Modeling Wildfire Behavior at the Continuum of Computing

1st Workshop on Parallel AI and Systems for the Edge (PAISE)

İlkay ALTINTAŞ, Ph.D.

UC San Diego

Chief Data Science Officer & Division Director of Cyberinfrastructure Research, Education and Development, **San Diego Supercomputer Center**

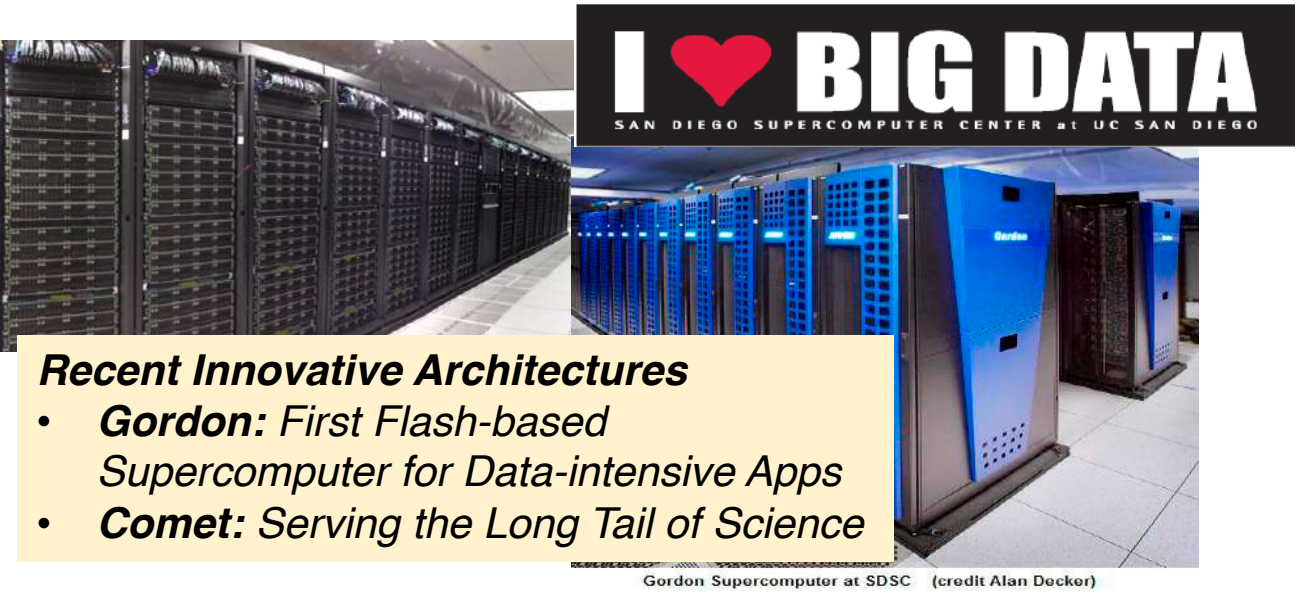
Fellow of Cyberinfrastructure and Founding Faculty, **Halıcıoğlu Data Science Institute**

Director**, Workflows for Data Science Center of Excellence**

A little about me…

SAN DIEGO SUPERCOMPUTER CENTER at UC San Diego

- Established as a national supercomputer resource center in 1985 by NSF
- A world leader in HPC, data-intensive computing, and scientific data management
- Current strategic focus on “Big Data”, “versatile computing”, and “life sciences applications”



Gordon Supercomputer at SDSC (credit Alan Decker)

Recent Innovative Architectures

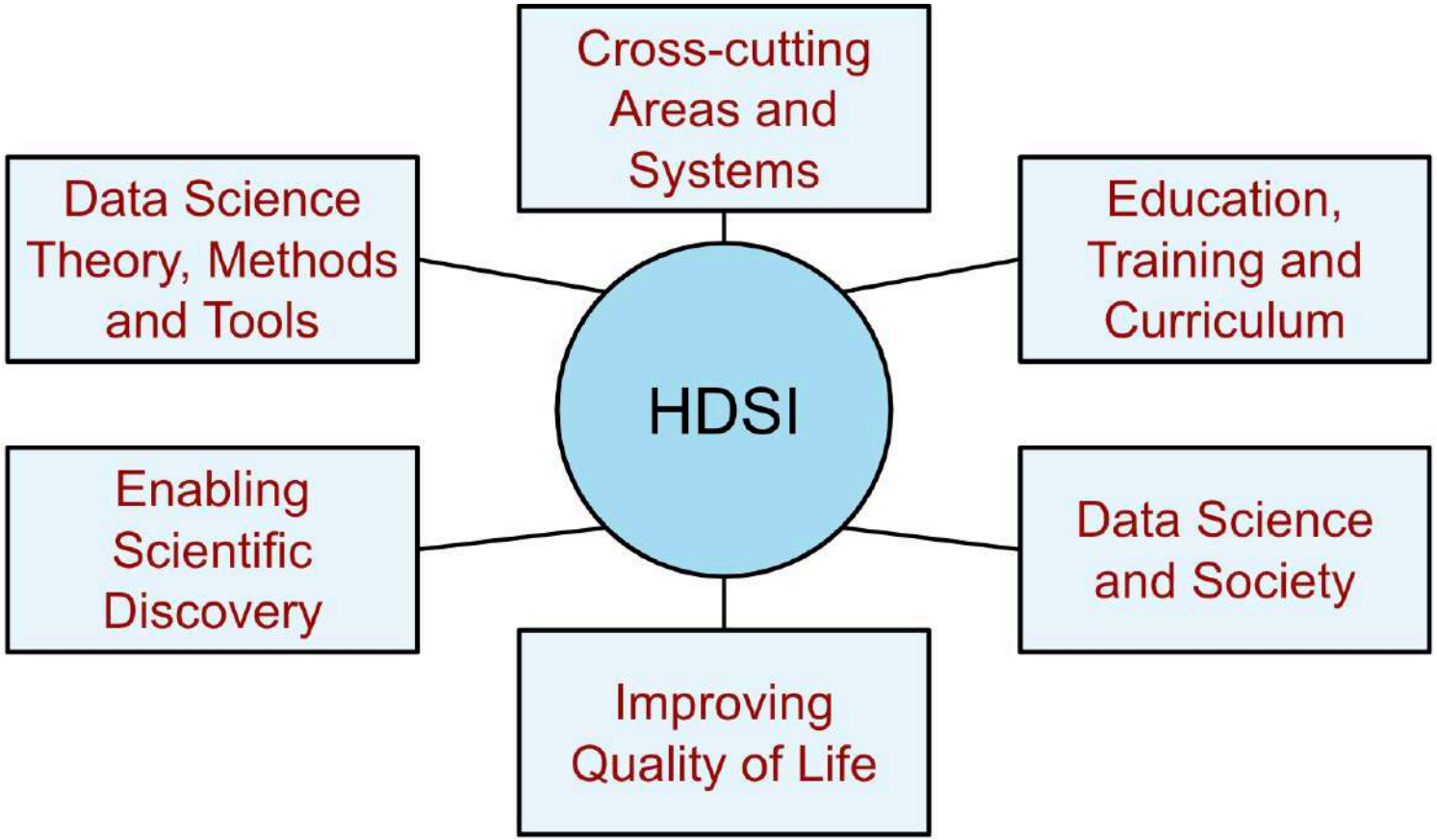
- ***Gordon:*** *First Flash-based Supercomputer for Data-intensive Apps*
- ***Comet:*** *Serving the Long Tail of Science*

Data Science Hub at SDSC

- **SDSC Data Gateway**
- **Collaborative experts and services for**
  - **Deep learning and AI**
  - **Machine learning**
  - **Blockchain**
  - **IOT**
  - **Data engineering and management services**
- **Heterogenous data science platforms**
- **Industry engagement opportunities**

Halıcıoğlu Data Science Institute

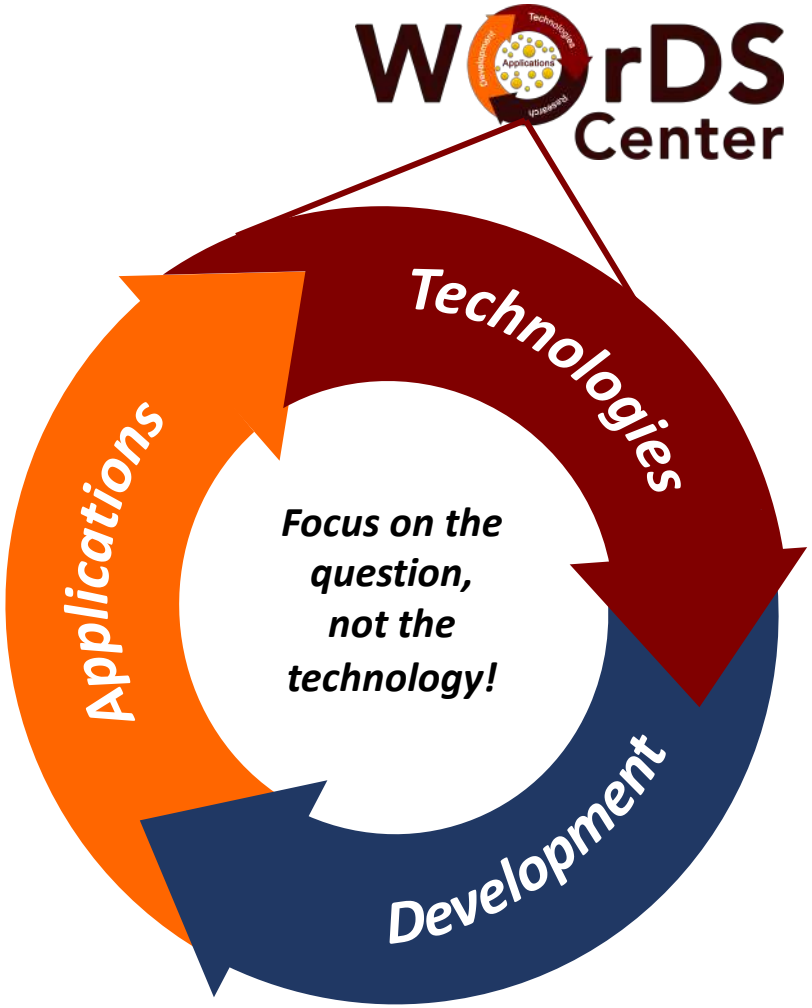
https://datascience.ucsd.edu



An academic unit that provides a home for curating the growth in Data Sciences as a discipline

- Deep expertise organized into clusters
- Manage new engagements to seed, cultivate and grow the practice of Data Sciences

Workflows for Data Science Center of Excellence at SDSC



http://WorDS.sdsc.edu

Mission:
Methodology and tool development
to enable collaborative workflow-driven science
and create solution architectures
op top of big data and advanced computing platforms.

Common Theme…

"Big" Data, Computational Science, Data Science, Cyberinfrastructure,

and Their Applications

i.e., the problems we are solving

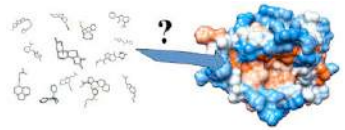
In most applications, utilization of Big Data often needs to be combined with Scalable Computing.

COMPUTING AT DIVERSE SCALES

"BIG" DATA

Enables dynamic data-driven applications

Computer-Aided Drug Discovery



Smart Manufacturing

Personalized Precision Medicine

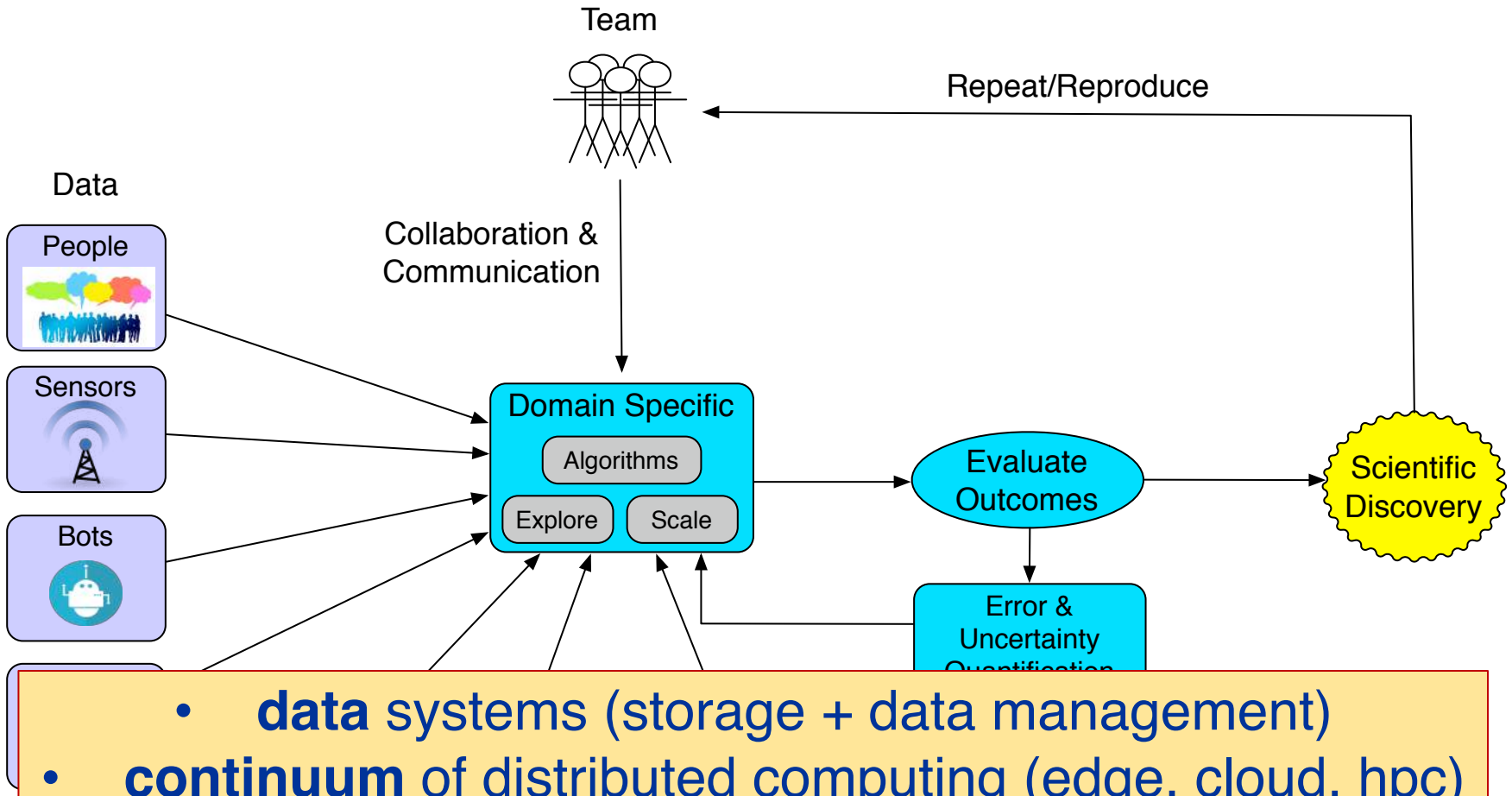
Smart Cities

Disaster Resilience and Response

Smart Grid and Energy Management

The problems we are solving are …

- data-driven
- heterogeneous
- collaborative



- **data** systems (storage + data management)
- **continuum** of distributed computing (edge, cloud, hpc)
- high speed **connectivity** (programmable networks)

What if the whole world was our computing environment?

New Opportunities for AI-Driven Approaches and Cyberinfrastructure

- Closing the loop between observation, experimentation and simulation
- Dynamic data-
- Real-time data
- Reactive syster
- Coupling the e
- New software and tools for AI-driven computing
- Intelligent security, integrity and privacy practices

All require getting value from data as fast as possible, some at the edge!

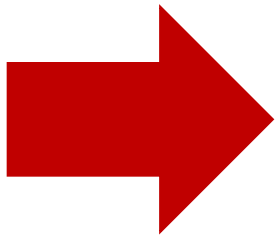
ng data and AI

otation

computations

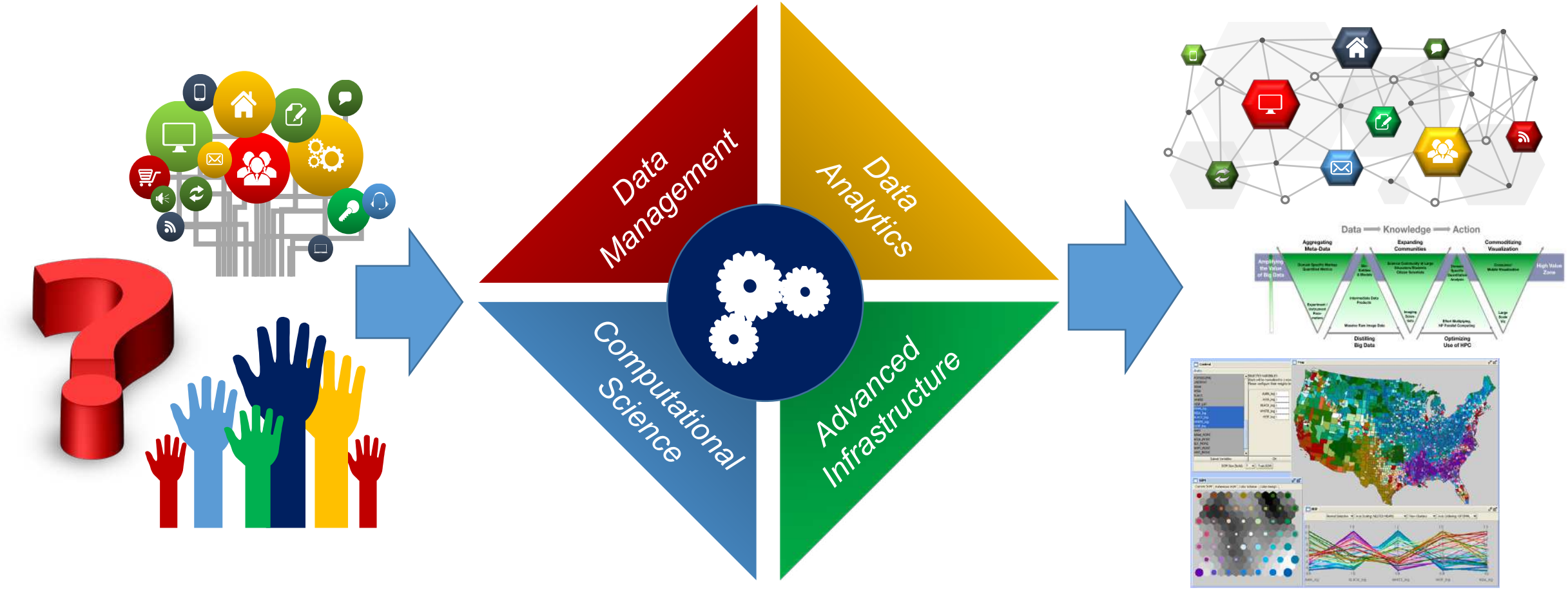
Going from Data to Discovery to Impact

Amplifying the Value of Data Related to **X**

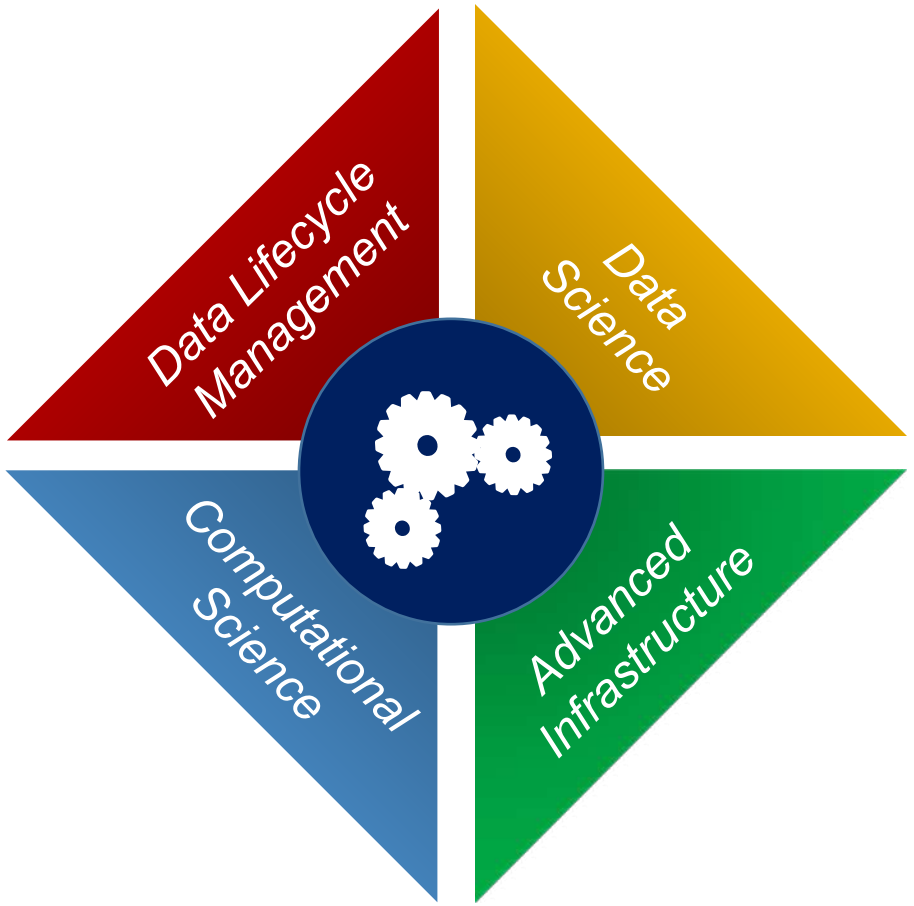


Benefit **Y** for Science, Business, Society or Education

We need to focus on the problems to solve.

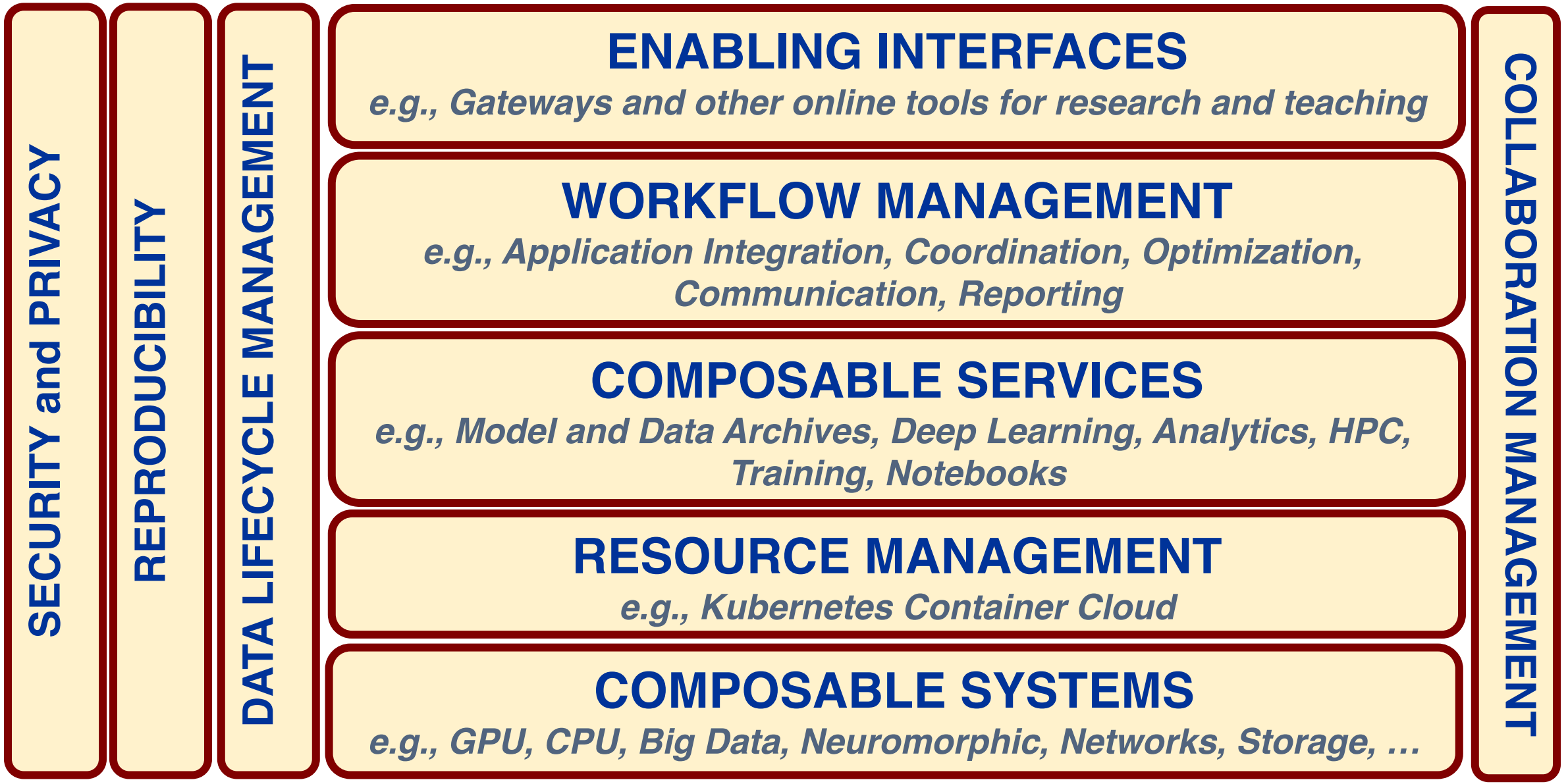


A Typical Collaborative Data Science Ecosystem



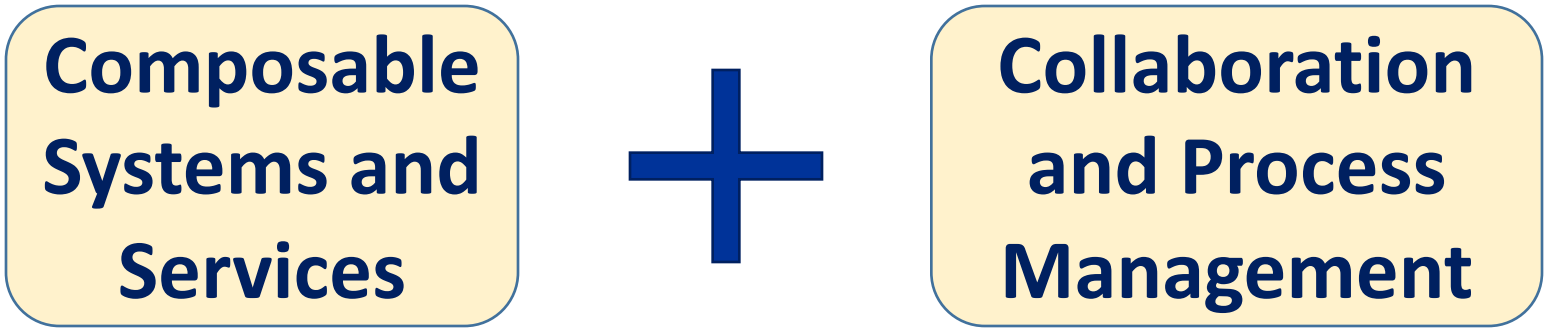
Data-driven problem solving requires:

- Heterogenous systems
- Data management
- Data-driven methods
- Scalable tools for dynamic coordination and resource optimization
- Skilled interdisciplinary workforce
- Collaborative culture and tools that enable groups to communicate


SECURITY and PRIVACY
REPRODUCIBILITY
DATA LIFECYCLE MANAGEMENT
ENABLING INTERFACES
e.g., Gateways and other online tools for research and teaching
WORKFLOW MANAGEMENT
e.g., Application Integration, Coordination, Optimization, Communication, Reporting
COMPOSABLE SERVICES
e.g., Model and Data Archives, Deep Learning, Analytics, HPC, Training, Notebooks
RESOURCE MANAGEMENT
e.g., Kubernetes Container Cloud
COMPOSABLE SYSTEMS
e.g., GPU, CPU, Big Data, Neuromorphic, Networks, Storage, …
COLLABORATION MANAGEMENT

The rest of my talk…

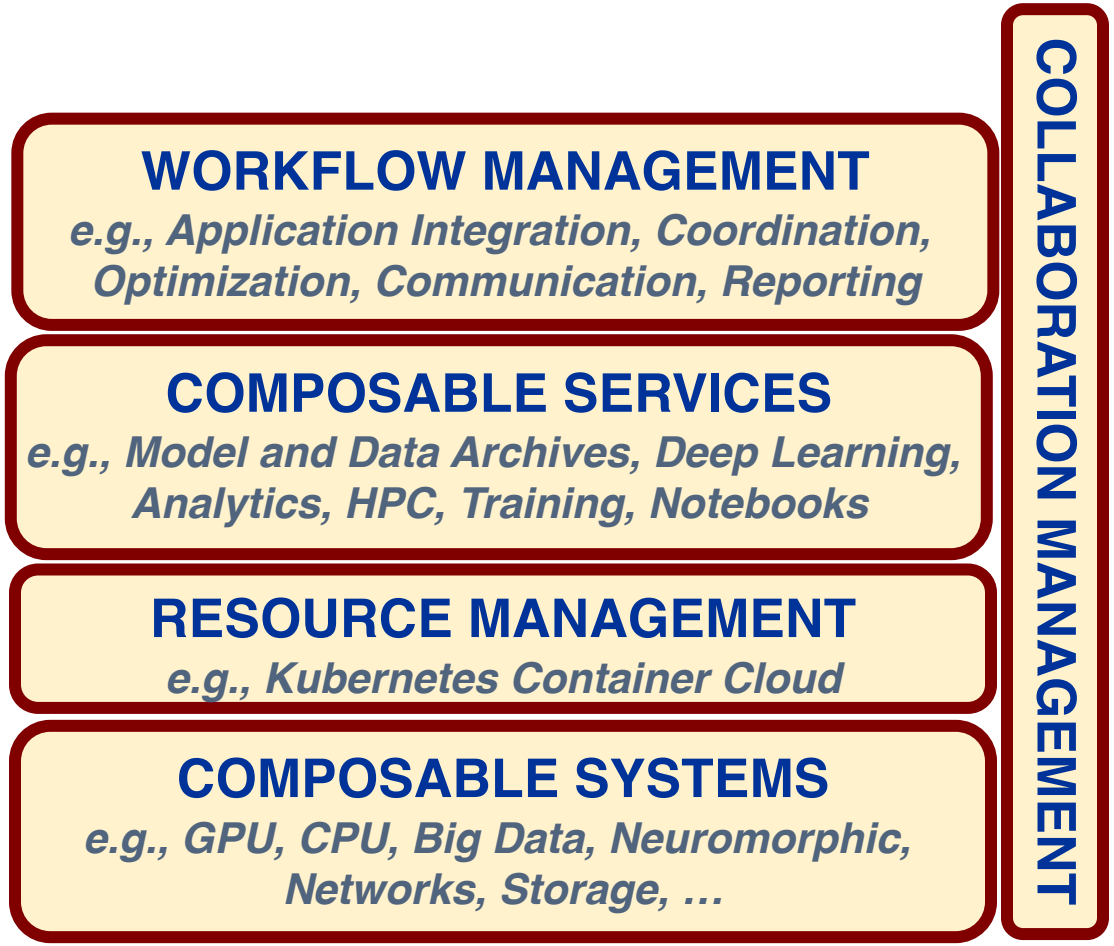
- Two general concepts for how problem solving happens on the continuum of computing



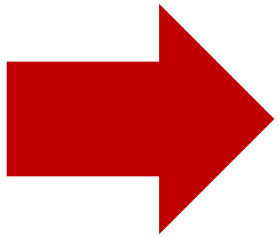
- Their applications to wildfires with some examples

Smart Composability from Systems to Services

- **Composable Systems**
  - dynamically measured and resourced
  - used as a resource based on need and availability
- **Resource Management**
  - Kubernetes integration
  - mapping tools for resource identification per task
- **Composable Services**
  - runs on composable systems (e.g., containers)
  - exposes a parametric interface for integration
  - continuously measured and profiled
- **Workflow Management**
  - focused on coordination and resource optimization
  - requires a number of AI-based tools and data processing to function
- **Collaboration Management**
  - focused on expertise integration and metric setting
  - applies methodologies for effective communication
  - provides tools measure and validate team activity



Amplifying the Value of Data Related to **X**



Benefit **Y** for Science, Business, Society or Education

What if X was wildfires?

WIFIRE

Best Data-Intensive System (End User focused) HPCwire Editors' Choice Awards 2014

Best Application of Big Data in HPC HPCwire Readers' Choice Awards 2014

Best Application of Big Data in HPC HPCwire Editors' Choice Awards 2014

ICCS 2015 Best Paper

CENIC 2018 Innovations in Networking Award for Experimental Applications

An Integrated Cyberinfrastructure for Scalable Data-Driven Monitoring, Dynamic Prediction and Resilience of Wildfires

wifire.ucsd.edu

Ilkay Altintas[1], Jessica Block[2], Daniel Crawl[1], Raymond de Callafon[3], Hans-Werner Braun[1], Michael Gollner[4], Larry Smarr[2], Jurgen Schulze[2] and Arnaud Trouve[4]

THINK GLOBAL ACT LOCAL

[1]San Diego Supercomputer Center, University of California San Diego, U.S.A.
[2]Qualcomm Institute, University of California San Diego, U.S.A.
[3]Dept. of Mechanical and Aerospace Engineering, University of California San Diego, U.S.A.
[4]Fire Protection Engineering Dept., University of Maryland, U.S.A.

NSF

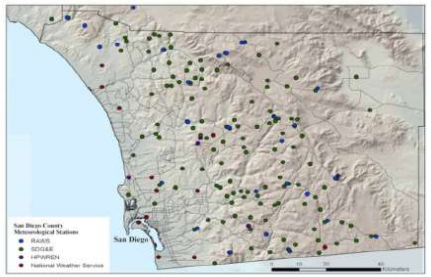
NSF-1331615

Situational awareness of and preparedness for wildfires heavily relies on understanding their Direction and Rate of Spread (RoS).

