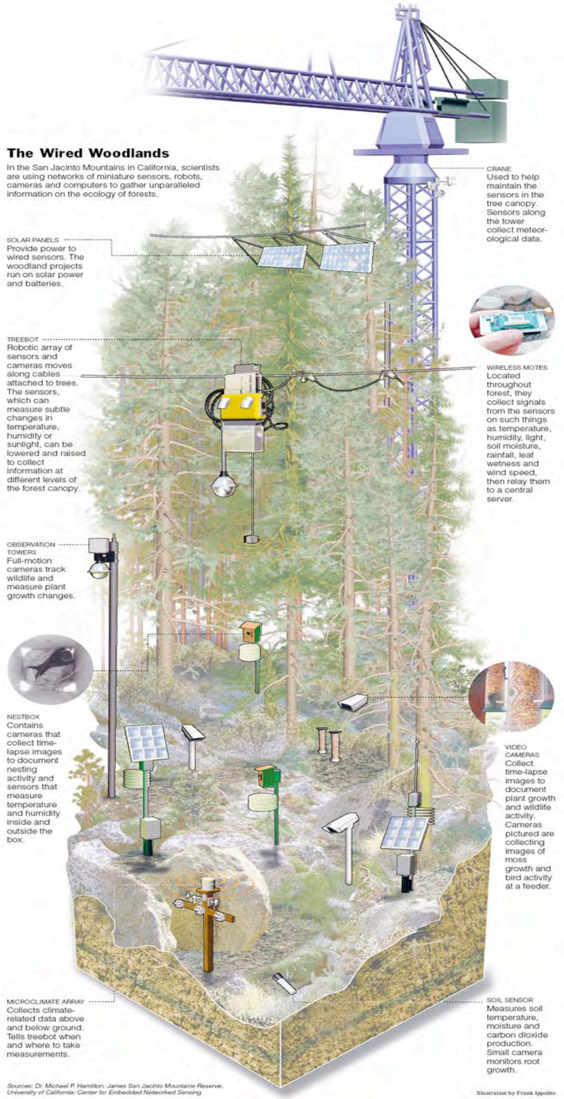


Mobile sensing systems: *From Ecosystems to Human Systems*

Professor Deborah Estrin,
UCLA Computer Science Department
destrin@cs.ucla.edu
in collaboration with:

Co-PIs: Jeff Burke (CENS/REMAP), Jeff Goldman (CENS), Ramesh Govindan (CENS/USC), Eric Graham (CENS), Mark Hansen (Statistics), Mary Jane Rotheram (Psychiatry/Semel), Mani Srivastava (EE/CENS), Ruth West (CENS)

Students and Staff: Betta Dawson, Hossein Falaki, Taimur Hassan, Donnie Kim, Olmo Maldonado, Min Mun, Nicolai Petersen, Nithya Ramanathan, Sasank Reddy, Jason Ryder, Vids Samanta, Katie Shilton, Nathan Yau, Eric Yuen

Work summarized here is that of students, staff, and faculty at CENS

We gratefully acknowledge the support of our sponsors, including the National Science Foundation, Nokia, Intel Corporation, Cisco Systems Inc., Sun Inc., Google, Microsoft Research, UC Micro, Crossbow Inc., Agilent, Conservation International, and the participating campuses.

YouTube Video on PEIR

Many critical issues facing science, government, and the public call for high fidelity and real time observations of the physical world

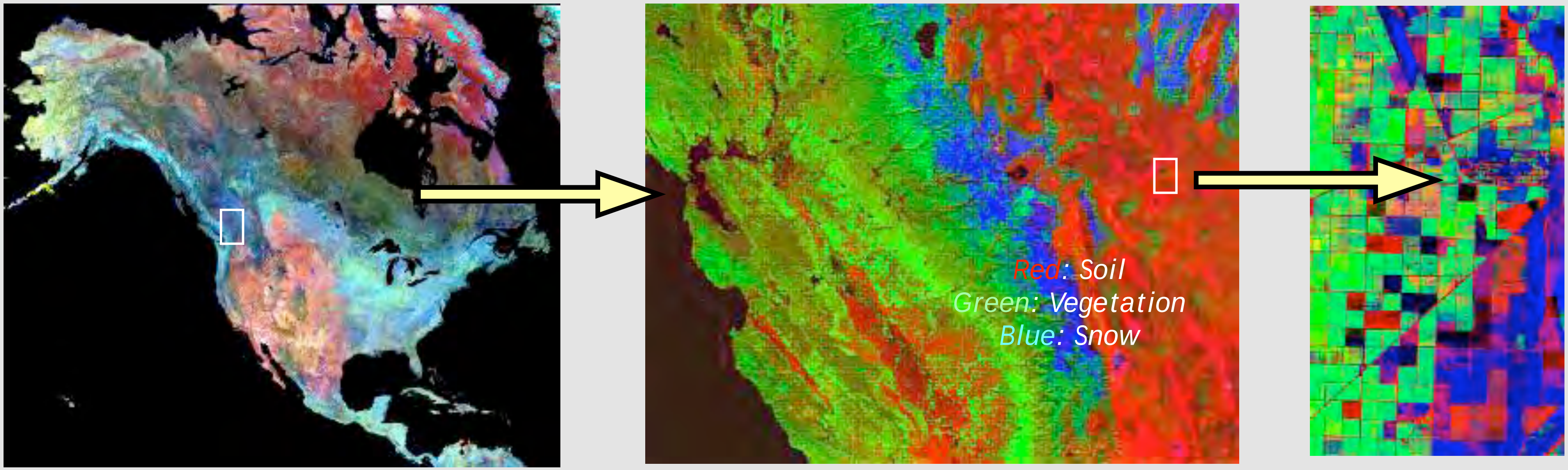
Embedded sensing systems:

reveal the previously unobservable

help us understand and manage interactions with physical world, scarce resources, and one another

Why *embedded* sensing?

- Remote sensing transformed observations of large scale phenomena
- Embedded (in situ) sensing transforms observations of spatially rich processes



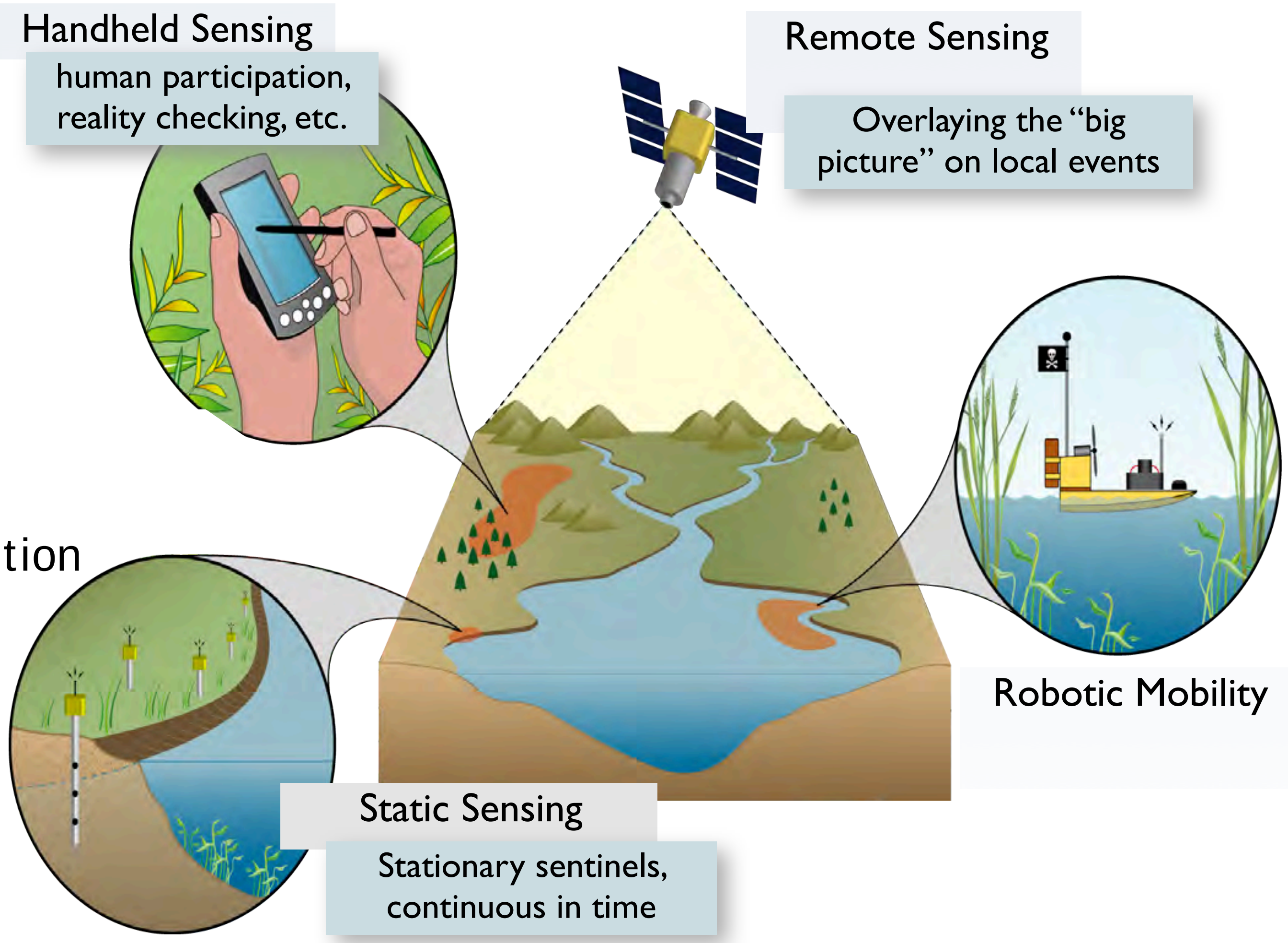
San Joaquin River Basin
Susan Ustin-Center for Spatial Technologies and Remote Sensing

Embedded networked sensing is revealing previously unobservable phenomena

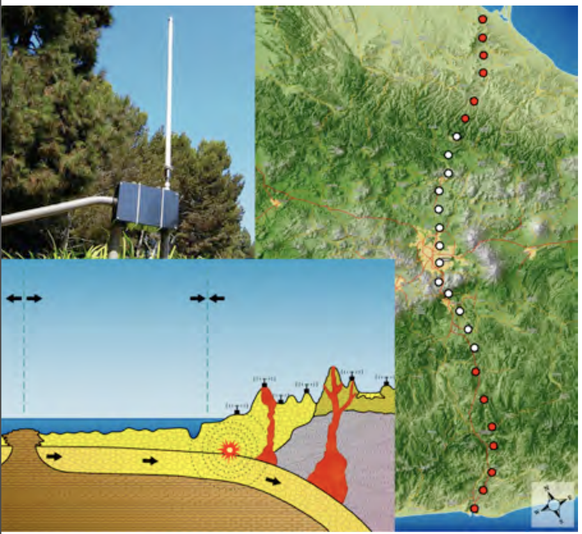
Embedded in the physical environment

Networked to share information/adapt function

Sensing physical world phenomena



Center-wide focus: embedded networked sensing

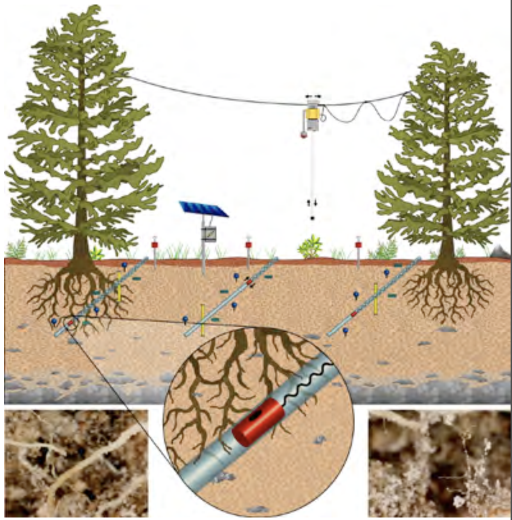


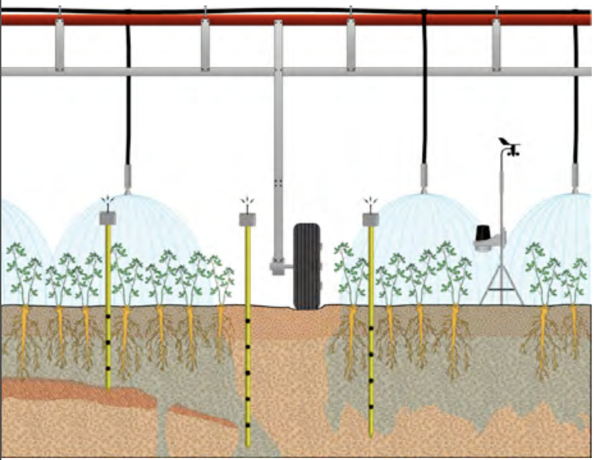
create programmable, distributed, multi-modal, multi-scale, multi-use observatories to address compelling science and engineering issues

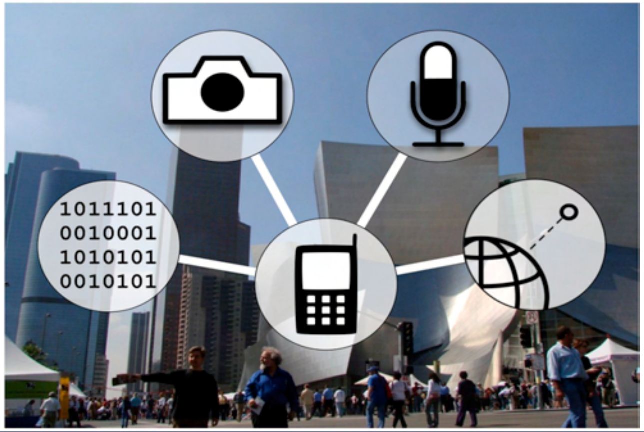
...and reveal the previously unobservable.

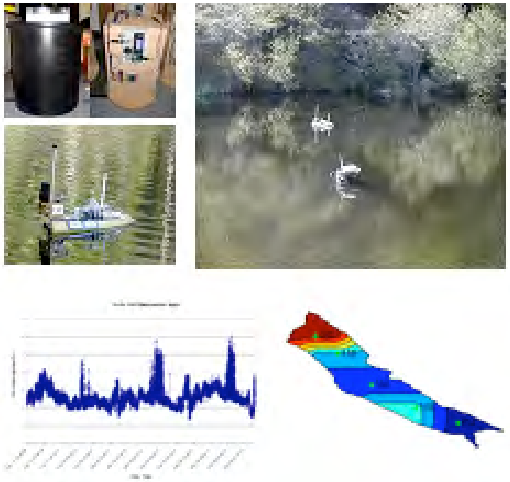
From the natural to the built environment...

From ecosystems to human systems...









