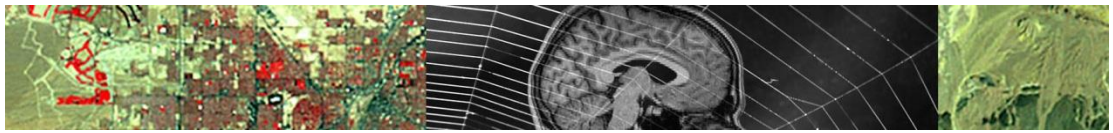
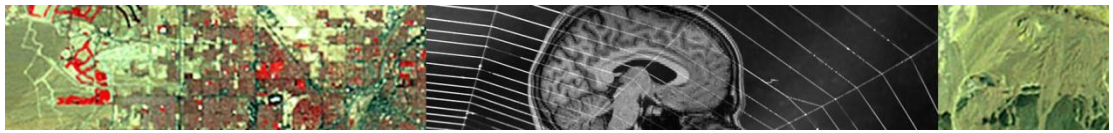
spatial@ucsb
perspectives for teaching and research



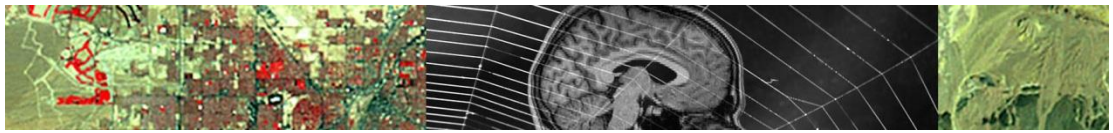
Big Data and GIS

Michael F. Goodchild
University of California
Santa Barbara



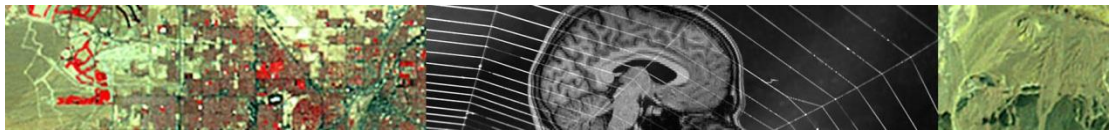
Two sections

- Data
  - Big Data, new geospatial sources
- Software
  - the Cloud, GIS as platform



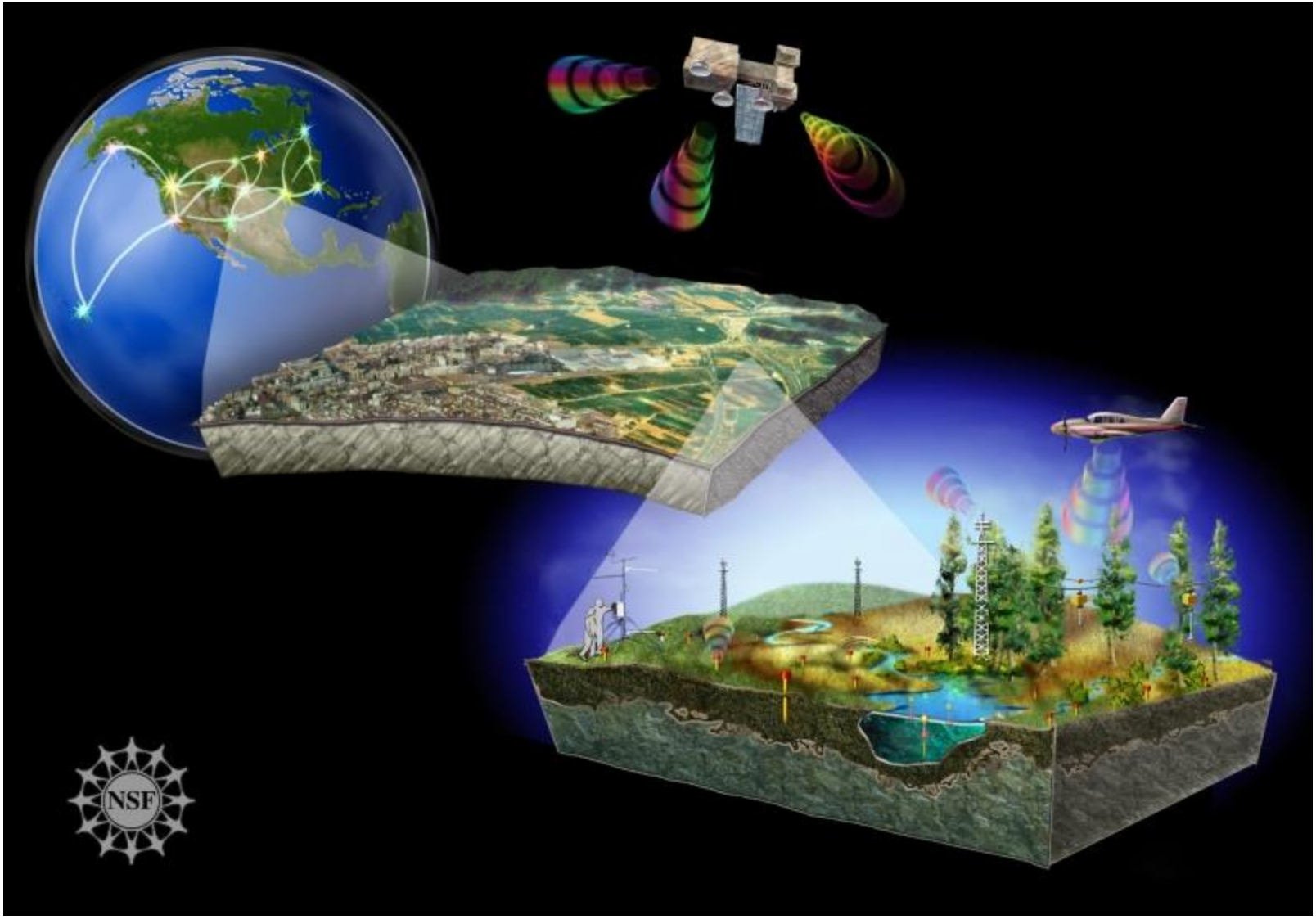
Overlapping ideas

- Data-driven science
  - an abundance of data
  - complex questions and problems
  - lack of faith in simple theories
    - all the simple discoveries have been made
    - the end of theory
  - the value of successful predictions
- Big Data
  - using multiple sources of data
  - not rigorously sampled
  - little quality control

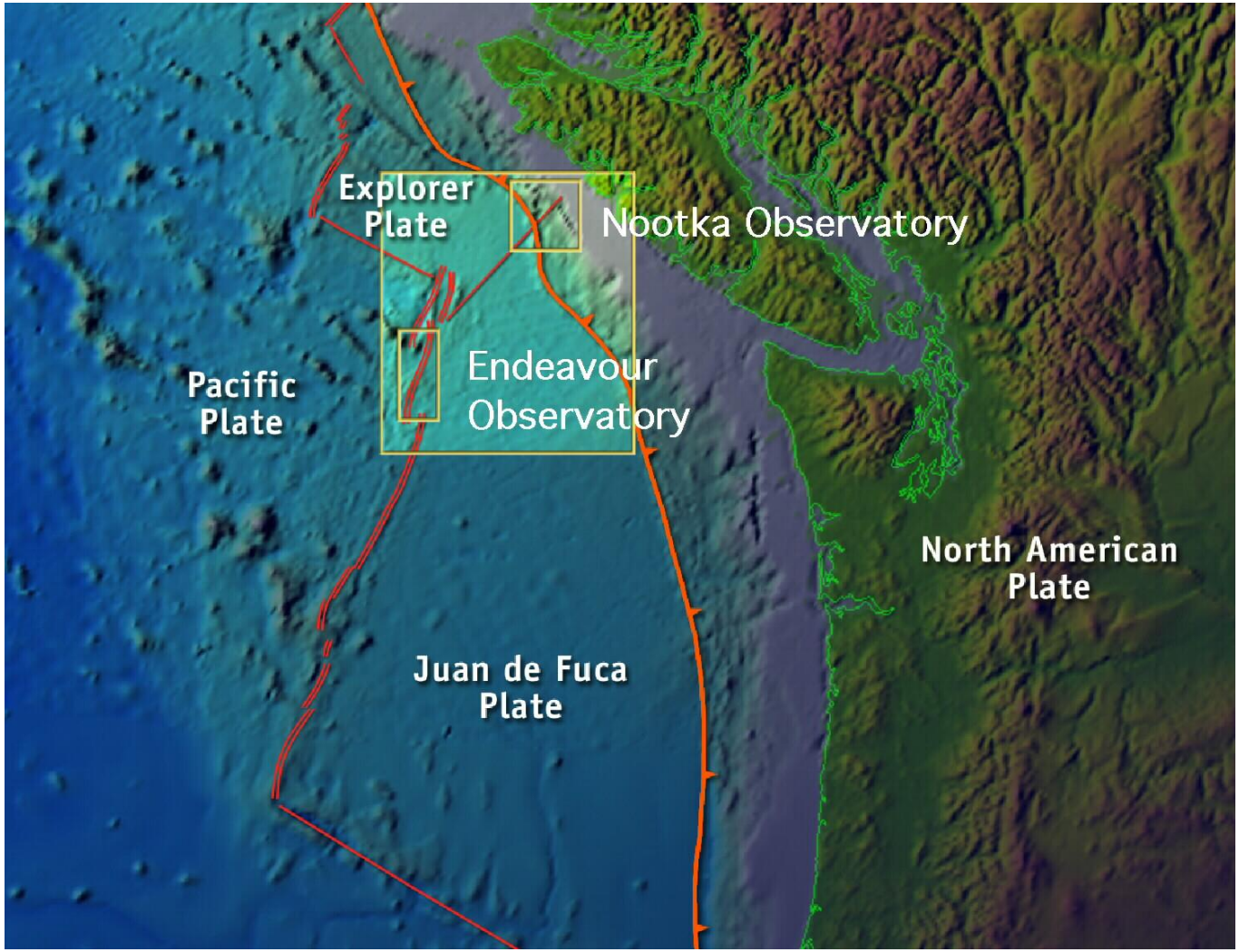


Volume

- Yes
  - Landsat since 1970s order exabytes
  - order 10^4 surveillance cameras in London
    - capturing order 10^6 faces for recognition and tracking
    - at 10 Hz
    - order 10^7 faces or 10^{12} bytes per second to be analyzed, identified, and linked into tracks
    - a petabyte computing problem
- But we divide and conquer
  - Landsat analysis by scene
  - partition the surveillance task into regions
  - we also aggregate, abstract, generalize



