Applied Cyber-GIS in the Age of Complex Spatial Health Data

Marta Jankowska, UCSD

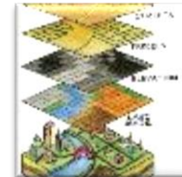
Complex Spatial Health Data

Big

- Volume
  - Person
  - Group
- Multiple data streams
- Data org. and processing beyond traditional software

Deep

- Multiple levels of info that are **connected**
- Allows to extract meaning and explore causal links



Contextual

- XY coordinates
- Place, exposure, spatial relations, context
- Does environment play a role?

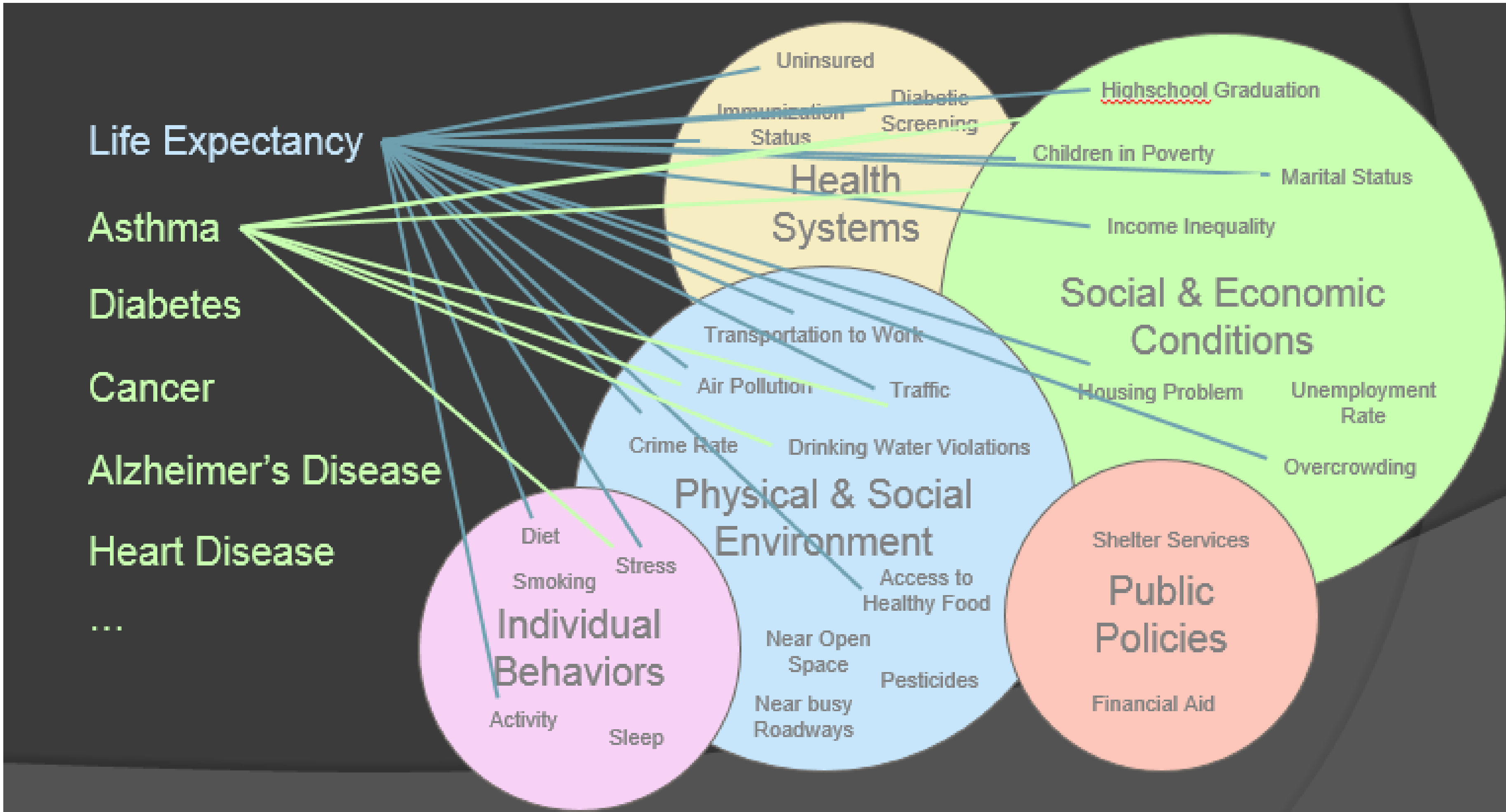
The Ideal Complex Spatial Health Data

- Health Outcomes
  - Electronic Medical Record
  - Self report
  - Sensor based
- Behavioral data
  - Survey based behaviors
  - Sensor based
  - App based
  - Life course data

- Omics
  - Genetics
  - Microbiome
  - Etc.
- Context
  - Exposures
  - Access

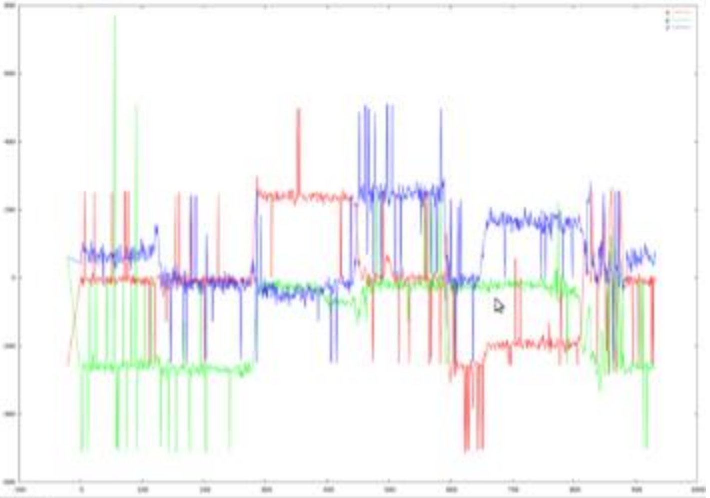
Targeted Individual Intervention

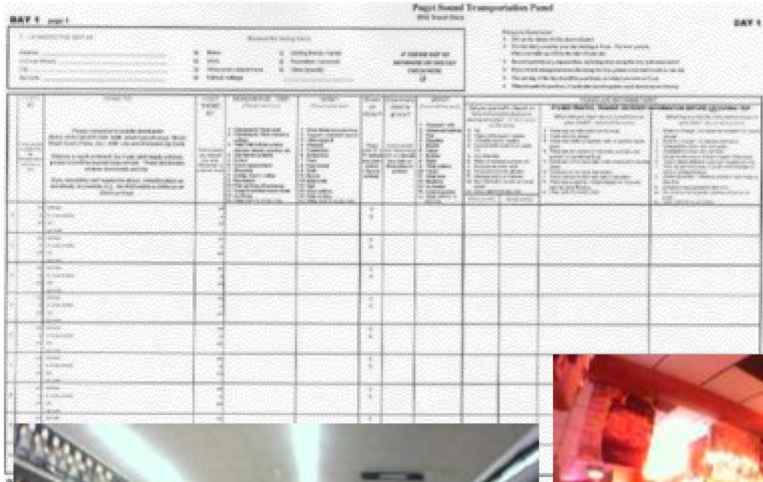
Population Level Health



The reality of Complex Spatial Health Data

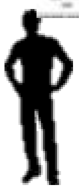
- Big data processing challenges w/ added dimensions
- Cross disciplinary communication
- HIPAA and security
- Who sees/gets the data?
- Modeling challenges
- Data organization/ infrastructure



