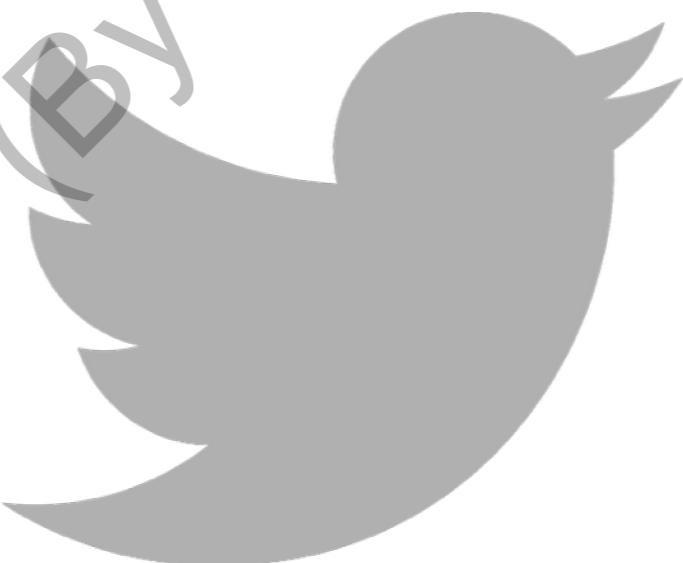




# Volunteered Geographic Information

Spatial Computing – University of Minnesota



# Volunteered Geographic Information

Spatial Computing – University of Minnesota

## Learning Objectives

1. Know the definition and historical context of volunteered geographic information (VGI)
2. Be able to produce VGI yourself!
3. Understand the socio-technical systems behind VGI.
4. Be qualified to assess the pros and cons of VGI for a given problem context.
5. **Technical Track:** Be able to use several types of VGI in your apps, websites, and other systems



**flickr**



Slides for Spatial Computing  
(By Brent Hecht)



**foursquare**





**flickr**



**foursquare**



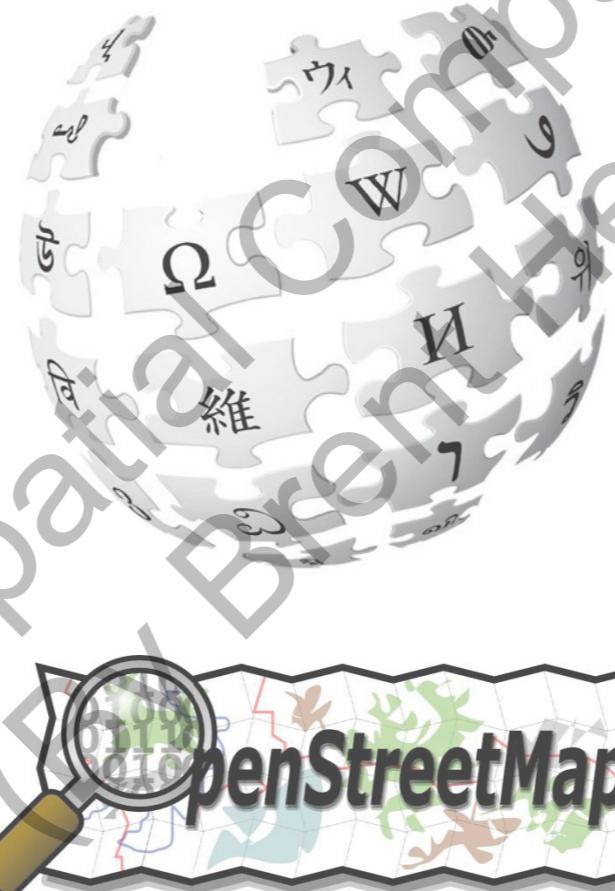
Slides for Spatial Computing  
(By Brent Hecht)



Slides for Spatial Computing  
(By Brent Hecht)



