Engaging the Public in Climate Science

Exploiting Crowdsourcing to Digitize and Analyze Climate Data

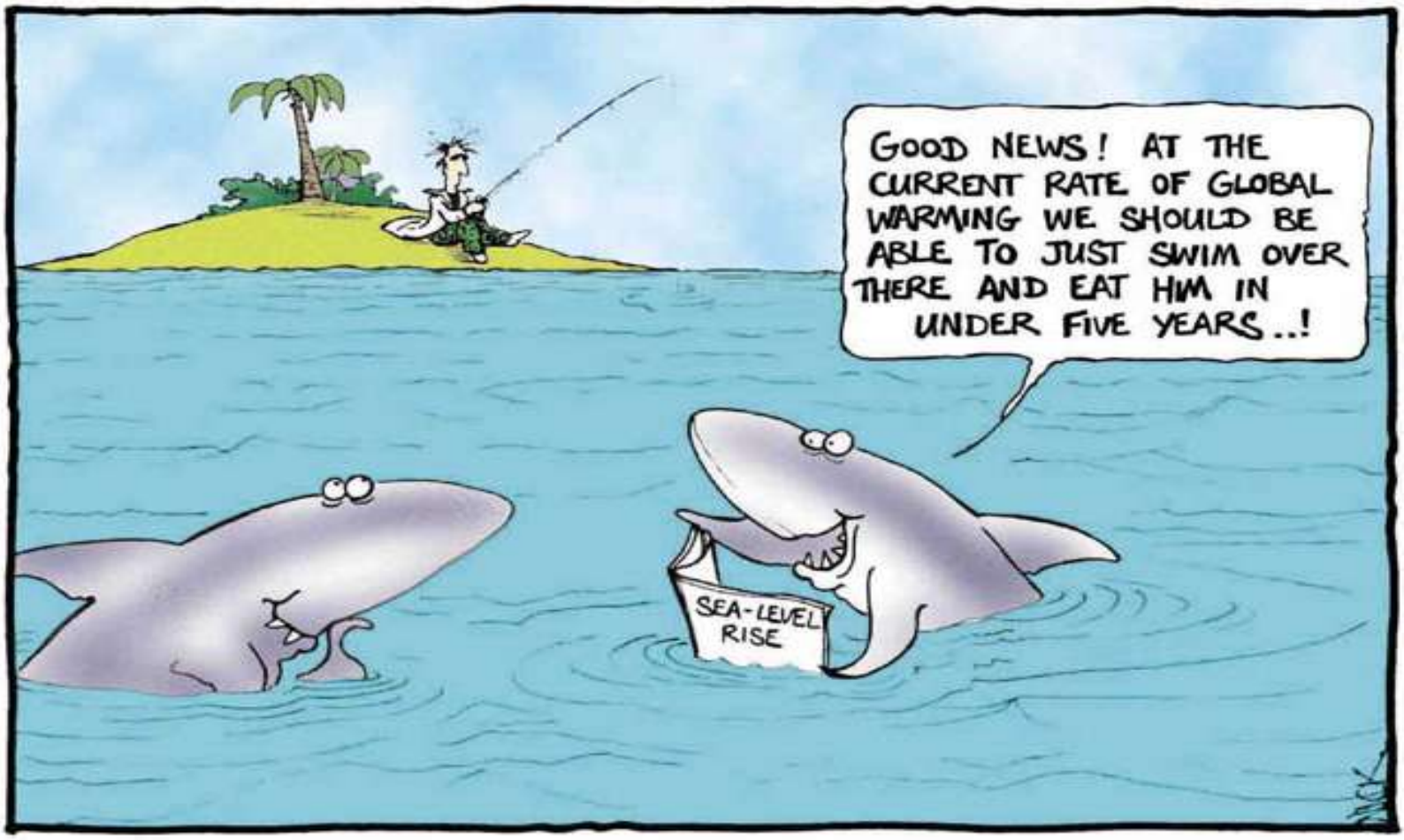
Scott Hausman
Deputy Director
National Climatic Data Center

13 June 2011

Board on Research Data & Information (BRDI)