Environmental sensors in science: the US National Science Foundation's NEON project: neoninc.org



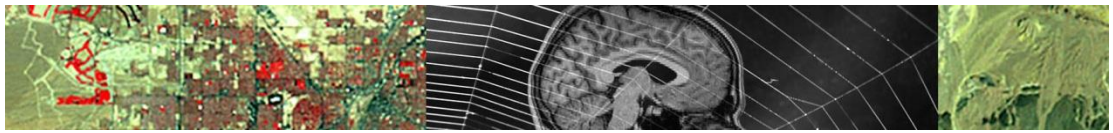
Underwater observatories: Monterey Bay Aquarium Research Institute: mbari.org

Mapping with kites, balloons, and micro-drones

step 1 Collect a Specimen

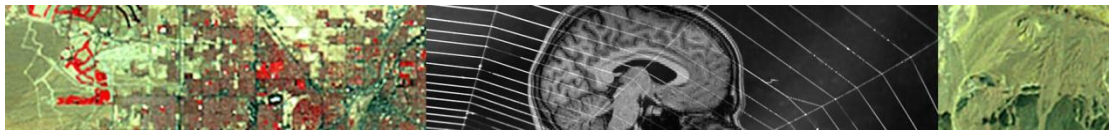
(1) Moisten a Q-tip with solution (like water or a mild detergent)
(2) Swab your surface with the moistened Q-tip
(3) Place the swab in a sealed container, like a plastic bag
(4) Record information about your specimen collection, such as date, time, geolocation, and weather.
(5) Mail your specimen to a lab that specializes in sequencing, such as this one

http://www.instructables.com/id/Mapping-Microbes/#step1



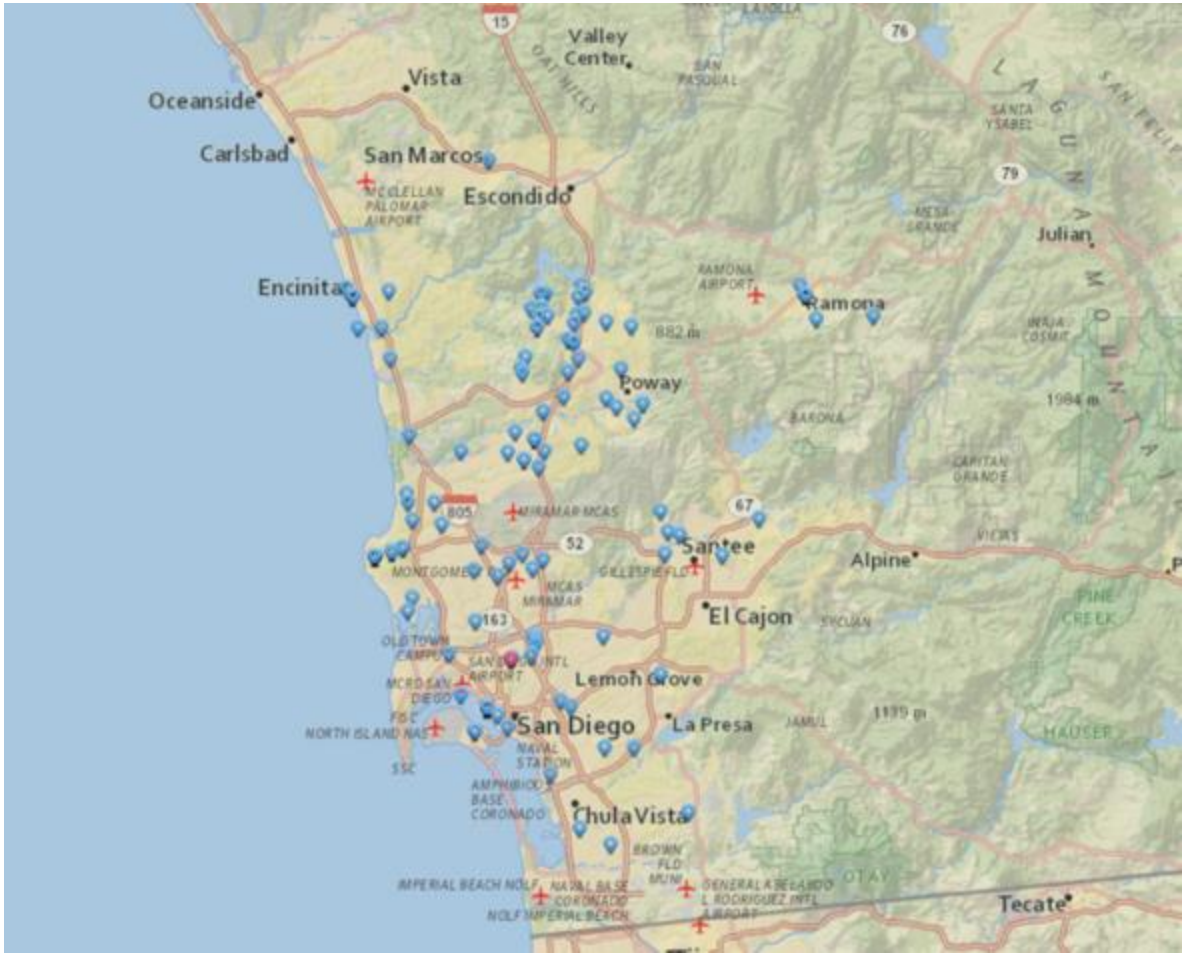
Velocity

- Big Data is also Fast Data
  - real-time sensing
    - submeter images available in near-real time
    - sensor networks
- Why might this be important?
  - science is typically slow, leisurely
  - a discovery that is only true now, and no longer true next week, is not of interest
    - and the same applies to location
  - fast data is not likely to advance science
    - at least as we currently understand science



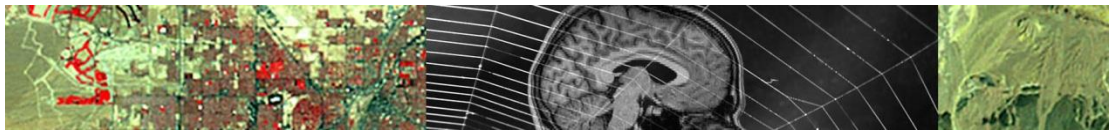
A broader view of science

- GIScience
  - the study of the fundamental properties of geographic information
    - the knowledge that is implemented in GIS
- The advancement of science in other domains through the application of GIScience knowledge
- The development of methods
  - that can be used to make discoveries
  - that can be used to solve real-world problems
    - that may be time-critical



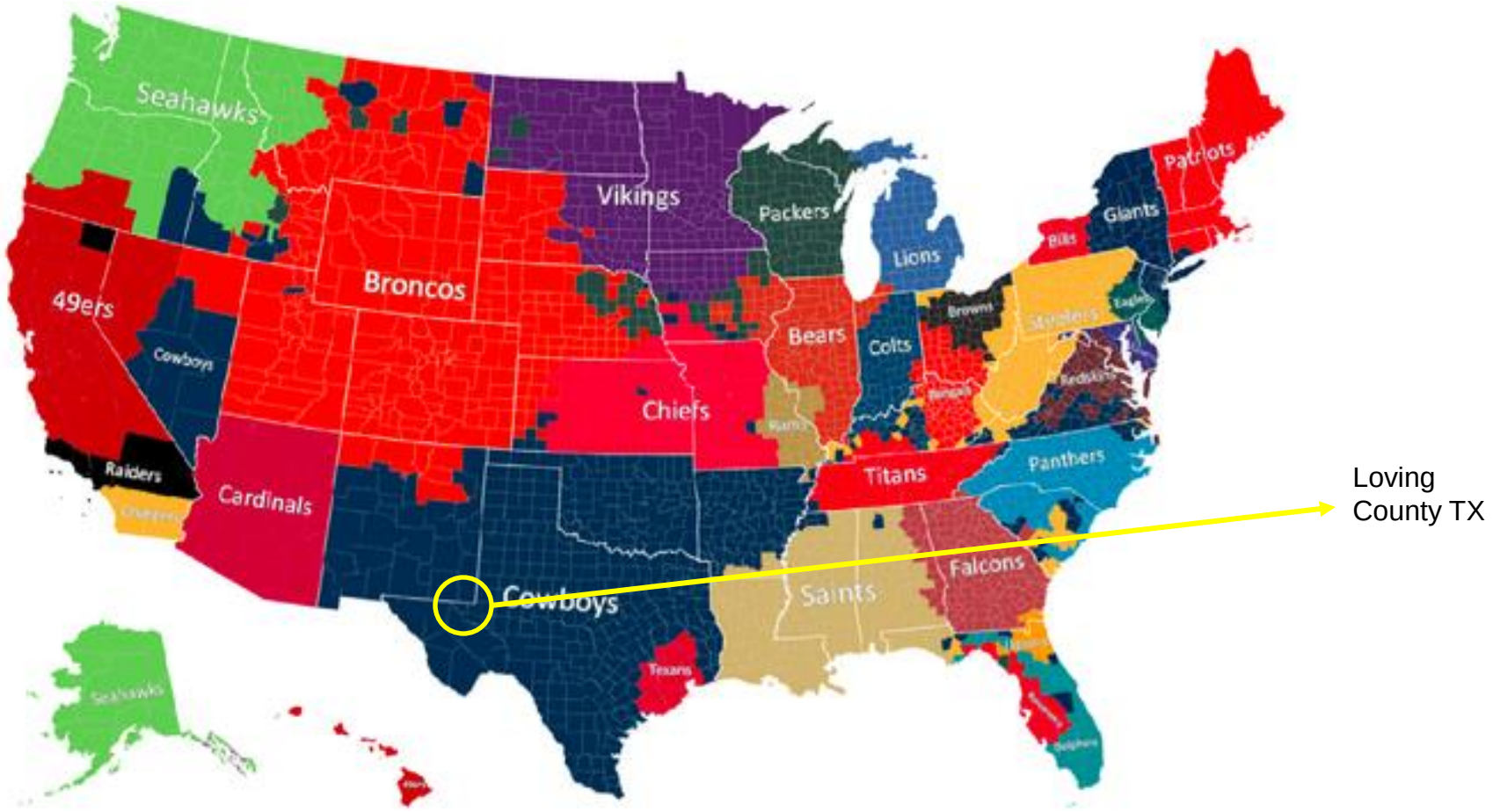
Tweets about wildfire in the San Diego area by Ming Tsou

http://www.engineering.sdsu.edu/sdsu_newscenter/news.aspx?s=75219



Variety

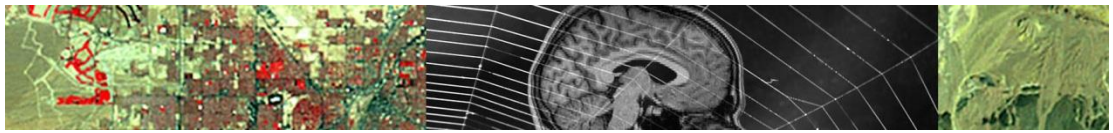
- Data from disparate sources
  - little or no quality control
  - no systematic experimental design
    - no recognizable sampling scheme
  - little or no documented provenance
- Perhaps a fourth V: veracity or validity?
  - or the lack of it



