Data Science Platforms for Applications with Societal Impacts

http://dsi.usc.edu/

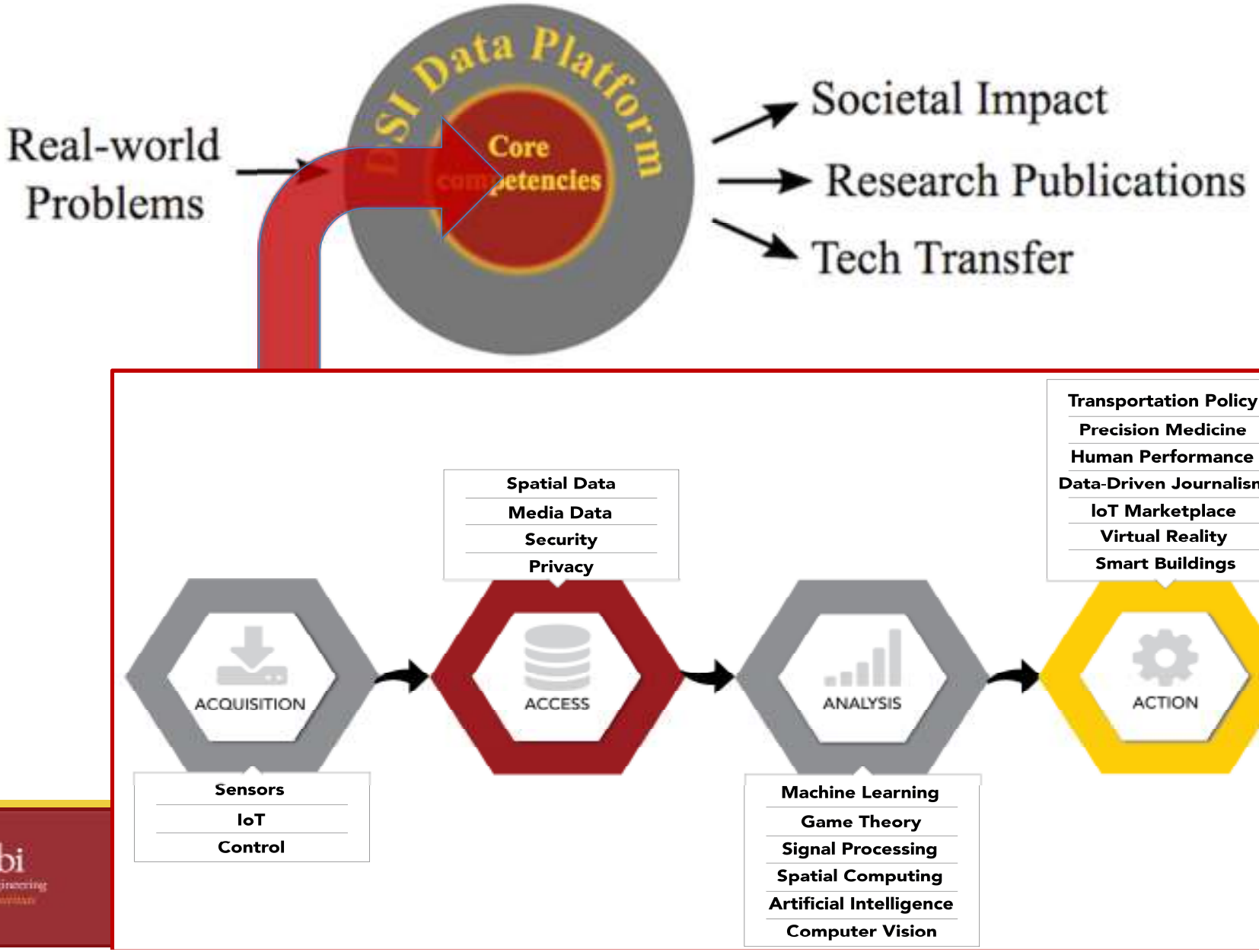
Cyrus Shahabi, Ph.D.
Professor of Computer Science, Electrical Engineering & Spatial Sciences
Chair, Department of Computer Science
Director, Data Science Institute (DSI)
Director, Informatics Program
Viterbi School of Engineering
University of Southern California
Los Angeles, CA 900890781
shahabi@usc.edu

OUTLINE

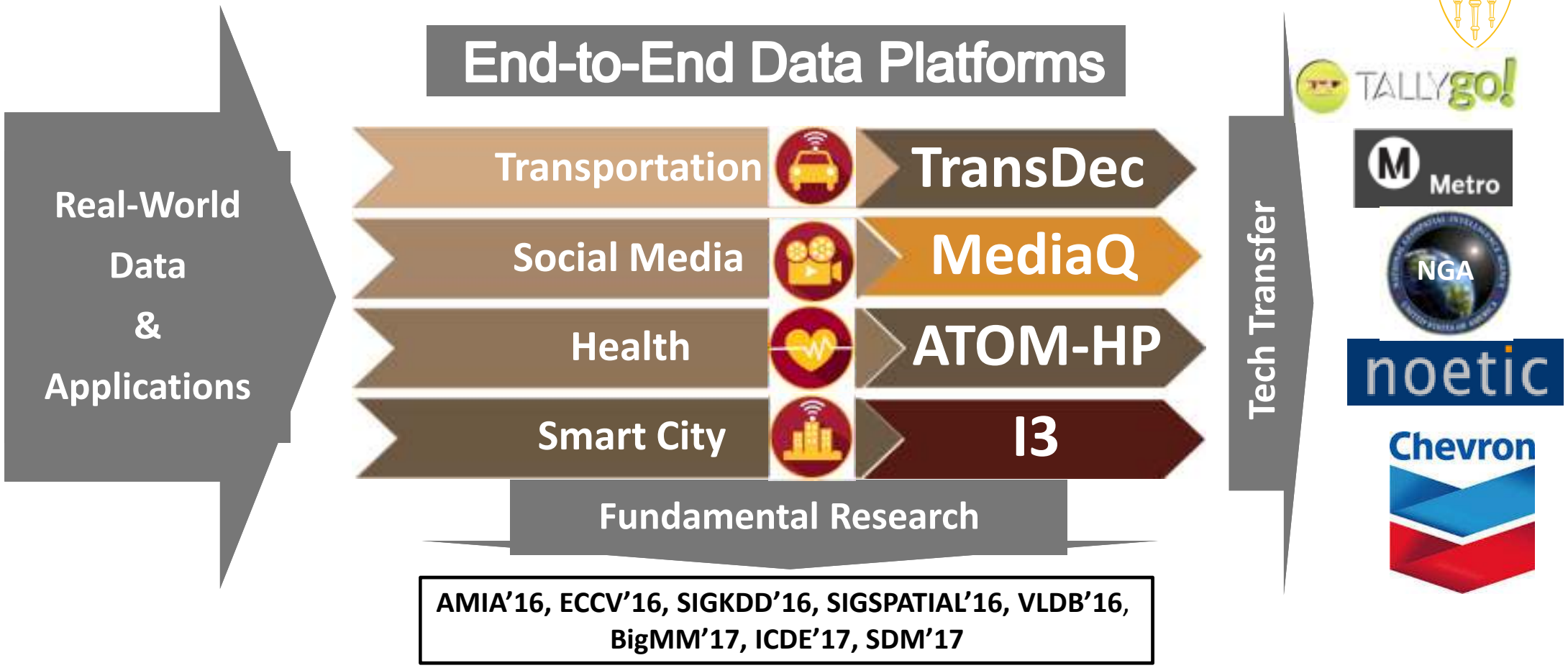
- DSI Overview
- Transportation Data Platform
- Social Media Data Platform
- Health Data Platform
- Smart City Data Platform
- Closing Remarks

OUTLINE

- DSI Overview
- Transportation Data Platform
- Social Media Data Platform
- Health Data Platform
- Smart City Data Platform
- Closing Remarks


Real-world Problems
SI Data Platform
Core competencies
Societal Impact
Research Publications
Tech Transfer
Transportation Policy
Precision Medicine
Human Performance
Data-Driven Journalism
IoT Marketplace
Virtual Reality
Smart Buildings
Spatial Data
Media Data
Security
Privacy
ACQUISITION
ACCESS
ANALYSIS
ACTION
Sensors
IoT
Control
Machine Learning
Game Theory
Signal Processing
Spatial Computing
Artificial Intelligence
Computer Vision