Photo of Harris Fire (2007) by former Fire Captain Bill Clayton

How do we Better Predict Wildfire Behavior?

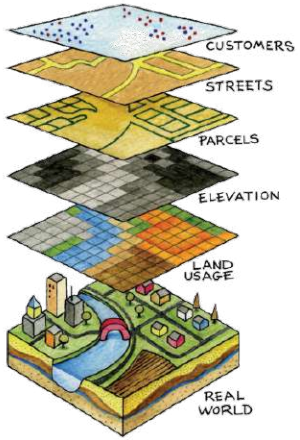
Fire Modeling Requires Many Types of Data



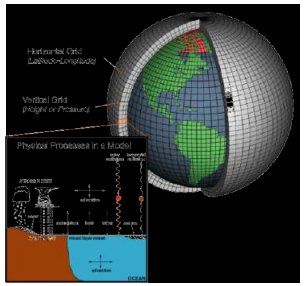
Ground-based real-time weather and camera imagery

Fire perimeter

Remote sensing

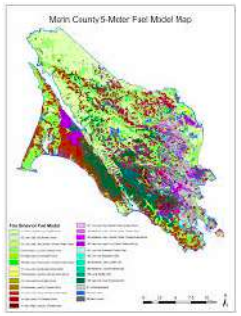


Landscape data



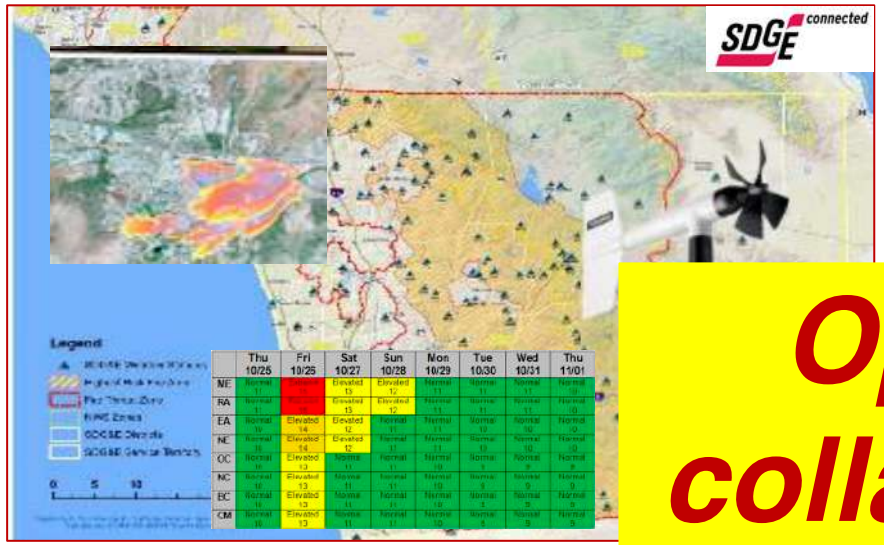
Weather forecast and modeling

Data-Driven Fire Modeling



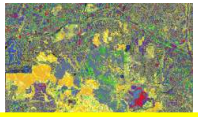
Land cover and fuel

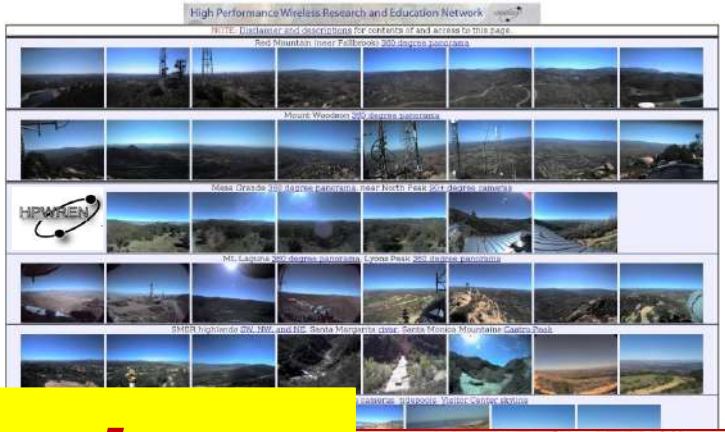
All modeling and observational methods are potentially useful.



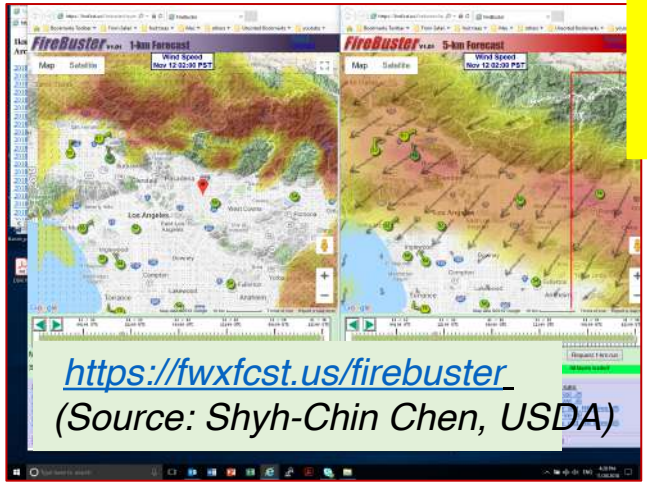
High-resolution satellite imagery

Vegetation classification

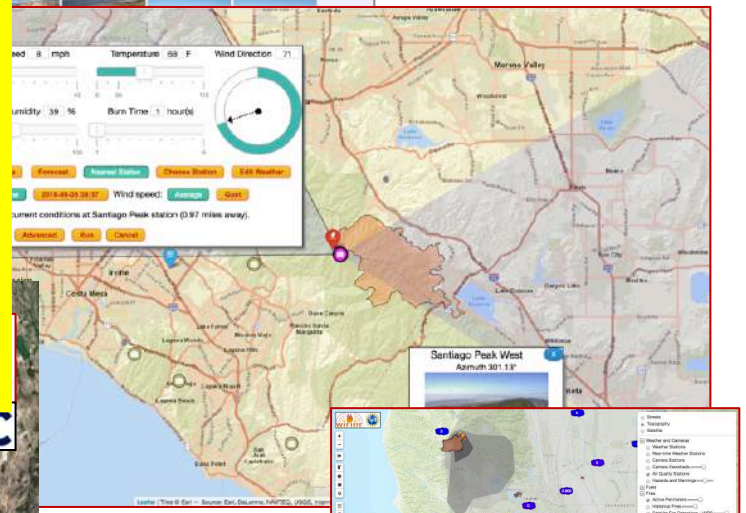


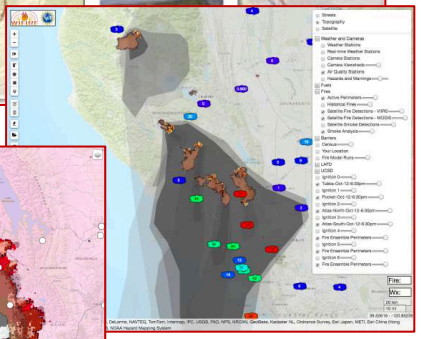


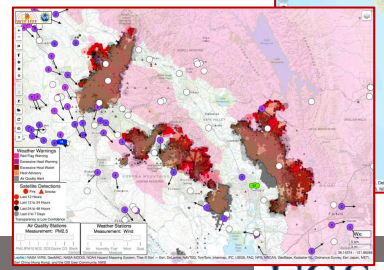
Opportunity to collaborate around models and data!



https://fwxfcst.us/firebuster
(Source: Shyh-Chin Chen, USDA)





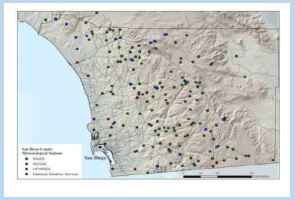


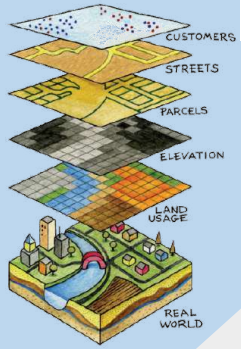
SDSC SAN DIEGO SUPERCOMPUTER CENTER

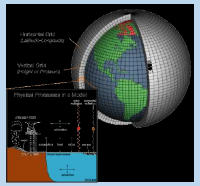
GENERAL ATOMICS

İ tıntaş, PhD (ialtintas@ucsd.edu)

UC San Diego
HALICIOĞLU DATA SCIENCE INSTITUTE







SOLUTION

*A neutral **intelligent** and **integrated** infrastructure to catalog, curate, exchange, analyze and communicate data*

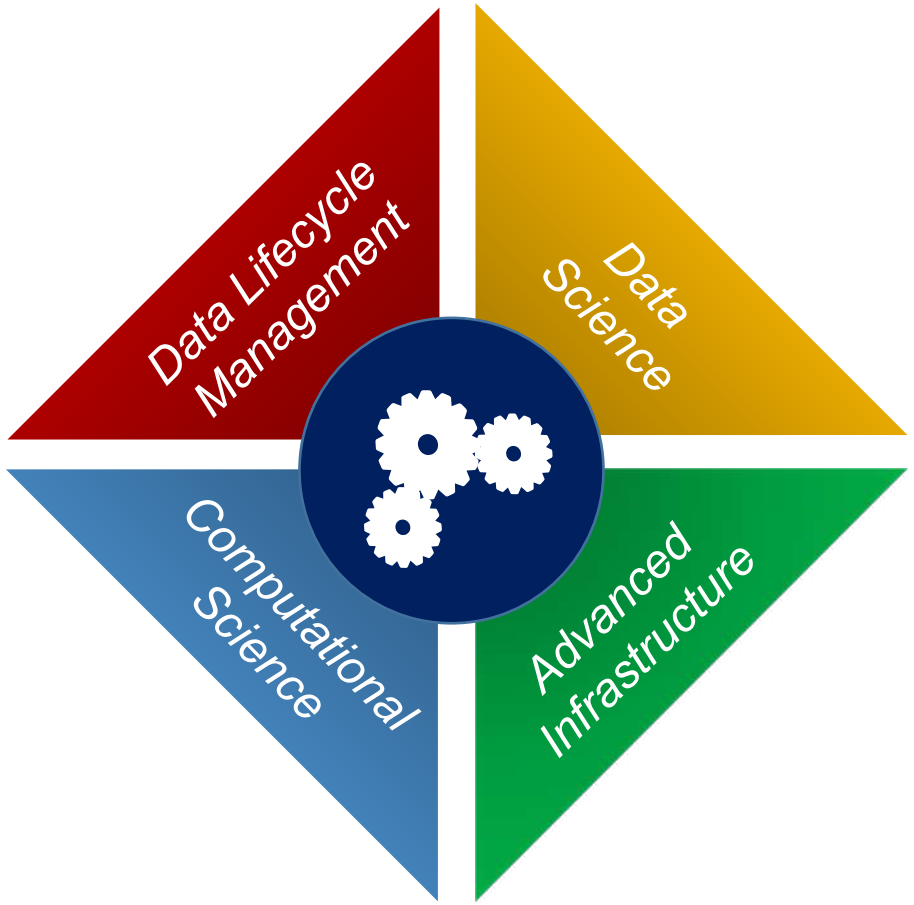
WILDFIRE INTELLIGENCE

A dynamic system integration of…..

real-time sensor networks, satellite imagery, near-real time data management tools, wildfire simulation tools, and connectivity to emergency command centers

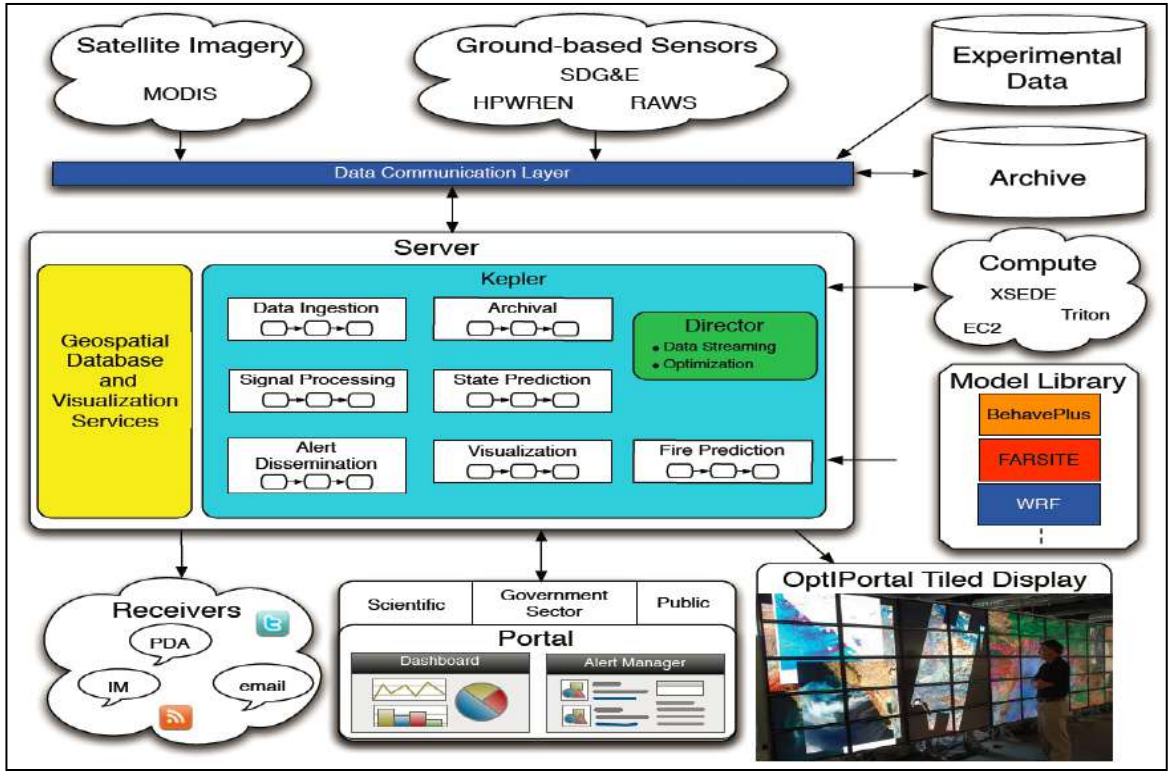
…. before, during and after a firestorm.

NSF Science Nation video available at:
https://www.youtube.com/watch?v=N4LAROiW5c8&t=2s

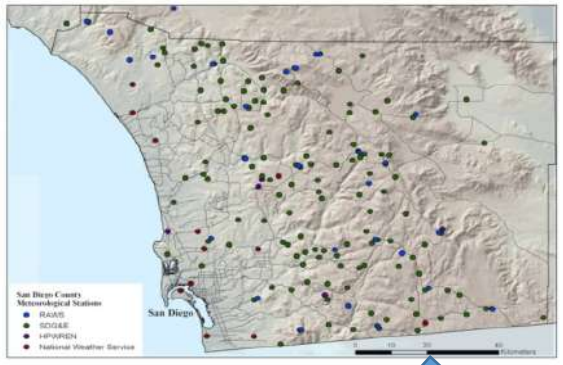

Data Lifecycle Management
Data Science
Computational Science
Advanced Infrastructure

WIFIRE

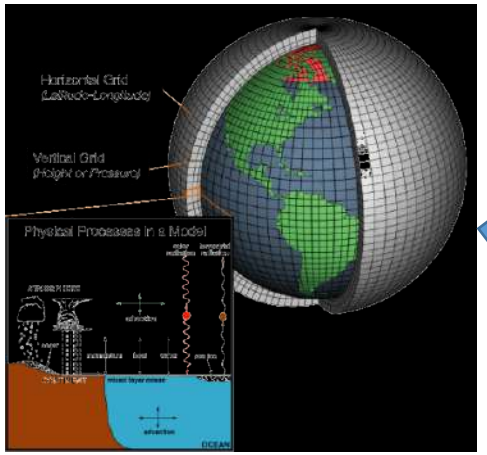
NSF


Satellite Imagery
MODIS
Ground-based Sensors
SDG&E
HPWREN
RAWS
Experimental Data
Archive
Data Communication Layer
Server
Kepler
Geospatial Database and Visualization Services
Data Ingestion
Archival
Director
Data Streaming
Optimization
Signal Processing
State Prediction
Alert Dissemination
Visualization
Fire Prediction
Compute
XSEDE
Triton
EC2
Model Library
BehavePlus
FARSITE
WRF
Receivers
PDA
IM
email
Scientific
Government Sector
Public
Portal
Dashboard
Alert Manager
OptIPortal Tiled Display

Fire Weather Monitoring and Prediction in WIFIRE

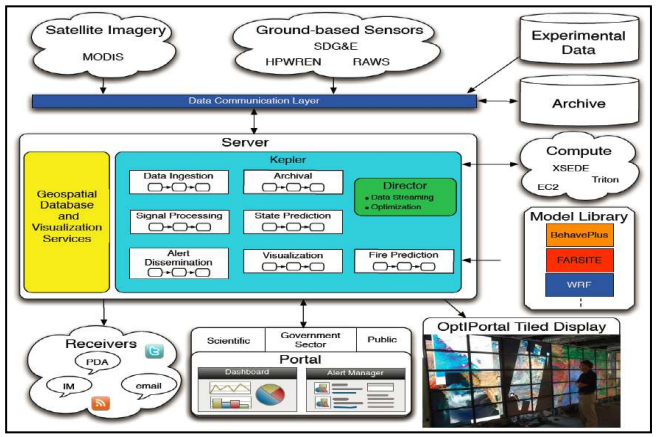


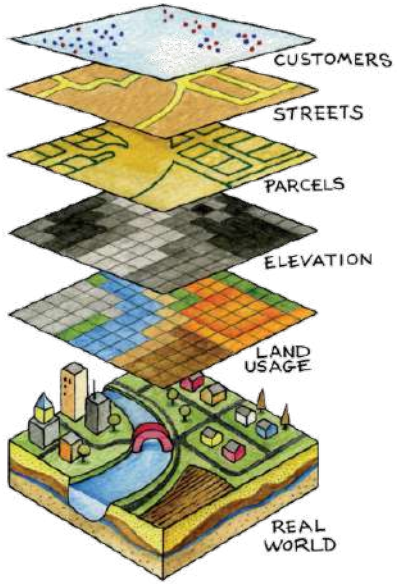
Real-time sensors



Weather forecast

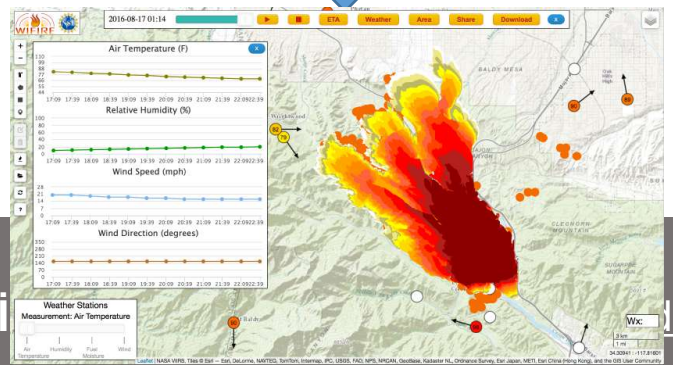
Fire perimeter





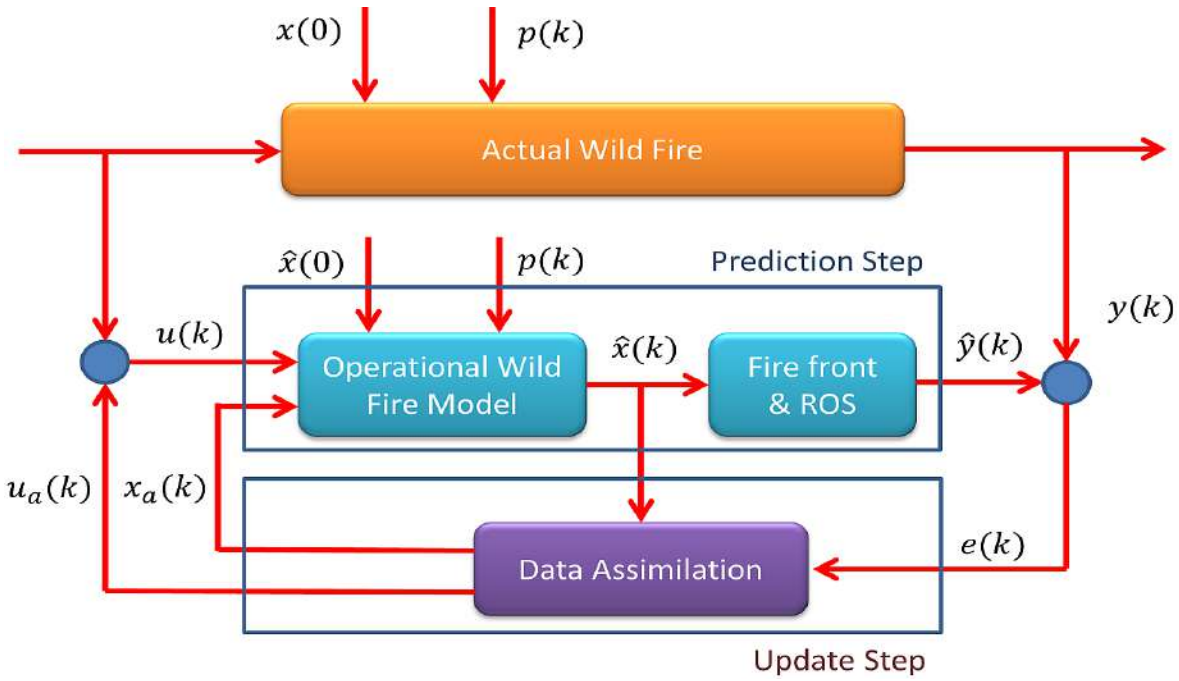
Landscape data

Monitoring & fire mapping



Closing the Loop using Big Data

-- Wildfire Behavior Modeling and Data Assimilation --

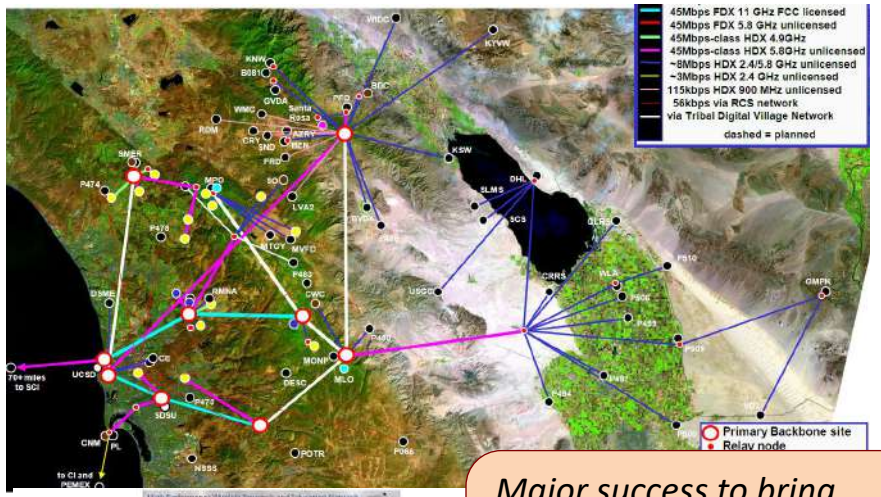


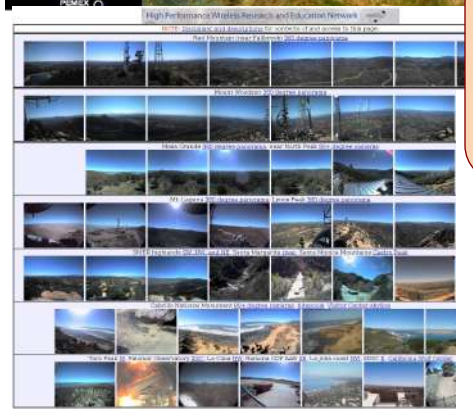
Conceptual Data Assimilation Workflow with Prediction and Update Steps using Sensor Data

- *a priori -> a posteriori*
  - Parameter estimation to make adjustments to the (input) parameters
  - State estimation to adjust the simulated fire front location with an a posteriori update/measurement of the actual fire front location

- Srivas, T., Artés, T., de Callafon, R., Altintas, I., **Wildfire Spread Prediction and Assimilation for FARSITE Using Ensemble Kalman Filtering**. doi:10.1016/j.procs.2016.05.328
- Srivas, T., de Callafon, R., Crawl, D., Altintas, I., **Data Assimilation of Wildfires with Fuel Adjustment Factors in FARSITE using Ensemble Kalman Filtering.** doi:10.1016/j.procs.2017.05.197

High Performance Wireless Research and Education Network



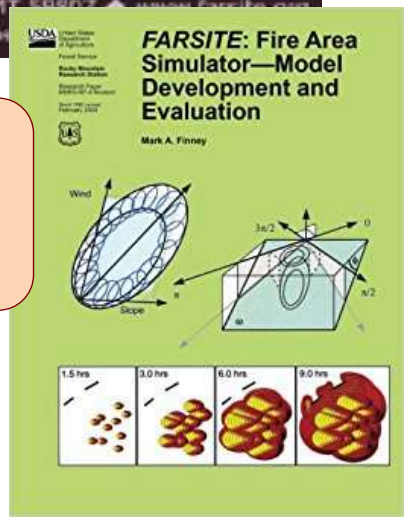


http://hpwren.ucsd.edu/cameras

Major success to bring internet to incident command in the field. Used in over 20 fires over time.

>160 Meteorological Sensors and Growing

FARSITE



Most popular operational fire behavior modeling system.

Use Case: Santa Ana Conditions

- *Santa Ana* winds lead to dangerous fire conditions in San Diego County
  - Oct. 2003:

    >700,000 acres burned
  - Oct. 2007:

    >500,000 acres burned

Santa Ana defined by:

Wind direction *> 10° and < 110°*

Wind speed *> 25mph*

Relative humidity *< 25%*

- ***Goal:*** *determine regions in San Diego County experiencing Santa Ana Winds*
- ***Solution:*** *Use WindNinja to compute wind conditions, post-process to find Santa Ana regions*

HPWREN Real-Time Weather Alerts

- *Weather stations measurements monitored Santa Ana conditions*
- *Alerts sent via email*

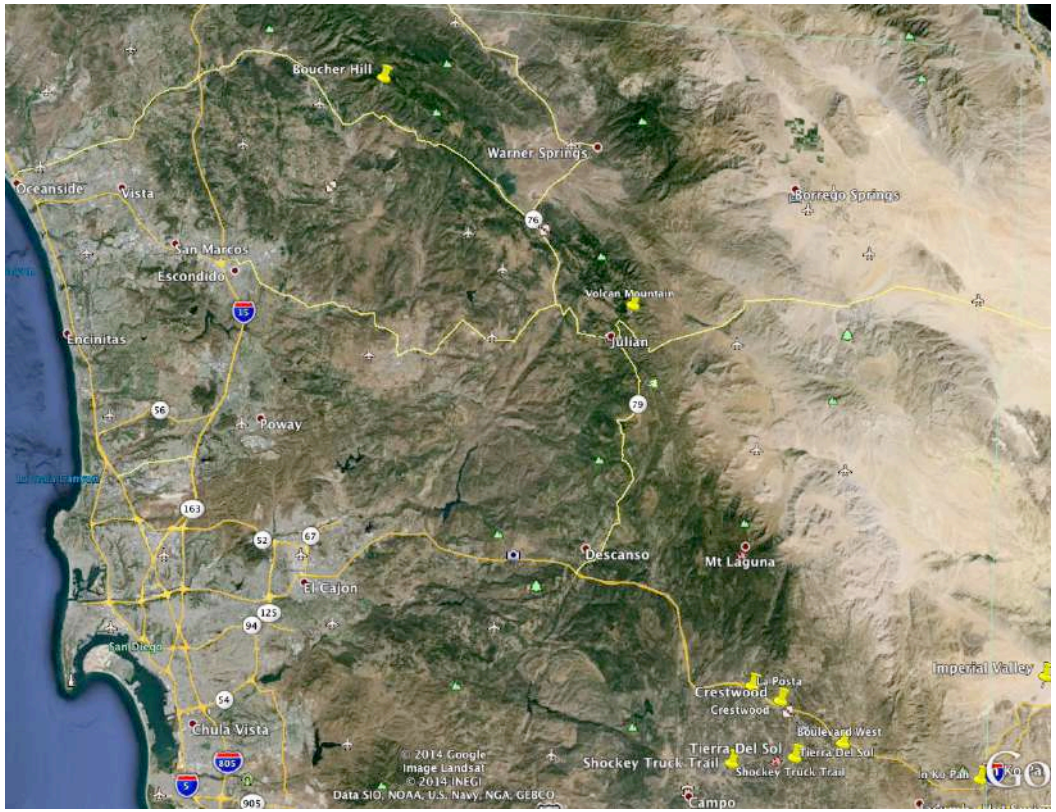
Santa Ana condition real-time weather sensor alert	Hans-Werner Braun	12/27/14 10:42 AM
Santa Ana condition real-time weather sensor alert	Hans-Werner Braun	12/27/14 10:52 AM
Santa Ana condition real-time weather sensor alert	Hans-Werner Braun	12/27/14 11:12 AM
Santa Ana condition real-time weather sensor alert	Hans-Werner Braun	12/27/14 11:22 AM
Santa Ana condition real-time weather sensor alert	Hans-Werner Braun	12/27/14 12:52 PM
Santa Ana condition real-time weather sensor alert	Hans-Werner Braun	12/27/14 1:13 PM
Santa Ana condition real-time weather sensor alert	Hans-Werner Braun	12/27/14 1:39 PM
Santa Ana condition real-time weather sensor alert	Hans-Werner Braun	12/28/14 7:20 AM
Santa Ana condition real-time weather sensor alert	Hans-Werner Braun	1/1/15 2:16 PM

From Hans-Werner Braun

Subject **Santa Ana condition real-time weather sensor alert** 1/7/

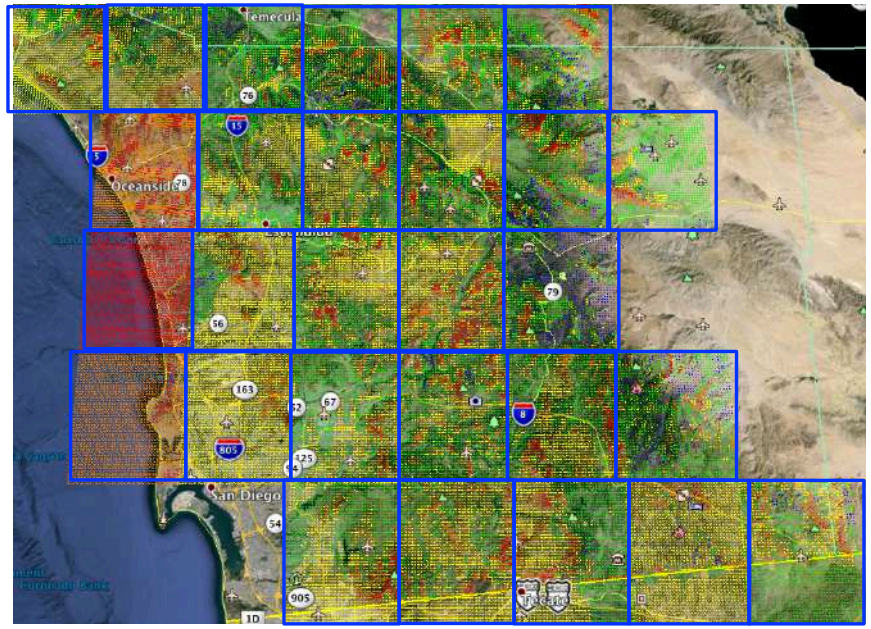
To l3@hpwren.ucsd.edu

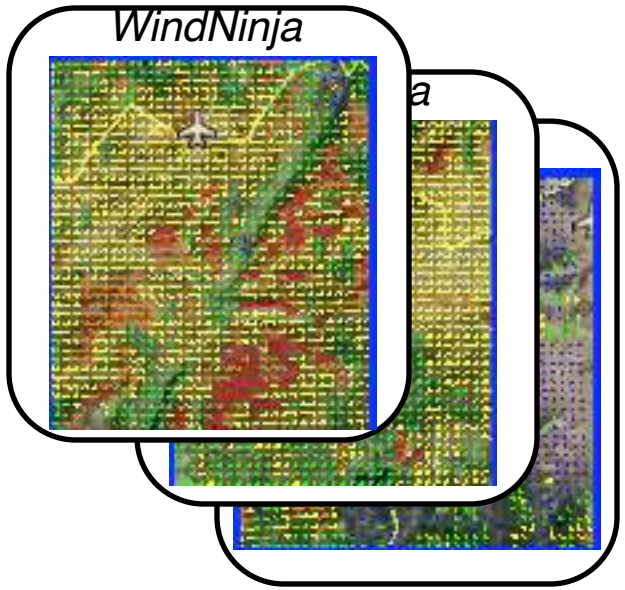
```
BMR: RH=15.9 WD=93 WS=28 AT=62 20150107.070022 Big Black Mountain
More details at http://hpwren.ucsd.edu/Sensors/
```