Outline

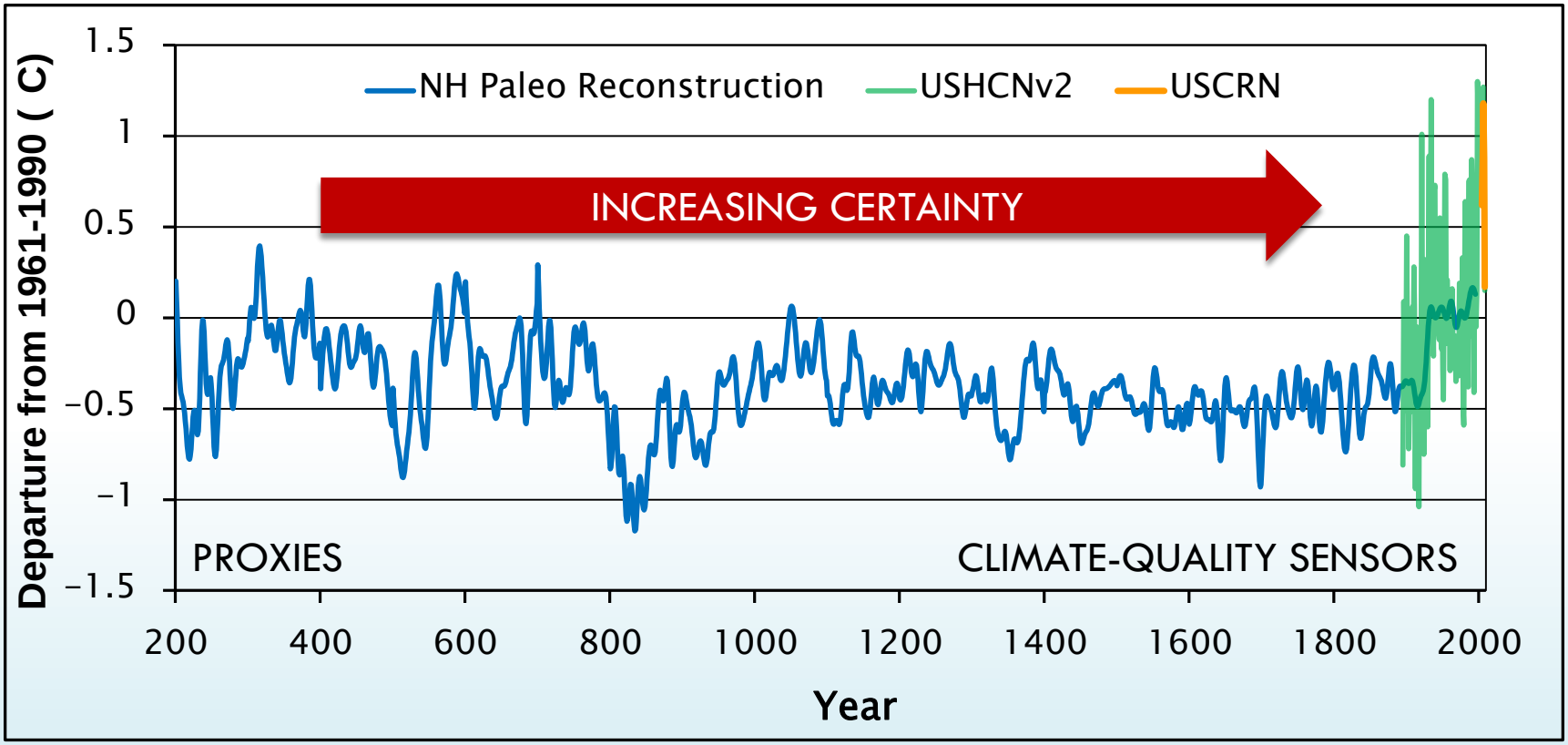
- Challenges of Monitoring Climate Change
- Recruiting Internet Citizen Scientists
- Partnering with Citizen Science Alliance
  - Project 1: Data Rescue for Surface Temperature Databank
  - Project 2: Tropical Cyclone Reanalysis