A Data Science Research Center



AMIA'16, ECCV'16, SIGKDD'16, SIGSPATIAL'16, VLDB'16, BigMM'17, ICDE'17, SDM'17

OUTLINE

- DSI Overview
- Transportation Data Platform
- Social Media Data Platform
- Health Data Platform
- Smart City Data Platform
- Closing Remarks

Transport-ation

ADMS: An Exclusive Contract w LA-Metro

Input Traffic Data | Data Processing | Storage | Analysis &Visualization

RIITS Los Angeles County

46 MB/min

StreamInsight

26 MB/min
11 TB/Year

ORACLE 11g

Spark
Lightning-Fast Cluster Computing

- Highway (4500+ sensors)
- Arterial (4700 + **9500 sensors**)
- Bus & Rail (2000+ buses)
- Event (~400 per day)
- Ramp meter
- CMS

Transit Ridership Data
4years of ~1M rows

Truck (WIM) Data
3 years of 10M rows

Inrix Probe Data
1 year of 400M rows

E.g., Traffic Forecasting (ICDM'13, KDD 16, SDM'17)

Transport-ation

ADMS Public Release

2011 ADMS RFP (Awarded to USC)	2011-2015 ADMS Developed (Research/Prototype by USC)	2015-2016 ADMS Extension (Awarded to USC)	2016-2021 ADMS Production (Awarded to Parsons/USC Tech Transfer of ADMS)

Metro

REVISED
PLANNING AND PROGRAMMING
APRIL 14, 2010

SUBJECT: CONTRACT NO. PS 4340-2301, ARCHIVE DATA MANAGEMENT SYSTEM (ADMS)

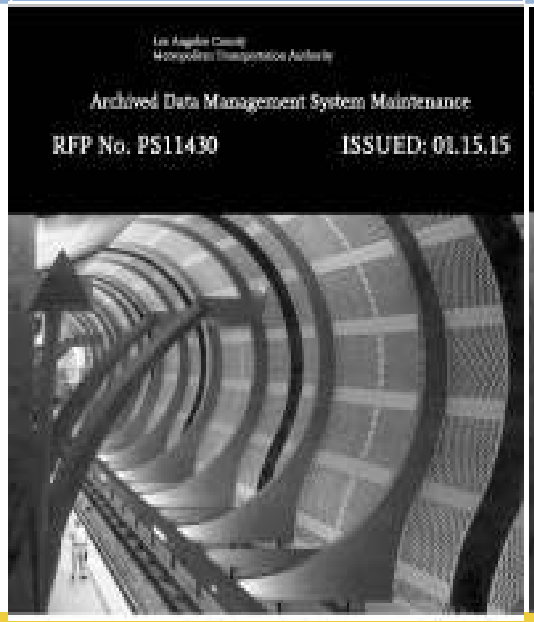
ACTION: AWARD A 3-YEAR FIRM FIXED PRICE CONTRACT TO METRANS TRANSPORTATION CENTER OF UNIVERSITY OF SOUTHERN CALIFORNIA

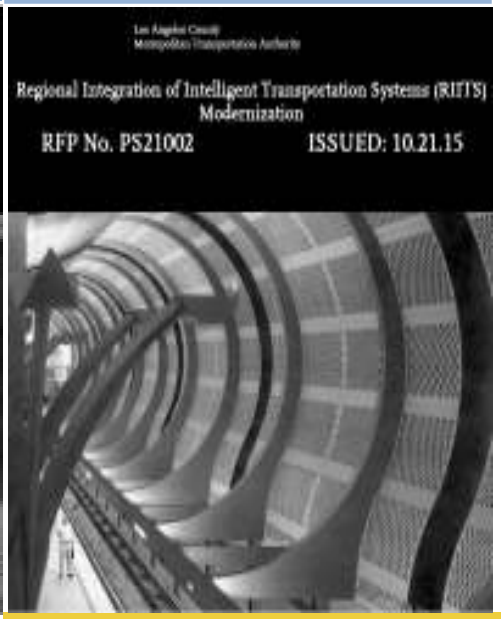
RECOMMENDATION

Authorize the Chief Executive Officer (CEO) to award a 3-year firm fixed price contract Contract No. PS 4340-2301, ADMS, to University of Southern California (USC) for Professional Services in an amount not to exceed $1,799,210, effective May 3, 2010

RATIONALE

Over the last four years, the Regional Integration of Intelligent Transportation System (RITS) network and program has been expanding to develop new interfaces with additional cities and transportation agencies such as Foothill Transit, Los Angeles Department of Transportation (LADOT)/Metro Rapid Bus, Los Angeles County Department of Public Works, California Highway Patrol, and Caltrans Districts 8 and 12. In addition, RITS continues to be the primary source of real-time traffic congestion data for Los Angeles County 511, Information Service Providers (ISPs), and third party data users for distribution of congestion data under separate formats for public consumption.

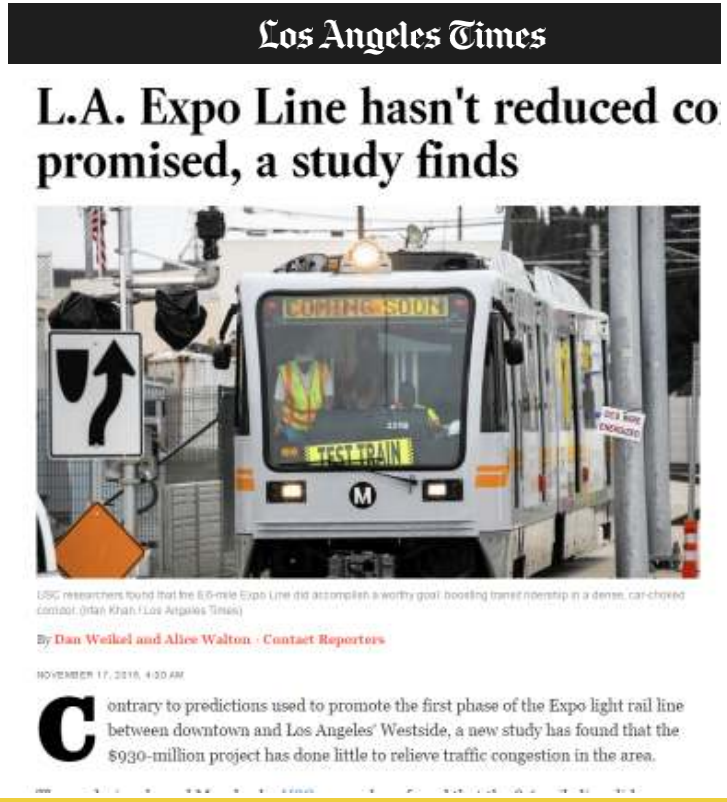




Policy- ADMS

- Collaboration between IMSC and Sol Price School of Public Policy

- Did Expo Line increase transit patronage?
- Did Expo Line impact traffic performance?
- Quasi-experimental design: Before/after and with/without



Los Angeles Times

L.A. Expo Line hasn't reduced co
promised, a study finds

By Dan Weikel and Alice Walton · Contact Reporters

Contrary to predictions used to promote the first phase of the Expo light rail line between downtown and Los Angeles' Westside, a new study has found that the $930-million project has done little to relieve traffic congestion in the area.