Spatial Coverage

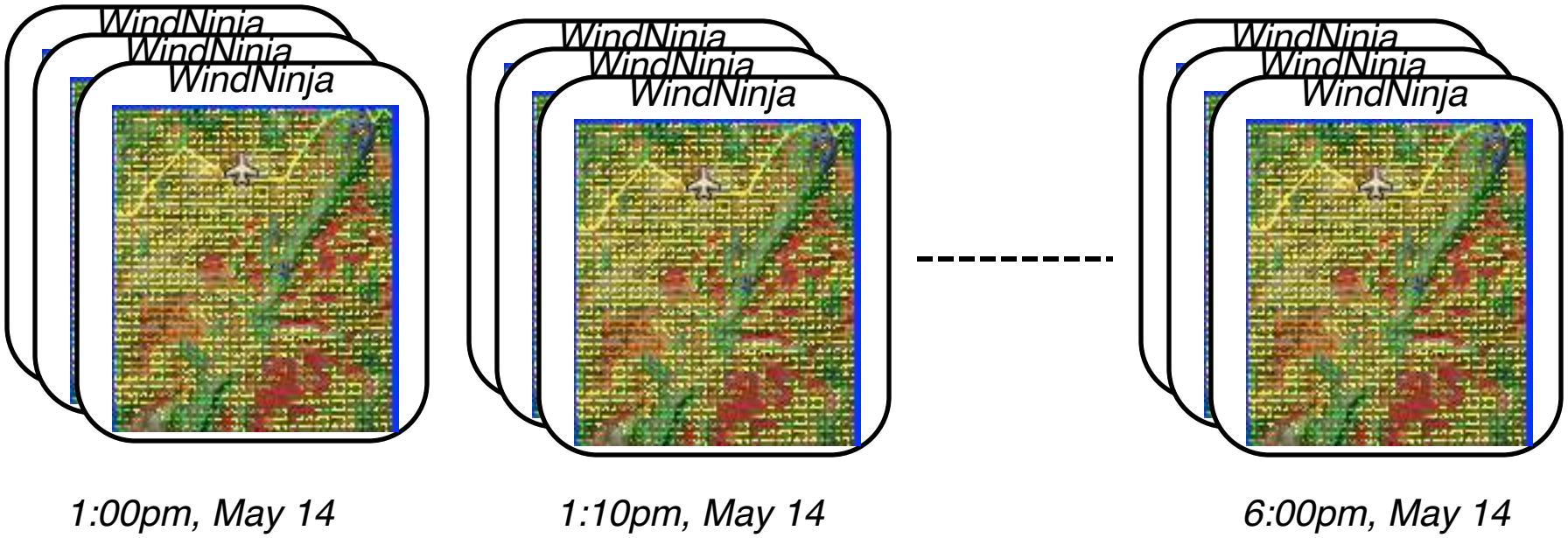
- WindNinja run on domain size up to 50x50km
  - Split SD County into tiles
  - Run WindNinja for each tile





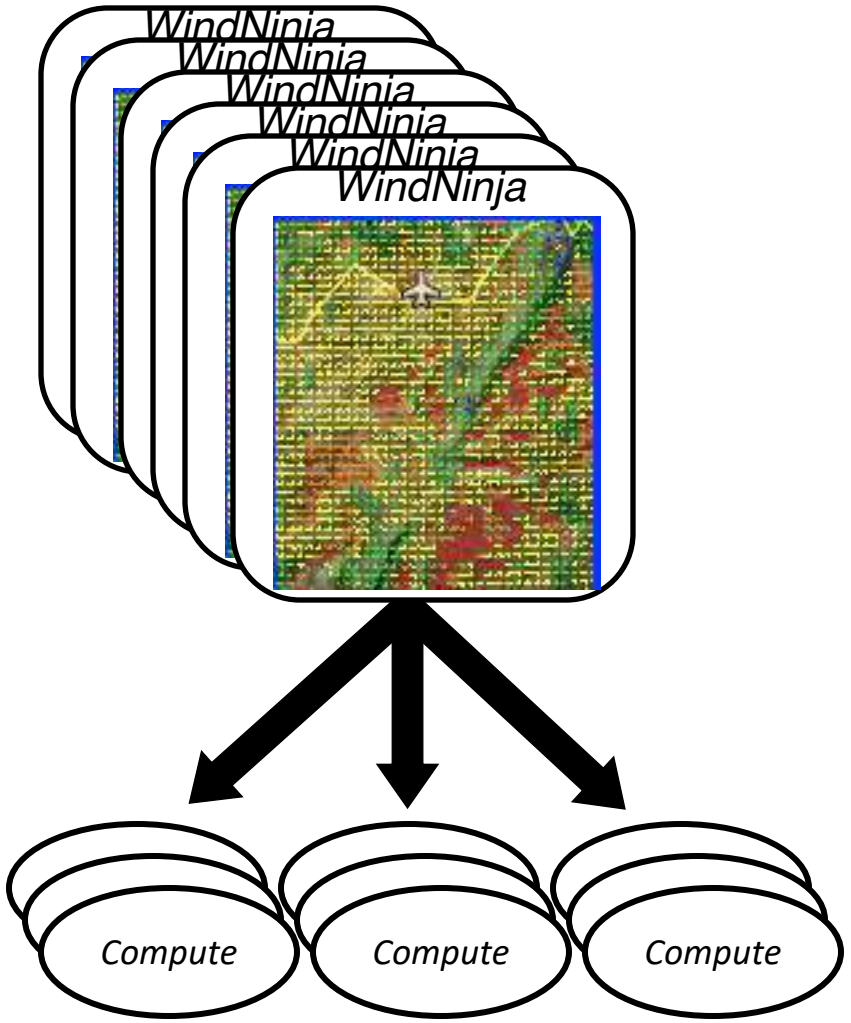
Temporal Coverage

- WindNinja calculates wind conditions for specific point in time
  - Run WindNinja for each timestamp



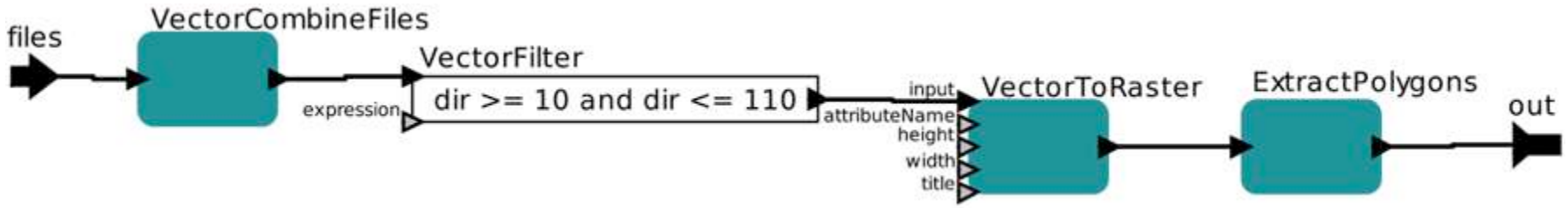
Execute in Parallel

- Run WindNinja for each tile
- Run WindNinja for each timestamp
- **Each execution is independent, so can be done in parallel**



Post-Processing WindNinja Output

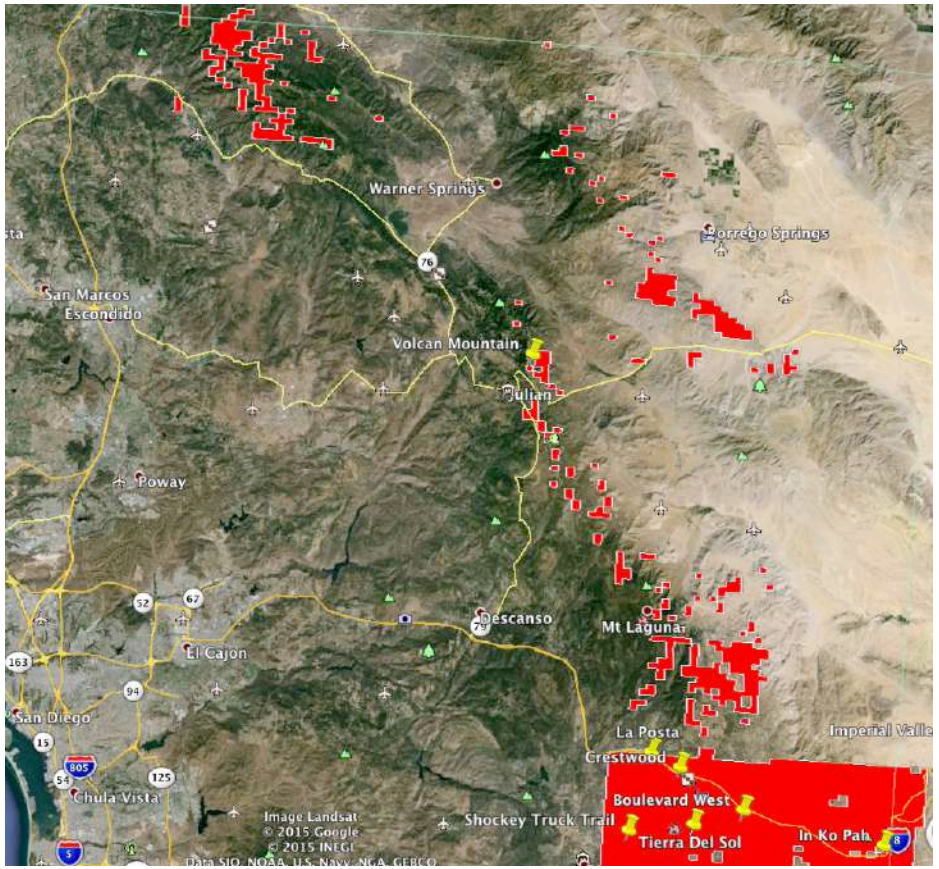
- WindNinja outputs wind direction and speed
- Process these outputs to find regions with Santa Ana winds



Union *Filter* *Rasterize* *Find Polygons*

Application Outputs

- Output shows Santa Ana regions
- Often much larger area surrounding weather station



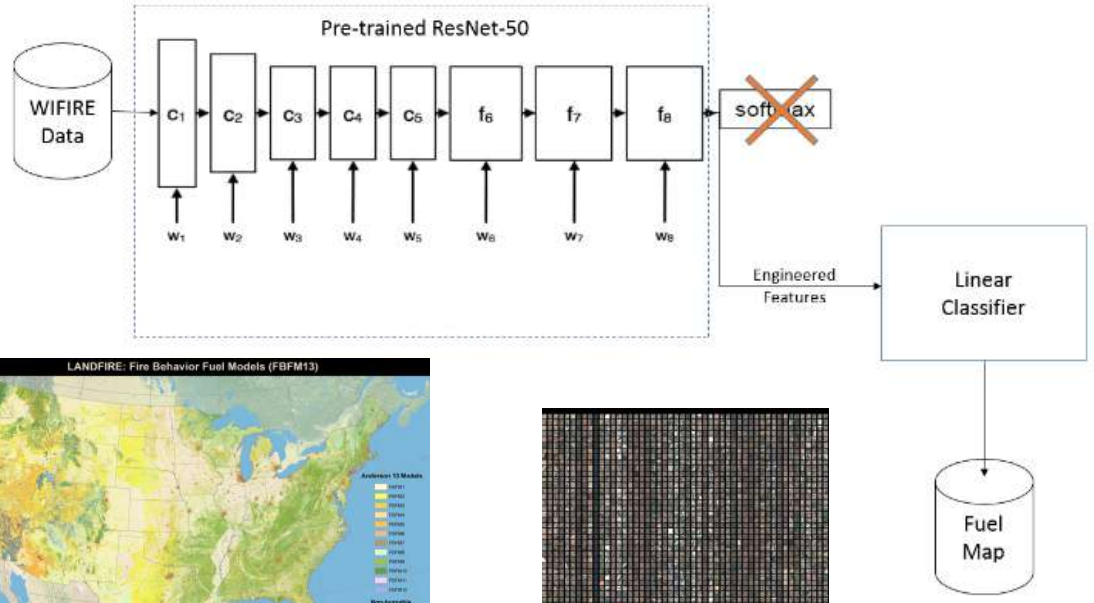
Some Machine Learning Case Studies

- Smoke and fire perimeter detection based on imagery
- Prediction of cific to location
- Prediction of eather history
- NLP for unde n radio communicati
- Deep learning h resolution fuel maps
- Classification e fuel maps (using Planet Labs satellite data)

All require periodic, dynamic and programmatic access to data!

Classification project to generate more accurate fuel maps

- Accurate and up-to-date vegetation maps are critical for modeling wildfire rate of speed, potential burn areas, and Targeting prevention strategies
- Challenge:
  - USGS Landfire provides the best available fuel maps every two years and is low resolution
- Approach:
  - Use high-resolution satellite imagery and deep learning methods to produce surface fuel
  - Create a classification model that will provide vegetation maps at greater frequency and resolution



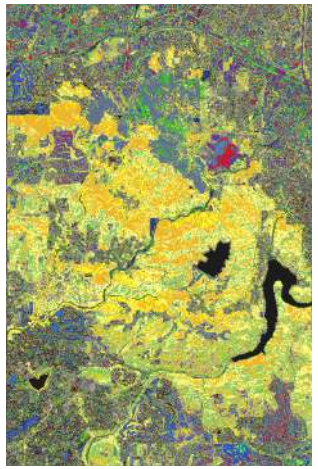
Cluster 1: Short Grass

Using AI for Integration of Planet Satellite Information to Monitoring and Prediction in WIFIRE

High-resolution satellite imagery

planet.

Vegetation classification



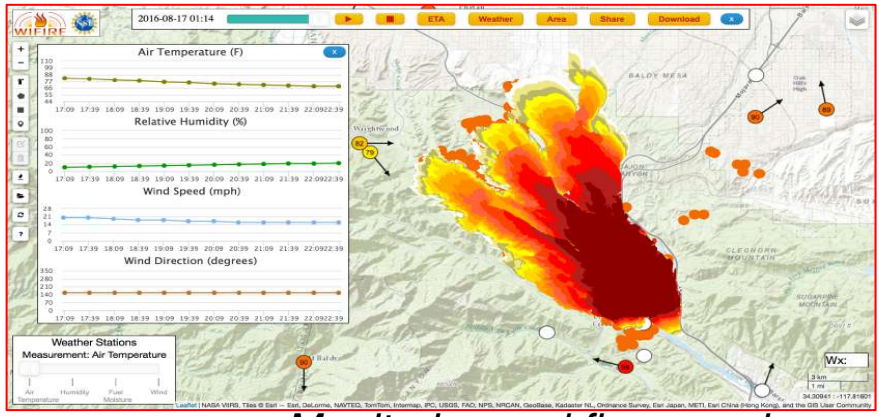
WIFIRE

Weather and cameras

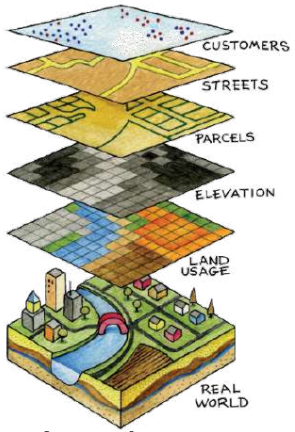
Fire perimeter

http://firemap.sdsc.edu

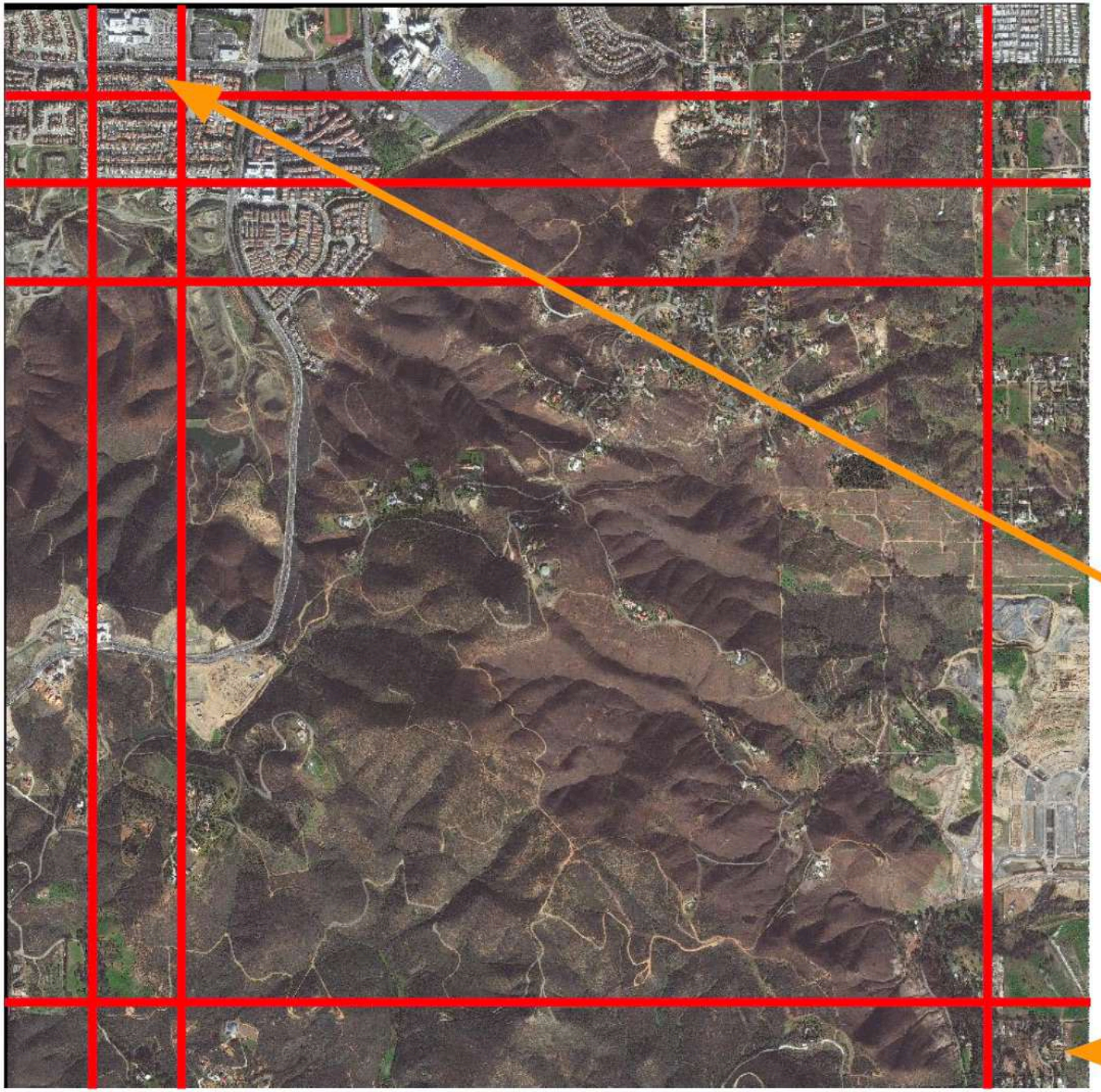
WIFIRE

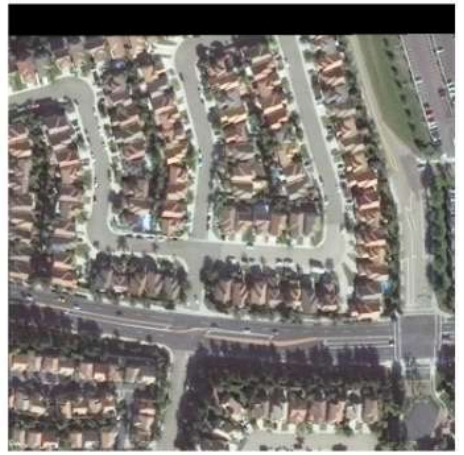


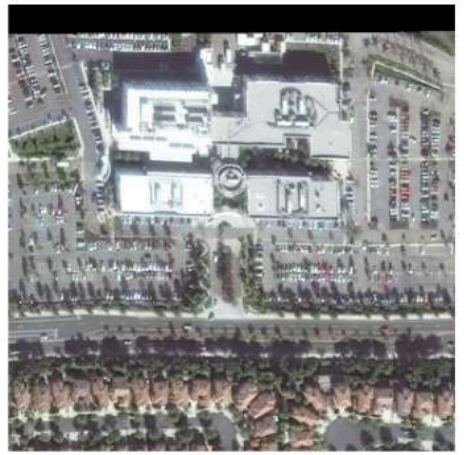
Monitoring and fire mapping

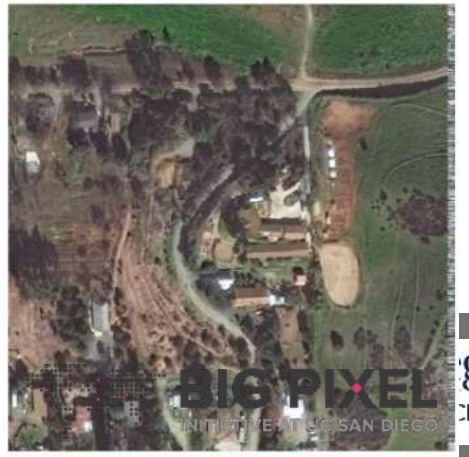


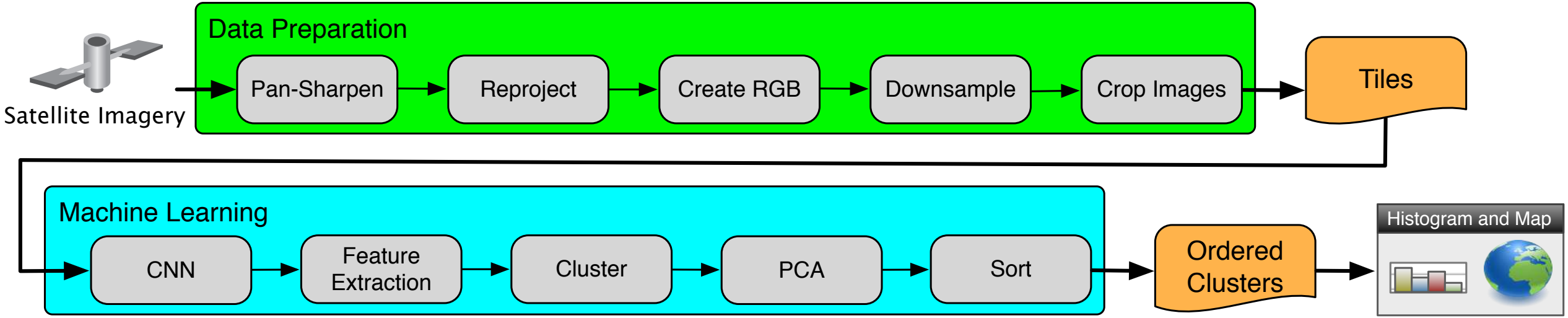
Landscape



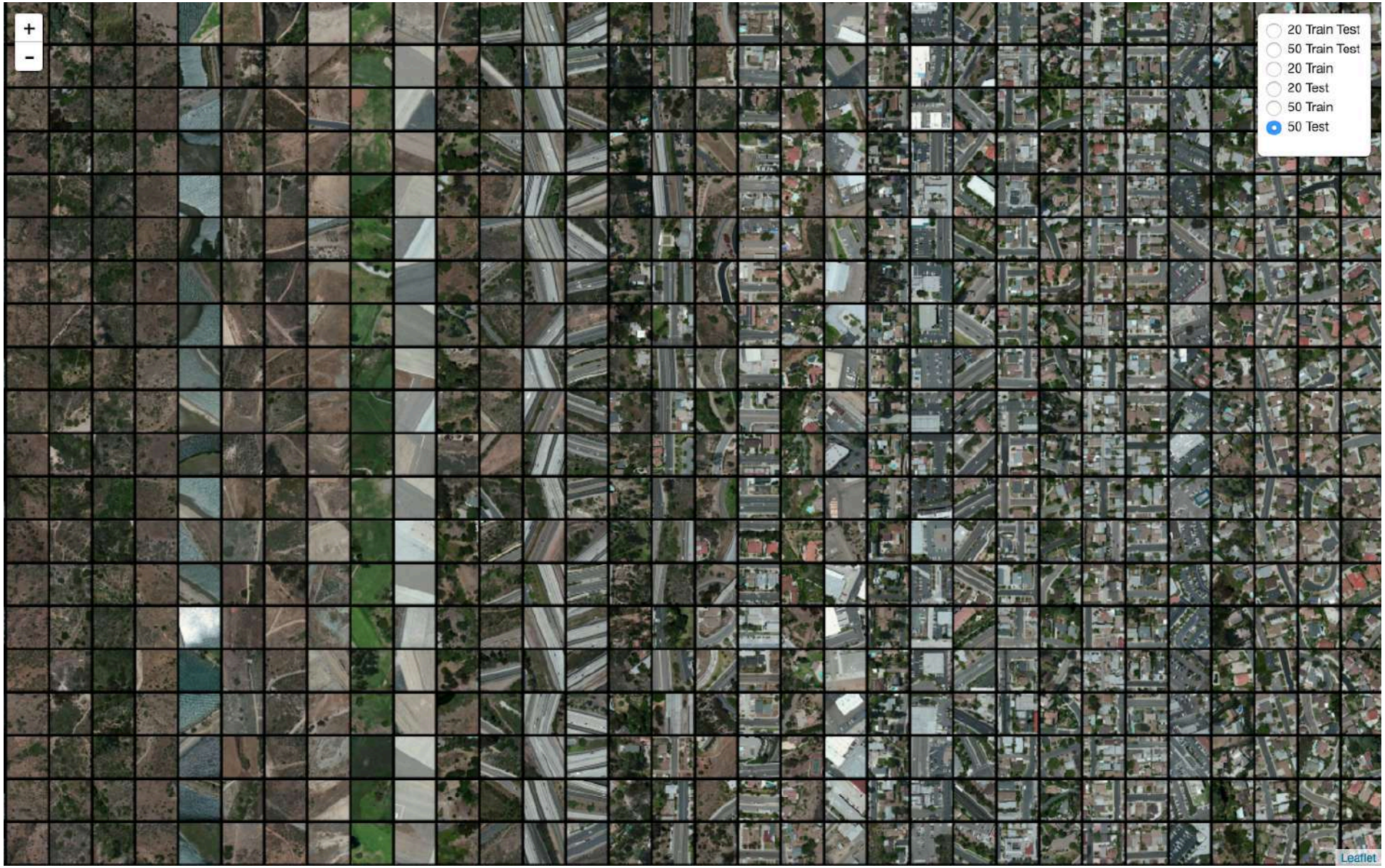






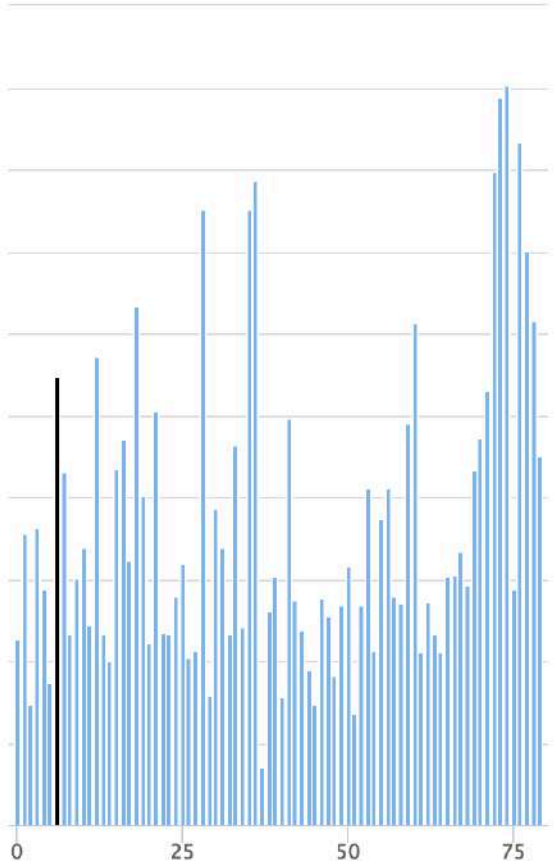


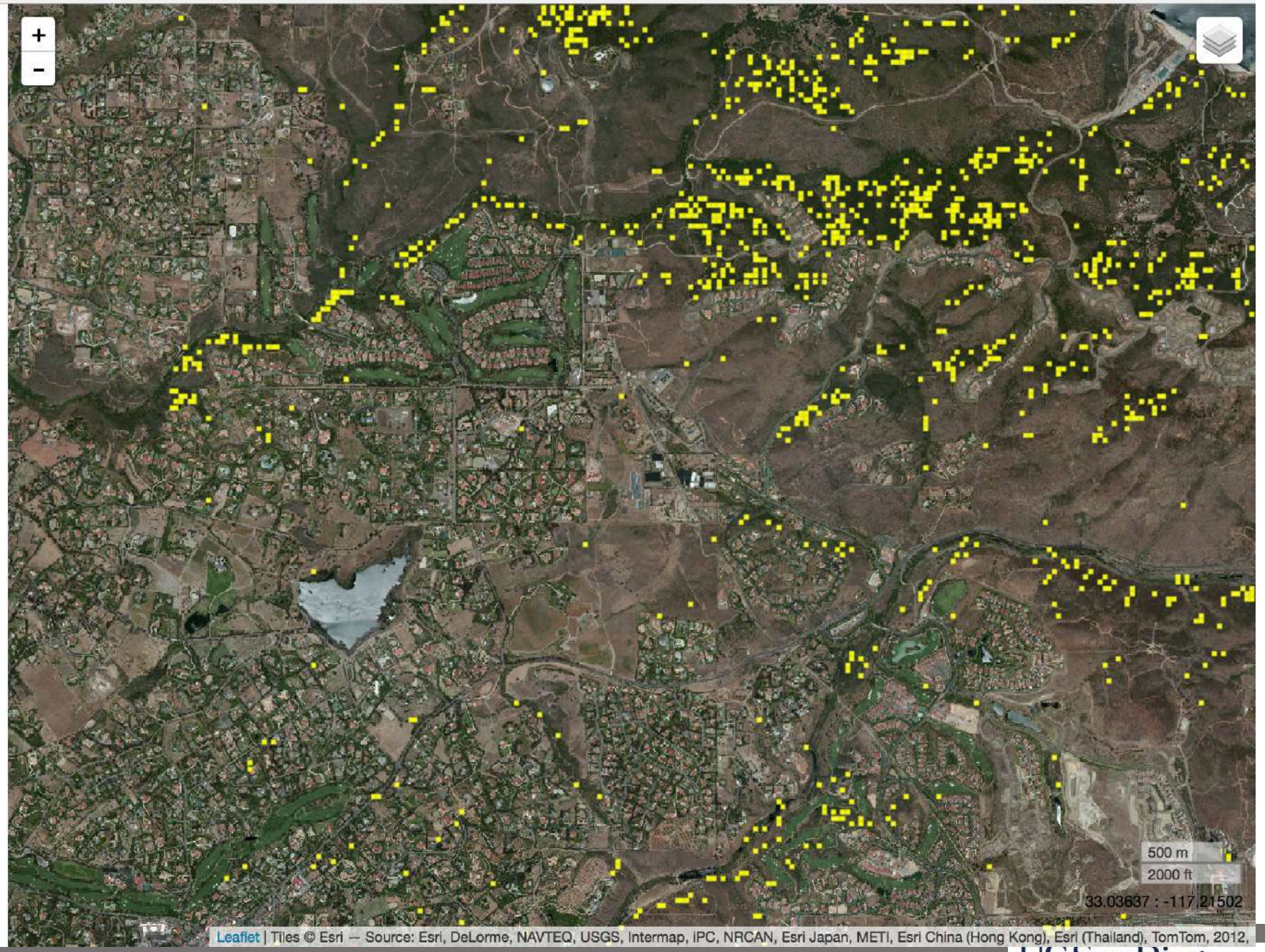
- *Resnet 152 CNN Architecture*
- *Caffe and SkiCaffe for feature extraction*
  - *Yields features vectors that are 1,000 dimensional*
- *k-means to cluster images based on features*
  - *We train the k-means on 10% of randomly selected images from the dataset*
- *PCA orders the clustering*


+
-
20 Train Test
50 Train Test
20 Train
20 Test
50 Train
50 Test
Leaflet

Cluster 6

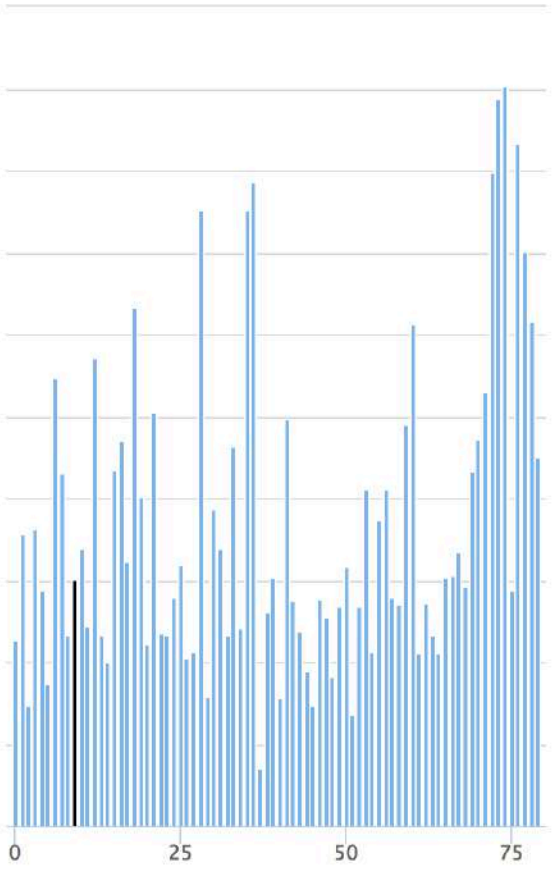
Click to lock, click twice to unlock.