"The Geography of NFL Fandom"

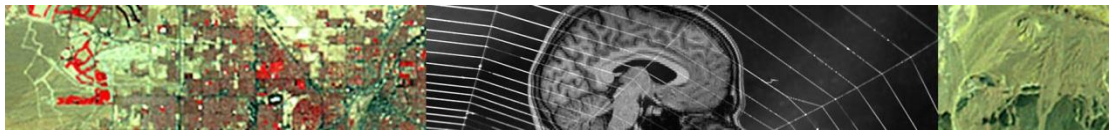
"This map displays Facebook fans of NFL teams across the United States. Each county is color-coded based on which official team page has the most 'Likes' from people who live in that county." (Facebook)

http://www.theatlantic.com/technology/archive/2014/09/the-geography-of-nfl-fandom/379729/



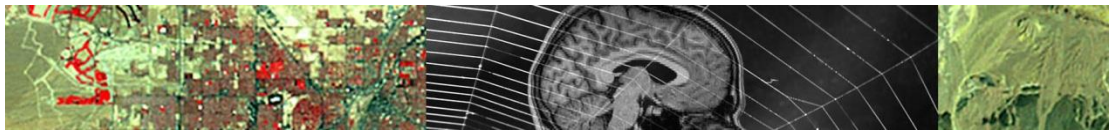
Loving County, TX

- **Loving County** is a county in the U.S. state of Texas. As of the 2010 census, the population was 82, making it the least populous county in the United States. Owing partly to its small and dispersed population, it also has the highest median per capita and household income of any county in Texas. Loving County has no incorporated communities; its county seat and only community is Mentone.



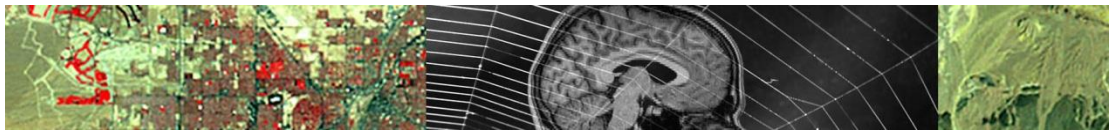
The value of Big Data

- H.J. Miller and M.F. Goodchild (2014) Data-driven geography. *GeoJournal*. DOI: 10.1007/s10708-014-9602-6.
- Hard science:
  - rigorous experimental design, hypothesis-driven, formal analysis of data quality
- Soft science:
  - data used for exploration, hypothesis generation, practical problem-solving
  - no attempt at formal sampling
    - no generalization from samples to populations
- Does Big Data have more value for soft science?



The successes of Big Data

- Predictions
  - the winner of the Eurovision Song Contest
  - the stock market
  - outcomes of elections
- Early warning from social media
  - flu outbreaks
  - wildfires



Spatial prediction

- Predicting where (and when)
  - something will happen
  - some condition will exist
- Where will this hurricane make landfall?
- Where is this criminal most likely to live?
- What is my best route given traffic conditions?
- Where is a new school most needed?

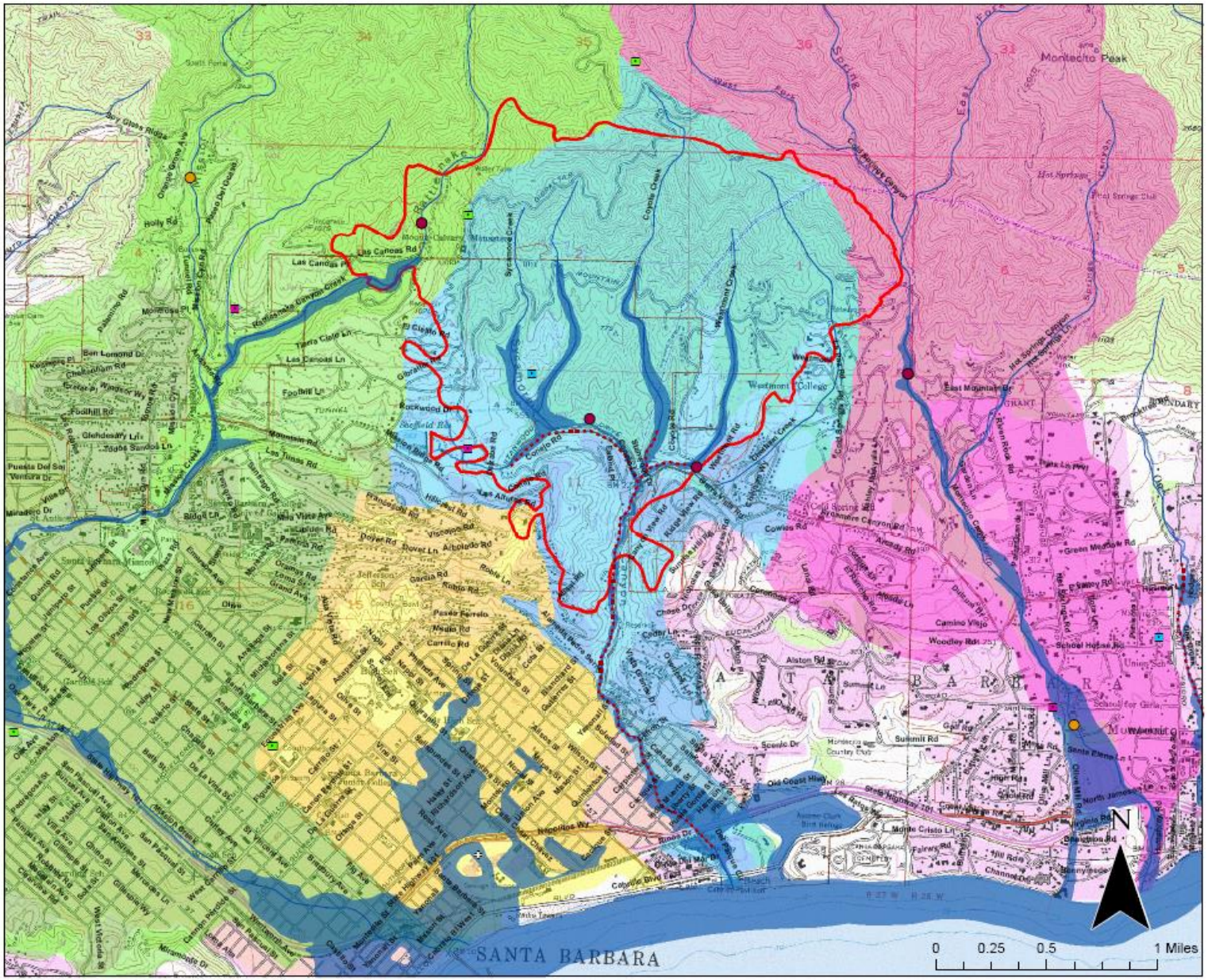
Tea Fire

Legend

- Debris Basin
- Debris Basin to be Cleaned
- Self-Recording Station
- Alert Station
- Observer Station
- Channel Clearing
- Sand Bags - 401 E. Yanonali St.
- Tea Fire Perimeter

Watersheds	Acres Burned	% of Watershed
Mission Canyon	183	2%
Laguna	8	1%
Sycamore Canyon	1,686	67%
Montecito Creek	64	1%