Transport-ation

Data Driven Journalism

http://www.nbclosangeles.com/news/local/USC-Freeway-LA-Traffic-Study_Los-Angeles-416848663.html

Transport-ation

Startup: TallyGo

- New business model (API)
  - LAFD Deployment
- Target is **Series-A** funding in 2017

United States Patent and Trademark Office
An Agency of the Department of Commerce

US Patent No. 9,286,793
Traffic prediction using real-world transportation data
March 15, 2016

US Patent No. 8,660,789
Hierarchical & exact fastest path computation in time-dependent spatial networks
February 2014

US Patent No. 8,566,030
Efficient K-nearest neighbor search in time-dependent spatial networks
October 2013

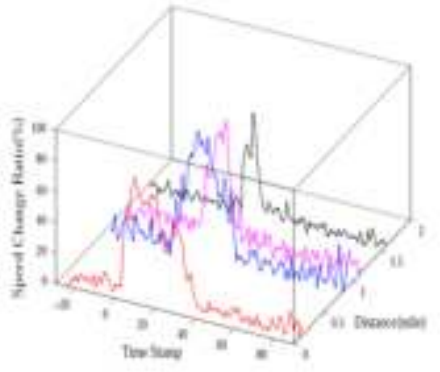
Transport-ation

Research: Traffic Forecasting

Single sensor

Time series analysis

ICDM'2012



Multi sensor

Latent Space -- *SIGKDD'2016*

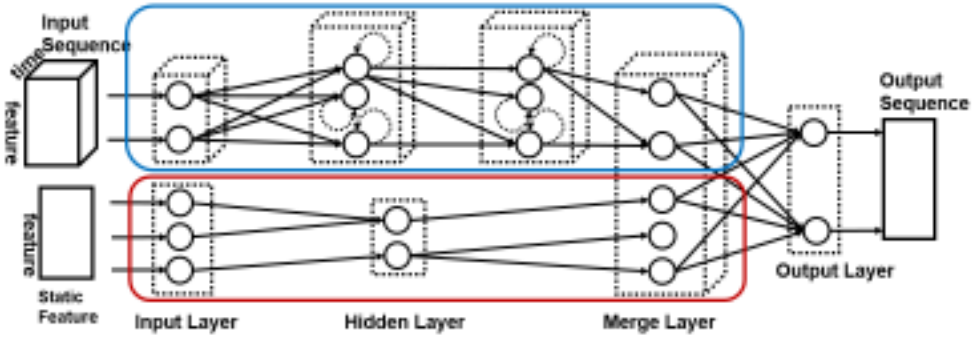
Graph matrix: G^{nxn} Latent properties: U^{nxk} and B^{kxk}

Single sensor

Causality

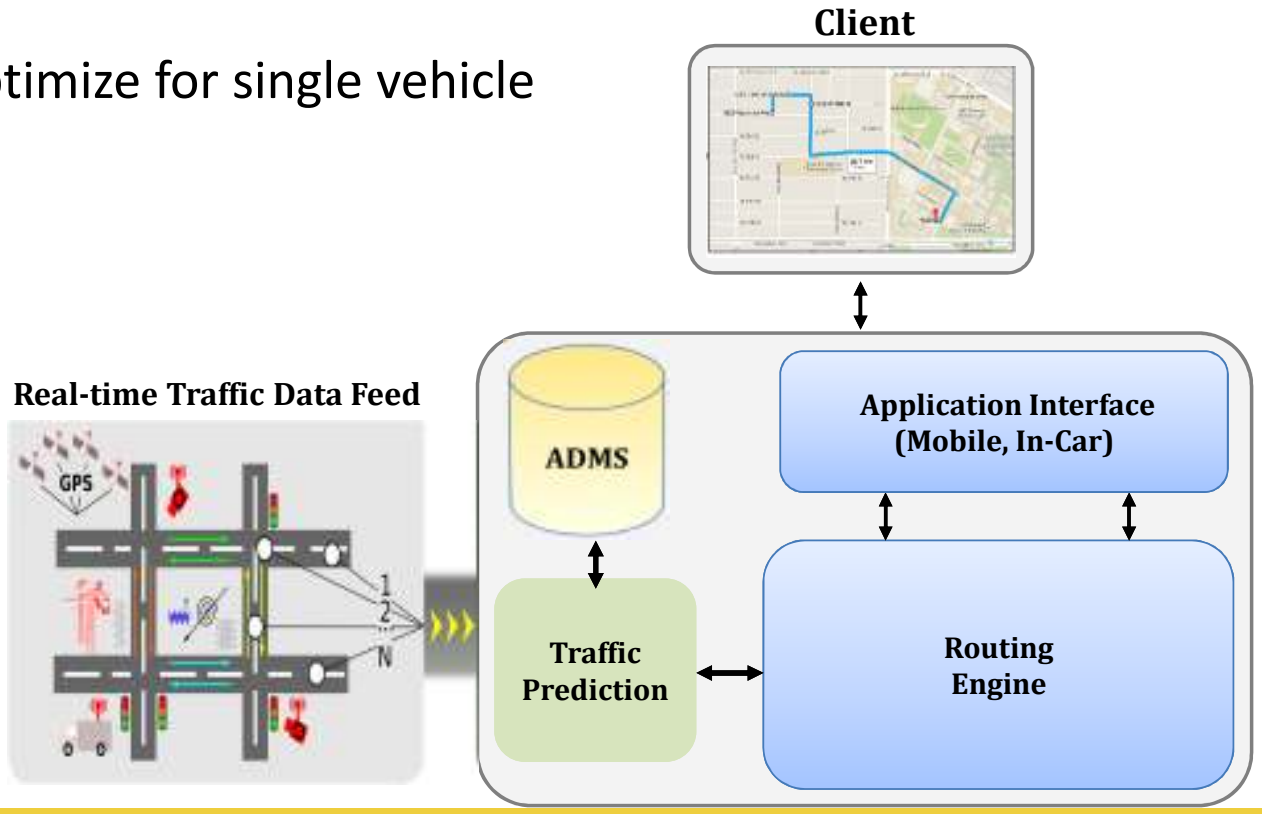
ICDM'2013

Multi sensor Deep Learning *SDM'2017*

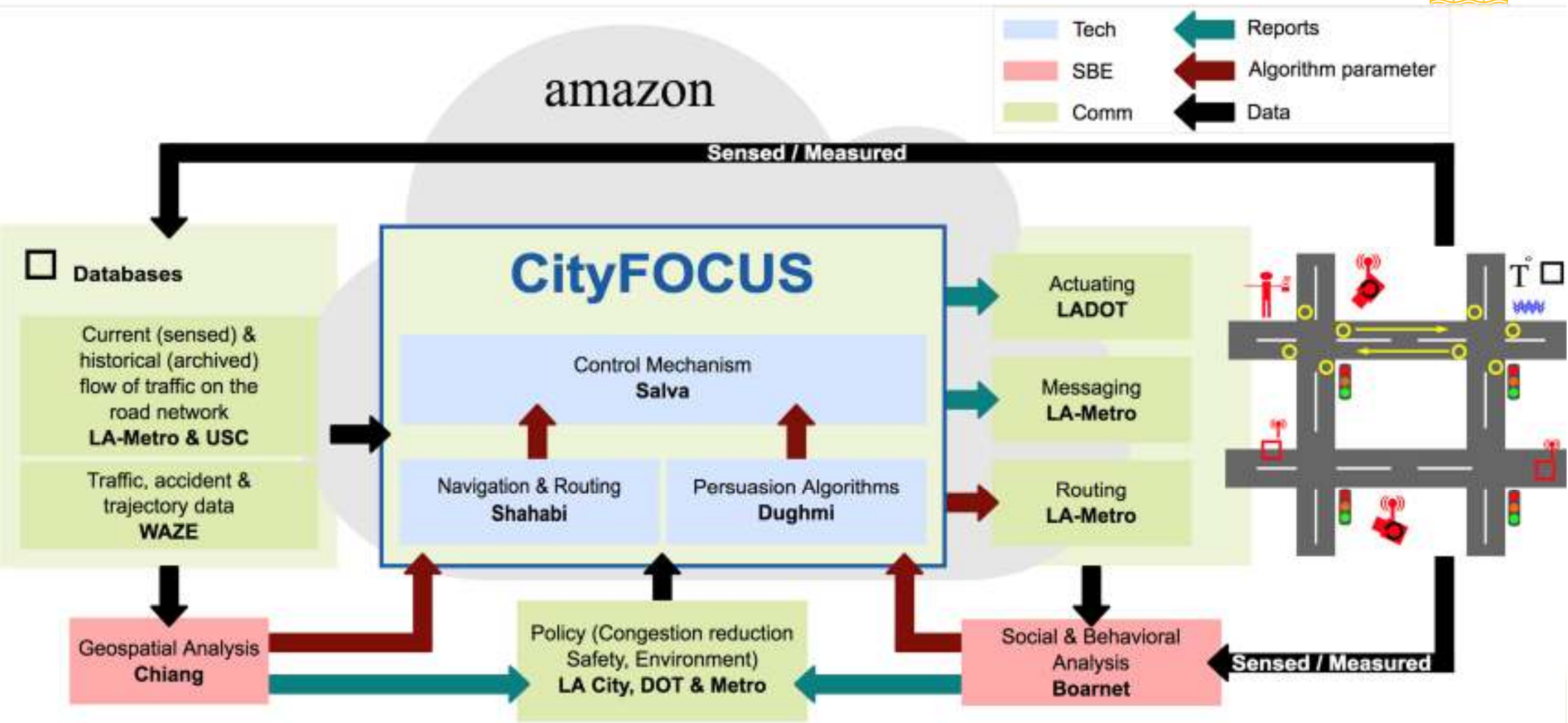


Open Problem

- **Current:** Optimize for single vehicle



Future: City Flow Optimization & Control Utilization System



OUTLINE

- DSI Overview
- Transportation Data Platform
- Social Media Data Platform
- Health Data Platform
- Smart City Data Platform
- Closing Remarks

