

# ENGAGING THE PUBLIC IN CLIMATE SCIENCE

## EXPLOITING CROWDSOURCING TO DIGITIZE AND ANALYZE CLIMATE DATA



**Scott Hausman**

*Deputy Director*

**National Climatic Data Center**

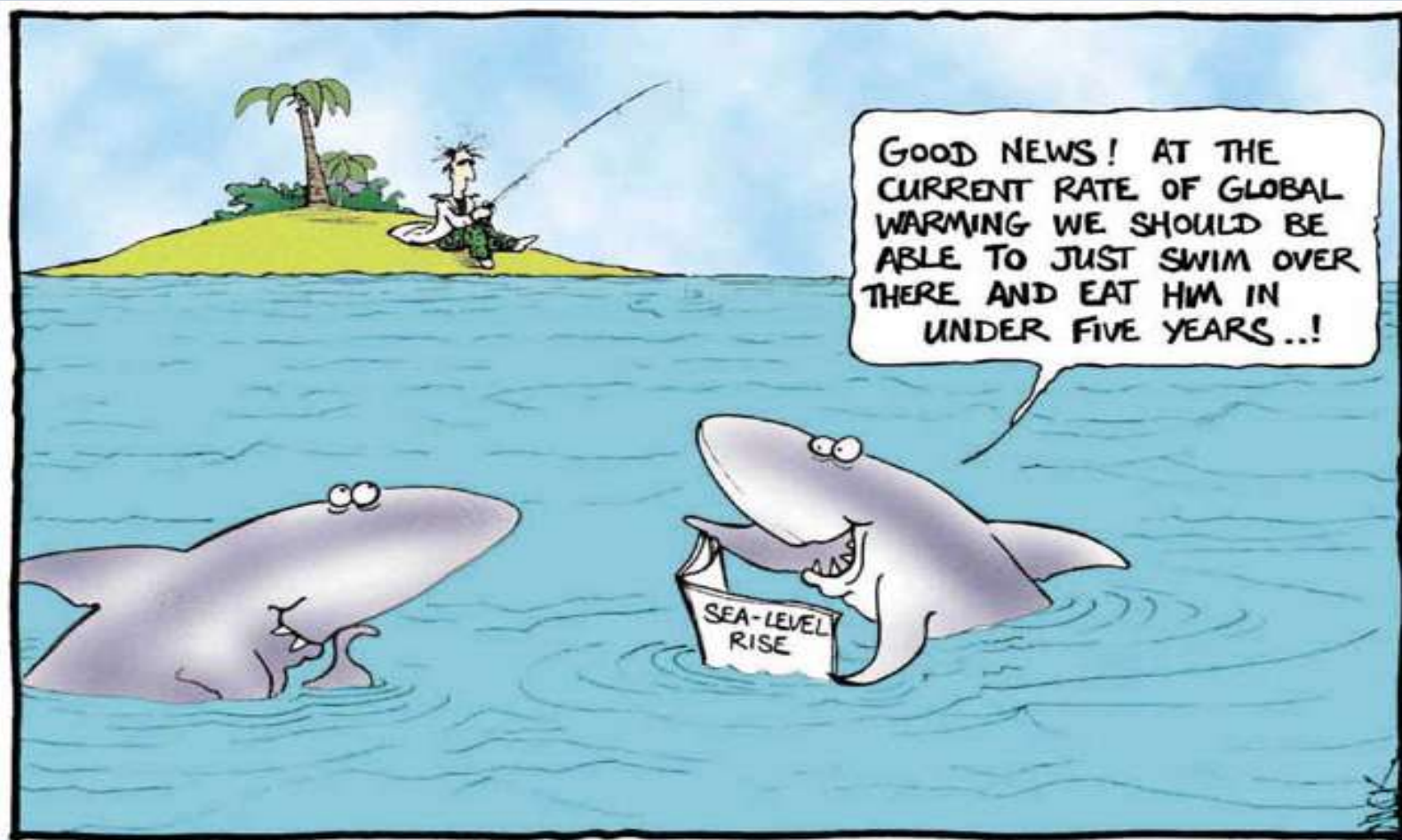
13 June 2011

Board on Research Data & Information (BRDI)

