# Data and Citizen Science

**Fred Roberts**
**Rutgers University**

# Putting this Workshop in Context: Mathematics of Planet Earth 2013

- A joint effort initiated by North American Math Institutes: **MPE2013**
- More than 100 partner institutes, societies, and organizations in UK, France, South Africa, Japan, and all over the world
- www.mpe2013.org

# Mathematics of Planet Earth 2013

- Activities world-wide throughout 2013
- Sponsorship by UNESCO
- Support from Simons Foundation
- Workshops, tutorials, competitions, distinguished lectures, educational programs

Math 2013

# Mathematics of Planet Earth Beyond 2013

- Problems of the planet do not go away in one year.
- We are organizing a series of events to continue beyond 2013.
- New initiative world-wide now called **MPE**
- In the US, we call it **MPE2013+**
- US National Science Foundation support
- We are delighted to have activities supported in other countries such as France, and especially partnership with LAMSADE.

# Mathematics of Planet Earth Beyond 2013

Goals of MPE2013+

- Involve mathematical scientists in addressing the problems of the planet
- Enhance collaborations between mathematical scientists and other scientists
- Involve students and junior researchers in the effort
- Encourage life-long commitment to working between disciplines to solve the problems of society

# Mathematics of Planet Earth 2013+

- Opening Introduction to Problems of the Planet and involve students and junior faculty: Arizona State University, Jan. 7-10, 2014
- Five ***Research Clusters***, beginning with workshops:
  - ***Sustainable Human Environments*** (Rutgers U.), April 23-25, 2014
  - ***Global Change*** (UC Berkeley), May 19-21, 2014
  - ***Data-aware Energy Use*** (UC San Diego), Sept. 29 – Oct. 1, 2014
  - ***Natural Disasters*** (GA Tech), May 13-15, 2015
  - ***Management of Natural Resources*** (Howard University), June 4-6, 2015

# Mathematics of Planet Earth 2013+

Follow-up cluster activities:

- Sustainable Human Environments cluster:
  - Pre-workshop: Urban Planning for Climate Events Sept. 2012; ***Post-workshop: This workshop***
  - Cluster activities of various kinds
- Natural Disasters cluster: working with several potential partners in Mexico and Colombia.
- Global Change cluster: considering a follow-up event at the National Center for Atmospheric Research (NCAR) and one at Old Dominion U. on communication of global change challenges

# Mathematics of Planet Earth 2013+

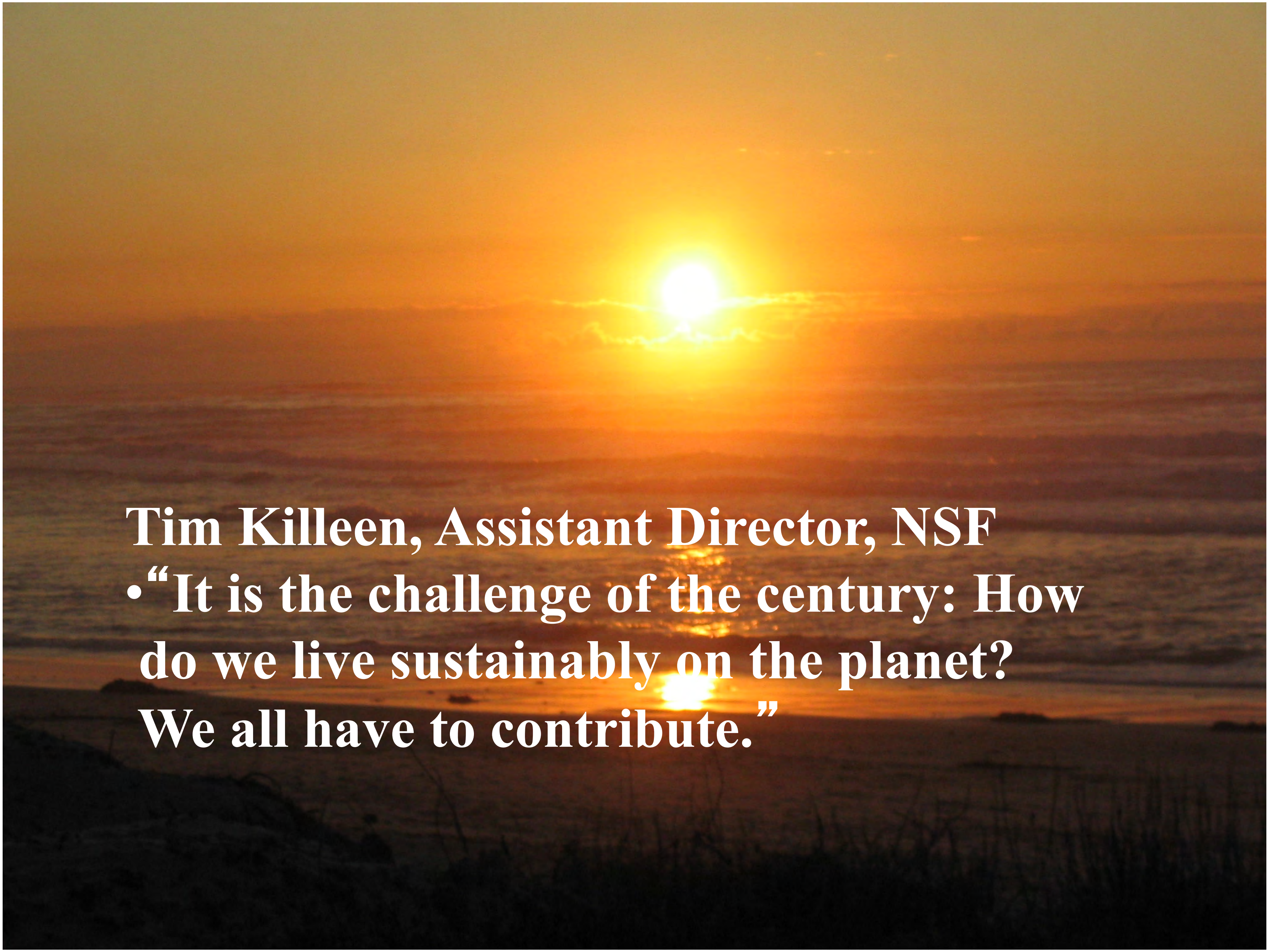
Follow-up cluster activities:

- Management of Natural Resources cluster:
  - Expecting a follow up in Africa (Ebola and lessons learned)
- All clusters:
  - Looking into possibility of research groups ("squares") at American Inst. of Mathematics (AIM)
- Many more follow-up workshops in the process of being scheduled.

# Mathematics of Planet Earth 2013+

- Education is a crucial piece of this and of the sustainability effort
  - Workforce development
  - Public literacy

- Need education at all levels, starting with K-12.
- Education issues in each workshop
- Special Education cluster: ***Education for the Planet Earth of Tomorrow***
- Cluster workshop: U. of Tennessee, Sept. 30 – Oct. 2, 2015.


Tim Killeen, Assistant Director, NSF
•"It is the challenge of the century: How do we live sustainably on the planet? We all have to contribute."

# Usefulness of Citizen Science

- A message reinforced at the workshop on Sustainable Human Environments: Engaging “ordinary” citizens can help with the development of science and the development of public policy.
- However, the challenge is to understand the quality of the data citizens provide and the implications of data quality for scientific advances and/or leading to public policy.

# Citizen Science & Natural Disasters

- Other data challenges:
  - How best to merge citizen science data with data from other data sets
  - How to collect and distribute citizen science data to make it most useful (and error-free)
  - How to use data produced from efforts distributed over space and time
  - How to keep things simple enough to minimize citizens' training needs while keeping data useful?
  - How can data provide "evidence" for decisions?
- This talk will address these data quality questions in the context of one class of applications: natural disasters.