Mobile Progress

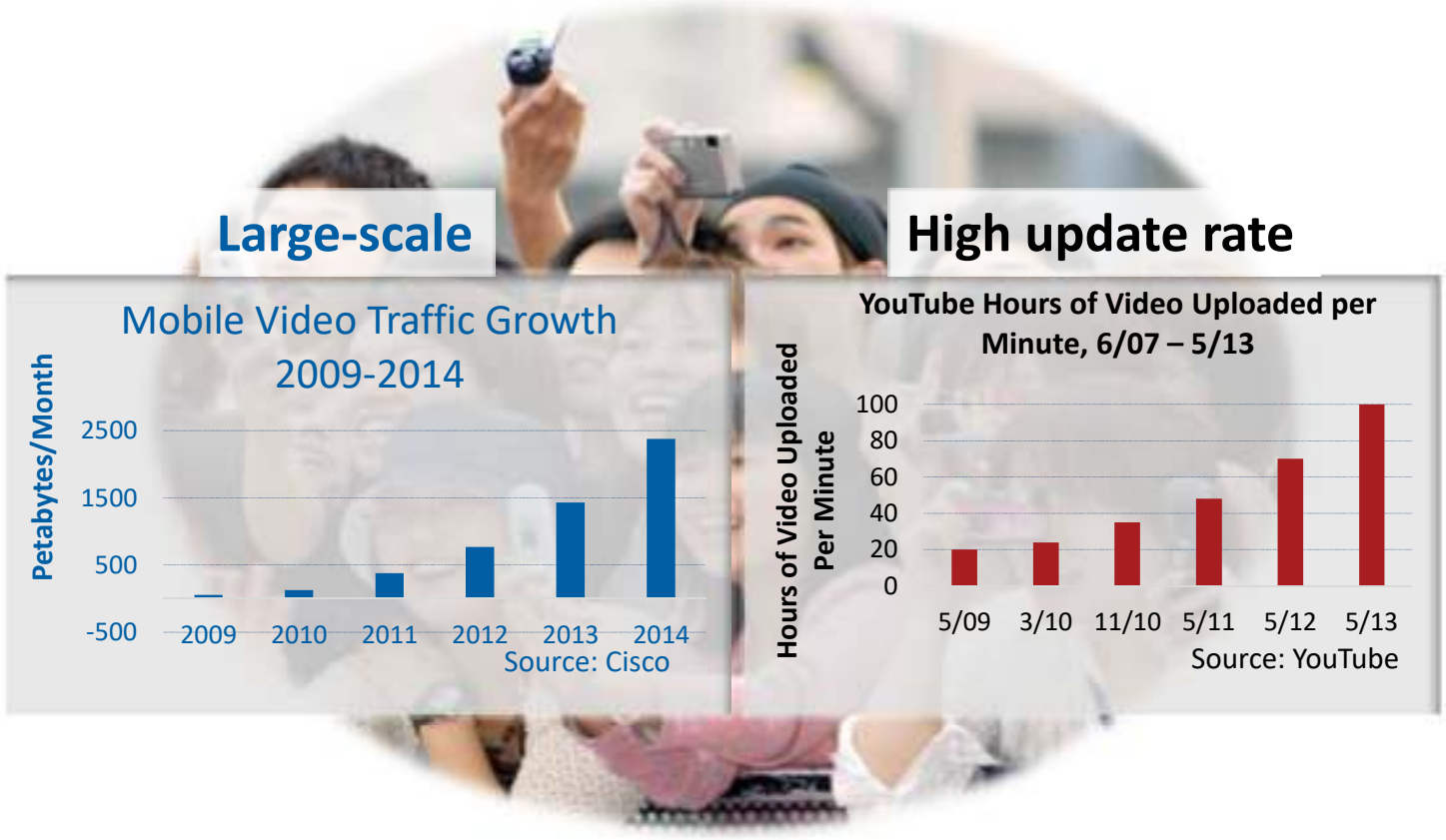
Ubiquity of mobile users — **6.5 billion** mobile subscriptions, 93.5% of the world population [1]

Technology advances on mobiles — Smartphone's **sensors**. e.g., video cameras

Network bandwidth improvements — From 2.5G (up to 384Kbps) to 3G (up to 14.7Mbps) and recently **4G** (up to 100 Mbps)

[1] http://mo

User-Generated Videos (UGVs)

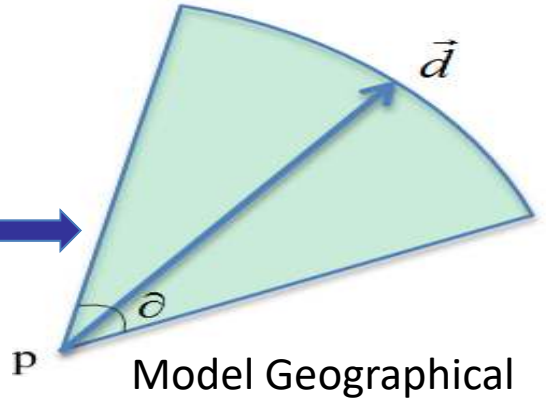


UGV and its Spatiotemporal Metadata

Metadata from Sensors

Record video using Camera with sensors (Mobile Apps)

GPS, Compass, Clock



Model Geographical Coverage of Video Scenes

p : camera location

$\vec{d}$: camera direction vector

∂ : viewable angle

t : timestamp

A. S. Ay, R. Zimmermann, and S. H. Kim. Viewable Scene Modeling for Geospatial Video Search. In ACM Intl. Conf. on MM, pages 309–318, 2008.

FOV Queries

Problem of **UGV** search → Problem of **FOV** search

Spatial queries on UGVs

- Range queries
  - *E.g., search videos overlapping with an area at USC.*
- Directional queries
  - *E.g., search videos directed towards the North.*

Range query in MediaQ
[Kim et al. MMSys14]

MediaQ Demo

http://mediaq.usc.edu/

National Science Foundation
WHERE DISCOVERIES BEGIN

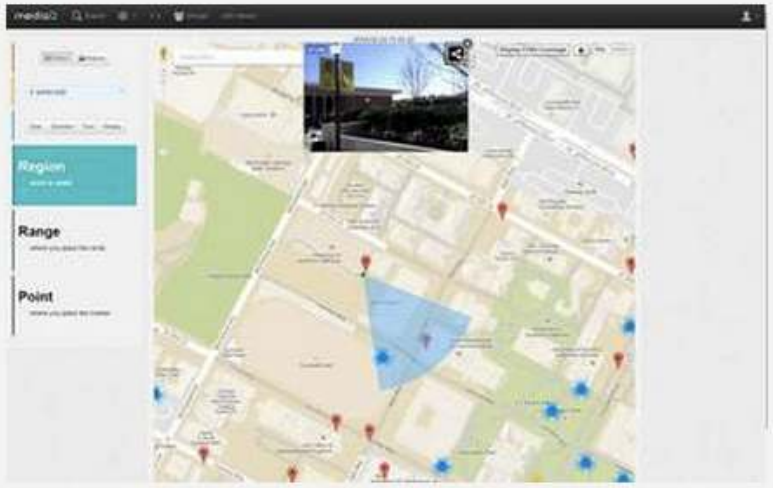
Application with Societal Impact

Disaster Response

Press Release 15-029

New U.S.-Japan collaborations bring Big Data approaches to disaster response

NSF and the Japan Science and Technology Agency announce joint support for 6 projects to improve future disaster management



USC's spatial crowdsourcing platform, MediaQ, collects pictures and videos during disasters.
Credit and Larger Version

March 30, 2015

When disaster strikes, it is critical that experts, decision makers and emergency personnel have access to real-time information in

GeoQ – NGA's Disaster Response Platform

GeoQ on GitHub - What we have accomplished so far

Launched April 4th 2014.

