Small, n=me, data and behavior change

Deborah Estrin
Professor, Computer Science, Cornell NYC Tech
Professor, Public Health, Weill Cornell Medical College
Co-founder, Open mHealth

destrin@cs.cornell.edu
work done with collaborators from
Cornell, UCLA, openmhealth.org, ...

CORNELL NYCTECH Open mHealth Weill Cornell Medical College

Small, n=me, data

The data we generate implicitly and explicitly across a myriad of systems, activities and encounters:

from mobile, cable TV, purchases, browsing, apps, entertainment, games, email, posts, texts and tweets....

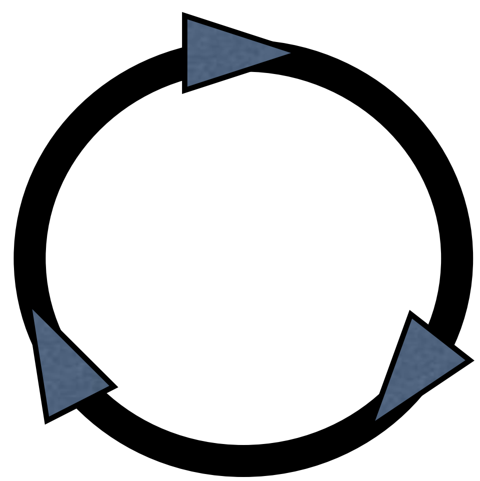
The role of mHealth and small data

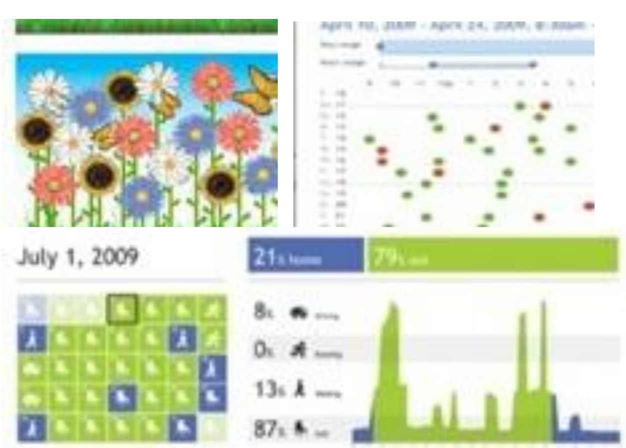
transform previously unmeasured behaviors and practices into personalized, evidence-based, and evidence-producing care

Photo: Marshall Astor, WWW

behaviors, activities, symptoms, side-effects, outcome measures, exposures

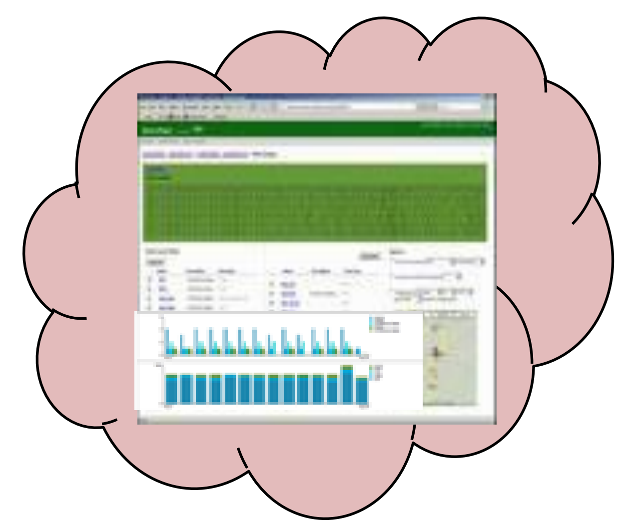
activity, location, communication, sleep, app usage, purchases; self-reports/EMAs