- City Boundary
- 100 year flood plain

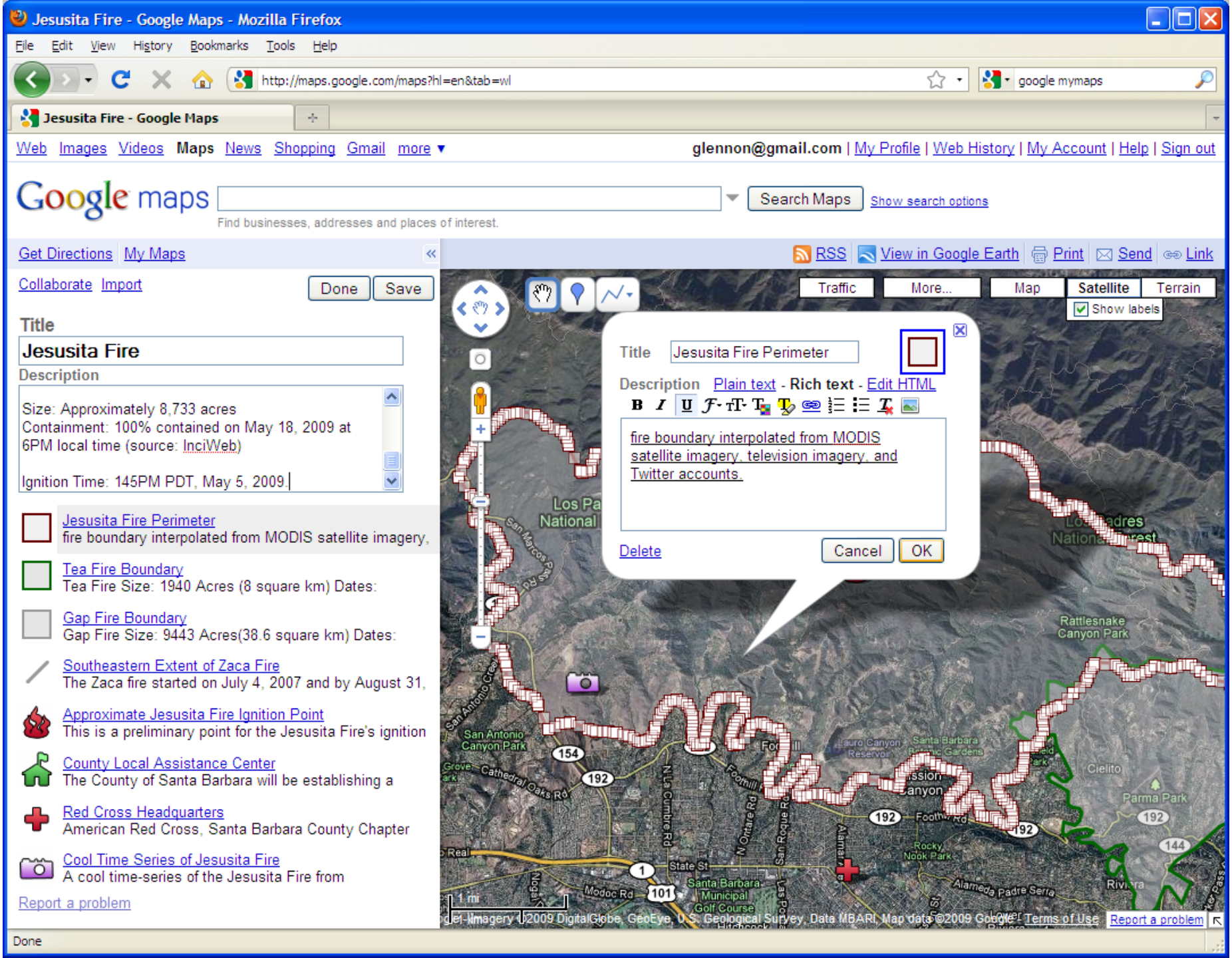



Timeline
Flipbook
List
Map
Full Screen
Embed Widget
Share:
Follow
FEMA: California Tea Fire
Nov 13, 2008
tea fire santa barbara | VPNewsLog
Nov 13, 2008 7:48 PM
The Tea Fire
Nov 13, 2008 11:24 PM
Fire destroys 80+ houses in Montecito, Santa ...
Nov 13, 2008
Santa Barbara "Tea" Fire
Nov 13, 2008 9:44 PM
Canned Spam in high demand - Crime Scene
Nov 13, 2008
Las Alturas Fire Montecito 2008 Tea Fire
Nov 13, 2008 8:42 PM
Wildfire Today: Update on Tea fire near Monte...
Nov 13, 2008
Lal Masjid of Bradistan «Pak Tea House
Nov 13, 2008 9:50 PM
BREAKING NEWS
MONTECITO
Tea Fire Westmont College
Nov 14, 2008 3:11 AM
Tea Fire (Santa Barbara & Mon...
Nov 14, 2008 7:39 AM
Tea Fire SB fr
Nov 14, 2008
Oprah's house Montecito: Oprah's house fire '...
Nov 13, 2008
Tea Fire Montecito/Santa Barbara pt. 1
Nov 14, 2008 12:00 AM
Santa Barbara Tea Fire Time Lapse
Nov 14, 2008 5:46 AM
Water Fill Up Helicopter Tea Fire Santa Barb...
Nov 14, 2008 10:51 AM
Montecito Fire (See Pic and Video Footage) - ...
Nov 13, 2008
Tea Fire Santa Barbara Helicopter Water Drop
Nov 14, 2008 1:55 AM
365 days: Abwarten und Tee trinken
Nov 13, 2008
Santa Barbara "Tea Fire" Helicopter
Nov 14, 2008 2:00 AM
3 pm
4 pm
5 pm
6 pm
7 pm
8 pm
9 pm
10 pm
11 pm
Nov 14, 2008
2 am
3 am
4 am
5 am
6 am
7 am
8 am
9 am
10 am
11 am
12 pm
1 pm
2 pm
SHOW SOURCES

Fire Starts
Red Cross announces shelter opening
Twitter hashtag jesusitafire established
Intermittent power dips
First visual by engine crew
First helicopter arrives
Public Access TV reporting no evacuations
Large flames observed on eastern flank
First written county press rel
First severity assessment
Spatial@ucsb learns about fire
Caltrans begins road closures
Ethan map created
First official information
Flickr images start being posted
County confirms reverse 911 system initiated
Assessment from Amateur expert
Official naming of Jesusita Fire
second helicopter arrives
First overhead news coverage of fire
Traffic jams near Mission due to road closures
Media outlets relay offi
First Relief Agency Response
Residents report receiving reverse 911 calls
KEYT television begins coverage
Mandatory Evacuations in Eff
IndyAlert system sends fire message
Citizen observations of firefighting efforts begin
Traditional news coverage begins
Citizen online post complaining about lack of news
Citizen online posting begins about fire's location
Jesusita Fire
14hr
15hr
16hr
17hr

Jesusita Fire Timeline

http://arogi.com/fireline


Jesusita Fire - Google Maps - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://maps.google.com/maps?hl=en&tab=wl
google mymaps
Jesusita Fire - Google Maps
Web Images Videos Maps News Shopping Gmail more ▾
glennon@gmail.com | My Profile | Web History | My Account | Help | Sign out
Google maps
Search Maps
Show search options
Find businesses, addresses and places of interest.
Get Directions | My Maps
Collaborate Import
Done Save
Title
Jesusita Fire
Description
Size: Approximately 8,733 acres
Containment: 100% contained on May 18, 2009 at 6PM local time (source: InciWeb)
Ignition Time: 145PM PDT, May 5, 2009.
Jesusita Fire Perimeter
fire boundary interpolated from MODIS satellite imagery,
Tea Fire Boundary
Tea Fire Size: 1940 Acres (8 square km) Dates:
Gap Fire Boundary
Gap Fire Size: 9443 Acres(38.6 square km) Dates:
Southeastern Extent of Zaca Fire
The Zaca fire started on July 4, 2007 and by August 31,
Approximate Jesusita Fire Ignition Point
This is a preliminary point for the Jesusita Fire's ignition
County Local Assistance Center
The County of Santa Barbara will be establishing a
Red Cross Headquarters
American Red Cross, Santa Barbara County Chapter
Cool Time Series of Jesusita Fire
A cool time-series of the Jesusita Fire from
Report a problem
RSS View in Google Earth Print Send Link
Traffic More... Map Satellite Terrain
Show labels
Title Jesusita Fire Perimeter
Description Plain text - Rich text - Edit HTML
fire boundary interpolated from MODIS satellite imagery, television imagery, and Twitter accounts.
Delete
Cancel OK
Los Padres National Forest
San Marcos Pass Rd
San Antonio Creek
San Antonio Canyon Park
Cathedral Oaks Rd
N La Cumbre Rd
Foothill Rd
N Ontare Rd
San Roque Rd
Alameda Padre Serra
Lauro Canyon Reservoir
Santa Barbara Botanic Gardens
Rattlesnake Canyon Park
Cielito
Parma Park
Rocky Nook Park
Riviera
State St
Modoc Rd
Santa Barbara Municipal Golf Course
154
192
101
1
144
1 mi
Imagery ©2009 DigitalGlobe, GeoEye, U.S. Geological Survey, Data MBARI, Map data ©2009 Google Terms of Use
Report a problem
Done

Hits	Source
595673	Jesusita Fire (Ethan)
188308	SBC Jesusita Fire Santa Barbara, CA (Robert O'Connor - fire news blog)
89214	Jesusita Fire Map (Randy - Independent.com)
67525	Jesusita Fire in Santa Barbara - LA Times map (Los Angeles Times)
27777	Map of burned homes in Santa Barbara (Los Angeles Times)
26330	Jesusita Fire Evacuation Areas: Approximation (COSB)
25454	Santa Barbara 'Jesusita Fire' (ABC7 Eyewitness News)
19592	Jesusita Fire - Santa Barbara (lanewspace)
2446	Santa Barbara Damaged Homes 2008 (Los Angeles Times, note: mapped for comparison with Jesusita)
2048	Jesusita Fire (longhairedhippy)
1314	Santa Barbara Fire Evacuation (Gary);
962	Jesusita Fire in Santa Barbara (ABC30 Action News)
788	Wildfire ~ Santa Barbara (Buffalo)
505	Closure map - Jesusita Fire in Santa Barbara (Los Angeles Times)
461	Untitled (Matthew, note: discovered via google.com.mx);
396	Jesusita Fire Structure Damage (Paul Bartsch);

Why did the Ethan map resonate?

Perhaps,

- maps as a synthesis and contextualization of many information inlets

- Editable by anyone

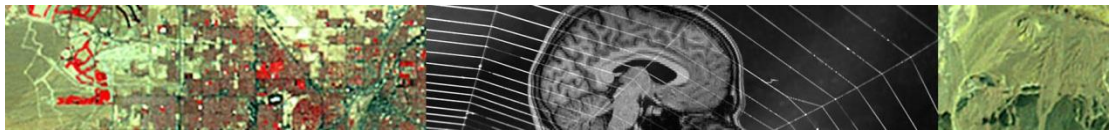
- Maintained by passionate, somewhat expert core group

- Interactive: map comment section offered community discussion

- Continuity (to contrast news sites that often created a new map for each new story)