Open Engineering for Data, Applications, Interfaces, Measurement and more

Transparency: Engaging all levels of government, industry, academia, and others through understanding, ideas, insight and lessons learned.

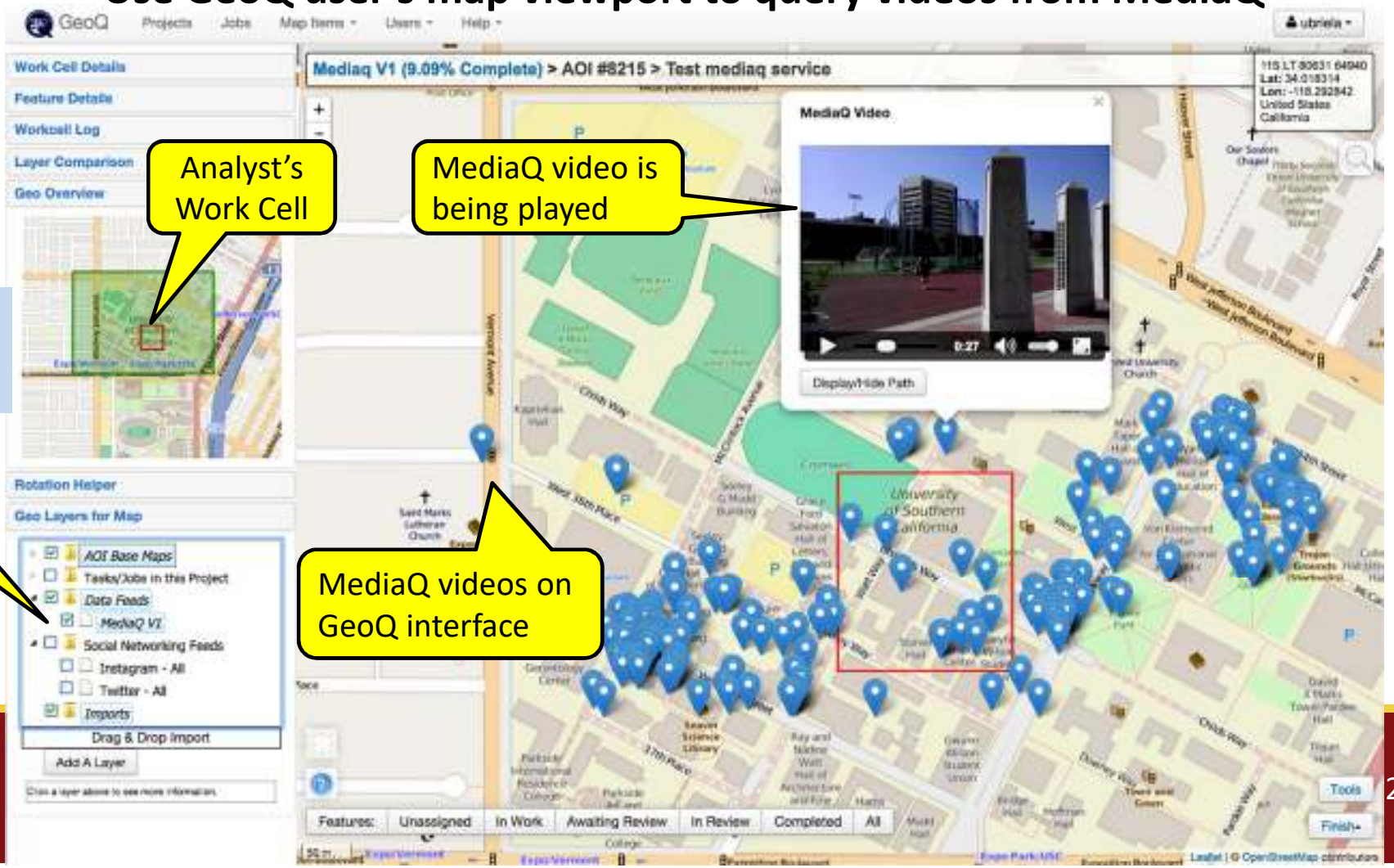
- **YouTube Video**
- Crisis Support GitHub Repositories: GeoQ, GeoEvents, RFI Generator, Gamification Server, 3 - Chef Installers
- **In The News**, CNN Video, White House Briefing, Forbes, Washington Post, World Bank, Humanitarian Open Street Map Team (HOT) Summit 2015, GEOINT Symposium 2014 GeoEnergy Summit, Top 30 Finalists for Igniting Innovation 2014, FCW Federal 100 Award 2016, 2015 Next Gov Bold Award 2015 and Peoples Choice and over 75 other articles
- GeoQ is in the top 10 government repositories on GitHub
- GeoQ Telecon
- **GeoQ GitHub Map**
- Briefed to POTUS (not bad for no funding)

Social Media

TechTransfer: MediaQ ➔ NGA's GeoQ

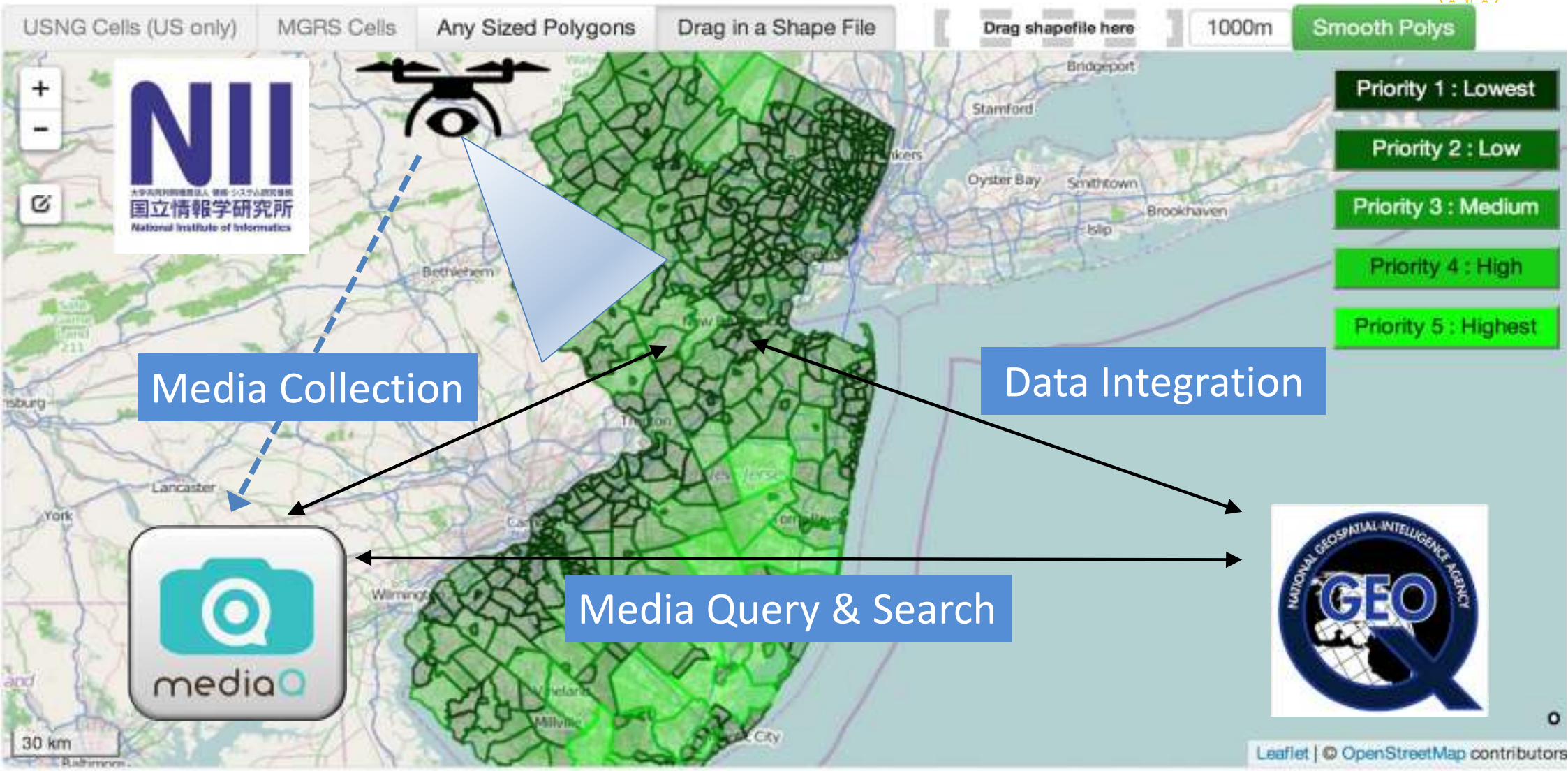
NGA GeoQ – USC MediaQ Integration:
Use GeoQ user's map viewport to query videos from MediaQ