# OUTLINE

---

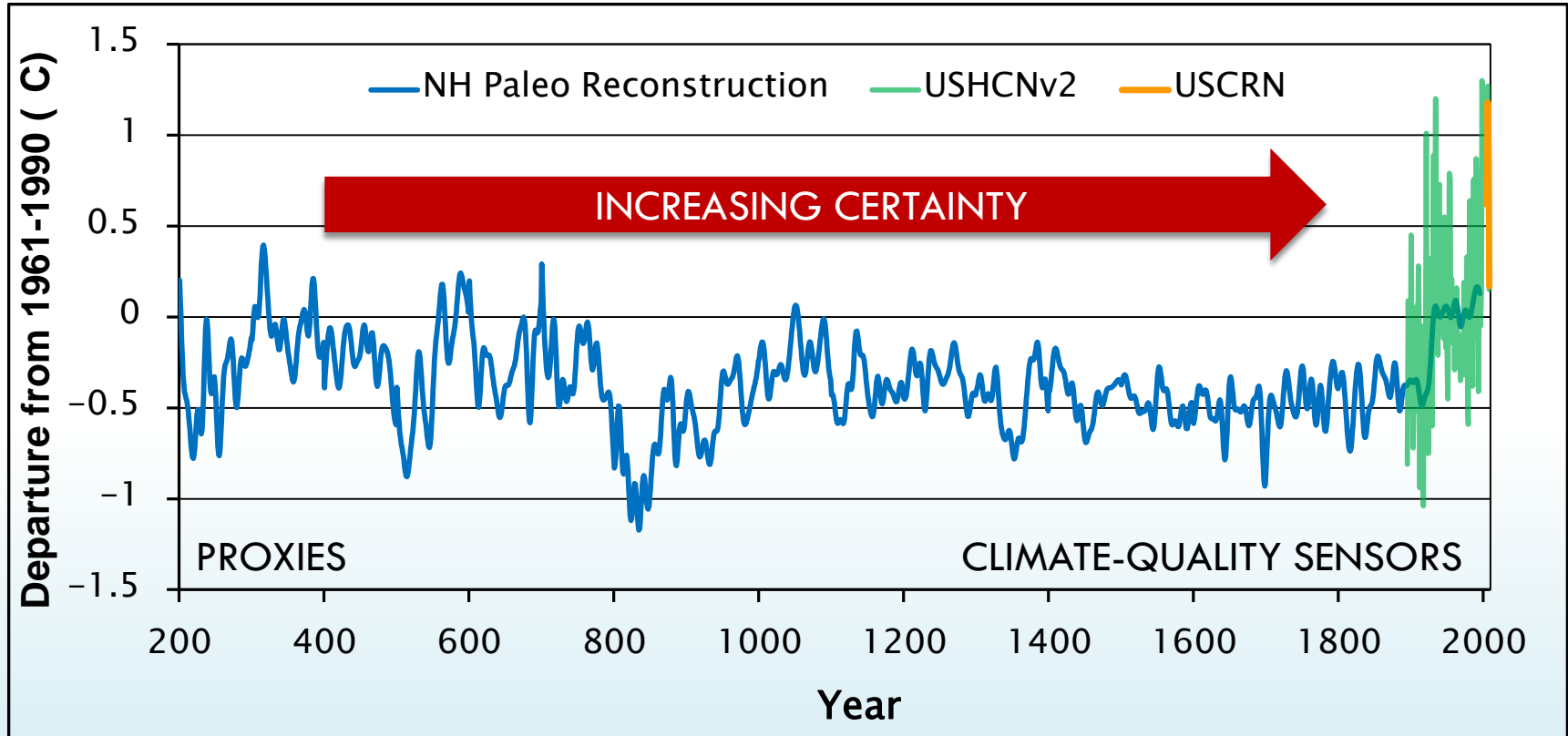
- Challenges of Monitoring Climate Change
- Recruiting Internet Citizen Scientists
- Partnering with Citizen Science Alliance
  - Project 1: Data Rescue for Surface Temperature Databank
  - Project 2: Tropical Cyclone Reanalysis



## CHALLENGES OF MONITORING CLIMATE CHANGE

Reducing Uncertainty to Improve Public Understanding

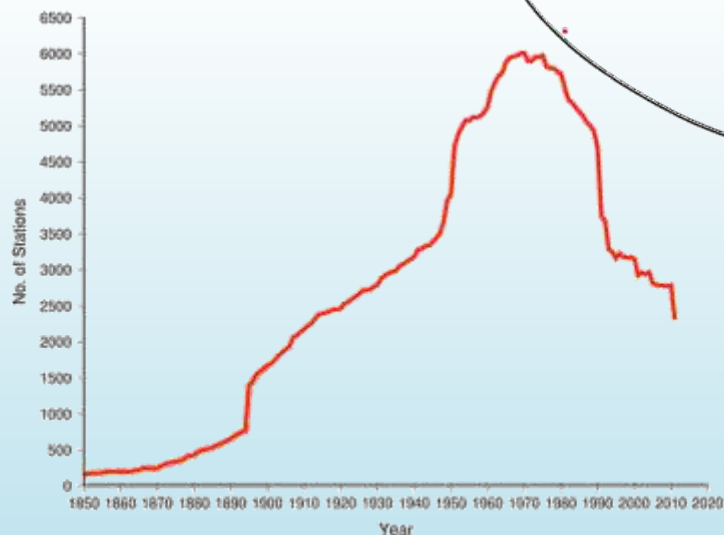
# OBSERVATION QUALITY AND UNCERTAINTY



Three temperature time series are displayed relative to the 1961-1990 normals ( $^{\circ}\text{C}$ ). The paleoclimate reconstruction from Mann et al. (2008) is for Northern Hemisphere land, while U.S. Historical Climatology Network v. 2 dataset from Menne et al. (2009) and U.S. Climate Reference Network temperature departures are for the continental U.S. only. A major goal is to provide consistent time series across paleoclimate and instrumental networks.

