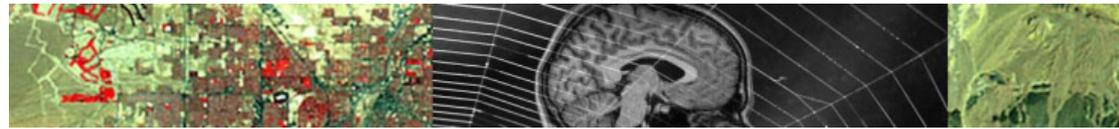


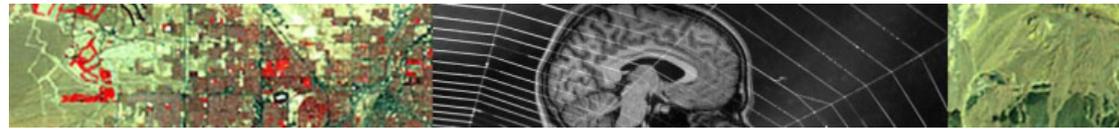
VGI Research Overview

Michael F. Goodchild
University of California
Santa Barbara



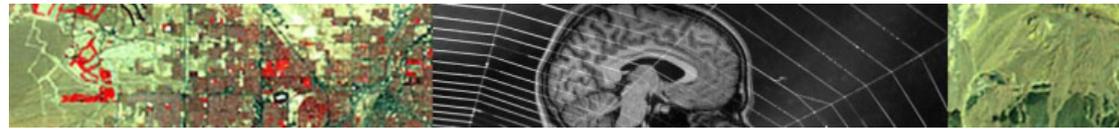
Two kinds of research

- Research about VGI
 - types of VGI
 - technical support
 - social context
 - data quality
- Research using VGI
 - exploring the world
 - answering questions
 - social and environmental science
 - citizen science



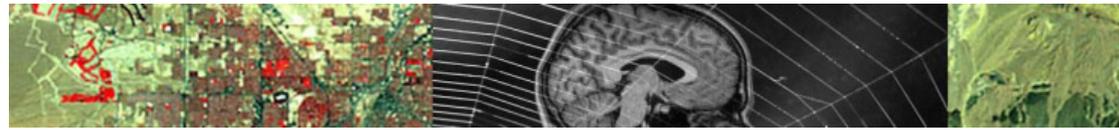
Types of VGI

- Volunteers replacing or augmenting traditional GI production
 - OSM, Wikimapia, Navteq, ...
- Volunteers producing new kinds of GI
 - Flickr, tweets, didyoufeelit, potholes, ...
- Involuntary or contributed GI
 - cellphone tracking, RFID, ...
- What are the limits to VGI?
 - types of GI that will never be crowdsourced
 - privacy, confidentiality



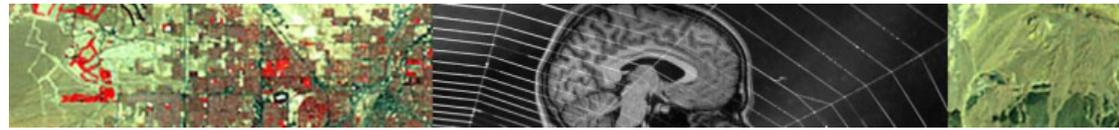
Technical infrastructure

- Generic VGI apps
- Esri VGI-related functions in ArcGIS Server
- Specialized services
 - OSM, Google MapMaker, ...
- Markup standards
 - schema.org
- What is the appropriate spatial organization of VGI production?



schema.org

- “Google, Microsoft and Yahoo! have decided to propose a common markup vocabulary, Schema.org, based on the Microdata format, simplifying the job of webmasters who want to give meaning to their web pages’ content.”
(June 2011)



Thing>Place>SportsActivityLocation

- BowlingAlley
- ExerciseGym
- GolfCourse
- HealthClub
- PublicSwimmingPool
- SkiResort
- SportsClub
- StadiumOrArena
- TennisComplex