Lessons from the field…

Early themes

- Thousands of small devices
  - Minimize individual node resource needs
  - Exploit large numbers
- Fully autonomous systems
  - In-network and collaborative processing for longevity: optimize communication

Current themes

- Systems of heterogeneous devices (capabilities, functions)
  - Combine in situ and server processing to optimize system
  - Mobility to overcome inevitable under-sampling with static sensing
  - Exploit multiple sensor types (e.g. imagers), multiple scales
- Humans and models *in the loop*
  - Coupled human-observational systems
  - Online observations achieved by combining direct measurements with server-side models, data, analysis
  - Participatory sensing leveraging mobile infrastructure

If you can't go to the field with the sensor you want, go with the sensor you have

present → *future*

abiotic ↓ *biotic*

Physical Sensors: Microclimate above and below ground

Chemical Sensors: gross concentrations

Chemical Sensors: trace concentrations

Acoustic and Image data samples

Acoustic, Image sensors with on board analysis

DNA analysis onboard embedded device

Organism tagging, tracking

Sensor triggered sample collection

- Commercially available autonomous devices available for physical and chemical measures only
- System designs need to compensate for lack of sensor specificity, sensitivity, availability...particularly wrt biological response variables
- Leverage proxy sensors and model based signal interpretation

Mobile Personal Sensing

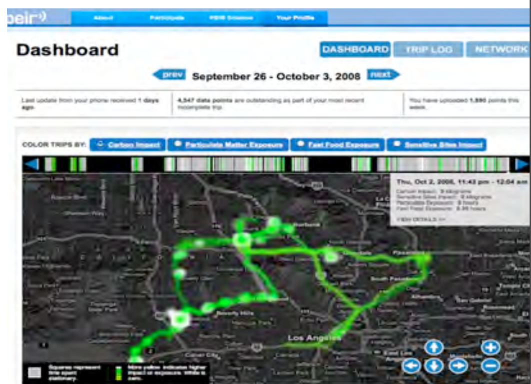
http://urban.cens.ucla.edu

Enabled by >3 x 10^9 mobile phone users, increasingly with…

- Digital imagers, location, bluetooth-connected sensors
- Automatic-geocoding of data
- Programmed, user, and server-initiated capture
- Server-side processing and presentation of personal data

Motivated by 6 x 10^9 people on planet earth and their concerns…

- Individual health and wellness
- Public health, urban planning, epidemiology
- Civic concerns (transportation, safety, culture…)
- Resource management



CENS CENTER FOR EMBEDDED NETWORKED SENSING
Burke, Estrin, Hansen, Ramanathan, Srivastava, West, et al

Participatory Sensing: Campaign Model
leveraging real-time, geo-coded, images

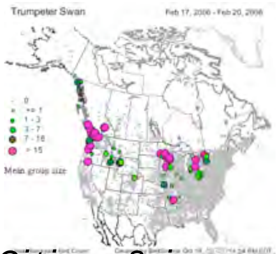
Distributed data gathering challenges as "Campaigns" -

- Spatially and temporally constrained systematic data collection operations.
- Exploring a single hypothesis, phenomena or theme.
- Using human-in-the loop sensing to gather data.
- With automatic and manual classification, auditing, and analysis.

Precedent - Community-Based Participatory Research

PhotoVoice
Caroline Wang, 1996

Citizen Science
World Water Quality Day



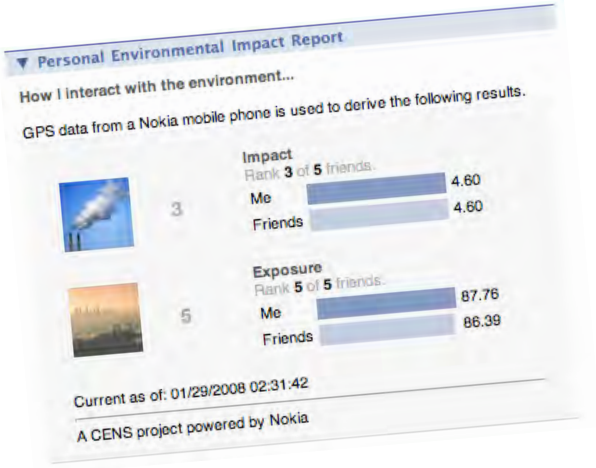
Citizen Science
Cornell e-Bird

Civic Participation
Video the Vote

Participatory GIS
Ctr for Neighborhood Knowledge

CENTER FOR EMBEDDED NETWORKED SENSING

Pilot Campaigns

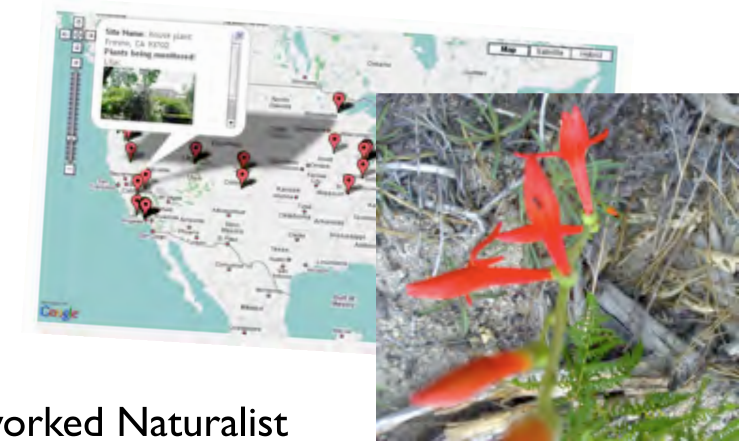


Personal Environment Impact Report

Campus Sustainability Initiatives

CycleSense

Networked Naturalist



DietSense as alternative to self-reporting
leveraging real-time, automated, personal, images

mobile phones

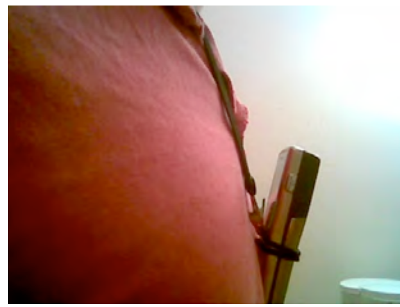
worn on a lanyard around the neck that automatically collect time-stamped images of food choices or purchases. Voice annotation, location stamping, and text message alerting will also be used.

participant data repository

receives annotated media collected by the devices, allows individuals private access to their own data before they are available to others, and supports filtering and alerting based on upload patterns and basic analysis of received data.

protocol management tools

enabling healthcare providers and researchers to easily author and automatically disseminate protocols for data collection to participants phones.



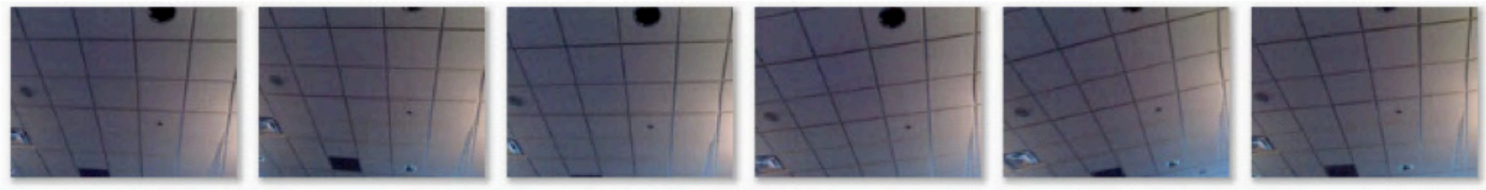
annotation, filtering, and analysis tools

available to both participants and researchers that provide efficient mechanisms to navigate, annotate, filter, and analyze the collected data, including the capability to export reports to common statistical software packages.

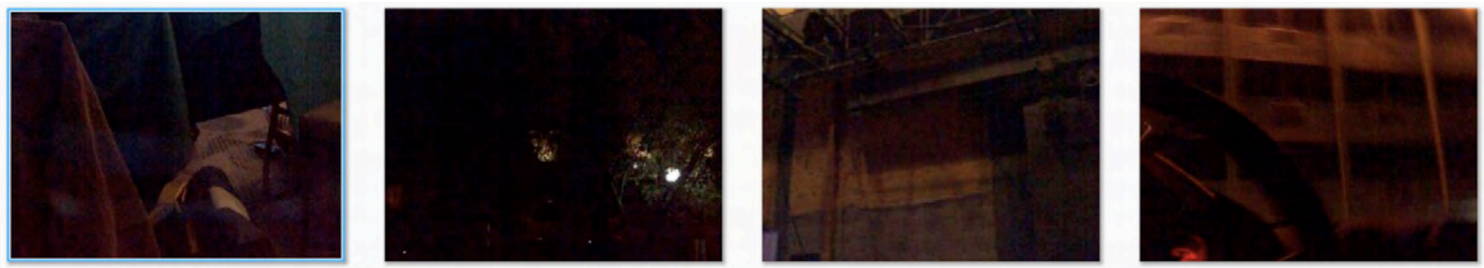
Kim, Kim, Petersen, Burke, Estrin, …
CENS CENTER FOR EMBEDDED NETWORKED SENSING

Identify likely useless images

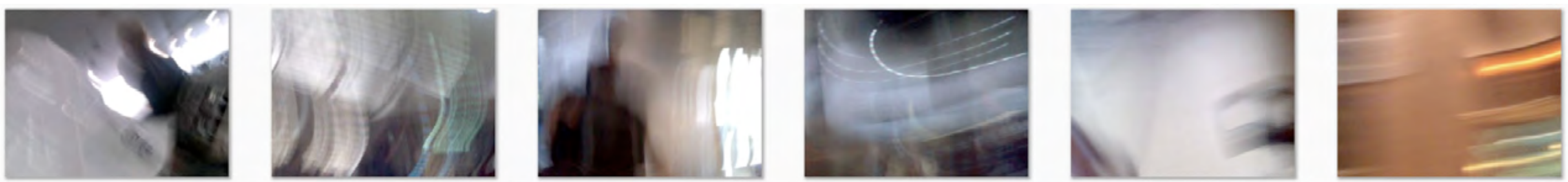
Redundant



Too Dark

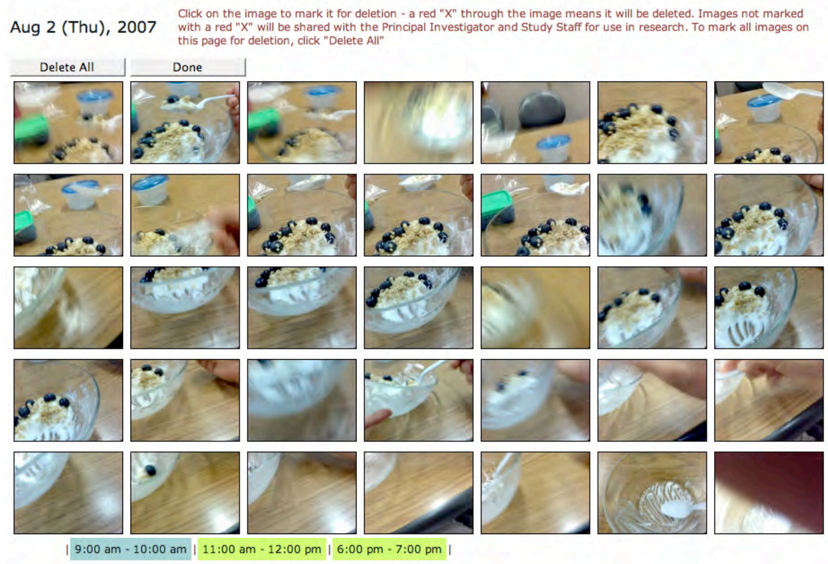


Too Blurry



Energetics Pilot - Continuous Image Capture for Dietary Recall (supplemented NIH funded study of recall)

- Objective - Improve the accuracy of 24-hour diet survey
- Pilot experiment in collaboration with Dr. Lenore Arab and DietDay(TM) application
- Observation day
  - Wear Nokia n80 during meal times
  - Take bio marker
- Recall day
  - Provided blood and urine sample
  - Document diet via 24-hour diet survey using diet day and energetics application to aid recall
- Analysis - Compare inferred dietary intake from survey with analysis of blood and urine sample

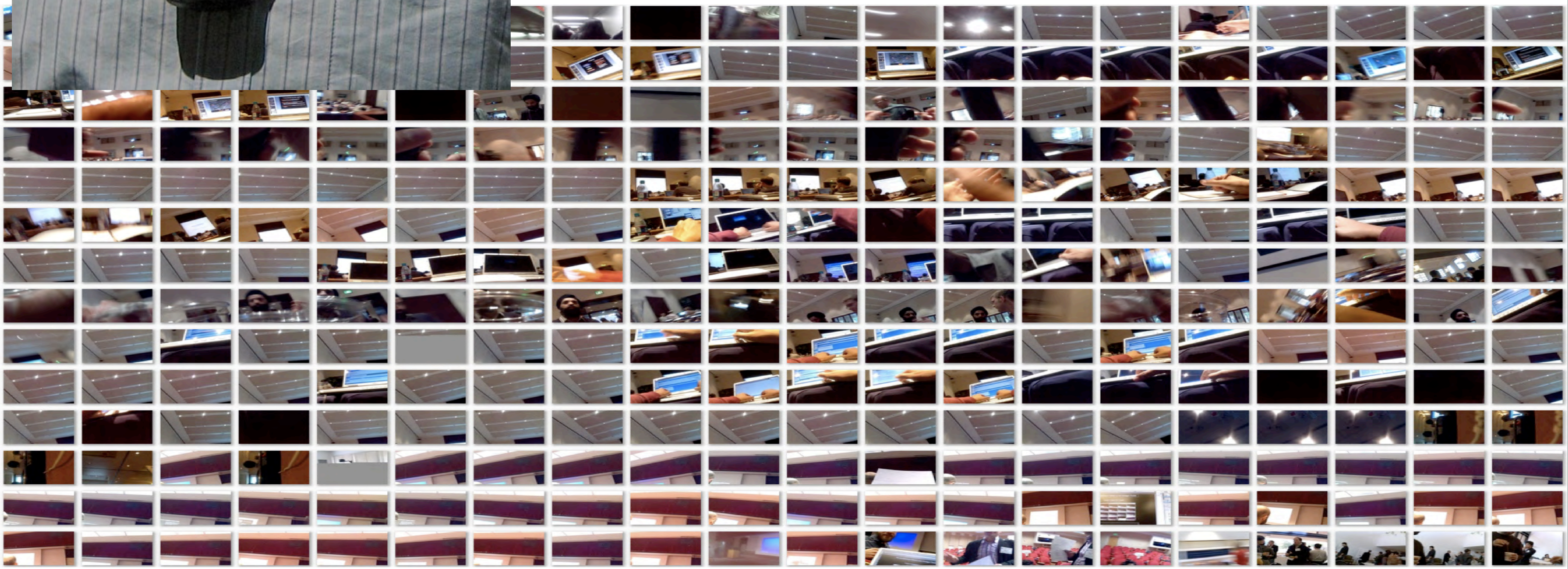


Staff and participant training user interface

Privacy concerns dominate system design

Automated pre-filtering to reduce number of viewed images

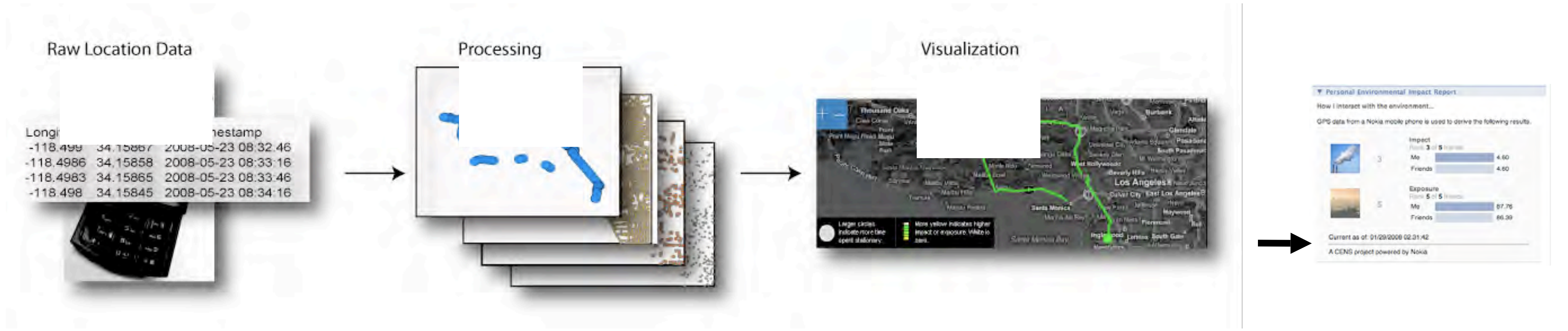
Privacy concerns dictate image viewing/tagging by individual, not by third party



Kim, Kim, Petersen, Arab, Burke, Estrin, ...