# FILLING SPATIAL GAPS FOR REGIONAL CHANGE DETECTION

- GHCNv3 has about 2700 reliable stations
- Rapid increase during WWII
- Gradual decrease in 80s, 90s as stations close



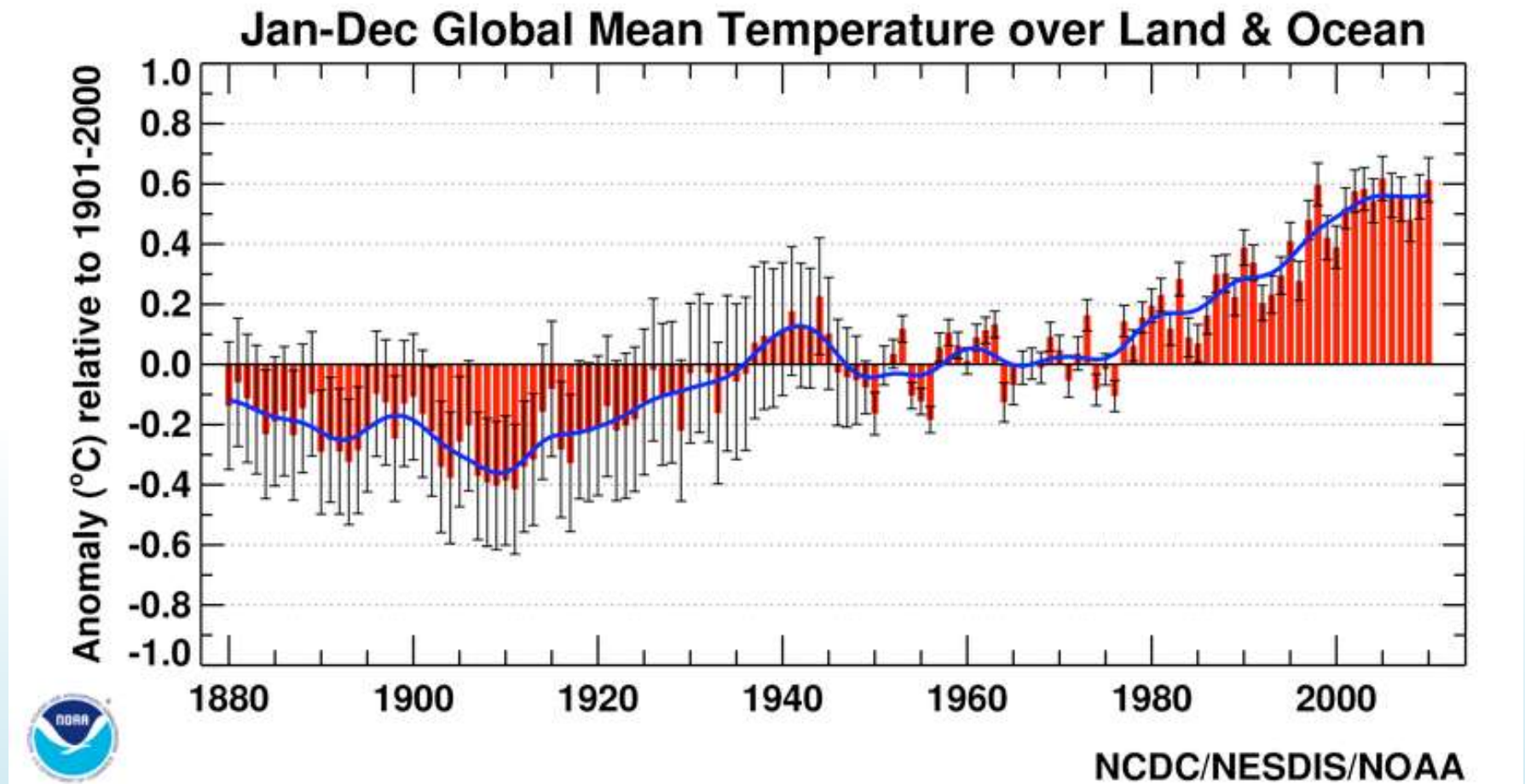
## Global Historical Climatology Network (Monthly) Stations Reporting Mean Temperature, 1961-1990



- Some data still not in digital form
- Rescuing data has been first priority
- Critical to understanding regional climate change.



# REDUCING UNCERTAINTY IN GLOBAL TRENDS



More observations needed to reduce uncertainty in historical record.