# Natural Disasters

- ***No part of the world is impervious to natural disasters***
  - Epidemics
  - Earthquakes
  - Floods
  - Hurricanes
  - Tornadoes
  - Wildfires
  - Tsunamis
  - Extreme temperatures
  - Drought
  - Oil spills


Nepal 2015: www.circleofblue.org

- ***Citizen science can help in predicting, monitoring, and responding to such events, and mitigating their effects.***

# Climate Events

- ***Example: Climate events***: Super Storms, heat, drought, floods – all could be increasing in number and severity.
- What can urban areas do to prepare for them?
- A topic of a predecessor workshop of this one.
  - Urban Planning for Climate Events, DIMACS, Rutgers University, Sept. 2013


# Climate Events

- Relevant to what I saw in Paris upon my arrival


ambafrance-nz.org


foei.org

# Urban Planning for Climate Events

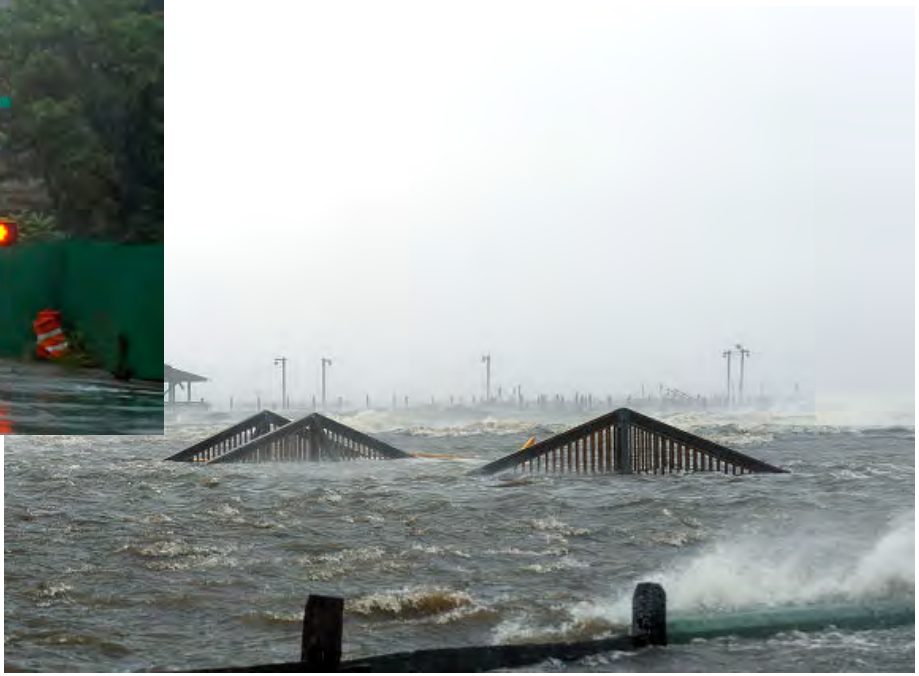
- Sustainable Human Environments: Urban Planning for Climate Change, Sept. 2013, at DIMACS-Rutgers University
- *What can urban areas do to prepare for/mitigate changes due to climate and in particular the effect of future climate events?*

# Extreme Events due to Global Warming

- We anticipate an increase in number and severity of extreme events due to global warming.
- More heat waves.
- More floods, hurricanes.

# Extreme Events due to Global Warming: More Hurricanes

Irene hits NYC – August 2011

# Extreme Events due to Global Warming: More Hurricanes

Irene hits NYC – August 2011

# Extreme Events due to Global Warming: More Hurricanes

Irene hits NYC – August 2011

# Extreme Events due to Global Warming: More Hurricanes

Sandy Hits NJ Oct. 29, 2013

My backyard

My block

# Extreme Events due to Global Warming: More Hurricanes

Sandy Hits NJ Oct. 29, 2013

My neighborhood

My block

# Extreme Events due to Global Warming: More Hurricanes

Sandy Hits NJ Oct. 29, 2013

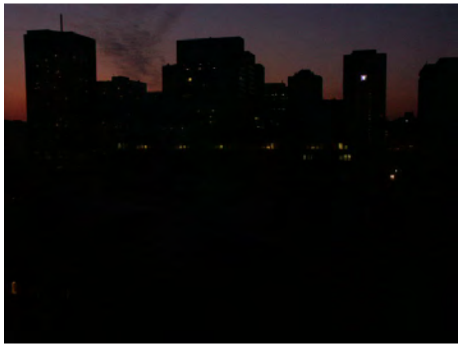
NJ Shore – from Jon Miller

# Extreme Events due to Global Warming

## *Future Storms*

- To plan for the future or intervene during an event, what do we need to do?
- Can citizen science help?
- We provide three examples from researchers at Rutgers University.
- We then speculate about other similar applications.

# Example I: Documenting Hazards that Could Lead to Loss of Power During Storms*



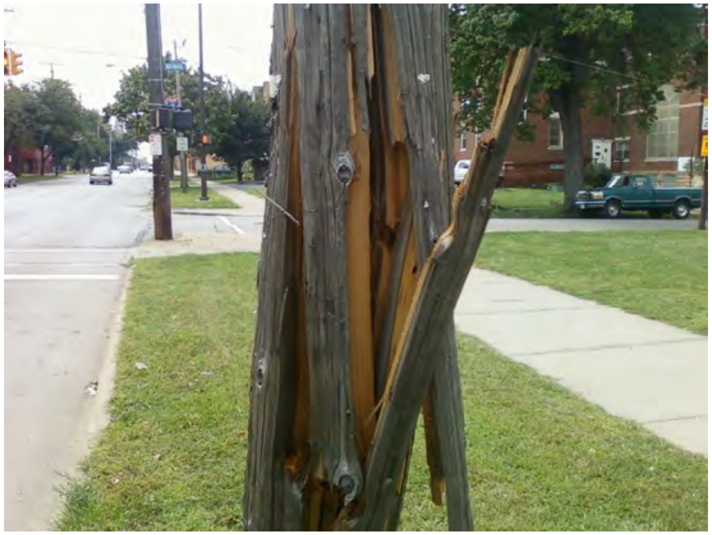
- Extreme weather events can lead to loss of power
- What can we do in advance to identify hazardous situations that could lead to loss of power during a major storm?
- Topic of a citizen science study by Yulong Yang, Michael Sherman, and Janne Lindqvist at Rutgers University.
- Sponsored by US National Science Foundation.

*Thanks to Janne Lindqvist for this example

# Example I: Documenting Hazards

- Sample hazards:
  - Tree branches in threatening position
  - Trees/branches on wires
  - Wire off pole/hanger
  - Wires twisted
  - Cracked/broken pole
  - Leaning/stressed pole

bartstreeservice.com



cleveland.com

# Example I: Documenting Hazards

- Early identification of such hazards can avoid serious problems during a storm
- This project involved ordinary citizens (mostly senior citizens) in a small community in New Jersey, USA, in this effort.
- The job of cataloguing hazards involves time-consuming manual labor.
- Large cities have well-established processes for this kind of thing – involving professional maintenance staff and even police officers.
- Large cities can arrange training time for people involved and coordination of efforts.
- But small cities cannot afford either of these things

# Example I: Documenting Hazards

- This project designed to:
  - Minimize training and coordination costs and salaries of professionals to do the work
  - Develop processes that are repeatable and scalable
- The project utilized technology: smartphones, that could make the process of documenting and reporting hazards relatively easy.
- Volunteers used smartphones to document and report hazards to a central server.
  - Smartphones provided, with app already installed
- Coordinators easily visualized and managed the data collected.

# Example I: Documenting Hazards

Shot of the smartphone app used in the project. Image from article by Yang, Sherman, and Lindqvist in Proccedings of 2014 IEEE Global Humanitarian Technology Conference

# Example I: Documenting Hazards

- Project recruited 8 volunteers.
- They walked around town collecting hazards.
  - 111 streets
  - 144 miles
- 349 hazards identified.
- 95% of them (333) fixed within 6 months.
- Very positive feedback from the town.
- Some specific designated days – on those, found lots of hazards
- On other days, doing “normal walking” but still found good number of hazards

# Example I: Documenting Hazards

- How the Smartphone App worked:
  - Designed to be easy to understand and use without much setup or training
    - ✓ Only training was to identify hazards
  - Designed to report:
    - ✓ Photo of hazard
    - ✓ When: Date/time
    - ✓ Who: Volunteer's ID
    - ✓ Where: Hazard location
- Time and location recorded automatically
  - Minimizes work for volunteer
  - Increases accuracy of data

# Example I: Documenting Hazards

- Because most hazards occur near a utility pole, information about nearest pole is important.
  - Provided by photo of the ID tag on the pole
- Comment section allowed extra information to be provided by volunteer.