CENS
CENTER FOR EMBEDDED NETWORKED SENSING

Geo-coding as primary (not just meta) data:
leveraging real-time location traces



Device data capture and interaction:

- software on mobile prompts/captures and uploads
- data types: location image, audio, text tagging, worn sensors
- UI on phone

Processing

- activity classification
- mapping
- integrate with other GIS and realtime data about built and natural environment
- index into models
- privacy relevant filtering

Visualization

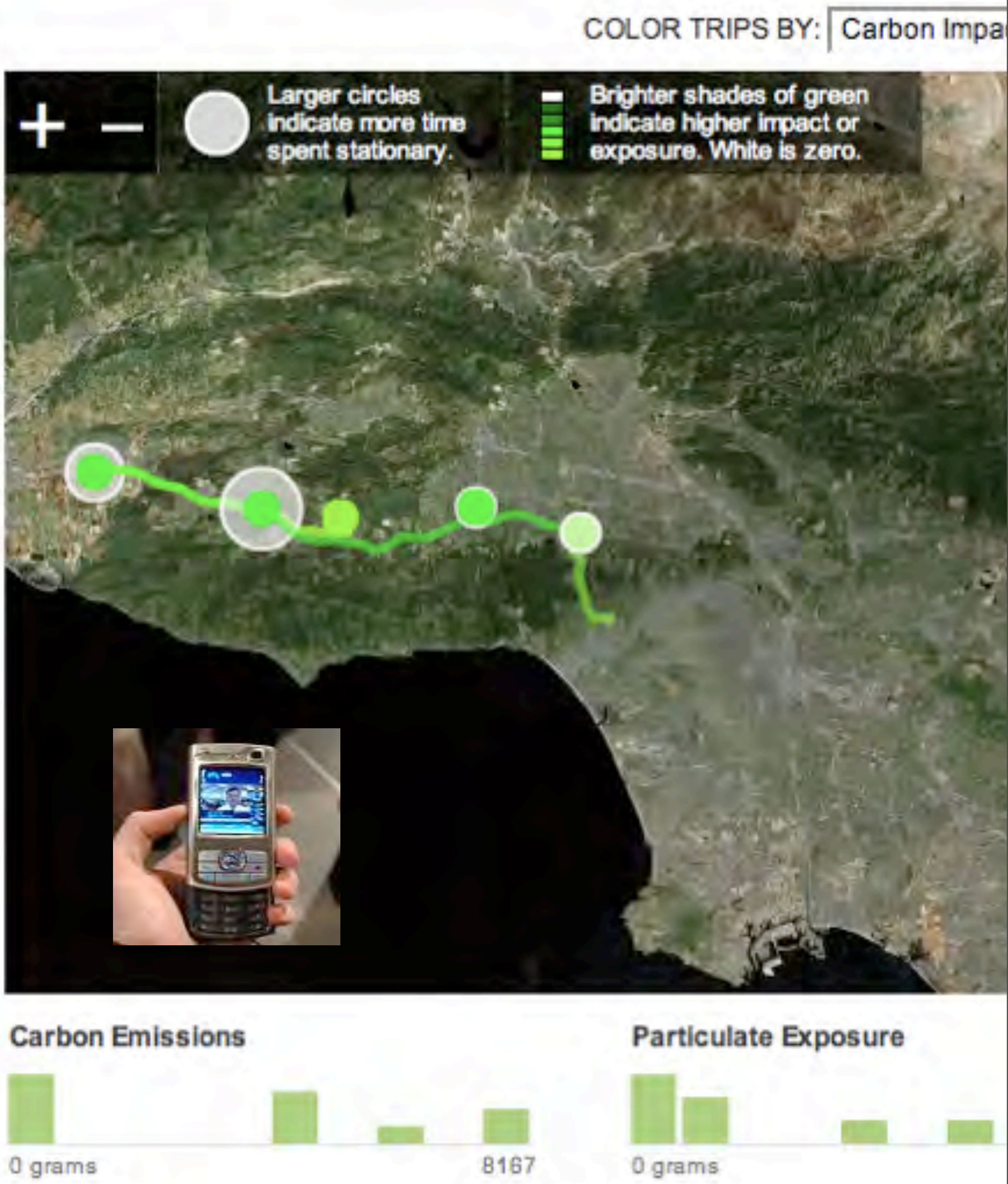
- for personal and professional insight
- legible, contextualized
- use/user configurable
- difficult to generalize
  - projects need support
  - platforms available for development

Sharing/Aggregating

- social networking
- web and device
- participatory privacy
- track data access for visibility/transparency

Imagine if….

Our everyday cell phones could show us *how we impact the environment, and how it impacts us*, just as they now alert us to traffic jams on the highway.



```
- <sensor type="tag">
    <prompt>Add An Annotation:</prompt>
  - <list>
      <item>Indoor</item>
      <item>Idle</item>
      <item>Walking</item>
      <item>Running</item>
      <item>Biking</item>
      <item>Freeway Driving</item>
      <item>Street Driving</item>
      <item>Bus</item>
    </list>
  </sensor>
```

Exposure Assessment

Personal Environment Impact Report

▼ Personal Environmental Impact Report

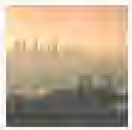
How I interact with the environment...

GPS data from a Nokia mobile phone is used to derive the following results.

3

Impact
Rank **3** of **5** friends.

Me		4.60
Friends		4.60



5

Exposure
Rank **5** of **5** friends.

Me		87.76
Friends		86.39

Current as of: 01/29/2008 02:31:42

A CENS project powered by Nokia

Exposure to Traffic-related Pollutants

Lifelong damage found in 13-year study of 3,600 Southland youngsters living within 500 yards of a highway.
The Los Angeles Times, 1/26/07

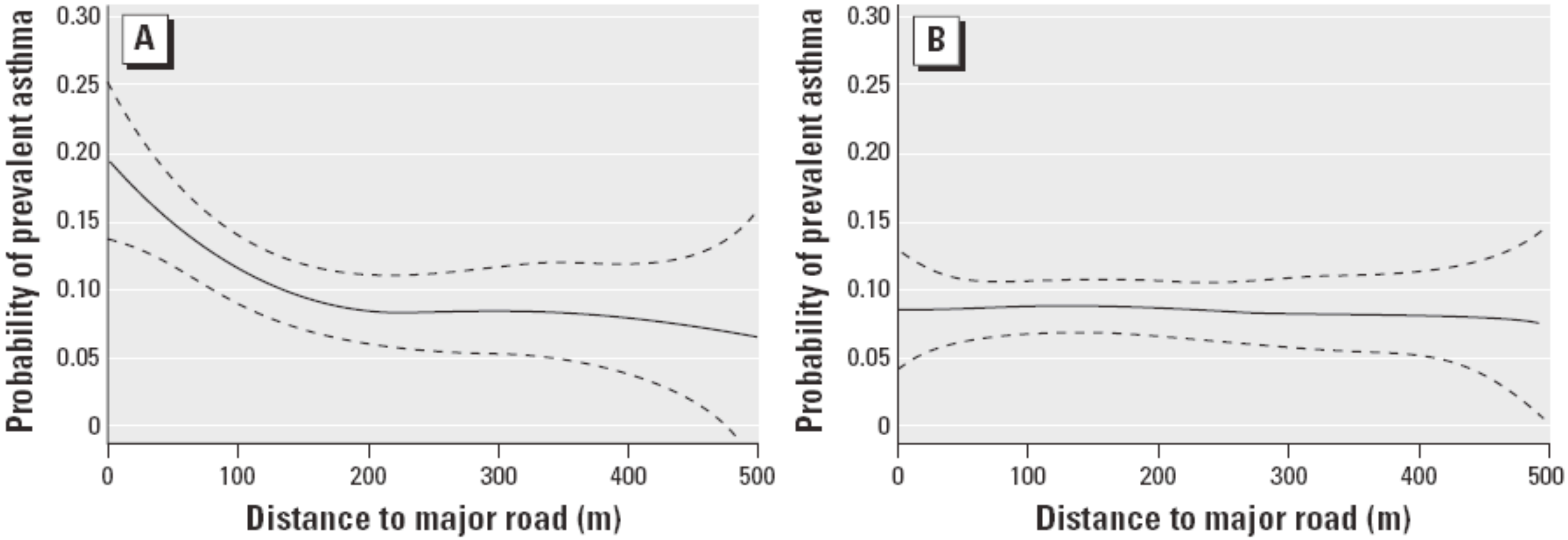
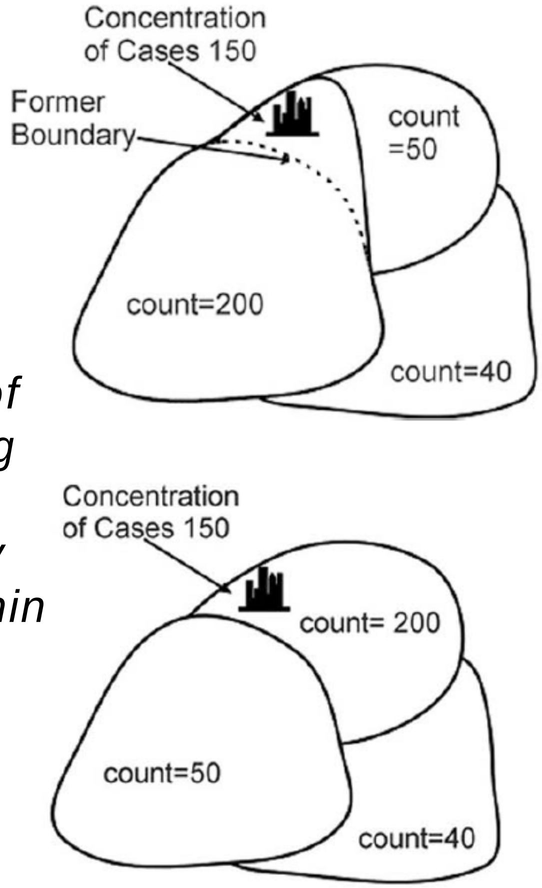


Figure 2. Prevalence of asthma by distance of residence to a major road within 500 m, among long-term (*A*) and short-term (*B*) residents with no family history of asthma. Dotted lines indicate 95% confidence interval.

Source: McConnell et al. Traffic, Susceptibility, and Childhood Asthma. Environ Health Perspect 114:766–772 (2006)

Implications of Scale and Zone Selection

- Health affected by “complex interactions between genetic and environmental factors”
- Measurement scale affects detection of relationships between exposure and health outcomes
    - Aggregation may obscure significant intra-county variation in exposure.
    - Disease incidence reported at county level … therefore, environmental exposure data should be aggregated at the same resolution



Example of the zoning effect on mortality events within a unit.

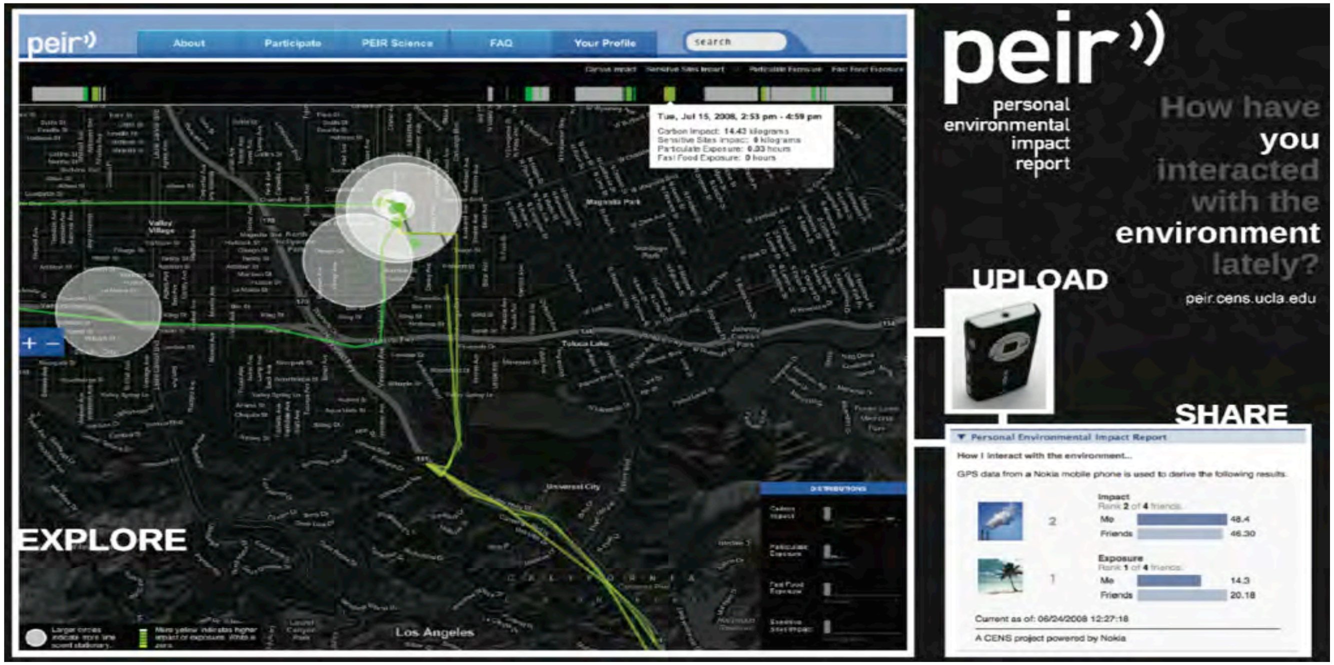
Source: Jerrett, Michael and Finkelstein, Murray (2005) 'Geographies of Risk in Studies Linking Chronic Air Pollution Exposure to Health Outcomes', Journal of Toxicology and Environmental Health, Part A, 68:13, 1207 - 1242

PEIR: Personal Environmental Impact Report

Personal, real-time, location traces....combined with micro-environmental models ...to provide personal exposure and impact assessment

Invite investigation of individual habits overtime:
...in relationship to others and the environment
...as seen in data and inferred from models.

http://peir.cens.ucla.edu



Mun, Yau, Burke, Estrin, Hansen, West, ...

peir personal environmental impact report

search

About | Participate | PEIR Science | FAQ | Your Profile

Powering Personal Choice for Global Impact

PEIR, the **Personal Environmental Impact Report**, is a new kind of online tool that allows you to use your mobile phone to explore and share how you impact the environment and how the environment impacts you.