# MAKING DATA ACCESSIBLE FOR DISCOVERY & ANALYSIS

- NOAA On-line Foreign Data Library
  - Over 75 countries and former colonies
  - Covers from 1830 to 1970s
- Digitally imaged through Climate Database Modernization Program(CDMP)
  - Millions of images; more accessible
  - Data is still essentially
- Additional 2000+ boxes of international data in NCDC physical archives
  - Similar numbers of data held elsewhere
  - None of it is usable without digitization
- [http://docs.lib.noaa.gov/rescue/data\\_rescue\\_home.html](http://docs.lib.noaa.gov/rescue/data_rescue_home.html)

Septembre 1878.

SIDI-BEL-ABBÉS.

Longitude 2°59' W. — Latitude 35°2'. — Altitude 476<sup>m</sup>,1.

DATES.	BAROMÈTRE (700+).			THERMOMÈTRE.					HUMIDITÉ RELATIVE.			VENT. Direction et force			SÉCULOSITÉ.			PLUIE.	REMARQUES.
	7 <sup>h</sup>	1 <sup>h</sup>	7 <sup>h</sup>	7 <sup>h</sup>	1 <sup>h</sup>	7 <sup>h</sup>	Min.	Max.	7 <sup>h</sup>	1 <sup>h</sup>	7 <sup>h</sup>	7 <sup>h</sup>	1 <sup>h</sup>	7 <sup>h</sup>	7 <sup>h</sup>	1 <sup>h</sup>	7 <sup>h</sup>		
1	22.4	19.4	20.6	21.6	31.1	28.2	18.4	35.0	82	36	56	Calme	ENE 5	Calme	6	0	1	.	
2	23.0	21.5	20.4	22.6	32.4	27.6	19.0	34.4	86	34	71	Calme	NNE 3	NNE 2	6	0	6	.	
3	20.9	19.5	19.2	21.4	31.0	27.3	18.0	36.8	87	36	53	NNE 1	ENE 3	ENE 3	3	6	6	.	
4	20.1	20.8	21.7	20.0	28.0	25.6	18.2	29.8	91	47	57	Calme	NE 2	NNE 1	1	8	6	0.0	
5	23.0	21.5	20.4	22.6	32.1	27.6	19.0	34.1	86	34	71	Calme	ENE 5	ENE 2	0	0	0	.	

# LIMITS TO OPTICAL CHARACTER RECOGNITION (OCR)

The collage consists of several historical weather-related documents and a central cartoon character:

- Top Left:** A form titled "Time entries on this farm are" with handwritten entries: "75<sup>th</sup> meridian time", "To convert to G. c. t. 5 hours", and "Height of barometer 16 feet (HSE)". A large number "12868" is written in the center, with "WEAR FROM SEA" below it.
- Top Right:** A document titled "H. C. Ship Berwickshire" dated "Saturday 17<sup>th</sup> April 1830". It contains a table with columns for "H", "K", "F", "Winds, &c.", and "Remarks". The table is partially filled with handwritten data.
- Center:** A cartoon character of a blue computer monitor with a face, arms, and legs, sitting at a desk with a keyboard and a mouse.
- Bottom Left:** A document titled "Diary of the Weather, Fort Amstel" dated "July 1820". It contains a table with columns for "Date", "Thermometer", "Cause of Weather", and "Remarks". The table is filled with handwritten data.
- Bottom Right:** A document titled "Diary of the Weather, Fort Amstel" dated "July 1820". It contains a table with columns for "Date", "Thermometer", "Cause of Weather", and "Remarks". The table is filled with handwritten data.





*"On the Internet, nobody knows you're a dog."*

©The New Yorker Collection 1993 Peter Steiner  
From cartoonbank.com. All rights reserved.

## RECRUITING INTERNET CITIZEN SCIENTISTS

We need carbon-based computing for imagery analysis.

# CITIZEN SCIENCE AND CROWDSOURCING

- Exploits the cognitive abilities of Human Computation!
- Novel mode of data collection:
  - Citizen Science = Volunteer Science = Participatory Science
  - e.g., VGI = Volunteer Geographic Information (Goodchild '07)
  - e.g., Galaxy Zoo @ <http://www.galaxyzoo.org/>
- Citizen science refers to the involvement of volunteer non-professionals in the research enterprise.
- The Citizen Science experience ...
  - must be engaging,
  - must work with real scientific data/information,
  - must not be busy-work (all clicks must count),
  - must address authentic science research questions that are beyond the capacity of science teams and enterprises, and
  - must involve the scientists.

Reference: Kirk Borne, Reference: Reaching Out with Eventful Astronomy, George Mason University

# EXAMPLE: RECAPTCHA

**reCAPTCHA™**

- WHAT IS reCAPTCHA
- GET reCAPTCHA
- PROTECT YOUR EMAIL
- MY ACCOUNT
- RESOURCES: DOCS & PLUGINS

reCAPTCHA IS A FREE ANTI-BOT SERVICE THAT HELPS DIGITIZE BOOKS.

steamboat train, from New  
this **morning** ran off the tra  
New-London. Four cars plung-

*morning* *upon*

Type the two words

reCAPTCHA™ stop spam, read books.