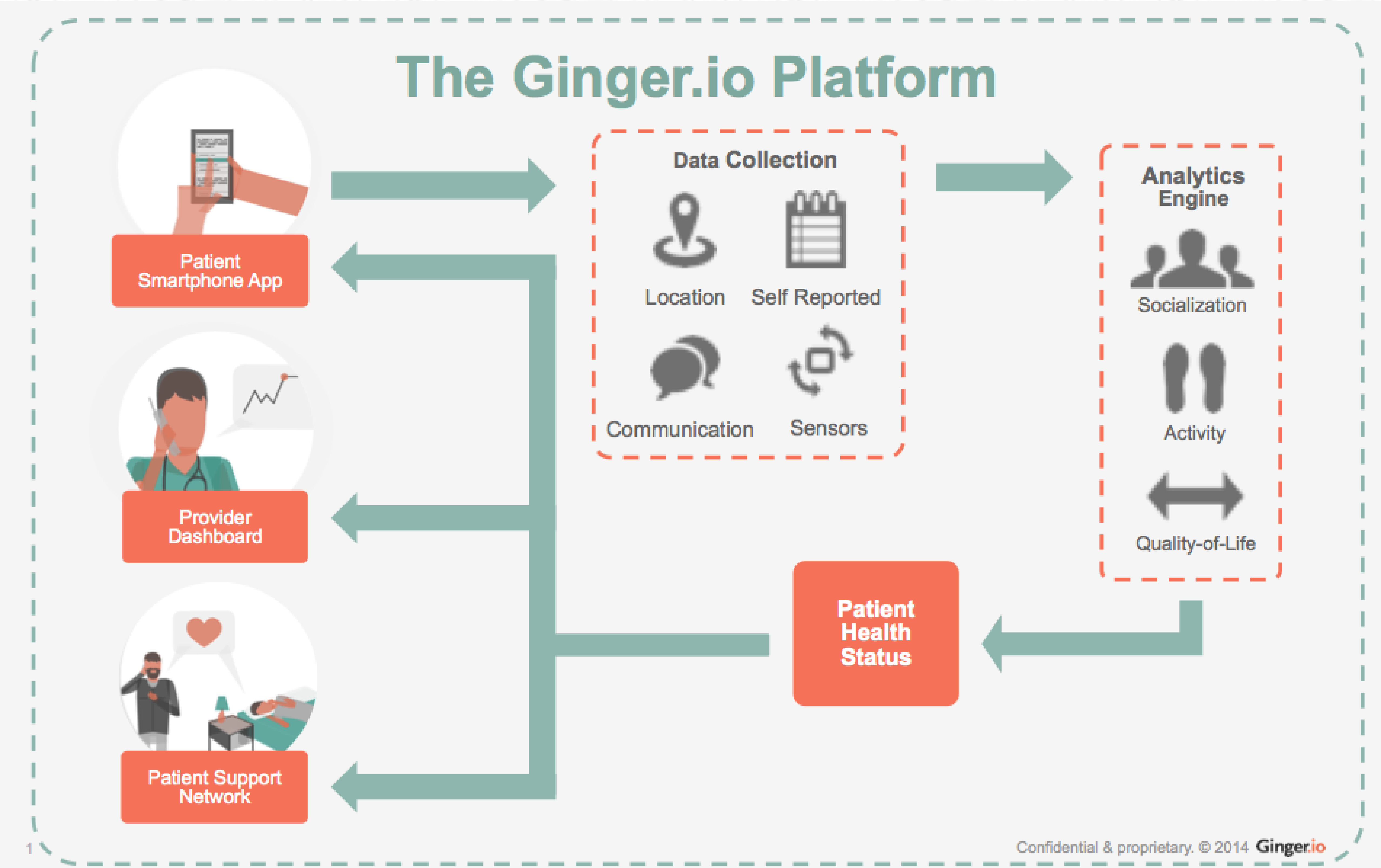


visualize, summarize, highlight; inform, advise, persuade

store, analyze, classify, fuse, mashup, filter, aggregate data



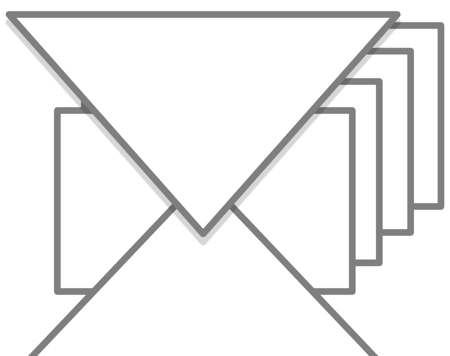
Rich communication and activity data: The Ginger.io Platform



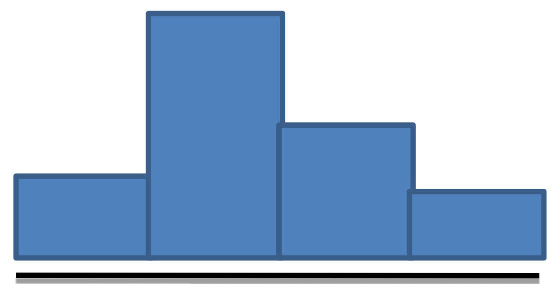
Beyond mobile...

Leveraging digital traces from diverse consumer services

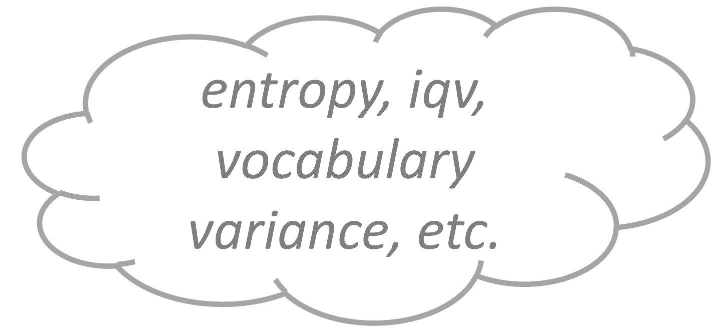
Example: communication language patterns as small data