Social Media  
VGI



Peer-Production  
VGI



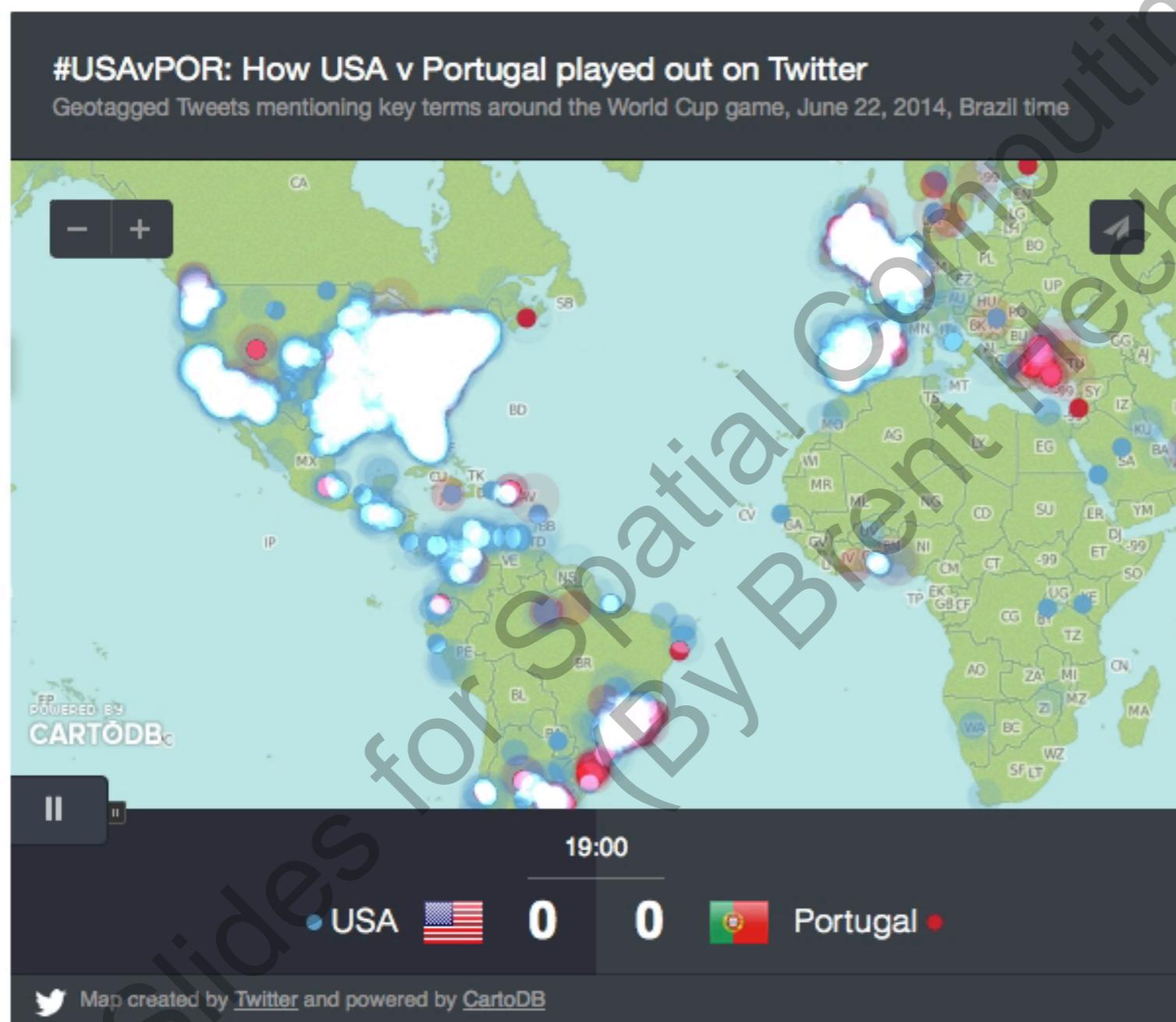
Citizen Science  
VGI





[www.washingtonpost.com/news/morning-mix/wp/2014/06/24/tweets-lit-up-the-map-through-u/](http://www.washingtonpost.com/news/morning-mix/wp/2014/06/24/tweets-lit-up-the-map-through-u/)

charts when goals were scored — especially at the end of the match, when USA's Clint Dempsey scored to go ahead 2-1 and Portugal's Silvestre Varela suddenly tied the game at the end.