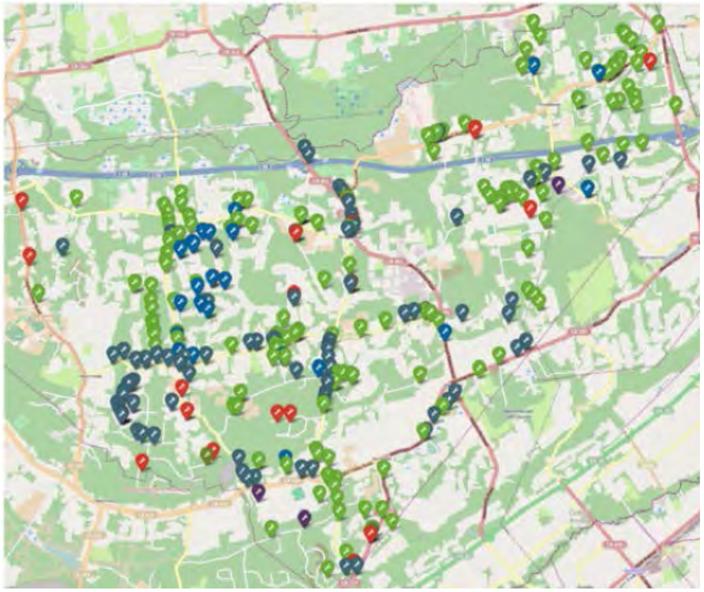

From Yang, Sherman, Lindqvist

# Example I: Documenting Hazards

- Other things to simplify the task:
  - One entry at a time
  - Clear instructions
  - Easy to go back to correct previous step(s)
  - Volunteer ID added in automatically if device used before
  - Allow preview, confirm, or retake the photo

# Example I: Documenting Hazards

- On the server:
  - Locations of reported hazards are all reported on a map with pushpins.
  - Pushpins in different colors depending on type of hazard found
  - Click on the pushpin to bring up key information: photo, street address.
  - Only those with appropriate identification (passwords) could access the site. Key: township committee members.



# Example I: Documenting Hazards

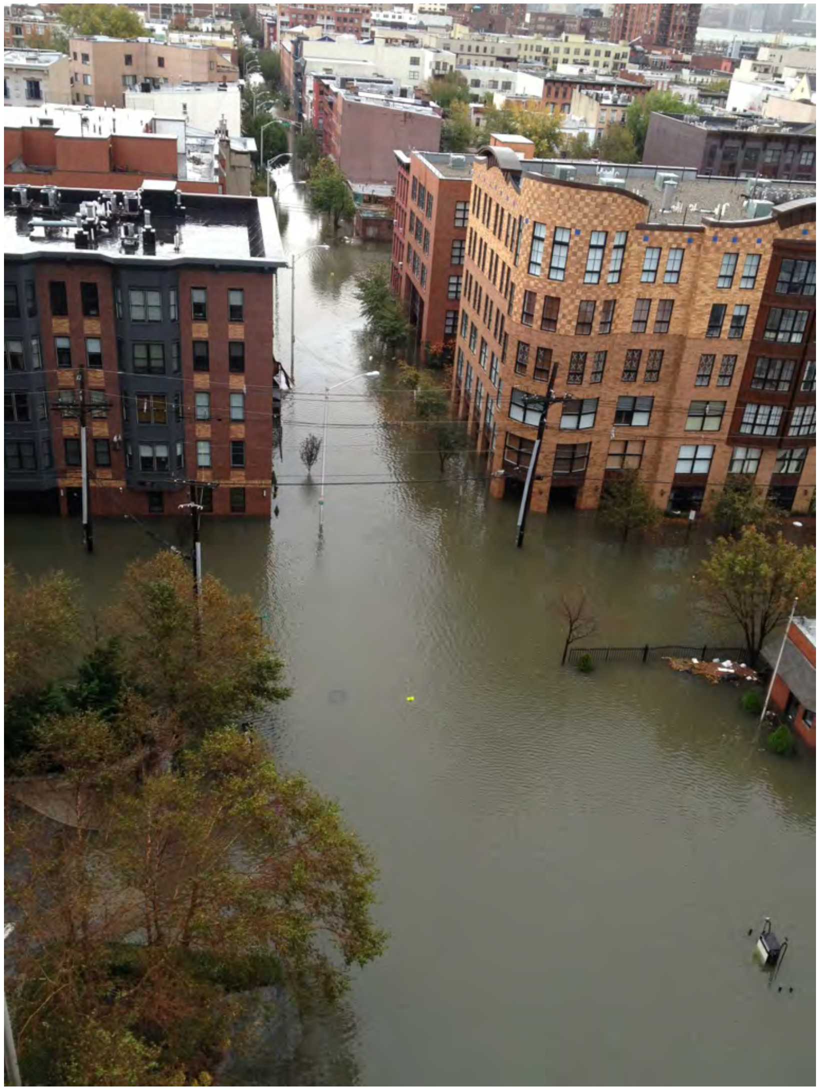
- Why did this work?
  - Considerable simplification in the procedure led to fewer errors:
    - ✓ Easy to learn
    - ✓ Easy to use
    - ✓ Many things automated
    - ✓ Easy corrections
    - ✓ Considerable attention to needs of citizen scientists
  - Thus, data accuracy not a problem.
  - Easy to apply even if going through normal daily life
  - Relatively small geographic area avoids distributed data
  - Inexpensive: only need a few smartphones & simple server
  - Data easy to put on the server
  - Data remains accessible on the server

# Example I: Documenting Hazards

- Minor issues:
  - Some volunteers, especially seniors, not used to small interactive area on touchscreen. Tablets might work better
  - Uploading failed in areas with poor network coverage – and not necessarily uploaded later.
    - ✓ Leads to omitted data
    - ✓ Easy to fix by caching data and uploading periodically when in the presence of a good network.

# Example II: Reporting Floods*

- Hoboken, New Jersey suffers from occasional severe flooding.
- A study at Rutgers looked at how citizens' reports of severity of flooding matched up with reports from trained experts.



*Thanks to Ryan Whytlaw for this example

betterwaterfront.org

# Example II: Reporting Floods

- Hoboken has a very old sewer system, which combines sewage from homes with rainwater runoff.
- When there is heavy rain, the system backs up into people's basements or into the streets.
- The study concerned a new "Stormwater Management Plan" for Hoboken.
- It was concerned with a "Health Impact Analysis": What is the effect on people's health of a large public works project?

# Example II: Reporting Floods

- The study sought to relate flooding level to levels of gastro-intestinal infections (GI) (and other health effects)
- When sewage backs up, it leads to more GI.
- How to understand frequency & severity of flooding?
- Flood data available from National Weather Service, which is part of National Oceanic & Atmospheric Administration (NOAA)
- National Weather Service data not adequate for description of flooding severity.
  - Data incomplete
  - Data not always local enough for a study of a particular local area

# Example II: Reporting Floods

- However, NJ State Climatologist Office does have a great deal of data. (The State Climatologist is a professor at Rutgers.)
- That data comes from citizen science.

# Example II: Reporting Floods

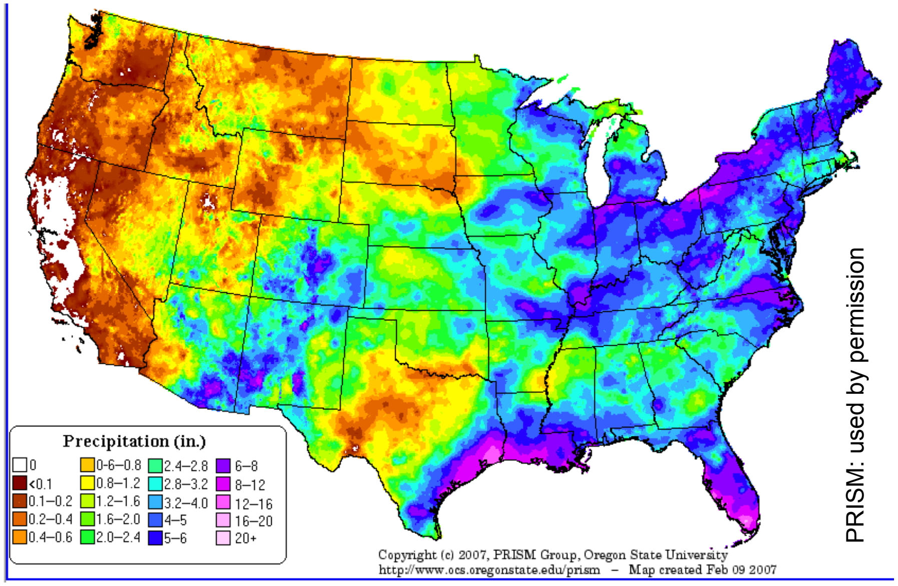
- NJ State Climatologist uses data from ***CoCoRaHS - the Community Collaborative Rain, Hail and Snow Network***.
- CoCoRaHS is a national volunteer network that uses citizens to report rainfall, snowfall, hail, flooding, and other weather data.
- 8,000 volunteers nationally
- Individuals take readings in their backyards, in schoolyards, etc.
- Data report via CoCoRaHS website.
- Web-based or in-person training.
- Use inexpensive rain gauges.
- Some report daily, some occasionally.
- The following slides from CoCoRaHS website.