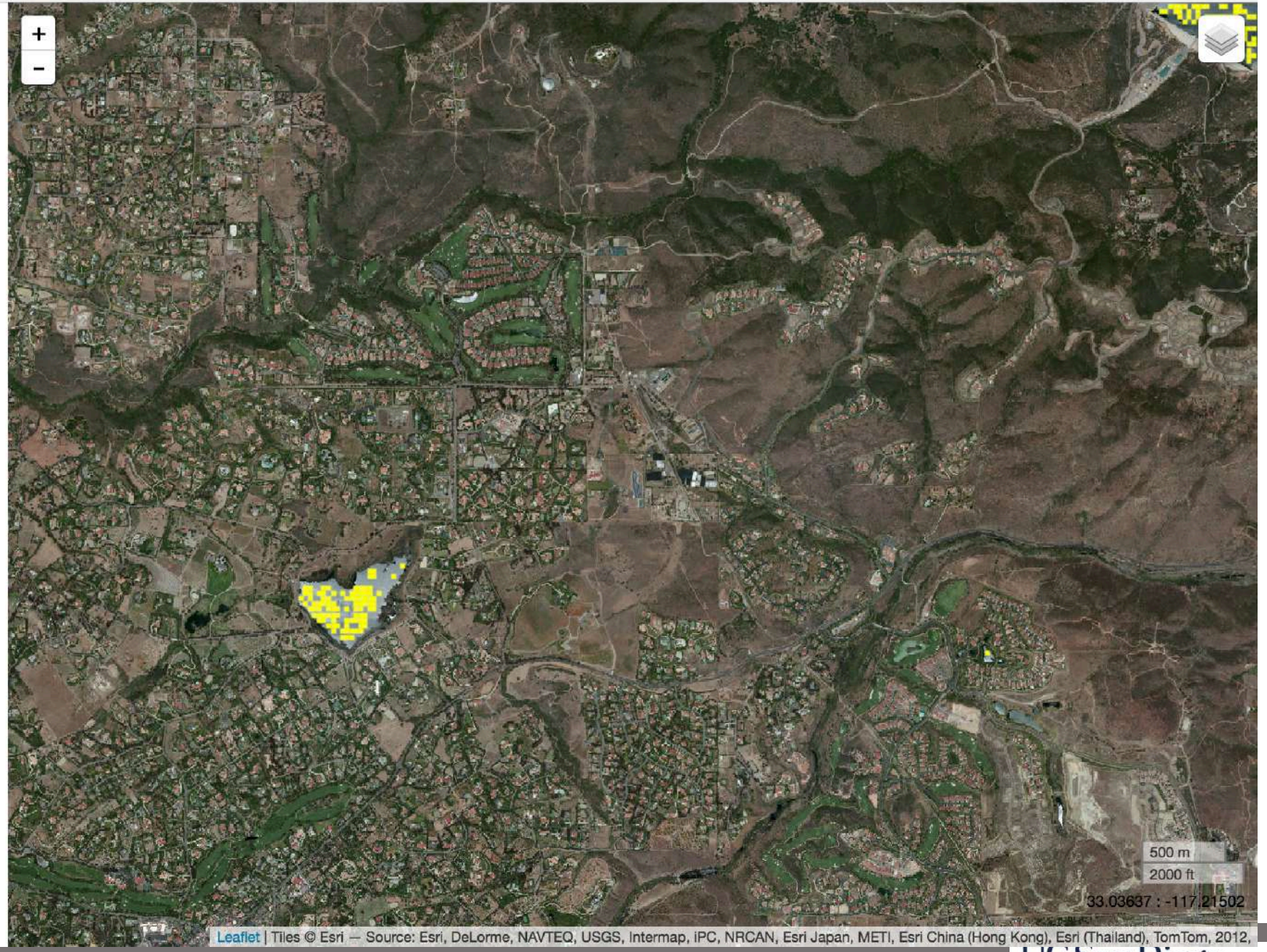




Cluster 9

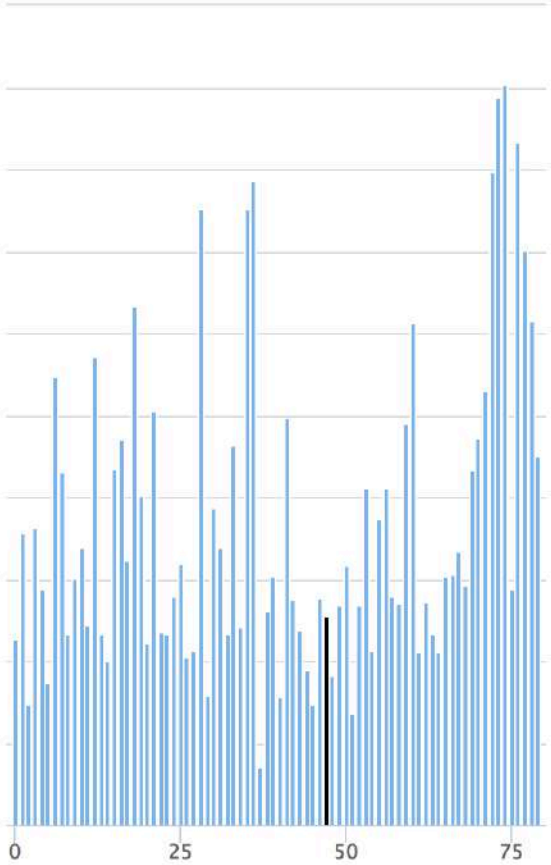
Click to lock, click twice to unlock.

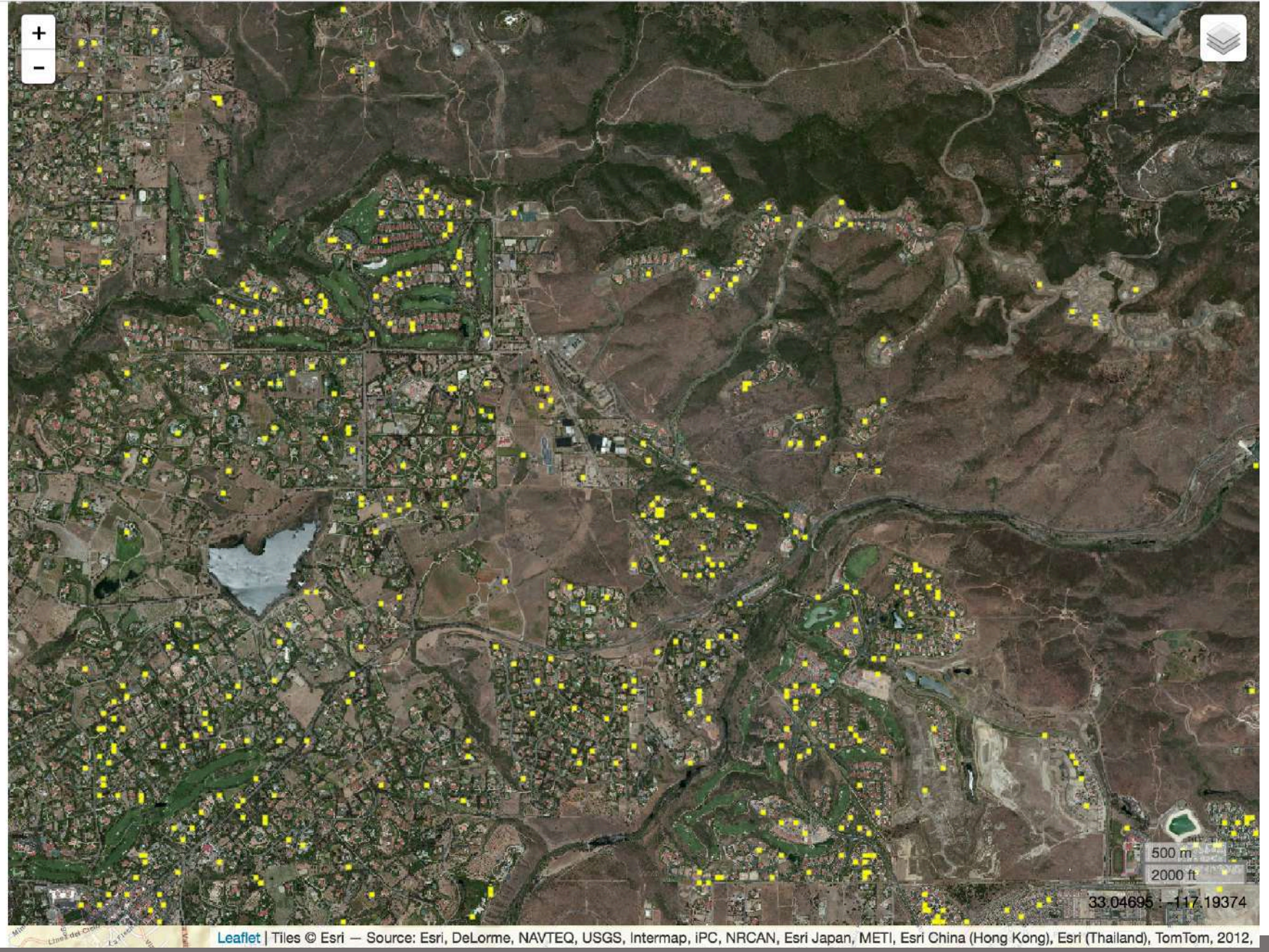




Cluster 47

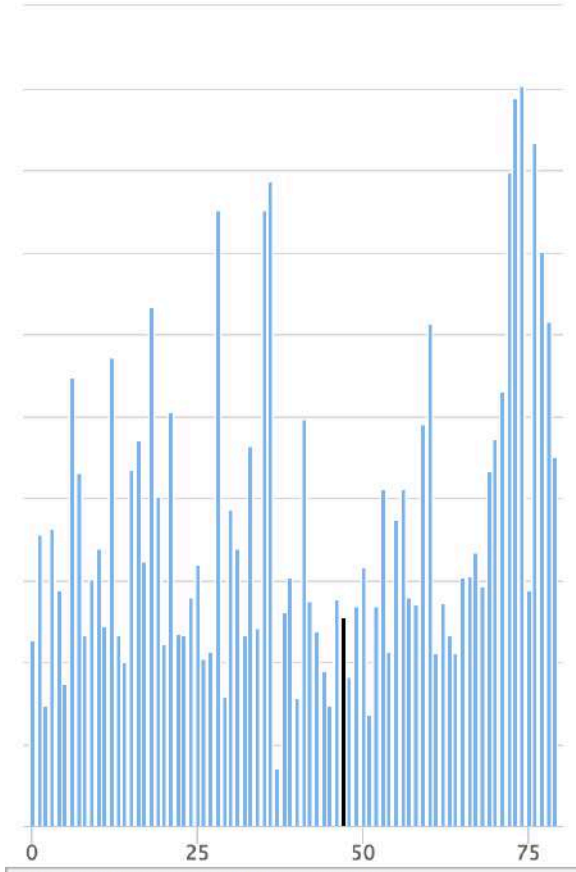
Click to lock, click twice to unlock.

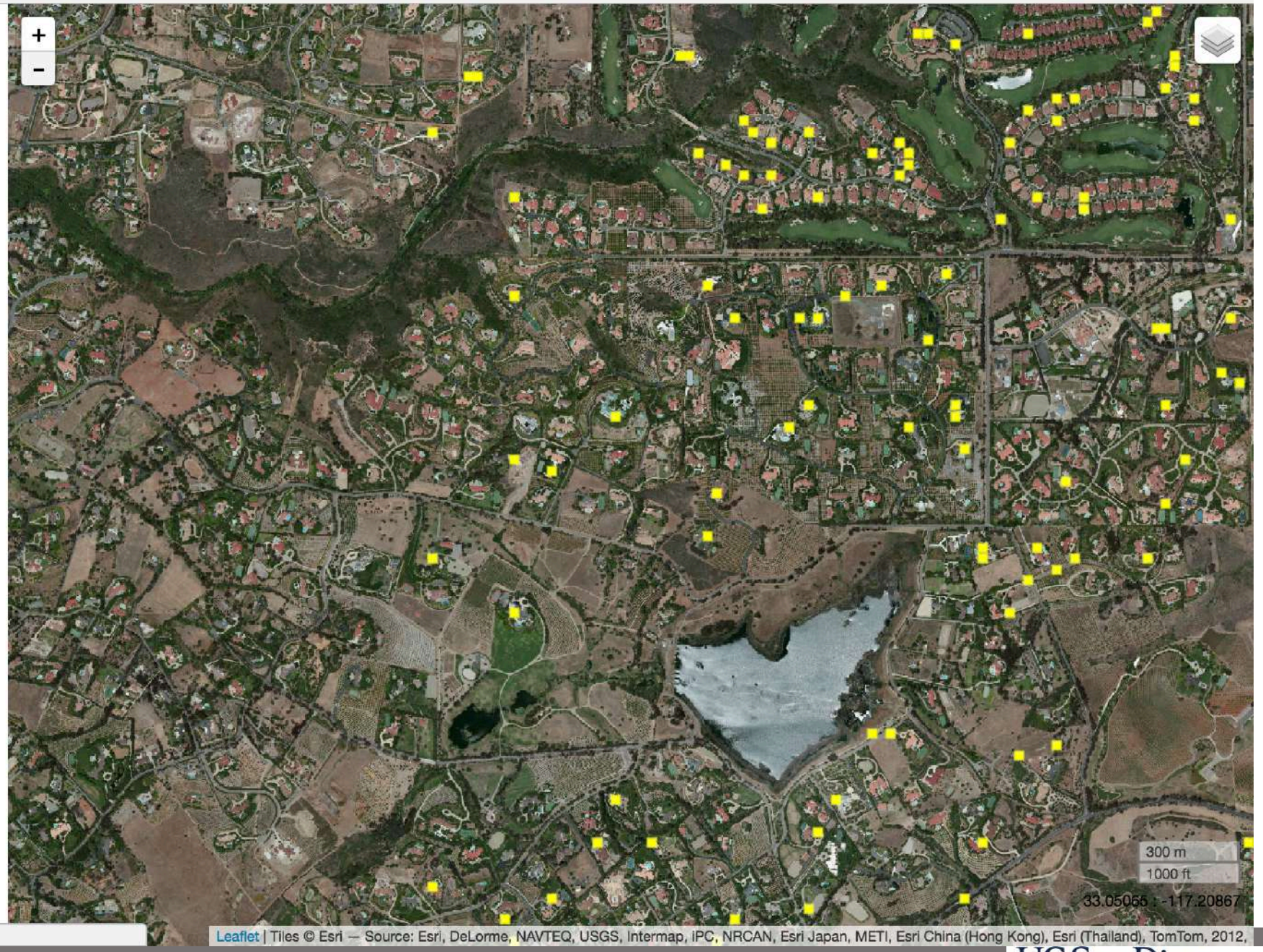




Cluster 47

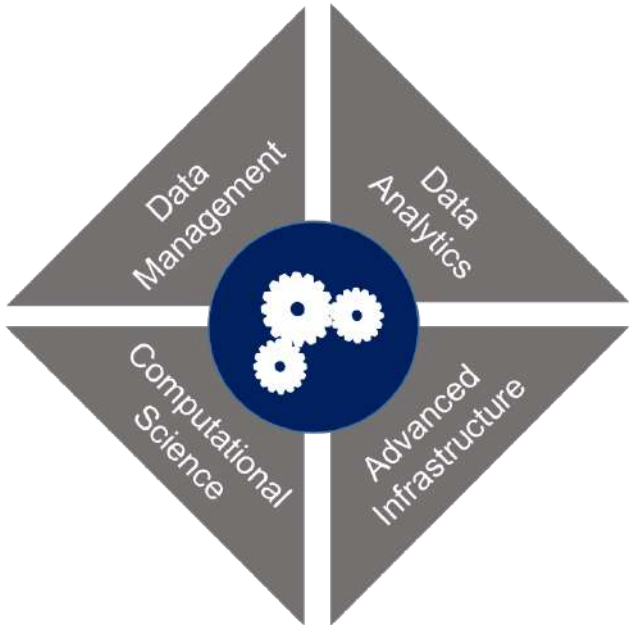
Click to lock, click twice to unlock.





https://satellite.sdsc.edu/mas/map.html#

Using workflows for process integration and maps as an enabling interface.



Kepler

https://kepler-project.org/

SECURITY

PROVENANCE

DATA LIFECYCLE MANAGEMENT

ENABLING INTERFACES
e.g., Gateways and other online tools for research and teaching

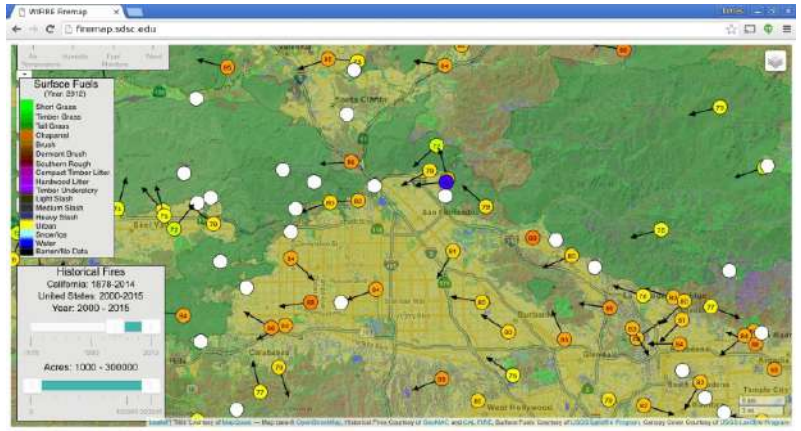
WORKFLOW MANAGEMENT
e.g., Application Integration, Coordination, Optimization, Communication, Reporting

COMPOSABLE SERVICES
e.g., Model and Data Archives, Deep Learning, Analytics, HPC, Training, Notebooks

RESOURCE MANAGEMENT
e.g., Kubernetes Container Cloud

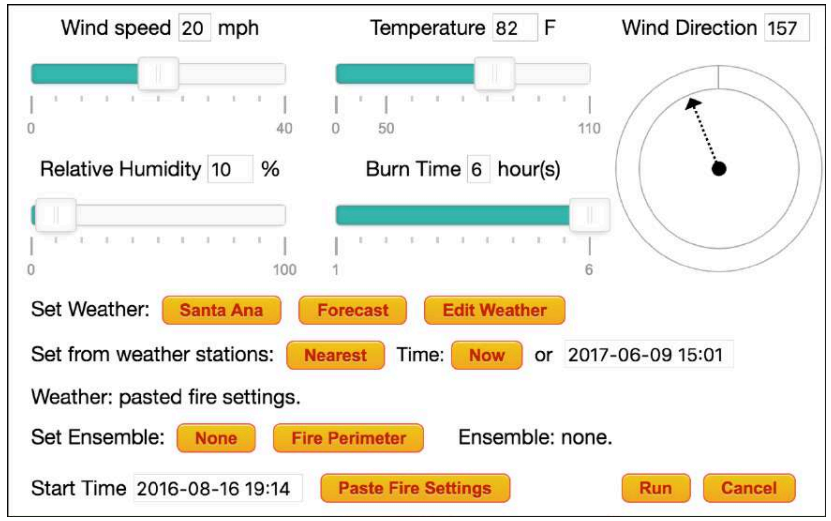
COMPOSABLE SYSTEMS
e.g., GPU, CPU, Big Data, Neuromorphic, Networks, Storage, …

COLLABORATION MANAGEMENT



Firemap Tool

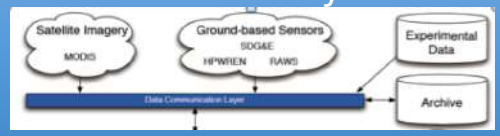
http://firemap.sdsc.edu



- A web-based GIS environment:
  - access information related to fire behavior
  - analyze what-if scenarios
  - model real-time fire behavior
  - generate reports
- Powered by WIFIRE
- Developed in partnership with LAFD

Firemap Web Interface

WIFIRE Data Interfaces

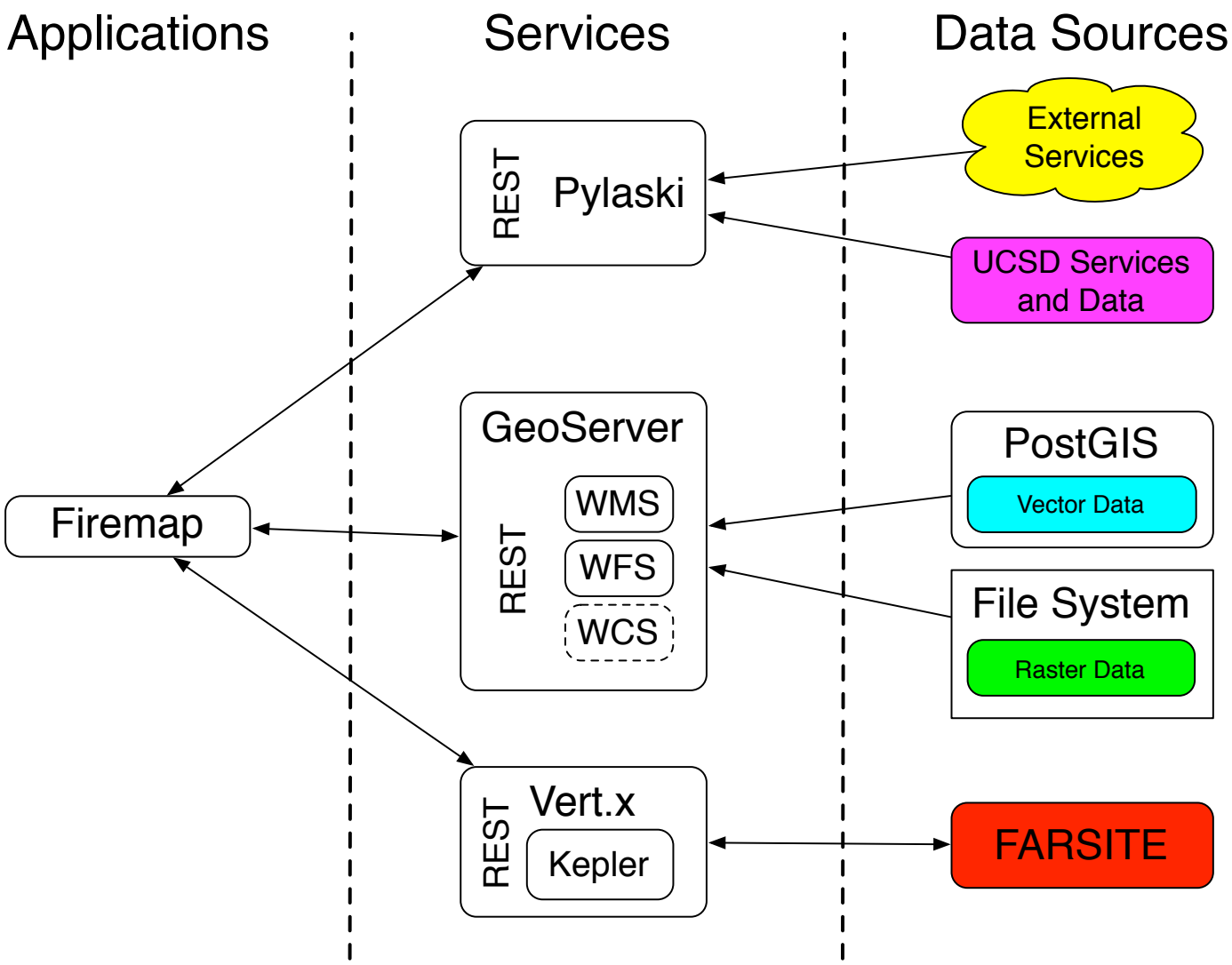


WIFIRE Workflows

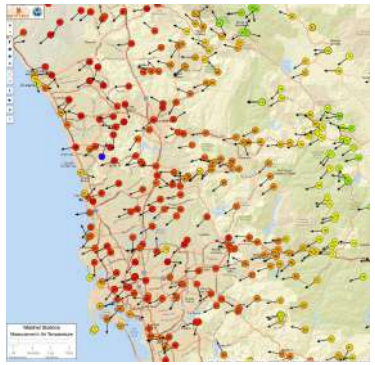
Computing Infrastructure

Firemap Overview

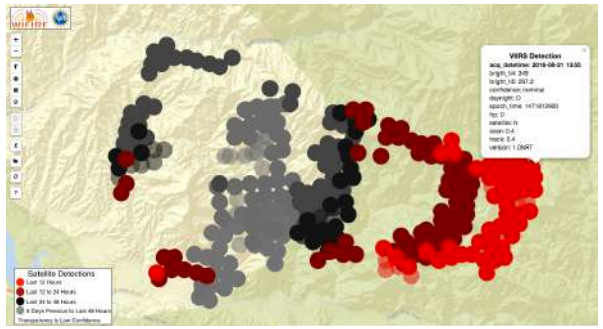
Architecture



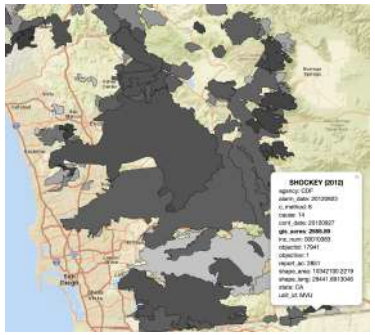
Firemap Layers



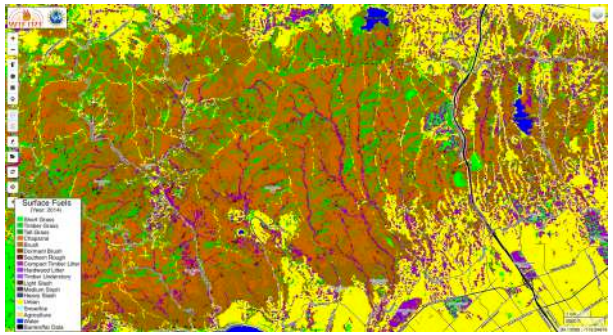
(a)



(b)

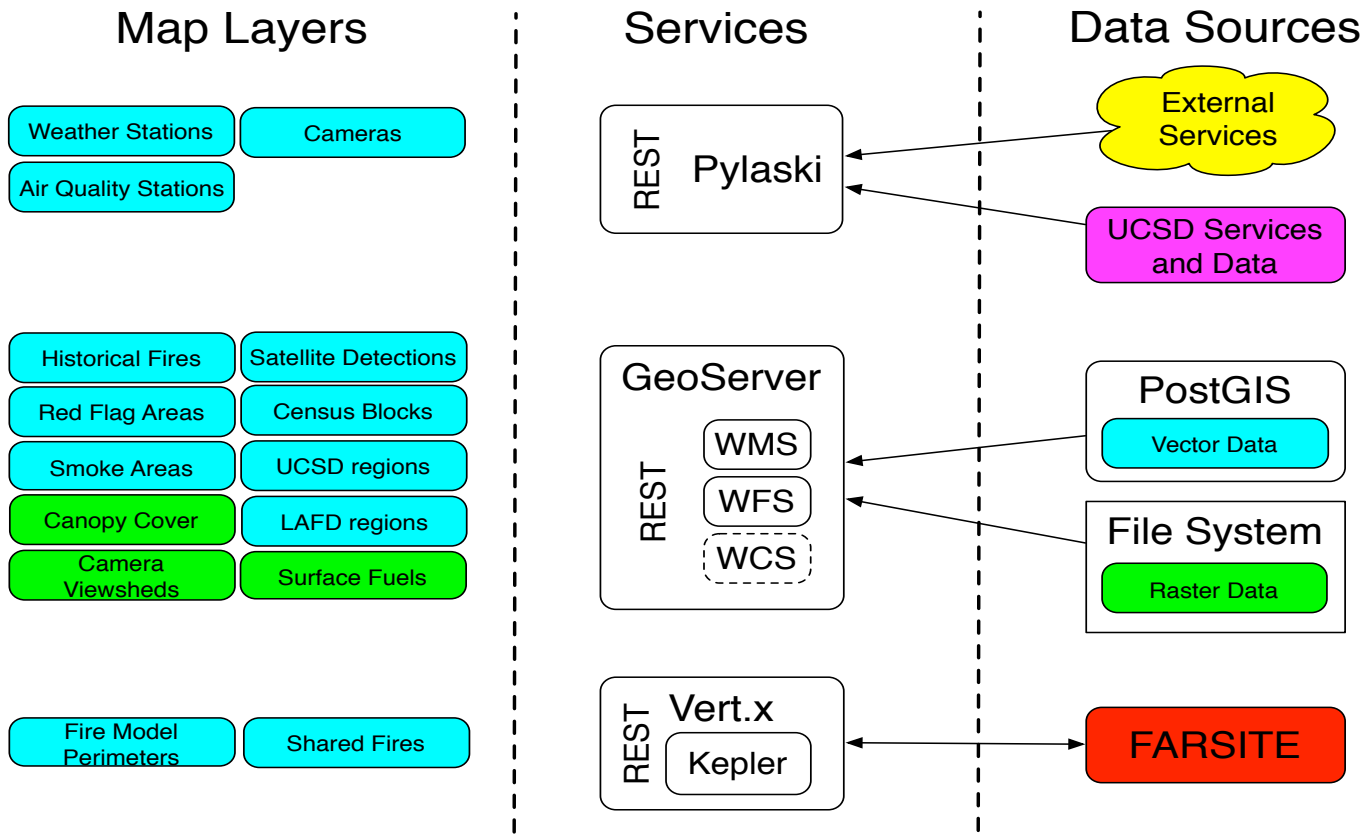


(c)



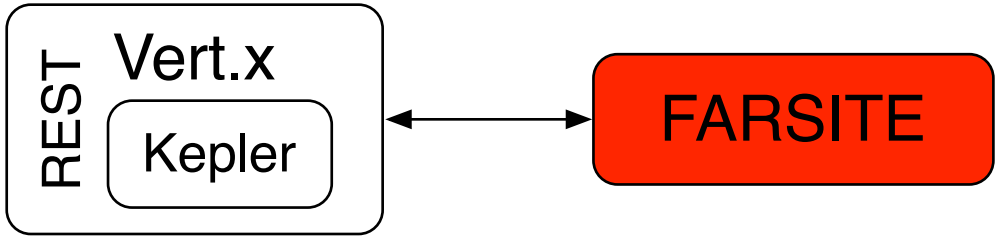
(d)

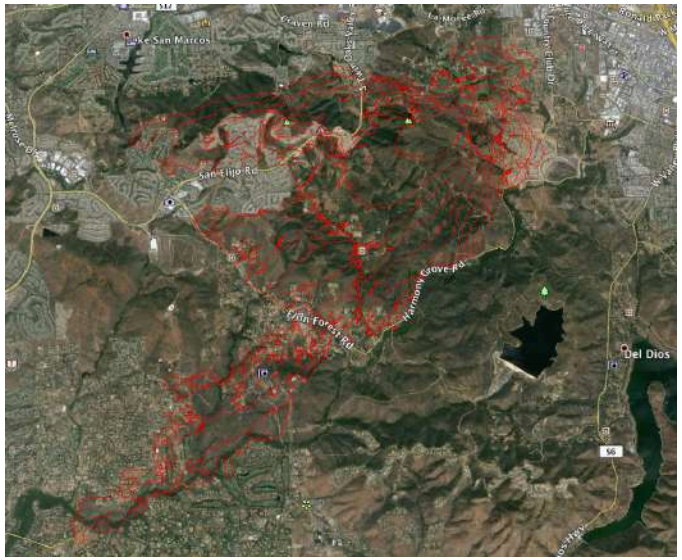
Figure 2: Firemap layers: (a) Weather Stations Layer showing MesoWest stations in San Diego County during Santa Ana conditions on 26 September, 2016; (b) VIIRS Layer showing 7 days of thermal detections of the Rey Fire, north of Santa Barbara, CA, during 14-21 August, 2016; (c) Historical Fires layer showing wildfires in San Diego County during 1970-2015; and (d) Surface Fuels Layer showing the vegetation fuels for the canyons northwest of Los Angeles in 2014.