IMSC provides APIs for integration



Social Media

Data Collection and Analysis in Disaster



OUTLINE

- DSI Overview
- Transportation Data Platform
- Social Media Data Platform
- Health Data Platform
- Smart City Data Platform
- Closing Remarks

Health

Analytical Technologies to Objectively Measure Human Performance (ATOM-HP)

Goal Evaluation of Human Performance in Cancer Patients

Peter Kuhn
Professor, Biological Sciences
The Bridge Institute at USC

USC Dornsife

Jorge Nieva, MD
Associate Professor of Clinical Medicine
USC Norris Comprehensive Cancer Center and Hospital

USC Norris Comprehensive Cancer Center

Joan E. BroderickSenior Behavioral Scientist; Associate Director, Center for Self-Report Science Center for Economic & Social Research

Paul Newton
Mathematics and Modeling

Cyrus Shahabi
Professor of Computer Science & Electrical Engineering

Sanjay Purushotham
IMSC

Luciano Nocera
IMSC

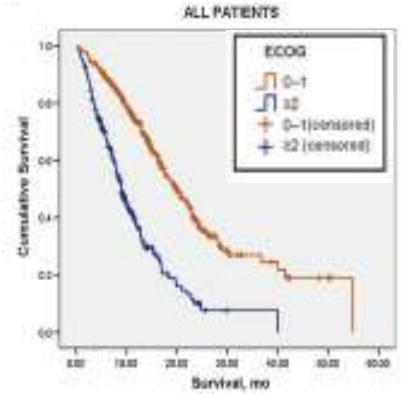
ATOM-HP: Human Performance Optimization
PSOC-0081

USC University of Southern California
RAPID REACTION Technology Office
noetic

Performance Status Scale

+ Performance Status remains best predictor of patient survival in patients with metastatic cancer: better than genomics, blood based biomarkers, imaging
− Evaluation limited to observations during visits

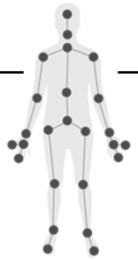
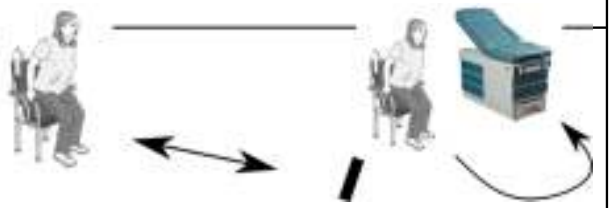
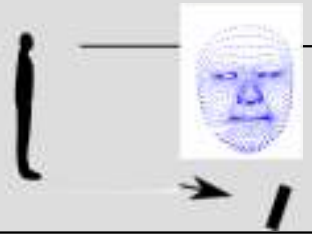
ECOG Performance Status Scale	
Grade	**Description**
0	Normal activity. Fully active, able to carry on all pre-disease performance without restriction.
1	Symptoms but ambulatory. Restricted in physically strenuous activity, but ambulatory and able to carry out work of a light or sedentary nature (*e.g.*, light housework, office work).
2	In bed <50% of the time. Ambulatory and capable of all self-care, but unable to carry out any work activities. Up and about more than 50% of waking hours.
3	In bed >50% of the time. Capable of only limited self-care, confined to bed or chair more than 50% of waking hours.
4	100% bedridden. Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair.
5	Dead.



Assessor pair	KPS	ECOG
Consultant–RMO	25 (63)	38 (92)
Consultant–nurse	33 (68)	32 (90)
Consultant–patient	23 (67)	38 (89)
RMO–nurse	33 (76)	51 (90)
RMO–patient	31 (68)	42 (86)
Nurse–patient	28 (74)	45 (93)

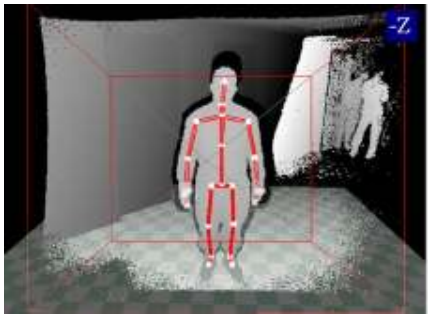
Health

ATOM-HP: Body Sensing

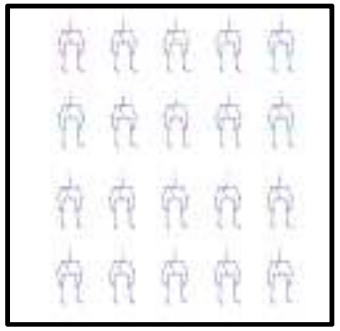
	Sensors	Data	Task
Clinical	Band 	Calories, Step Count, Heart Rate (mean, peak, min)	In the field: 60 days 8AM-8PM
	Kinect 	Raw files -> skeleton data 	Clinic: 1. Chair to Table 2. Get-Up and Go 
Military	Band 	Calories, Step Count, Heart Rate (mean, peak, min)	In the field: 5 days - all day
	Kinect 	Raw files -> skeleton, face mesh, face parameters (e.g., eye open, engaged)	Controlled environment Walk and Talk 

Research: Integrated Micro & Macro Data Analysis

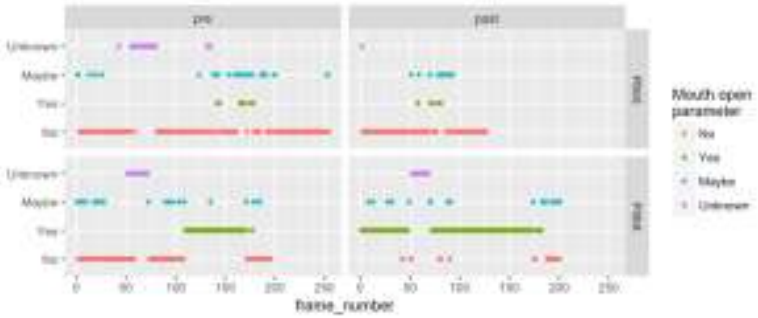
Skeleton Data

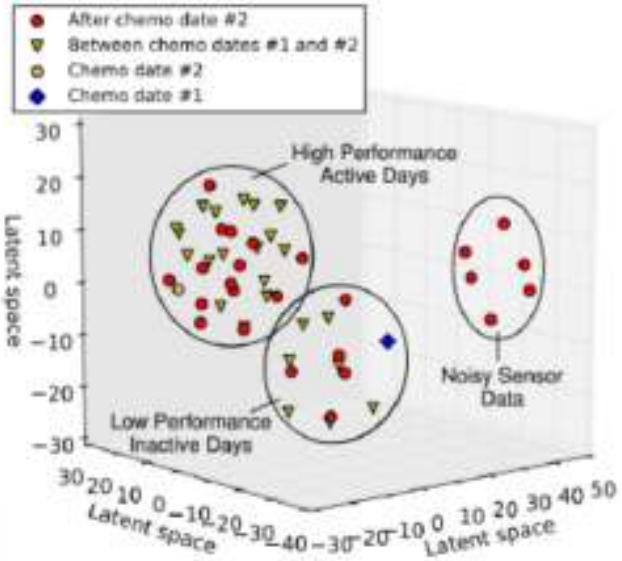


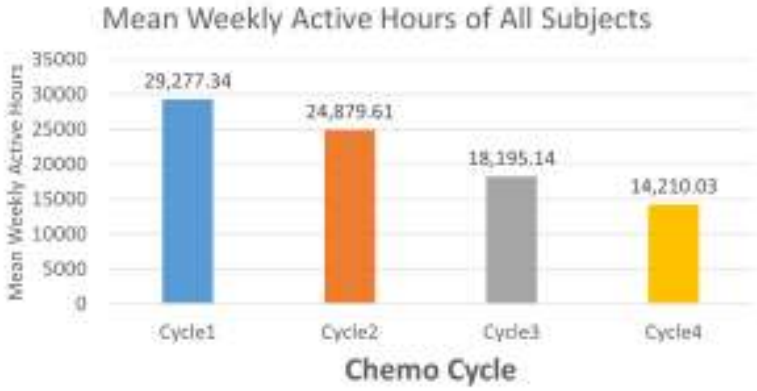
GFT features capture



Happy : No
Engaged : Yes
WearingGlasses : No
LeftEyeClosed : No
RightEyeClosed : No
MouthOpen : Yes
MouthMoved : Yes
LookingAway : No
FaceYaw : -10
FacePitch : 0
FacenRoll : -5