CoCoRaHS
"Because every drop counts"
CoCoRaHS
Community Collaborative Rain, Hail & Snow Network
NOAA
National Weather Service

# WHY COCORAHS ??

**Precipitation is important and highly variable**

**Data sources are few and rain gauges are far apart**




PRISM: used by permission


Measurements from many sources are not always accurate (especially snow)


There is almost no quantitative data being collected about hail

Storm reports can save lives

*"CoCoRaHS is a national grassroots, non-profit, community-based, high-density precipitation network*


→ made up of volunteers of all backgrounds and ages . . .


. . . who take daily measurements of "<u>just precipitation</u>" right in their own backyards"



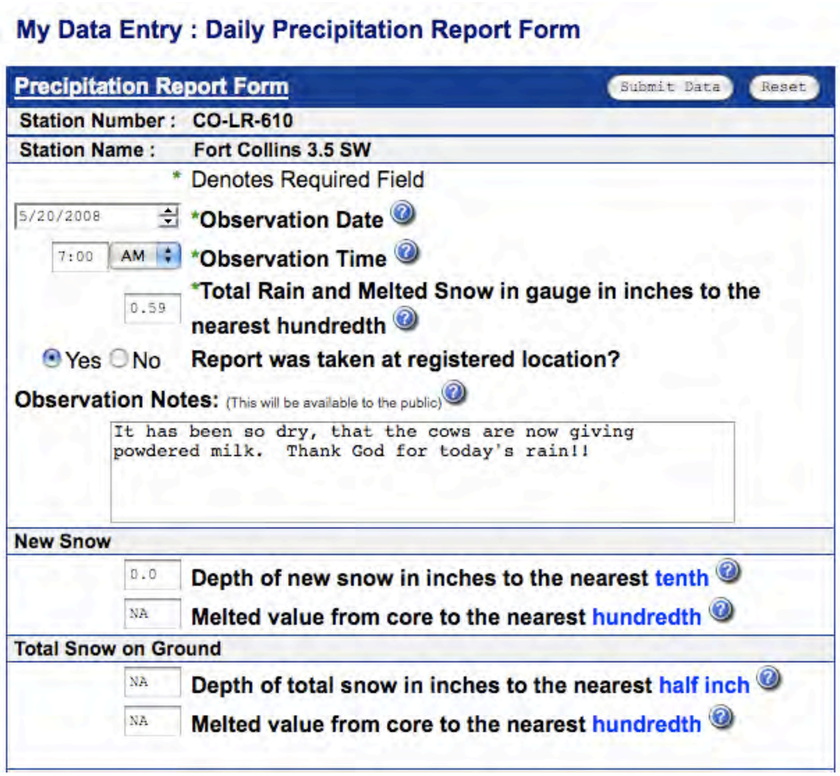
Once trained, our volunteers collect data using low-cost measurement tools . . .

4-inch diameter
high capacity rain gauges


Aluminum foil-wrapped
Styrofoam hail pads


Training is important to assure accurate, high quality data

# and <u>report</u> their daily observations on our interactive Web site: www.cocorahs.org



My Data Entry : Daily Precipitation Report Form

**Precipitation Report Form** — Submit Data | Reset

Station Number : CO-LR-610

Station Name : Fort Collins 3.5 SW

* Denotes Required Field

5/20/2008 — *Observation Date

7:00 AM — *Observation Time

0.59 — *Total Rain and Melted Snow in gauge in inches to the nearest hundredth

Yes No — Report was taken at registered location?

Observation Notes: (This will be available to the public)

It has been so dry, that the cows are now giving powdered milk. Thank God for today's rain!!

**New Snow**

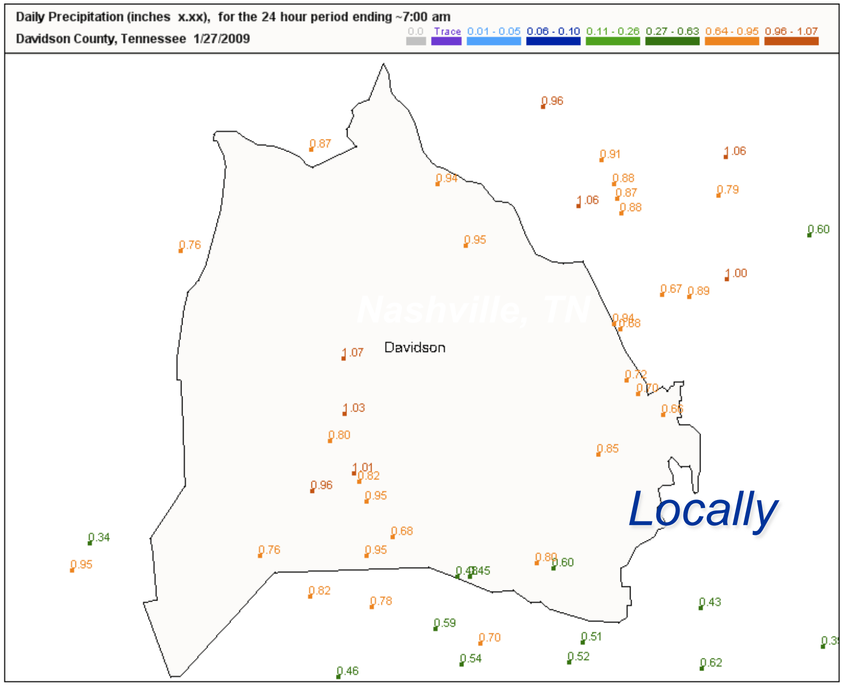
0.0 — Depth of new snow in inches to the nearest tenth

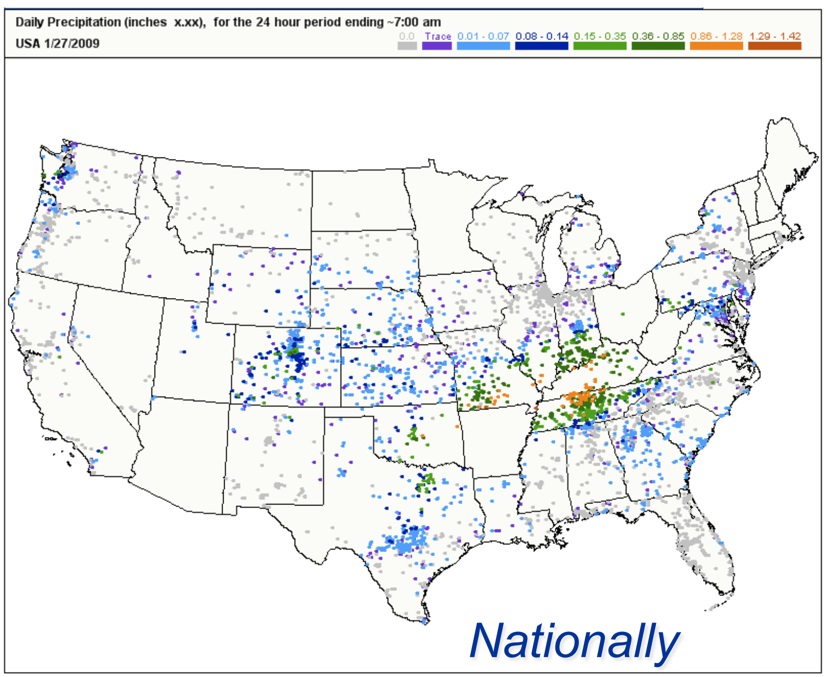
NA — Melted value from core to the nearest hundredth