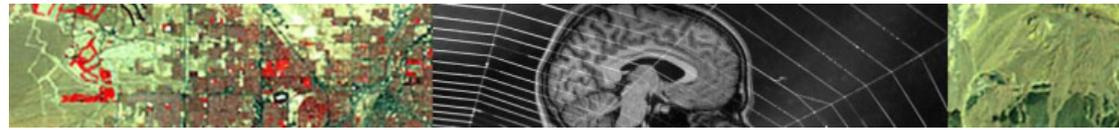
Mark and Turk, Progress on Yindjibarndi ethnophysiology, *COSIT 2007*



Two examples of *Wundu*

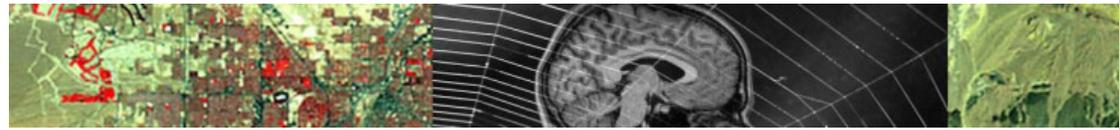


Two examples of *Garga*



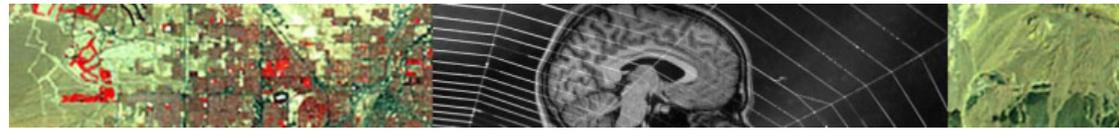
Names as social constructions

- Where is it?
- What is it called?
 - The English Channel
 - La Manche
- *Who calls it that?*
 - Whose national mapping agency made the map?
 - defined spatially?
- What type of feature is it?
 - *In whose system of types?*
 - defined spatially?



The VGI quality problem

- No quality control, no metadata, no standards
 - none of the guarantees of authoritative data
 - no prospect of conventional analysis
- What can we do to assure quality?
- Three solutions



The crowd solution

- Linus's Law
 - the more eyes to review, the more accurate
 - works for popular facts
- 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.

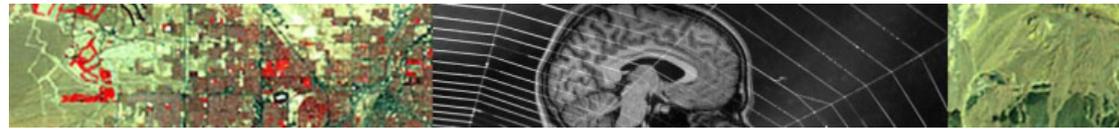
This article is protected.
Category: [golf course](#)

Address: [Glen Annie Road](#), 405

[Permalink to this place](#)

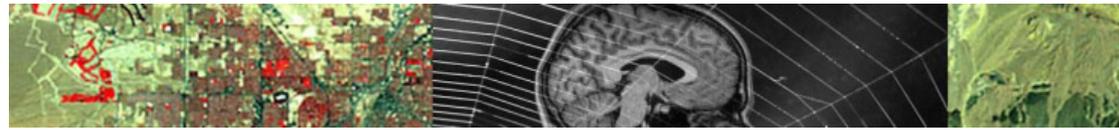


...and I'm a Mormon
I'm Bruce Summerhays. I'm a pro golfer, a husband, & I'm a Mormon.
youtube.com/mormon



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

Ho

You're on page
1 of 1

1 photo taken here

+ Show detail



from [ramoncoler](#)

1 of 1

Map

[Hybrid](#)

[Satellite](#)