Challenges of Monitoring Climate Change

Reducing Uncertainty to Improve Public Understanding

Observation Quality and Uncertainty



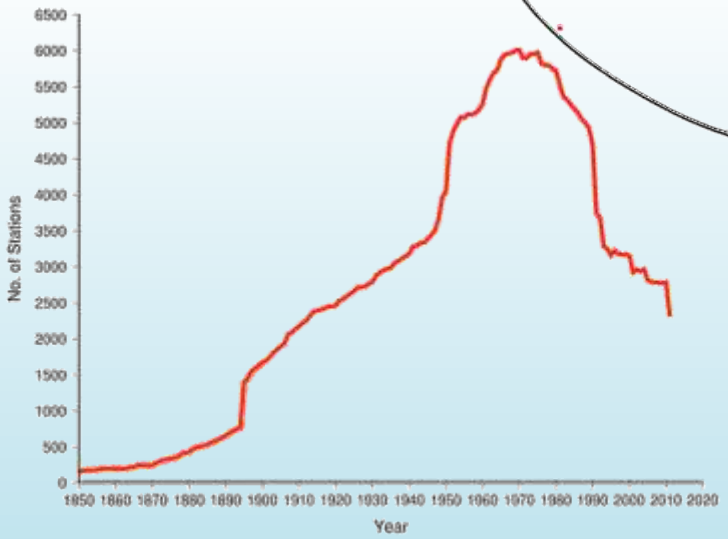
Three temperature time series are displayed relative to the 1961-1990 normals (°C). The paleoclimate reconstruction from Mann et al. (2008) is for Northern Hemisphere land, while U.S. Historical Climatology Network v. 2 dataset from Menne et al. (2009) and U.S. Climate Reference Network temperature departures 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

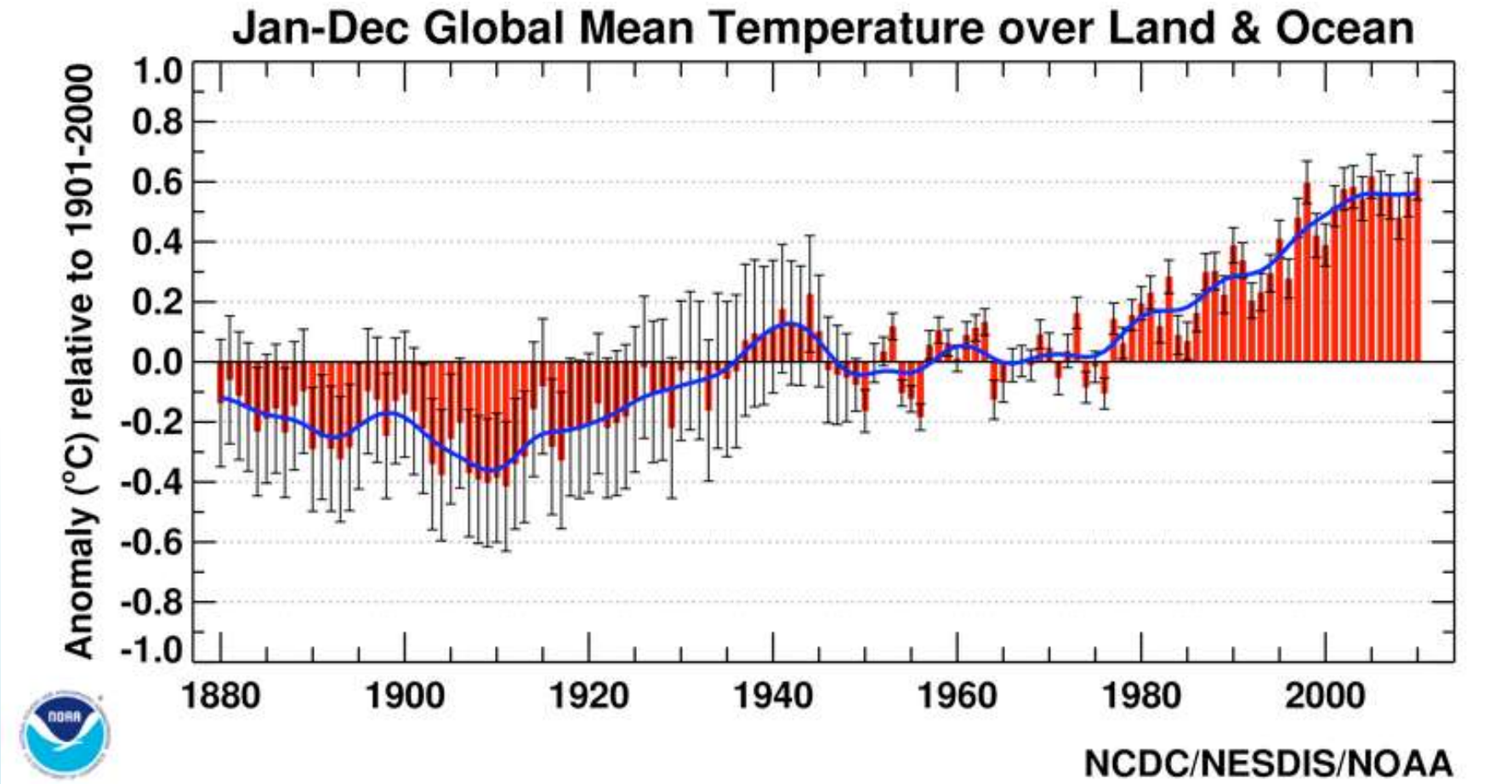
Global Historical Climatology Network (Monthly)