**Total Snow on Ground**

NA — Depth of total snow in inches to the nearest half inch

NA — Melted value from core to the nearest hundredth




***Volunteer's observations are immediately available in map and table form for the public to view.***

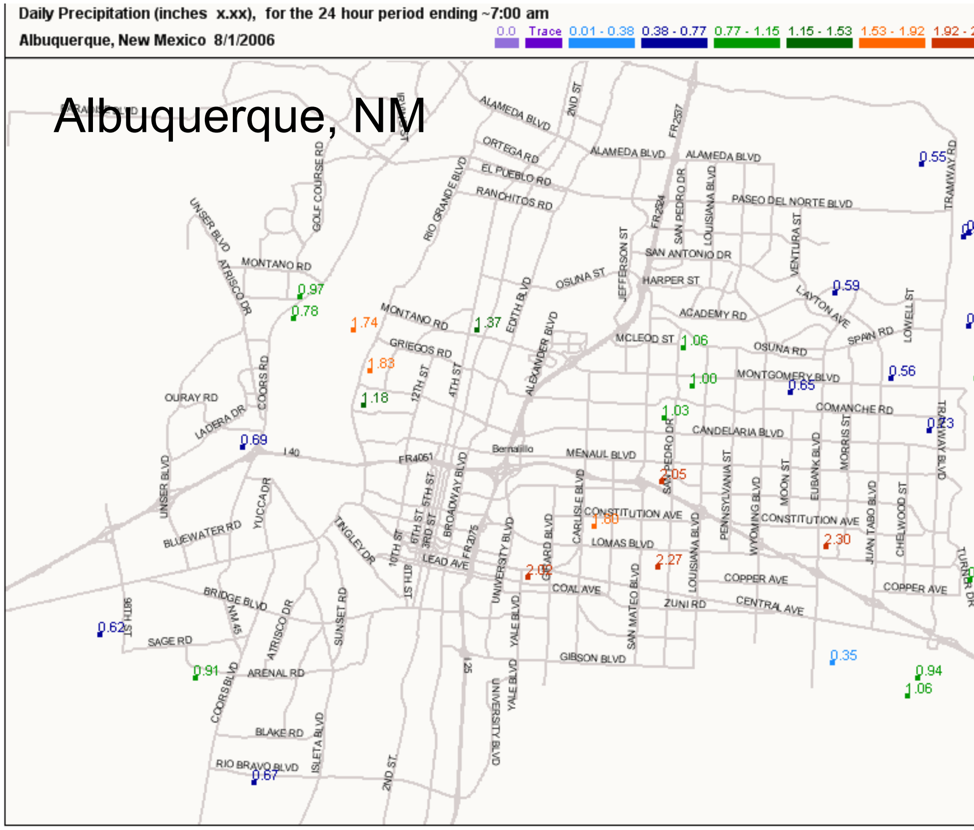
| Date | Time | Station Number | Station Name | Total Precip .Ins | New Snow .In | Total Snow .In | State | County | View |
|---|---|---|---|---|---|---|---|---|---|
| 1/19/2009 | 7:00 AM | RI-PR-10 | Woonsocket 0.3 W | 1.00 | 11.5 | NA | RI | Providence | |
| 1/19/2009 | 7:00 AM | RI-PR-7 | Cranston 1.9 E | 0.84 | 7.0 | 9.5 | RI | Providence | |
| 1/19/2009 | 8:00 AM | RI-PR-13 | Pawtucket 1.4 NE | 0.83 | 8.7 | 13.0 | RI | Providence | |
| 1/19/2009 | 7:00 AM | RI-PR-11 | Providence 3.0 ENE | 0.82 | 9.0 | 12.0 | RI | Providence | |
| 1/19/2009 | 8:00 AM | RI-WS-8 | Saunderstown 2.2 NW | 0.73 | 6.0 | NA | RI | Washington | |
| 1/19/2009 | 8:30 AM | RI-KN-2 | East Greenwich 2.3 ESE | 0.67 | 5.5 | 9.5 | RI | Kent | |
| 1/19/2009 | 7:00 AM | RI-NW-3 | Jamestown 2.6 NNW | 0.62 | 5.4 | 11.0 | RI | Newport | |
| 1/19/2009 | 7:00 AM | RI-WS-7 | North Kingstown 3 N | 0.59 | 5.0 | 9.0 | RI | Washington | |
| 1/19/2009 | 8:30 AM | RI-KN-1 | Coventry Center | 0.54 | 8.6 | 10.0 | RI | Kent | |
| 1/19/2009 | 7:00 AM | RI-WS-6 | Narragansett Pier 0.5 N | 0.49 | 4.2 | 7.5 | RI | Washington | |
| 1/19/2009 | 8:00 AM | RI-WS-9 | Charlestown 3.9 NNW | 0.46 | 6.2 | 9.0 | RI | Washington | |
| 1/19/2009 | 9:00 AM | RI-WS-1 | Hope Valley 3.7 S | 0.34 | 3.9 | NA | RI | Washington | |
| 1/19/2009 | 9:30 AM | RI-WS-5 | Kingston 0.5 NW | 0.31 | 3.4 | 8.0 | RI | Washington | |
| 1/19/2009 | 11:59 PM | RI-PR-14 | Woonsocket 1.3 ESE | 0.13 | 2.0 | 12.0 | RI | Providence | |
| 1/19/2009 | 7:00 AM | RI-NW-4 | Middletown 1.1 SW | NA | 3.0 | 7.0 | RI | Newport | |

CoCoRaHS

Community Collaborative Rain, Hail & Snow Network

CoCoRaHS's main focus is to provide:

# precipitation *data* . . .




Daily precipitation maps:
Rainfall, Hail and Snowfall

| Date | Time | Station Number | Station Name | Total Precip .in | New Snow .in | Total Snow .in | State | County | View |
|---|---|---|---|---|---|---|---|---|---|
| 2/14/2007 | 7:00 AM | MD-GR-1 | Mc Henry 4.0 SSE | 2.85 | 6.7 | 12.5 | MD | Garrett | |
| 2/14/2007 | 11:59 PM | MD-MG-8 | Gaithersburg 2 WNW | 2.80 | 4.2 | 4.0 | MD | Montgomery | |
| 2/14/2007 | 10:00 AM | MD-CR-7 | Westminster 1.0 W | 2.10 | 5.5 | 5.5 | MD | Carroll | |
| 2/14/2007 | 7:40 AM | MD-MG-1 | Montgomery Village 1.3 SSW | 2.05 | 4.1 | 3.0 | MD | Montgomery | |
| 2/14/2007 | 5:44 AM | MD-WH-1 | Williamsport 2.8 ENE | 1.92 | 2.6 | 5.0 | MD | Washington | |
| 2/14/2007 | 7:15 AM | MD-CR-3 | Mount Airy 0.2 SE | 1.90 | 5.1 | 5.0 | MD | Carroll | |
| 2/14/2007 | 7:00 AM | MD-CR-6 | Taneytown 3.2 NE | 1.83 | 5.0 | NA | MD | Carroll | |
| 2/14/2007 | 7:00 AM | MD-HW-2 | Sykesville 1.7 SSE | 1.78 | 5.0 | 5.0 | MD | Howard | |
| 2/14/2007 | 7:00 AM | MD-HW-12 | Sykesville 2.6 SE | 1.61 | 0.0 | NA | MD | Howard | |
| 2/14/2007 | 8:00 AM | MD-MG-3 | Potomac 0.9 NNW | 1.54 | 3.2 | NA | MD | Montgomery | |
| 2/14/2007 | 7:00 AM | MD-MG-2 | Redland 0.8 NNE | 1.52 | 4.5 | 4.5 | MD | Montgomery | |
| 2/14/2007 | 7:00 AM | MD-PG-37 | Brandywine 6.7 ESE | 1.49 | T | T | MD | Prince George's | |
| 2/14/2007 | 7:00 AM | MD-PG-1 | Bowie 0.5 E | 1.47 | 1.0 | 1.5 | MD | Prince George's | |
| 2/14/2007 | 7:00 AM | MD-SM-3 | Leonardtown 0.6 NE | 1.42 | 0.0 | NA | MD | St. Mary's | |
| 2/14/2007 | 7:00 AM | MD-CH-7 | Waldorf 3.2 SW | 1.40 | 0.8 | 0.7 | MD | Charles | |
| 2/14/2007 | 7:00 AM | MD-HW-11 | Columbia 1.7 W | 1.40 | 3.2 | 3.5 | MD | Howard | |
| 2/14/2007 | 7:00 AM | MD-PG-7 | Camp Springs 1.6 NNW | 1.38 | 1.8 | NA | MD | Prince George's | |
| 2/14/2007 | 4:00 PM | MD-BL-7 | White Hall 3.5 NE | 1.38 | NA | NA | MD | Baltimore | |
| 2/14/2007 | 7:00 AM | MD-CV-1 | Marlton 6.0 E | 1.37 | 0.3 | 0.0 | MD | Calvert | |
| 2/14/2007 | 7:00 AM | MD-SM-4 | Charlotte Hall 3.6 ENE | 1.37 | 0.3 | T | MD | St. Mary's | |
| 2/14/2007 | 7:00 AM | MD-MG-24 | White Oak 1.2 N | 1.35 | 2.5 | 2.0 | MD | Montgomery | |
| 2/14/2007 | 7:00 AM | MD-PG-35 | Brandywine 2.5 NNW | 1.35 | 1.0 | 1.4 | MD | Prince George's | |
| 2/14/2007 | 7:00 AM | MD-WC-1 | Vienna 11.3 SSW | 1.35 | 0.0 | NA | MD | Wicomico | |
| 2/14/2007 | 7:00 AM | MD-PG-6 | Friendly 1.0 N | 1.32 | 2.4 | 1.8 | MD | Prince George's | |

Daily data
in table form

This data allows CoCoRaHS to supplement existing networks and provide many useful results to scientists, resource managers, decision makers and other end users on a timely basis.