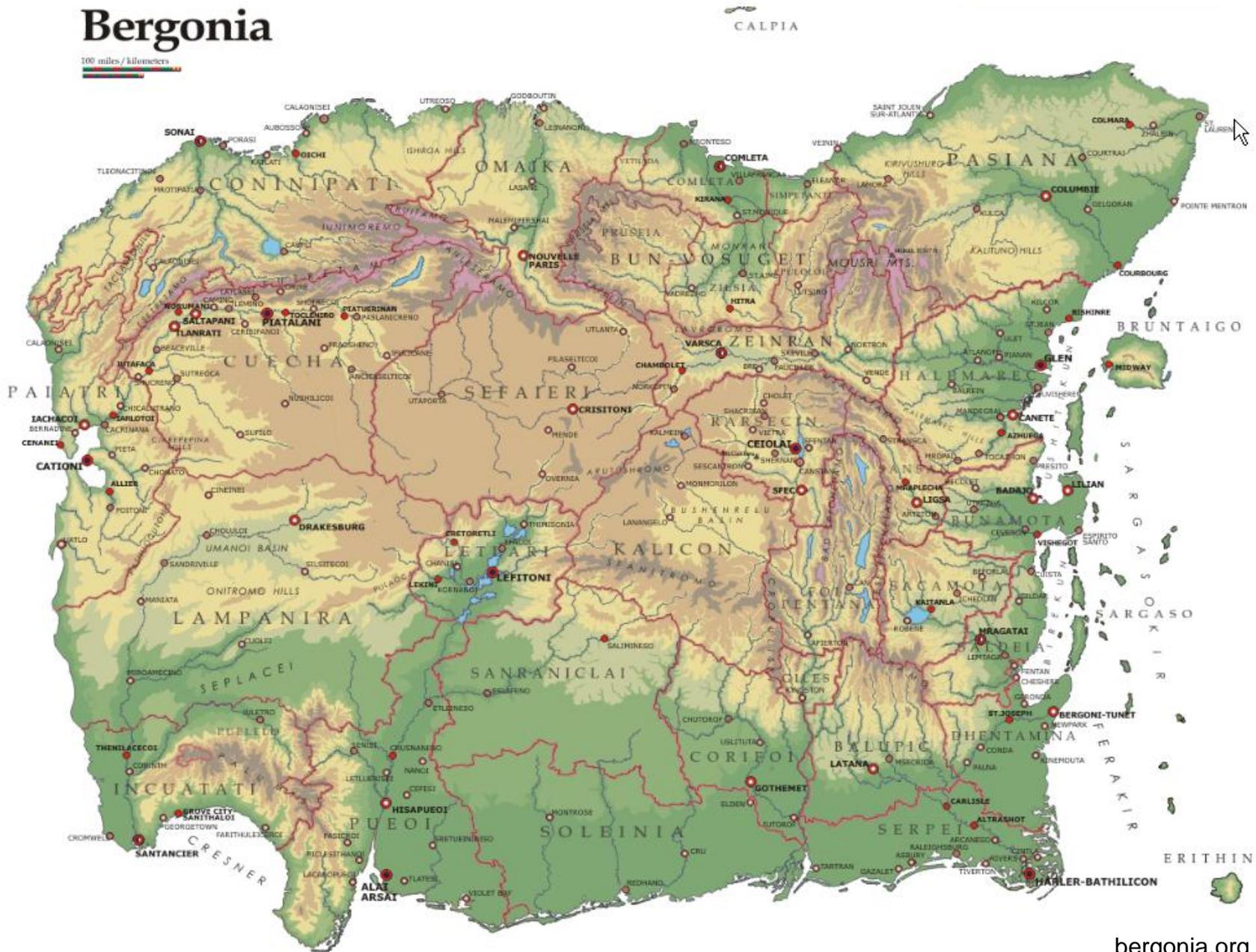


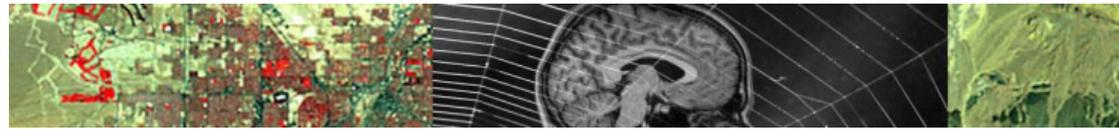
Search

You're looking at **All geotagged photos matching "restaurants".**

Bergonia

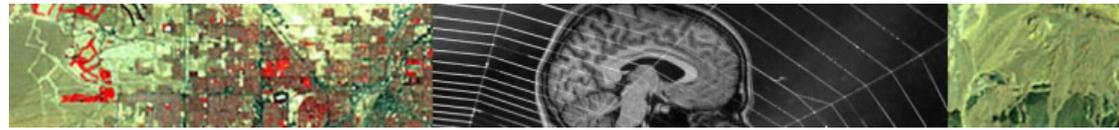
100 miles/kilometers





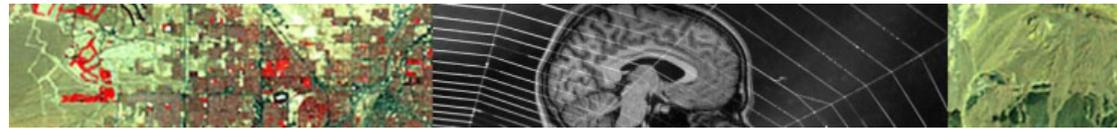
Formalizing geographic knowledge

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



Synthesis as the new analysis?

- How to assemble disparate data
 - into a form analogous to the traditional authoritative map?
 - or the traditional statistical database?
- From text, images, spatial data
- With metadata, provenance
- How to analyze such data?