→ LEARN HOW reCAPTCHA WORKS

**USE reCAPTCHA ON YOUR SITE**

- STRONG SECURITY**
- ACCESSIBLE TO BLIND USERS**
- 30+ MILLION SERVED DAILY**

**NEW** See how accurate reCAPTCHA is at digitizing content!

Blog | About Us | Contact | FAQs | Terms | Privacy  
© 2011 Google, all rights reserved.

**Scanned type** → **This aged portion of society were distinguished from**

**OCR reads as** → **"niis aged pntkm at society were distinguished frow."**

# EXAMPLE: ZOONIVERSE

CITIZEN SCIENCE  ALLIANCE



## About the Zooniverse

The Zooniverse is home to the internet's largest, most popular and most successful citizen science projects. Our current projects [are here](#) but plenty more are on the way. If you're new to the Zooniverse, we suggest picking a project and diving in - the same account will get you into all of our projects, and you can keep track of what you've contributed by watching 'My Zooniverse'.

The Zooniverse and the suite of projects it contains is produced, maintained and developed by the [Citizen Science Alliance](#). The [member institutions](#) of the CSA work with many academic and other partners and rely on the efforts and ability of volunteers to help process data that confronts them.

The Zooniverse began with a single project in 2007. The Galaxy Zoo team had expected a small number of volunteers, but were overwhelmed by the response to the project. When the project was buckling under the strain, they set about creating the Zooniverse.

Galaxy Zoo was important because it produced many unique scientific results, ranging from those using classifications that depended on human eyes. This commitment to producing real results from your time - is at the heart of everything we do.

## Zooniverse Activity

Total Volunteers: **430,804**



Recent Zooniverse activity

## Live Projects

[planethunters.org](#)

[THE MILKY WAY PROJECT](#)

[MOON ZOO](#)

[GALAXY ZOO HUBBLE](#)

[oldWeather](#)

[SOLAR STORMWATCH](#)

[GALAXY ZOO](#)  
UNDERSTANDING COSMIC MERGERS

[GALAXY ZOO](#)  
THE HUNT FOR SUPERNOVAE

[ZOOSHOP](#)

WEAR THE BADGE  
SPREAD THE WORD

[GO SHOPPING](#)



ADLER  
PLANETARIUM

JOHNS HOPKINS  
UNIVERSITY

UNIVERSITY  
OF MINNESOTA



The University of  
Nottingham



UNIVERSITY OF  
OXFORD



ROYAL OBSERVATORY  
GREENWICH



# EXAMPLE: GALAXYZOO.ORG

- ~260,000 participants (and growing)
- ~1 million galaxies have been labeled (classified)
- ~180 million classifications have been collected



# EXAMPLE: OLDWEATHER.ORG

oldWeather

HOME VESSELS TUTORIAL TRANSCRIBE ABOUT BLOG FORUM GET STARTED

1. Follow vessels 2. Digitise pages 3. Get promoted

Project Statistics  
Old Weather transcriptions so far

**68 %** OF THE LOGS COMPLETED

517217 PAGES DONE  
118 SHIPS COMPLETE

Google

Old Weather: Our Weather's Past, the Climate's Future

The image is a screenshot of the Old Weather website. At the top, there's a navigation bar with links: HOME, VESSELS, TUTORIAL, TRANSCRIBE, ABOUT, BLOG, FORUM, and a prominent 'GET STARTED' button. Below the navigation bar, there are three icons representing the project's goals: '1. Follow vessels' (a sailboat), '2. Digitise pages' (a person at a computer), and '3. Get promoted' (a person at a ship's wheel). To the right, a 'Project Statistics' section displays '68 % OF THE LOGS COMPLETED' in large red text, with '517217 PAGES DONE' and '118 SHIPS COMPLETE' below it. In the center, a black and white photograph of a large steamship is shown, with a small red icon of a ship's wheel and a red line pointing to it. At the bottom, a large red banner contains the text 'Old Weather: Our Weather's Past, the Climate's Future'. A 'Google' logo is visible on the left, and a 'Terms of Use' link is on the right.

# EXAMPLE: OLDWEATHER

## Cadet hausmans

0 weather reports on 0 pages contributed to this voyage. **30 weather reports more** for promotion to **Lieutenant**

2. LOCATION

4. OTHER

5. FINISH

## HMS Teutonic

Active: Atlantic convoys

Armed Merchant Cruiser - [Learn more](#)

## Map and timeline

DATE  
//

POSITION  
lat:60.14N  
lng:5.27W





"It's black, and it looks like a hole.  
I'd say it's a black hole."

## PARTNERING WITH CITIZEN SCIENCE ALLIANCE

Lessons learned from astronomy applied to climatology.