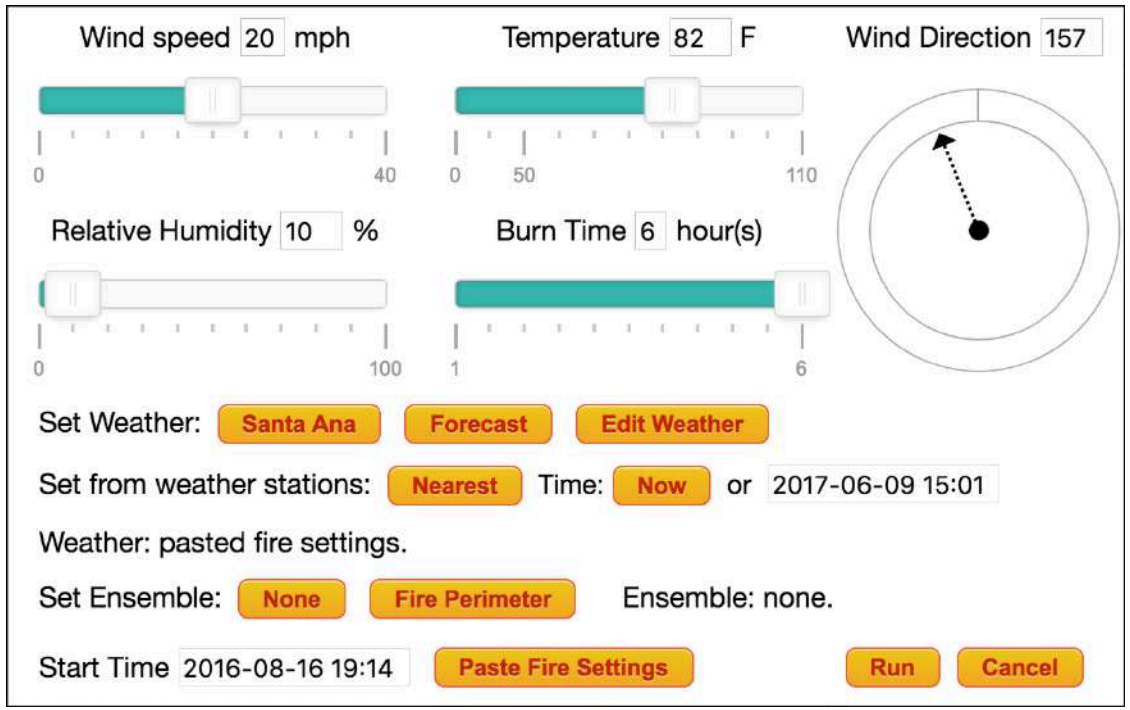
Kepler WebView

- REST service to run Kepler workflows
- FARSITE workflow
  - Inputs: ignition, weather, wind, landscape
  - Outputs: perimeters

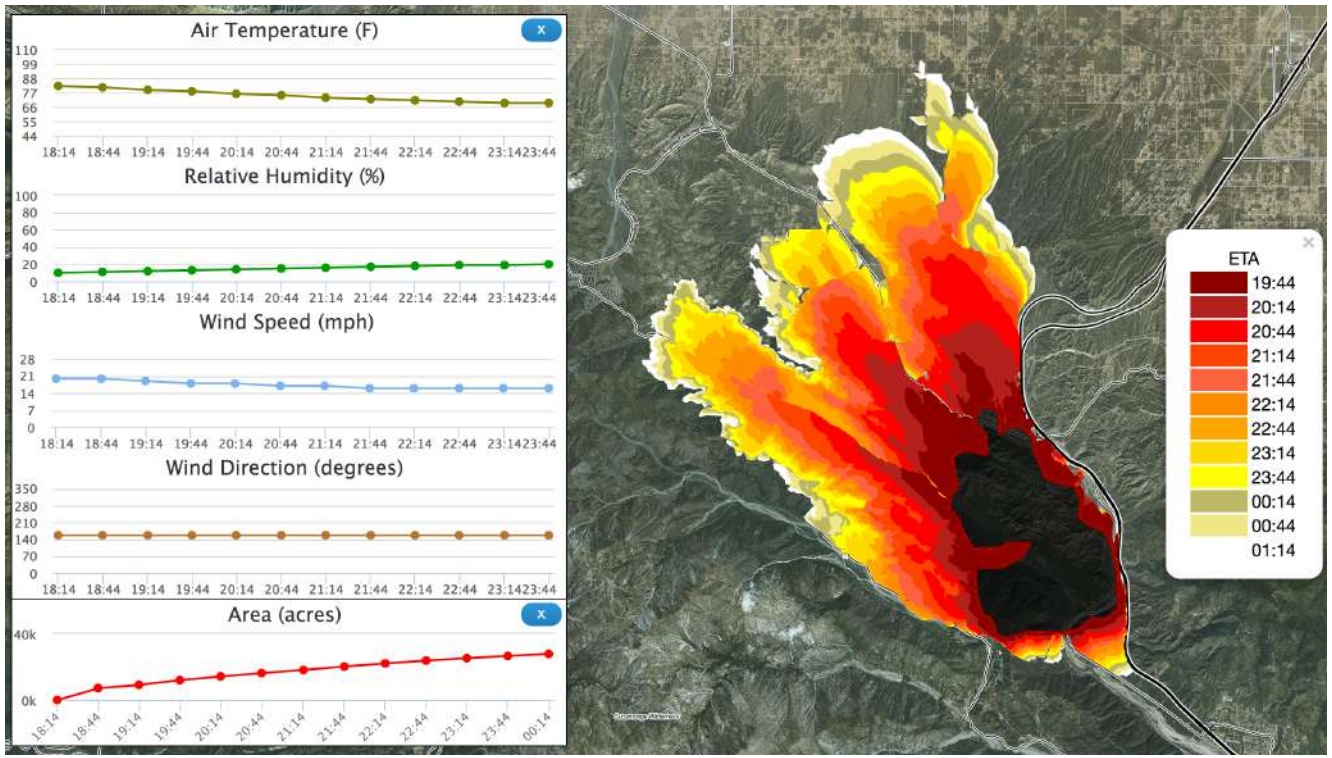




Firemap Predictive Modeling

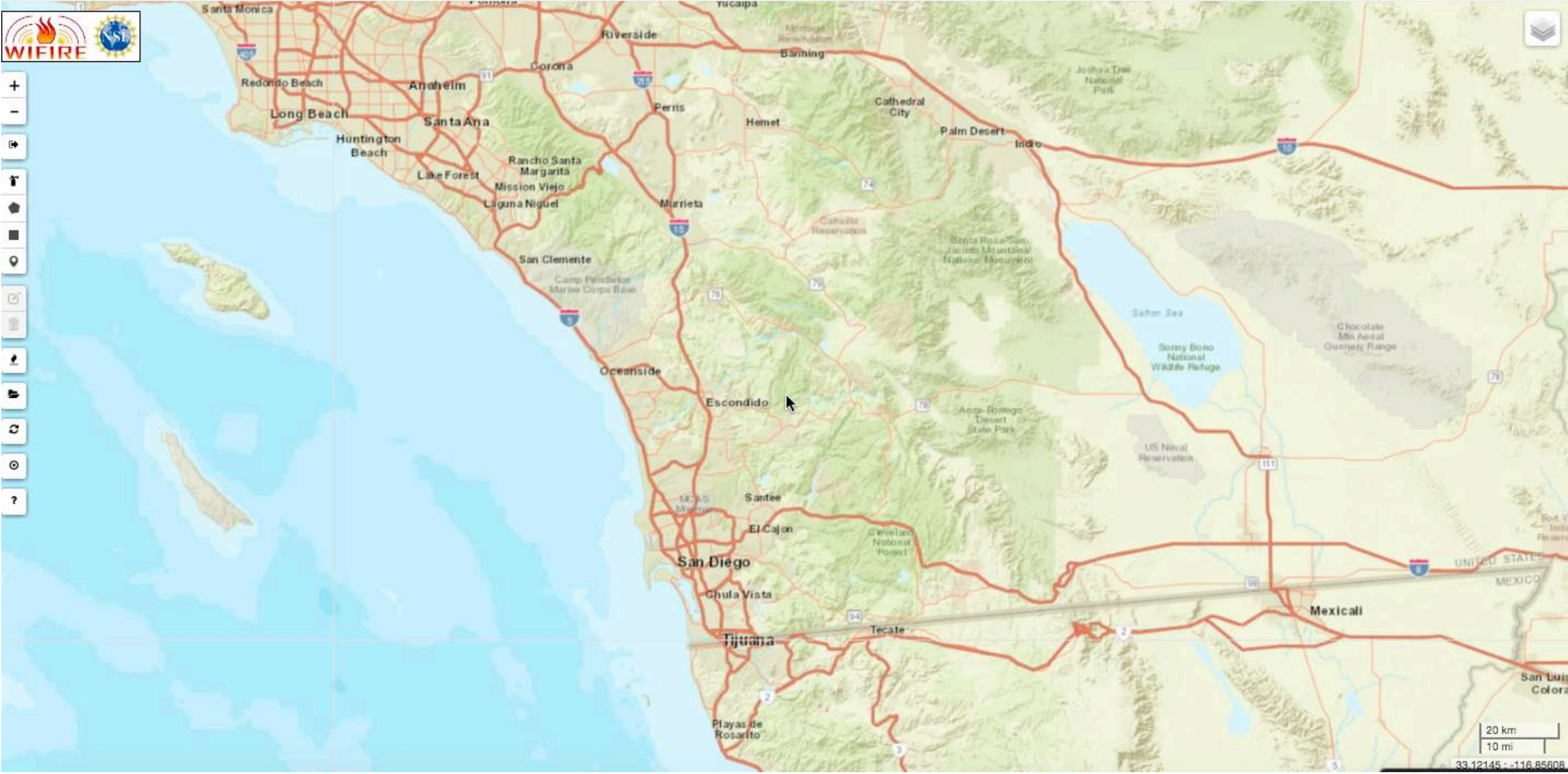


(a)

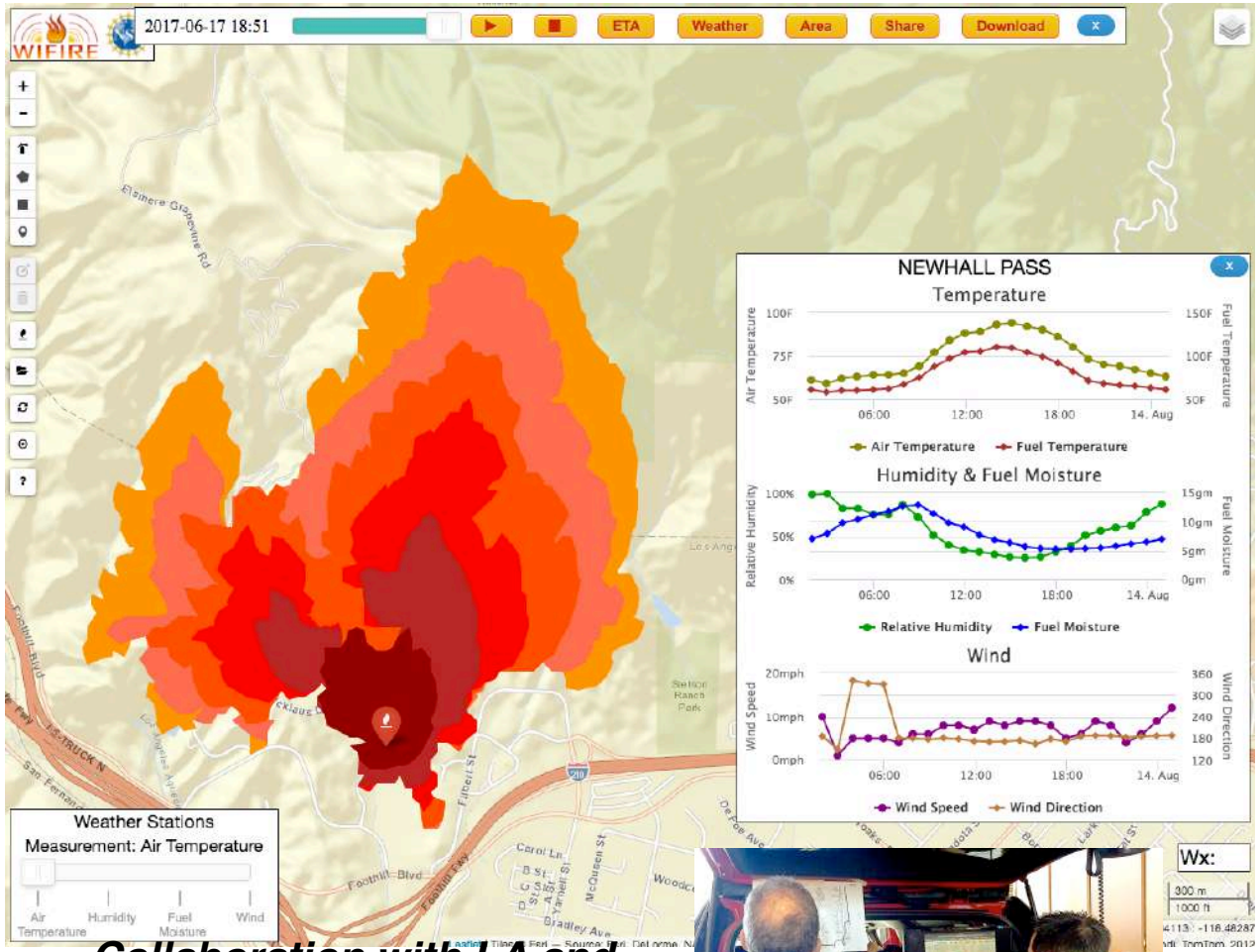


(b)

Figure 3: Firemap predictive modeling: (a) the configuration dialog specifying the simulation time, and weather conditions; (b) predicted fire spread for the Blue Cut Fire showing the weather conditions, fire perimeters, area, and arrival times.


WIFIRE
Santa Monica
Redondo Beach
Long Beach
Anaheim
Santa Ana
Huntington Beach
Lake Forest
Rancho Santa Margarita
Mission Viejo
Laguna Niguel
Riverside
Corona
Perris
Hemet
Banning
Cathedral City
Palm Desert
Indio
Murrieta
San Clemente
Camp Pendleton Marine Corps Base
Oceanside
Escondido
Santee
El Cajon
San Diego
Chula Vista
Tijuana
Tecate
Mexicali
Playas de Rosarito
Joshua Tree National Park
Cahuilla Reservation
Santa Rosa San Jacinto Mountains National Monument
Salton Sea
Sonny Bono National Wildlife Refuge
Anza-Borrego Desert State Park
US Naval Reservation
Cleveland National Forest
Chocolate Mts Aerial Gunnery Range
20 km
10 mi
33.12145 : -116.85608

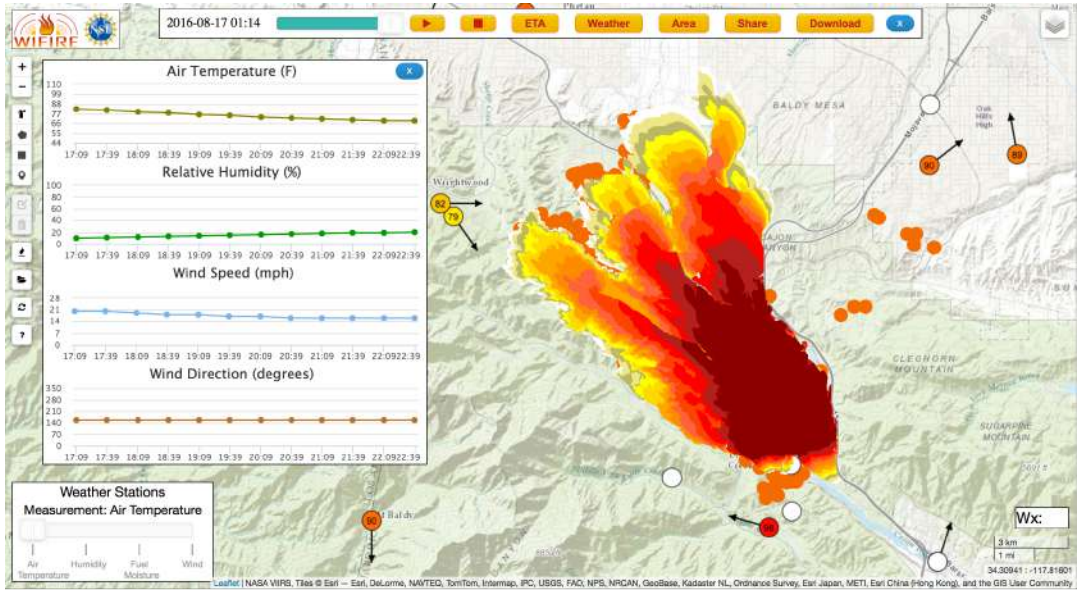
Data-Driven Fire Progression Prediction Over Three Hours

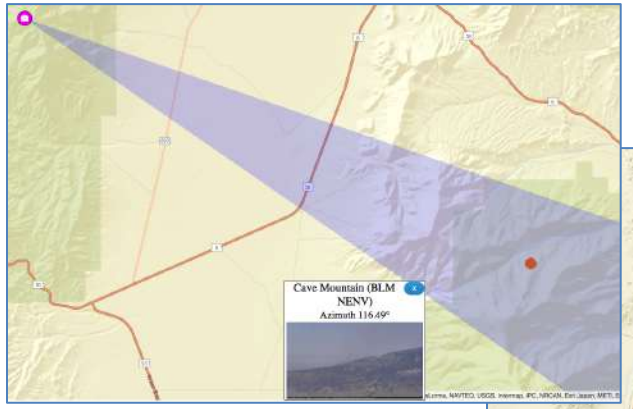


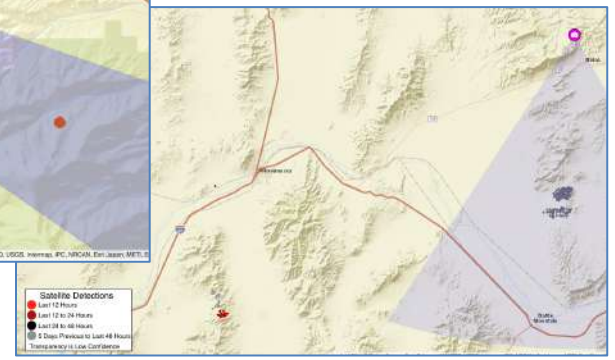
Collaboration with LA and SD Fire Departments

http://firemap.sdsc.edu

August 2016 – Blue Cut Fire

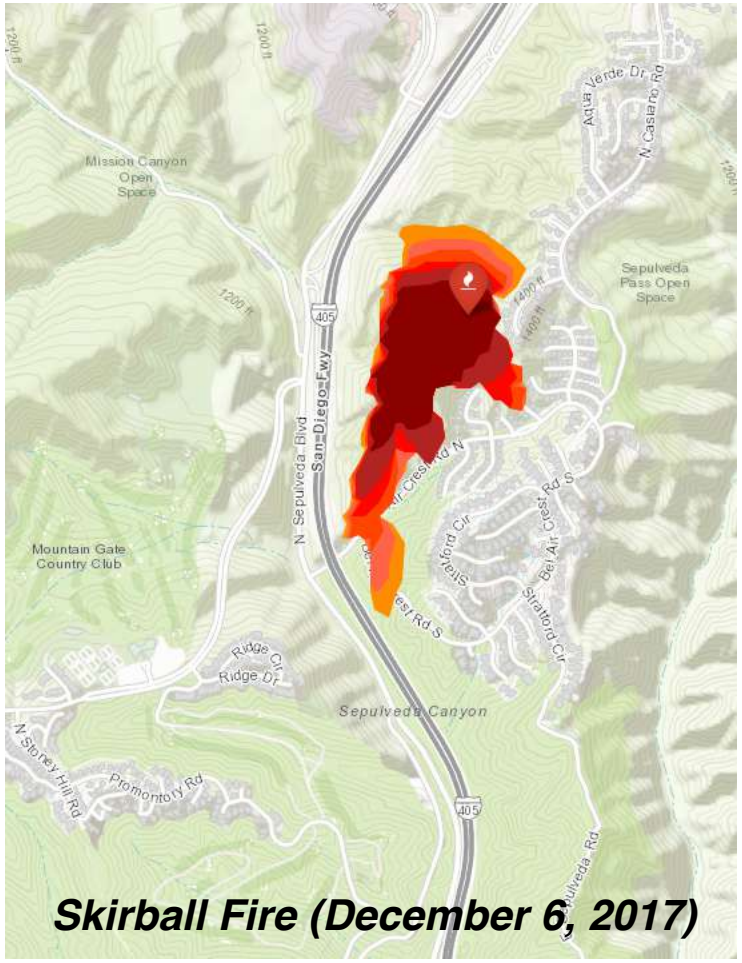






Tahoe and Nevada Bureau of Land Management Cameras: 20 cameras added with field-of-view

Los Angeles Fires in December 2017

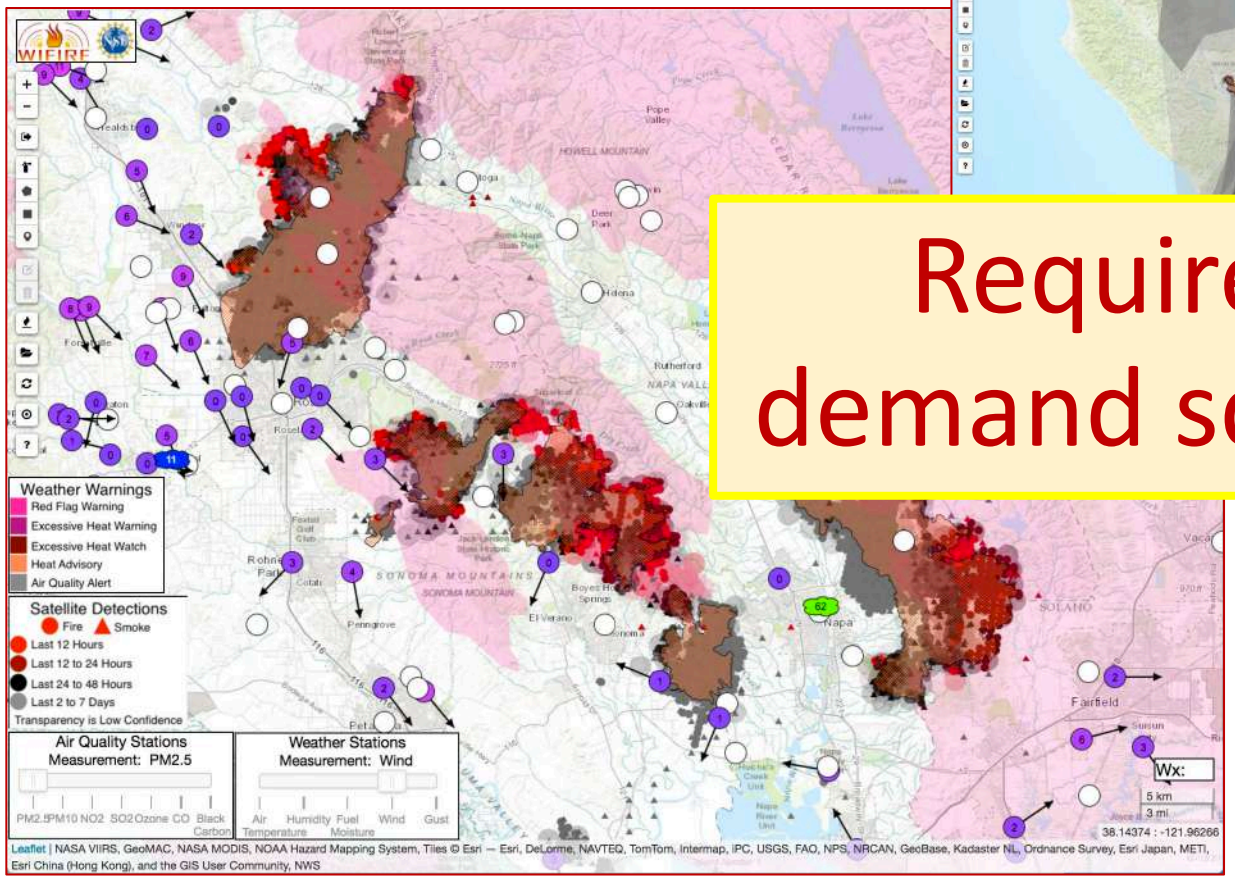


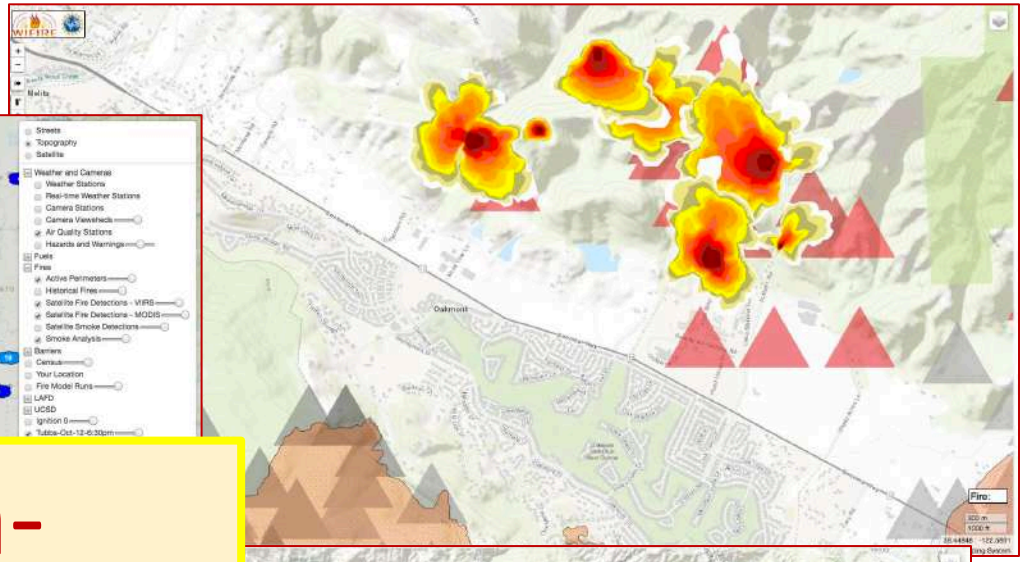
Creek Fire (December 5, 2017)

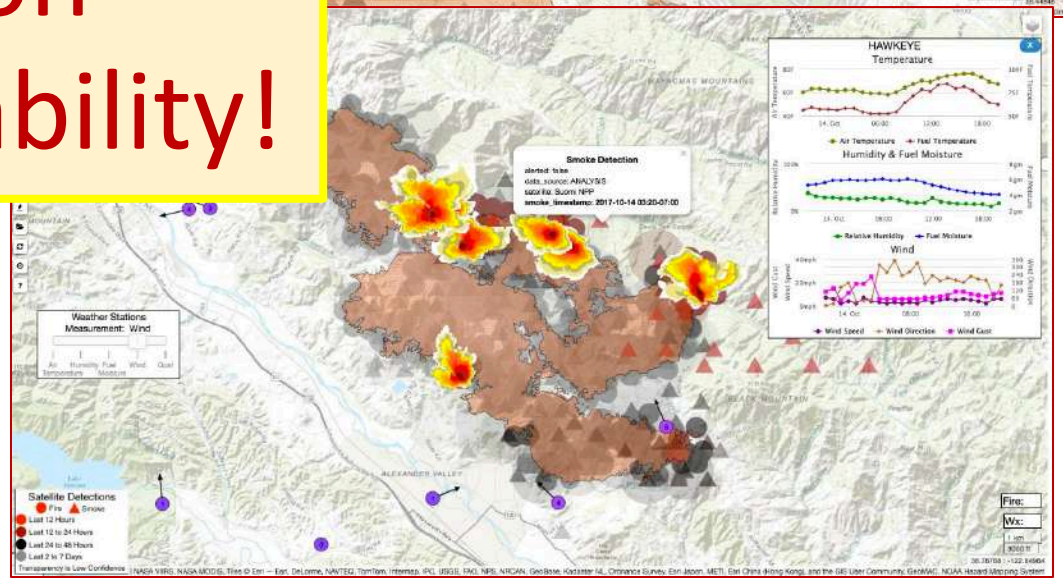
CA Fires 10/2017 through 12/2017

800K+ unique visitors and 8M+ hits

http://firemap.sdsc.edu







Required on-demand scalability!

NSF HPWREN, NASA Satellites and other public information

NSF

WIFIRE

Data Lifecycle Management

Data Science

Computational Science

Advanced Infrastructure

Real-time and periodical geospatial data reduction, curation and transformation

Kepler

PPoDS

jupyter

GDAL

Keras

TensorFlow

Spark

WMS

WFS

WCS

OGC

KML

GML

SLD

NSF Comet, PRP, XSEDE, CHASE-CI, Kepler CORE, REAP, GEON, SEEK, RoadNET, bioKepler, DISCOSci, …

Ability to combine fire modeling tools and data assimilation capabilities

https://firefly.cerfacs.fr/en/system-overview/

hadoop

NVIDIA

amazon web services

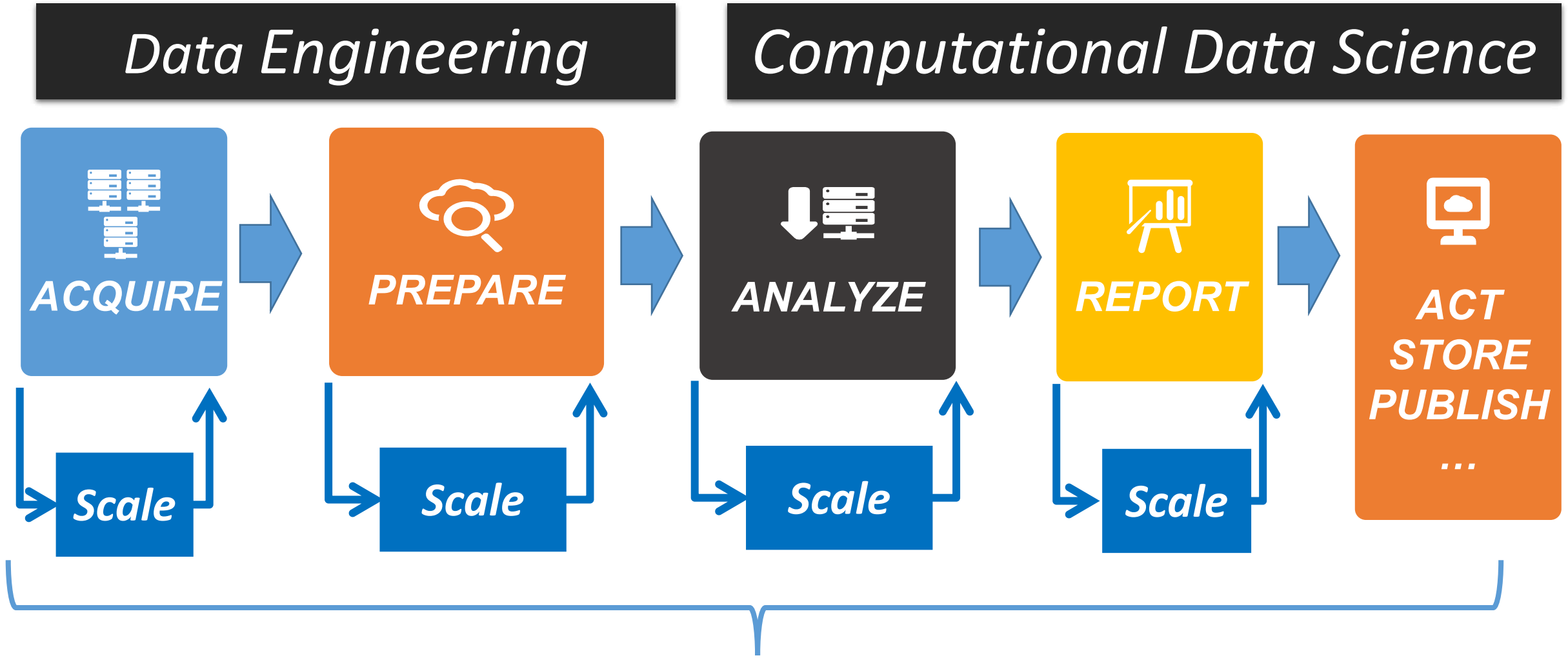
Spark

Application integration requires the expertise of a collaborative team.