Twitter also charted mentions of "Clint Dempsey," showing a spike to

- 1 weapons inside hospitals, mosques a...
- 2 Dedication: The octopus that spent nearly 4 1/2 years sitt...
- 3 Sixteen-foot swells reported in once-frozen region of Arctic Ocean
- 4 Two new mysterious craters emerge in Siberia, deepening gi...
- 5 Cheesecake Factory owns a health advocate's list of me...

Don't miss a single story!

SUBSCRIBE >

## trove trending articles

Drew Barrymore's Half-Sister Found Dead With 'Dozens Of White...

5 things to know about Ebola outbreak in W. Africa

Orlando Bloom Throws Punch at Justin Bieber in Ibiza Restaura...

# Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors

Takeshi Sakaki

The University of Tokyo  
Yayoi 2-11-16, Bunkyo-ku  
Tokyo, Japan

sakaki@biz-model.t.u-tokyo.ac.jp

Makoto Okazaki

The University of Tokyo  
Yayoi 2-11-16, Bunkyo-ku  
Tokyo, Japan

m.okazaki@biz-model.t.u-tokyo.ac.jp

Yutaka Matsuo

The University of Tokyo  
Yayoi 2-11-16, Bunkyo-ku  
Tokyo, Japan

matsuo@biz-model.t.u-tokyo.ac.jp

## ABSTRACT

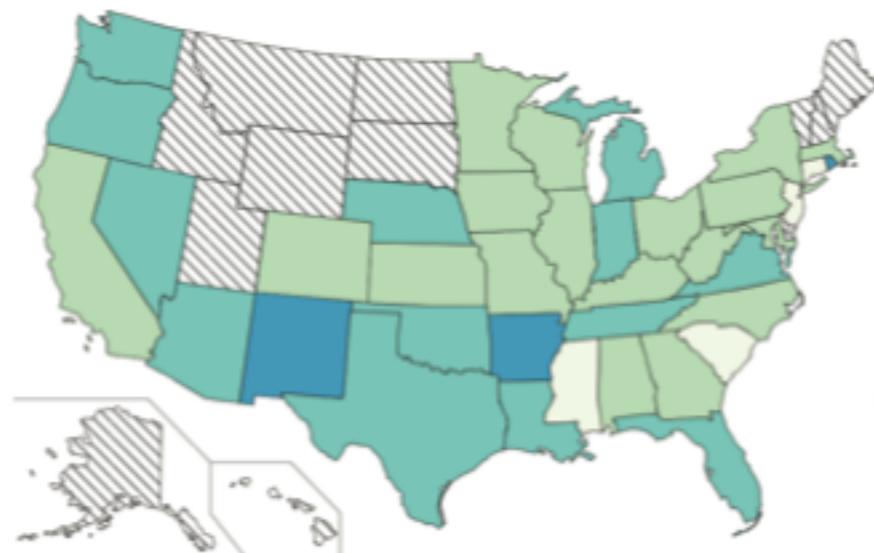
Twitter, a popular microblogging service, has received much attention recently. An important characteristic of Twitter is its real-time nature. For example, when an earthquake occurs, people make many Twitter posts (*tweets*) related to the earthquake, which enables detection of earthquake occurrence promptly, simply by observing the tweets. As described in this paper, we investigate the real-time interaction of events such as earthquakes, in Twitter, and propose an algorithm to monitor tweets and to detect a target event. To detect a target event, we devise a classifier of tweets based on features such as the keywords in a tweet, the number of words, and their context. Subsequently, we produce a probabilistic spatiotemporal model for the target event that can find the center and the trajectory of the event location. We consider each Twitter user as a *sensor*

currently estimated as 44.5 million worldwide<sup>1</sup>. Monthly growth of users has been 1382% year-on-year, which makes Twitter one of the fastest-growing sites in the world<sup>2</sup>.