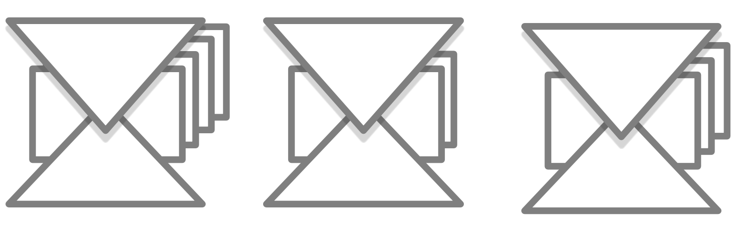
The goal is to transform text communication…



…into a "bag of words" model (a categorical distribution)…

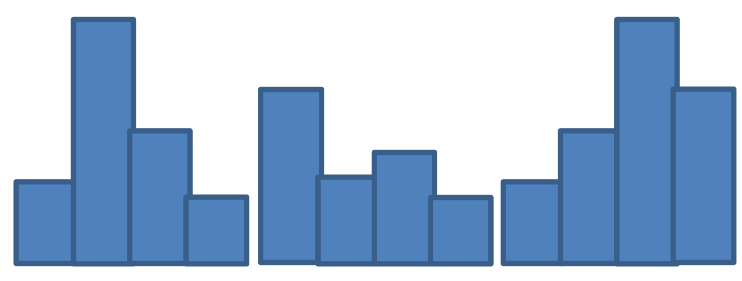


…on which descriptive statistics can be computed.



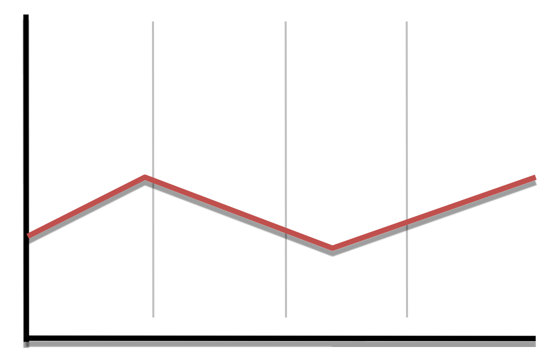
Sept Oct Nov

Since we have a progression of emails over time…



Sept Oct Nov

…we can examine how the resulting distributions change…

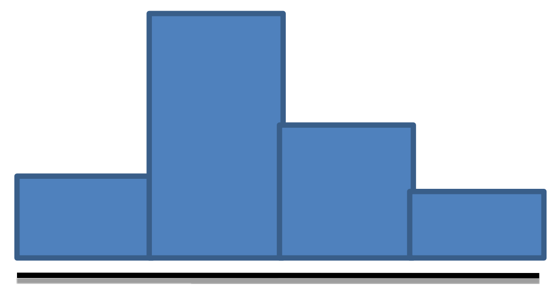


…and plot the statistics for each time frame, resulting in a time series which can then be correlated against other time series.

F. Alquaddoomi

Example: consumer transaction patterns as small data

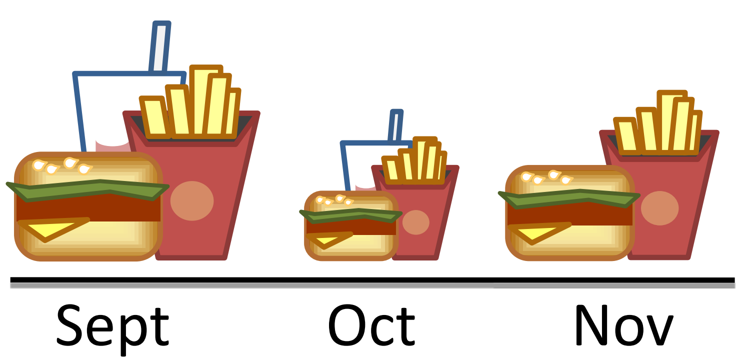
The goal is to transform purchasing patterns…



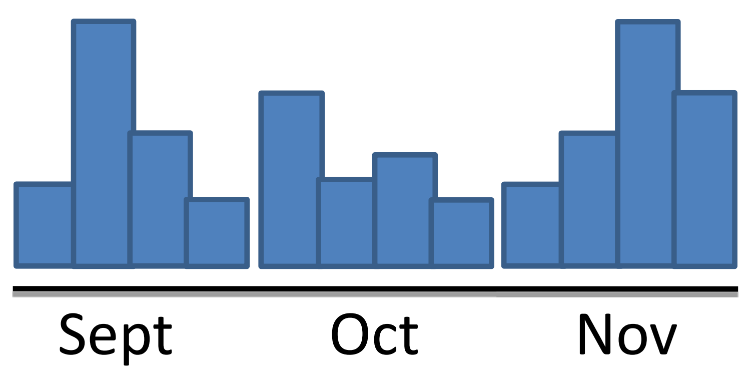
…into a consumption model (a categorical distribution of restaurant, fast food, drug store, etc. purchases)…

fast food, processed food, carb + carbon count…

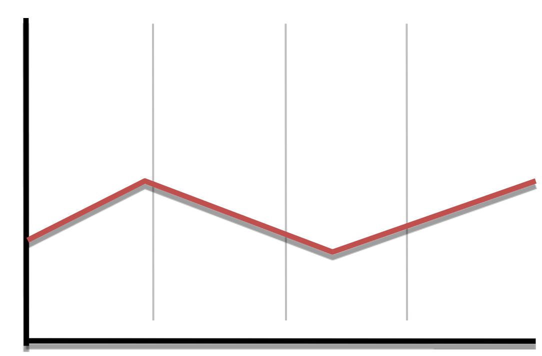
…on which descriptive statistics can be computed.