What's unique about PEIR? Taking a step beyond a "footprint calculator" that relies only on your demographics, PEIR uses location data that is regularly and securely uploaded from your mobile phone to create a dynamic and personalized report about your environmental impact and exposure.

How PEIR Works

PEIR gives you greater control over your environmental impact and exposure by allowing you to interactively explore how it creates its results from your activity patterns.

Join PEIR

We're currently in private beta, but **sign up** to get notified when PEIR is open for new user registration. Also, feel free to **check out the demo**.

FEATURED

Press Release: UCLA Researchers Create PEIR Using Cell Phones as Sensors

We unveiled our new tool this week to help you understand your relationship with the environment, and we want to know what you think.

PEIR News

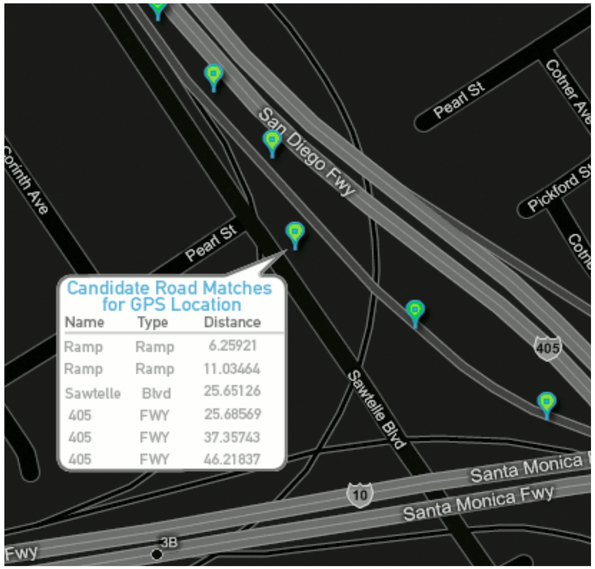
PM 2.5 Exposure Model Values

Since the PEIR launch, we have noticed some small bugs related to our modeling computations, but instead of blocking

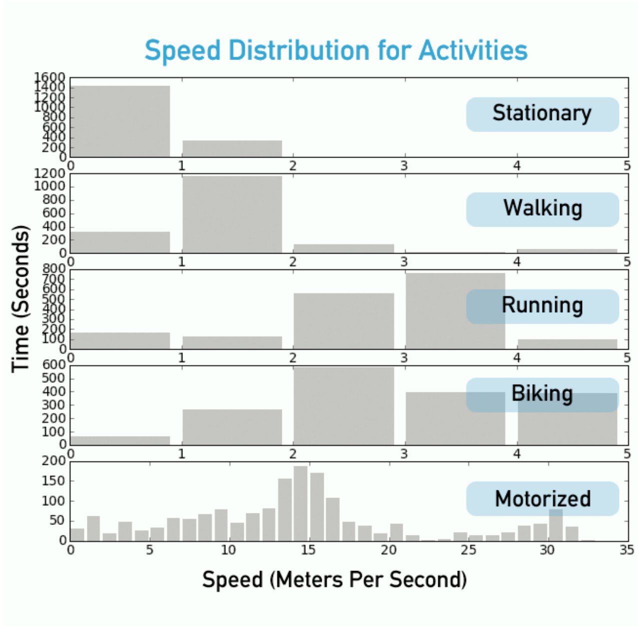
PEIR processing steps

- Users location traces are sampled and timestamped
- Activity annotation and trip chunking
- CO_2 and PM2.5 emissions are computed as a function of speed and weather conditions using California Air Resources Board Emission Factor (EMFAC) model
- Sensitive sites impact is computed using PM2.5 emission and location information
- PM2.5 exposure is computed using historic traffic conditions (SCAG traffic data).
- Fastfood exposure is computed using location information

Activity estimated from location trace on secure servers



Map Matching



Activity Classification

Annotate GPS trace with type of activity

- Now: Still, Walking, Driving; Soon: Bicycling; Someday: Public transportation

Process

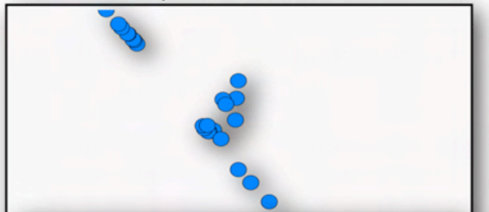
- Filter anomalous GPS points; Map match freeway; Speed feature from GPS reading.; Decision tree 6 scenarios (speed/freeway combinations); HMM recognition; Trip chunking w/configurable dwell time (10 mins).

CENS
CENTER FOR EMBEDDED NETWORKED SENSING

Location-Activity Trace processed through scientific models

Location Trace Processing

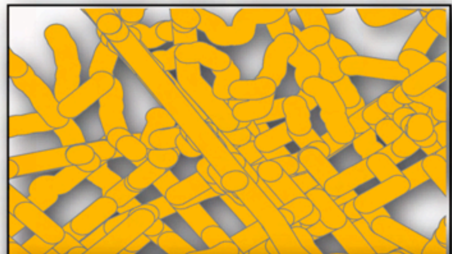
Location Trace + Weather
+ Activty Classification



Lat.	Lon.	Temp (K).	R.H.	Act. Class
34.048257	-118.447998	282.039	80%	Driving
34.045795	-118.445858	282.039	80%	Driving
34.02637	-118.429833	284.817	79%	Driving

+

Road Buffers

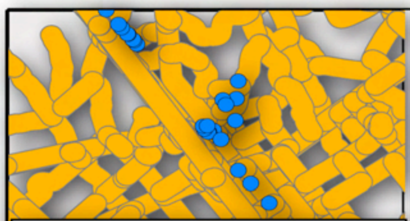


Name	Length (Mi)	A-B Car	A-B LDT	A-B HDT	A-B Speed
Westwood	0.38	2550	13	7	20
Westwood	0.36	3776	22	20	19
Wilshire	0.5	5095	36	25	23

+

EMFAC

Trip Aggregation

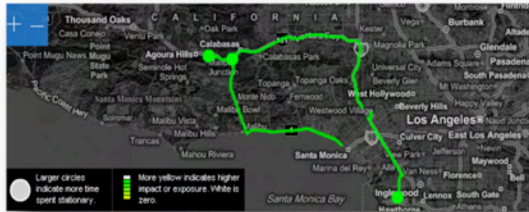


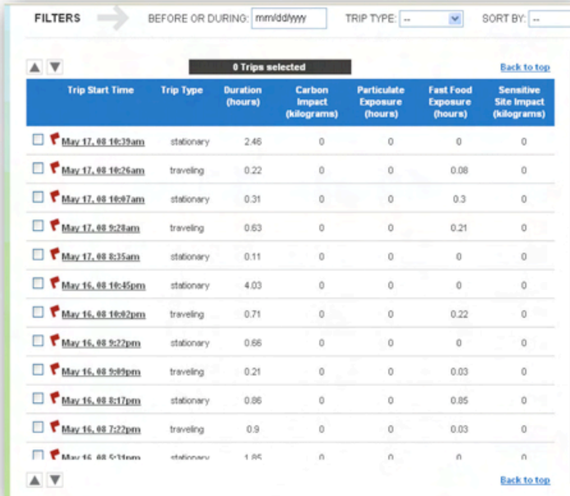
Lat.	Lon.	PM 2.5 (g/sec)	PM 2.5 (µg/m3/sec)
34.048257	-118.448	1.471601	0.443403
34.045795	-118.44586	2.533434	0.335633
34.02637	-118.42983	1.992753	0.360976

Trip Summary

Trip	Avg. PM 2.5 Exposure	time spent over 0.112716 (hours)
2432	0.470002	0.15
2423	0.235333	0.27
2410	0.210576	0.04

PEIR U.I.





FILTERS BEFORE OR DURING: mm/dd/yyyy TRIP TYPE: -- SORT BY: --

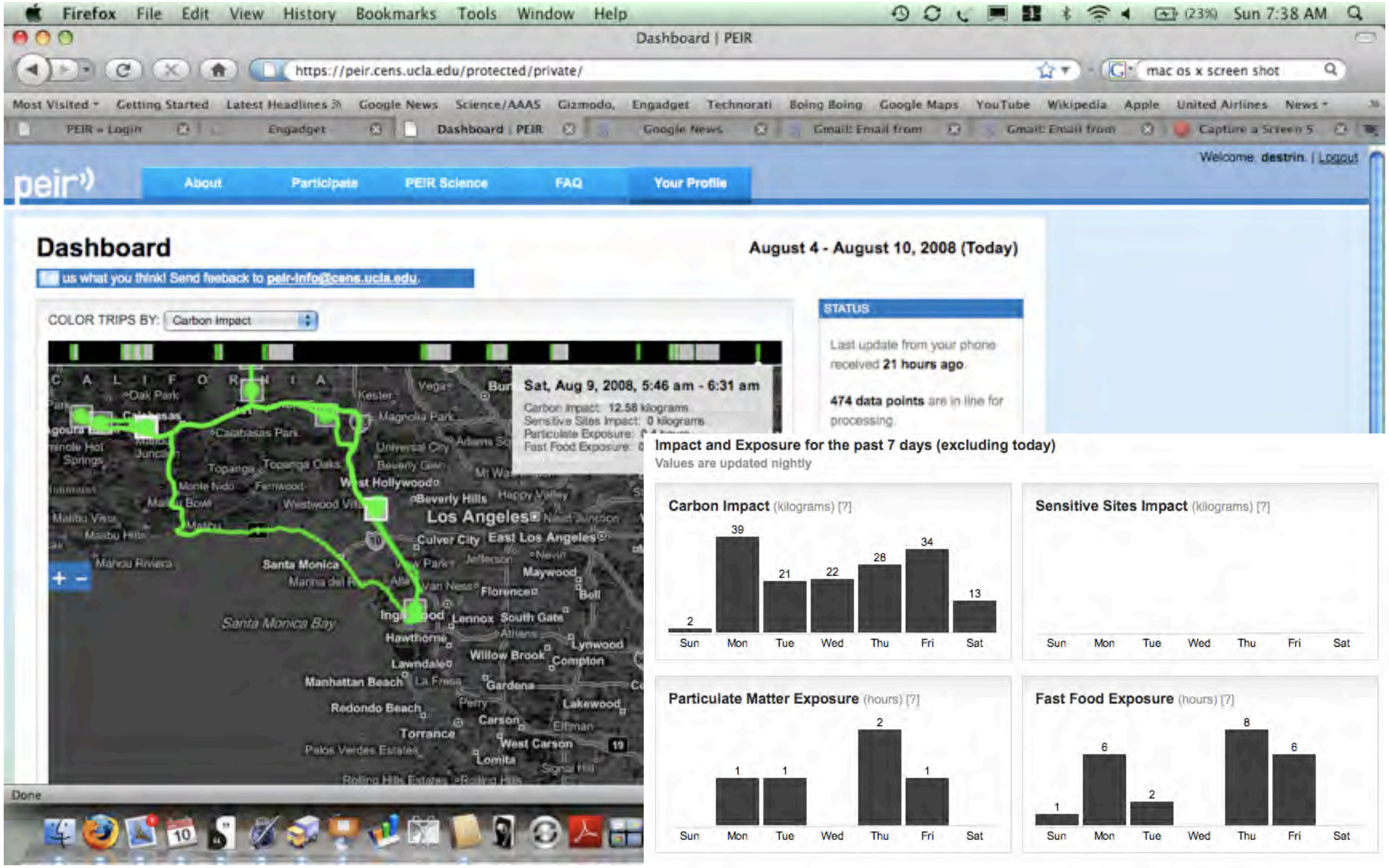
0 Trips selected

Back to top

Trip Start Time	Trip Type	Duration (hours)	Carbon Impact (kilograms)	Particulate Exposure (hours)	Fast Food Exposure (hours)	Sensitive Site Impact (kilograms)
May 17, 08 10:39am	stationary	2.46	0	0	0	0
May 17, 08 10:26am	traveling	0.22	0	0	0.08	0
May 17, 08 10:07am	stationary	0.31	0	0	0.3	0
May 17, 08 9:28am	traveling	0.63	0	0	0.21	0
May 17, 08 8:35am	stationary	0.11	0	0	0	0
May 16, 08 10:45pm	stationary	4.03	0	0	0	0
May 16, 08 10:02pm	traveling	0.71	0	0	0.22	0
May 16, 08 9:22pm	stationary	0.66	0	0	0	0
May 16, 08 9:09pm	traveling	0.21	0	0	0.03	0
May 16, 08 8:17pm	stationary	0.86	0	0	0.85	0
May 16, 08 7:22pm	traveling	0.9	0	0	0.03	0
May 16, 08 5:11pm	stationary	1.85	0	0	0	0