- Multi-disciplinary scientific and technological expertise
- Integration of many scales of computing
- Integration of big and small experimental datasets
  - Can be historical or real time
- Usage of individual or community developed legacy tools
- Methods to manage and interpret data
- Modeling and simulation
- Gateways for visualization and dashboard
- Long term active and passive storage needs

An Ecosystem that Enables Needs and Best Practices

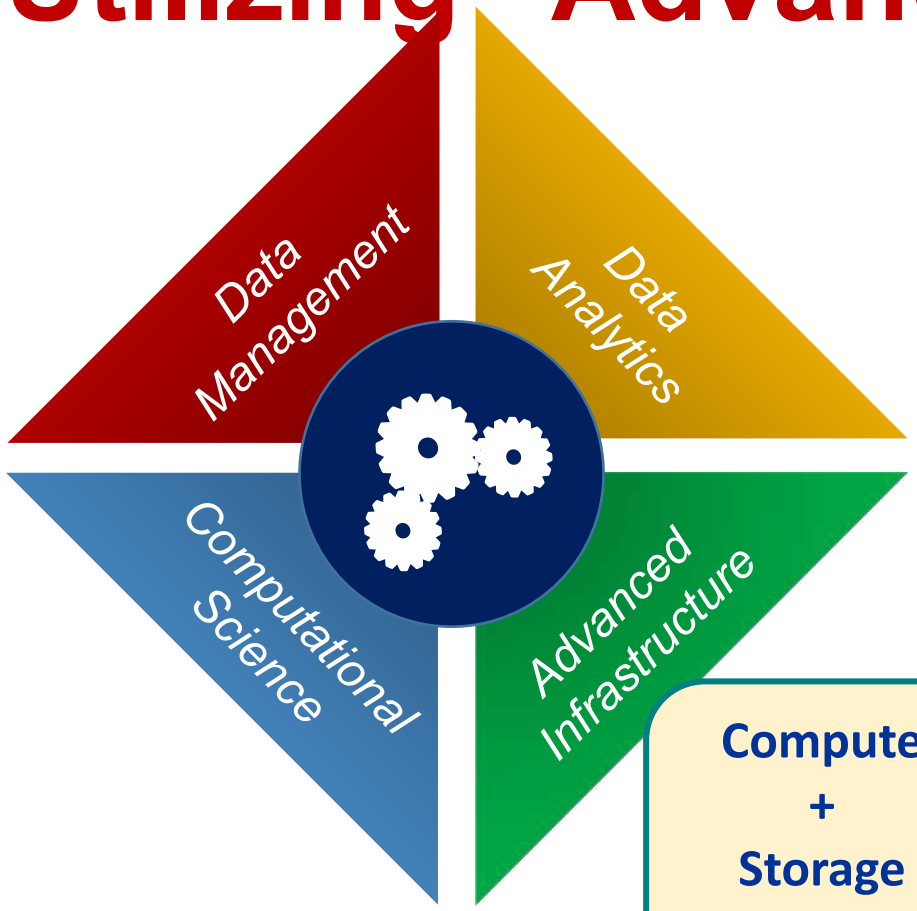
- data-driven
- scalable
- dynamic
- process-driven
- collaborative
- accountable
- reproducible
- interactive
- heterogeneous
- includes many different expertise


Data Engineering
Computational Data Science
ACQUIRE
PREPARE
ANALYZE
REPORT
ACT
STORE
PUBLISH
...
Scale
Scale
Scale
Scale
Continuous Iteration, Integration and Programmability

Continuous Iteration, Integration and Programmability at the Continuum of Computing

May 24th, 2019 – İlkay Altıntaş, PhD (ialtintas@ucsd.edu)

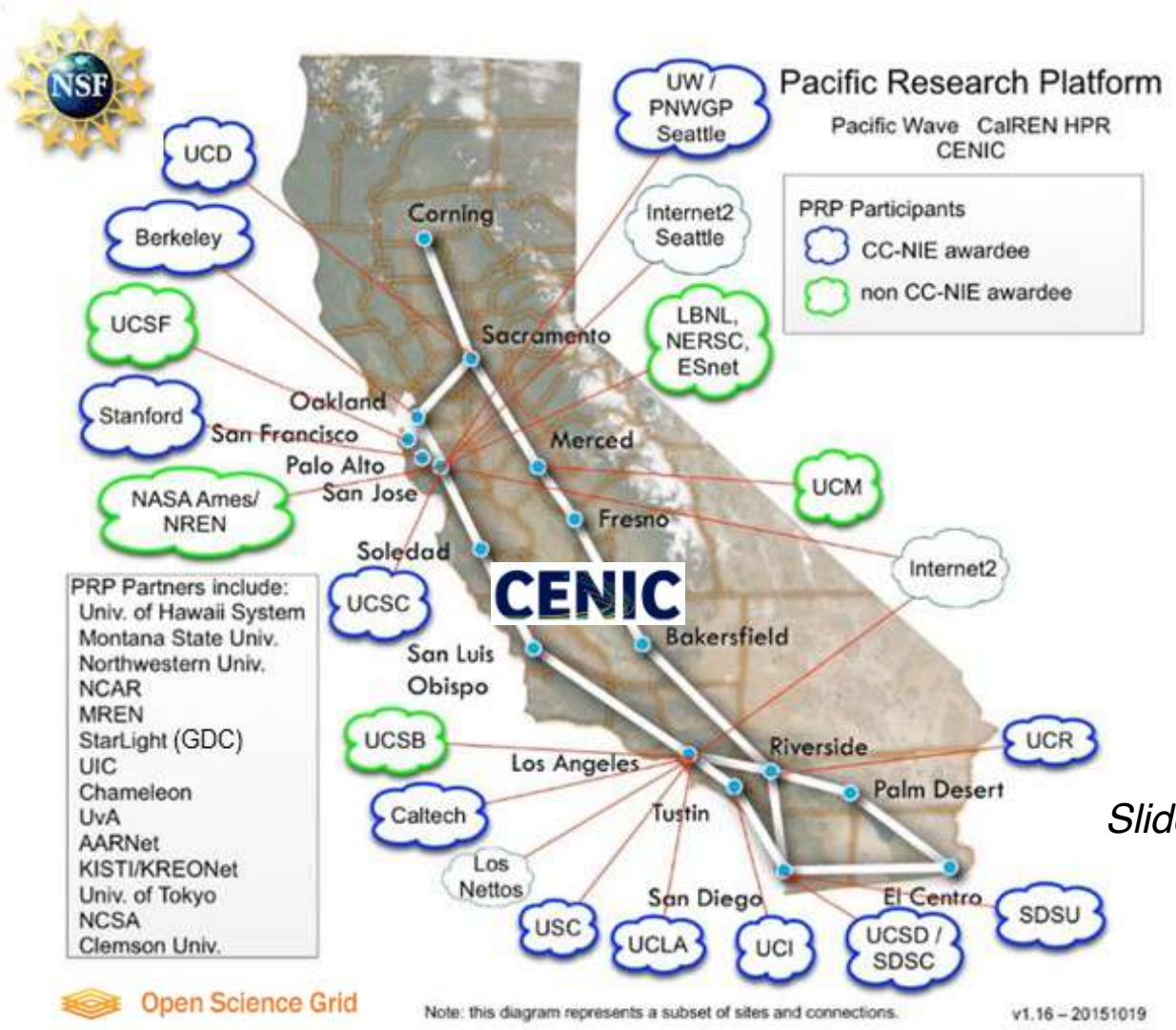
Utilizing “Advanced Cyberinfrastructure”



NSF Comet, PRP, XSEDE, CHASE-CI, Kepler CORE, REAP, GEON, SEEK, RoadNET, bioKepler, DISCOSci, …

Compute + Storage + Network

The Pacific Research Platform Creates a Regional End-to-End Science-Driven "Big Data Superhighway" System



NSF CC*DNI Grant
$5M 10/2015-10/2020

PI: Larry Smarr, UC San Diego Calit2
Co-PIs:

- ***Camille Crittenden, UC Berkeley CITRIS,***
- ***Tom DeFanti, UC San Diego Calit2,***
- ***Philip Papadopoulos, UCSD SDSC,***
- ***Frank Wuerthwein, UCSD Physics and SDSC***

Letters of Commitment from:

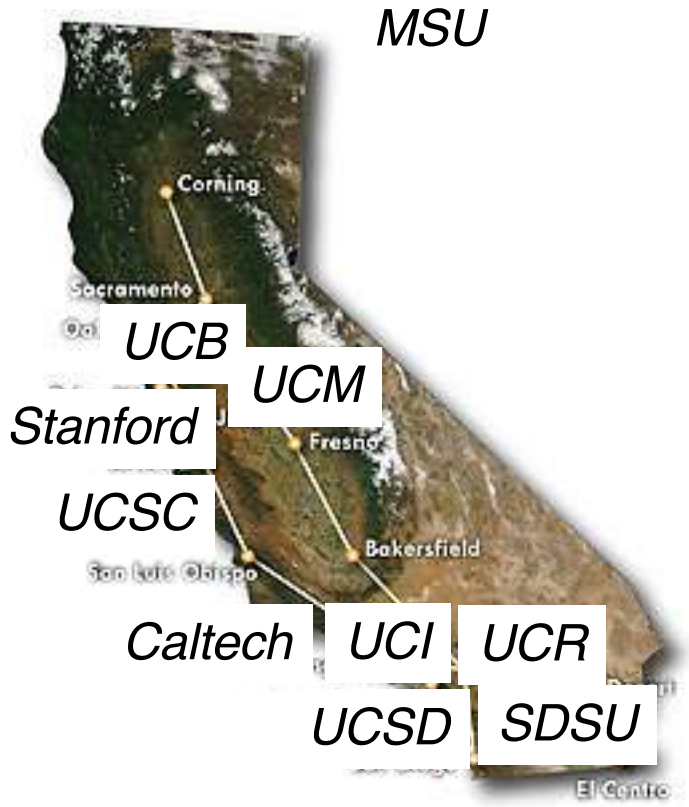
- *50 Researchers from 15 Campuses*
- *32 IT/Network Organization Leaders*

Slide Source: Larry Smarr, UCSD

Disk-to-Disk
10-100 Gbps

Source:
John Hess, CENIC

NSF CHASE-CI Grant Creates a Community Cyberinfrastructure Adding a Machine Learning Layer Built on Top of the Pacific Research Platform



CI-New: Cognitive Hardware and Software Ecosystem Community Infrastructure (CHASE-CI)

For the Period September 1, 2017 – August 31, 2020

SUBMITTED – January 18, 2017

PI: Larry Smarr, Professor of Computer Science and Engineering, Director Calit2, UCSD
Co-PI: Tajana Rosing, Professor of Computer Science and Engineering, UCSD
Co-PI: Ken Kreutz-Delgado, Professor of Electrical and Computer Engineering, UCSD
Co-PI: Ilkay Altintas, Chief Data Science Officer, San Diego Supercomputer Center, UCSD
Co-PI: Tom DeFanti, Research Scientist, Calit2, UCSD

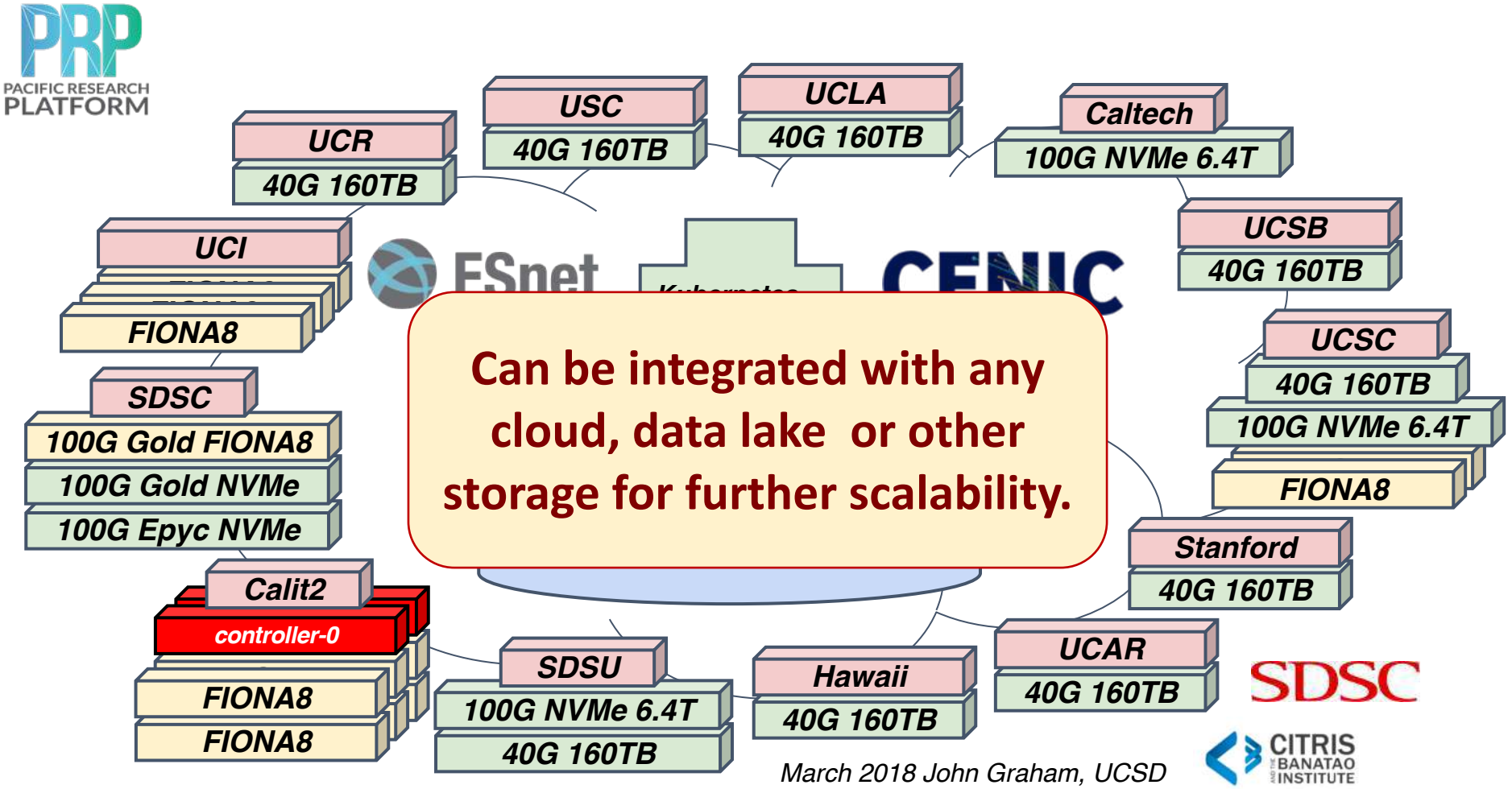
Slide Source: Larry Smarr, UCSD

NSF Grant for High Speed "Cloud" of 256 GPUs
For 30 ML Faculty & Their Students at 10 Campuses
for Training AI Algorithms on Big Data

JACOBS SCHOOL OF ENGINEERING

SDSC UC San Diego

Running Kubernetes/Rook/Ceph On PRP Allows Deployment of Distributed PB+ of Storage for Posting Science Data.



*Five Racked **FIONAs** at Calit2*

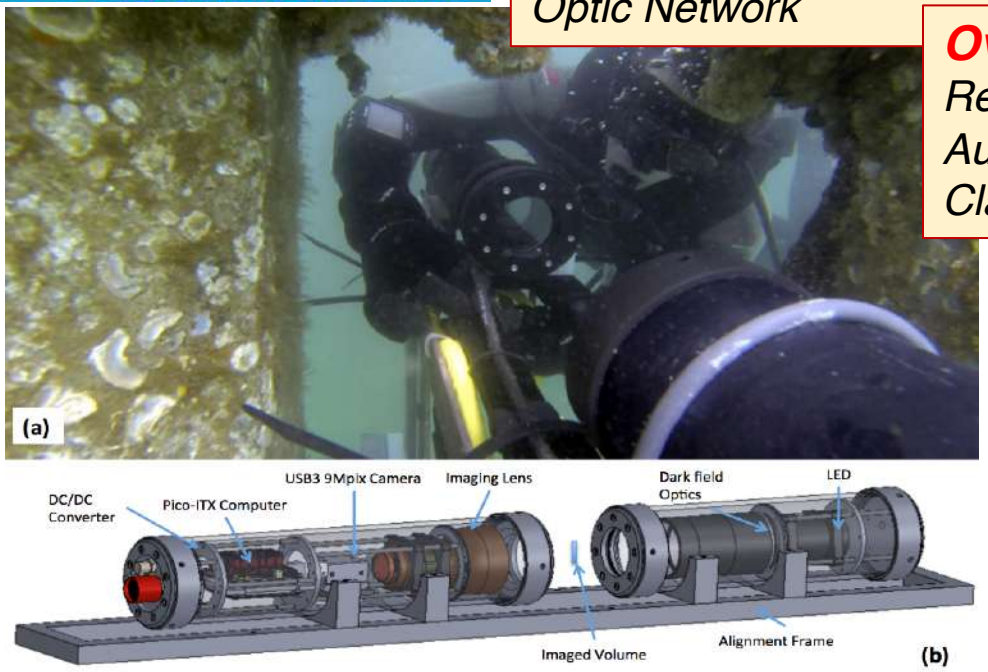
- *Each Contains:*
  - *Dual 12-Core CPUs*
  - *96GB RAM*
  - *1TB SSD*
  - *2 10GbE interfaces*
    - *Total ~$10,500*
  - *With 8 GPUs*
    - *total ~$18,500*

Phil Papadopoulos, SDSC & Tom DeFanti, Joe Keefe & John Graham, Calit2

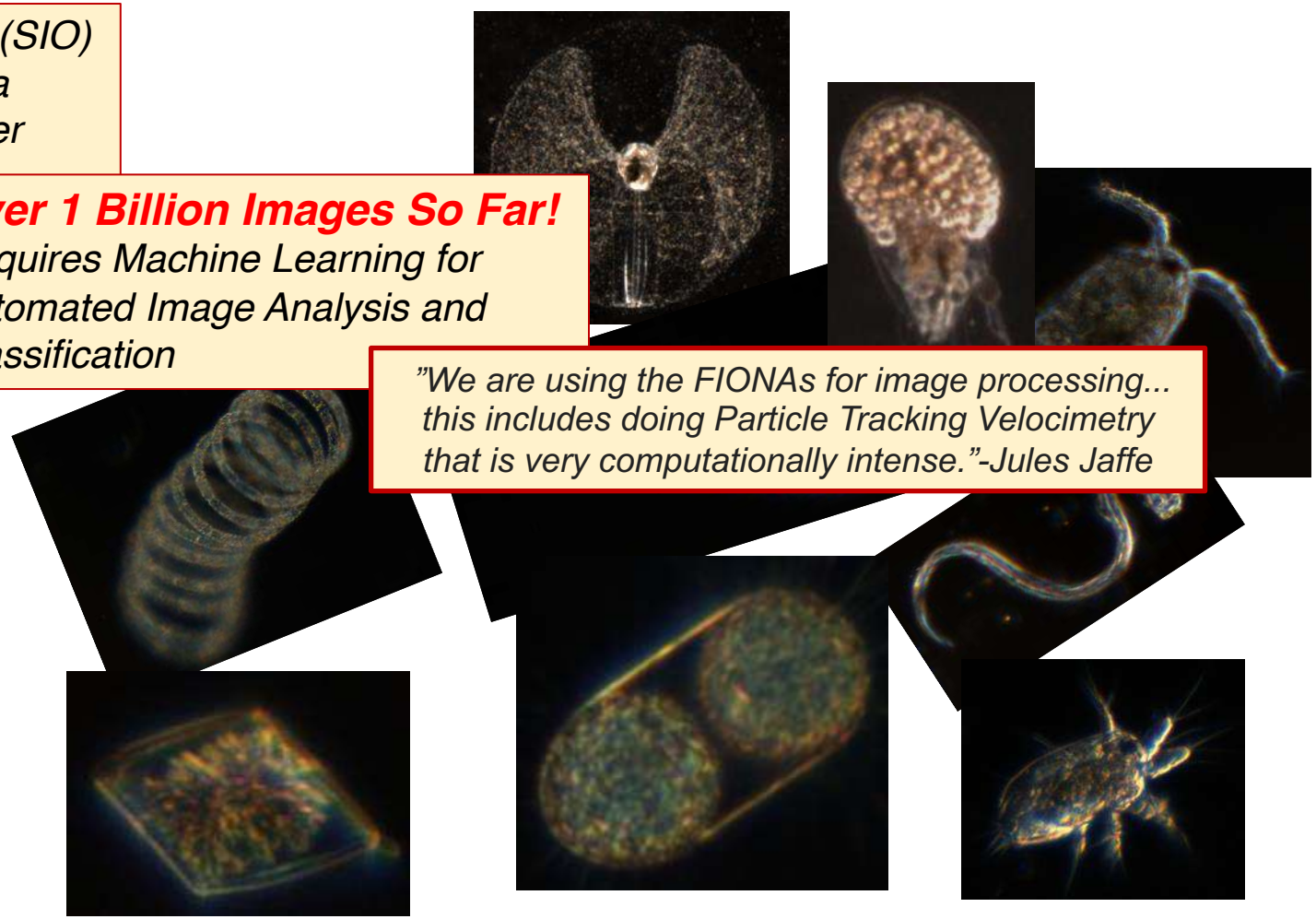
Source: Larry Smarr, Calit2

Streaming Image Processing using PRP and CHASE-CI

UC San Diego Jaffe Lab (SIO)
Scripps Plankton Camera
Off the SIO Pier with Fiber Optic Network



Over 1 Billion Images So Far!
Requires Machine Learning for Automated Image Analysis and Classification



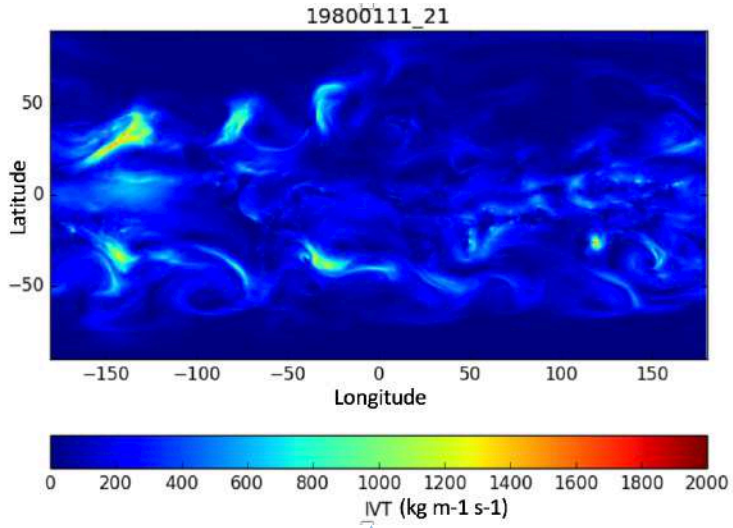
"We are using the FIONAs for image processing... this includes doing Particle Tracking Velocimetry that is very computationally intense."-Jules Jaffe

Source: Larry Smarr, Calit2, Jules Jaffe, SIO

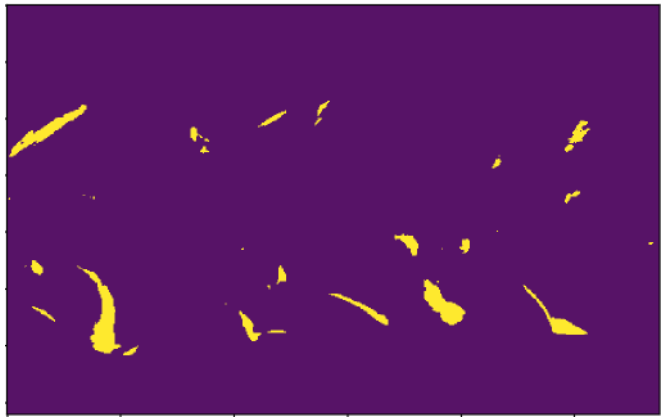
Scott Sellars Rapid 4D Object Segmentation of NASA Water Vapor Data - "Stitching" in Time and Space

Slide Source: Larry Smarr, UCSD

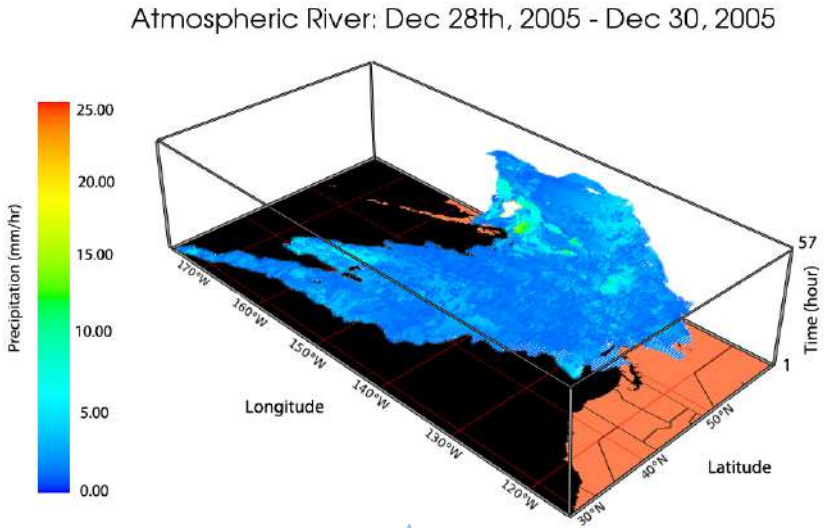
*NASA *MERRA v2 – Water Vapor Data Across the Globe*



Object Detection, Segmentation and Tracking



4D Object Constructed (Lat, Lon, Value, Time)



Scott L. Sellars[1], John Graham[1], Dima Mishin[1], Kyle Marcus[2] , Ilkay Altintas[2], Tom DeFanti[1], Larry Smarr[1], Joulien Tatar[3], Phu Nguyen[4], Eric Shearer[4], and Soroosh Sorooshian[4]

[1]Calit2@UCSD; [2]SDSC; [3]Office of Information Technology, UCI; [4]Center for Hydrometeorology and Remote Sensing, UCI

PRP Shortened Scott's Workflow From 19.2 Days to 52 Minutes - 532 Times Faster!

Slide Source: Larry Smarr, UCSD

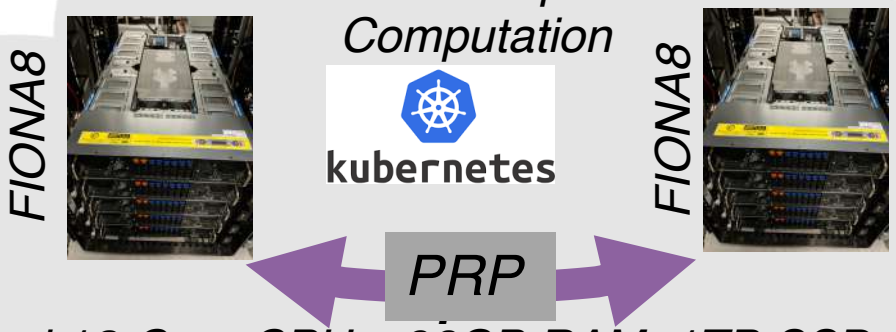
TensorFlow

v4
11/2018

Nautilus

**Altintas et al., 2019 (IPDPS SNACS 2019)*

Download/Preparation
Computation



FIONA8, Dual 12-Core CPUs, 96GB RAM, 1TB SSD, **8 GPUs*

v4 Key Steps	Wall-Clock Time (hrs)
K8s Download/Prep	.6
K8s Compute (50 GPUs)	24.6
*(64 Pods, i.e. workers)	
Total	25 (**1.01 days**)

**bottlenecks: 3D CNN Training/Inference!*
https://gitlab.nautilus.optiputer.net/connect/connect-workflow

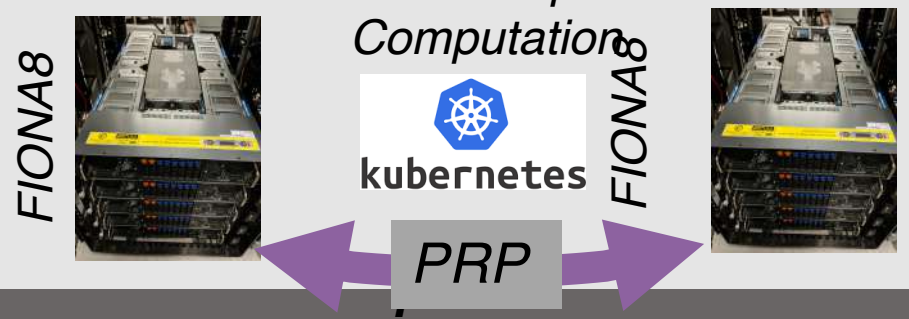
Cython

v5
2/2019

Nautilus

**Sellars et al., 2019 (submitted to eScience 2019)*

Download/Preparation
Computation

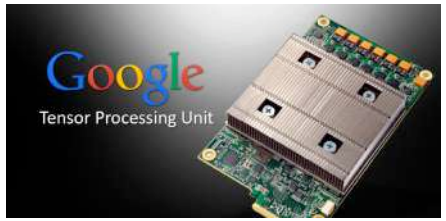


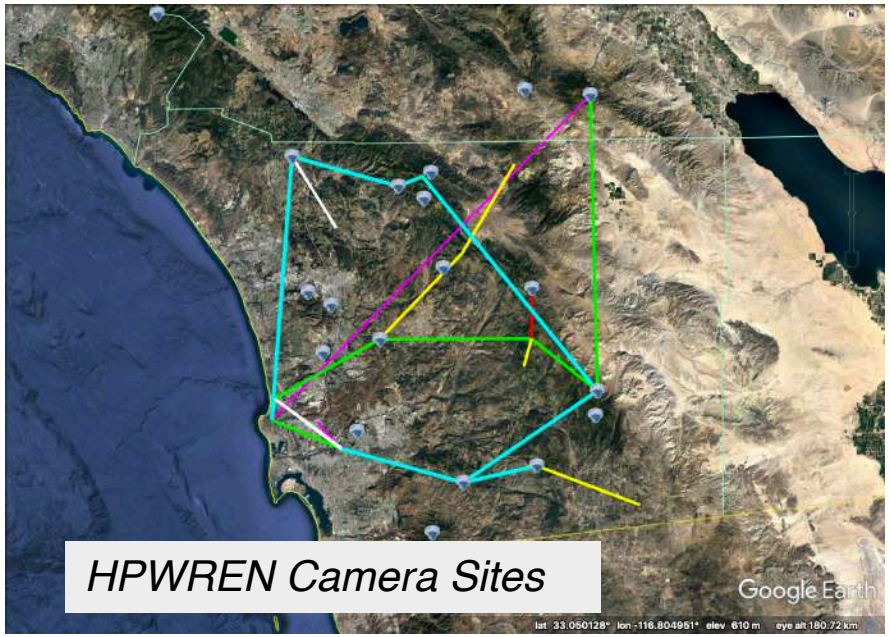
v5 Key Steps	Wall-Clock Time (hrs)
K8s Download/Prep	.75
K8s Compute (25 CPUs)	.11
*(35 Pods, i.e. workers)	
Total	.86 (**52 mins**)

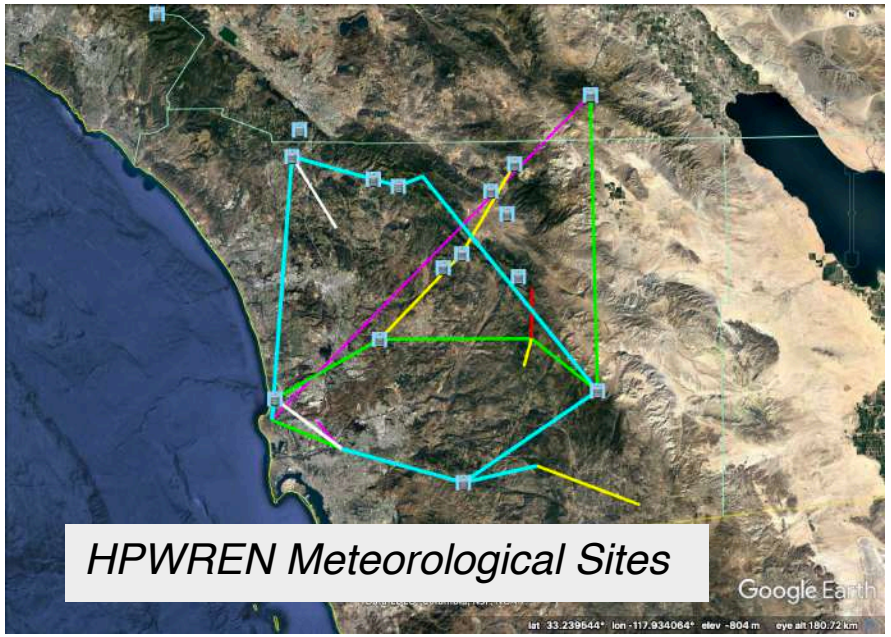
**bottlenecks: Downloading data (again)!*
https://gitlab.nautilus.optiputer.net/connect/hyper-connect

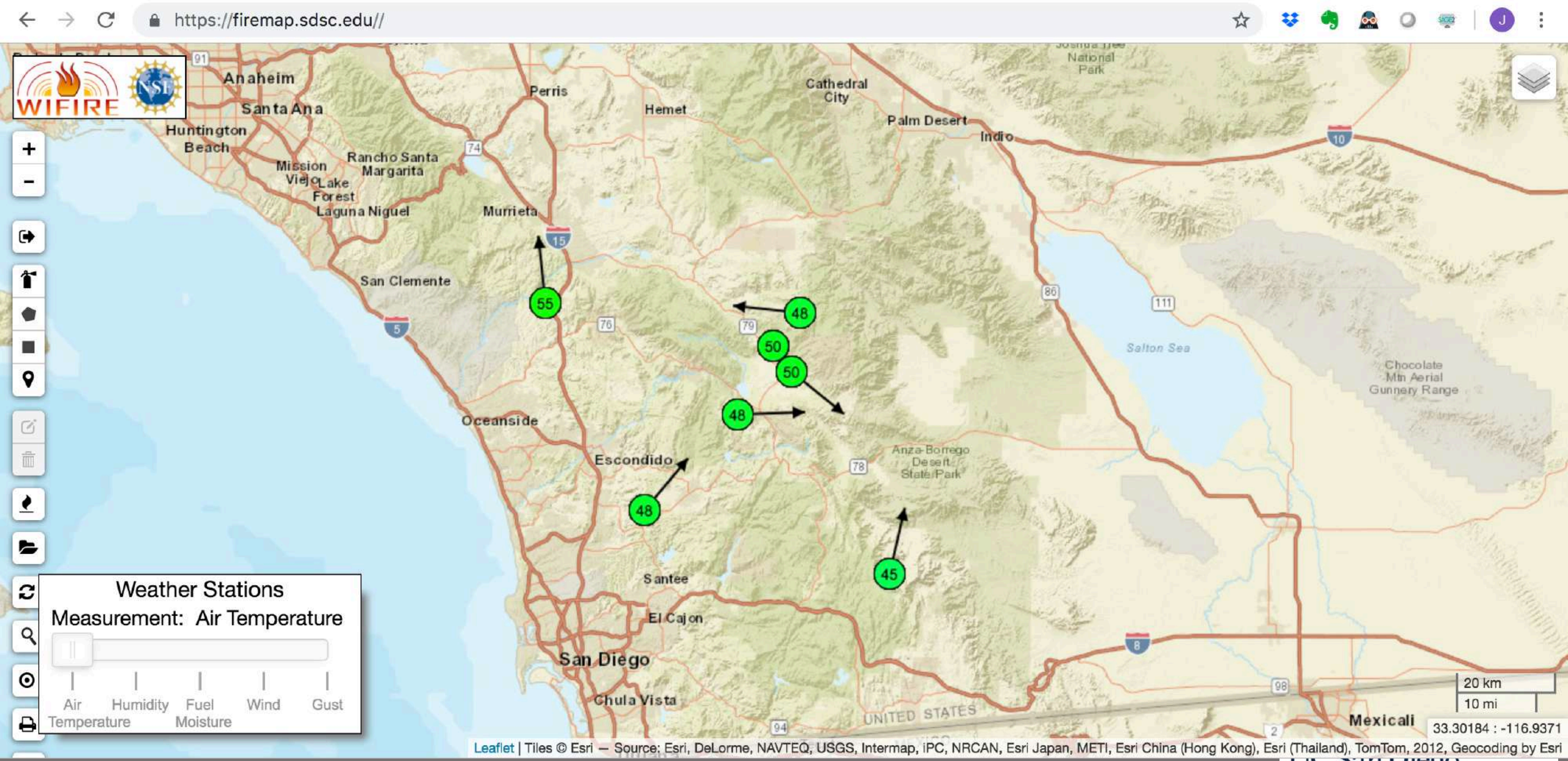
Ongoing Experiments to Extend CHASE-CI to the Edge of HPWREN