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.

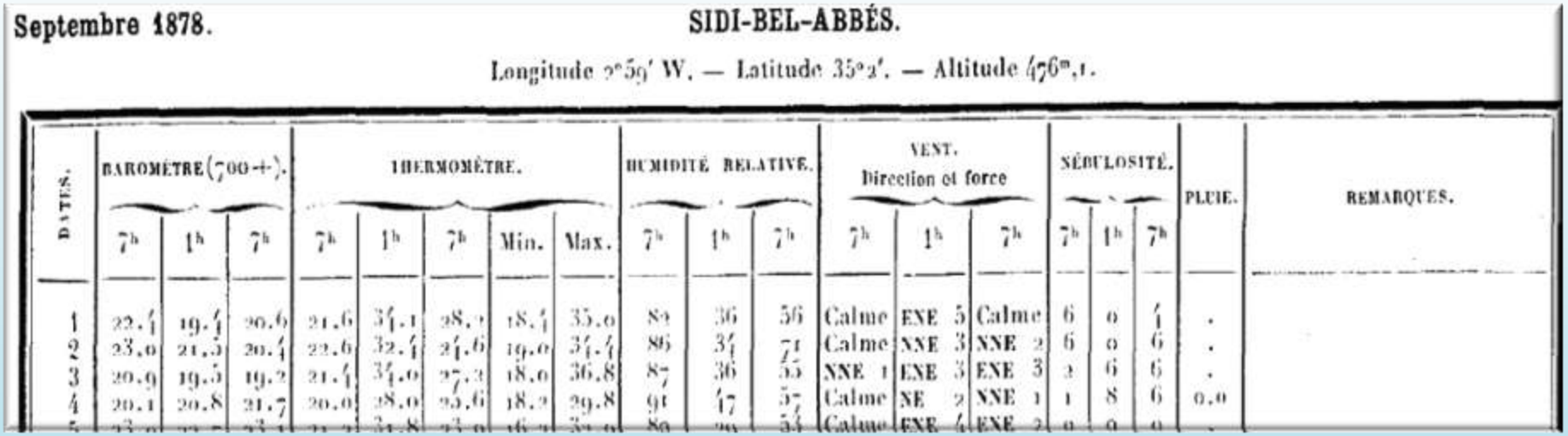
Reducing Uncertainty in Global Trends



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

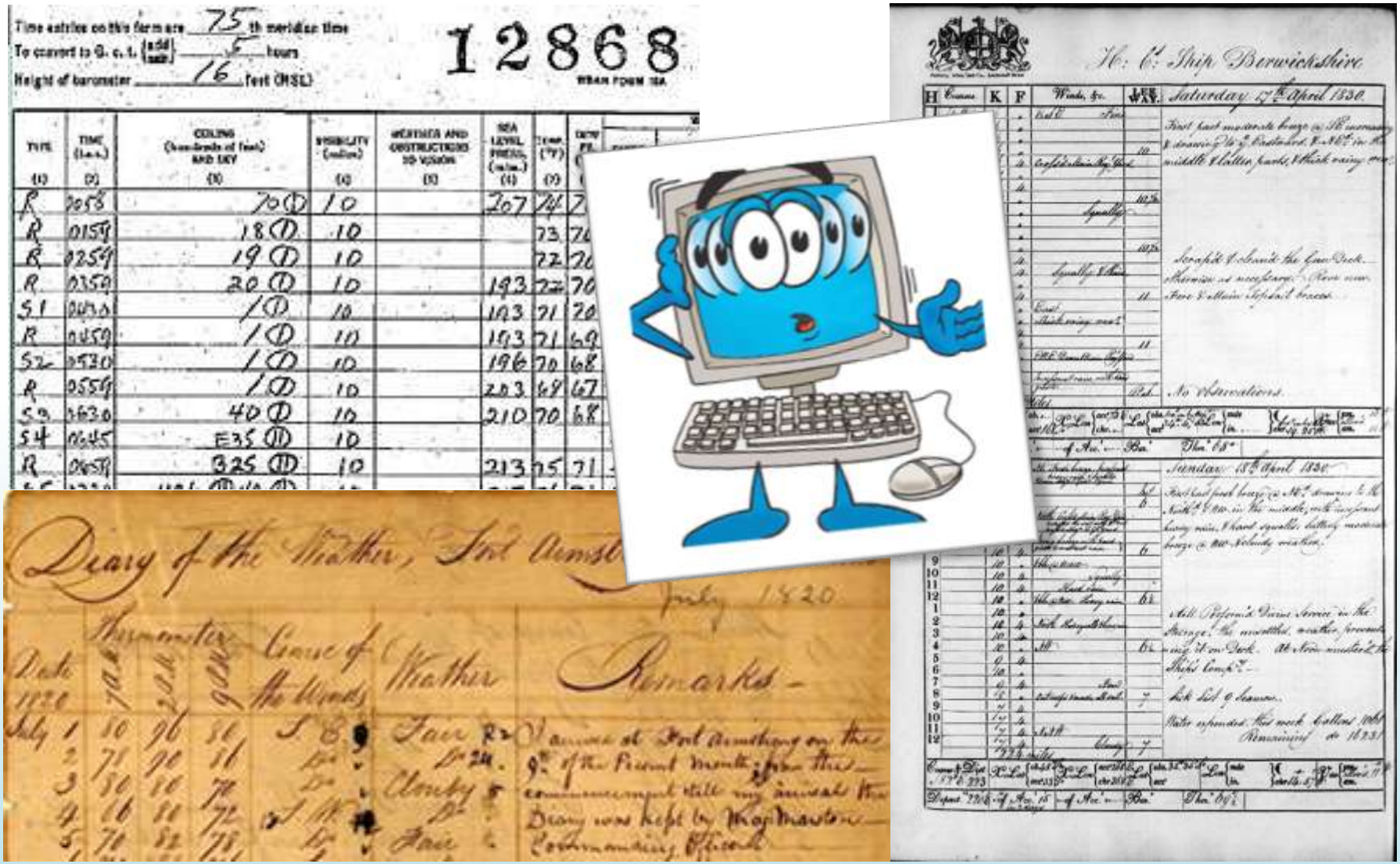


Septembre 1878. SIDI-BEL-ABBÉS.

Longitude 0°59′ W. — Latitude 35°2′. — Altitude 476m,1.

DATES.	BAROMÈTRE (700+).			THERMOMÈTRE.					HUMIDITÉ RELATIVE.			VENT. Direction et force			NÉBULOSITÉ.			PLUIE.	REMARQUES.
	7h	1h	7h	7h	1h	7h	Min.	Max.	7h	1h	7h	7h	1h	7h	7h	1h	7h		
1	22.1	19.1	20.6	21.6	31.1	28.0	18.1	35.0	89	36	56	Calme	ENE 5	Calme	6	0	1	.	
2	23.0	21.3	20.1	22.6	32.1	21.6	19.0	34.1	86	31	71	Calme	NNE 3	NNE 2	6	0	6	.	
3	20.9	19.5	19.2	21.1	34.0	27.3	18.6	36.8	87	36	55	NNE 1	ENE 3	ENE 3	2	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	22.2	23.1	21.2	31.8	23.0	16.1	32.0	89	39	54	Calme	ENE 4	ENE 2	0	0	0	.	