Back to top

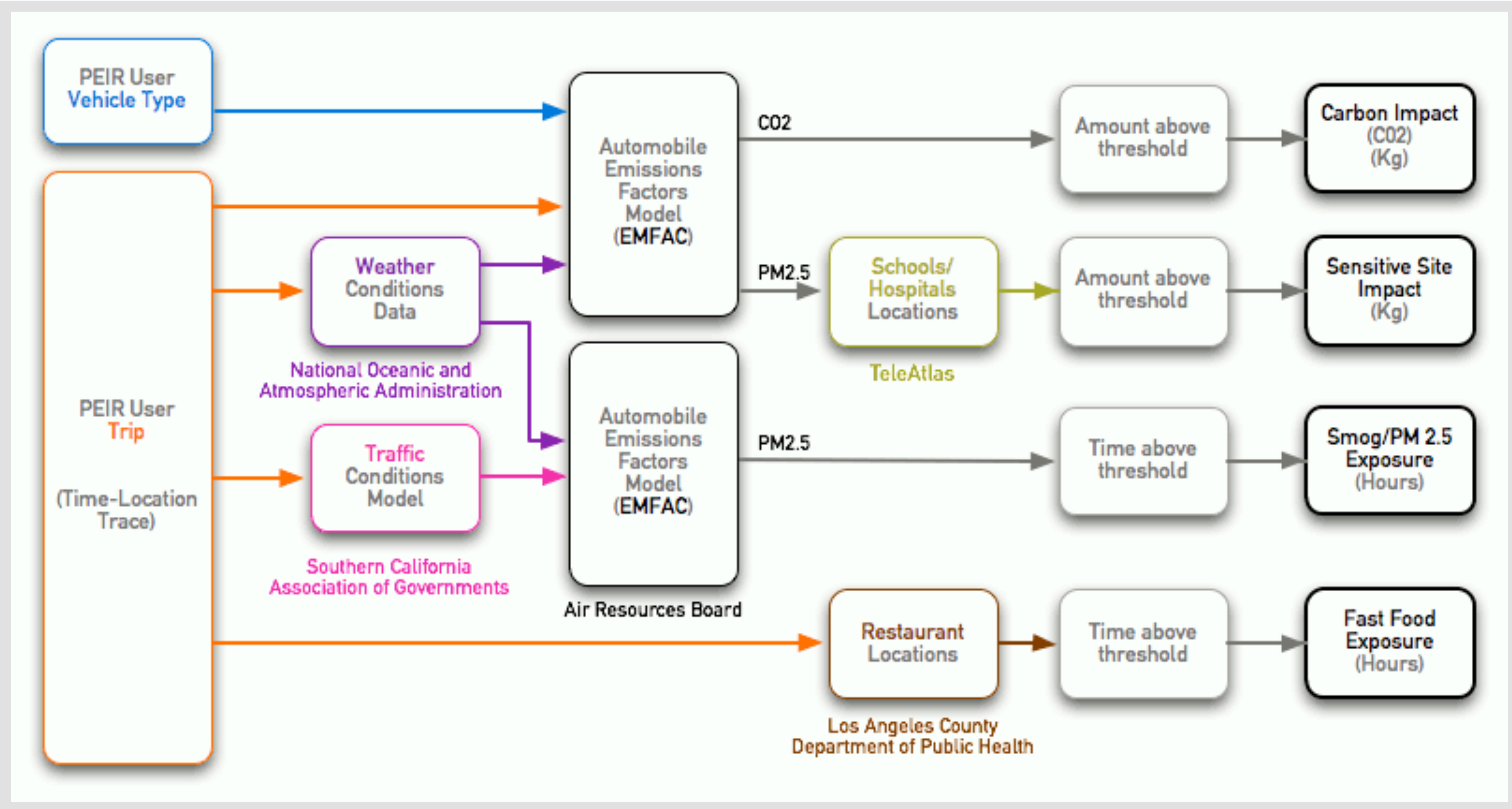
A week in PEIR



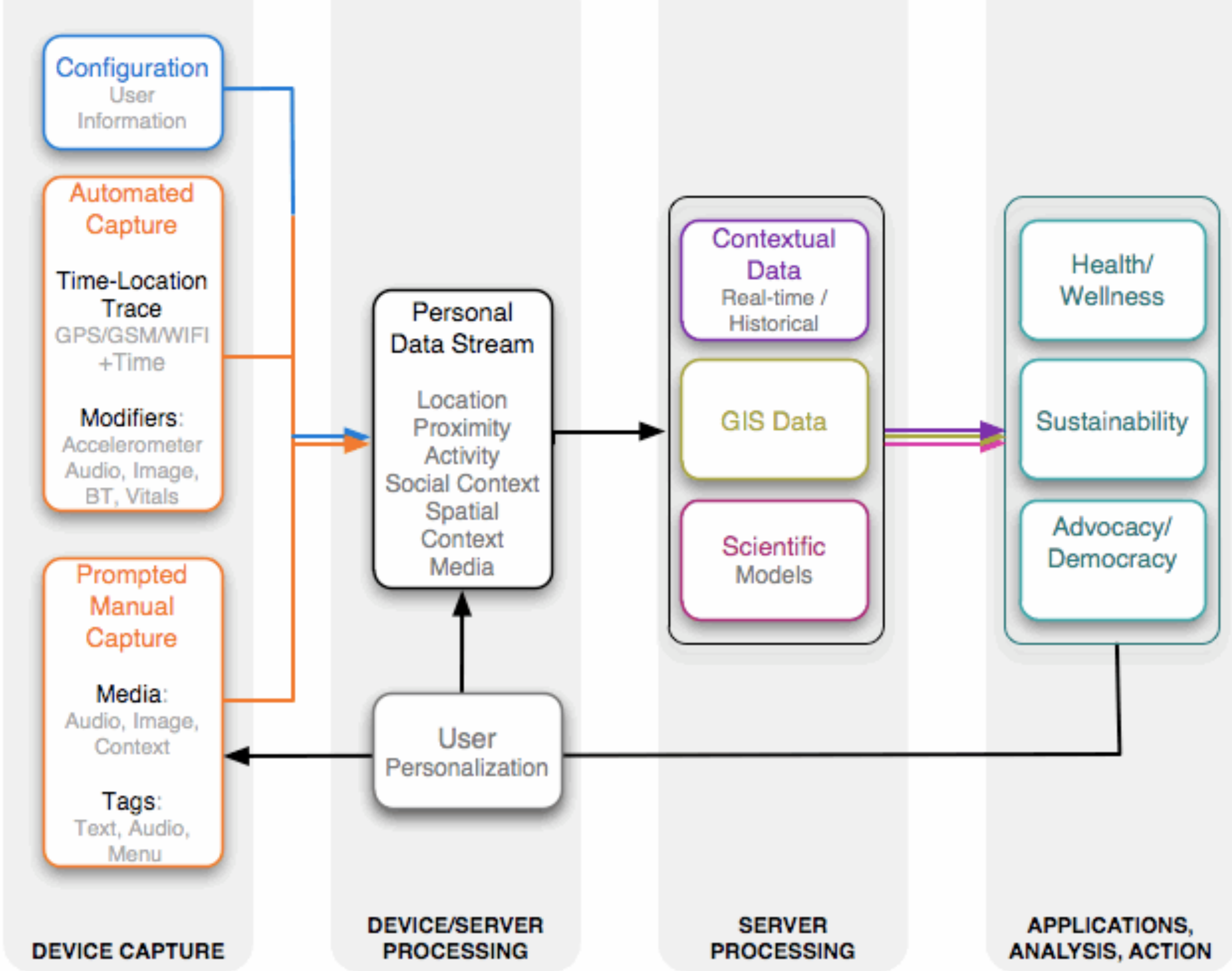
- User interface designed to promote data exploration and legibility
- User's data exploration begins with trip log
  - trip list sortable by model (e.g., most carbon impact/most particulate matter exposure)
  - calendar used to advance directly to specific points in time

CENS
Yau, Hansen, Burke, West, Estrin, Naik, Chandler, Mun, …
CENTER FOR EMBEDDED NETWORKED SENSING

System Challenge: Robust, Modular, Scalable Processing pipeline



Emerging mobile personal sensing system architecture



Mobile personal sensing for health and wellness

INFERENCE
Multi-scale Measurement to Meaning

- **Epidemiology Cohort/Longitudinal**
- **Intervention & Treatment Outcomes**
- **Personalized Health Management Behavior Change**

Data Privacy / Visualization

PROCESSING
Device and Web Services

Location · Activity Type · Social Context · Spatial Context · Media

Annotation / Analysis

CAPTURE
Mobile Devices

Time-location trace: (GPS or GSM/Wifi + Time)
Modifiers: Motion, Audio, Image, Proximity, Biometrics

Automatic / Prompted-Manual

CENS CENTER FOR EMBEDDED NETWORKED SENSING

In collaboration with Mary Jane Rotheram et al at Global Center for Children and Families

Epidemiology

Large scale longitudinal cohort and smaller targeted-population studies

- collect and analyze location/activity data and media-rich ecological momentary assessment at the individual and population scale
- monitor behaviors and risks for every single study participant for the entire duration of the study, continuously

Critical for studies aiming to understand dynamic, multilevel influences on health

- structural, environmental, community, neighborhood, institutional (school, clinic setting), family, and individual levels

Preliminary analyses can be executed in real time as the study progresses

- enabling additional experimental components to be introduced over time
- to gather the data needed to decipher the multilevel web of mediating and moderating influences
- multiple participants can report about the same event and person, with discrepancies clarified in real time
- personal, community, and national health outcomes

Treatment, prevention, and intervention research

Automatic monitoring of biomarkers, behaviors, cognitive and emotional states

- may be automatically transmitted to health providers and stored in personal health accounts to be routinely analyzed in standardized ways, while simultaneously documenting relevant and participant-approved aspects of their location and activity traces.

Link traces to self-perceptions, attributions, and relationship evaluations

- can substantially increase our understanding of medical conditions and offers an opportunity to dramatically improve the quality of care.

Increased capacity to evaluate the efficacy and side effects of particular treatment regimens

- from anti-depressants to chemotherapy, both in clinical practice settings and in research trials.

Real time monitoring of patient pain, fatigue, physical functioning, emotional distress, and social role participation

- may allow for better allocation of health care resources, especially for the 5% patients who currently utilize about 80% of family medicine visits for conditions easily managed by patients.

Similar to the just-in-time inventory management systems

- used by private enterprise and to the impact of precision agriculture on farm production--Each of these innovations transformed their industries into more environmentally-sound and economically-affordable practices.

Individual health self-management

Chronic diseases, typically resulting from five habits

- how much & what we eat, exercise, alcohol use, & smoking
- account for 50% of the global disease burden

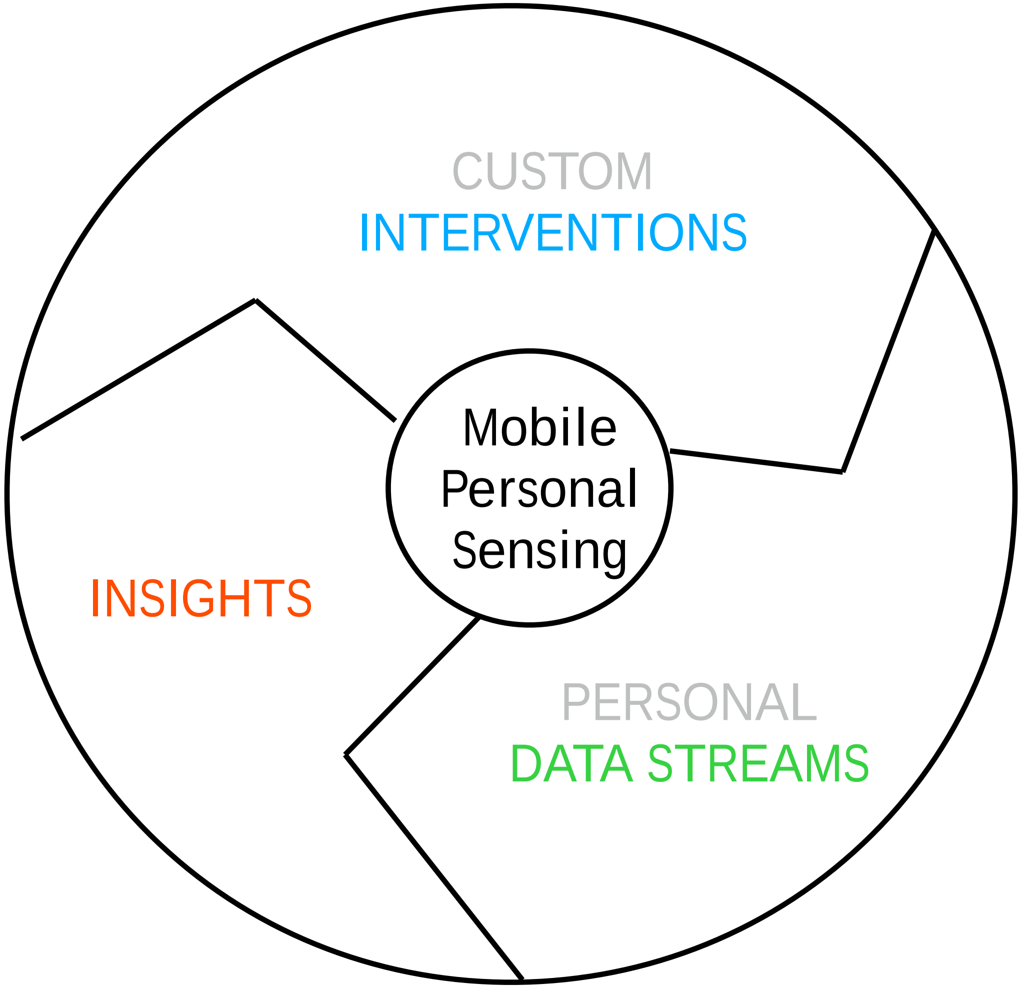
Personal mobile devices can be programmed to constantly provide personalized coaching

- for behavior change to adopt healthy routines, avoid relapse, and monitor their health status (e.g., adjust insulin doses for diabetics)
- including adherence with medical regimens and prescriptions

Provide individual feedback about the efficacy of behavioral changes and/or health status

- enabling individuals to adjust their medications and dosages as a tool for drug titration.