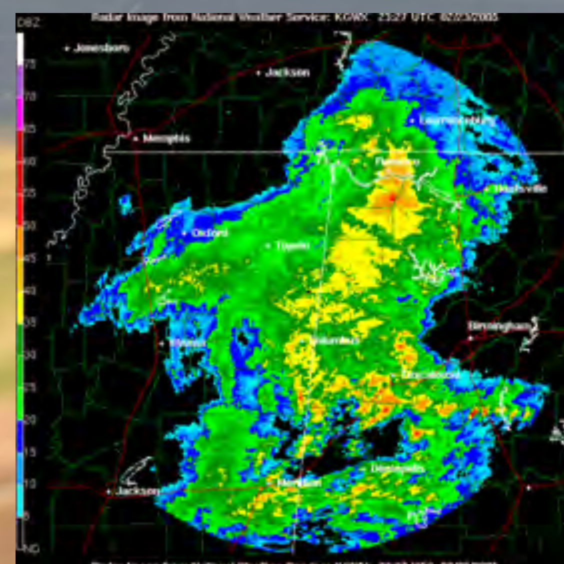
# . . . as well as ***educational opportunities***




# CoCoRaHS data is used by many

- National Weather Service
- Other Meteorologists
- Hydrologists
- Emergency Managers
- City Utilities
  - -Water supply
  - -Water conservation
  - -Storm water
- Insurance adjusters
- USDA—Crop production
- Engineers
- Scientists studying storms
- Mosquito control
- Farm Service Agency
- Ranchers and Farmers
- Outdoor & Recreation

- Teachers and Students
  - – Geoscience education tool
  - – Taking measurements
  - – Analyzing data
  - – Organizing results
  - – Conducting research
  - – Helping the community



# CoCoRaHS hopes to one day achieve a network of . . .

one observer <u>**every square mile**</u> in <u>**urban**</u> areas

one observer <u>**every 36 square miles**</u> in <u>**rural**</u> areas

# How Can you become part of the network?

## Five easy steps

*Simply sign-up on the CoCoRaHS web page* ***www.cocorahs.org***

*Obtain a 4" plastic rain gauge (info available on web site)*

*View the "training slide show" or attend a training session*

*Set-up the gauge in a "good" location in your backyard*

*Start observing precipitation and report on-line daily*


Just 5 minutes a day!
It's easy and fun!
We're
Cuckoo
For
CoCoRaHS!
www.cocorahs.org

# Example II: Reporting Floods

- To avoid data quality issues, the national effort has an automated quality control mechanism that highlights “suspect” reports.
  - This in recognition of fact that volunteer data is not always accurate
- Individual users, such as NJ State Climatologist, also use manual quality control


colostate.edu

# Example II: Reporting Floods

- The effort to use CoCoRaHS flooding reports from Hoboken in the Health Impacts Analysis study failed.
- Why?
- While National Weather Service/NOAA data on flooding in Hoboken was incomplete, so was CoCoRaHS data.
- Also, NOAA and CoCoRaHS use different categories of flooding in their reports.

# Example II: Reporting Floods

- NOAA categories (3):
  - Minor Flooding – minimal or no property damage, but possibly some public threat or inconvenience
  - Moderate Flooding - some inundation of structures and roads near streams. Some evacuations of people and/or transfer of property to higher elevations are necessary
  - Major Flooding - extensive inundation of structures and roads. Significant evacuations of people and/or transfer of property to higher elevations
- CoCoRaHS flooding data categories (4):
  - Minor (typical). Street or field flooding
  - Unusual street or field flooding (only see this every few years)
  - Severe flooding
  - Extreme (never seen it this bad before)

# Example II: Reporting Floods

- CoCoRaHS's flood severity definitions were developed with a focus on making the system user friendly (easy to understand)
  - Volunteers range in age significantly and include fairly young ones
- Unfortunately, then, definitions don't match up with the government used definitions.

# Example II: Reporting Floods

- Summary:
  - Citizen science data set and more formal data set hard to match up
  - Simplifying things for citizen scientists can lead to problems with data
  - Incomplete data, even for small area
  - Data collection mostly not automated
  - Instructions not precise

# Example III: Using Social Media to Gain Situational Awareness*

- Social media can help to provide situational awareness in case of developing emergencies
- Projects at the CCICADA Center (Ji and Wallace, Hovy and Metzer, Nelson & Pottenger) analyzed data from Twitter from Haitian Earthquake of 2010 & Japanese Earthquake and Tsunami of 2011 – over a billion tweets

*Thanks to Christie Nelson and others at CCICADA for this example

CCICADA

Command, Control, and Interoperability Center for Advanced Data Analysis

# Example III: Using Social Media

- During a disaster, people send social media messages with various types of requests.
- Help may be requesting information, medical aid, etc.
- Work in these projects has found:
  - Great diversity of communication
  - Interesting characteristics of network spread
  - People coordinate in different ways
  - People follow typical sequences when communicating in emergency situations
- Understanding typical sequence allows crisis responders and others to identify "relapses," pick out anomalies, etc.

CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis

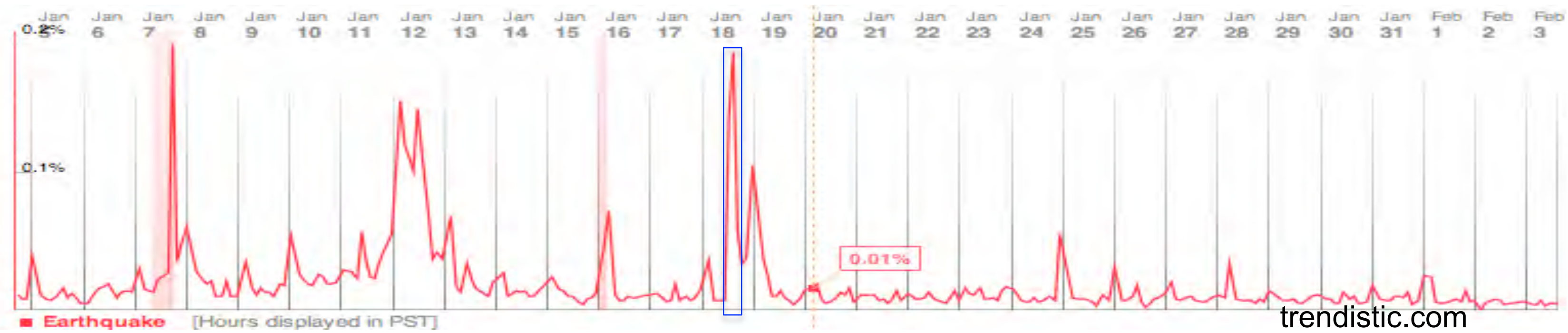
# Example III: Using Social Media

- CCICADA researchers developed tools to learn "topic signatures" that indicate when an event of a given type occurs
  - Discover 'burst' of topic-related words in timespan
  - Identify relevant tweets
  - Extract main ones to build a summary
  - Monitor as event unfolds
  - Pick out anomalies, etc.