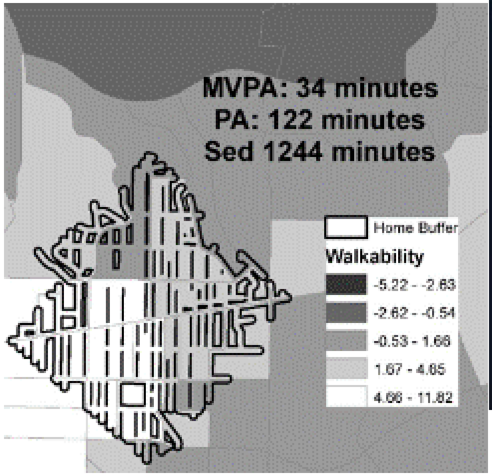


heart

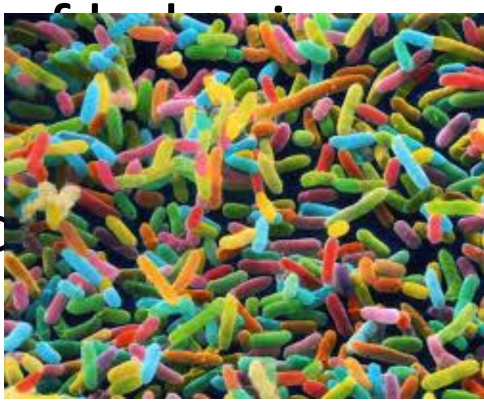
ables

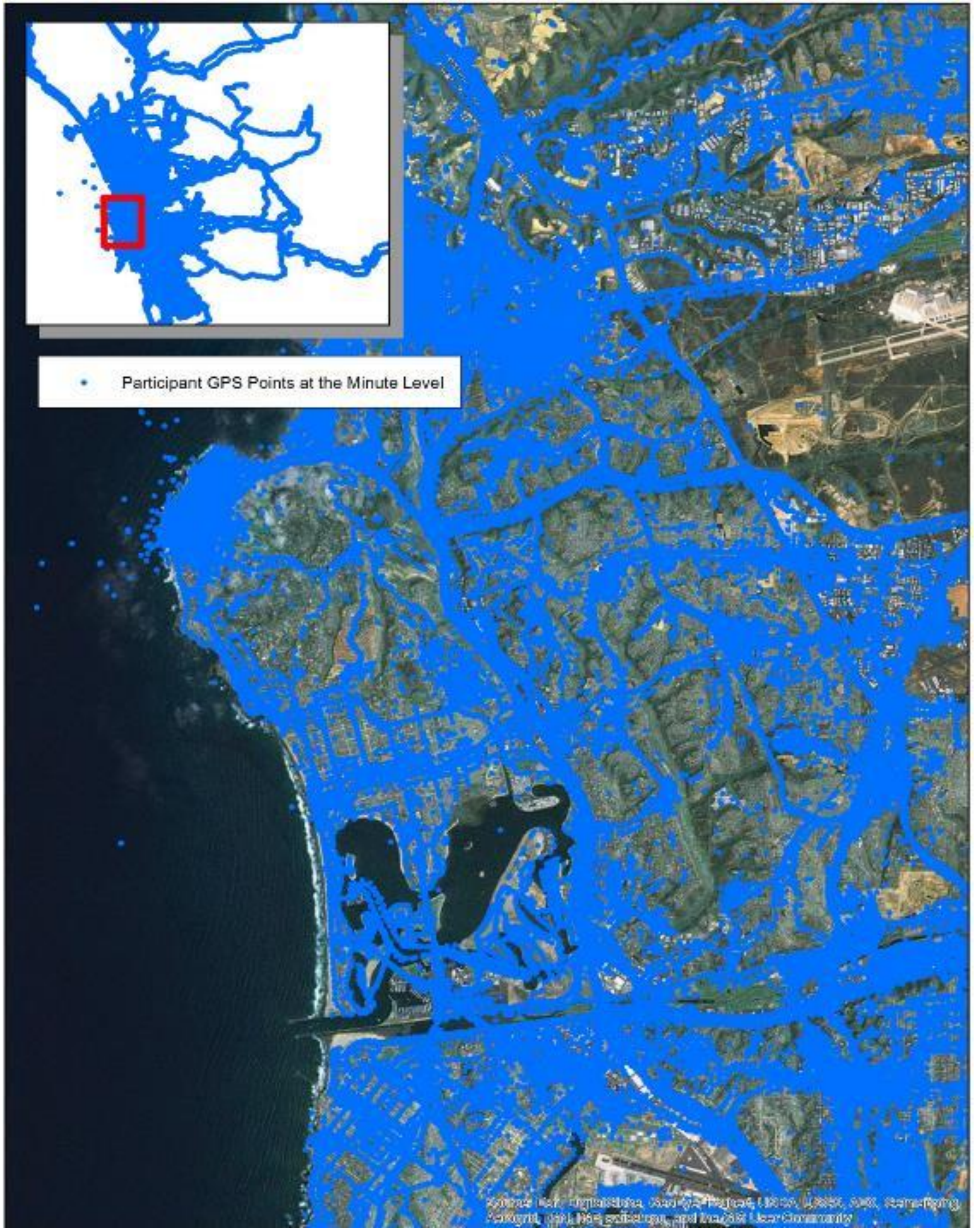


x 7 days x 10 hours



- Photography (SenseCam)
- longitudinal d
- Microbiome
- Genetic data

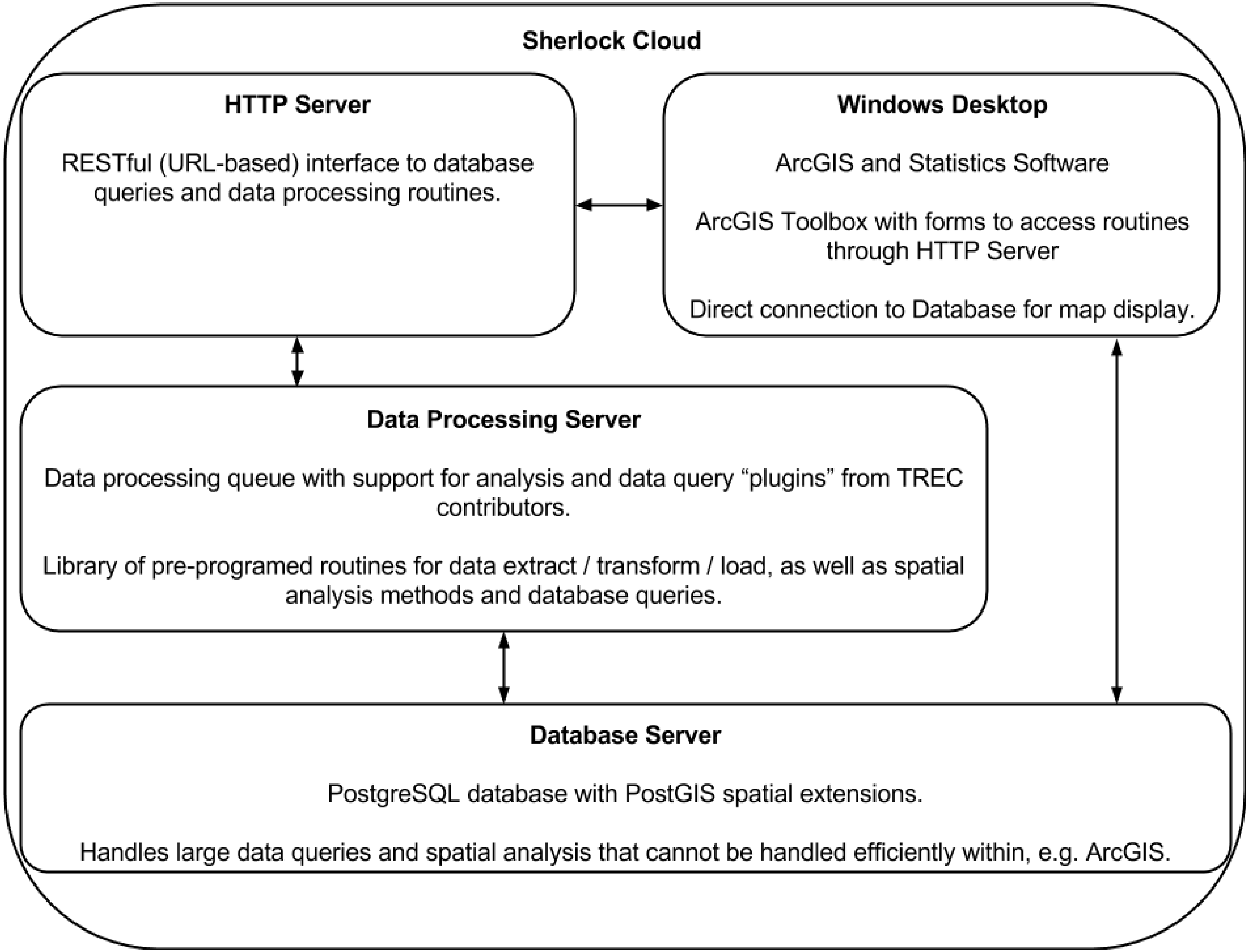


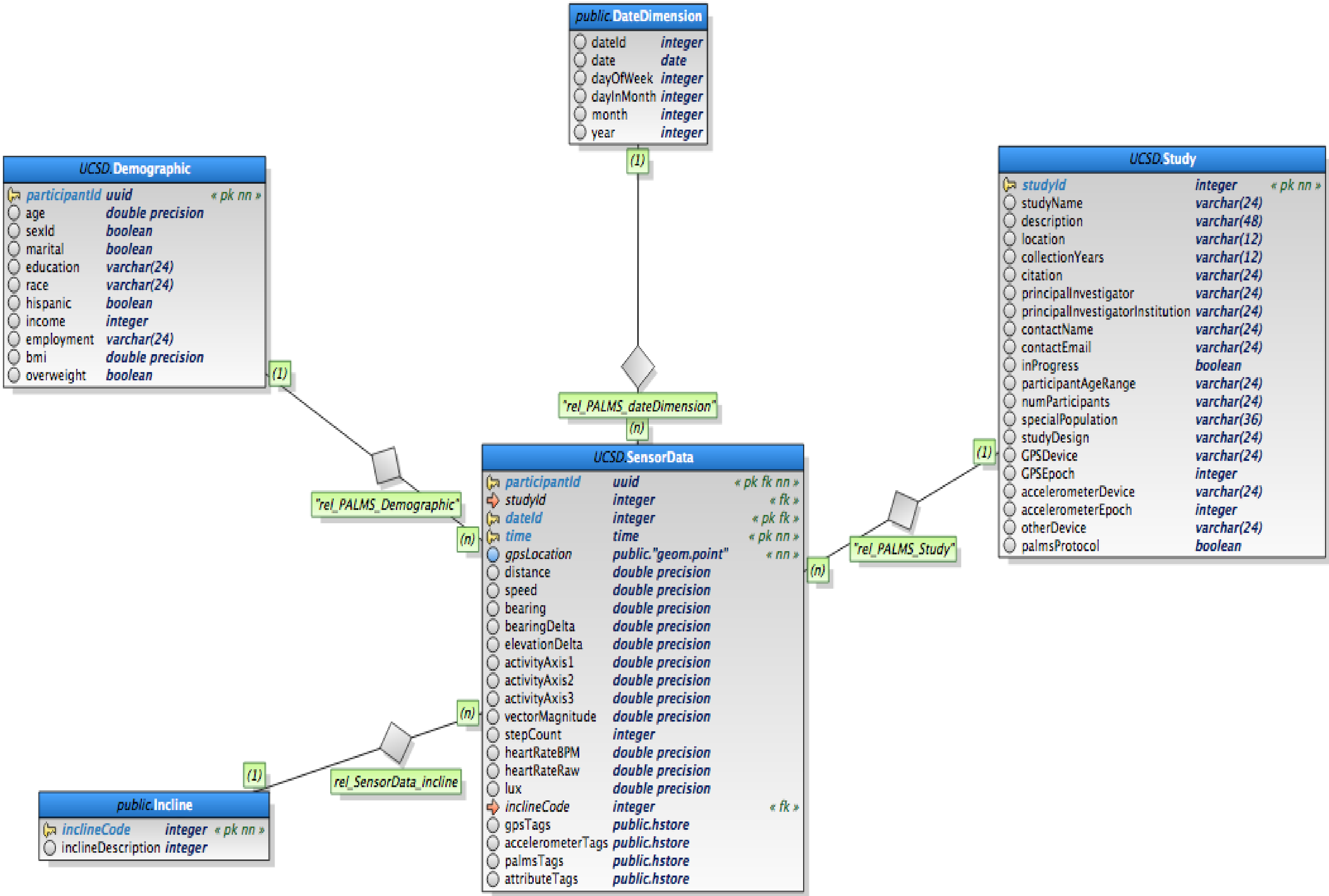

Participant GPS Points at the Minute Level

How Can CyberGIS help?

- CyberGIS as defined by ESRI: “GIS detached from the desktop and deployed on the web, with the associated issues of hardware software, data storage, digital networks, people, training and education.”
- Goldberg *et al.* Spatial-Health CyberGIS Marketplace
  - confidentiality and privacy protections
  - real-time analytic methods
  - data standardization
  - comprehensive end-to-end ecosystem architecture
- In addition:
  - need for shareable workflows to promote inter-field collaboration
  - diverse data type integration
  - replicability of analytic processes.

Step 1: Hosted HIPAA compliant Geodatabase




public.DateDimension
dateId integer
date date
dayOfWeek integer
dayInMonth integer
month integer
year integer
(1)
"rel_PALMS_dateDimension"
(n)
UCSD.Demographic
participantId uuid « pk nn »
age double precision
sexId boolean
marital boolean
education varchar(24)
race varchar(24)
hispanic boolean
income integer
employment varchar(24)
bmi double precision
overweight boolean