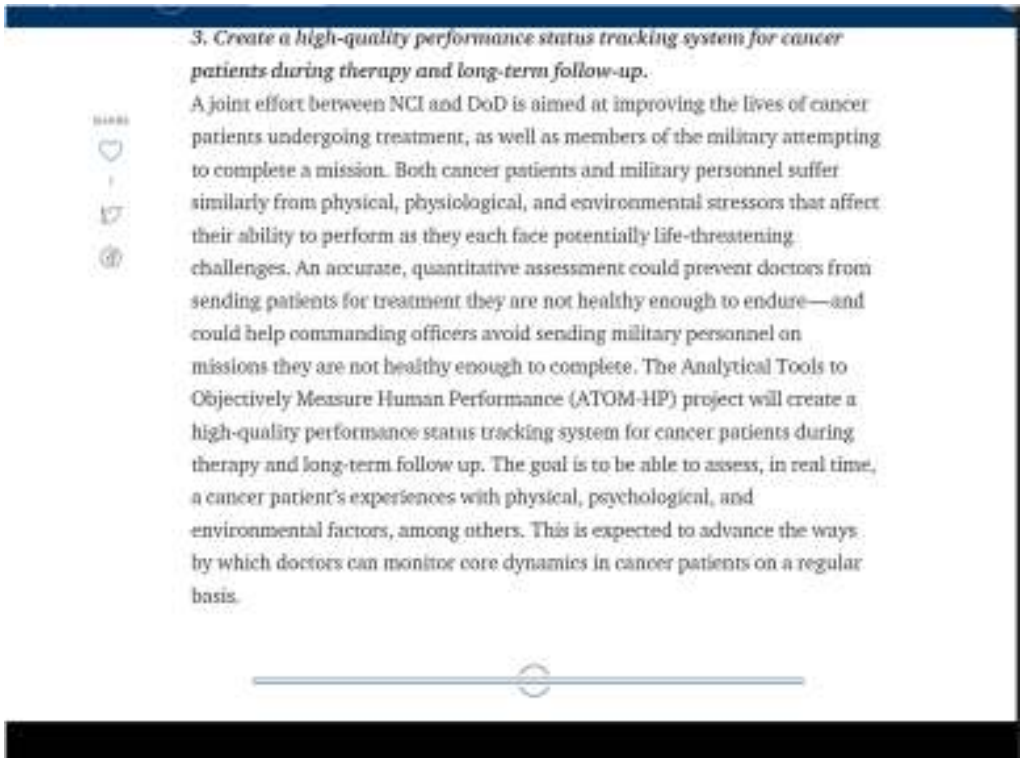


ATOM-HP Demo: SXSL 2016 at White House

Health

ATOM-HP in the News

The ATOM-HP is a formal recommendation by the Whitehouse. It was presented to President Obama by Vice President Biden in Sep'16 as the final outcome of the moonshot.



3. Create a high-quality performance status tracking system for cancer patients during therapy and long-term follow-up.

A joint effort between NCI and DoD is aimed at improving the lives of cancer patients undergoing treatment, as well as members of the military attempting to complete a mission. Both cancer patients and military personnel suffer similarly from physical, physiological, and environmental stressors that affect their ability to perform as they each face potentially life-threatening challenges. An accurate, quantitative assessment could prevent doctors from sending patients for treatment they are not healthy enough to endure—and could help commanding officers avoid sending military personnel on missions they are not healthy enough to complete. The Analytical Tools to Objectively Measure Human Performance (ATOM-HP) project will create a high-quality performance status tracking system for cancer patients during therapy and long-term follow up. The goal is to be able to assess, in real time, a cancer patient's experiences with physical, psychological, and environmental factors, among others. This is expected to advance the ways by which doctors can monitor core dynamics in cancer patients on a regular basis.

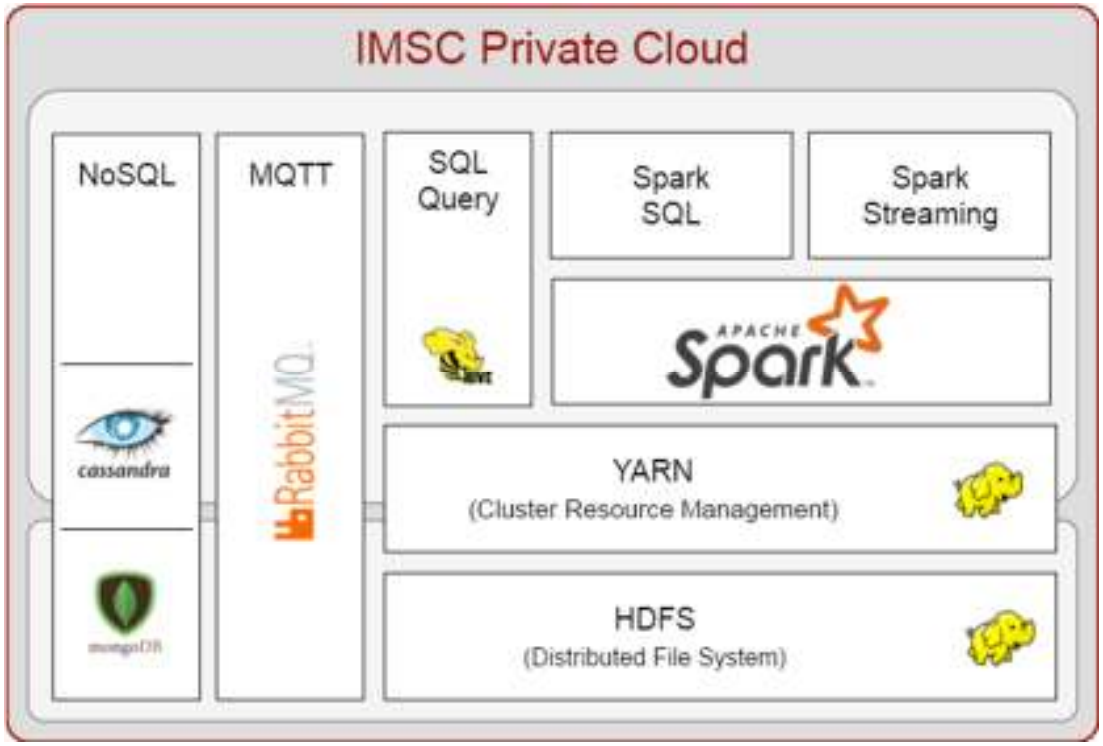
USC Viterbi School of Engineering

http://http://www.nbclosangeles.com/on-air/as-seen-on/Wearable-Tech-Improves-Cancer-Treatment_Los-Angeles-395200891.html