# INTERNATIONAL CROWDSOURCING COLLABORATION

- Establishing a long-term partnership with Citizen Science Alliance (Zooniverse)
  - Effort led by scientists from the Cooperative Institute for Climate and Satellites in North Carolina (CICS-NC)
- In short-term, developing prototype capabilities
  - International exchange of scientific expertise and technology transfer
- Long-term goal of joining the CSA and developing climate crowdsourcing applications
- Two initial projects are proposed
  - Data Rescue for Surface Temperature Databank
  - Tropical Cyclone Reanalysis

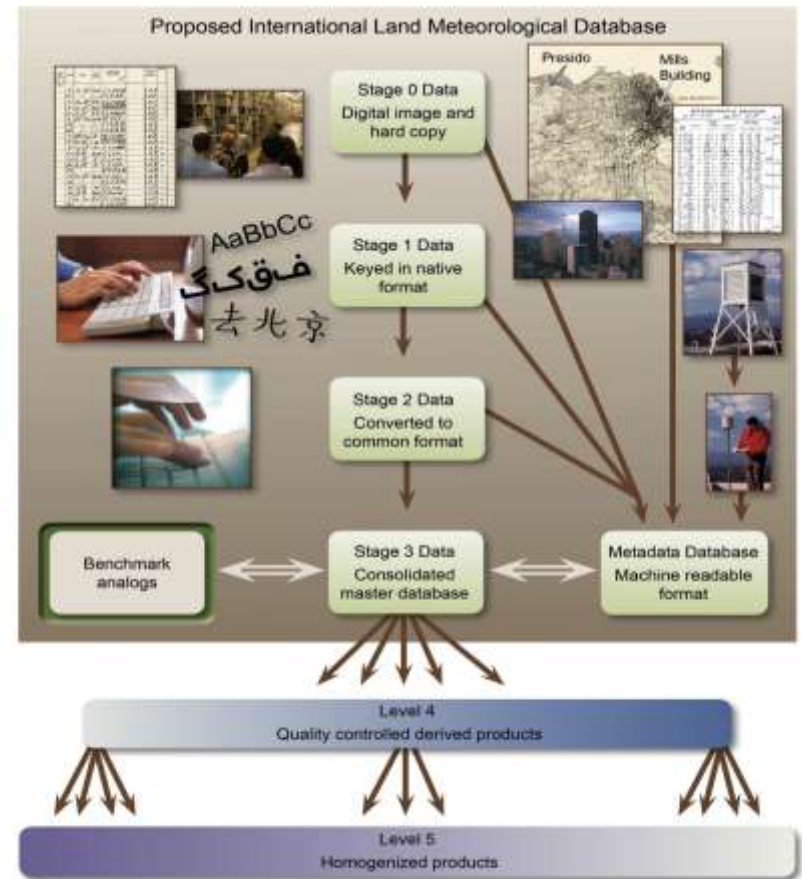


CITIZEN SCIENCE  ALLIANCE

# PROJECT 1: DATA RESCUE FOR SURFACE TEMPERATURE DATABANK

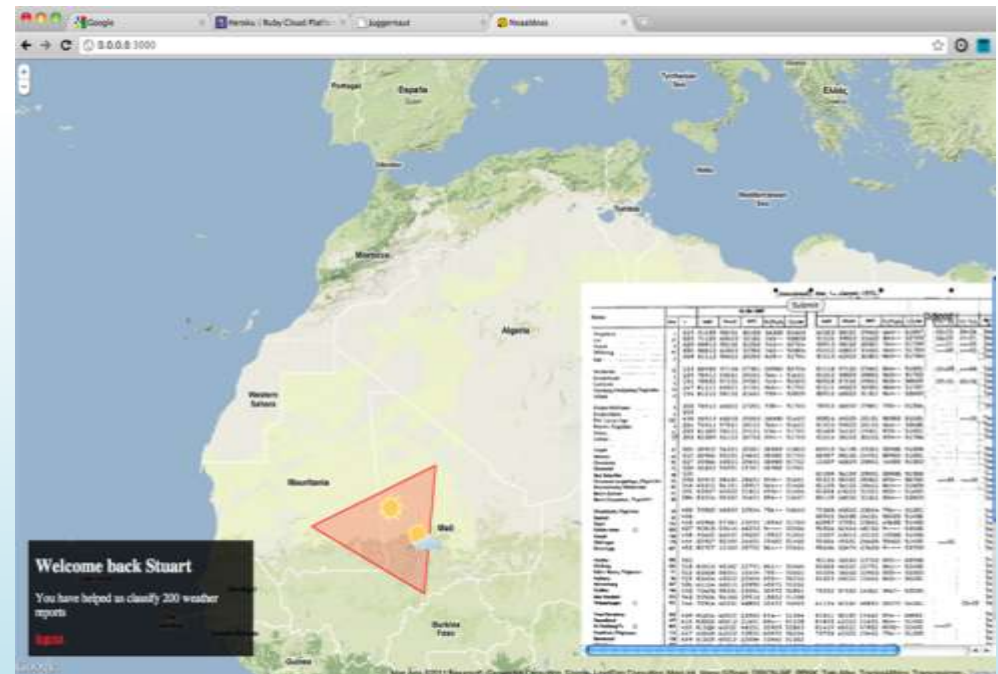
# GLOBAL DATABANK

- Society expects openness and transparency in the understanding of the (un)certainly on how climate has changed and how it will continue to change
- UK Meteorological Office (UKMO) proposed a new International Analysis of Land-Surface Air Temperature Data
  - Endorsed by World Meteorological Organization (WMO) Commission for Climatology, February 2010
- *International Surface Temperature Initiative*
  - Established at workshop in Exeter, UK in September, 2010
  - NOAA/NCDC is leading in establishing and potentially hosting the *data bank*
  - Data rescue is a priority of the initiative