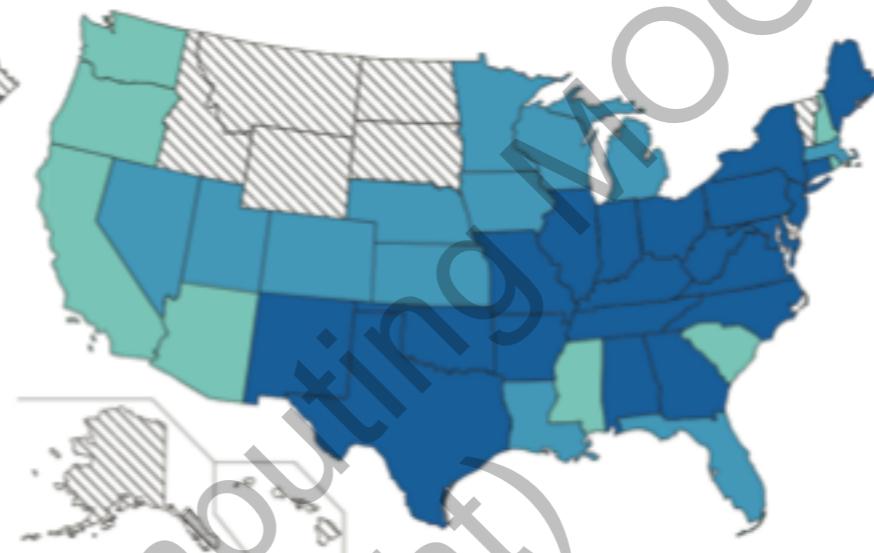
Some studies have investigated Twitter: Java et al. analyzed Twitter as early as 2007. They described the social network of Twitter users and investigated the motivation of Twitter users [13]. B. Huberman et al. analyzed more than 300 thousand users. They discovered that the relation between friends (defined as a person to whom a user has directed posts using an "@" symbol) is the key to understanding interaction in Twitter [11]. Recently, boyd et al. investigated *retweet* activity, which is the Twitter-equivalent of e-mail forwarding, where users post messages originally posted by others [5].

Twitter is categorized as a micro-blogging service. Microblogging is a form of blogging that allows users to send

watery  
helsinki  
**mold**  
watering  
faucet  
lolss  
sneezes  
sneezy  
teary  
**bloom**

