MD2K Center of Excellence for Mobile Sensor Data-to-Knowledge

Advancing biomedical discovery and improving health through mobile sensor big data

Cornell Tech ♦ Georgia Tech ♦ U. Memphis ♦ Northwestern ♦ Ohio State ♦ Open mHealth ♦ Rice ♦ UCLA ♦ UC San Diego ♦ UC San Francisco ♦ UMass Amherst ♦ U. Michigan

Provenance-based Data Analytics CI for High-frequency Mobile Sensor Data (mProv)

Emre Ertin (Ohio State), Zach Ives (UPenn), Santosh Kumar (Memphis), Jim Rehg (Georgia Tech), Ida Sim (UCSF), Mani Srivastava (UCLA)

Growing Potential of High-Frequency Mobile Sensor Data

HEALTH AND WELLBEING

intelligent city

smart energy

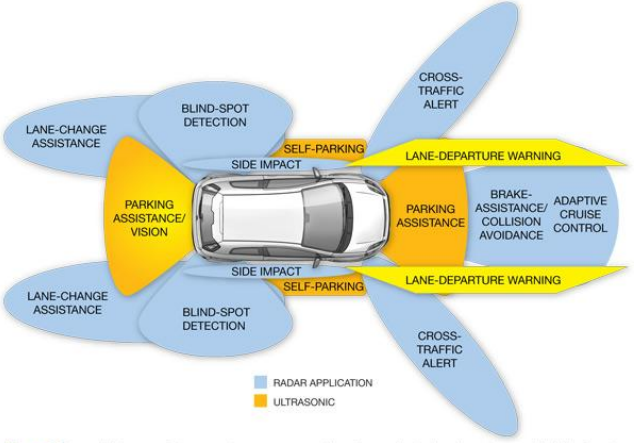
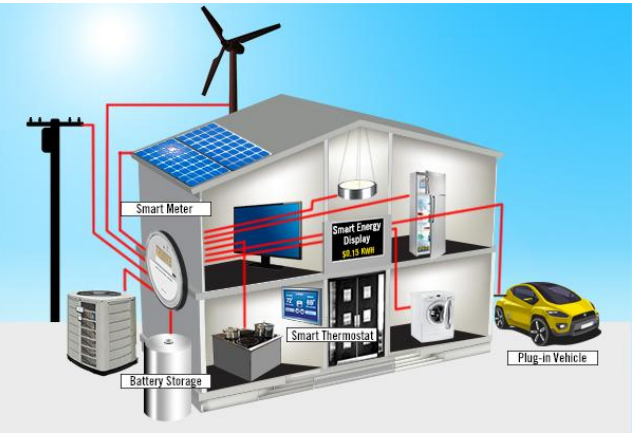
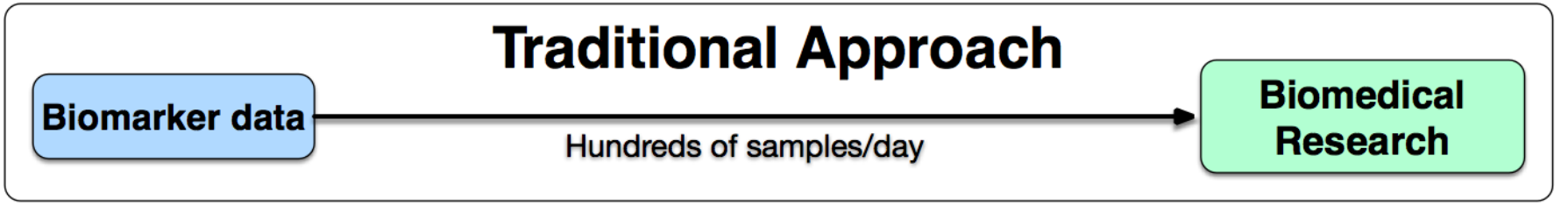


