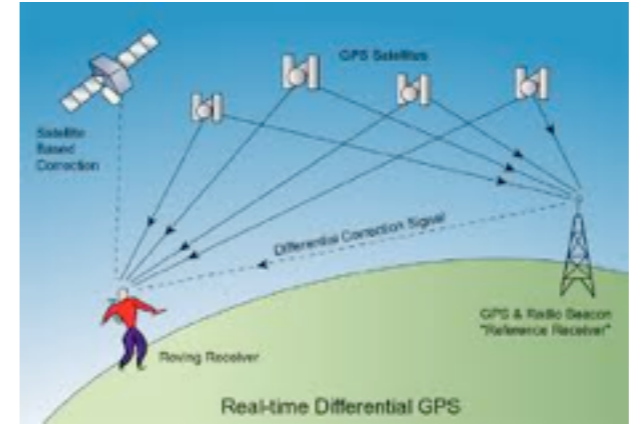
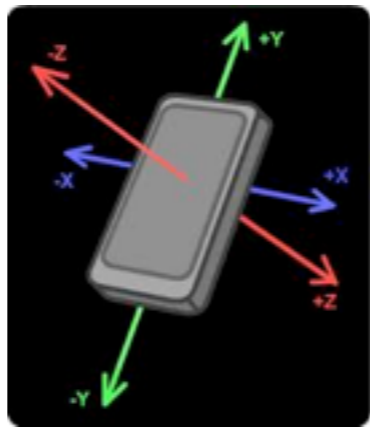


# Passive activity and location traces





# Communication and activity data: Ginger.io check engine light



Ginger.io

Anmol Madan (Logout)  
Help

Back to All Studies

## MDD & Diabetes

Overview Participants Analytics Insights Settings Invite Participants

Last Name	Number	Status	Phone	Surveys	Contact	GBI
Battle, Ginger	W1ZPSXT	GOOD	✓	✓	📞 📄	
Bleekburn, Vilhelmina	ORTBBHZ	BAD	✓	✓	📞 📄	
Dobos, Lore	ZRBK1FG	GOOD	✓	✓	📞 📄	
Evoff, Er	OS6WXP	GOOD	✓	✓	📞 📄	
Gartner, Angeliki	0BZ1SRW	GOOD	✓	✓	📞 📄	
Gentle, Glead	3BFXTW	GOOD	✓	✓	📞 📄	
Gulyás, Hedvika	LGPNZL	BAD	✓	✓	📞 📄	
Hoffman, Philippe	T4A7BMS	BAD	✓	✓	📞 📄	
Makar, Cenek	NB83XOU	GOOD	✓	✓	📞 📄	
Mulyan, Dominicus	PKW7WP8	GOOD	✓	✓	📞 📄	
O'Donnell, Charlotte	ZCZYULE	GOOD	✓	✓	📞 📄	
Overton, Theodore	ZYC8J0Z	GOOD	✓	✓	📞 📄	
Righi, Monique	C5SR77X	GOOD	✓	✓	📞 📄	
Roy, Weland	FDRXZV	BAD	✓	✓	📞 📄	