- <http://www.surface-temperatures.org>



# DATA DIGITIZATION THROUGH CROWDSOURCING

- Proposed method similar to oldWeather.org
  - Multiple redundant keying of historical images
  - Similar to ReCAPTCHA, use human to refine OCR results
- Digitized data placed in The Databank
  - Retain the redundant values as valuable metadata
  - Full provenance and version tracking – anyone can go back to the raw data value at any time
- Finding the right motivation is challenge
  - Citizen scientist must understand the impact they're making

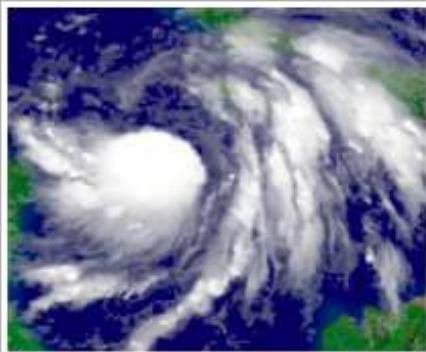




# PROJECT 2: TROPICAL CYCLONE REANALYSIS

# TROPICAL CYCLONE REANALYSIS

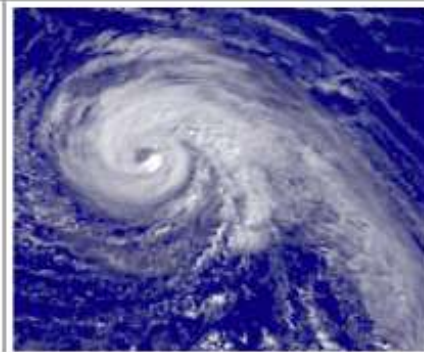
- Historical intensity records of tropical storms are based on regional methods, leading to basin-to-basin differences
  - Methods have also changed with time and even differ by forecaster
- Since the late-1970s we have archived satellite images
- Use the 'crowd' to create a consistently analyzed historical record across the globe.
- Dvorak (intensity) classification is well-suited because it can be easily translated to a GalaxyZoo-type analysis and is the world-wide standard



Tropical Storm Wilma at T3.0



Tropical Storm Dennis at T4.0



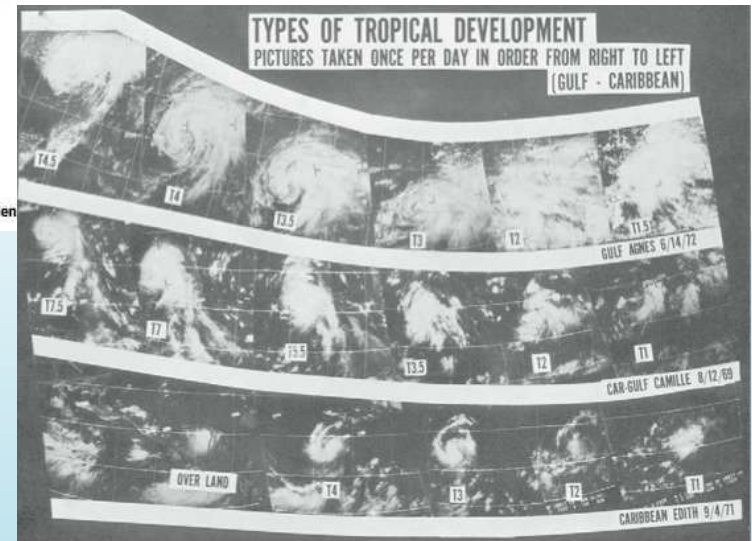
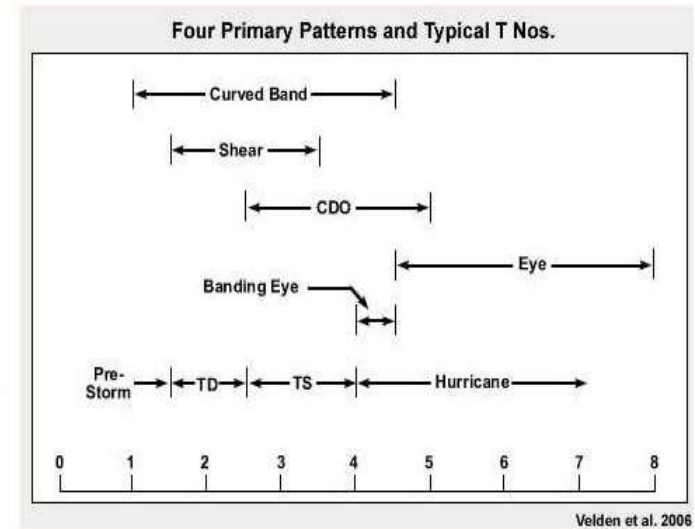
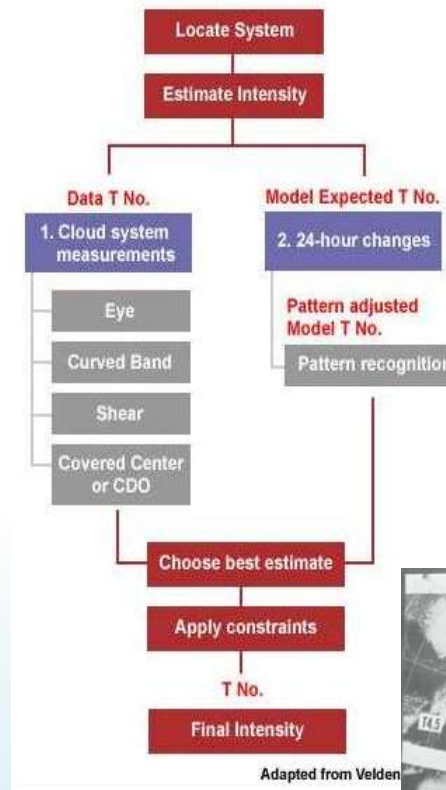
Hurricane Jeanne at T5.0