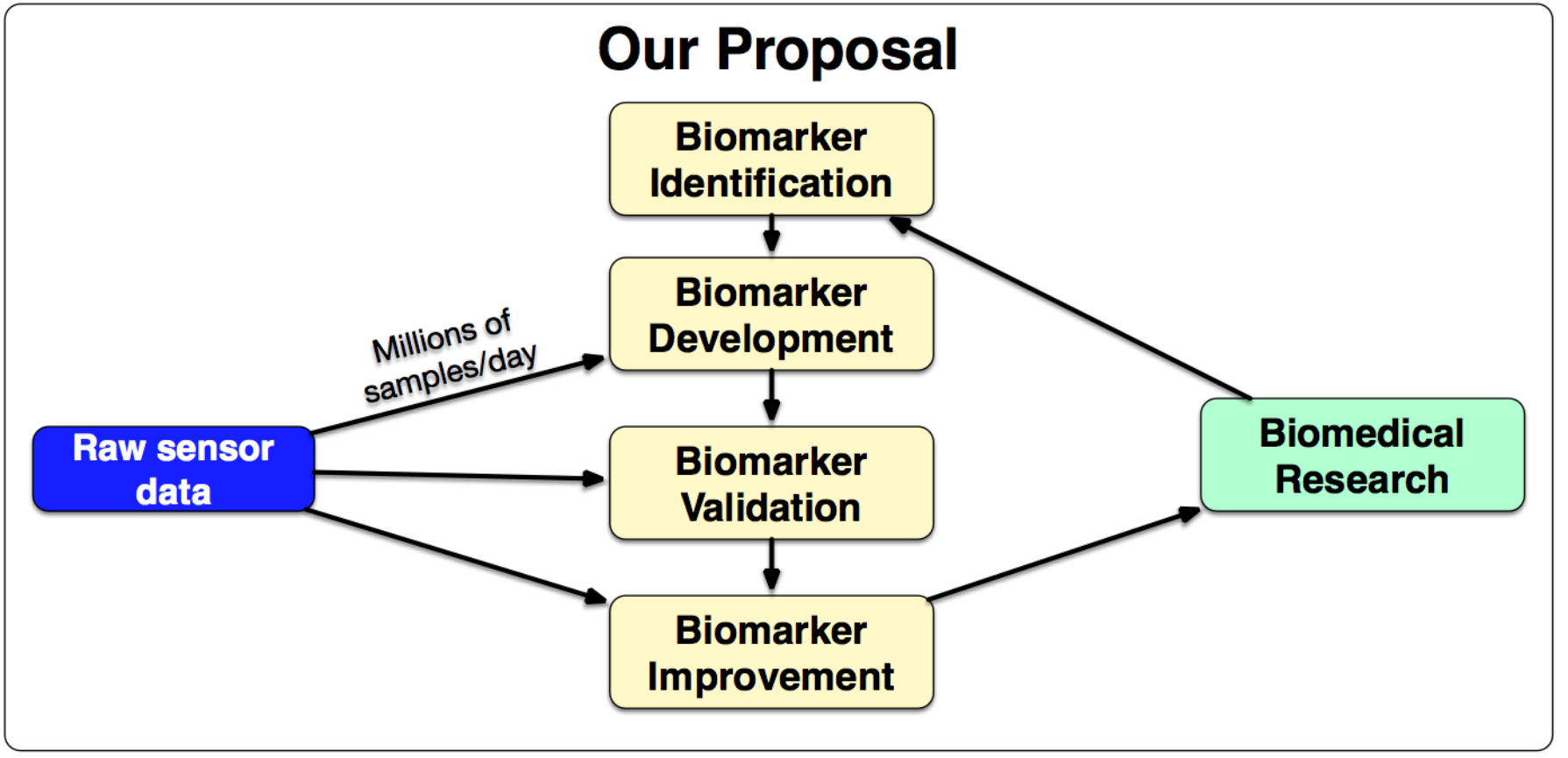
Figure 2 Several driver-assistance systems are currently using radar technology to provide blind-spot detection, parking assistance, collision avoidance, and other driver aids (courtesy Analog Devices).