(1)
"rel_PALMS_Demographic"
(n)
UCSD.Study
studyId integer « pk nn »
studyName varchar(24)
description varchar(48)
location varchar(12)
collectionYears varchar(12)
citation varchar(24)
principalInvestigator varchar(24)
principalInvestigatorInstitution varchar(24)
contactName varchar(24)
contactEmail varchar(24)
inProgress boolean
participantAgeRange varchar(24)
numParticipants varchar(24)
specialPopulation varchar(36)
studyDesign varchar(24)
GPSDevice varchar(24)
GPSEpoch integer
accelerometerDevice varchar(24)
accelerometerEpoch integer
otherDevice varchar(24)
palmsProtocol boolean
(1)
"rel_PALMS_Study"
(n)
UCSD.SensorData
participantId uuid « pk fk nn »
studyId integer « fk »
dateId integer « pk fk »
time time « pk nn »
gpsLocation public."geom.point" « nn »
distance double precision
speed double precision
bearing double precision
bearingDelta double precision
elevationDelta double precision
activityAxis1 double precision
activityAxis2 double precision
activityAxis3 double precision
vectorMagnitude double precision
stepCount integer
heartRateBPM double precision
heartRateRaw double precision
lux double precision
inclineCode integer « fk »
gpsTags public.hstore
accelerometerTags public.hstore
palmsTags public.hstore
attributeTags public.hstore
(n)
rel_SensorData_incline
(1)
public.Incline
inclineCode integer « pk nn »
inclineDescription integer