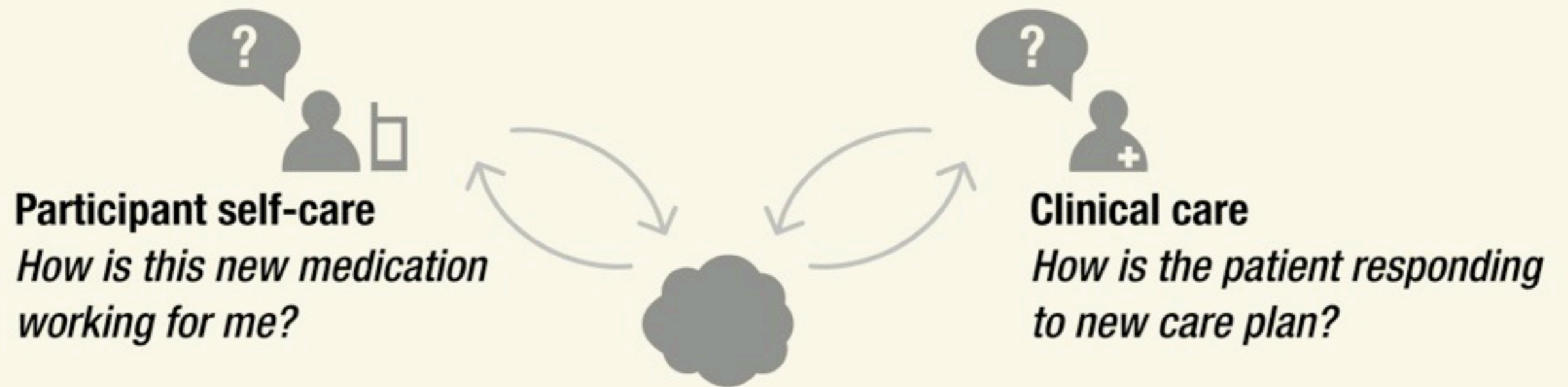
<http://ginger.io/the-platform/>

# Small data: digital traces from diverse internet services

- internet search
- mobile carriers
- social media
- online commerce
- maps
- cable TV
- utilities (elec, water)
- games, entertainment



# *What can digital behavior tell about personal health?*



**n=me small data**

# small data: Key challenges

# small data: Key challenges

- Getting the data
  - Personal data APIs (data liberation to the consumer!)
  - **Convincing/incentivizing service providers**

# small data: Key challenges

- Getting the data
  - Personal data APIs (data liberation to the consumer!)
  - **Convincing/incentivizing service providers**
- Data processing, inference, fusion, modeling
  - diverse, noisy, lossy data
  - signal processing, machine learning, natural language...

# small data: Key challenges

- Getting the data
  - Personal data APIs (data liberation to the consumer!)
  - **Convincing/incentivizing service providers**
- Data processing, inference, fusion, modeling
  - diverse, noisy, lossy data
  - signal processing, machine learning, natural language...
- Data and API standards
  - app model and economy

# small data: Key challenges

- Getting the data
  - Personal data APIs (data liberation to the consumer!)
  - **Convincing/incentivizing service providers**
- **Data processing, inference, fusion, modeling**
  - **diverse, noisy, lossy data**
  - **signal processing, machine learning, natural language...**
- Data and API standards
  - app model and economy
- Personal data vaults
  - Security models and mechanisms, usability
  - Policy questions re. ownership, access, rights