Command, Control, and Interoperability Center for Advanced Data Analysis

# Example III: Using Social Media Challenge : Detect Events

Metzler et al. 2011




Query: q

earthquake

Step 1:  q' = EXPAND (q)

**Event signature:** *earthquake, earthquake., earthquake.., magnitude, epicenter, earthquake.., foreshocks, usgs, tsunami, indonesia, …*

Step 2: TS = RANK_TIMESPAN(q')

"Burstiness" of term *w* in timespan *T*: $\beta_T(w) = \frac{P(w|T)}{P(w)}$

Step 3: SUMMARIZE(TS)

**Timespan 1**: Jan 18, 2010 at 22 UTC, for 9 hours

**Summary tweets:**

*i.#IWO Latest on Pakistan earthquake tonight: 7.3 mag quake in SW Pakistan http://bit.ly/guYOry*

*ii."@cnnbrk: USGS: An earthquake with preliminary magnitude of 7.4 strikes southwestern Pakistan http://on.cnn.com/gQcnRa*

*iii.Major quake hits Pakistan: An earthquake with a preliminary magnitude of 7.4 struck Wednesday morning in southwe… http://bit.ly/iildWP*

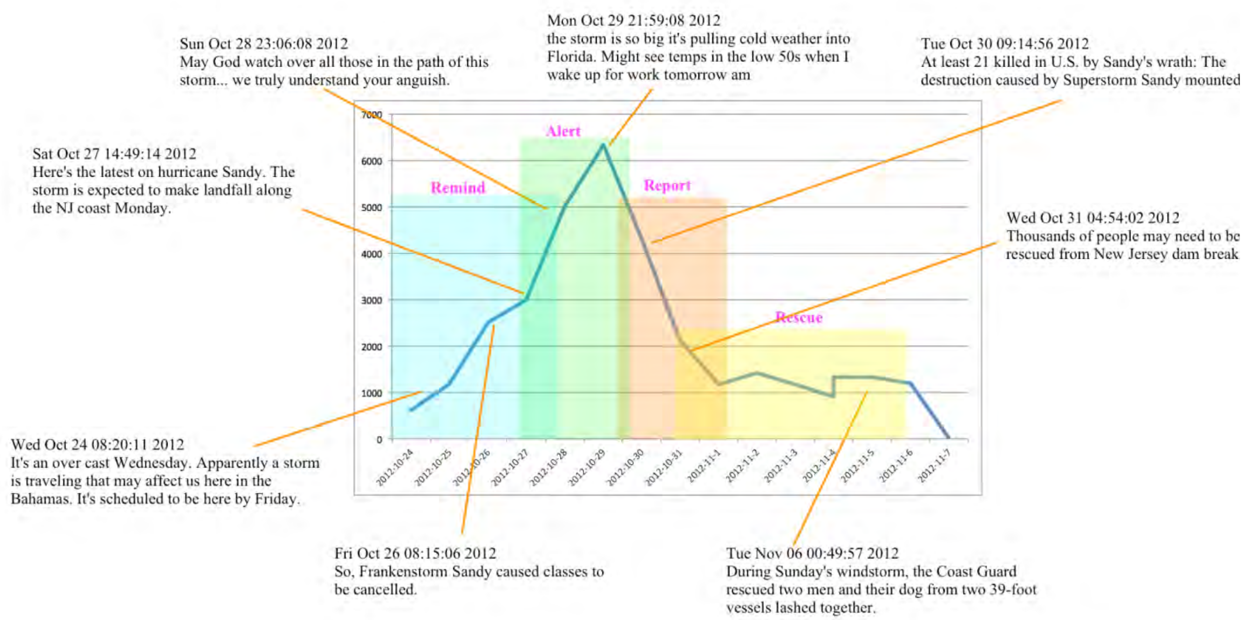
Command, Control, and Interoperability Center for Advanced Data Analysis

# Example III: Using Social Media Challenge 2. Track Event Stages

- Model 1: simple linear timeline
  - Earthquake:
    - 3 mins of fear
    - 30 mins of family & friends outreach
    - 3 hours of planning immediate actions
    - 3–5 days of discussion about repair activities
- Model 2: multiple timelines, depending on relationship of tweeters to event
  - It's different if you are there than if you just heard about it

**10 topics and 10 top words for each topic (generated by LDA)**

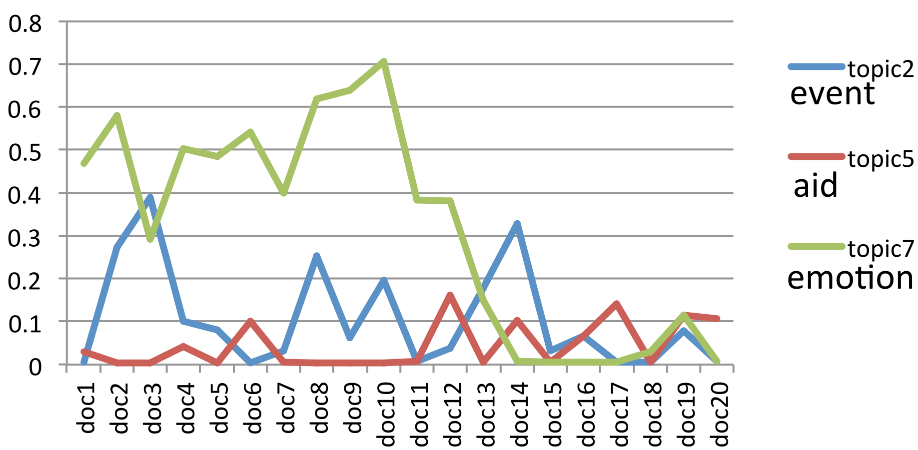
| topic0 | topic1 | topic2 | topic3 | topic4 | topic5 | topic6 | topic7 | topic8 | topic9 |
|---|---|---|---|---|---|---|---|---|---|
| stupid | hate | magnitude | condolences | tsunami | donate | depth | feel | large | upgraded |
| supermoon | destroy | damage | bodies | affected | text | epicenter | felt | crack | triggering |
| search | teen | weird | islands | people | relief | coast | shit | photo | large |
| total | violence | make | found | hit | redcross | offshore | thought | showing | effects |
| utter | gangs | started | worst | prayers | victims | miles | shaking | death | unbelievable |
| offensive | protect | rocked | city | thoughts | red | strikes | tonight | toll | issued |
| post | depth | hurricane | sweeping | news | cross | survey | big | raise | flash |
| bollocks | thing | news | living | massive | support | bst | time | police | feed |
| site | stopped | back | slam | coast | record | region | scary | earth | suffered |
| response | | fault | snow | pray | strongest | revised | back | unleashed | larger |
| | | event | | sympathy | aid | location | emotion | | |



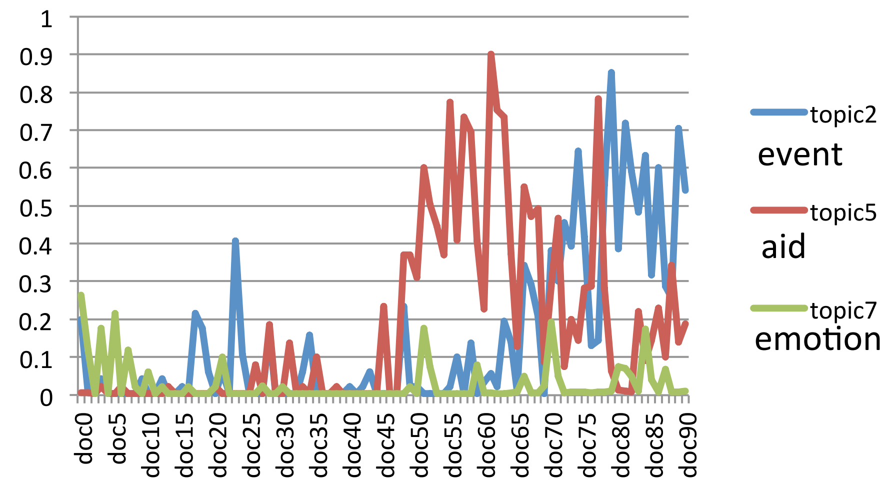

Command, Control, and Interoperability Center for Advanced Data Analysis

# Example III: Using Social Media: Two Earthquakes on Twitter

S. California, Jun 2010




Japan, Mar 2011




Each ‘document’ is a bucket of 100 tweets, sorted in time order. Initial discussion about emotions, then focus shifts to aid

Less about emotion: English-language tweeters were not so present in Tokyo.
More about event: implications for Dai-ichi Nuclear Plant

**Need to determine details of location and participants of events**

# Example III: Using Social Media Real Time Optimization of Emergency Response Using Social Media

- CCICADA Project by Nelson and Pottenger
- Goal: Use social media data to identify most important items requested during an emergency
- Created a framework that:
  - Grouped messages by location (clustering)
  - Determined top requests by location using machine learning (Higher Order Naïve Bayes – HONB – or Higher Order Latent Dirichlet Assn. – HO-LDA)
  - Allocated aid based on integrated social media geolocations requests received
- Applied ideas to social media data from 2010 Haitian Earthquake