Study	Description/Population	n	F, M, (%F)	Age(SD)	%Hispanic	%Employed (full-time)	Mean BMI	Valid(Acc,GPS, Both)	Home Location Y,N
Context	Overweight and obese adults 21-60 years old	71	55, 16, (77.4%)	42(10)	26, 45, (37.0%)	40, 31, (56.3%)	33(3)	448, 576, 389	67, 4
DIAL	N=40 women who did not meet the strict eligibility criteria of the MENU or RfH studies	37	37, 0, (100%)	57(15)	3, 34, (8.1%)	20, 17, (54.1%)	33(4)	276, 325, 270	37, 0
MIPARC	351 residents over the age of 65 living in Continuing Care Retirement Communities (CCRC) in SD County n=307 participants, n=44 peer leaders	347	247, 100, (71%)	84(6)	5, 332, (1.4%)	0, 347, (0.0%)	NA	7924, 9207, 7073	334, 13
RfH	6-month randomized controlled trial of metformin, lifestyle intervention, or both, among a sample of 340 postmenopausal, overweight/obese breast cancer survivors.	126	126, 0, (100%)	61(7)	17, 108, (13.6%)	44, 82, (34.9%)	31(6)	946, 1117, 852	126, 0
RfH Memory	Postmenopausal, normal weight cancer survivors	40	40, 0, (100%)	63(7)	5, 35, (11.1%)	10, 30, (23.3%)	22(2)	294, 325, 254	39, 1
SDSU Cycling	SDSU students, faculty, and staff who were cyclists	33	8, 25, (24%)	NA	1, 32, (3%)	27, 6, (81.8%)	NA	105, 201, 105	32, 1
Sensecam Cycling	N=40, healthy, working adults or students from the UCSD Commuter Cycling Network	40	12, 28, (30%)	36(12)	0, 40, (0%)	31, 9, (77.5%)	23(3)	103, 154, 101	40, 0
PALMS007	Latino population	42	20, 22, (47.6%)	27(11)	42, 0, (100%)	3, 39, (7.1%)	26(6)	260, 204, 169	41, 1
SAGE	N = 40 participants from the Stein SAGE study, selected to vary in physical functioning based on the SF-36 measure, with 10 participants from each of four decades 60-100.	40	16, 24, (40%)	78(10)	3, 37, (7.5%)	4, 36, (11.1%)	25(3)	231, 282,228	40, 0
Community of Mine	Adults living in a geographically diverse set of neighborhoods throughout San Diego County	700			50%				

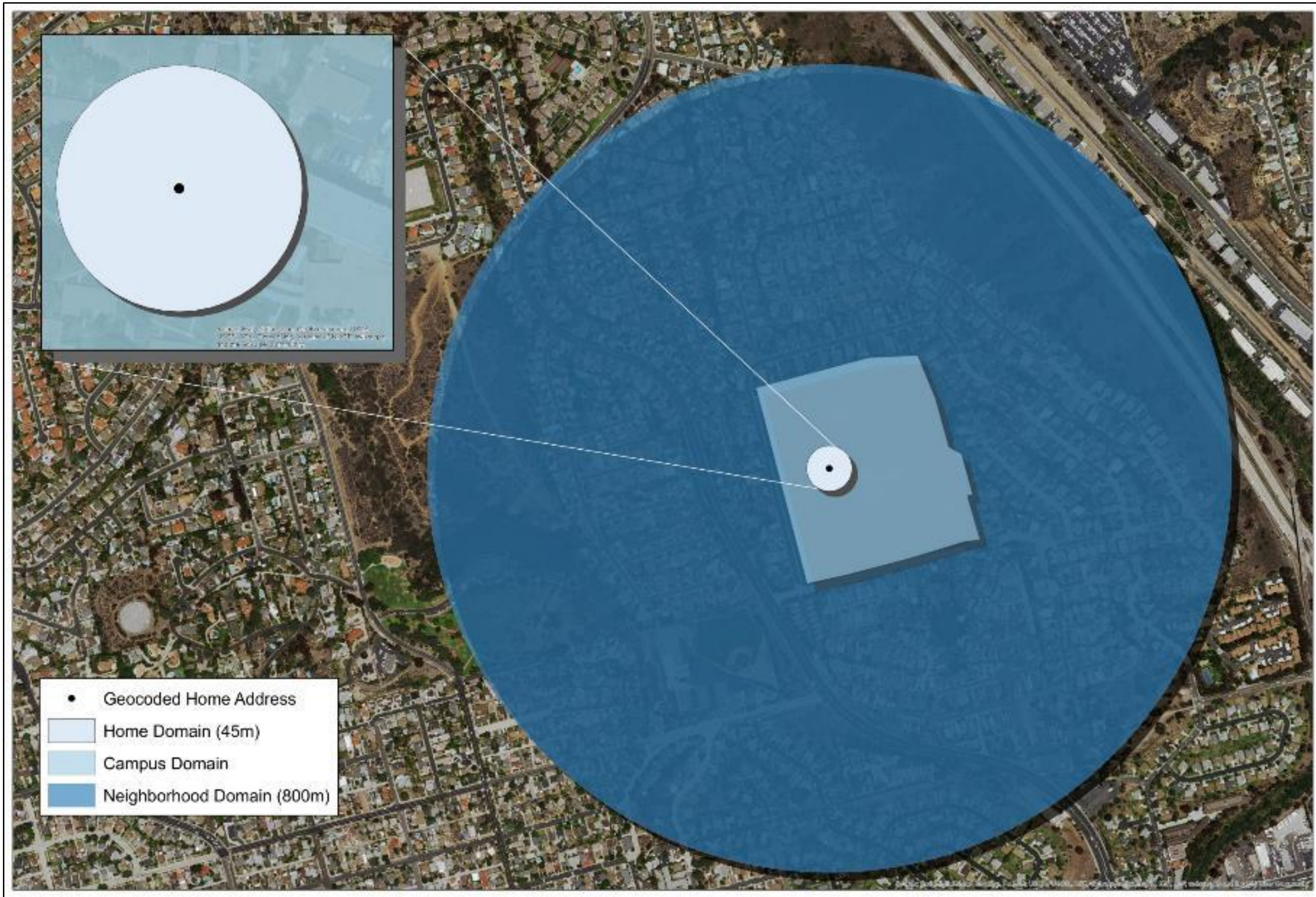
Change in Lifespace Over Time Within Retirement Communities: A Walking Intervention

Kristin Meseck, Marta Jankowska, Suneeta Godbole,
Jasper Schipperijn, Katie Crist, Michelle Black,
Loki Natarajan, Jacqueline Kerr