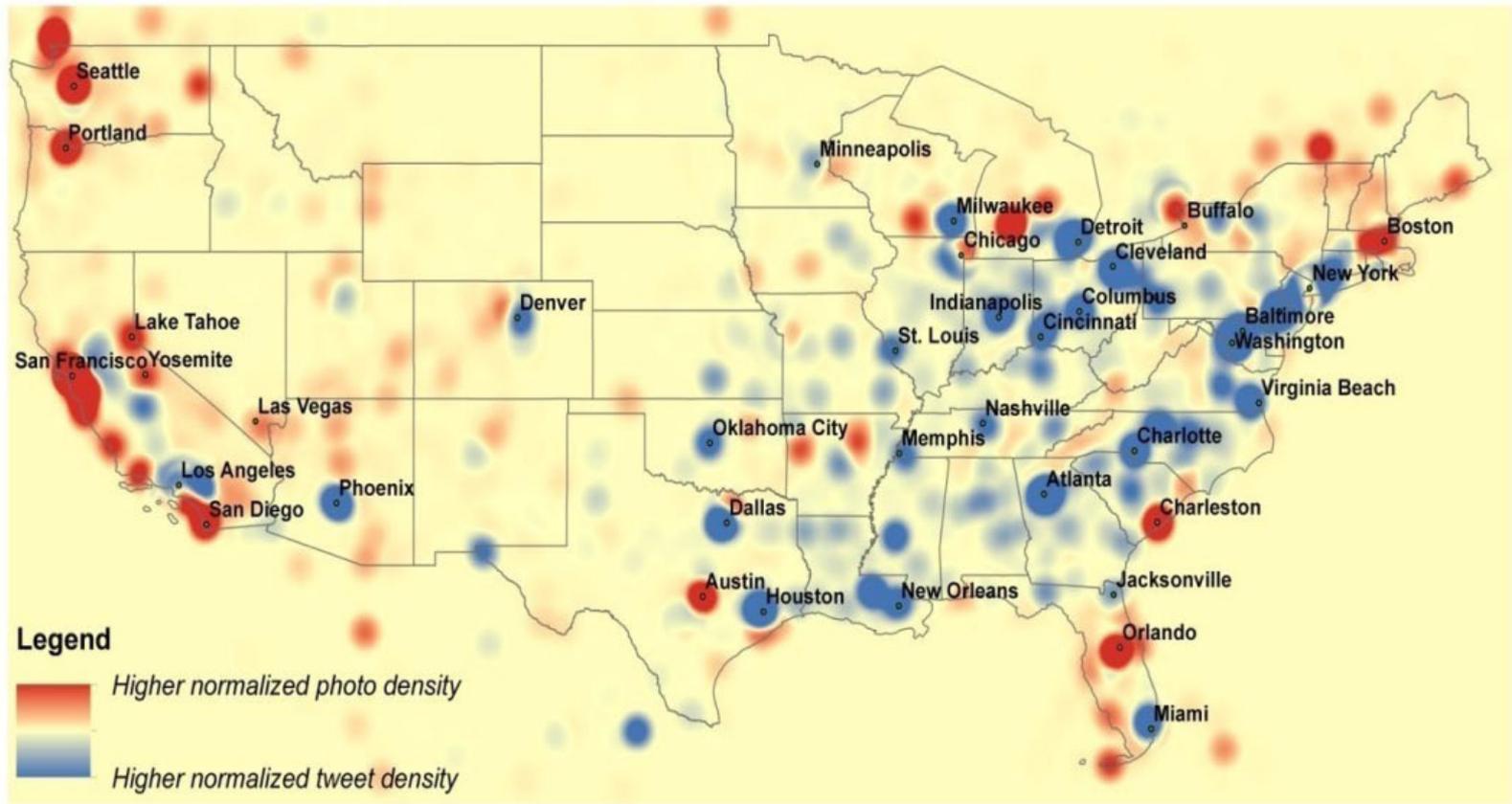
The general problem

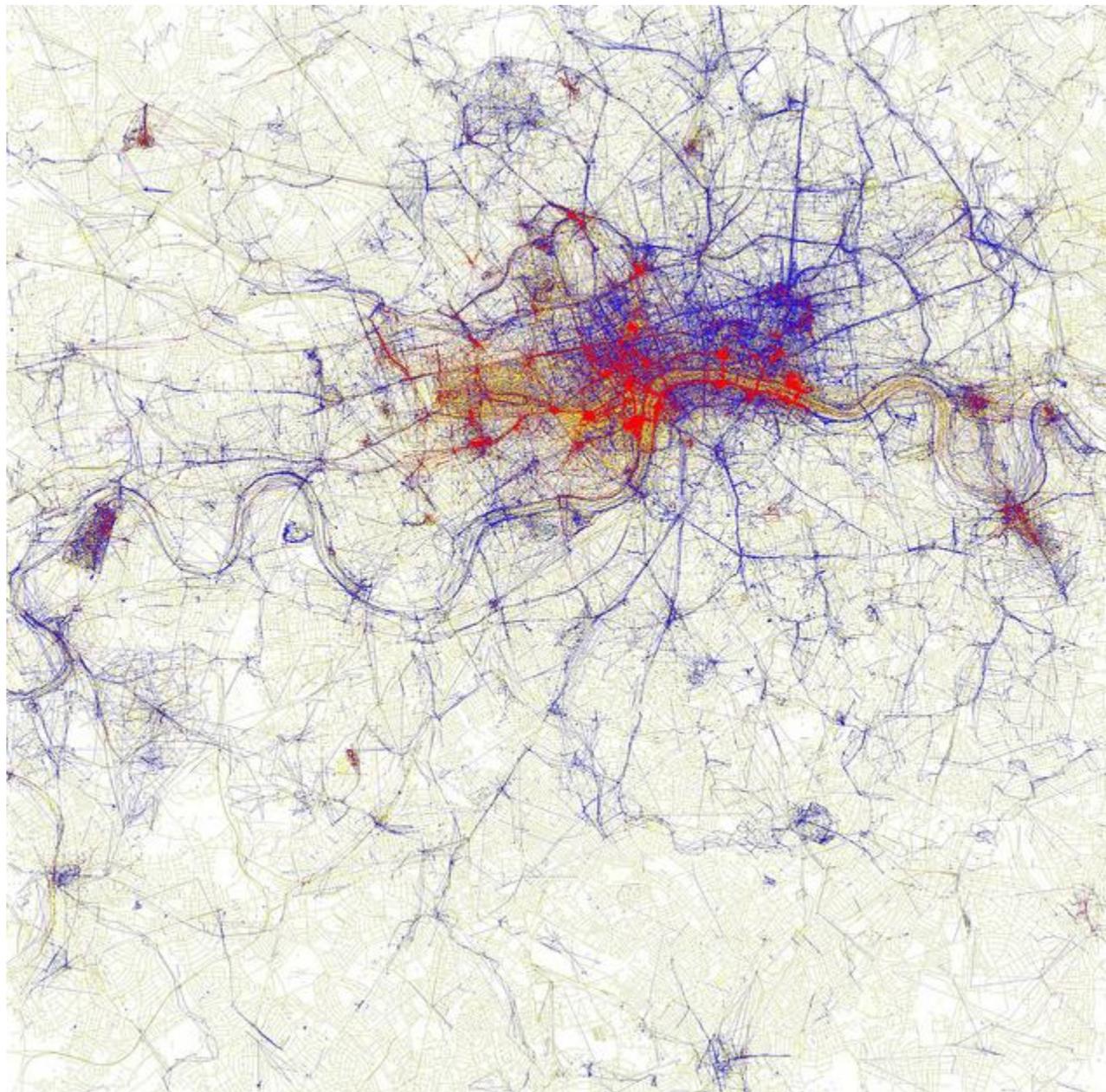
- Consider an atom of unary geographic information $\langle \mathbf{x}, \mathbf{z} \rangle$
- All such atoms are incomplete
 - measurement error in coordinates
 - classification uncertainty
 - “this location is oak savanna”
 - attribute error
 - misnamed street
- Synthesis attempts to resolve incompleteness
 - and to integrate atoms into polygons, polylines