Health applications are a natural focal point for research using sensor data

Enabling mHealth Data Science Research

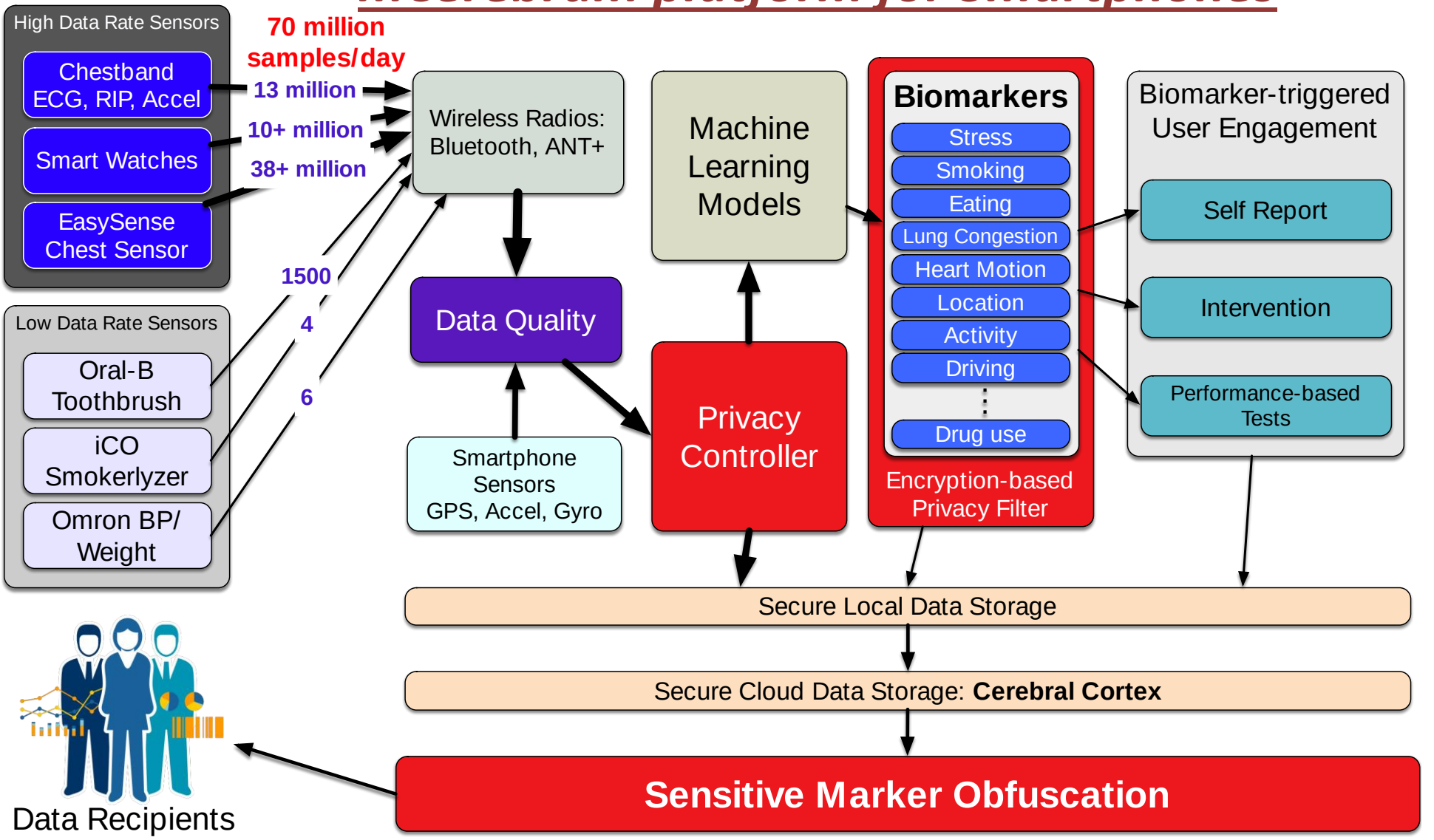





CORNELL NYC TECH
UCLA
Georgia Institute of Technology
UMASS AMHERST
UCSF University of California San Francisco
THE UNIVERSITY OF MEMPHIS
Open mHealth
THE UNIVERSITY OF MEMPHIS
UCLA
Computer Science
UNIVERSITY OF MICHIGAN
MD2K
Center of Excellence for Mobile Sensor Data-to-Knowledge
Medicine
Behavioral Health
NORTHWESTERN UNIVERSITY
OHIO STATE
Electrical Engineering
RICE
OHIO STATE
UCLA
UC San Diego

MD2K Mobile Software Platform (open-source)

mCerebrum platform for Smartphones

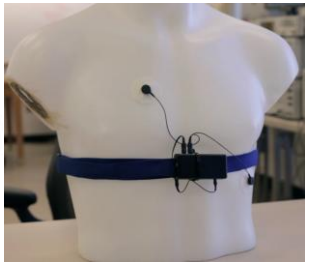


User Studies Using MD2K Software

Study	Target	Participants	Length (days)	Participant-days	Expected samples
Northwestern	Smoking, Stress	210	14	2,940	203 billion
Ohio State	CHF	225	30	6,750	437 billion
UCLA	Oral health	157	180	22,595	878 billion
Vermont	Smoking	90	14	1,260	87 billion
Rice	Smoking	300	30	9,000	622 billion
Utah	Smoking	300	30	9,000	622 billion
Johns Hopkins	Cocaine	25	7	175	11 billion
Totals		**1,307**	**-**	**51,720**	**2.86 trillion (~150TB)**

Barriers in Conducting Research with High-frequency Mobile Sensor Data

Due to lack of data sharing, everyone needs to collect their own data



Sharing of raw mobile sensor data can accelerate research, but *provenance infrastructure* is needed to enable reproducibility and comparative analysis

Velocity	Variety	Volume	Variability	Veracity	Validation