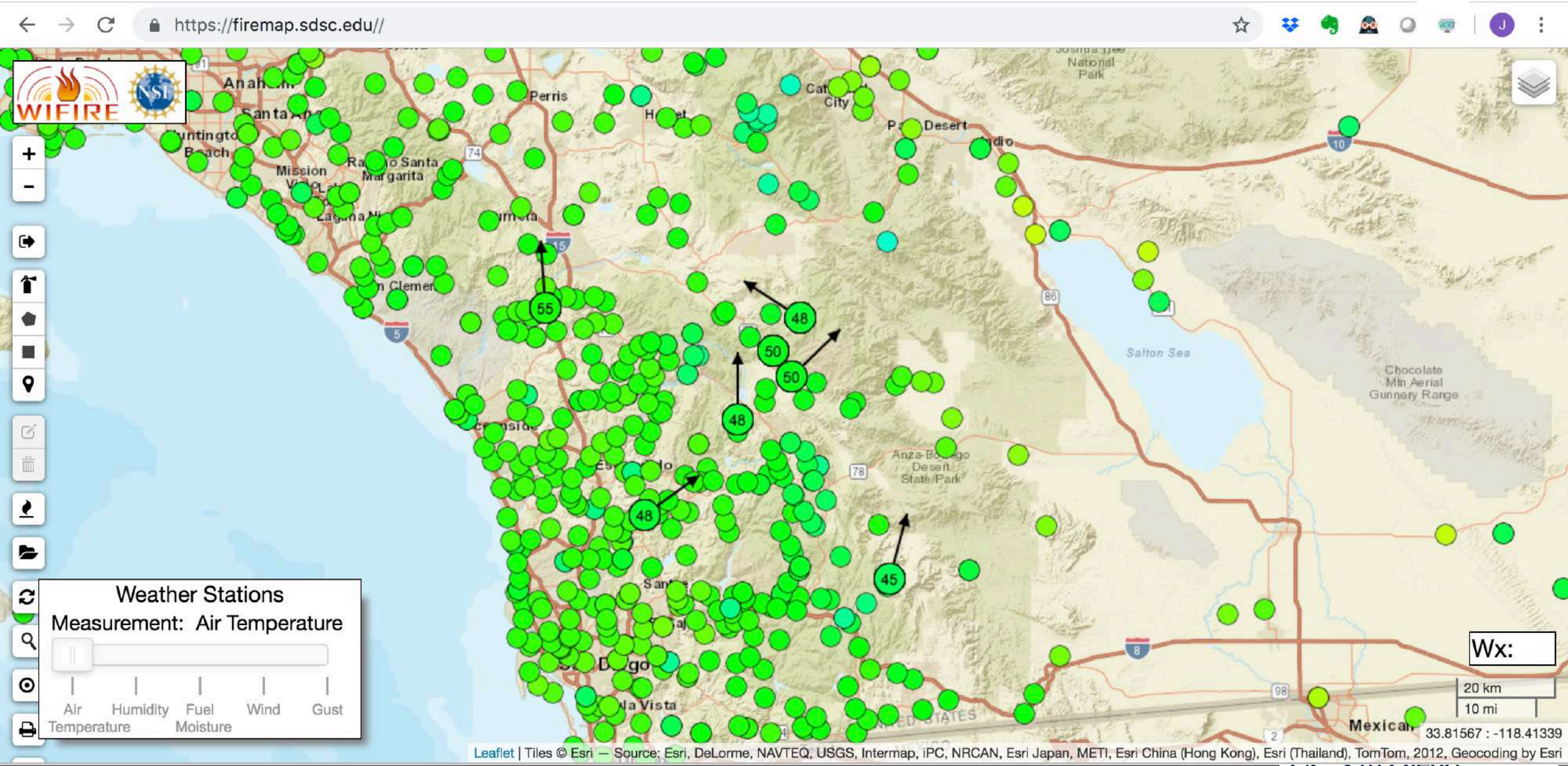
- Ground-based Smoke and Fire Detection from Web Camera Imagery at the edge to generate Alerts
- Continuous weather modeling at the edge (sensor -> higher resolution) to Weather Alerts
- Fuel and Soil Moisture Alerts

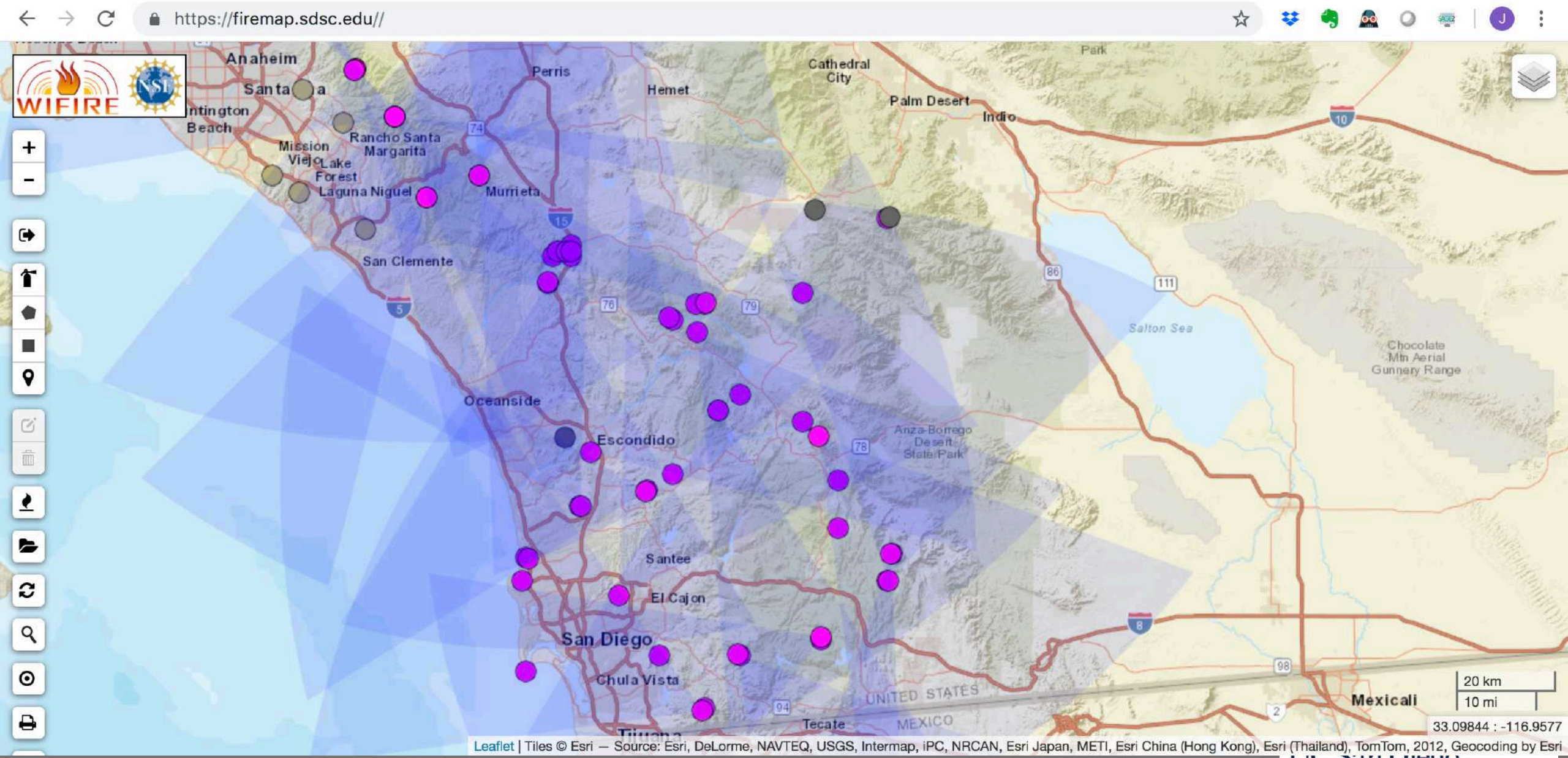


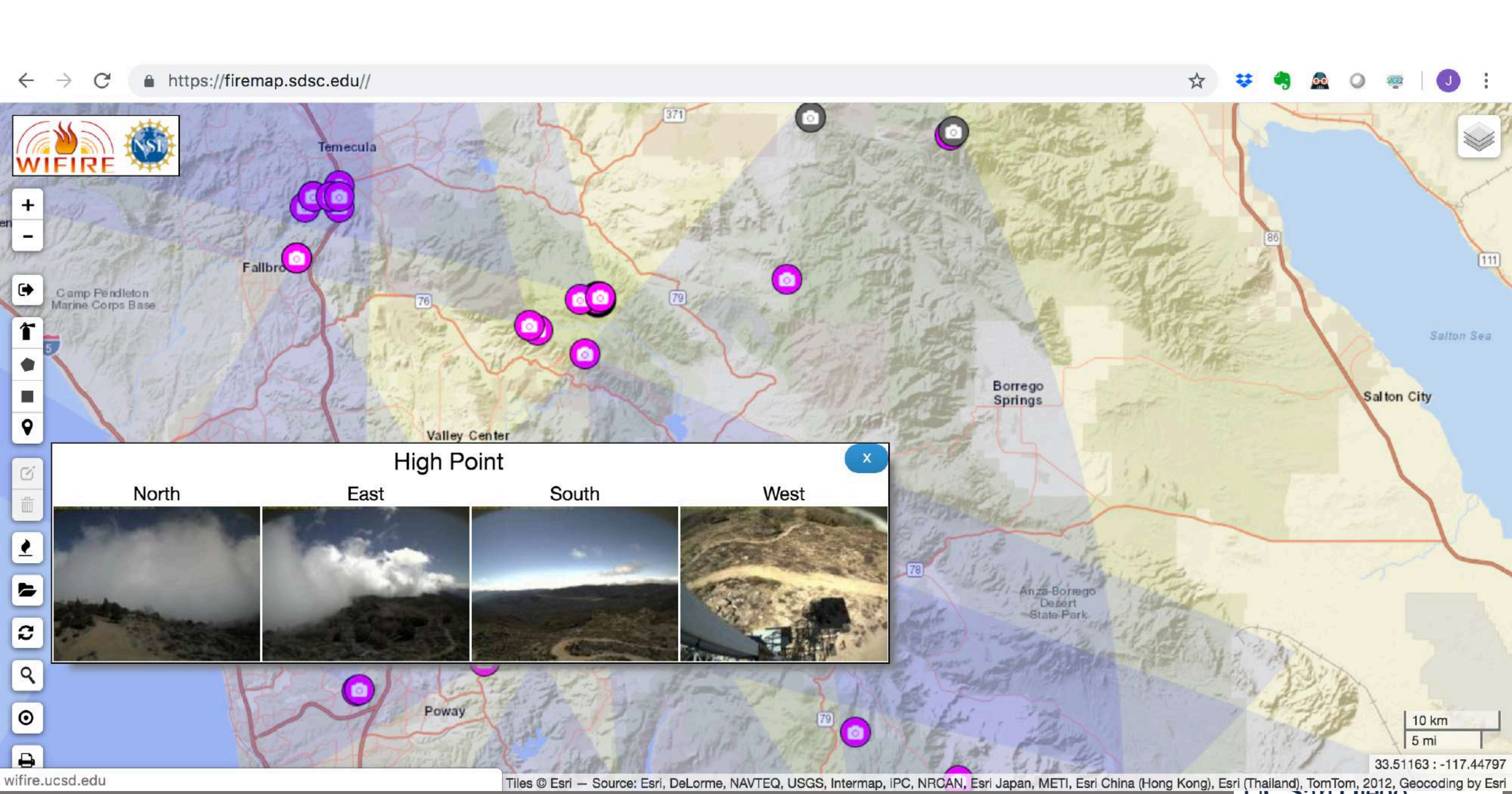




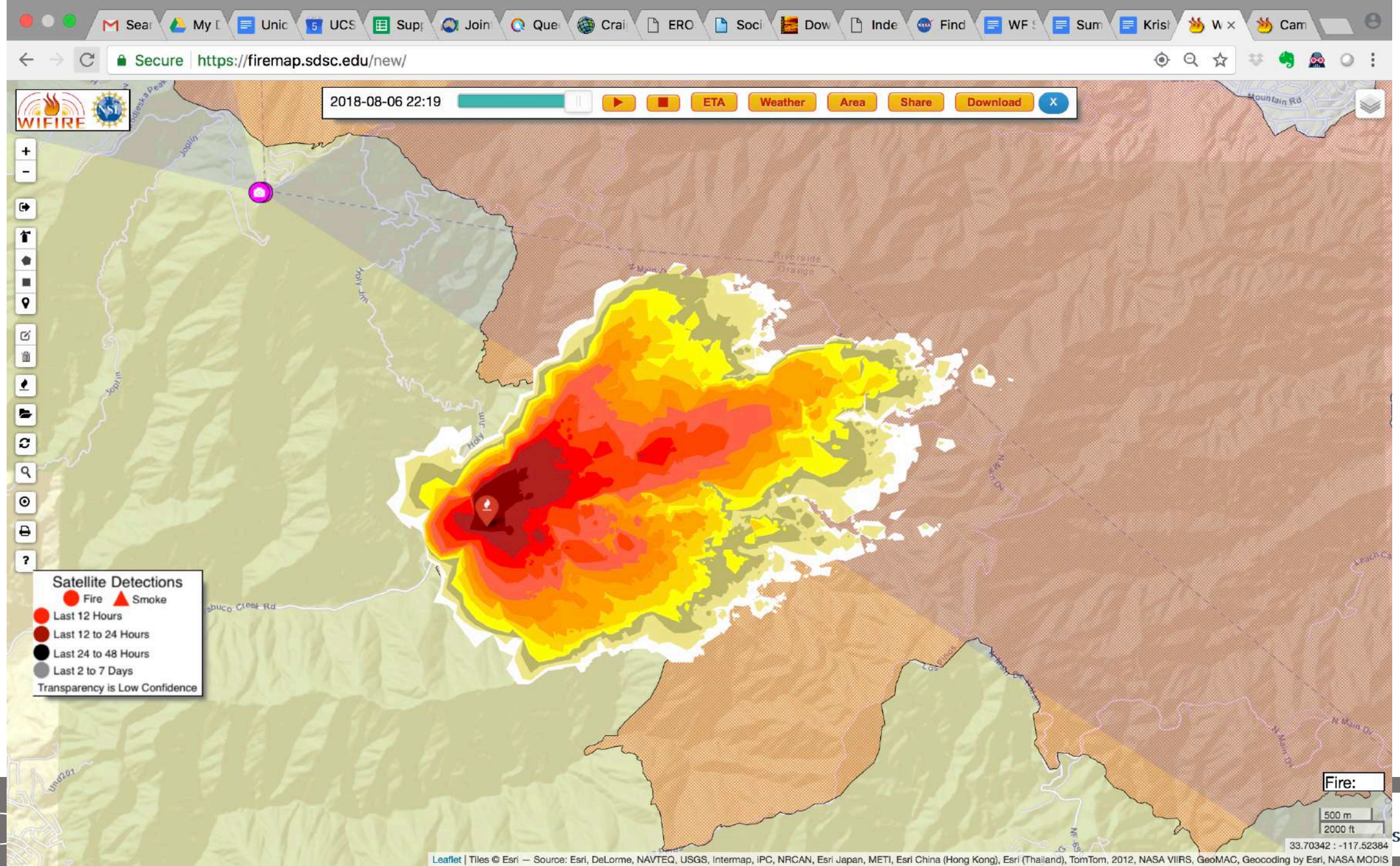

https://firemap.sdsc.edu//
WIFIRE
NSF
Anaheim
Santa Ana
Huntington Beach
Mission Viejo
Lake Forest
Laguna Niguel
Rancho Santa Margarita
Perris
Hemet
Murrieta
San Clemente
Oceanside
Escondido
Santee
El Cajon
San Diego
Chula Vista
Cathedral City
Palm Desert
Indio
Joshua Tree National Park
Salton Sea
Chocolate Mtn Aerial Gunnery Range
Anza-Borrego Desert State Park
Mexicali
UNITED STATES
55
48
50
50
48
48
45
Weather Stations
Measurement: Air Temperature
Air Temperature
Humidity
Fuel Moisture
Wind
Gust
20 km
10 mi
33.30184 : -116.9371
Leaflet | Tiles © Esri — Source: Esri, DeLorme, NAVTEQ, USGS, Intermap, iPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), TomTom, 2012, Geocoding by Esri

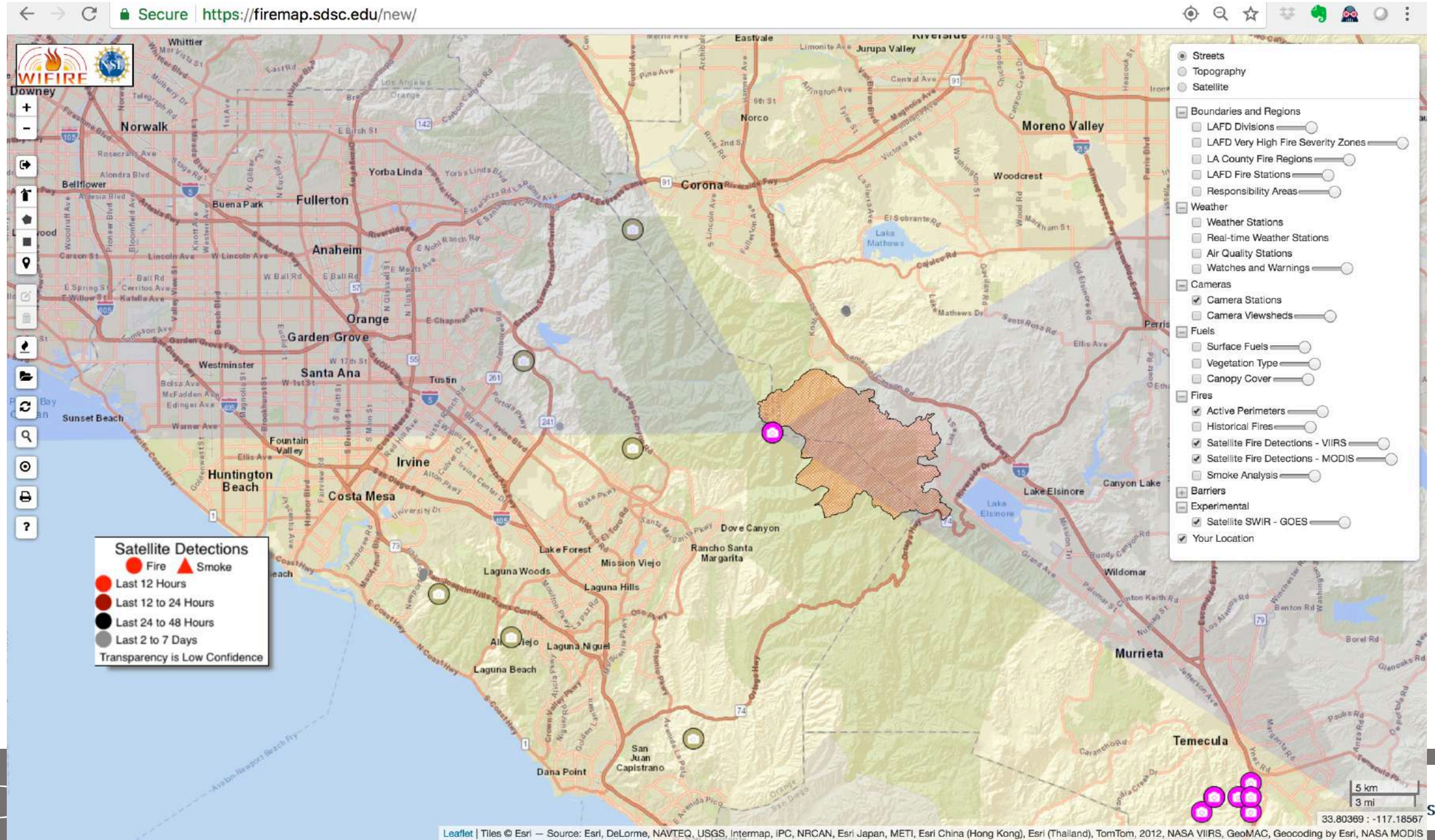

https://firemap.sdsc.edu//
WIFIRE
NSF
Weather Stations
Measurement: Air Temperature
Air Temperature
Humidity
Fuel Moisture
Wind
Gust
Wx:
20 km
10 mi
33.81567 : -118.41339
Leaflet | Tiles © Esri — Source: Esri, DeLorme, NAVTEQ, USGS, Intermap, iPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), TomTom, 2012, Geocoding by Esri

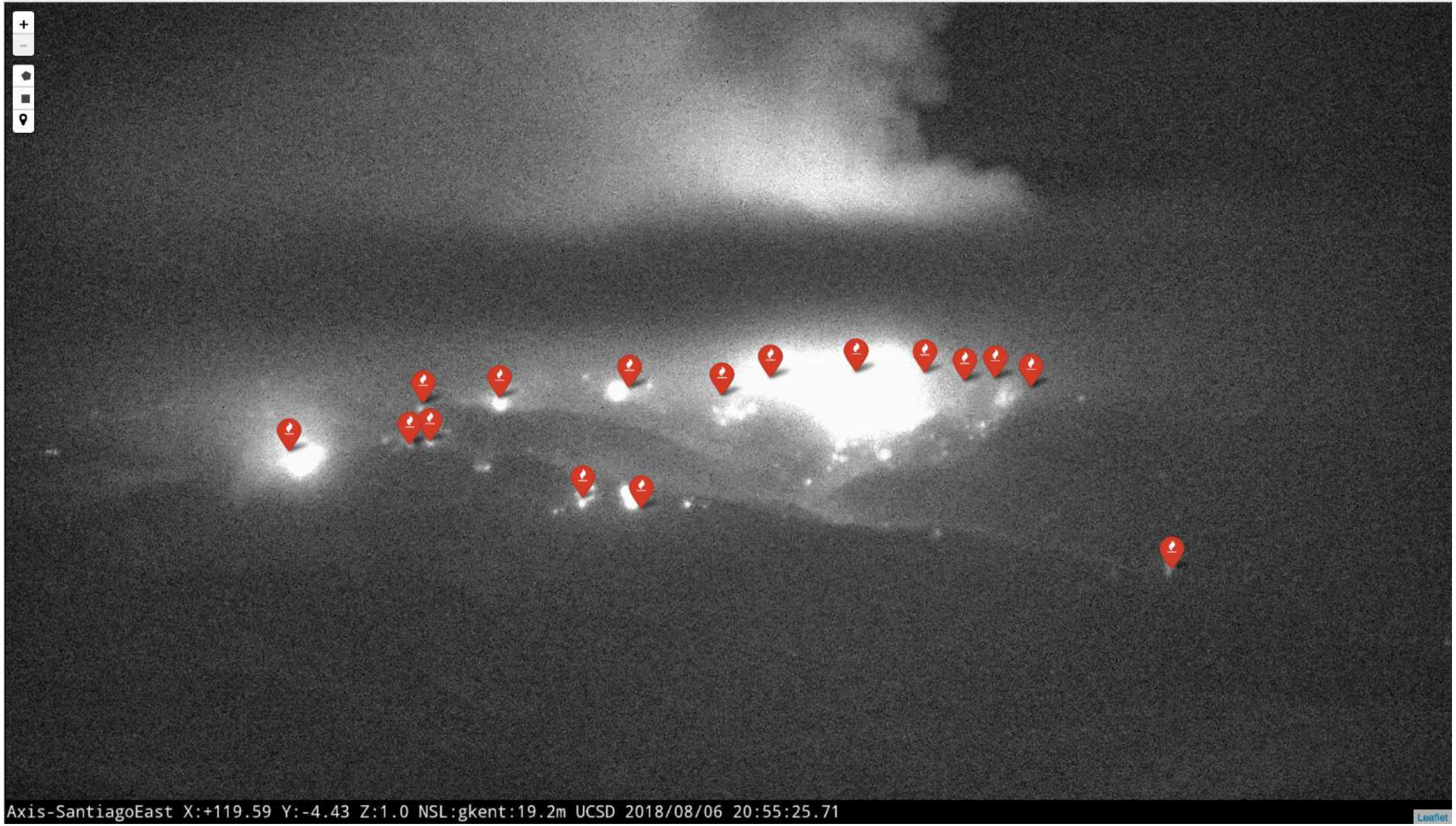

https://firemap.sdsc.edu//
WIFIRE
NSF
Anaheim
Perris
Hemet
Cathedral City
Palm Desert
Indio
Park
Rancho Santa Margarita
Mission Viejo
Lake Forest
Laguna Niguel
Murrieta
San Clemente
Oceanside
Escondido
Salton Sea
Chocolate Mtn Aerial Gunnery Range
Anza-Borrego Desert State Park
Santee
El Cajon
San Diego
Chula Vista
Tecate
UNITED STATES
MEXICO
Mexicali
20 km
10 mi
33.09844 : -116.9577
Leaflet | Tiles © Esri — Source: Esri, DeLorme, NAVTEQ, USGS, Intermap, iPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), TomTom, 2012, Geocoding by Esri

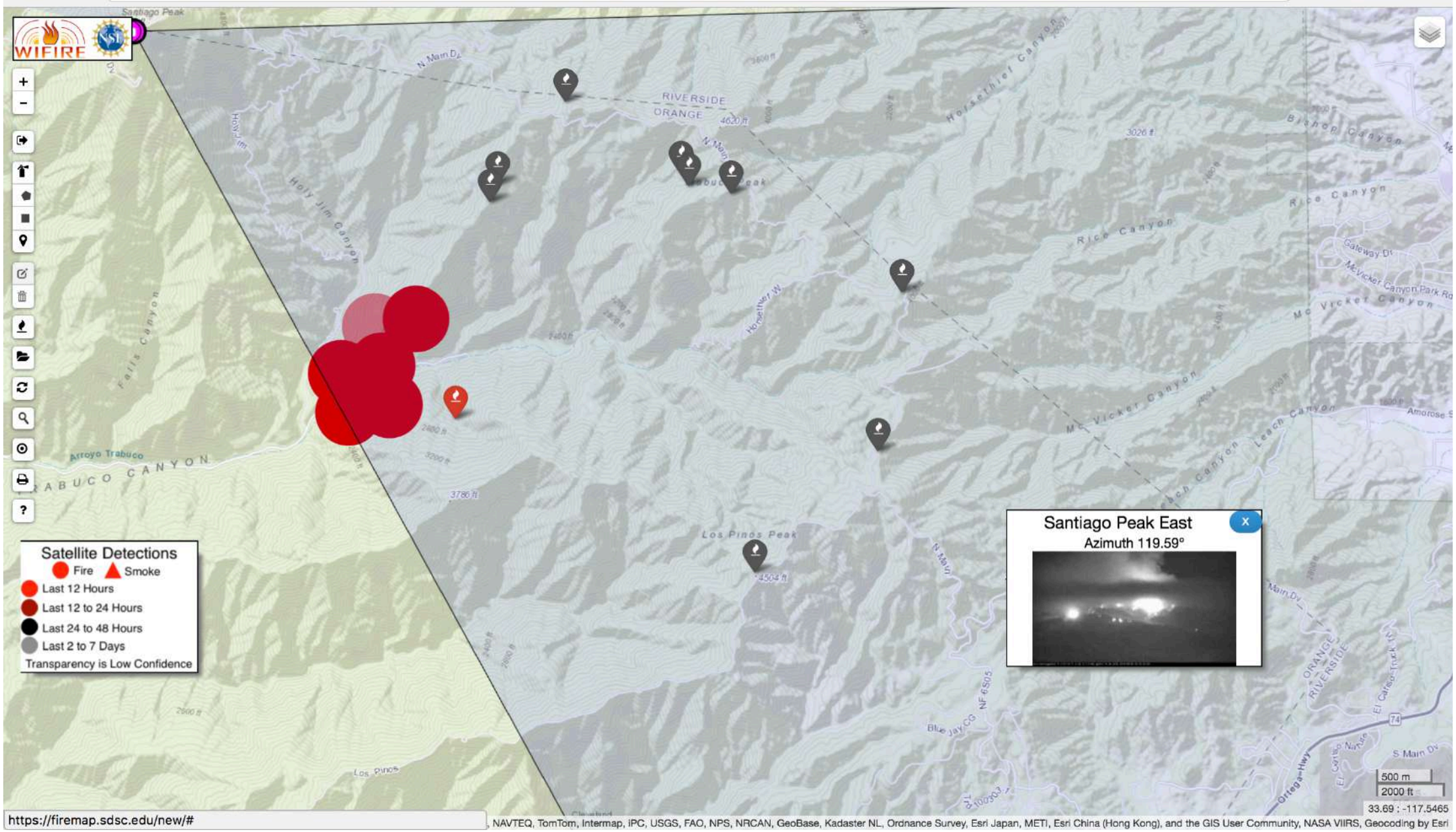

https://firemap.sdsc.edu//
WIFIRE
Temecula
Fallbro
Camp Pendleton
Marine Corps Base
Valley Center
Poway
Borrego
Springs
Salton City
Salton Sea
Anza-Borrego
Desert
State Park
371
76
79
78
86
111
5
High Point
North
East
South
West
X
10 km
5 mi
33.51163 : -117.44797
wifire.ucsd.edu
Tiles © Esri — Source: Esri, DeLorme, NAVTEQ, USGS, Intermap, iPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), TomTom, 2012, Geocoding by Esri

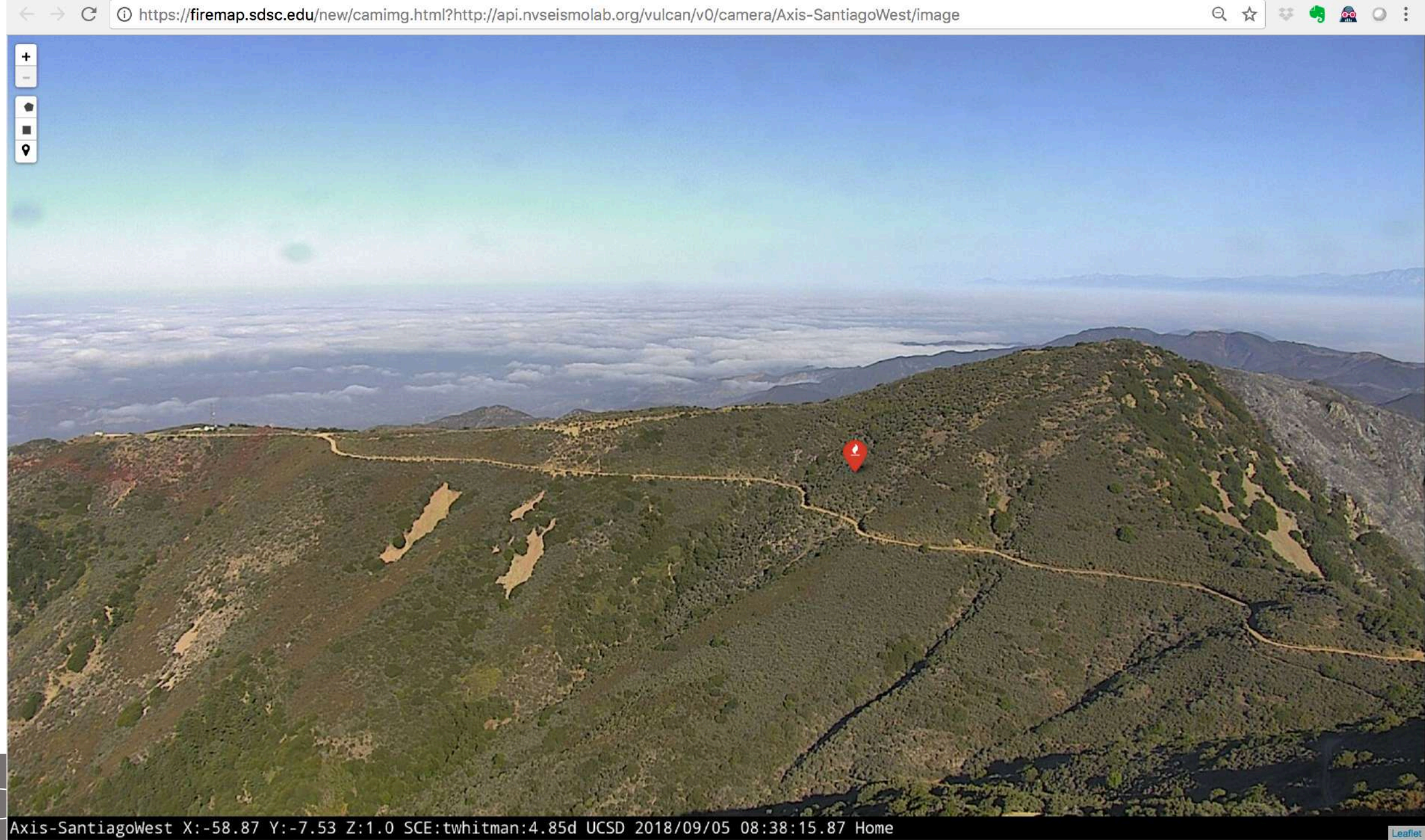

20180806.150230 Santiago Peak East, http://hpwren.ucsd.edu a3
HPWREN
Axis-SantiagoEast X:+130.86 Y:-3.90 Z:1.0 OCFA:bnorton:49.8m2018/08/06 15:02:31.05 UC San Diego/HPWREN http://hpwren.ucsd.edu