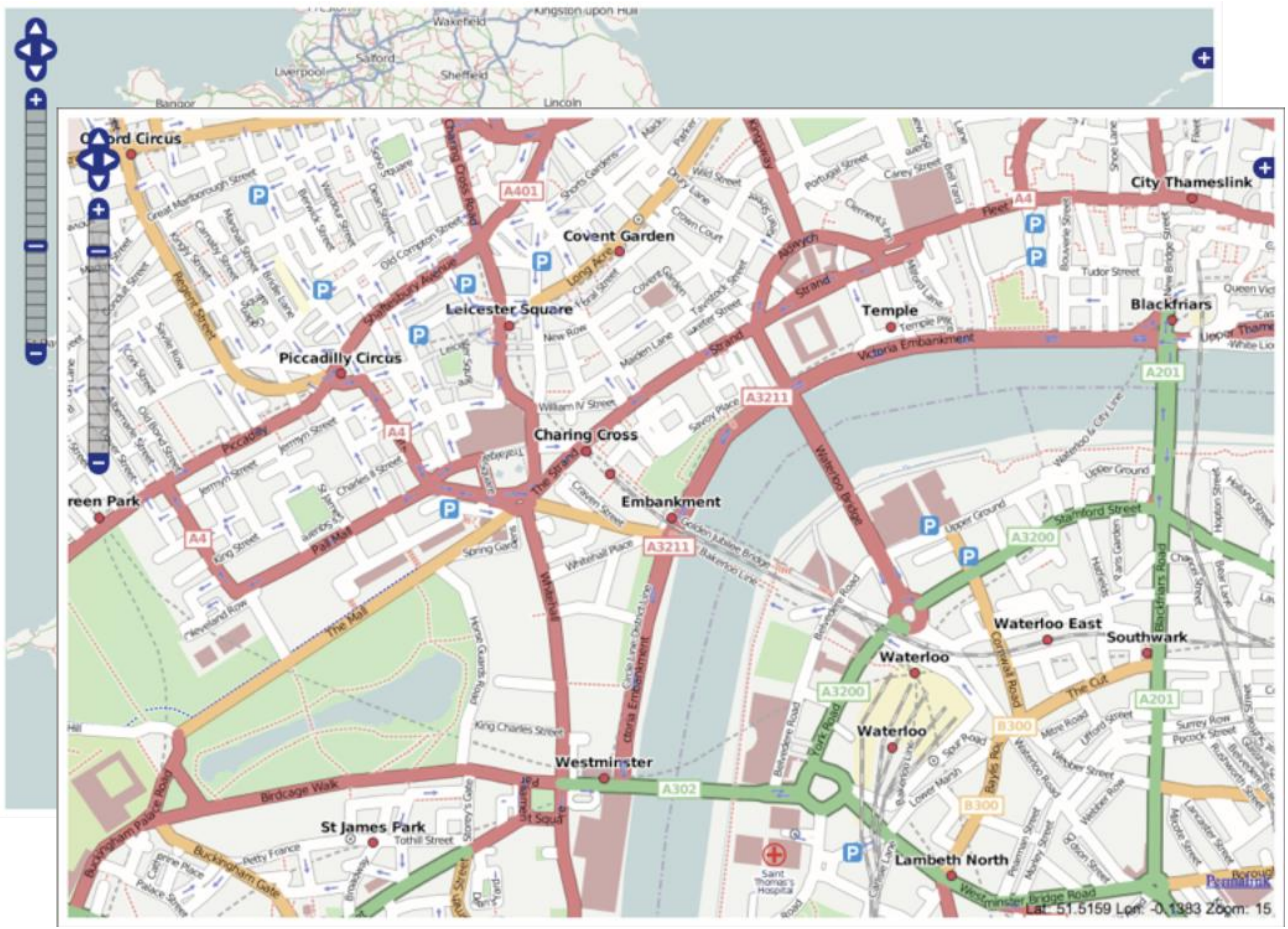
- Fast, continuous updates

- Ubiquity and ease of Google Maps UI

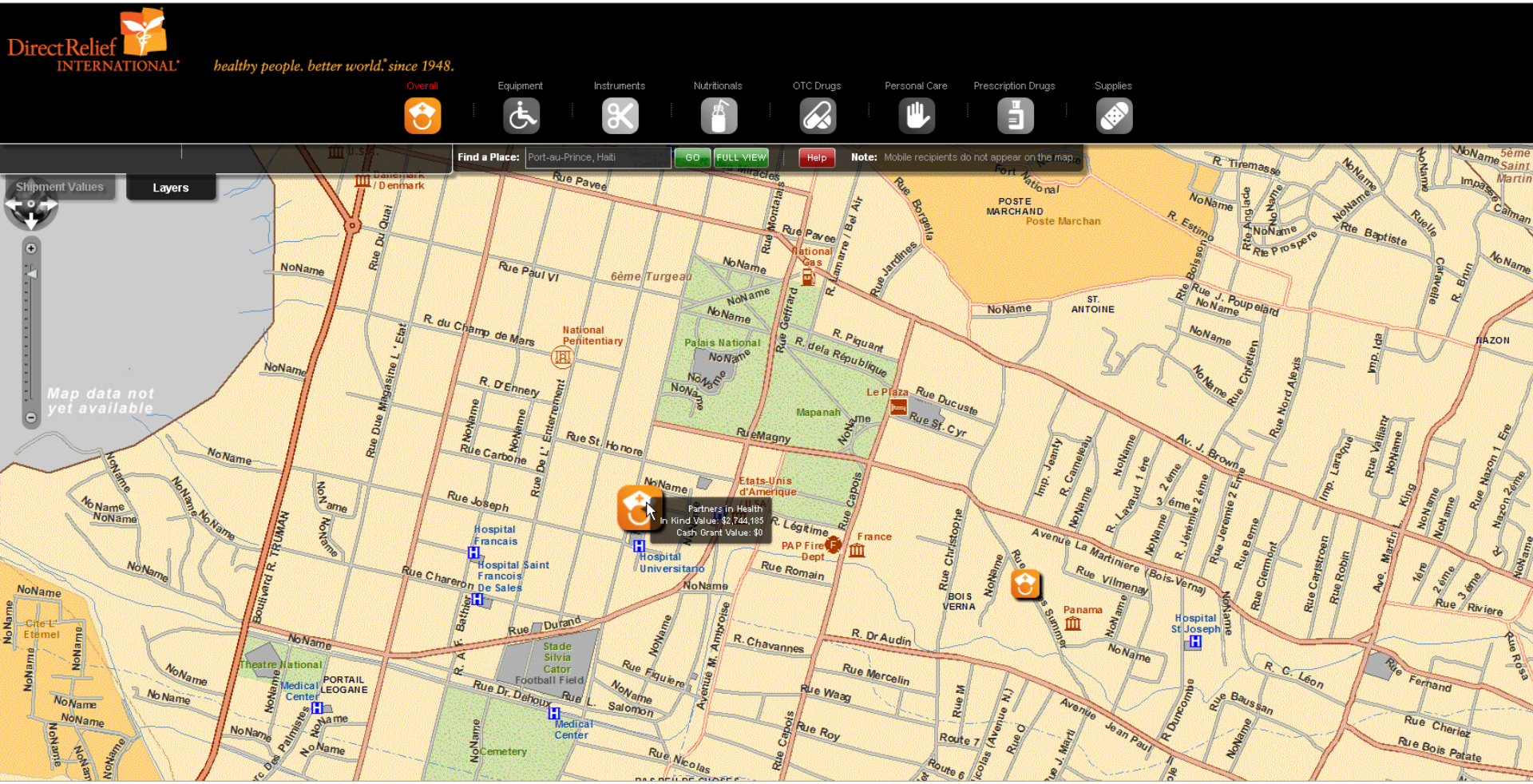


Neogeography

- Citizens as both users and producers of geospatial data
  - volunteered geographic information
  - crowdsourcing
  - multiple sources, little quality control, not rigorously sampled
- Maps for the individual
  - user-centered
  - transitory
  - the view from the ground
  - delivered on small devices

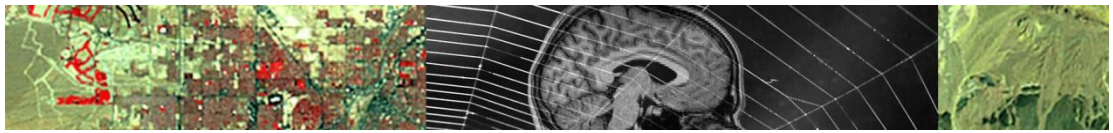


www.openstreetmap.org



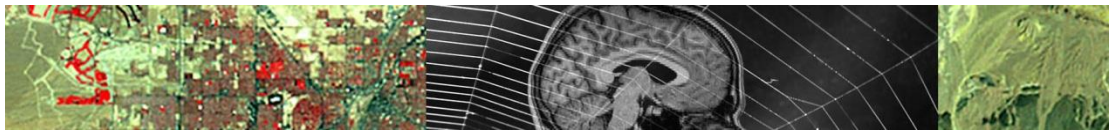
http://www.directrelief.org/Flash/HaitiShipments/Index.html

nationalmap.gov



“Small” data

- Rigorously sampled to provide representation
- Quality control throughout
- Abundant metadata
- Generalizable results
- Census data

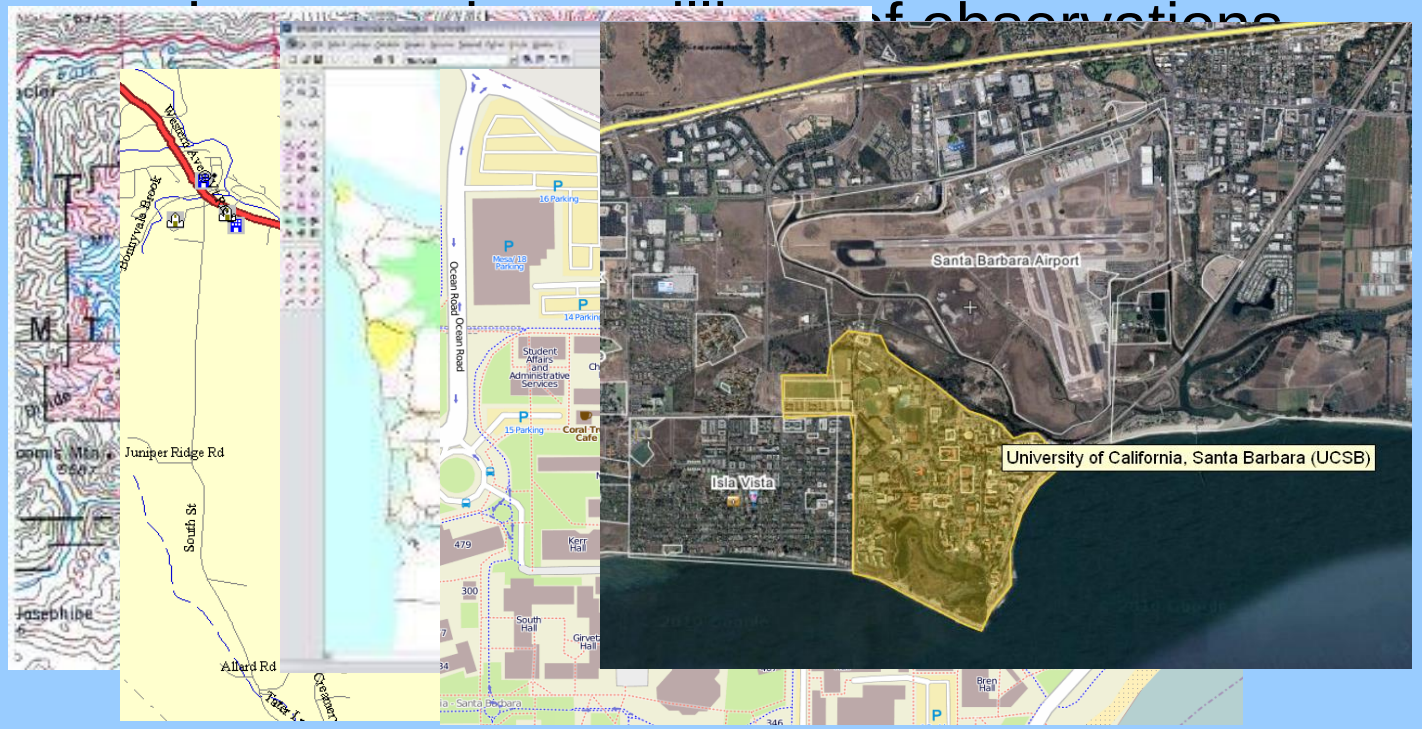


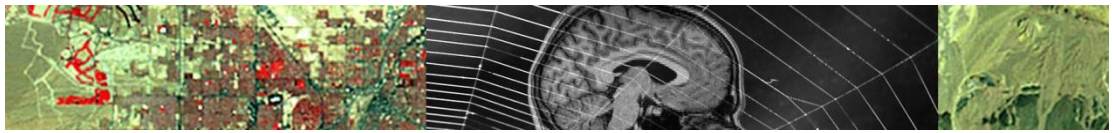
Big data

- The Three Vs
  - volume, variety, velocity
- No sampling scheme
  - samples often self-identified
- No generalizability
- No metadata
- No quality control
  - though quality can be excellent
  - more up-to-date