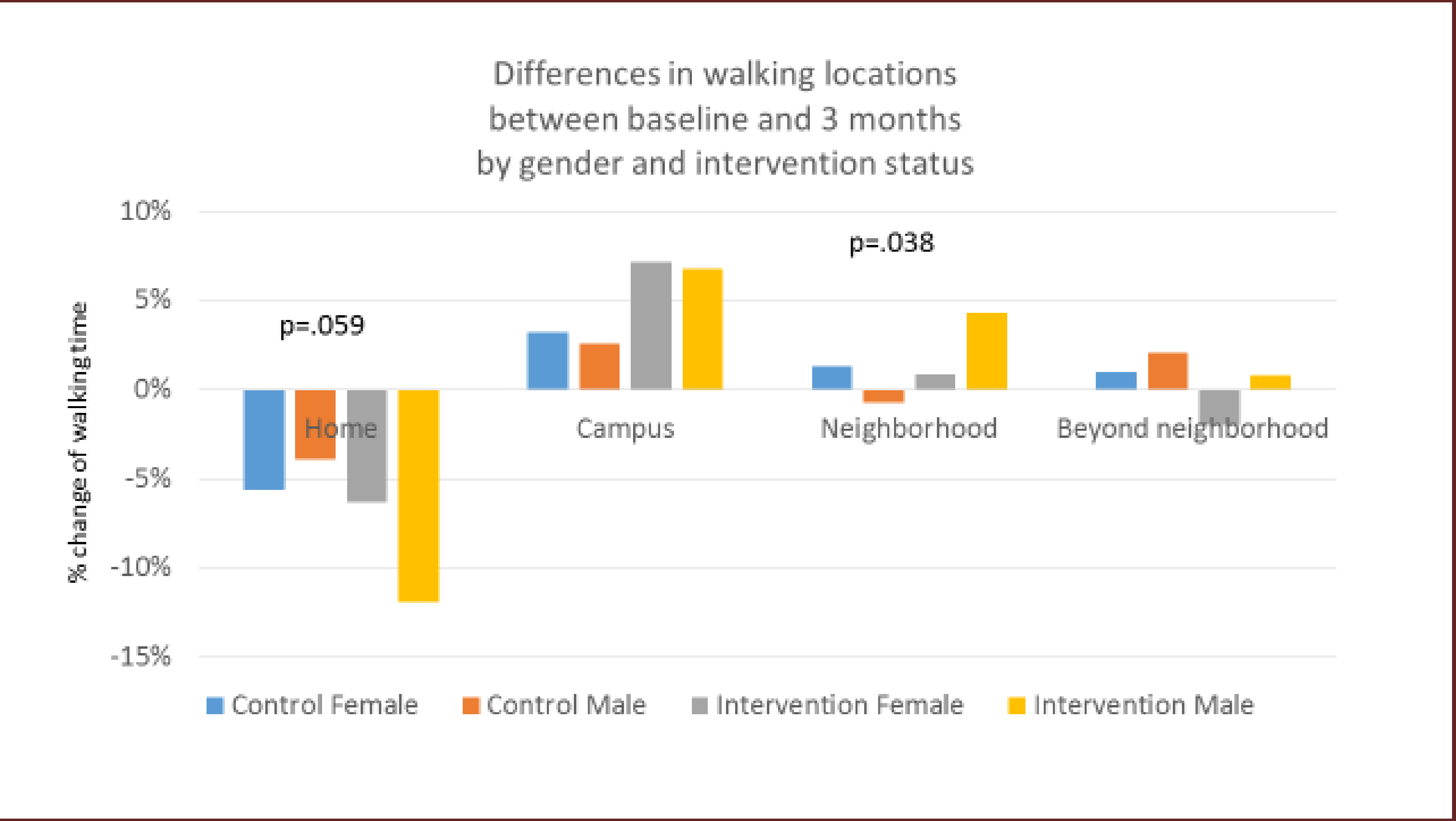
UC San Diego
SCHOOL OF MEDICINE

FAMILY MEDICINE & PUBLIC HEALTH
UC SAN DIEGO

Methods: Lifespaces



Results



Differences in walking locations
between baseline and 3 months
by gender and intervention status
% change of walking time
10%
5%
0%
-5%
-10%
-15%
p=.059
p=.038
Home
Campus
Neighborhood
Beyond neighborhood
Control Female
Control Male
Intervention Female
Intervention Male

Activity hotspots, one school, ≥ 5 min of bout time per recess period

Step 2: SPACES

Data and Analytical Considerations

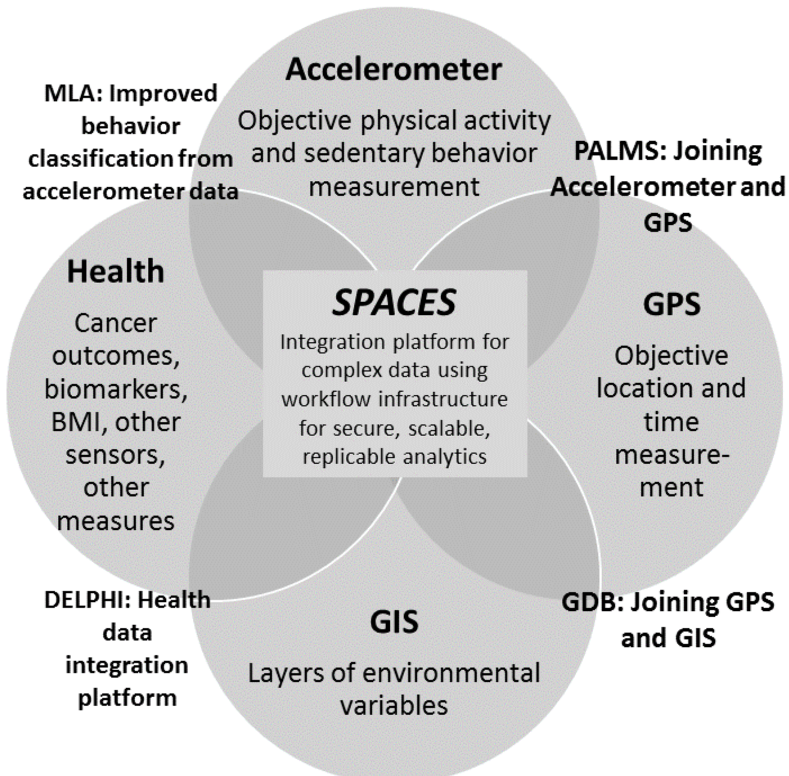
Validity and reliability
- Representation of behaviors
- Representation of spatial context

Compliance and technology
- GPS accuracy and treatment of missing data
- Participant compliance consistency

Spatial and temporal data
- Specific considerations for spatial and temporal data (high dimensional)
- GIS data considerations