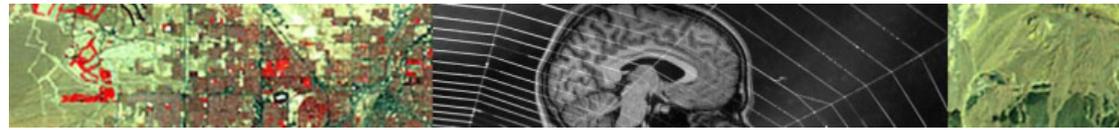
Research using VGI

- What research can be done with VGI?
 - wrong question?
- What kinds of research questions in social and environmental science can VGI be used to answer?
 - better question?
- What limitations are imposed by the nature of VGI?
 - good question!



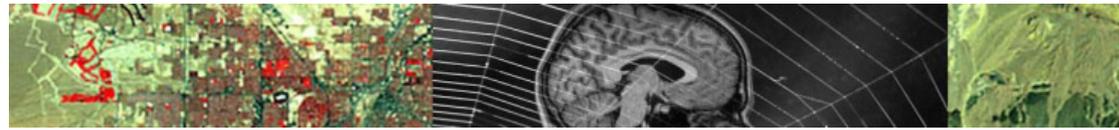


Eric Fischer, <http://www.flickr.com/photos/walkingsf/4671589629/>



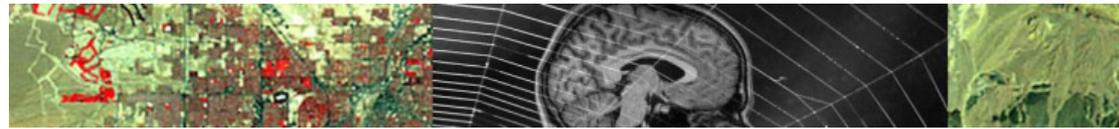
VGI limitations

- Sampling will never be random
 - unlike traditional sources used in science
 - samples will never be representative
 - of any larger and more general population
- Quality will always be variable
 - and often unknown