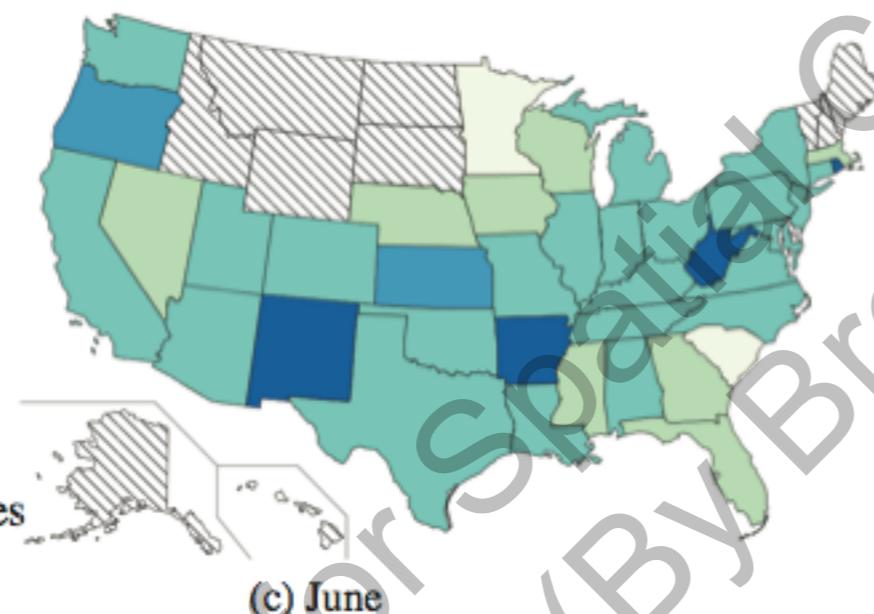


(a) February

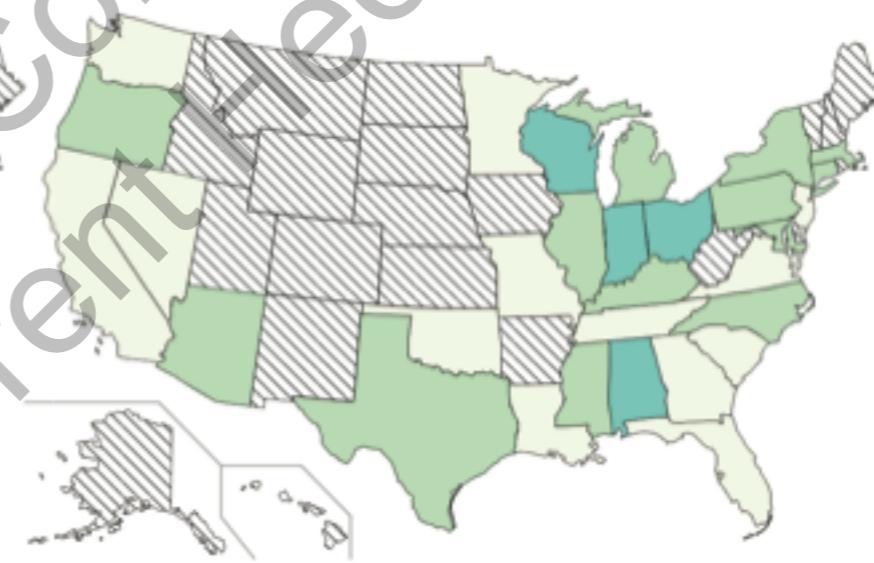


(b) April

**hayfever**  
grass  
watering  
watery  
claritin  
humidity  
zyrtec  
bonkers  
**mold**  
antihistamines



(c) June



(d) August

**pollen**  
zyrtec  
claritin  
spring  
watering  
**trees**  
watery  
itching  
**bloom**  
grass

**dust**  
catherine  
pinche  
buildings  
**mold**  
hadd  
gato  
cessation  
meditating  
**ragweed**

# You Are What You Tweet: Analyzing Twitter for Public Health

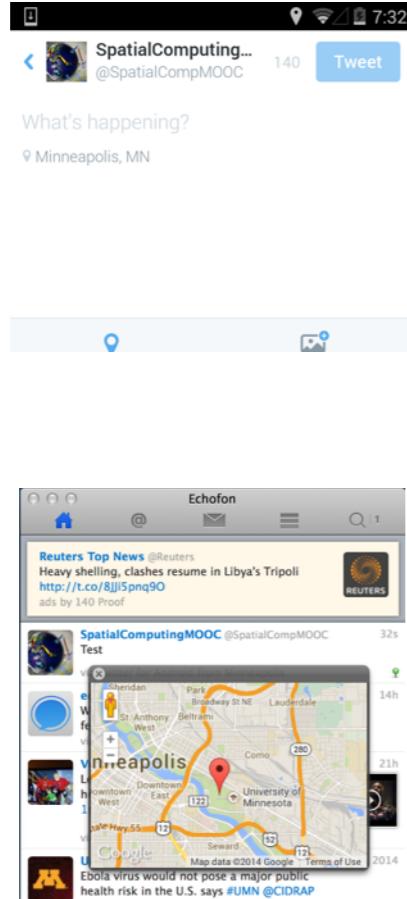
(Paul and Dredze 2011)

All these cool applications start with the lowly smartphone Twitter application...