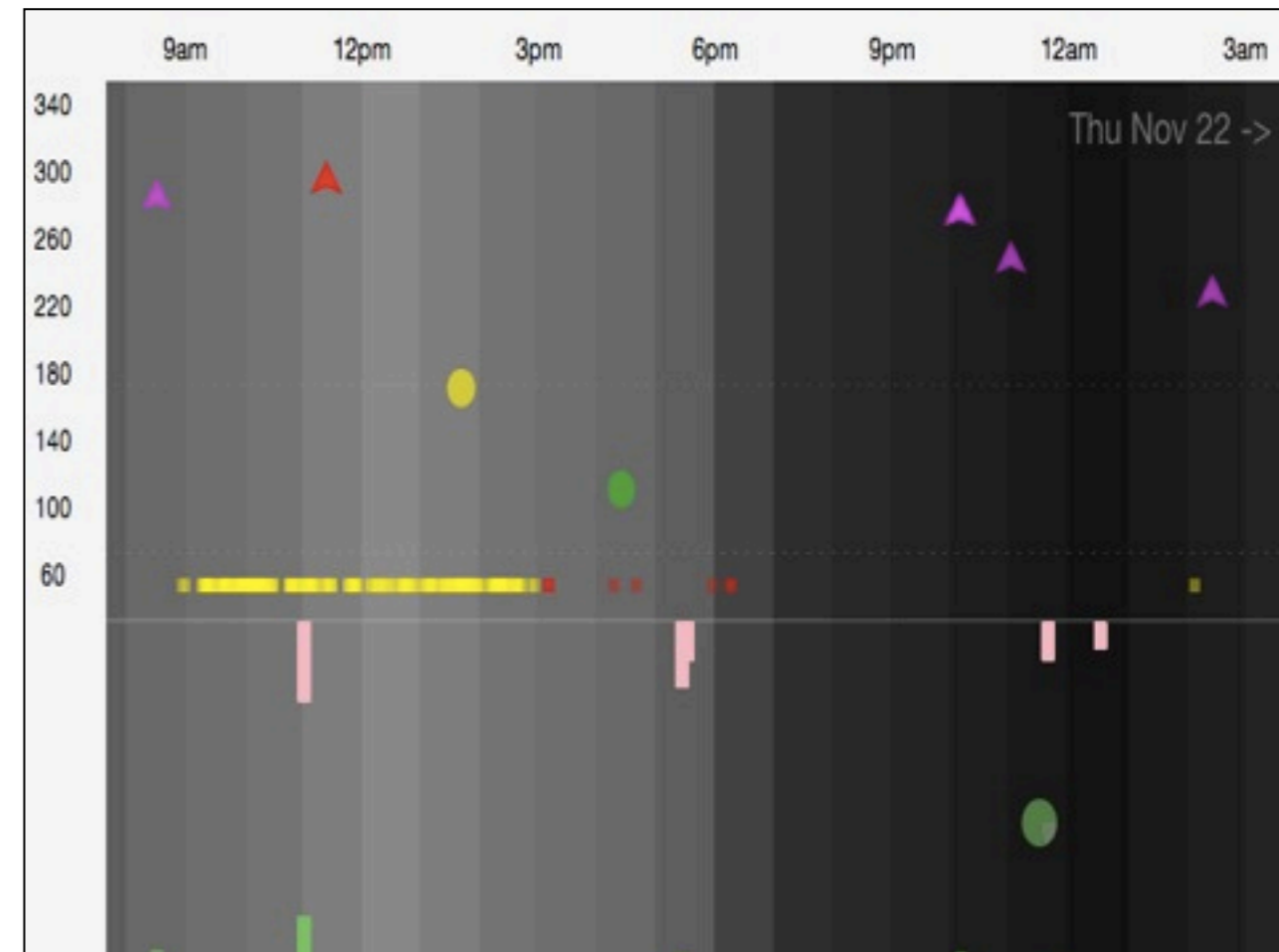


# Key challenge: making data clinically and personally actionable

## Clinician dashboard: PTSD



## Clinician dashboard: T1Diabetes



<http://greendotdiabetes.com>

# Deriving behavioral biomarkers ...from sensor, app, and small data streams

**Behavioral biomarker:**

*specific behavioral trait used  
to measure or indicate the  
effects or progress of a  
disease, illness, or condition*

# Deriving behavioral biomarkers ...from sensor, app, and small data streams

**Behavioral biomarker:**

*specific behavioral trait used to measure or indicate the effects or progress of a disease, illness, or condition*

**state classification**

- sedentary/ambulatory
- at home/work
- using apps
- communicating

# Deriving behavioral biomarkers ...from sensor, app, and small data streams

**Behavioral biomarker:**

*specific behavioral trait used to measure or indicate the effects or progress of a disease, illness, or condition*

## **summarization**

- ambulatory/sedentary cumulative and durations, walking speed
- sleep times, social interactions
- time spent key locations
- diameter of day



## **state classification**

- sedentary/ambulatory
- at home/work
- using apps
- communicating

# Deriving behavioral biomarkers ...from sensor, app, and small data streams

**Behavioral biomarker:**

*specific behavioral trait used to measure or indicate the effects or progress of a disease, illness, or condition*

## behavioral biomarker

- in-person variance/patterns of fatigue, pain, depression, insomnia, cognitive function...