Since we have a progression of spending patterns over time…



…we can examine how the resulting distributions change…



…and plot the statistics for each time frame, resulting in a time series which can then be correlated against other time series.

Aspiration: small data meets behavioral science

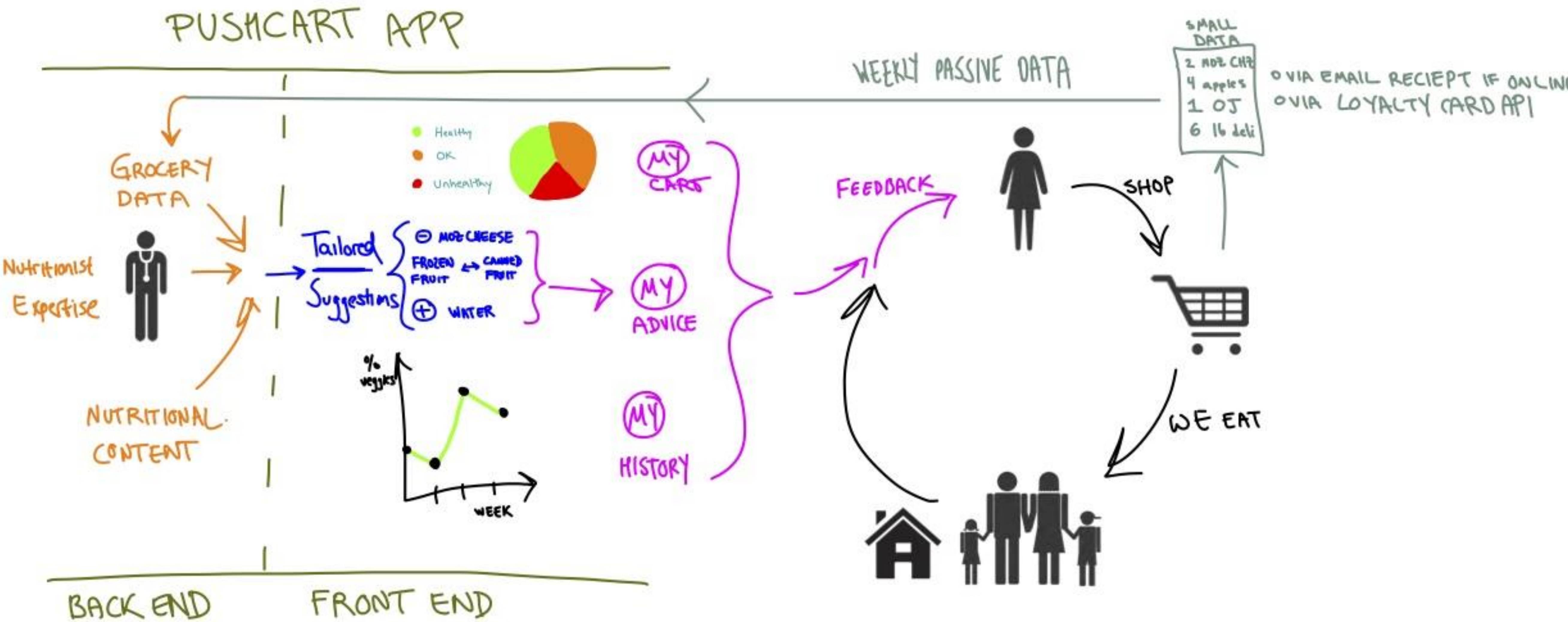
- *Create a new body of **small-data techniques to fuel applications** that measure, motivate, and provide opportunities for people to improve their individual and family lives.*

- *Study and support behavior change by helping individuals with information **recall and processing** challenges that interfere with significant and sustained change (recency, saliency, recall, delayed discounting).*

  - Martha in her self-care (weight management, daily back stretches, yoga with meditation) to help her reduce dependence on pain meds
  - The Smiths in their family food choices
  - Sean in his media consumption and sleep habits