Other data types
- Travel diaries
- Omics
- Life trajectories
- Sensors

Data Integration and Output: SPACES



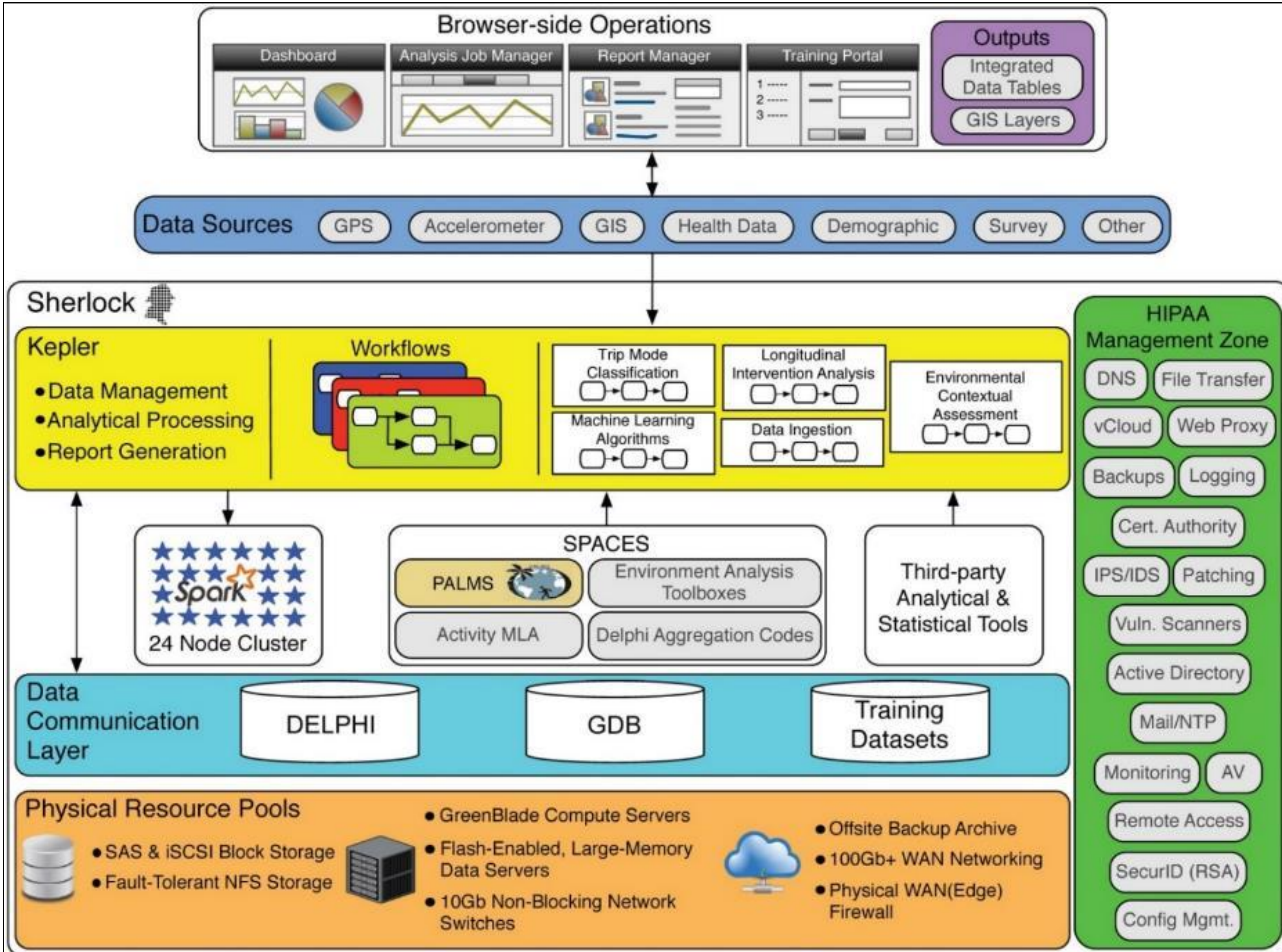
Analytic Workflow Examples: Kepler

Demographic parameters
- Testing PA cut points for older adults
- Comparing MLA to PALMS outputs for specific ethnic groups

Exposure assessment
- Identifying locations of interest
- Activity space analysis
- Creating kernel density weighted exposure measures

Behavioral assessment
- Activity bouts
- Machine learning behavior classification
- Behaviors in specific locations/environments

Time-based analysis
- Intervention outcomes
- Environmental change
- Lifecourse analysis



Browser-side Operations
Dashboard
Analysis Job Manager
Report Manager
Training Portal
Outputs
Integrated Data Tables
GIS Layers
Data Sources
GPS
Accelerometer
GIS
Health Data
Demographic
Survey
Other
Sherlock
Kepler
•Data Management
•Analytical Processing
•Report Generation
Workflows
Trip Mode Classification
Longitudinal Intervention Analysis
Environmental Contextual Assessment
Machine Learning Algorithms
Data Ingestion
Spark
24 Node Cluster
SPACES
PALMS
Environment Analysis Toolboxes
Activity MLA
Delphi Aggregation Codes
Third-party Analytical & Statistical Tools
Data Communication Layer
DELPHI
GDB
Training Datasets
Physical Resource Pools
• SAS & iSCSI Block Storage
• Fault-Tolerant NFS Storage
• GreenBlade Compute Servers
• Flash-Enabled, Large-Memory Data Servers
• 10Gb Non-Blocking Network Switches
• Offsite Backup Archive
• 100Gb+ WAN Networking
• Physical WAN(Edge) Firewall
HIPAA Management Zone
DNS
File Transfer
vCloud
Web Proxy
Backups
Logging
Cert. Authority
IPS/IDS
Patching
Vuln. Scanners
Active Directory
Mail/NTP
Monitoring
AV
Remote Access
SecurID (RSA)
Config Mgmt.

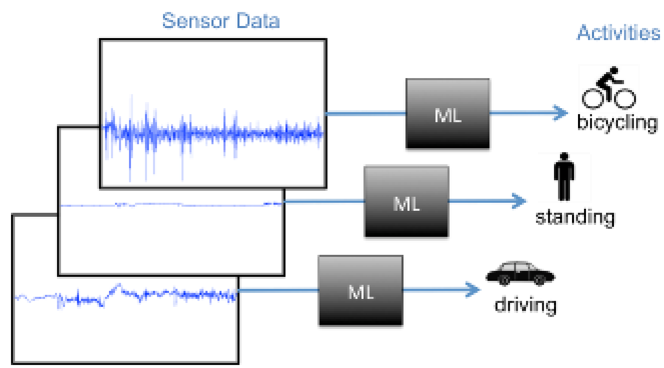
iWatch

CENTER FOR WIRELESS & POPULATION HEALTH SYSTEMS

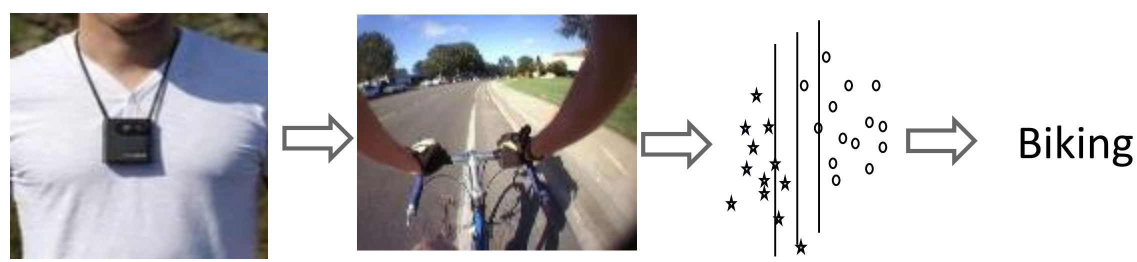
Validating machine-learned classifiers of sedentary behavior & physical activity

Purpose:

1. Validate machine-learned algorithms to classify patterns of accelerometer data to better discriminate types of sedentary behaviors and physical activity.