Traditional geographic information

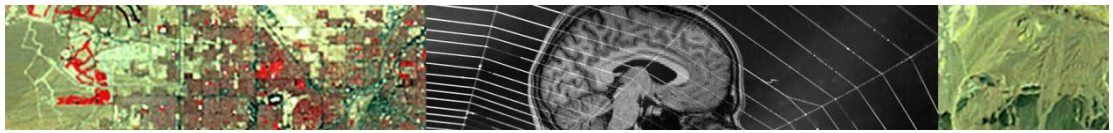
- Maps, atlases, globes
  - highly synthesized, compiled, abstracted
  - often rich quantitative attributes





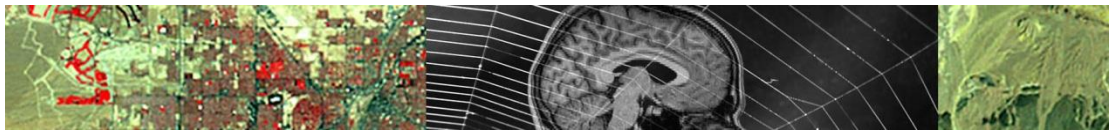
Hidden synthesis

- By experts in traditional authoritative production of geographic information
  - the process by which observations are synthesized into statements about points, polylines, or polygons is typically hidden
  - no provenance



The new synthesis

- Software
- Actions of volunteers
- Disparate purported facts of varying quality and reliability
- How to harden?



The crowd solution

- Linus's Law
  - the more eyes to review, the more accurate
  - works for popular facts
- Corroborating reports
  - but how do we know the information is the same?
  - distinct georeferences
  - distinct descriptions
- Geographic facts may be obscure
  - little-known areas of the world
    - or not so obscure

Menu

World / USA / California / Isla Vista , 2km from center

Coordinates: 34°25'11"N 119°52'38"W

Glen Annie Golf Course (Goleta)

405 Glen Annie Rd.
Santa Barbara, CA 93117
(805) 968-6400

www.glenanniegolf.com/golf/proto/glenanniegolf/

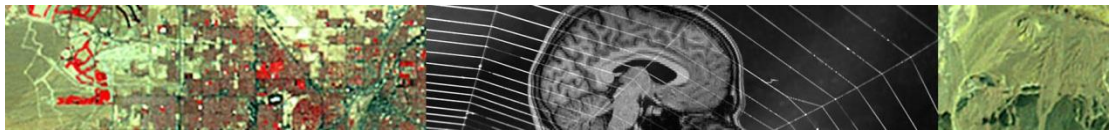
Glen Annie Golf Club is a championship golf course with first class amenities situated in the rolling foothills of scenic Santa Barbara. This challenging golf layout is enhanced by breathtaking panoramic views of the Pacific Ocean and Channel Islands and is always maintained with the highest standards.

...and I'm a Mormon
I'm Bruce Summerhays. I'm a pro golfer, a husband, & I'm a Mormon.
youtube.com/mormon

This article is protected.
Category: golf course

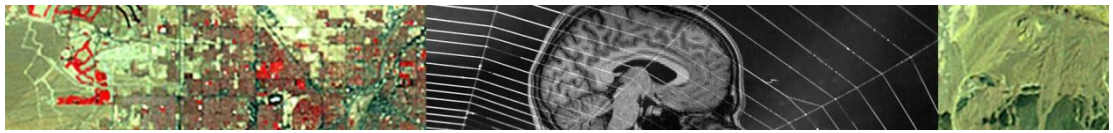
Address: Glen Annie Road, 405

Permalink to this place



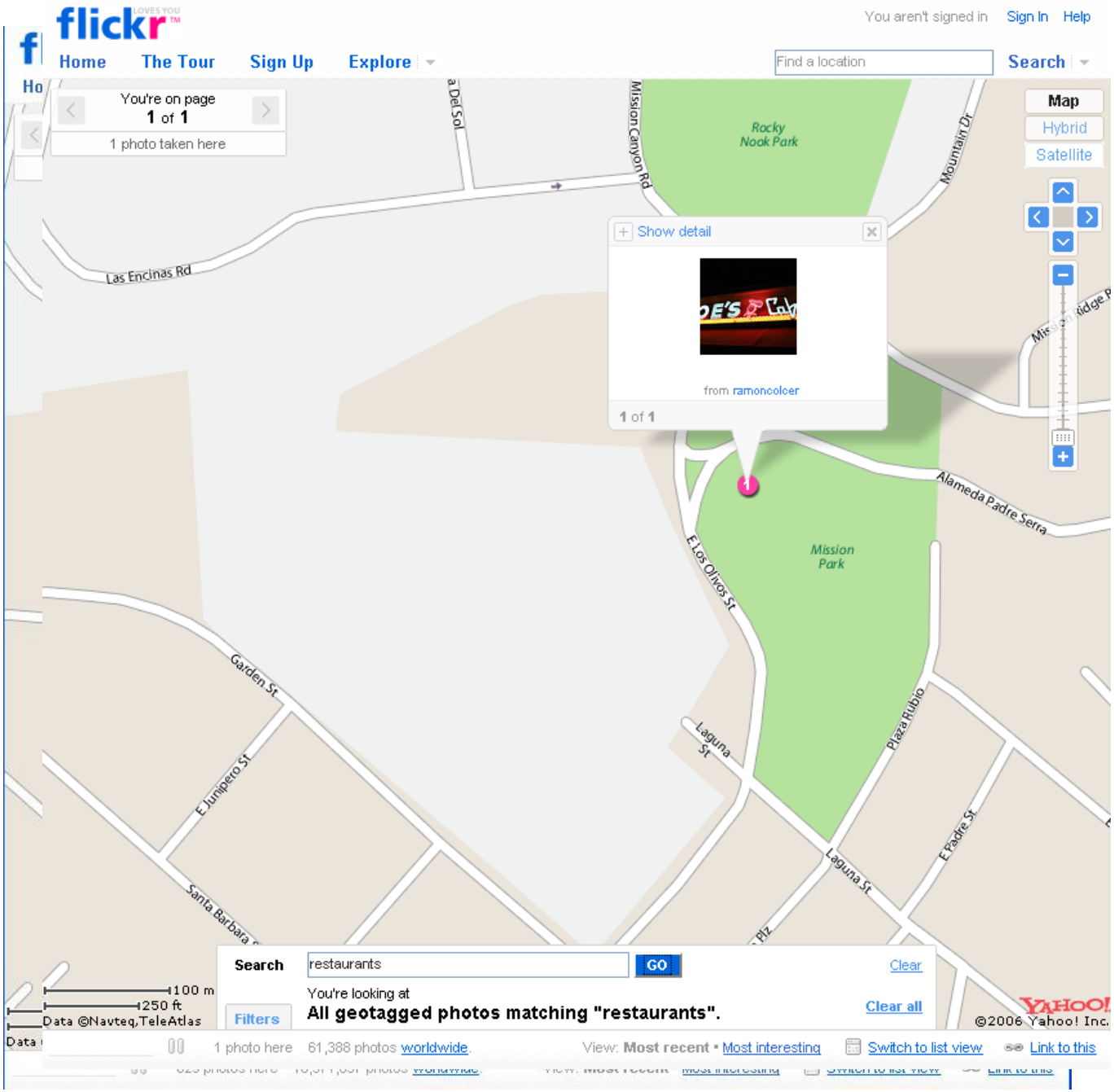
The social solution

- Who can be trusted?
- A hierarchy of moderators and gate-keepers
  - all volunteered facts referred up the hierarchy
- A social structure
  - promotion based on track record
  - heavy, accurate contributors promoted
  - e.g., Wikipedia, OSM
  - top levels of Google MapMaker reserved for Google staff