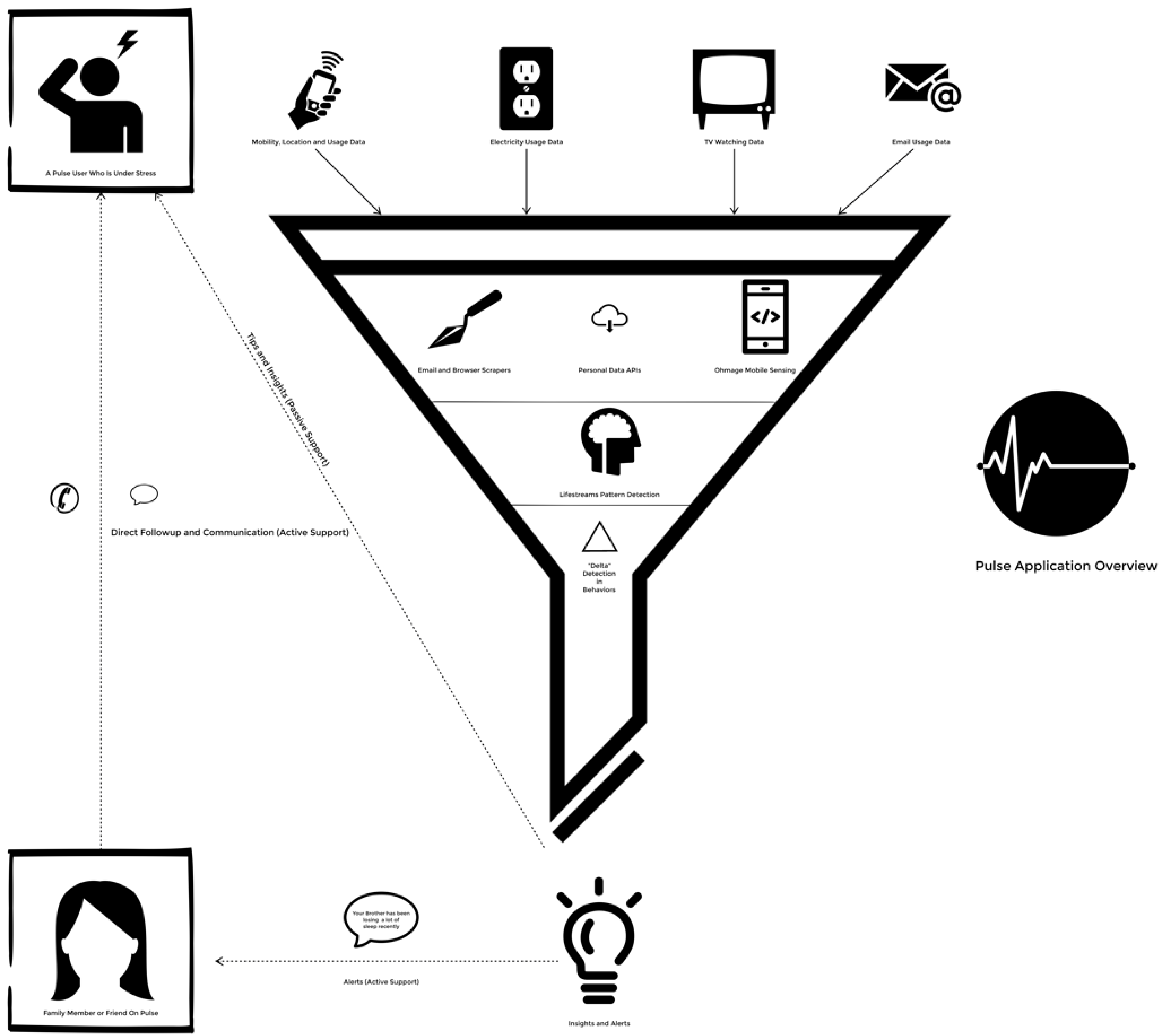
- ***Make small data available, accessible, and actionable for individuals***

Example: personalized advice and feedback on weekly grocery shopping to promote healthier home food environment