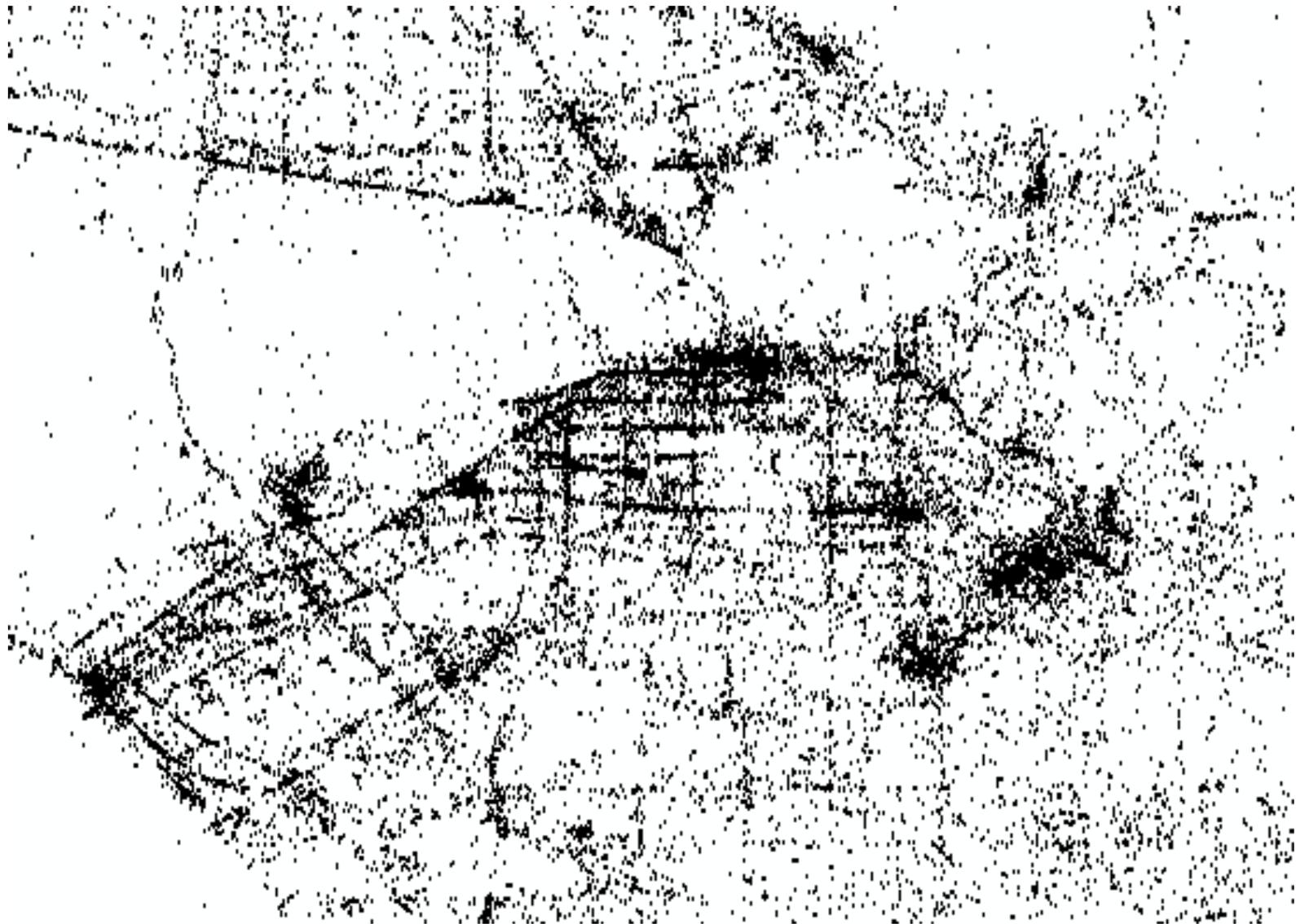
Two phases of science

- The soft phase
 - exploratory
 - study area selection
 - hypothesis-generating
 - compare ethnography
 - no generalizability
- The hard phase
 - confirmatory
 - hypothesis-testing
 - generalizability important



VGI in the hard phase

- A sample of VGI may be representative of volunteers
 - but volunteers are not representative of society
 - but if we can identify the demographics of volunteers
 - then perhaps we can weight appropriately
- The content of VGI may be independent of its demographics
 - citizen science
 - anyone can count birds