# Example III: Using Social Media

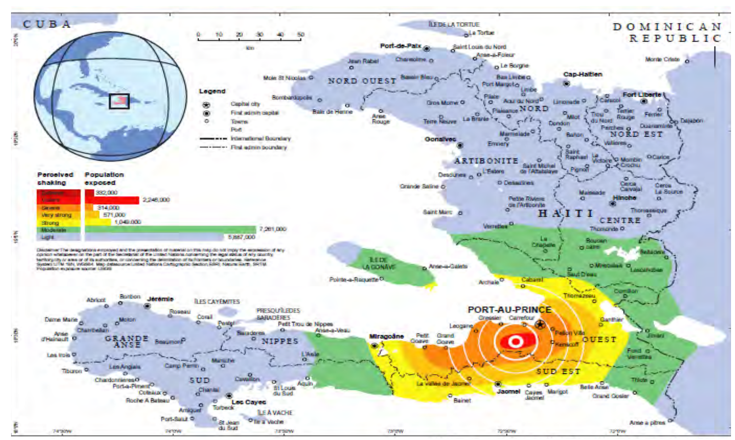
- Approach developed using Ushahidi data obtained during the 2010 Haitian Earthquake.
- Ushahidi, Inc.:
  - Non-profit software company that develops free and open-source software for information collection, visualization, and interactive mapping.
  - Founded following Kenya's disputed presidential election in 2007.
  - Uses crowdsourcing for social activism and public accountability

Ushahidi volunteers manually determining aid requests remotely from the US

# Example III: Using Social Media

- During a disaster, people may send social media messages for “help” of some kind
  - Help may be requesting information or aid, etc.
  - Looked at Haitian Earthquake of 2010
  - Ushahidi social media and text message dataset
    - ✓3,358 messages over 45 days (Jan 13, 2010 – Feb 26, 2010)
    - ✓Data included social media messages along with texts sent to an emergency number, and geolocation

Haiti after the earthquake




Haiti earthquake intensity map

# Example III: Using Social Media

- Created a framework that:
  - Grouped messages by location (**clustering**)
  - Determined top requests by location (**machine learning**)
    - ✓Potential "aid" requests (class labels): Hospital Clinics Operating; Services Available; Medical Emergency; Security Threats; People Trapped; Medical Supply; Water Shortage; Food Shortage; Help; Hygiene (water); Human Remains; Shelter; Vital Lines (Infrastructure); Fuel Shortage; Clothing; Damaged Structure; Power Outage; Persons News; Other
  - Allocated aid based on traditional methods and pre-existing facility locations that integrated the social media geolocations and requests (**resource allocation model**)

# Example III: Using Social Media

- Haiti – Individual Social Media Messages and Pre-existing Facilities *(illustrative)*


Individual social media messages received one at a time (like what Ushahidi does)

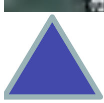
Hospital or Clinic

x **Social Media Message**

*Resource requested*

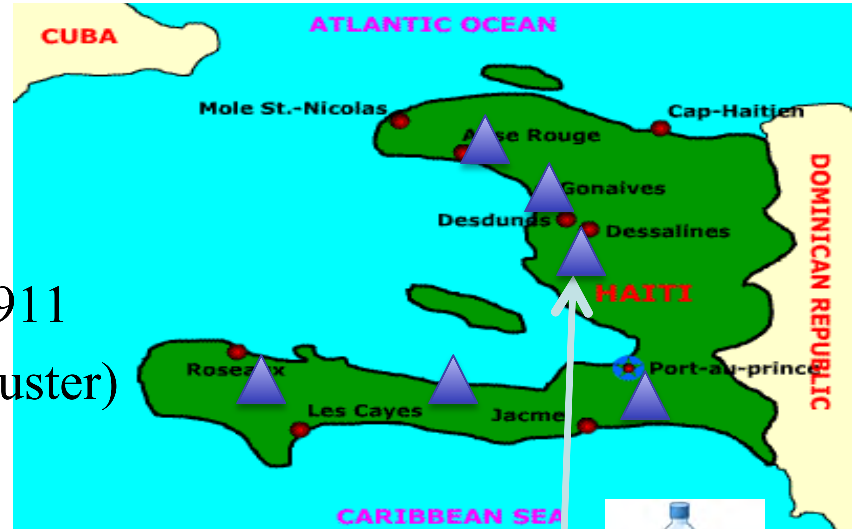
# Example III: Using Social Media

- Haiti – Social Media Messages – locations and requests


 (location, resource)

- New locations of groups of people based on social media messages, along with their “need” requests

# Example III: Using Social Media




911
(cluster)


"we do not have water we are in Santo 15 thank you"

After several messages are received, first they are **CLUSTERED**

*Supervised method*

HONB learned on the cluster

Class: Water

*Predict resource need*

OR

*Unsupervised method*

HO-LDA learned on the cluster

Topic 1
*(water, liquid, fluid, thirsty)*
*Words matched to category*

*with xx% probability associated*

Mixed integer programming resource allocation model

**Shelter *a* needs Water**

*Messages received; Messages CLUSTERED; Model learned with either HO-LDA or HONB; then predicted need put into RESOURCE ALLOCATION Model*

# Example III: Using Social Media

- Challenges using Social Media:
  - Data during emergencies is often inconsistent or conflicting
  - Could be due to noise or malicious intent
  - ***Not sure what you can trust***
- Need to develop computational tools to address problem of trustworthiness in such contexts
- Need find appropriate degree of "trust" in claims made.
- Need precise definitions of and metrics for factors contributing to trust: accuracy, completeness, bias

# Example III: Using Social Media

- Social media is being used to gain situational awareness and for political activism (as by Ushahidi).
- Very distributed information can be handled by social media.
- Good to use additional features – not just content of messages, but also things like geolocation that are obtained automatically without participants having to provide it.
- Good to use other data sources such as information from sensors, news reports, etc.
- But, to repeat: Big obstacle is TRUST.
- Without Trust, data cannot provide evidence.

# Example III: Using Social Media

- In this way, citizen science (crowdsourcing) can be used to ***make decisions during an emergency***.
- Or to be used to review previous emergencies and ***make policy*** based on what was discovered.
- The former requires rapid usage of data and data trust becomes a critical issue
- The latter can be done more slowly, and there may be more time to evaluate the trustworthiness of the data and to select subsets of it that are trustworthy.

# Discussion: For What Issues Concerning Natural Disasters Could Citizen Science Help?

- Here a few questions from CCICADA studies
- Evacuations during an emergency:
  - How can we get early warning to citizens that they need to evacuate**?** ***Social media clearly relevant***
  - How can we plan such evacuations effectively? ***Citizen input could be gathered; information about where people have gone in past could be obtained***


# Discussion

- More on evacuations:
  - Can plans for who should go to which shelter be implemented quickly; can we get information to people quickly? ***Social media***.
- Can we develop evacuation plans that are quickly modifiable given data from evacuation centers, traffic management, flood reports, etc.? ***Citizens report in. Social media and other ways***.
- Far from what happens in evacuations today.

# Discussion

- Subway flooding during storms.
  - What subways will be flooded? ***If the question were "are being flooded," citizen science clearly relevant.***
  - How can we protect against such flooding? ***Could citizen science be used to gather input from past storms to determine most vulnerable subways?***


# Discussion

- How can we plan placement of utility lines to minimize down time? ***Could we use citizen science to gather information about downed lines during storms to gather data to plan ahead?***

# Discussion

- How can we plan for getting people back on line after a storm? Who gets priority?
- Answer may depend on other needs that depend on getting your business on line. (Can't pump gas without power.) ***Citizen science could help with information from previous storms.***
- Answer may depend on special needs of people living in a given home. ***Citizen science could help even during a storm – using social media***.



Bringing in help from out of state

# Discussion

- How can we set priorities for cleanup? ***Citizen reports from earlier storms or social media from current one.***

# Discussion

- What supplies are needed during an emergency?
  - Water, Food, Fuel, Generators, Chainsaws?
- How and where can we stockpile them?
- What are good methods for getting these to those who need them in an efficient way?
- ***For citizen science, see Nelson-Pottenger example.***


# Discussion

- How can we tell who needs what kinds of goods during an emergency? ***This is Nelson-Pottenger*.**
- How can we locate stockpiles so as to be "agile" in allocating the resources when needed? ***Use citizen reports from previous storms.***
- E.g.: strategic national stockpile of medicines for emergencies from CDC (Centers for Disease Control): how do we decide what medicines to include, how many doses, where to keep them? ***This is less likely to benefit from citizen science??***

Source: cdc.gov

# Discussion

- How do we plan emergency rescue vehicle routing to avoid rising flood waters while still minimizing delay in provision of medical attention and still getting afflicted people to available hospital facilities? ***Citizen science from reporting where previous floods keep roads open***


***Data is Important.***

***We need to understand the limits it puts on citizen science and its usefulness.***