A. Baum

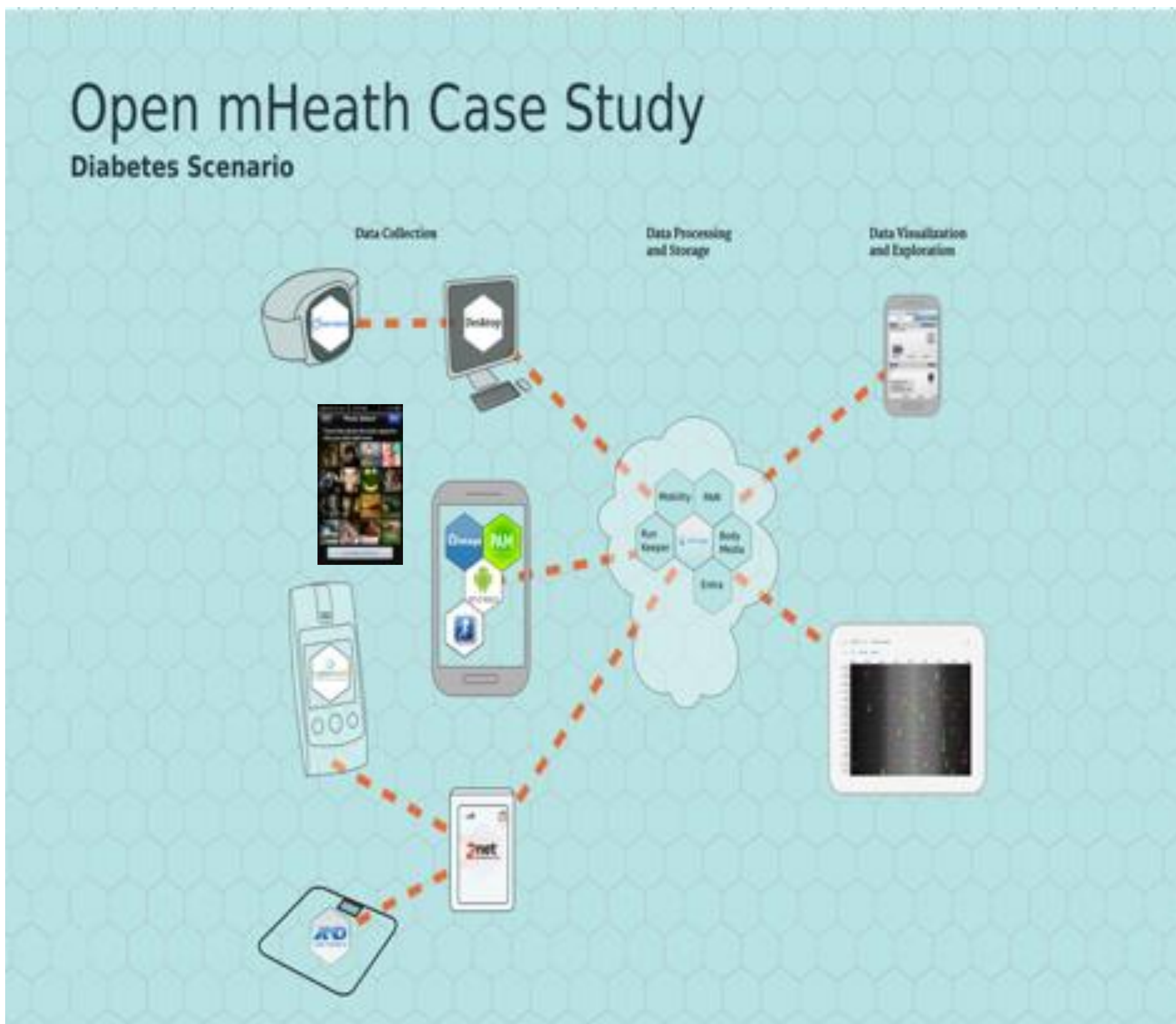
Example: Pulse

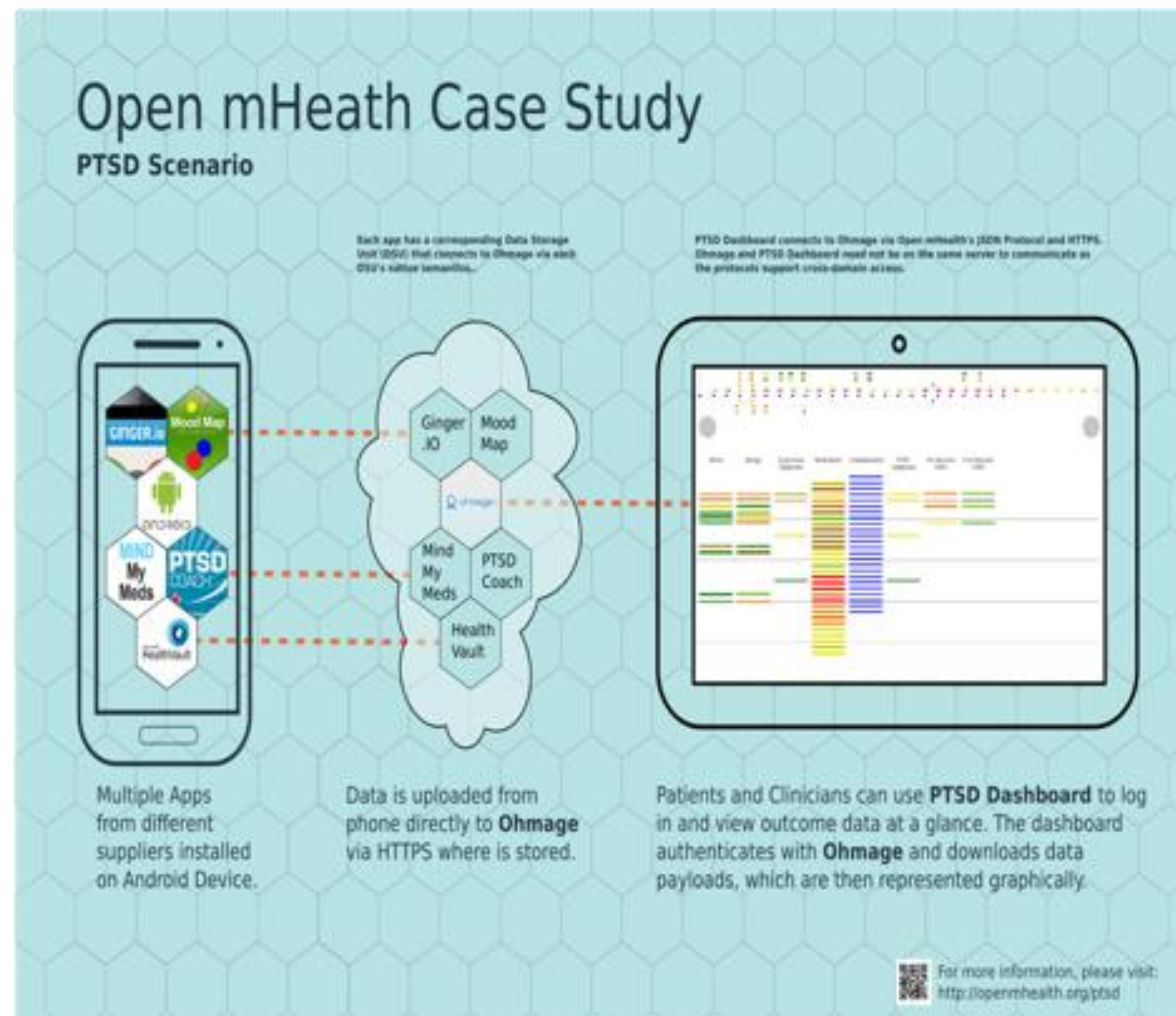


L. Gunasekara

Open mhealth (openmhealth.org)