2. To develop machine learned algorithms to classify behaviors using images collected by the SenseCam (Computer Vision).

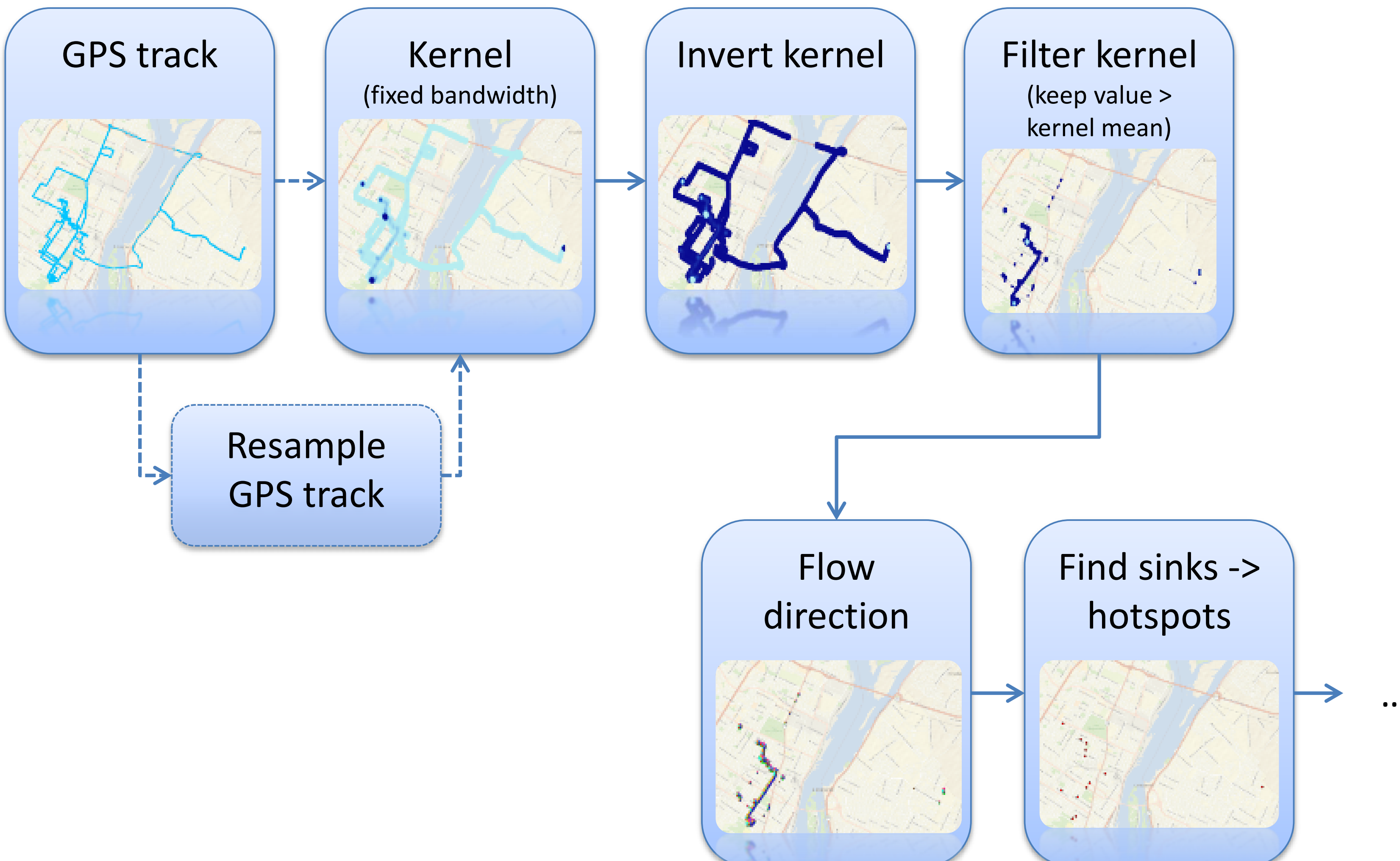


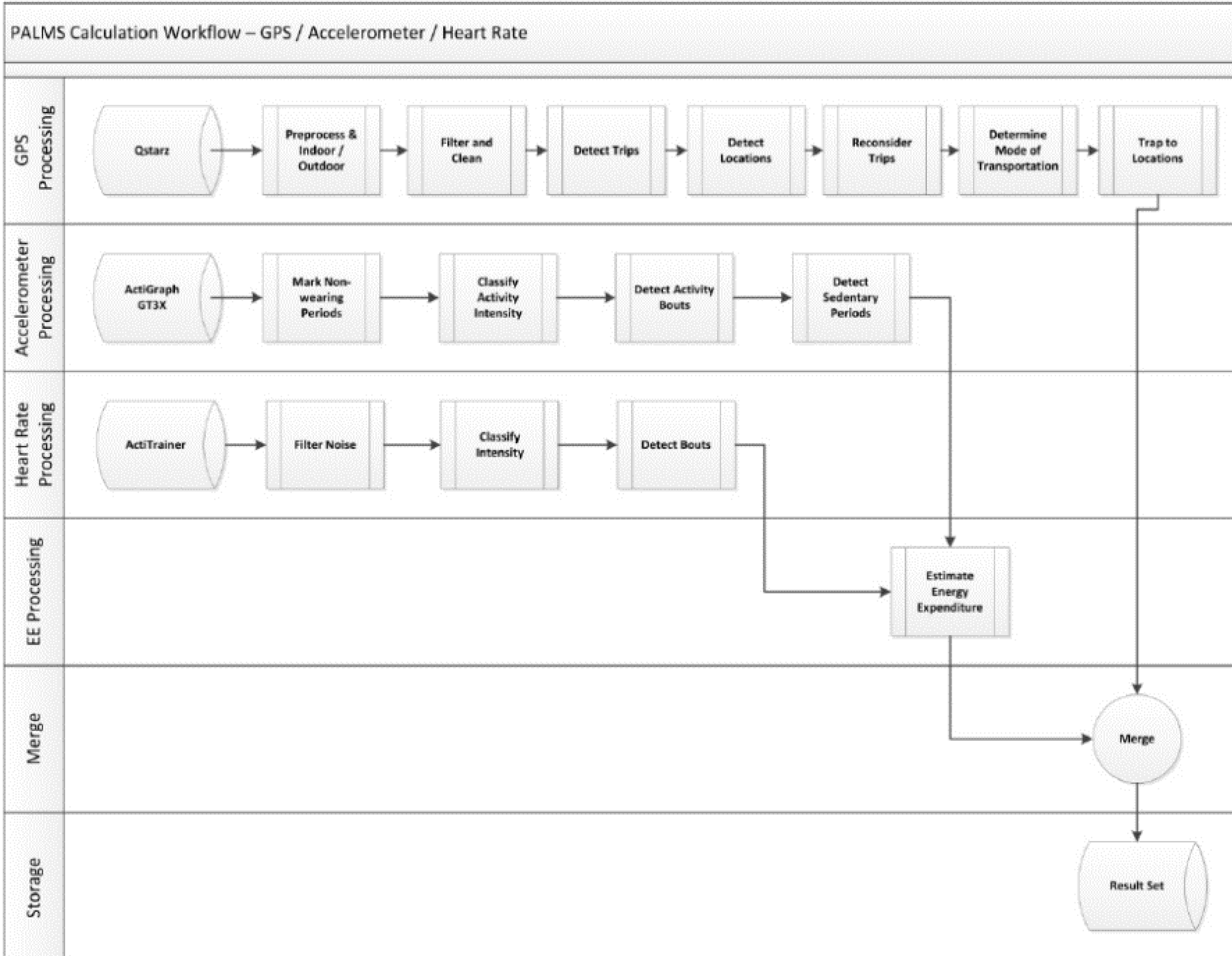
Jacqueline Kerr, Kevin Patrick, Jim Sallis, Simon Marshall, Loki Natarajan, Serge Belongie, Gert Lanckriet, Mohammad Moghimi, Katherine Ellis

Funded by NIH/NCI Grant 1 R01 CA164993-01

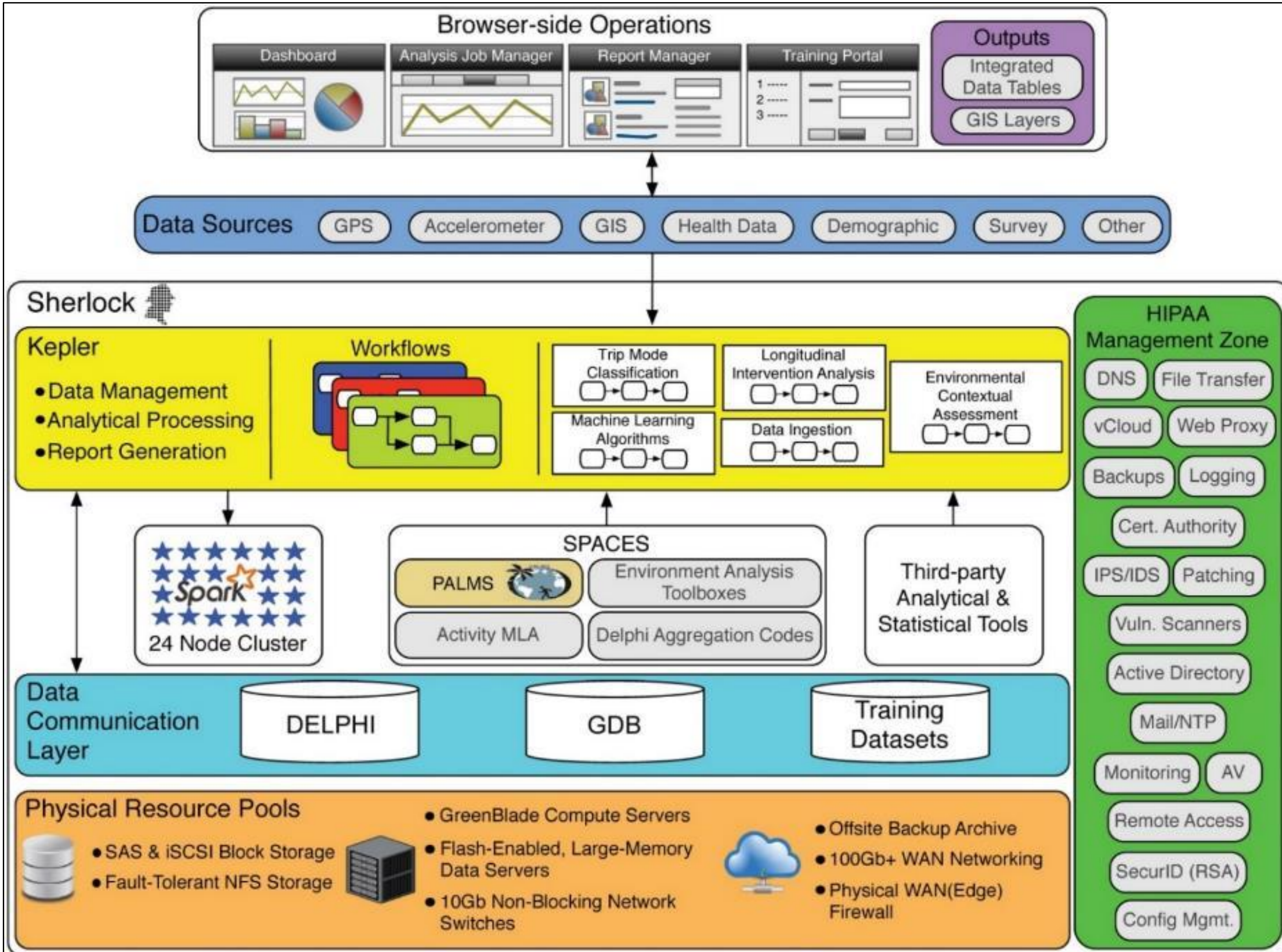
Hotspot detection algorithm (1)