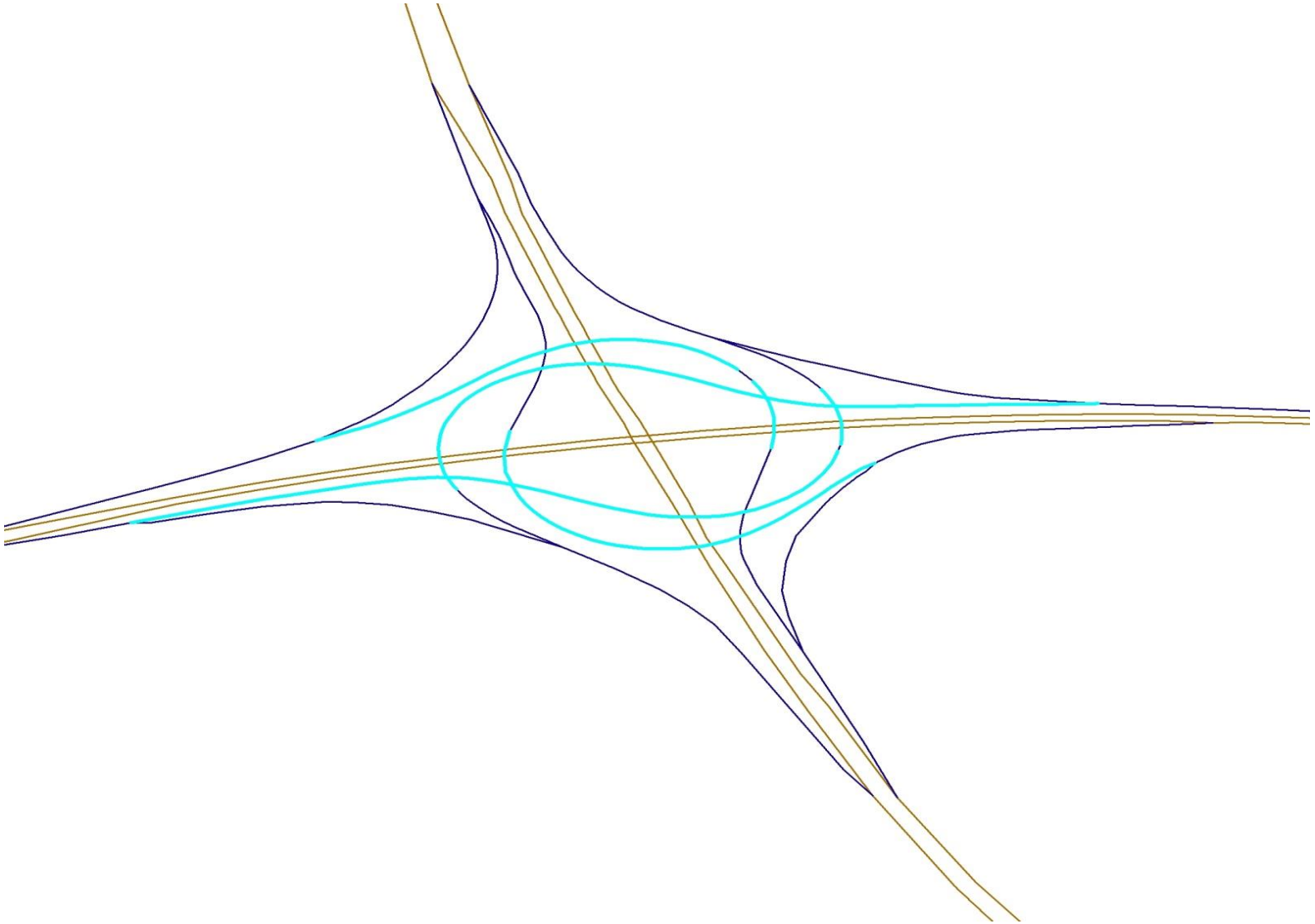
The geographic solution

- How can we know if a purported geographic fact is false?
  - because it violates the rules by which the geographic world is constructed
  - the syntactic rules
  - compare language rules, the sentence structure of English
- What are those rules?
  - essential, fundamental geographic knowledge



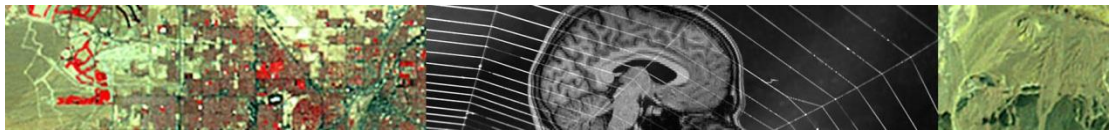
www.flickr.com

Angle larger than 30 degrees



Bergonia

bergonia.org

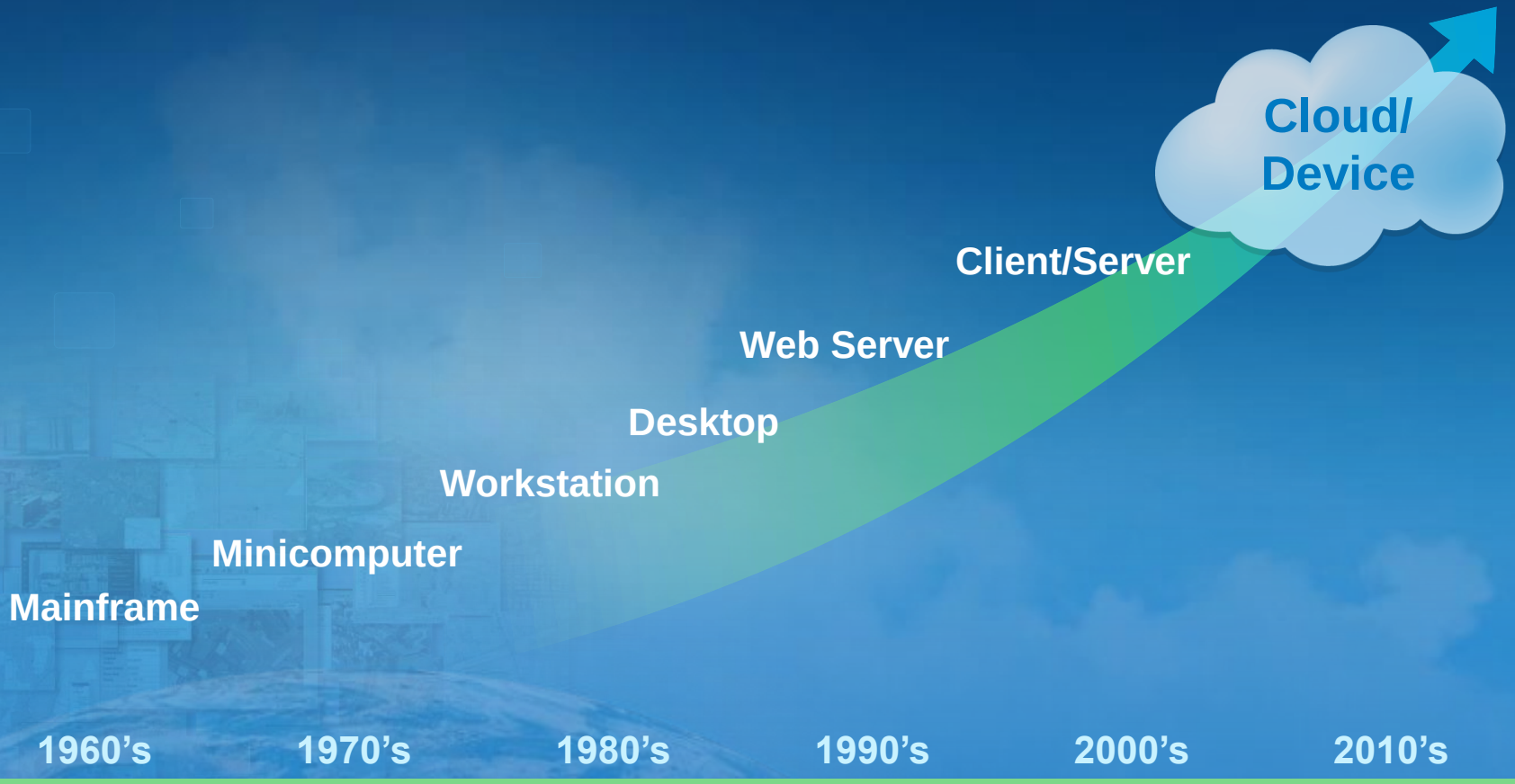


Formalizing geographic knowledge

- To enable automated triage of asserted facts
- To enable automated synthesis
  - into products that more closely resemble the traditional ones