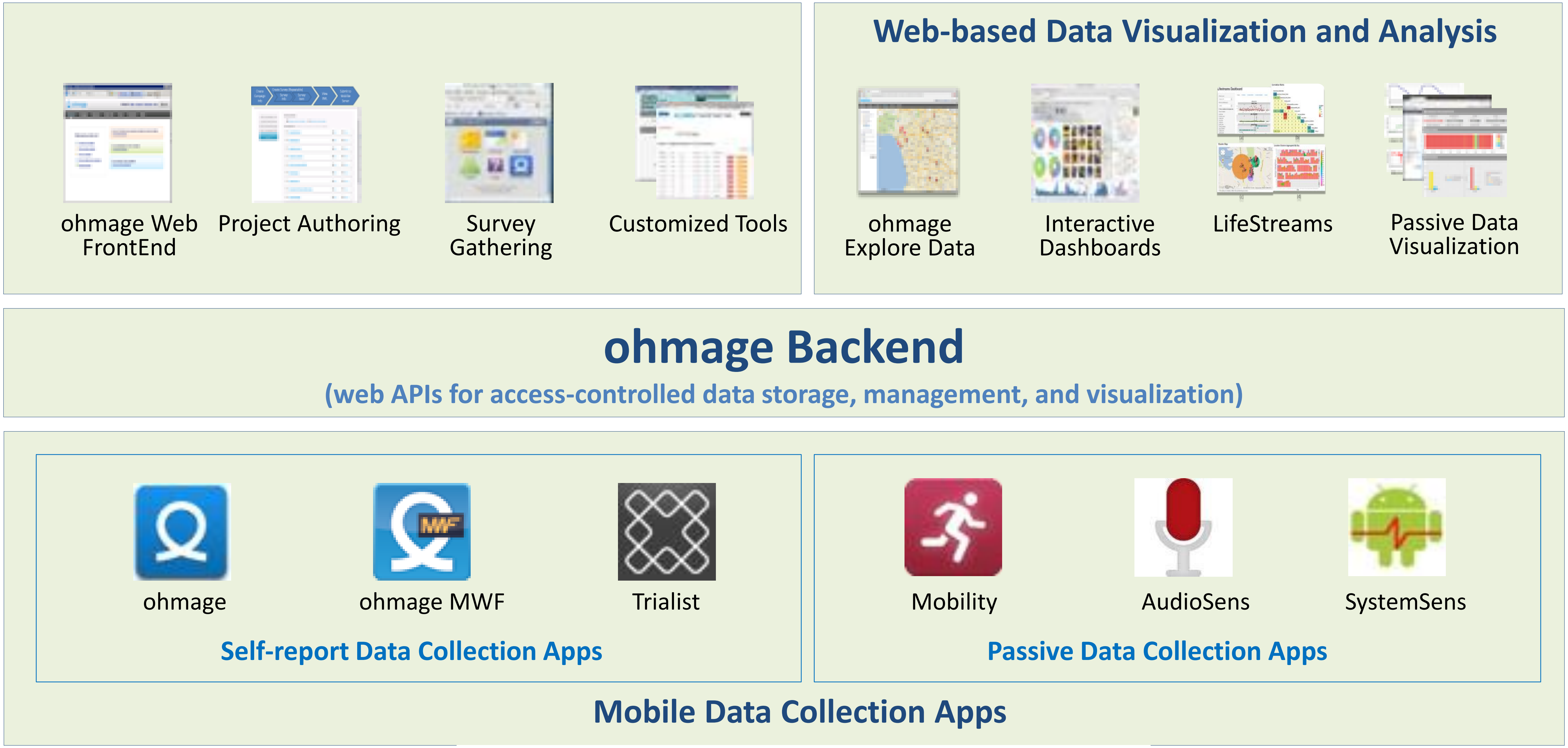
no one data stream tells the story...
its about modularity, integration, fusion, and sense-making