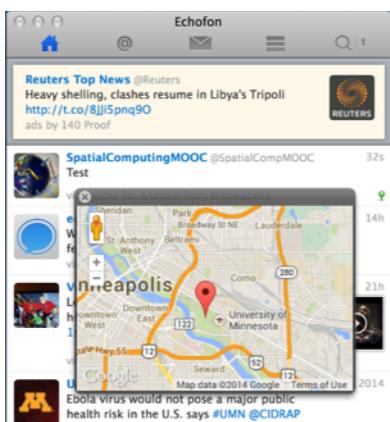




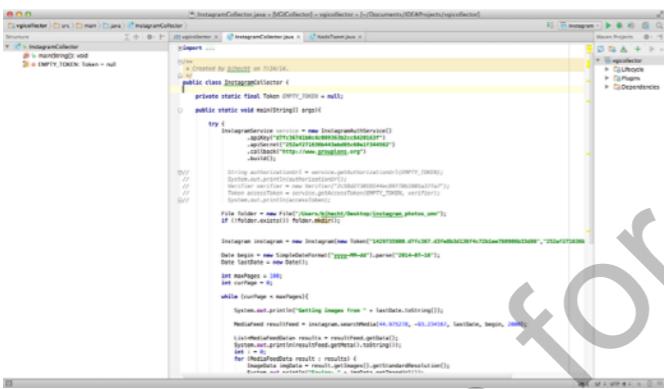




You create content, geotag it, and upload it!



Your friends and contacts (or the general public) can see your geotagged content



In many cases, programmers can download your geotagged content (assuming it's public)

# Social Media VGI



tripadvisor®



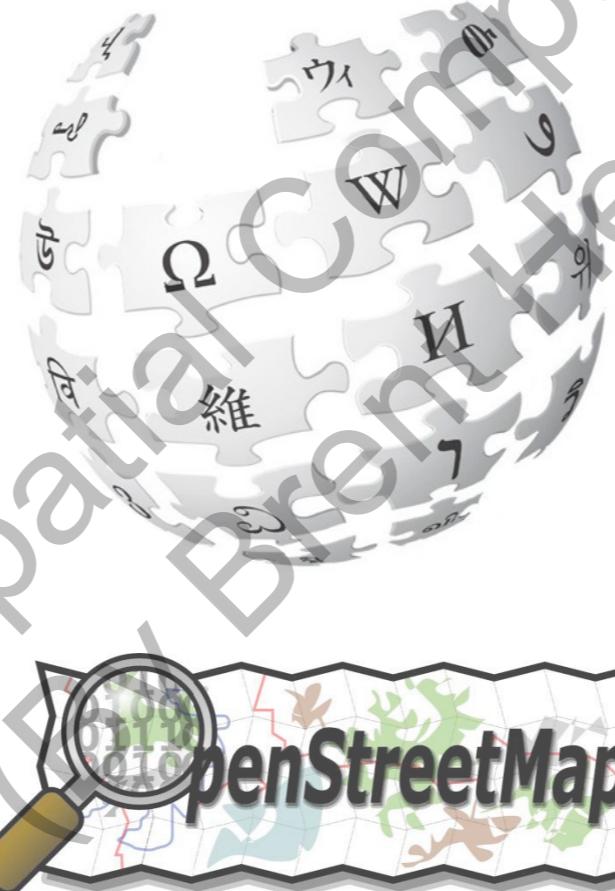
foursquare

facebook

flickr

Social Media  
VGI

Cyclopath  
discover your path



Peer-Production  
VGI

ZOONIVERSE  
REAL SCIENCE ONLINE



Audubon  
CHRISTMAS BIRD COUNT

Citizen Science  
VGI



Social Media  
VGI



Peer-Production  
VGI



Citizen Science  
VGI

---

“Citizen science is scientific research conducted, in whole or in part, by amateur or nonprofessional scientists, often by crowdsourcing and crowdfunding”

---



**English Wikipedia**  
Peer-Produced Encyclopedia  
The World

Many citizen science projects rely  
heavily on measurements that are  
geographically-referenced!