Hurricane Emily at T6.0

# DVORAK TECHNIQUE

- Subjective estimate of tropical cyclone (TC) intensity based solely on visible and infrared satellite images
- TCs of similar intensity tend to have certain characteristic features, and as they strengthen, they tend to change in appearance in a predictable manner
- A "T-number" and a Current Intensity (CI) value are assigned to the storm
  - 1 - minimum intensity
  - 8 - maximum intensity
- Developed in 1973 by Vernon Dvorak



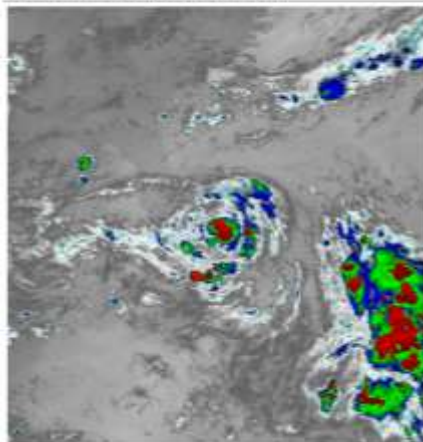
## Tropical Cyclone Reanalysis Prototype



*The Galaxy zoo community has provided over 250 million classifications through the galaxy zoo website. Hurricanes like galaxies are still best classified by humans, the citizen science community can produce an equally valuable dataset for meteorologists and climatologists to help them understand these extreme events*

## PROTOTYPE OF HURRICANE INTERFACE

### Storm classification



Is the bulk of the storm

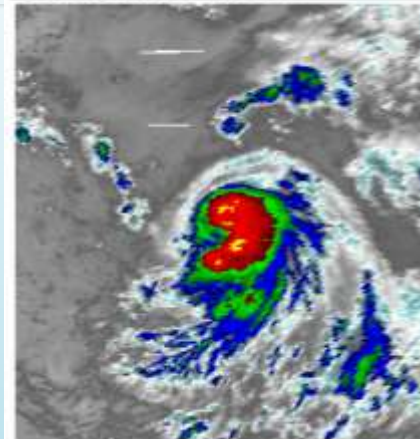
[around the eye](#)

[to the left of the eye](#)

[to the right of the eye](#)

History  
[Revised](#) [Revised](#) [Revised](#)

### Storm classification



Is the eye of the storm

[Obscured](#)

[Obscured](#)

[Partly Obscured](#)



# SUMMARY

---

- Critical environmental data has been digitally “rescued” for long-term preservation, but essentially remains “lost” to scientific inquiry
- Croud sourcing offers tremendous potential to not only leverage online “cranial capacity”, but also to engage the public in science
- NOAA is collaborating with Citizen Science Alliance, which is leading the way in crowdsourcing, to fill climate data voids and enable climate science

# Scott Hausman


**Deputy Director**

**NOAA's National Climatic Data Center (NCDC)**

151 Patton Avenue, Room 557

Asheville, NC 28807-5002

 828-271-4848

 828-271-4246

 828-450-9188

[Scott.Hausman@noaa.gov](mailto:Scott.Hausman@noaa.gov)

[www.ncdc.noaa.gov](http://www.ncdc.noaa.gov)