OUTLINE

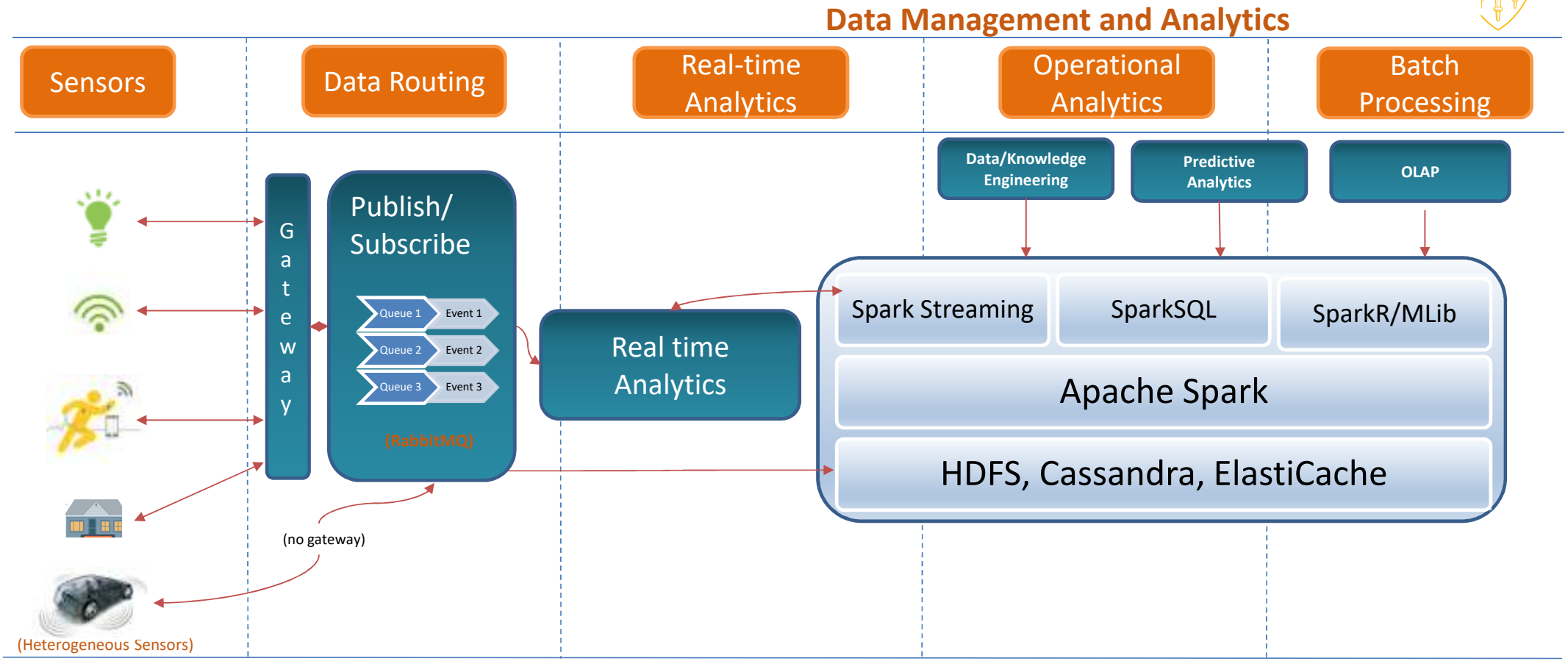
- DSI Overview
- Transportation Data Platform
- Social Media Data Platform
- Health Data Platform
- Smart City Data Platform
- Closing Remarks

DSI Private Cloud



Moving all our datasets into a single platform for Data and Code Sharing!

DSI Uniform Data Platform for IoT



OUTLINE

- DSI Overview
- Transportation Data Platform
- Social Media Data Platform
- Health Data Platform
- Smart City Data Platform
- Closing Remarks

Team

Director

Cyrus Shahabi
Director
(213) 740-8162

Associate Directors

Luciano Nocera
Associate Director
(213) 740-0478

Seon Ho Kim
Associate Director
(213) 821-0876

Yao-Yi Chiang
Associate Director

Research Associates

Anna Farzindar
Research Associate
Health Informatics
Advisor

Erick Moen
Research Associate

Technical Staff

Chrysovalantis Anastasiou
Systems Administrator

Kaila Su Young Kim
Assistant Researcher

Yijun Lin
Data Scientist
(Research)

Administrative Staff

Daisy Tang
Research
Administrator
(213) 740-8945

Affiliated Faculty

Aleksandra Korolova
Assistant Professor, Computer Science
(213) 740-

Antonio Ortega
Professor, Electrical Engineering
(213) 740-2320

Bhaskar Krishnamachari
Associate Professor, Electrical Engineering
(213) 821-2528

Carolee J. Winstein
Professor of Biokinesiology and Physical Therapy

Gerard Medioni
Professor, Computer Science
(213) 740-6440

Gerard Power
Assistant Professor of Clinical Marketing

Hao Li
Assistant Professor, Computer Science
941 Bloom Walk, SAL 244, USC

John P. Wilson
Professor of Sociology and Spatial Sciences

Joon-Ho Choi
Assistant Professor, Building Science in Architecture
(213) 740-4576

Jorge Nieva, MD
Associate Professor of Clinical Medicine

Ketan Savla
Assistant Professor & John and Dorothy Shea Early Career Chair in Civil Engineering

Kiran K. Dhanireddy, MD
Assistant Professor of Clinical Surgery
Director of Transplant Quality

Leslie Ann Saxon, MD
Professor of Medicine
Executive Director, Center For Body Computing

Milind Tambe
Professor, Computer Science
(213) 740-6447

Muhamaad Naveed
Assistant Professor, Department of Computer Science, USC

Patrick Joseph Lynett
Professor, Civil Engineering
(213) 740-3133

Peter Kuhn
Associate Director of The Bridge@USC

Ram Nevatia
Professor, Computer Science
(213) 740-6427

Shaddin Dughmi
Assistant Professor of Computer Science, USC

Yan Liu
Assistant Professor, Computer Science
(213) 740-4371

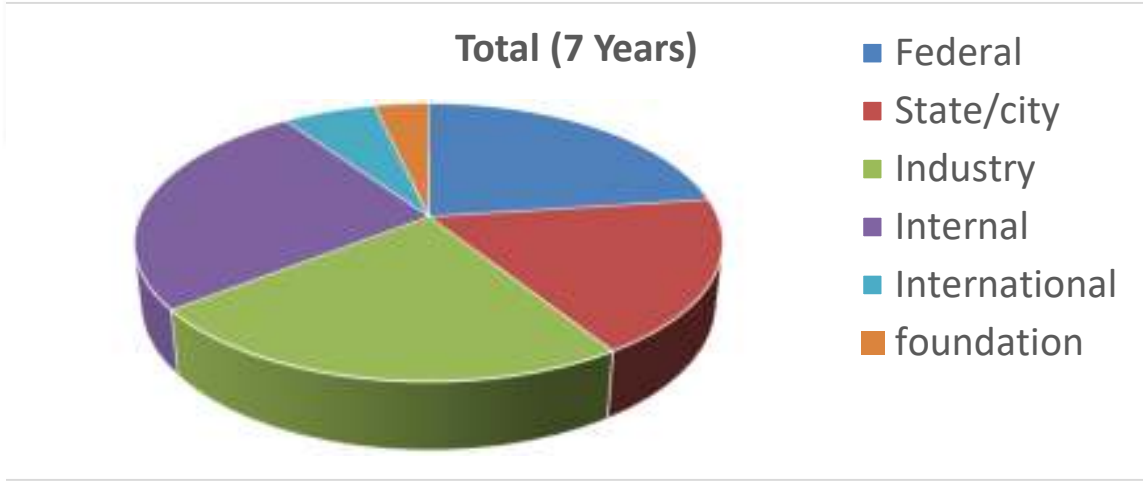
USC Viterbi
School of Engineering
Data Science Institute

Thank you

Sponsors

Internal Sponsors

In-Kind Support



Impact – Workforce

Graduates in the last 5 years

- **PhD**

Afsin Akdogan	Google
Huy Pham	Oracle
Bei Pan	Microsoft
Houtan Shirani-Mehr	Uber
Ali Khodaei	facebook
Leyla Kazemi	Microsoft
Ugur Demiryurek	Apple
Ling Hu	Google
Songhua Xing	Oracle
Ali Khoshgozaran	Tilofy

- **Selected MS and Undergrad**

Shireesh Asthana	facebook
Jiayun Ge	Microsoft
Yu Sun	facebook
Ashley Luo	amazon
Jingyi Du	Microsoft
Nicholas Bopp	Microsoft
Junyuan Shi	Qualcomm
Colin Gu	Teradata
Vanessa Kuroda	NASA
Ning Jiang	PayPal

DSI Value Add to its Partners

- Our vision, expertise, background & experience in
  - Fundamental and applied research
  - Multidisciplinary research
  - Integrated system development
- Our test-beds
- Government/Federal customers
- Industry Partners
- Global Reach
- Educational Presence