Individual health self-management: potential role of mobile personal sensing



Personal Data Streams

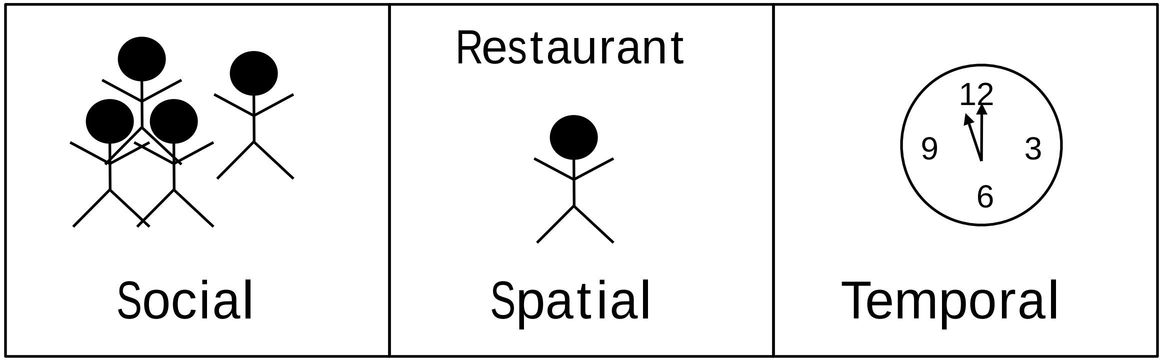
- automated identification of social context using Bluetooth traces; location; and activity
- user prompted to record audio, video, and images in significant contexts

User-tailored interventions

- triggered by significant social context, location, dime, or aggregate activity level
- dynamic and reconfigurable

Behavioral modification with mobile personal sensing: Obesity intervention example

Contextual Triggers



Contextual Triggers

Prompted data collection

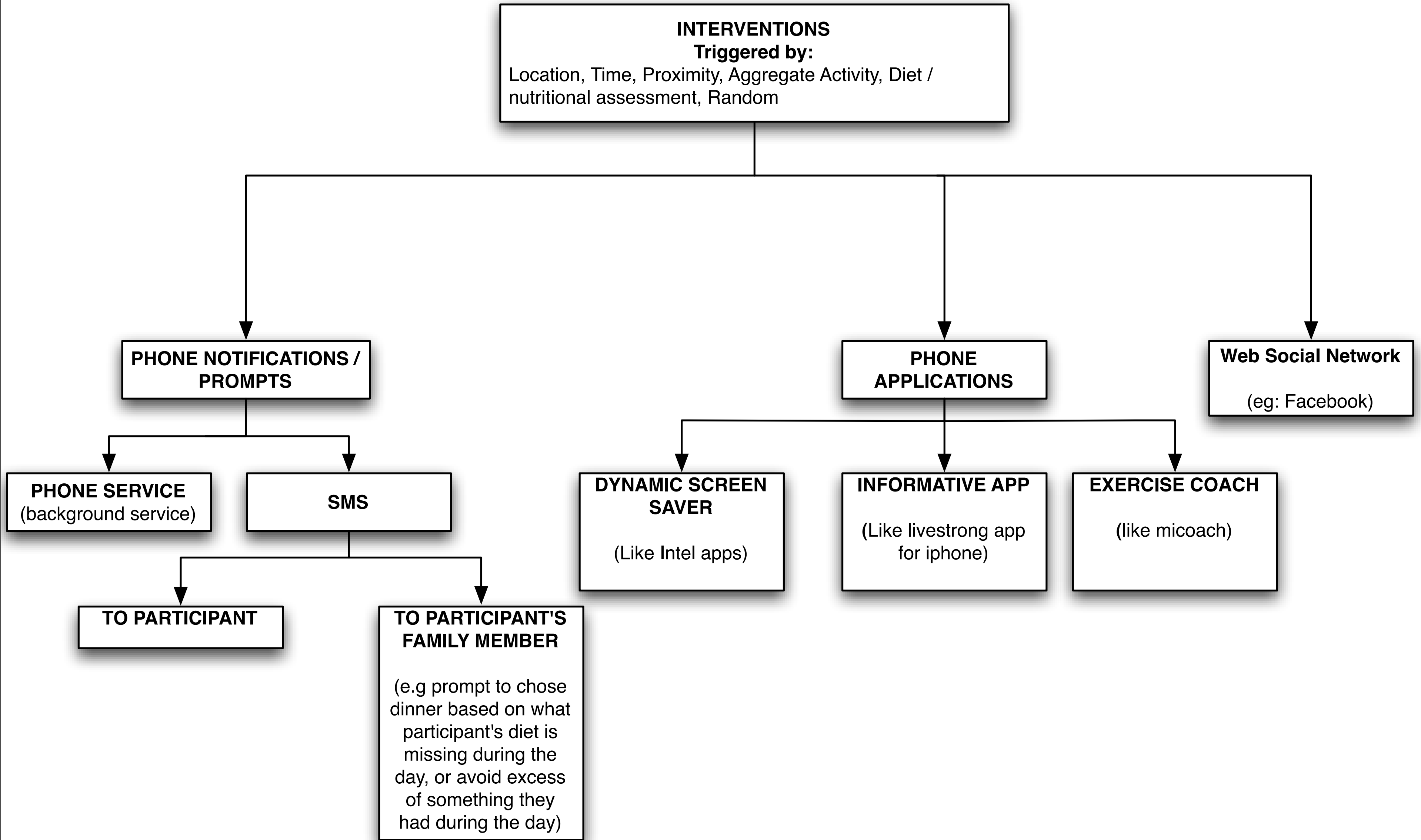
Reminders

User insights into behavior

Personalized interventions

User designs intervention

Implementing intervention software



Services needed to shift the locus of control to the user

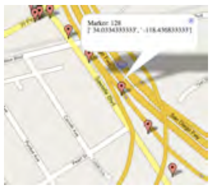
For these systems to be accepted and used, they will need to offer a degree of personalization and privacy that matches the intimate nature of their use.

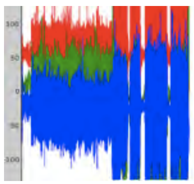
- A privacy preserving data collection service
  - collect data from devices on-board the mobile platform
  - give the user the ability to release, obscure, hide, or delete data
- An adaptive event-detection service
  - identify a variety of events
  - incorporate user feedback to adapt to individual environments

Services and interfaces strive to be non-invasive by automating key pieces of functionality--services involve the user only as needed

Many potential applications: activity and mobility profiles for those aging in place

- Observe patterns and trends in *indicative* activities of aging participants:
  - timing and frequency of trips to store, social activities, exercise routines
  - daily patterns of time spent in kitchen, dining area, TV room, bath/bedroom...
- Outdoor: time series of GPS and cell tower data points, combined with map matching





- Indoors: accelerometers and bluetooth stumbling

Automatic data collection from consumer grade devices
(mobility, proximity, image, acoustic signatures)
+
Legible presentation via Web based applications
=
Consumer-oriented, incrementally-adoptable, affordable, usable, individualized, solutions

Comparing Mobility Profile for Similarity

- Build a base mobility profile from context information (location/activity).
- This profile can be represented as an "association matrix" that captures the amount of time spent in a particular context during a time period.

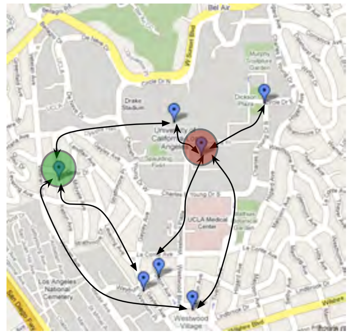
Amount of time spent at work.

	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.	Sun.
8 a.m.	0.1	0.0	0.1	0.2	0.1	0.0	0.0
9 a.m.	0.1	0.0	0.1	0.3	0.1	0.0	0.0
10 a.m.	0.1	0.0	0.1	0.4	0.1	0.0	0.0
11 a.m.	0.1	0.0	0.1	0.1	0.1	0.0	0.0
12 p.m.	0.1	0.0	0.1	0.0	0.1	0.0	0.0

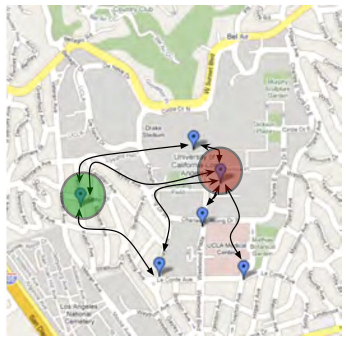
- Perform Singular Value Decomposition to obtain the "eigenbehaviors" (main column signatures in the profile)

- Compare periods of mobility information by calculating the similarity of eigenbehaviors for different time periods.

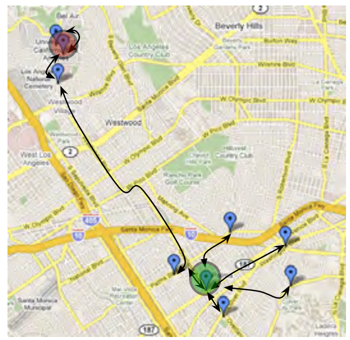
Base Profile



Natural Variations
(Participant at Different Restaurants/Stores)

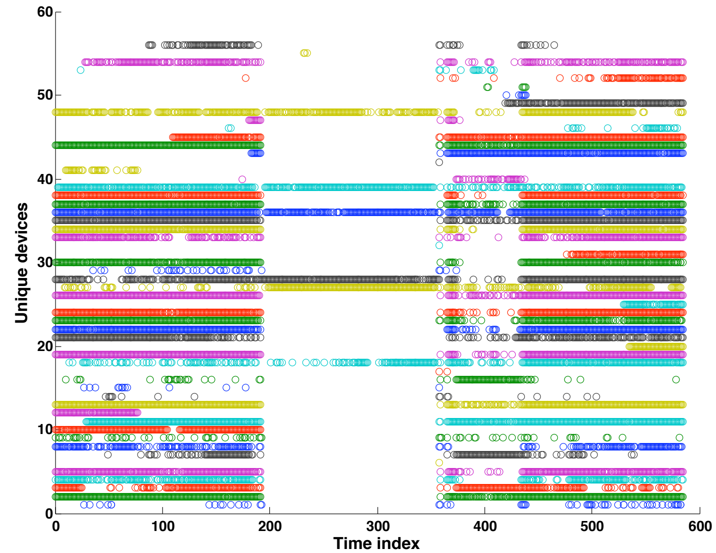


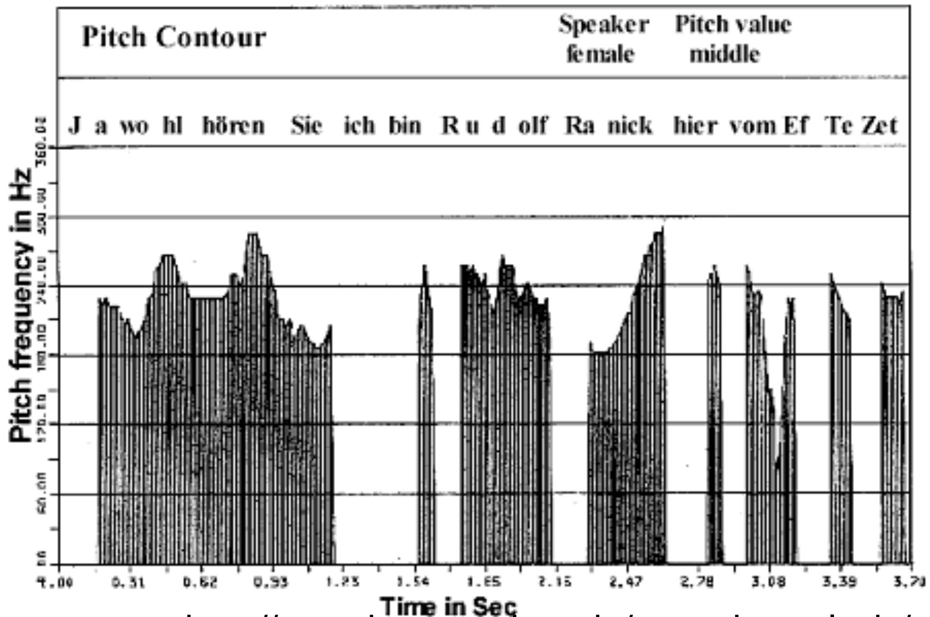
Re-learning Needed
(Participant Moved)



Social interaction: an interesting indicator at all stages of life

- Co-location interaction patterns give insights for families
- Near term: use bluetooth proximity
- Mid term: Estimate frequency, duration, trends in human communication using audio samples
  - Programed automatic capture of short audio snippets (avoid content)
  - Processed locally/on-server to detect patterns of interactive communication (distinguish from TV, Radio; phone, in person)
- Observe aggregate data to identify sudden or significant changes in social contact and interaction

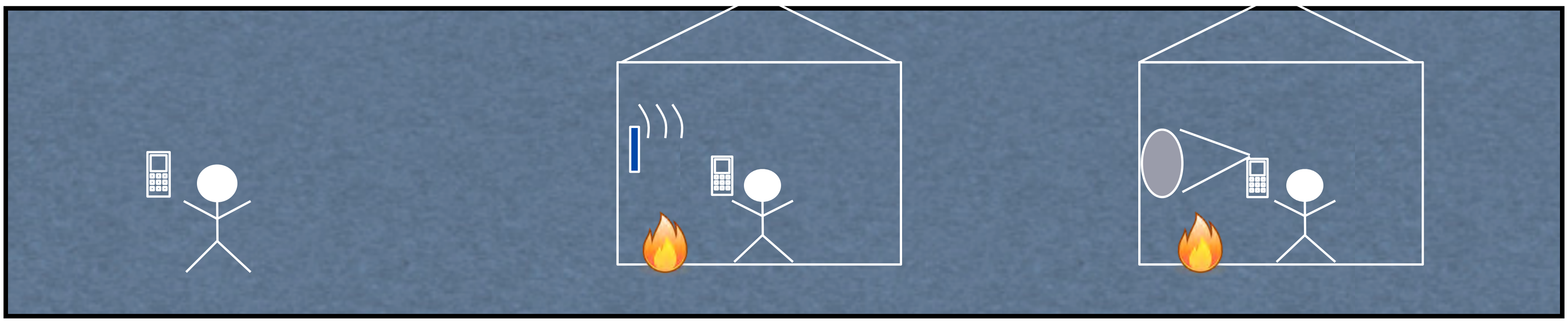




http://www.kt.tu-cottbus.de/speech-analysis/

Many potential applications: monitor villagers' pollution exposure before and after introduction of clean cook stoves

profile participants' daily activities and exposure to indoor air pollution in unprecedented detail using mobile phone based location and proximity traces



Outdoor Activities are inferred from GPS and accelerometer data

Duration of exposure to cooking fires inferred when a user is in range of a Bluetooth temperature sensor in the kitchen.

Pollution Levels inferred from images of a special filter installed in the house

Epidemiologists at Sri Ramachandra University will deploy the cell-phone tool along with surveys and professional observation to evaluate Project Surya's impacts on the health of villagers.

Algorithmic Challenge: Determining Transportation Mode On Mobile Phones

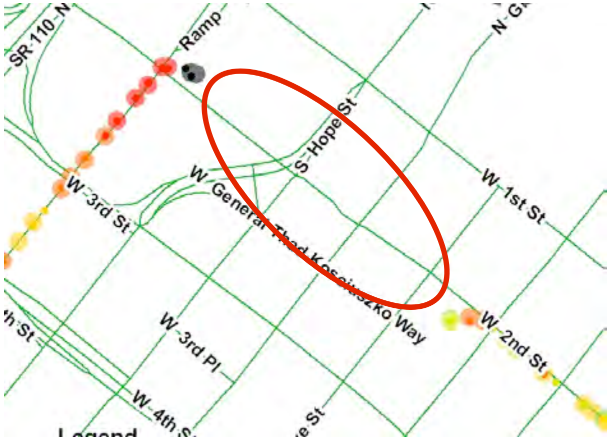
Mobile phones as instruments to understand physical processes in the world - a tool for introspection into the habits and situations of individuals and communities.

