Small, n=me, data

Deborah Estrin
Professor, Computer Science, Cornell NYC Tech
Professor, Healthcare Policy and Research, Weill Cornell Medical College
Co-founder, Open mHealth

destrin@cs.cornell.edu

http://smalldata.tech.cornell.edu
Small data are....
the data an individual generates implicitly,
across a myriad of systems, activities and encounters.
The role of mHealth and small data

data transform previously unmeasured behaviors and practices into personalized, evidence-based, and evidence-producing care

behaviors, activities, symptoms, side-effects, outcome measures, exposures
activity, location, communication, sleep, app usage, purchases; self-reports/EMAs

visualize, summarize, highlight; inform, advise, persuade
store, analyze, classify, fuse, mashup, filter, aggregate data

Photo: Marshall Astor, WWW
How can mHealth and small data improve health and healthcare?

Participant self-care
How is this new medication working for me?

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

mhealth data
How can mHealth and small data improve health and healthcare?

Participant self-care
How is this new medication working for me?

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

mhealth data
Passively-recorded activity, location
How can mHealth and small data improve health and healthcare?

mhealth data

Passively-recorded activity, location

Mobile apps: data, metadata, usage
How can mHealth and small data improve health and healthcare?

Participant self-care
How is this new medication working for me?

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

mhealth data

Passively-recorded activity, location

Mobile apps: data, metadata, usage

“Real Sensor” streams too
Behavioral biomarkers
specific behavioral traits to measure progress of disease and treatment

state classification
• sedentary/ambulatory
• at home/work
• app analytics (games, media...)
• communication
Behavioral biomarkers
specific behavioral traits to measure progress of disease and treatment

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

**state classification**
- sedentary/ambulatory
- at home/work
- app analytics (games, media...)
- communication

Today
17.4 h 55 min 0 min 0 min
Behavioral biomarkers
specific behavioral traits to measure progress of disease and treatment

**behavioral biomarker**
- individual’s patterns; relevance is symptom and condition dependent
- ‘function, fatigue, pain, depression, insomnia, cognition, self-medication...

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

**state classification**
- sedentary/ambulatory
- at home/work
- app analytics (games, media...)
- communication

---

**Hours at home per day**

**Walking periods > 6 min per day**

**Today**
- 17.4 h
- 55 min
- 0 min
- 0 min

---

Sunday, May 11, 14
Example: The Ginger.io Platform

The Ginger.io Platform

Data Collection
- Location
- Self Reported
- Communication
- Sensors

Analytics Engine
- Socialization
- Activity
- Quality-of-Life

Patient Health Status

Provider Dashboard

Patient Smartphone App

Patient Support Network
Mining small data to fuel health(ier) behaviors

*Let's build small-data fueled apps, engagements, and games that measure, motivate, and mitigate behavior change*

Martha’s food choices, daily stretches, yoga, to help her reduce dependence on pain meds

Sean’s media consumption to help him rebalance his sleep habits
The TMI Challenges of small data

Transparency: Make data available to service subscribers (personal data APIs)

Sense-Making: Develop algorithms and methods to make small data actionable

Selective Sharing: Individual as nexus of control over data use and sharing.
Open mHealth (openmhealth.org)

no one data stream tells the story

data interoperability, e.g., from clouds

mobile integrated 'solutions/systems'
Acknowledgments

Collaborators

• **small data lab@Cornell Tech**: Faisal Alquaddoomi, Aaron Baum, Michael Carroll, Lucky Gunasekara, Andy Hsieh, John Jenkins, Cameron Ketcham, JP Pollak, Josh Selsky

• **Cornell CIS**: Tanzeem Choudhury, Geri Gay, Thorsten Joachims, Jon Kleinberg, James Landay, Phoebe Sengers, Eva Tardos

• **Open mHealth**: David Haddad, Anna de Paula Hanika, Ida Sim

• **Hospital for Special Surgery**: Vivian Bykerk, Alana Levine, Stephen Lyman, Jane Salmon

• **NYGenomeCenter**: Toby Bloom, Robert Darnell, Dana Green

• **Weill Cornell Medical College**: Jessica Ancker, Mary Charlson, Curtis Cole, Neel Mehta, Erica Phillips, Cary Reid, Phillip Say, Daniel Stein

• **VA**: Julia Hoffman.

Sponsors/Partners

• Cornell: Cornell Tech, WCMC CHiPs

• Federal funding: NSF (STC and CISE SCH), NIH/NHLBI

• Commercial: Google, Intel, IBM, United Health Group, Verizon, Time Warner Cable

• Foundations/NGOs: RWJF, Verizon Foundation, MacArthur Open Governance Network

• UCLA: CENS, Health Sciences

Earlier contributors and collaborators

• **UCLA**: Jeff Burke, Scott Comulada, Mark Hansen, Nithya Ramanathan, Hongsuda Tangmunarunkt, Mary Jane Rotheram-Borus, Dallas Swendeman; Many PhD and MS students (CS and Stat)