Find hotspots





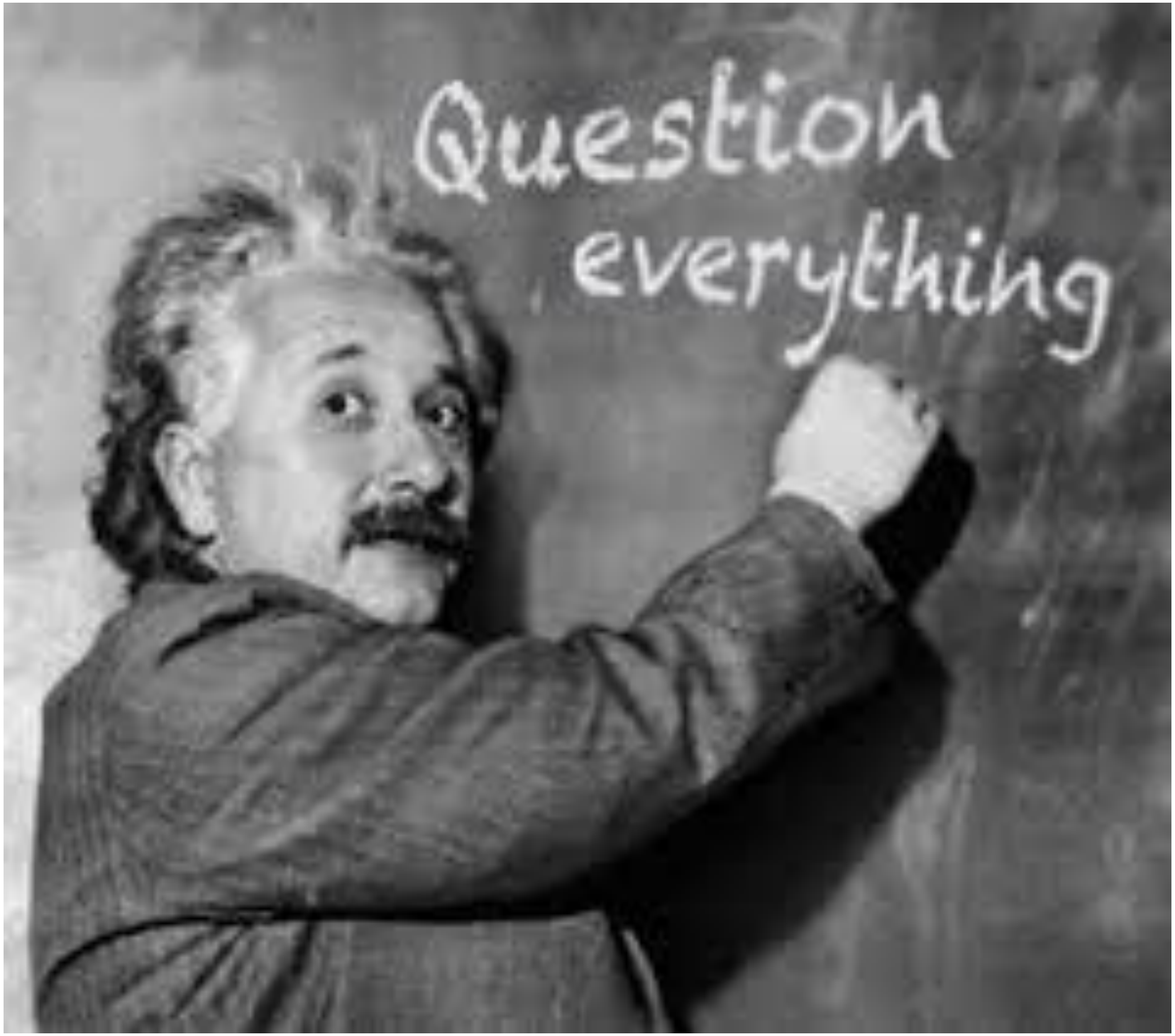
PALMS Calculation Workflow – GPS / Accelerometer / Heart Rate
GPS Processing
Qstarz
Preprocess & Indoor / Outdoor
Filter and Clean
Detect Trips
Detect Locations
Reconsider Trips
Determine Mode of Transportation
Trap to Locations
Accelerometer Processing
ActiGraph GT3X
Mark Non-wearing Periods
Classify Activity Intensity
Detect Activity Bouts
Detect Sedentary Periods
Heart Rate Processing
ActiTrainer
Filter Noise
Classify Intensity
Detect Bouts
EE Processing
Estimate Energy Expenditure
Merge
Merge
Storage
Result Set



Browser-side Operations
Dashboard
Analysis Job Manager
Report Manager
Training Portal
Outputs
Integrated Data Tables
GIS Layers
Data Sources
GPS
Accelerometer
GIS
Health Data
Demographic
Survey
Other
Sherlock
Kepler
•Data Management
•Analytical Processing
•Report Generation
Workflows
Trip Mode Classification
Longitudinal Intervention Analysis
Environmental Contextual Assessment
Machine Learning Algorithms
Data Ingestion
HIPAA Management Zone
DNS
File Transfer
vCloud
Web Proxy
Backups
Logging
Cert. Authority
IPS/IDS
Patching
Vuln. Scanners
Active Directory
Mail/NTP
Monitoring
AV
Remote Access
SecurID (RSA)
Config Mgmt.
Spark
24 Node Cluster
SPACES
PALMS
Environment Analysis Toolboxes
Activity MLA
Delphi Aggregation Codes
Third-party Analytical & Statistical Tools
Data Communication Layer
DELPHI
GDB
Training Datasets
Physical Resource Pools
• SAS & iSCSI Block Storage
• Fault-Tolerant NFS Storage
• GreenBlade Compute Servers
• Flash-Enabled, Large-Memory Data Servers
• 10Gb Non-Blocking Network Switches
• Offsite Backup Archive
• 100Gb+ WAN Networking
• Physical WAN(Edge) Firewall

SPACES goals

- Increase provenance in not only workflows and processing procedures, but also data formats and structures
- Provide a secure computing environment for sensitive data and studies
- Make CyberGIS and complex computer infrastructures more accessible to public health and behavioral researchers (not have to worry about ‘big’ data)
- Allow for collaboration between diverse disciplines to advance discovery and knowledge creation


Question
everything