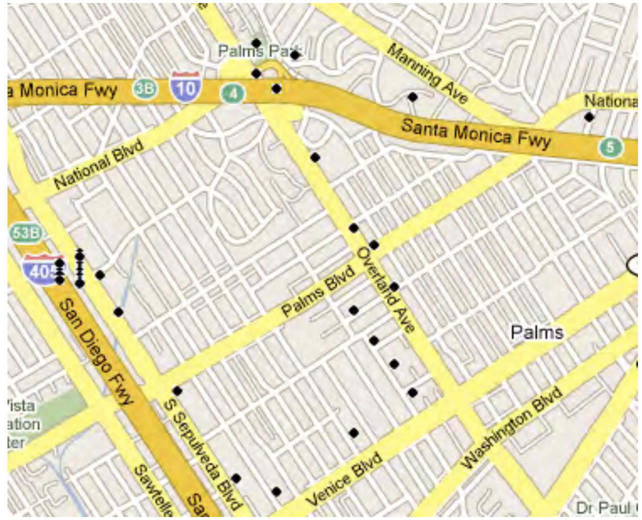
Many of these applications rely on contextual information about an individual such as transportation mode: stationary, walking, running, biking, motorized transport

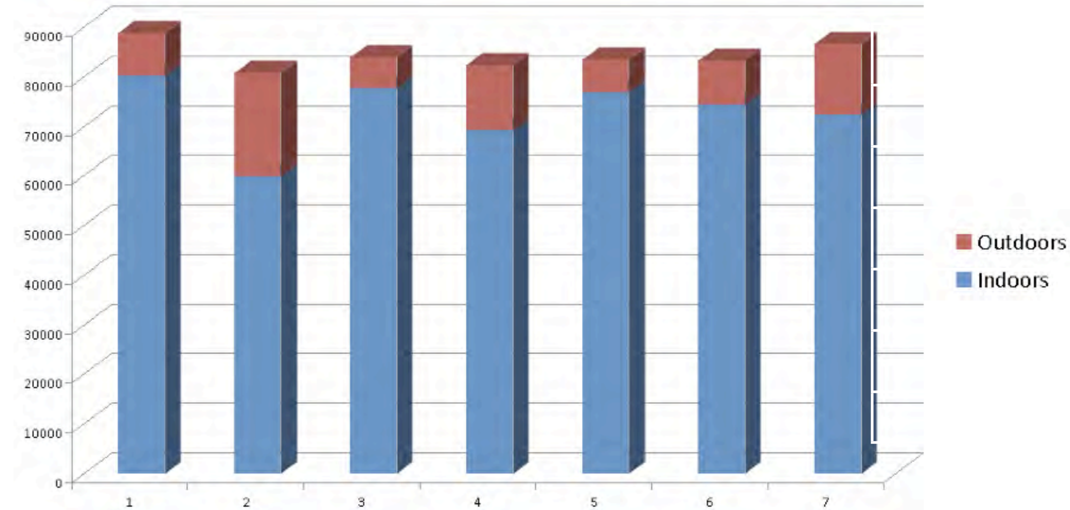
Extensive prior art

GPS, Contextual Models	Patterson 03 Liao 04,05,07 Zheng 08	- Models are too complicated to perform other tasks - GIS data is not always readily available
GSM	Anderson 06 Sohn 06	- A large portion of standard mobile phones does not release the information of multiple cell towers in range - They did not attempt to leverage smaller cell-size data such as Wi-Fi
Bluetooth	Tapia 04	- Bluetooth data is inappropriate to infer mobility states with different speed values because it is not practical to have static Bluetooth sensors distributed ubiquitously in outdoor settings. Also, it is difficult to distinguish whether an individual is moving or if the environment around him or her is changing
Wi-Fi	Bahl 00(RADAR) Ladd 02 Krumm 04(LOCADIO) Griswold 02 Muthukrishnan 06	- Wi-Fi data targets indoor environments with known access points and tower locations for localization.

Drawbacks of Using only GPS Data: coverage indoors/built areas, power draw







Activity	Power(Watts)
Phone Idle	0.054
GSM Sampling	0.056
GSM, WiFi Sampling	0.23
GPS Outdoor Sampling	0.407
Accelerometer Sampling	0.111

Satellite Visibility Variation

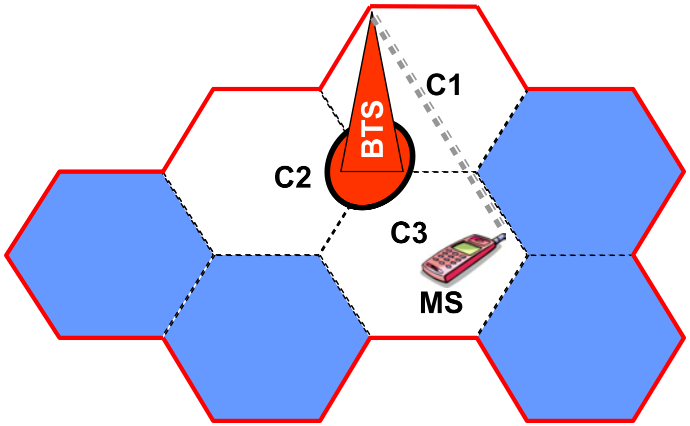
Wilshire	Palms	UCLA	Marina Del Ray	East Culver City





Poor	Moderate	Very Poor	Very Good	Good

Consider GSM and Wi-Fi



Cell tower locations can be used to **roughly indicate a user's location**.

Cell sizes in urban areas are **small/ medium** and density of BTSs is **high** [*cell-ID location technique,limits and benefits: an experimental study, WMCSA 04*].

- Note that we do not try to find a user's exact location using location of WiFi access points. So neither a priori knowledge nor estimated location of access points are required.

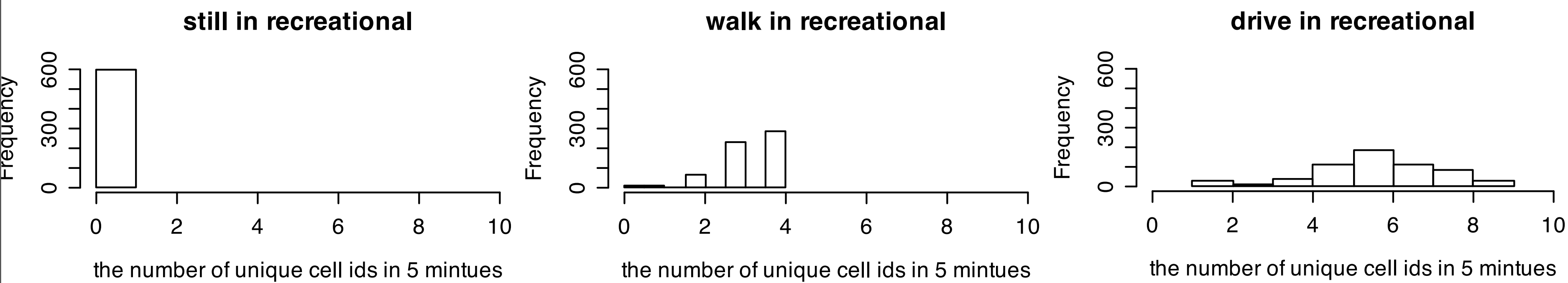
GSM Features

- Number of Unique Cell IDs (C unique,w)
- Number of Cell ID Changes (C changes,w)
- Residence Time in a Cell Footprint (C residence)

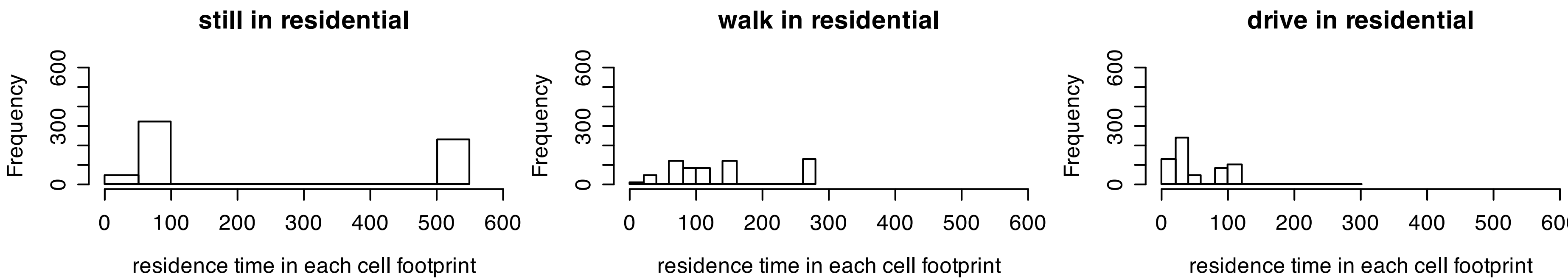
Time	1	2	3	4	5	6	7	8	9	10	11	12
Cell ID	1	1	none	1	1	2	2	1	2	2	2	3

Feature Values at time 10 are,
C *unique* = 2 (I and 2)
C *norm_unique* = 2/9 (#valid points = 9 due to no cell id at time 3)
C *changes* = 3 (1->2, 2->1, 1->2)
C *norm_changes* = 3/9
C *residence time* = 3
Where window size = 10

Number of Unique Cell IDs



Residence Time in a Cell Footprint



Wi-Fi Features

- Duration of Dominant Wi-Fi Access point visibility
- Proportion of Duration of Dominant Wi-Fi Access point visibility
- Signal Strength Variance

When we have measurement (Rx^a_1, Rx^a_2, Rx^a_3) at time a and (Rx^b_1, Rx^b_2, Rx^b_3) at time b, the value will be calculated as: $\sqrt{(Rx^a_1-Rx^b_1)^2 + (Rx^a_2-Rx^b_2)^2 + (Rx^a_3-Rx^b_3)^2}$

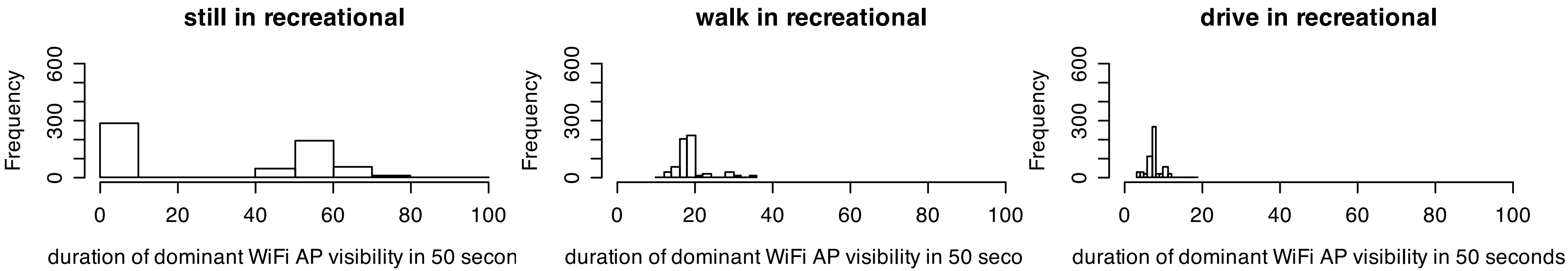
Time	MAC
1	{ **1**, 2, 3, 4, 5 }
2	{ 2, **1**, 3, 4, 6 }
3	{ **1**, 6, 2, 4, 5 }
4	{ **1**, 7, 4, 3, 5 }
5	{ 2, 7, **1**, 4, 5 }
6	{ 2, **1**, 3, 4, 5 }
7	{ **1**, 3, 2, 4, 5 }
8	{ **1**, 2, 3, 4, 5 }
9	{ **1**, 2, 3, 4, 5 }
10	{ 2, 3, **1**, 4, 5 }
11	{ 2, 3, **1**, 4, 6 }
12	{ 2, 3, 4, 6, 5 }

Feature Values at time 10 are,

WF *dominant* = 10 where the most dominant WiFi AP is 1
WF second_dominant = 9 where the second dominant WiFi AP is 2
WF *dominant_proportion* = 1 (10/10)
WF second_*dominant_proportion* = 0.9 (9/10)

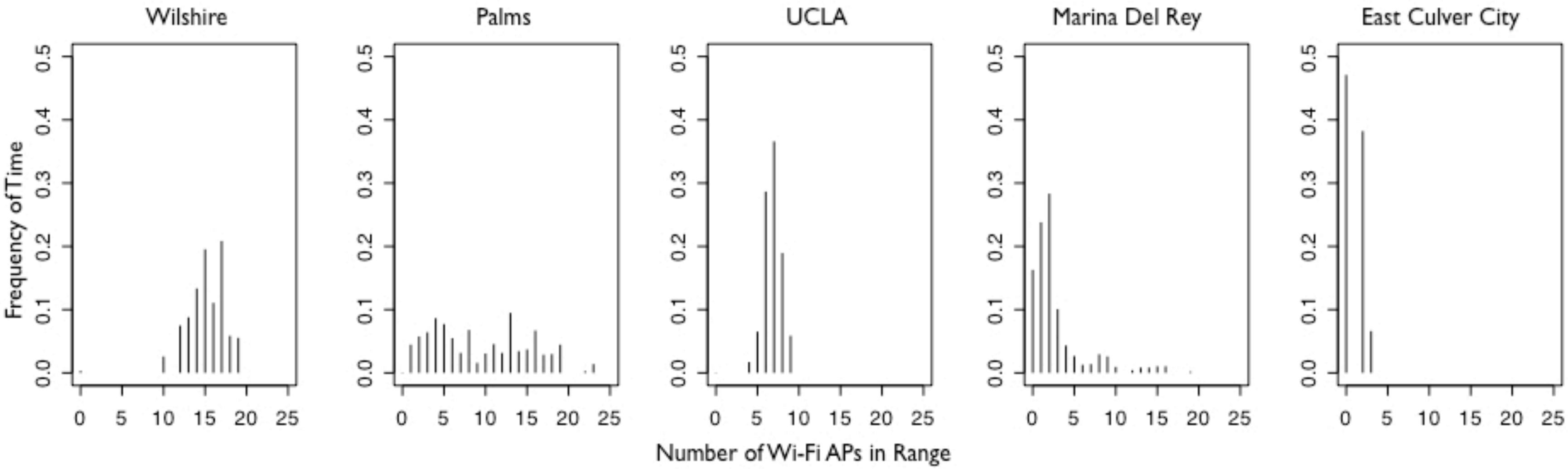
Where window size = 10

Proportion of Duration of Dominant Wi-Fi Access point visibility









High

Medium High

Medium

Low

Very Low

Experimental Setup

Hardware: Nokia N95
Software: A custom application written in Python S60

Every second, the application captured
“the primary cell tower ID, surrounding Wi-Fi beacons and GPS locations”

Areas

	Wilshire	Palms	UCLA	Marina Del Ray	East Culver City
Wi-Fi Density	High	Medium High	Medium	Low	Very Low
GPS Satellite Visibility	Poor	Moderate	Very Poor	Very Good	Good
Environmental Type	Commercial	Residential	Public	Recreational	Industrial

Participants

One user collected data in five different regions for fifteen minutes each for stationary, walking, and driving.