## summarization

- ambulatory/sedentary cumulative and durations, walking speed
- sleep times, social interactions
- time spent key locations
- diameter of day

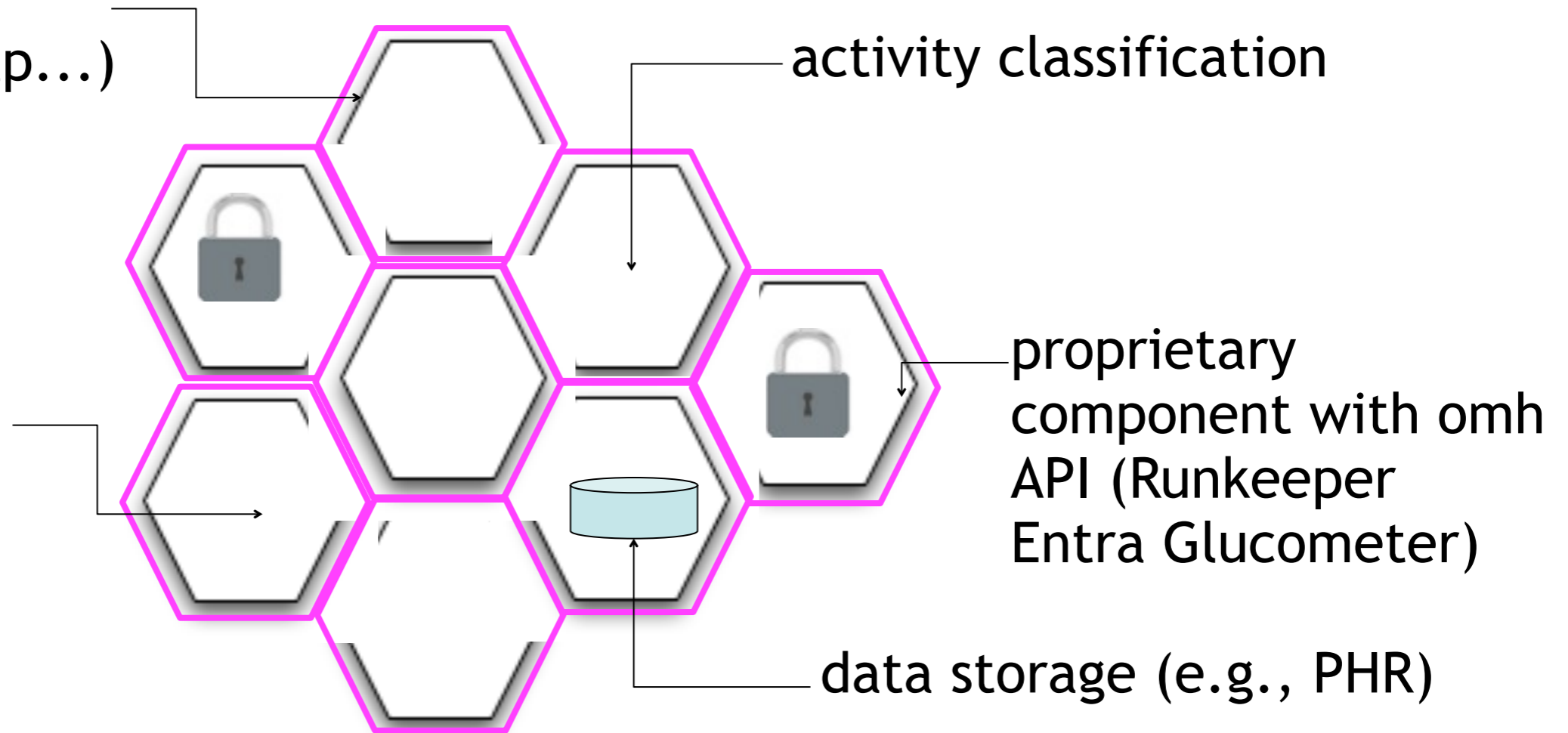
## state classification

- sedentary/ambulatory
- at home/work
- using apps
- communicating

# open architecture to promote modularity, data interoperability and software reusability

mood App  
(PAM, moodmap...)