Hundreds of samples/sec per sensor	Tens of sensors per sensor	Gigabytes per day per person	Variations in attachment, placement, signal quality	Multiple biomarkers from same sensor	Sources of validation for specific biomarkers

mProv: Provenance CI for High-frequency Mobile Sensor Data

- mProv is developing *data models*, *metadata standards*, *API's*, and *runtime support* for annotating sensor data streams with
  - Source – sensor type, placement, sampling rate, continuous/episodic
  - Semantics – number, probability, class/category;
  - Provenance – features and rules applied to obtain a biomarker;
  - Validation – specificity, sensitivity, benchmark, gold standard;
  - Privacy – user controls exercised and applicable privacy policies
- mProv will enable replay, interpretability, comparative analysis, and reproducibility

The mProv Team

Scientific Leadership	**Santosh Kumar (PI, U Memphis, analytics); Zachary Ives (Co-PI, U Penn, provenance); Ida Sim (Co-PI, UCSF, metadata); Mani Srivastava (Co-PI, UCLA, privacy)**
Scientific Consultants	Emre Ertin (*Ohio State*, sensor quality); *Open mHealth* (API integration); James M. Rehg (*GA Tech*, sensor data analytics)
Advisory Panelists	**CISE Advisory Panel:** Tanzeem Choudhury (*Cornell*); Vasant Honavar (*Penn State*); David Kotz (*Dartmouth*); **Health Advisory Panel:** Brian Bot (*Sage Bionetworks*); Nick Anderson (*UC Davis*); **Industry Advisory Panel:** Joe Corkery (*Google*); Mike O'Reilly (*Apple*)
Collaborators	Madeleine Ball (Open Humans Project, Participant Recruitment); Gary Wolf (Quantified Self, Participant Recruitment)