Open mHealth

ohmage open source platform

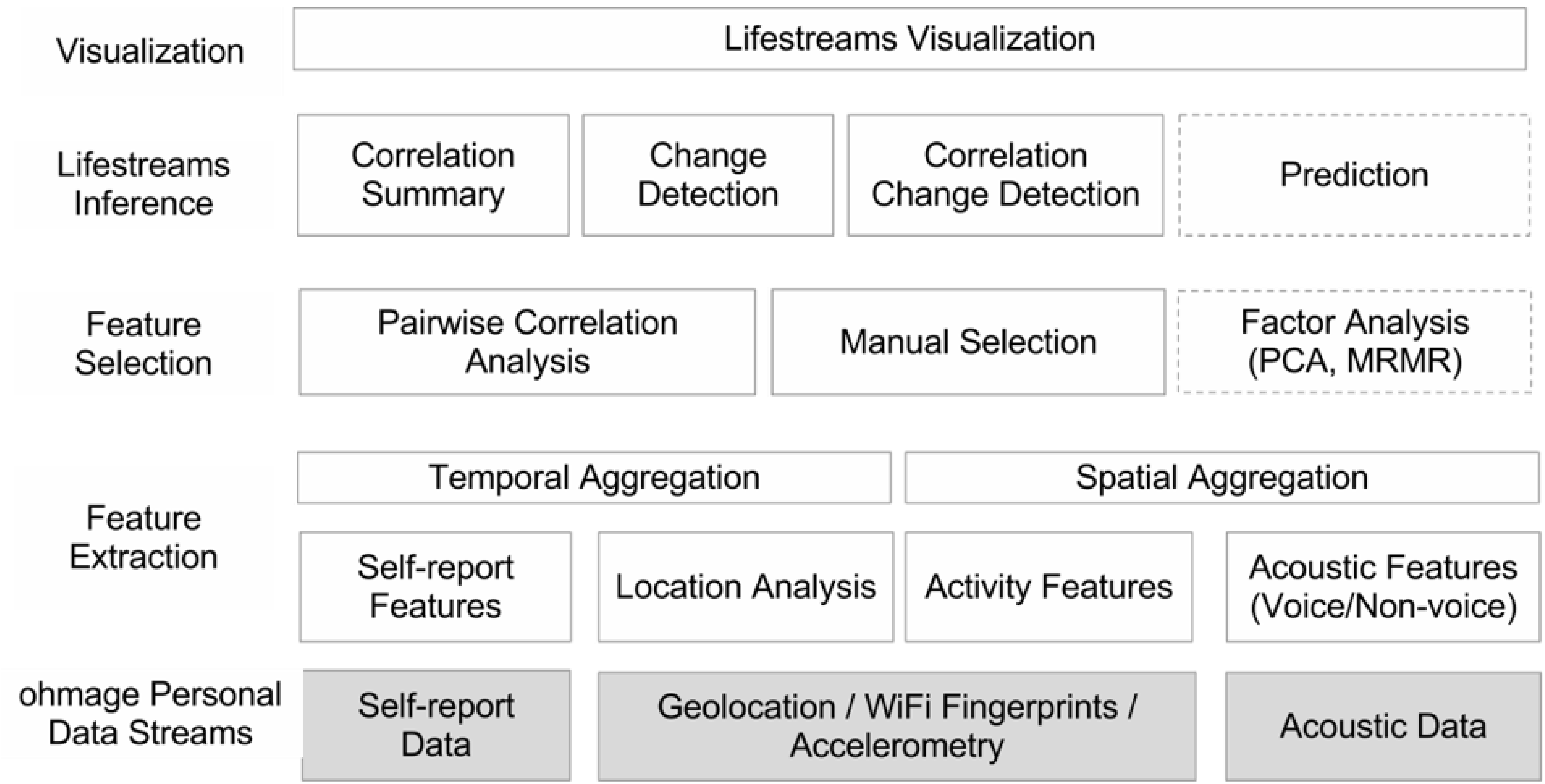


App Name	Platform	Download	Docs *
ohmage client	Android 2.2+	Play Store	
ohmage MWF Android client **	Android 2.2+	Play Store	
ohmage MWF iOS client **	iPad/iPhone	App Store	
Mobility client	Android 2.2+	Play Store	

* = tutorial, = written guide, = frequently asked questions

** the MWF versions currently only support self-report, whereas the ohmage Android client is full-featured, supporting collection of continuous background data streams as well as self-report

Modular tools to identify, iterate, share building blocks for behavioral biomarkers, sensemaking



(*Lifestreams, Hsieh et al, Sensys 2013)*

Hsieh, Tangmunarunkit, Ramanathan, et al

The TMI Challenges of small data

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

Selective Sharing: Individual as nexus of control over data use and sharing.

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

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
- UCLA: CENS, Health Sciences

Earlier contributors and collaborators

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