Slides for Spatial Computing MOOC  
(By Brennenstuhl)

Seafloor Explorer

www.seafloorexplorer.org/?utm\_source=Zooniverse Home&utm\_medium=Web&utm\_campaign=MOOC

Home About Science Classify Profile FAQ Talk Blog

Thanks!

Latitude 41.355192°  
Longitude -68.757508°  
Depth 15.24 M  
Altitude 1.72 M  
Salinity 32.454 PSU  
Temperature 5.555° C  
Speed 4.70 kts

Add to my favorites

Would you like to discuss this image in Talk?

Yes No

Share this image on...

Facebook Twitter

SEAFLOOR EXPLORER

Latitude 41.355192°  
Longitude -68.757508°  
Depth 15.24 M  
Altitude 1.72 M  
Salinity 32.454 PSU  
Temperature 5.555° C  
Speed 4.70 kts

Add to my favorites

Would you like to discuss this image in Talk?

Yes No

Share this image on...

Facebook Twitter

Seafloor Explorer

www.seafloorexplorer.org/?utm\_source=Zooniverse Home&utm\_medium=Web&utm\_

# The Science of Seafloor Explorer

## HabCam (Habitat Mapping Camera System)

HabCam is a cabled optical and acoustic imaging system that is "flown" from a ship traveling at 5 kts at an altitude of 1 to 3 meters off the bottom at depths to 250m while collecting high resolution still images at a rate of six images per second.

Imaging at this rate provides ~50% overlap to allow for construction of image mosaics of the seafloor. A track approximately 100 nautical miles in length and 259,200 m<sup>2</sup> in area is imaged each 24 hour day while at sea. Over 30 million images (>30 TB) have been collected in less than one year of sea time, which demonstrates how quickly this quantity of data can accumulate.

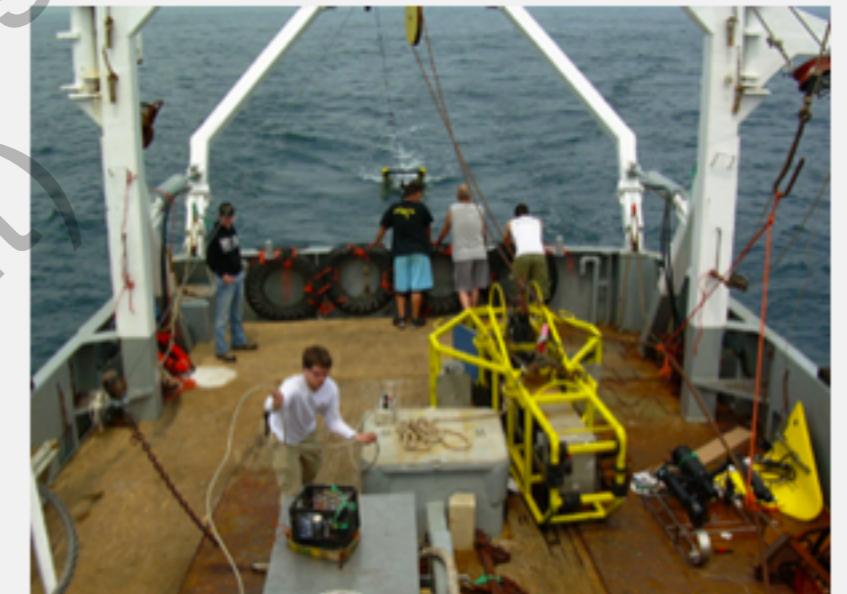
## Fauna, flora, benthic maps and more

For the past five years, study areas along the northeast continental shelf have been revisited seasonally, with measurements of all visible macrofauna, and characterization of benthic fauna and flora, oceanic properties (salinity, temperature, nutrients) and substrate type, providing the baseline of an exceptional and unique ecological time series.