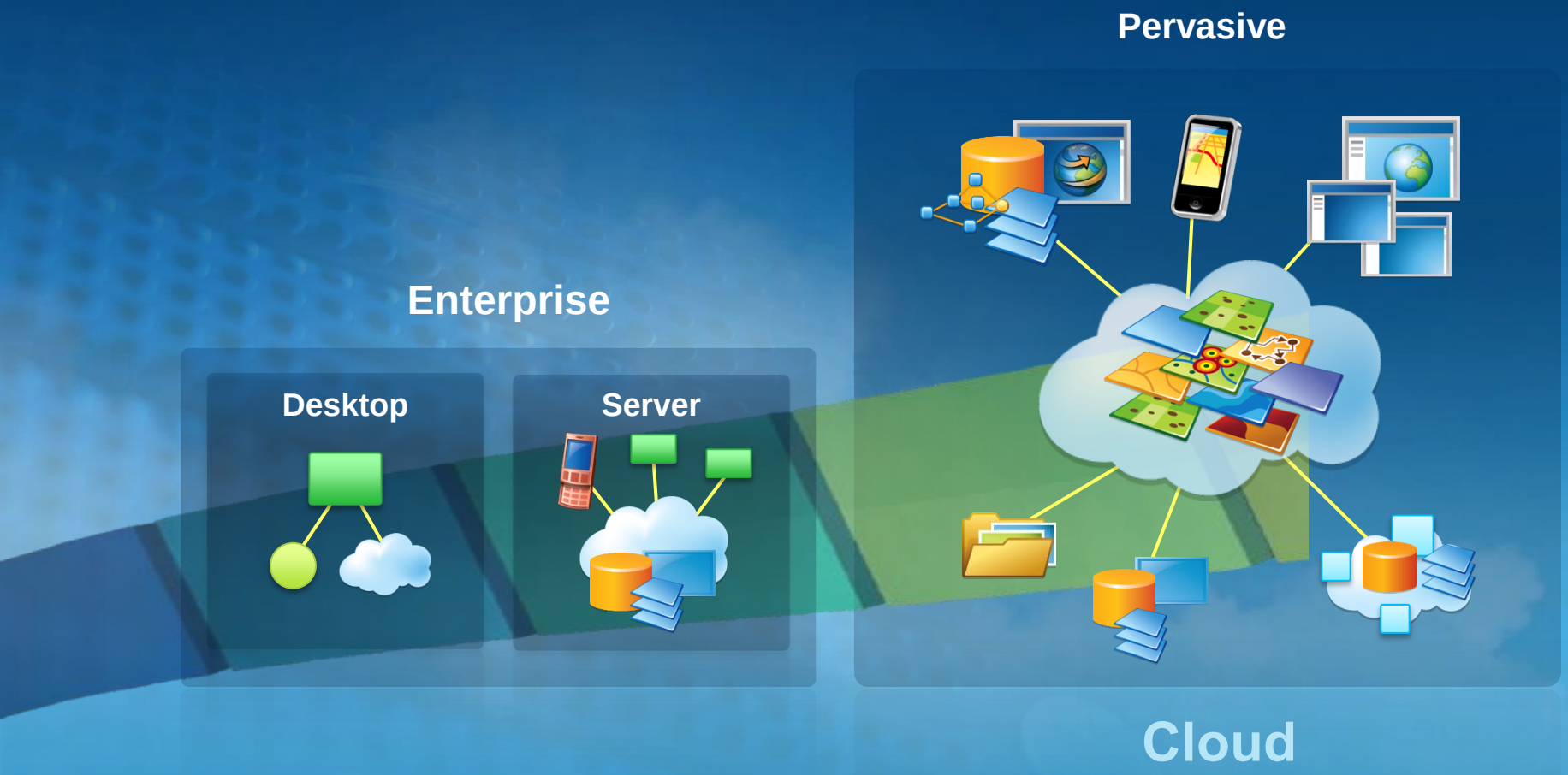
GIS Is Transforming Into a Platform

Integrating Software and Services



This New Platform

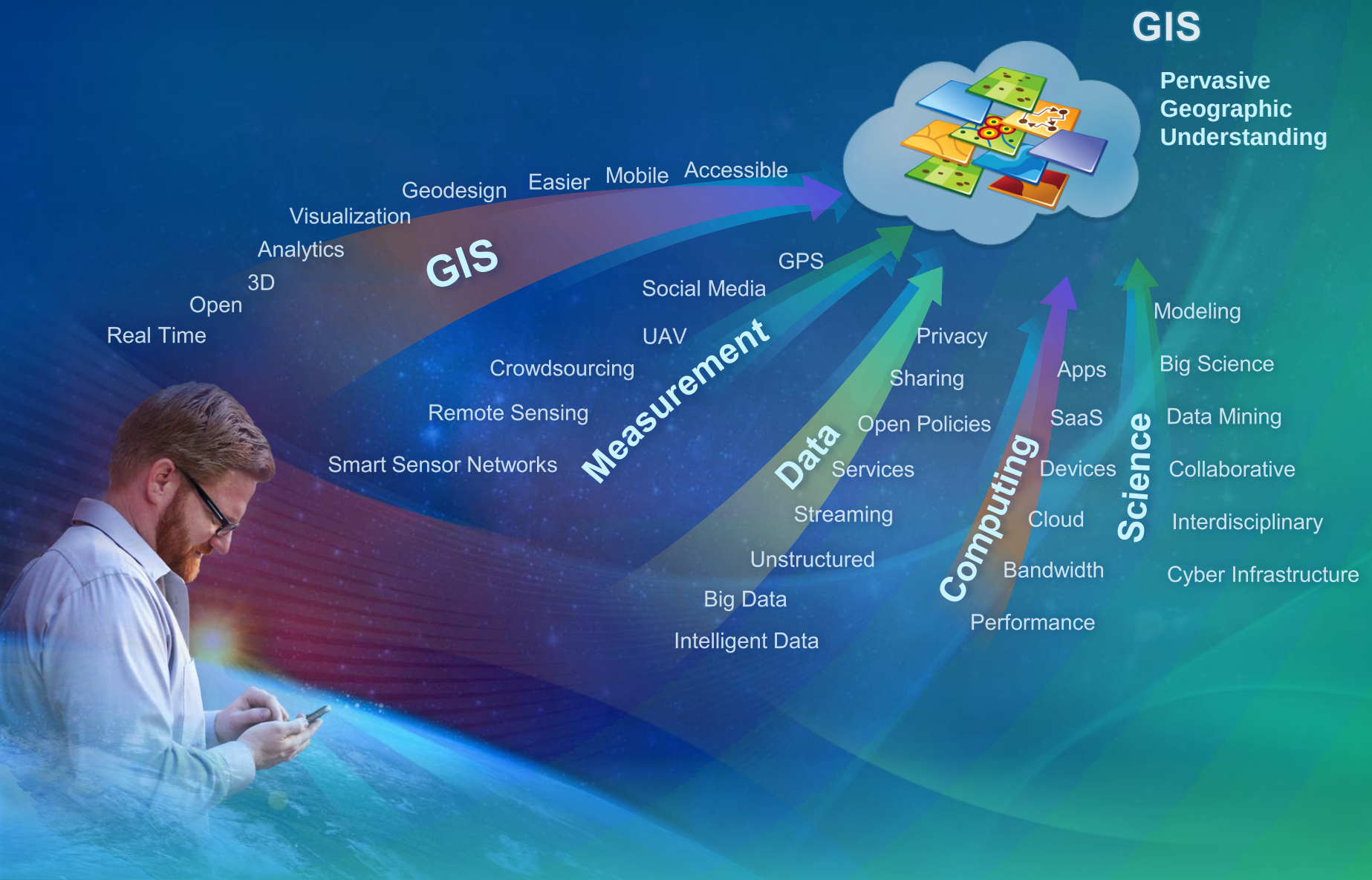
Connects and Leverages Existing GIS Investments



Providing Mapping and GIS As a Service

To the Entire Organization

This Platform Leverages Many Trends



Enabling Pervasive Access

Integrating Traditional GIS with a Whole New World of Apps

Enabling New User Experiences

- **Land Records**
  - [x] Parcel Boundary
  - [ ] Parcel Lines
  - [ ] Line Length -- Parcel Lines
  - [ ] Subdivisions
  - [ ] Subdivision Text
  - [ ] Acreage Text
  - [ ] Dimensions
  - [x] Easement Text
- **Administrative**
  - [x] County Boundary
  - [x] City Limits
  - [ ] OverLayZones
  - [ ] LandUse2004
  - [ ] Landuse2009
  - [ ] Future Land Use Plan
  - [ ] Voter Precincts
  - [ ] Fire Fee Districts
  - [ ] TownShip
  - [ ] Elementary School Districts
  - [ ] Middle School Districts
  - [ ] High School Districts
  - [ ] Agricultural District
  - [ ] Sanitary District
  - [ ] Postal
  - [x] Jurisdiction Text
  - [ ] Zoning
  - [ ] NC House Districts
  - [ ] Electoral Districts
  - [ ] Voting Tabulation District

The Platform Integrates All Types of Geospatial Information

Using Web Maps to Normalize the Information . . .

Web Maps Are Fundamental

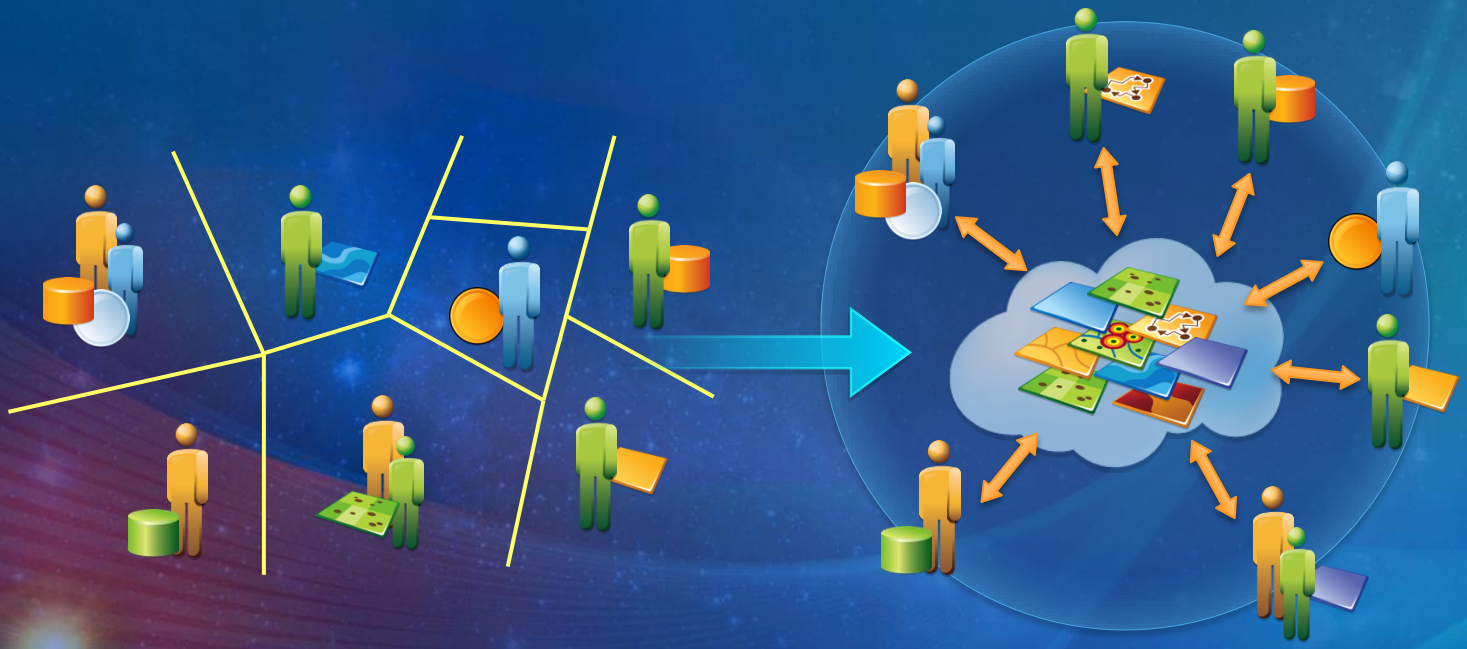
Providing a New Medium for Organizing and Publishing

Supporting Visualization,
Query, Editing, and Analysis

Making Geographic Information Available Anywhere

This Platform Transforms Organizations

Breaking Down Barriers Between Workflows, Disciplines, and Cultures

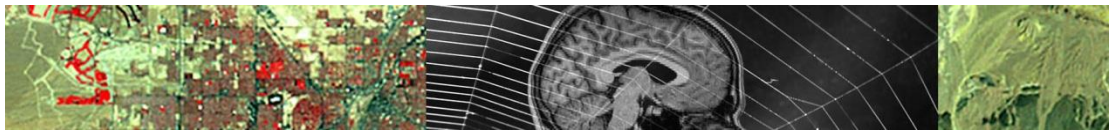


Enabling Collaboration, Sharing, and Holistic Approaches

GIS Is at a Major Turning Point

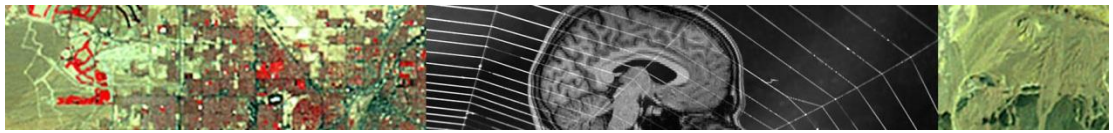
Becoming a Platform

Enabling Wide Scale Access and Use of GIS



Summary

- Big Data is relevant to GIS:
  - in the soft stages of science
  - in solving time-critical problems
  - in spatial prediction
- Big Data requires a change of scientific perspective
  - science driven by data rather than theory
  - *all* the data, not just the best data
  - prediction as a legitimate activity



Summary (2)

- We need to develop ways to harden Big Data
  - at electronic speed
- Synthesis may be the most important activity in GIScience in the future
- GIS is becoming a platform
  - an integrated set of Cloud services
  - ubiquitous access across all devices
  - making it easy to develop new applications
  - but with many open questions about privacy, data management