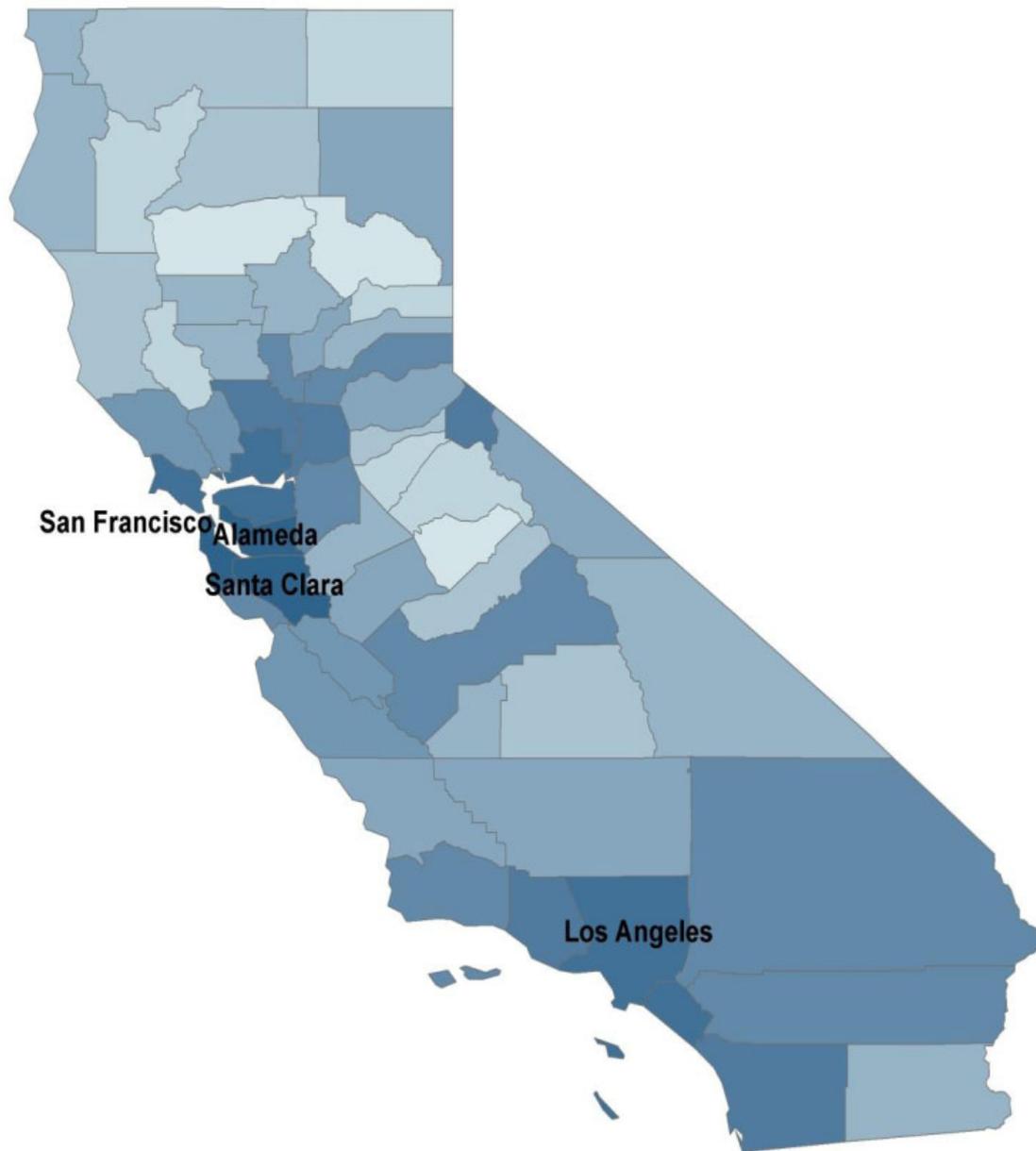


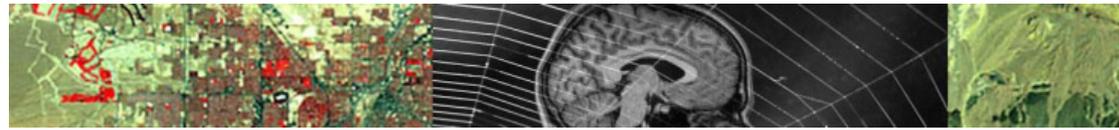
Linna Li: Density of geo-located tweets in Los Angeles, Jan1 to Feb 25, 2011

Table 1. Sample loading matrix for PLS components in the tweet density model

Explanatory variables	Component				
	1	2	3	4	5
White alone	-0.74	0.52	-0.15	0.14	-0.08
Black or African American alone	0.12	-0.11	0.00	-0.08	0.06
American Indian and Alaska Native alone	-0.03	0.05	0.07	-0.05	0.07
Asian alone	0.53	-0.02	0.10	-0.08	-0.02
Native Hawaiian and Other Pacific Islander alone	0.01	0.00	0.00	0.00	0.00
Some other race alone	0.10	-0.41	-0.03	0.12	-0.03
Two or more races:	0.01	-0.02	0.00	-0.05	-0.01

First-component scores





Summary

- Distinct forms of VGI
- Many interesting topics for research about VGI
- VGI as a tool for exploring the world
- VGI in science
 - a cautionary tale