Key to extracting useful ecological information from this ever expanding library of image data is development of tools for rapid and accurate segmentation and classification of benthic organisms and substrate, together with visualization of images and their metadata through a Geospatially-explicit, queryable database. As scientists and ecosystem managers, we need to be able to ask questions like "*what is the current distribution and abundance of sea scallop and yellowtail flounder on Georges Bank, and how have they changed over the past few years?*" And, "*where is the invasive tunicate Didemnum vexillum currently co-located with gravel substratum and what is the potential for its' spread to new areas*".

## Enter Seafloor Explorer

The data now exist to answer such ecologically critical questions, but it is buried in hundreds of TBs of images that need to be processed through a defined and scalable pipeline of tools.

[Seafloor Explorer](#) is allowing us to capture data on the distributions of sea scallop and other commercially important species as well as defining the substrate and habitat in which they live. This information has never before been acquired on such an expansive (1000s of km) yet high resolution (1mm pixel resolution) scale. While manual classification currently pushes our knowledge and understanding of these distributions, the development of tools for automated segmentation and classification is lagging orders of magnitude behind the rate at which image data are currently being acquired and manually classified.

Using data from [Seafloor Explorer](#) we can now begin to build training sets of images and data that will provide the foundation for automated machine vision approaches to target classification from HabCam images. These tools must be developed if the untapped wealth of information available in optical imagery is to be fully realized in

The screenshot shows a web browser window displaying the 'TIGER Products' page from the United States Census Bureau. The URL in the address bar is <https://www.census.gov/geo/maps-data/data/tiger.html>. The page has a header with the Census Bureau logo and navigation links for Topics, Geography, Library, Data, About the Bureau, and Newsroom. A search bar is also present. The main content area features a section titled 'Geography' with a sub-section 'Maps & Data'. Under 'Maps & Data', there is a 'TIGER Products' section. This section includes a definition of TIGER (Topologically Integrated Geographic Encoding and Referencing), a list of product types, and a table comparing TIGER/Line Shapefiles and TIGER/Line Geodatabases.

## TIGER Products

**TIGER** = Topologically Integrated Geographic Encoding and Referencing

TIGER products are spatial extracts from the Census Bureau's MAF/TIGER database, containing features such as roads, railroads, rivers, as well as legal and statistical geographic areas. The Census Bureau offers several file types and an online mapping application. Our products are:

- [TIGER/Line Shapefiles - New 2013 Shapefiles](#)
- [TIGER/Line Geodatabases](#)
- [TIGER/Line with Selected Demographic and Economic Data](#)
- [Cartographic Boundary Shapefiles](#)
- [KML - Cartographic Boundary Files](#)
- [TIGERweb](#)

Product	Best For...	File Format	Type of Data	Level of Detail	Descriptive Attributes	Vintages Available
<a href="#">TIGER/Line Shapefiles</a>	Most mapping projects--this is our <b>most comprehensive dataset</b> . Designed for use with GIS (geographic information systems).	Shapefiles (.shp) and database files (.dbf)	Boundaries, roads, address information, water features, and more	Full detail (not generalized)	Extensive	2006 - 2013, CD 113
<a href="#">TIGER Geodatabases</a>	Useful for users needing national datasets or all major boundaries for by state. Designed for use in ArcGIS.	Geodatabase (.gdb)	Boundaries, roads, address information, water features, and	Full detail (not generalized)	Limited	2013

# Traditional Geographic Information