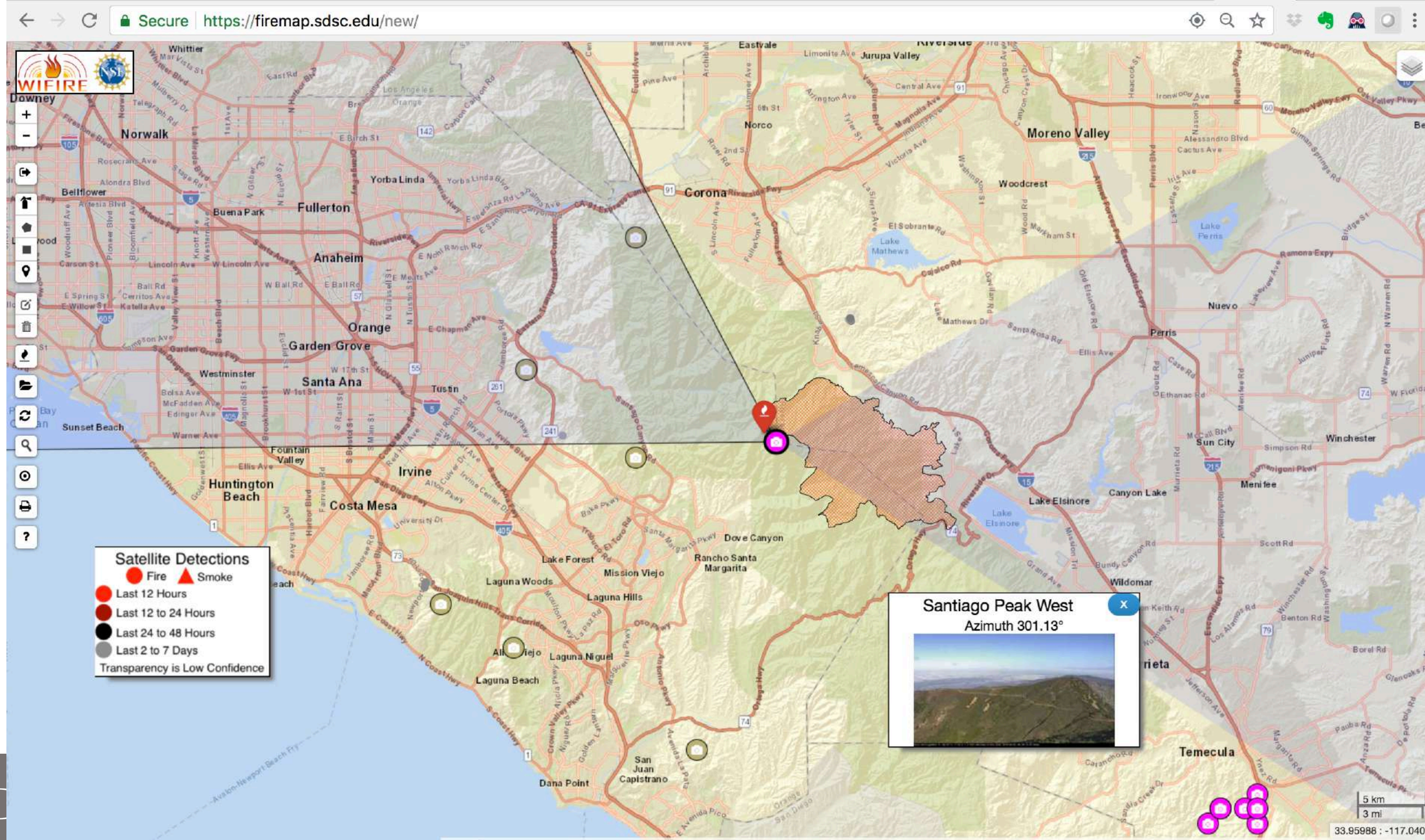

Secure https://firemap.sdsc.edu/new/
WIFIRE
2018-08-06 22:19
ETA
Weather
Area
Share
Download
X
Satellite Detections
Fire
Smoke
Last 12 Hours
Last 12 to 24 Hours
Last 24 to 48 Hours
Last 2 to 7 Days
Transparency is Low Confidence
Fire:
500 m
2000 ft
33.70342 : -117.52384
Leaflet | Tiles © Esri — Source: Esri, DeLorme, NAVTEQ, USGS, Intermap, iPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), TomTom, 2012, NASA VIIRS, GeoMAC, Geocoding by Esri, NASA MODIS

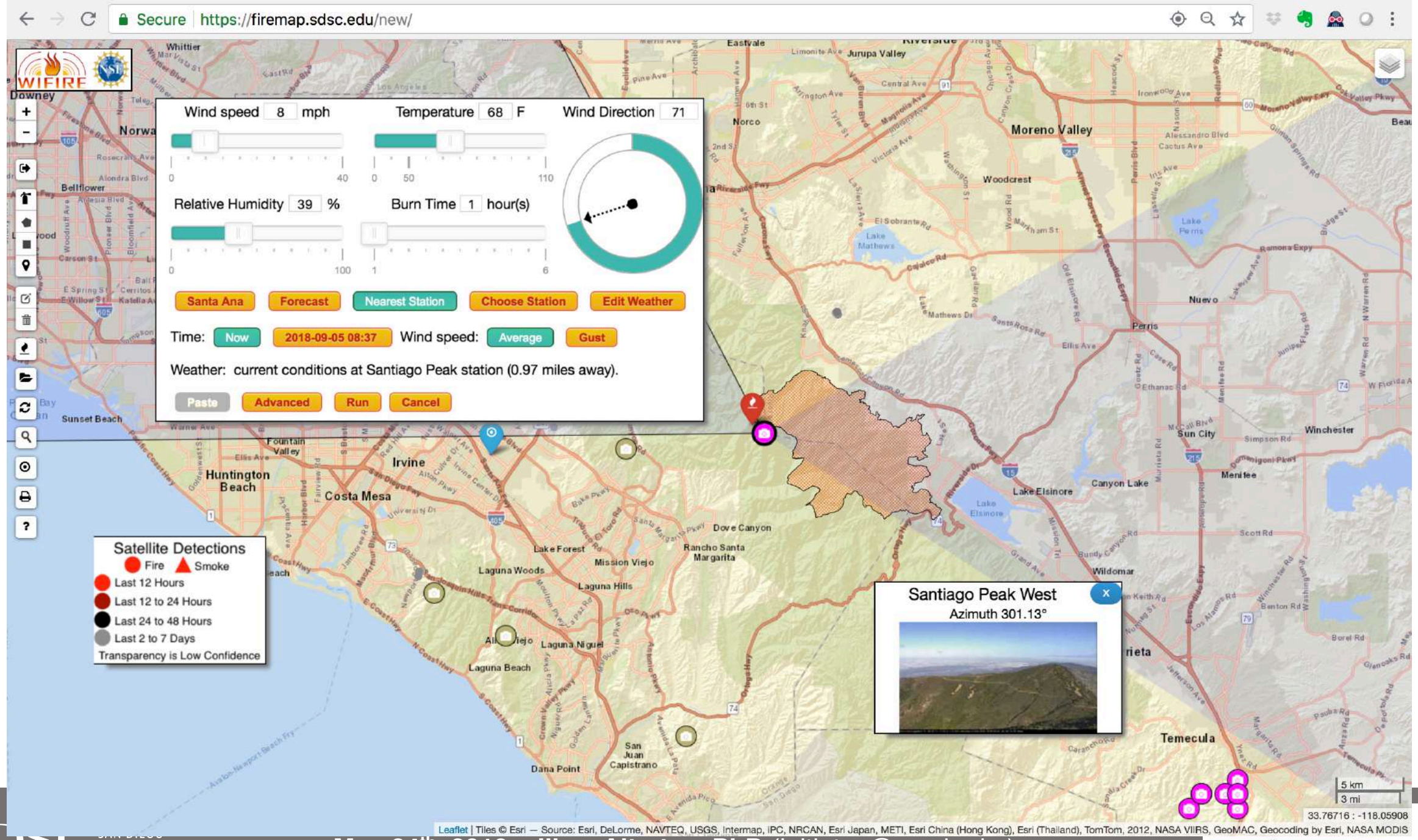

Secure https://firemap.sdsc.edu/new/
WIFIRE
Satellite Detections
Fire
Smoke
Last 12 Hours
Last 12 to 24 Hours
Last 24 to 48 Hours
Last 2 to 7 Days
Transparency is Low Confidence
Streets
Topography
Satellite
Boundaries and Regions
LAFD Divisions
LAFD Very High Fire Severity Zones
LA County Fire Regions
LAFD Fire Stations
Responsibility Areas
Weather
Weather Stations
Real-time Weather Stations
Air Quality Stations
Watches and Warnings
Cameras
Camera Stations
Camera Viewsheds
Fuels
Surface Fuels
Vegetation Type
Canopy Cover
Fires
Active Perimeters
Historical Fires
Satellite Fire Detections - VIIRS
Satellite Fire Detections - MODIS
Smoke Analysis
Barriers
Experimental
Satellite SWIR - GOES
Your Location
Whittier
Downey
Norwalk
Bellflower
Buena Park
Fullerton
Yorba Linda
Anaheim
Orange
Garden Grove
Westminster
Santa Ana
Tustin
Sunset Beach
Fountain Valley
Huntington Beach
Irvine
Costa Mesa
Lake Forest
Mission Viejo
Laguna Woods
Laguna Hills
Laguna Niguel
Laguna Beach
Dana Point
San Juan Capistrano
Dove Canyon
Rancho Santa Margarita
Eastvale
Jurupa Valley
Norco
Corona
Moreno Valley
Woodcrest
Lake Elsinore
Canyon Lake
Wildomar
Murrieta
Temecula
5 km
3 mi
33.80369 : -117.18567
Leaflet | Tiles © Esri — Source: Esri, DeLorme, NAVTEQ, USGS, Intermap, iPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), TomTom, 2012, NASA VIIRS, GeoMAC, Geocoding by Esri, NASA MODIS


Axis-SantiagoEast X:+119.59 Y:-4.43 Z:1.0 NSL:gkent:19.2m UCSD 2018/08/06 20:55:25.71
Leaflet


WIFIRE
Satellite Detections
Fire
Smoke
Last 12 Hours
Last 12 to 24 Hours
Last 24 to 48 Hours
Last 2 to 7 Days
Transparency is Low Confidence
Santiago Peak East
Azimuth 119.59°
RIVERSIDE
ORANGE
Arroyo Trabuco
TRABUCO CANYON
Los Pinos Peak
Rice Canyon
Mc Vicker Canyon
Leach Canyon
Holy Jim Canyon
Falls Canyon
Horsethief Canyon
Bishop Canyon
Gateway Dr
McVicker Canyon Park Rd
Amorose S
N Main Dr
S Main Dr
Ortega Hwy
El Cariso Truck Tr
Blue Jay CG
NF 6S05
500 m
2000 ft
33.69 : -117.5465
https://firemap.sdsc.edu/new/#
NAVTEQ, TomTom, Intermap, iPC, USGS, FAO, NPS, NRCAN, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), and the GIS User Community, NASA VIIRS, Geocoding by Esri


https://firemap.sdsc.edu/new/camimg.html?http://api.nvseismolab.org/vulcan/v0/camera/Axis-SantiagoWest/image
Axis-SantiagoWest X:-58.87 Y:-7.53 Z:1.0 SCE:twhitman:4.85d UCSD 2018/09/05 08:38:15.87 Home
Leaflet

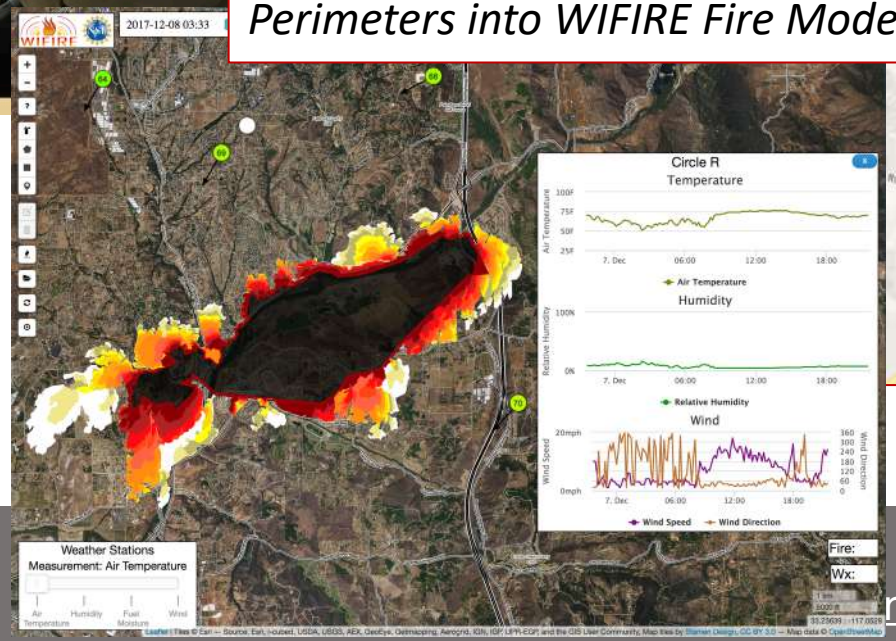

Secure https://firemap.sdsc.edu/new/
WIFIRE
Satellite Detections
Fire
Smoke
Last 12 Hours
Last 12 to 24 Hours
Last 24 to 48 Hours
Last 2 to 7 Days
Transparency is Low Confidence
Santiago Peak West
Azimuth 301.13°
Whittier
Downey
Norwalk
Bellflower
Buena Park
Fullerton
Yorba Linda
Anaheim
Orange
Garden Grove
Westminster
Santa Ana
Tustin
Sunset Beach
Fountain Valley
Irvine
Huntington Beach
Costa Mesa
Lake Forest
Laguna Woods
Mission Viejo
Laguna Hills
Laguna Niguel
Laguna Beach
Dana Point
San Juan Capistrano
Dove Canyon
Rancho Santa Margarita
Eastvale
Jurupa Valley
Norco
Corona
Moreno Valley
Woodcrest
Lake Mathews
Lake Perris
Nuevo
Perris
Sun City
Lake Elsinore
Canyon Lake
Menifee
Wildomar
Temecula
Winchester
5 km
3 mi
33.95988 : -117.040

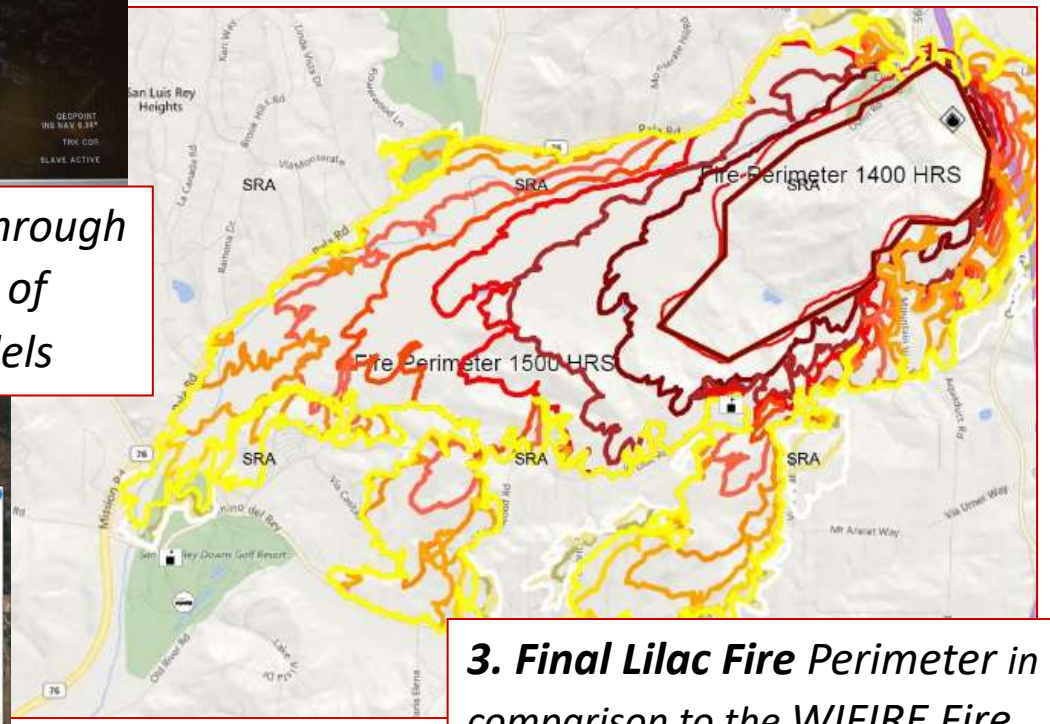

Secure https://firemap.sdsc.edu/new/
WIFIRE
Wind speed 8 mph
Temperature 68 F
Wind Direction 71
Relative Humidity 39 %
Burn Time 1 hour(s)
Santa Ana
Forecast
Nearest Station
Choose Station
Edit Weather
Time: Now 2018-09-05 08:37 Wind speed: Average Gust
Weather: current conditions at Santiago Peak station (0.97 miles away).
Paste
Advanced
Run
Cancel
Satellite Detections
Fire Smoke
Last 12 Hours
Last 12 to 24 Hours
Last 24 to 48 Hours
Last 2 to 7 Days
Transparency is Low Confidence
Santiago Peak West
Azimuth 301.13°
Moreno Valley
Perris
Lake Elsinore
Temecula
Irvine
Huntington Beach
Costa Mesa
Laguna Beach
Dana Point
San Juan Capistrano
5 km
3 mi
33.76716 : -118.05908
Leaflet | Tiles © Esri — Source: Esri, DeLorme, NAVTEQ, USGS, Intermap, IPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), TomTom, 2012, NASA VIIRS, GeoMAC, Geocoding by Esri, NASA MODIS

Partnership with General Atomics Lilac Fire (Dec 7-16, 2017)

1. SDAIRS *team at work at the SDFD Emergency Command and Dispatch Center*

2. Lilac Fire *Perimeter Capture through the GA Aircraft and Assimilation of Perimeters into WIFIRE Fire Models*





3. Final Lilac Fire *Perimeter in comparison to the WIFIRE Fire Progression Model in* ***SCOUT***

ntas@ucsd.edu)

Systems integration requirements…

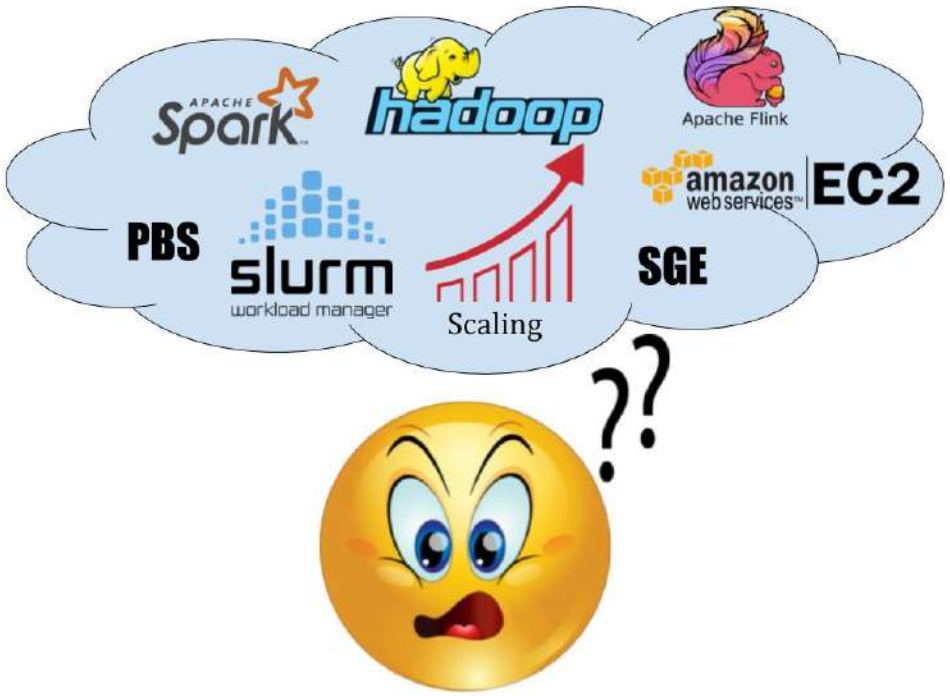
Dynamic composability matters.

Systems are only useful if groups can integrate them into applications.

Tools that enhance teamwork need to be coupled with AI systems.

Dynamic composition requires dynamic network, compute, storage and resource combined with intelligent software for steering applications.

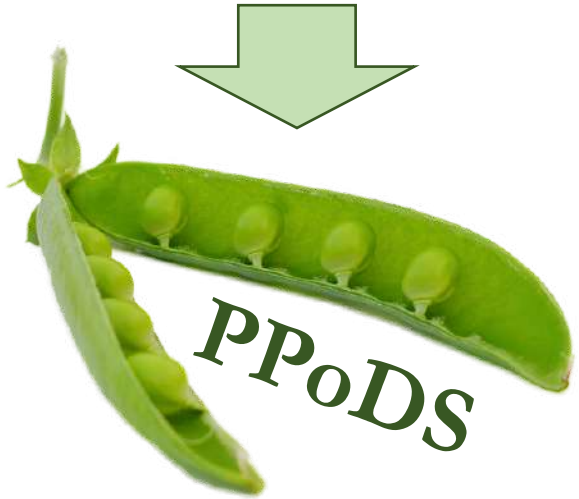
Data to Discovery Challenges



How it looks to Researchers

- Collaboration environment
- Distributed Teams
- Explore to Scale
- Keeping up with latest advancements in computer science, data science and cyber-infrastructure
- Lack of an intelligent end-to-end workflow execution system.

Data



Process for Practice of Data Science

Product

Programmability

Ease of use, iteration, interaction, re-use, re-purpose

Scalability

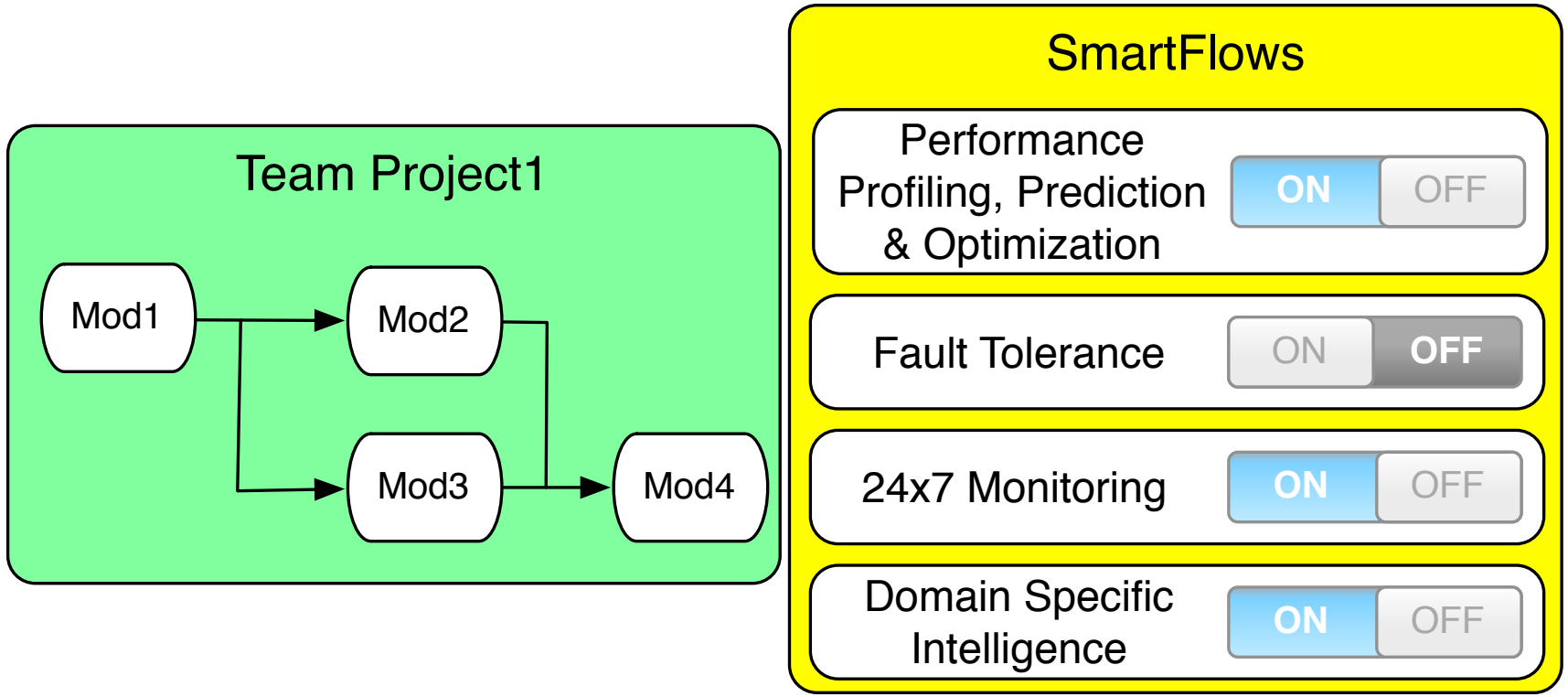
From local experiments to large-scale runs

Reproducibility

Ability to validate, re-run, re-play

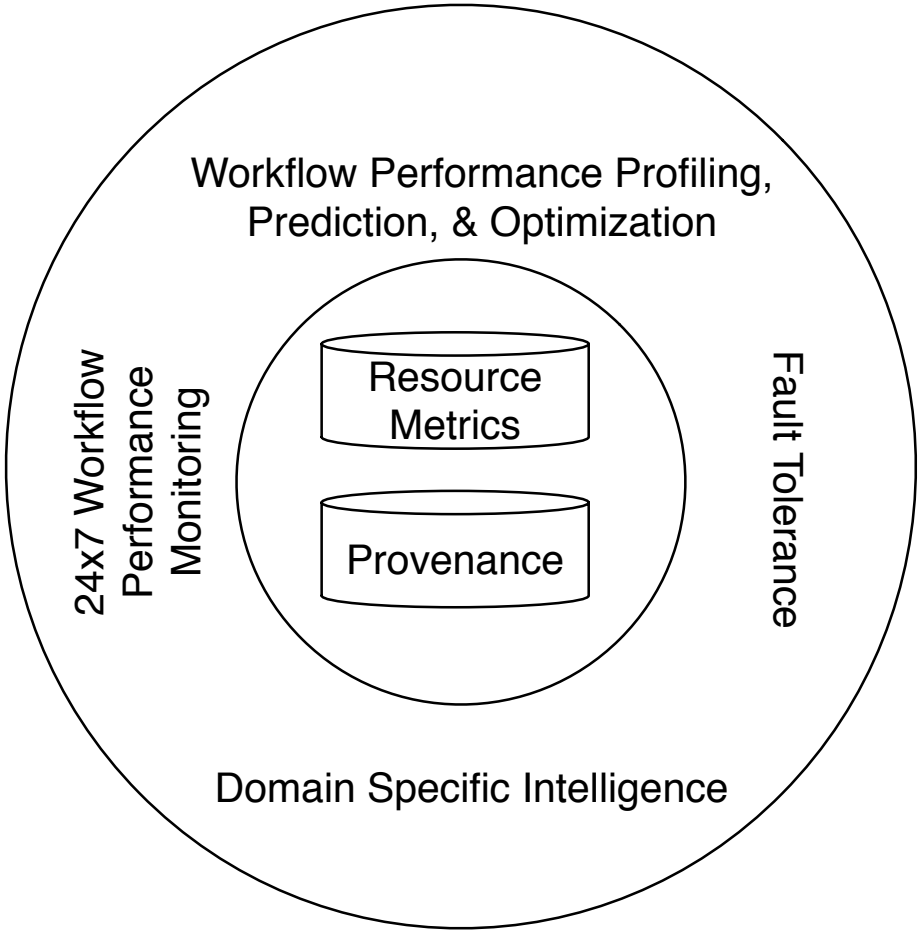
Integrate the workflow-driven, scalable and reproducible science with collaborative and exploratory nature of the scientific process

PPoDS – Workflow Design Canvas



Workflow Design Canvas: enables exploration through an iterative continuous-improvement workflow design by collecting test metrics. Users can also select SmartFlow services for their analysis.

SmartFlows Suite of Services



Intelligent Predictive Control: *SmartFlows, a suite of services that provide intelligent orchestration of workflows, ensuring fault tolerance, 24x7 monitoring, ability to remotely control workflow states, and automatic scaling of workflows.*

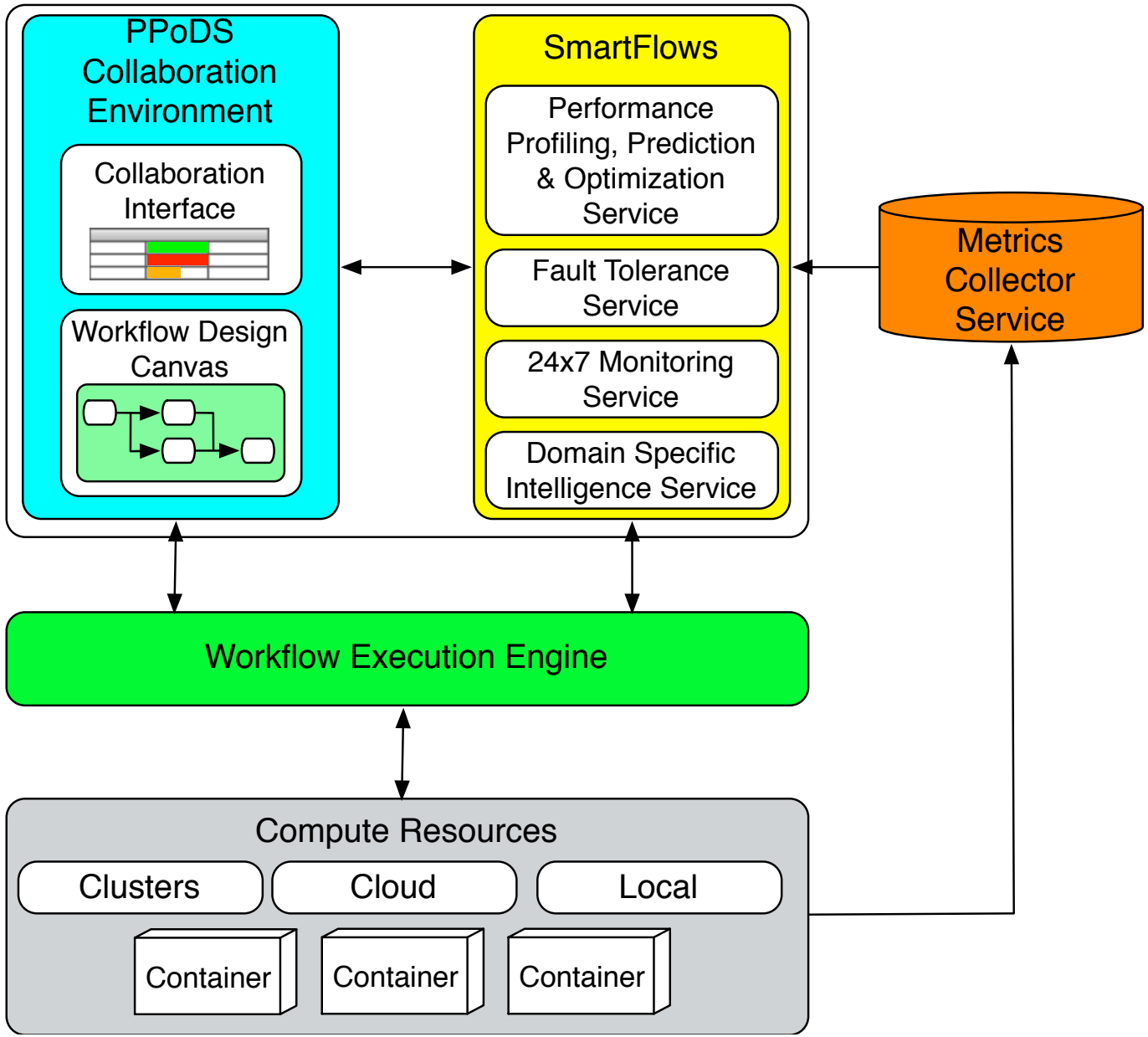
PPoDS – Team Collaboration Interface

Team Project 1

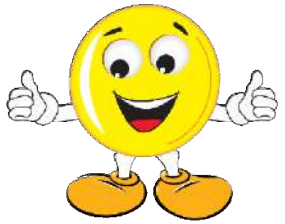
Task	Ownership	Progress	Test Metrics	Remarks
Mod1	Nancy	100%	Show Report	
Mod2	Kia	Error	Show Report	
Mod3	Sam	45%	Show Report	

Mod3 Metrics Performance
Compute Resource Details
CPU cores (hours:minutes:seconds)
Memory Usage (maximum, average)
I/O Reads and Writes

Continuous Improvement: An iterative continuous-improvement methodology called PPODS for enabling metric-driven workflow design.



Collaborative Workflow Environment Architecture



Note: Any workflow execution system can be plugged into this framework

Ending Remarks

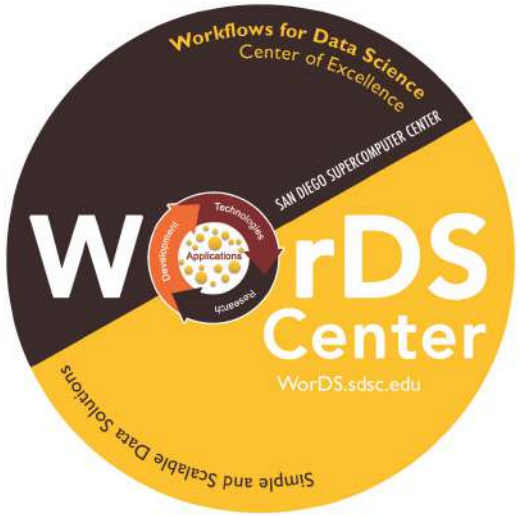
- Parallel AI on the Edge needs a full cyberinfrastructure and continuum computing/data architecture to support it.
  - Storage, programming models, and end-to-end data pipelines are a big part of it!

- Lots of new research challenges
  - Scalable processing methods at the continuum (also think downscale!)
  - Dynamic composability
  - Middleware
  - Collaboration methodologies

UC San Diego

Contact: *Ilkay Altintas, Ph.D.*
Email: ialtintas@ucsd.edu

SDSC SAN DIEGO SUPERCOMPUTER CENTER



https://words.sdsc.edu/publications

NBCR

it²

Questions?

NIH National Institutes of Health

NSF

U.S. DEPARTMENT OF ENERGY | Office of Science

The presented work is collaborative work with many wonderful individuals, and parts of it are funded by NSF, DOE, NIH, UC San Diego and various industry partners.