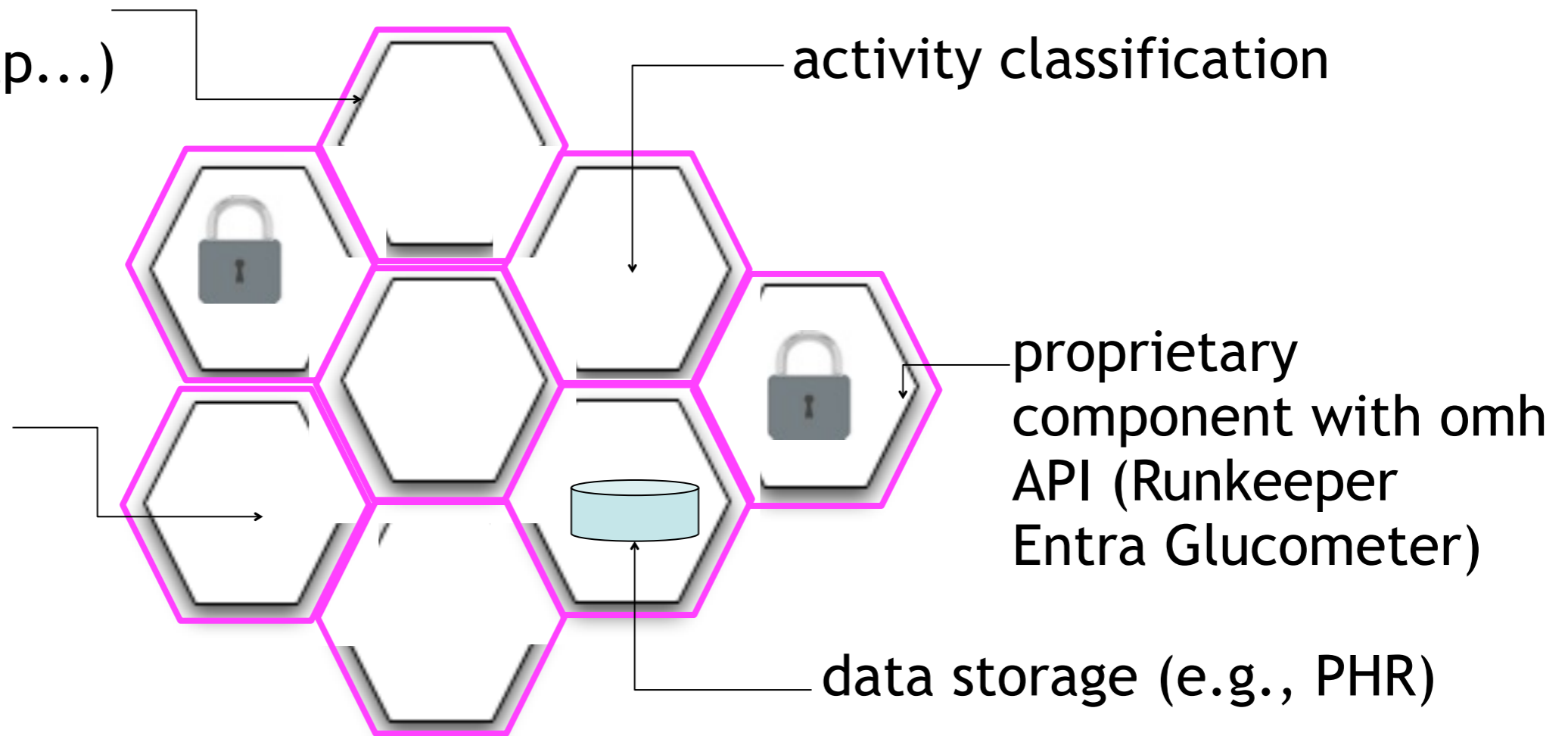
graphing  
significant  
changes in  
mobility



# open architecture to promote modularity, data interoperability and software reusability

mood App  
(PAM, moodmap...)

graphing  
significant  
changes in  
mobility



Open mHealth APIs enable independently developed apps and tools to generate and access data in a uniform way

Promote integrative mHealth solutions





By table, 5-10 minute task

Describe a small data source and derivable biomarker, that would be relevant to your SCH projects.

# small data

(as presented at [TEDMED](#))

**CORNELL**  
**NYCTECH**

## please package up my small data for me!

By entering your email address below, we will together send a message to Internet, Social Media, and Mobile service providers to package our small data for us. You will be the customer for the data about you; I will be the customer for the data about me.

Your Email Address:

Notify Me

- \* Your contact information will not be shared with any other parties.
- \* We will share only the aggregate number of people who have registered interest in getting access to their small data.
- \* We will use your email only to contact you regarding updates to our small data initiative.



<http://smalldata.tech.cornell.edu/>