	Being Stationary		Walking		Driving		All	
	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision
GSM,Wi-Fi	92%	66.30%	66.10%	84.20%	82.90%	90%	**80.30%**	**80.20%**
GSM	70.70%	76.30%	71.20%	59.40%	68.80%	80.30%	70.23%	72%
Wi-Fi	60.60%	75.90%	61.40%	64.60%	84.30%	70.10%	68.77%	70.20%
GPS	92.50%	81.60%	91.20%	93.50%	91.30%	98.90%	91.67%	91.30%

Leveraging Accelerometer Data for Fine Grained Classification

Transportation mode classifier
- decision tree followed by discrete HMM
- distinguishes among stationary, walking, running, biking, motorized transport

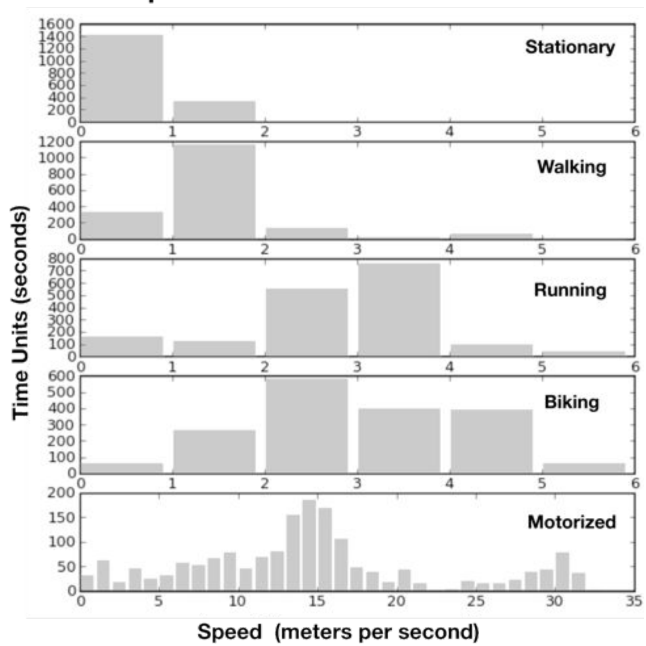
GPS receiver and 3 axis accelerometer as sensors

System does not have strict position/ orientation requirements--worn outside or inside of clothes

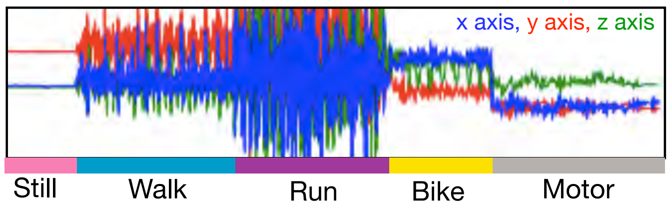
General classifier performance on par with user-specific and location-specific instances.

High accuracy levels in general
- greater than 93% - with user experiment witg 16 individuals

Speed Distribution for Activities



Accelerometer Data of User Carrying Cell Phone in their Pocket



Comparing Classifiers

Decision tree (DT) + Discrete HMM (DHMM) is the best classifier

- Easy to develop for and low computation overhead
- Size of DT Tree: 31, # of Leaves: 16, Height: 7
- DHMM fixes misclassification due to lack of state history knowledge

	Still	Walk	Run	Bike	Drive	Overall
DT	97.2	88.4	91.9	85.3	93.4	91.3
K Means	99.7	75.3	81.0	34.8	63.2	70.8
KNN (243)	97.2	77.4	51.2	51.2	95.3	83.0
Naive Bayes	96.6	88.0	84.2	84.2	92.9	90.9
SVM	97.4	86.9	87.1	87.1	89.4	90.7
CHMM	97.5	79.0	94.7	63.5	95.9	86.1
DT+DHMM	97.8	90.8	94.4	90.6	94.5	93.6

Mun, Reddy, et al

Activity classification future work

Adaptive mobility classification system

- different types of sensor data in various situations: e.g. when Wi-Fi APs are too sparse, only use GSM data; accelerometer when GPS speed and map matching makes inference ambiguous
- activate location and activity monitoring to capture outside events: avoid power draw of uniformly sampling when GPS has fix; trigger based on detected GSM-changes.

Opportunities to tune classification method.

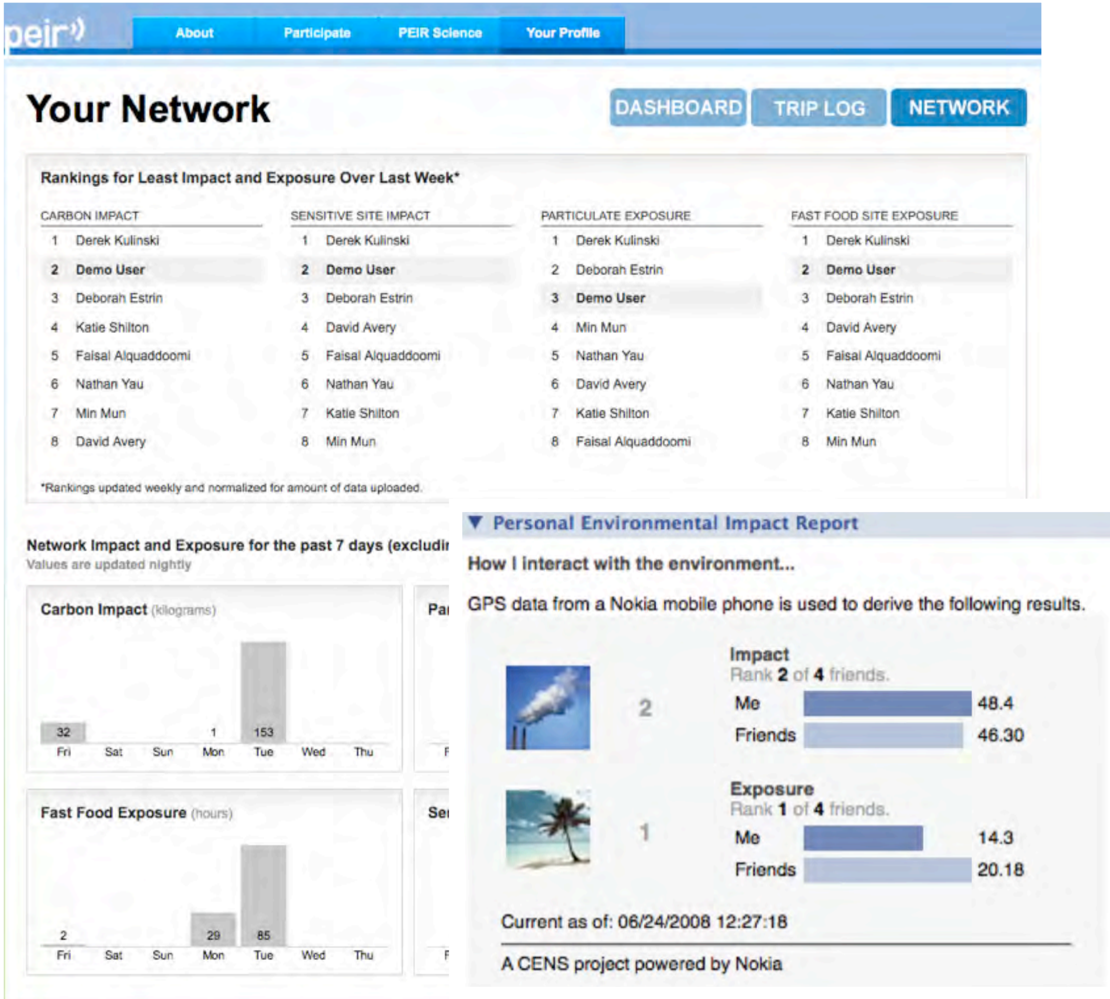
-Could user input or monitoring usage improve accuracy?
-How should we handle cases where features are not available?
-Could cost of capturing/processing features be incorporated?
-Does using different devices models affect the models?

Post-process to filter out unlikely series of activities

Research challenge for location/activity trace based systems: *individual control of time/space accountability*

- Location traces are revealing
  - Prevents "little white lies" for convenience, social cohesion
  - Makes omissions impossible
  - Might create chilling effect on legal but stigmatized activities
- Full disclosure is not inevitable
  - Selective sharing, hiding, remembering
  - Information flow control in supporting systems
- Abuse never preventable
  - Need strong audit trails
  - Legibility/transparency
  - Laws concerning fair use

Designing for privacy from the ground up



Share derived statistics instead of raw traces

- detailed data only accessible to individual

Simple example

- peir Facebook app/widget

Research challenges

- selective sharing and retention
- model-equivalent substitute data
- system transparency and audit-trail wrt data use and provenance

Technical/research challenges

Challenge 1: Information flow control for personal data streams

systems architecture in which the processing web/graph/pipeline is designed to maintain the integrity of both data and inference, with interfaces that are expressive enough to communicate application and data handling requirements--three dimensions key to supporting personal data streams:

- uncertainty to maintain confidence (e.g., error) measures through the stages of processing, aggregation, and inference.
- capabilities to express restrictions on which data can be shared and used, and the conditions under which data export should be subject to specified selective sharing filters.
- audit trail to support system transparency so that individuals can automatically and legibly extract records-of-access, use, inference, and manipulation of their data, at multiple points in the system.

can every datum that exists in the system have self-describing encoding of its uncertainty, capabilities, and audit trail?

- fixed policies and norms of confidence, privacy and transparency are not imposed on all users uniformly, but rather, that all data have associated metrics according to which the user can make personal judgement or negotiate terms.
- next steps include identifying relevant policies already defined by the HIPPA and privacy preserving data mining, database, and medical informatics communities, and using those policies as test cases for our proposed mechanisms.
- use of third party clearinghouses and contract mechanisms, will be key to defining the ecosystem of institutional, social, legal, and technical components, needed to serve the individual, as well as a rich array of health and other personal services.

Technical/research challenges (cont.)

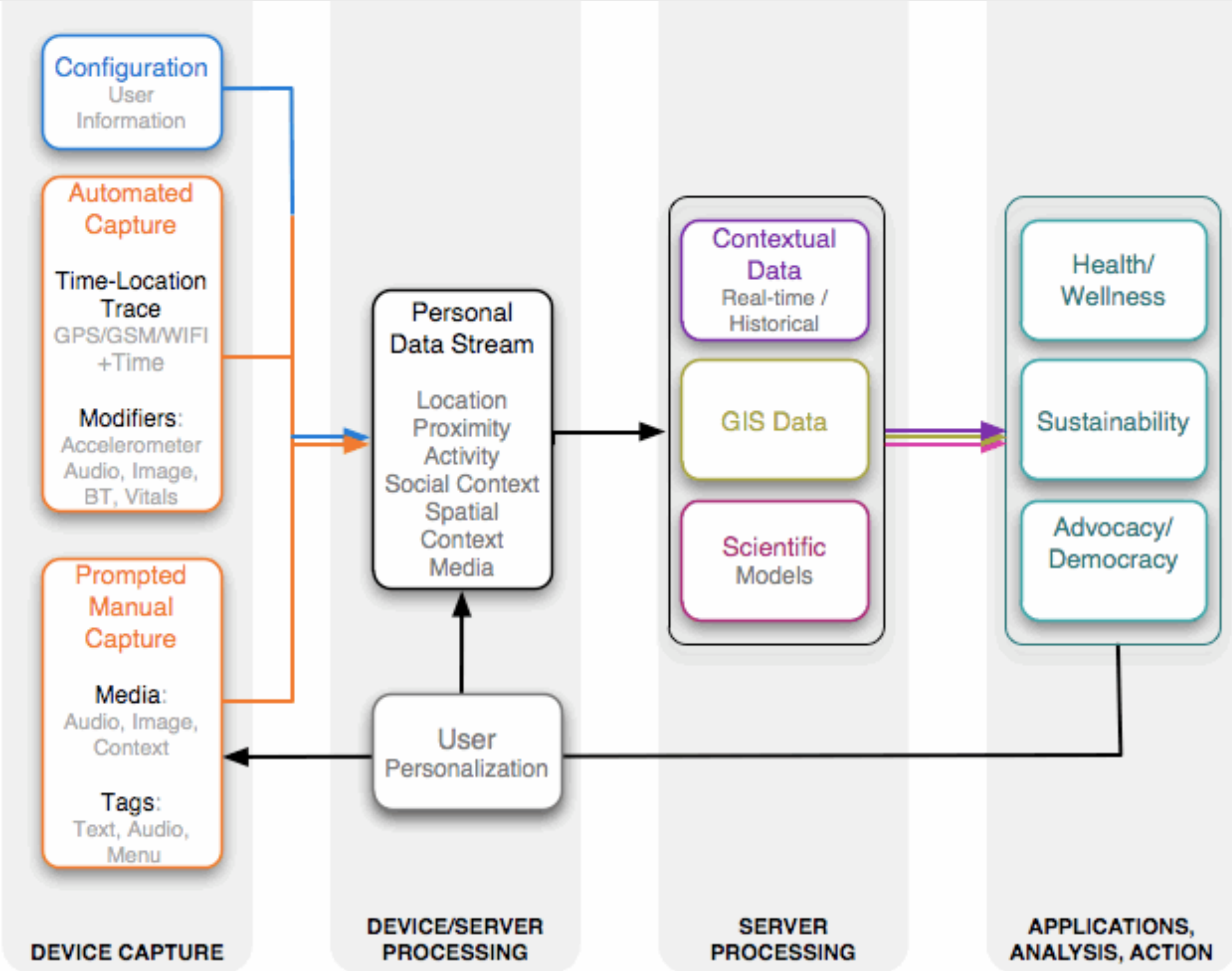
Challenge 2: Adaptive and personalizable components to run interactively and on behalf of individual users.

- activity classification and inferences of personal states rely on machine learning/training techniques, to improve accuracy of inference algorithms--training tasks will be important and visible element of user experience--creating usable adaptive and trainable interfaces for data capture and inference will be key.
- allow users to experience the full benefits of personal technologies by supporting them in the authoring of their interactions and applications--e.g., personal coaching application that user configures to trigger an intervention when the user enters a personally-and-dynamically-specifiable high risk context (spatial, social-proximity, ...), for some previously established undesirable behavior (eating, drinking, ..). How do we create legible, usable interfaces to expose this sort of capability to the user?

Challenge 3: Split device programming and runtime support

- use of both mobile and web based processing--on the phone for latency or privacy reasons, but in many cases web based information is needed to make most sense of data (maps, models, data sets, aggregation functions, etc.)
- even when a function is best supported on the device in steady state, training the local algorithm might benefit from web based processing and calibration.
- split device programming concern is not just at design time--during the daily life of an application the processing will shift back and forth between the device and the web infrastructure.

Emerging mobile personal sensing system architecture



Conclusion

If you can’t go to the field with the sensor you want…
go with the sensor you have! (Anon)

The power of the Internet, the reach of the phone (Voxiva)

Acknowledgments-Sponsors

- NSF: OIA, NETS-FIND Program, CRI Program, CISE, Engineering, Bio
- Nokia, Cisco, Sun, Intel, Samsung, Google, MSR, Crossbow, Agilent
- UC Micro, Participating campuses (UCLA, UCR, UCM, USC, Caltech)
- Wilson Foundation, Conservation International