Limits to Optical Character Recognition (OCR)



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

Recruiting Internet Citizen Scientists

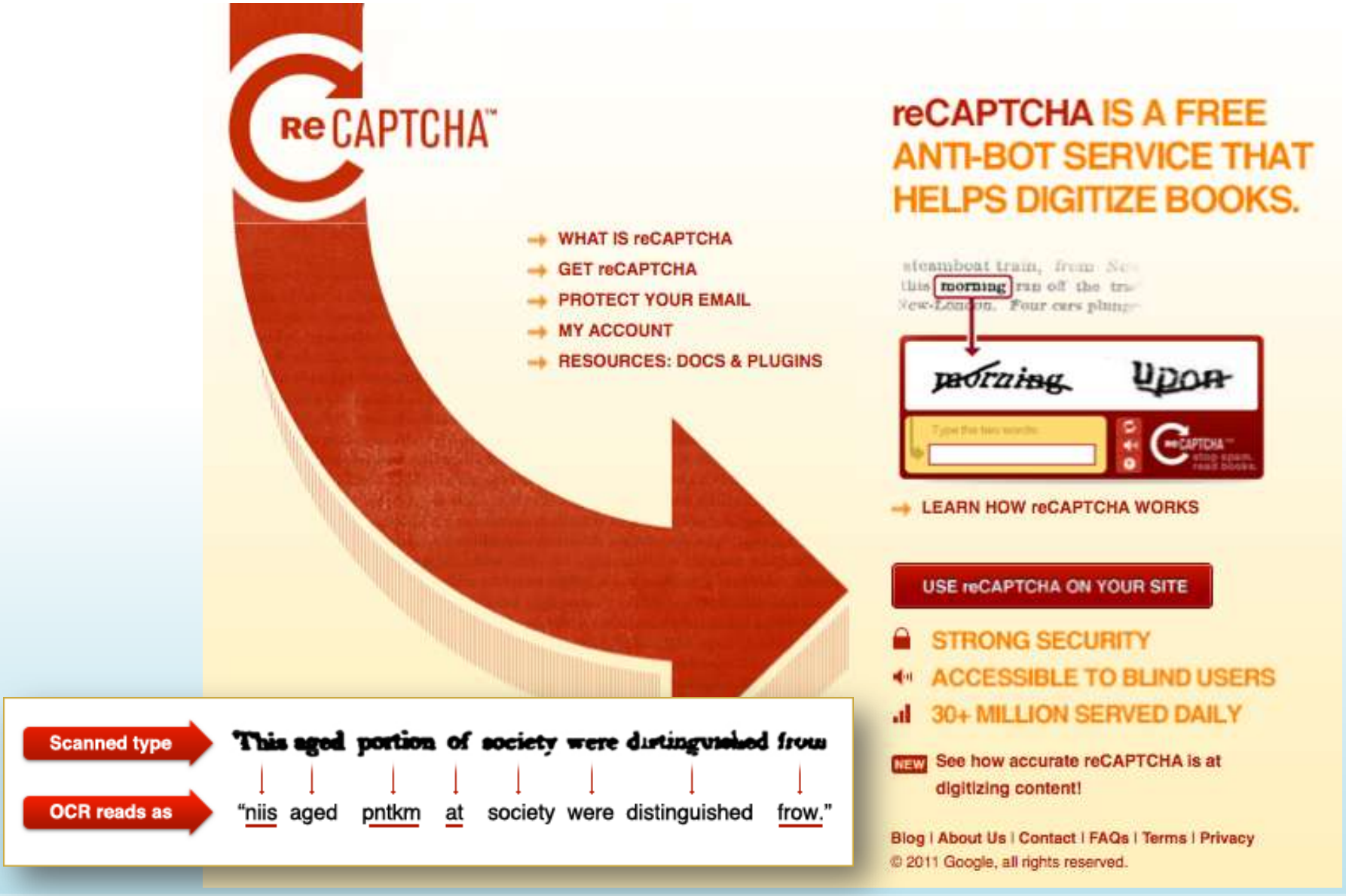
We need carbon-based computing for imagery analysis.

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



EXAMPLE: ZOONIVERSE

CITIZEN SCIENCE ALLIANCE

ZOONIVERSE
REAL SCIENCE ONLINE

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 a
efforts and ability of volunteers to he
data that confronts them.

The Zooniverse began with a single
2007. The Galaxy Zoo team had expe
overawed by the response to the p
buckling under the strain, they set ab

Galaxy Zoo was important because
many unique scientific results, rang
those using classifications that deper
This commitment to producing real r
your time - is at the heart of everythi

Zooniverse Activity

Total Volunteers: 430,804

Recent Zooniverse activity

Live Projects

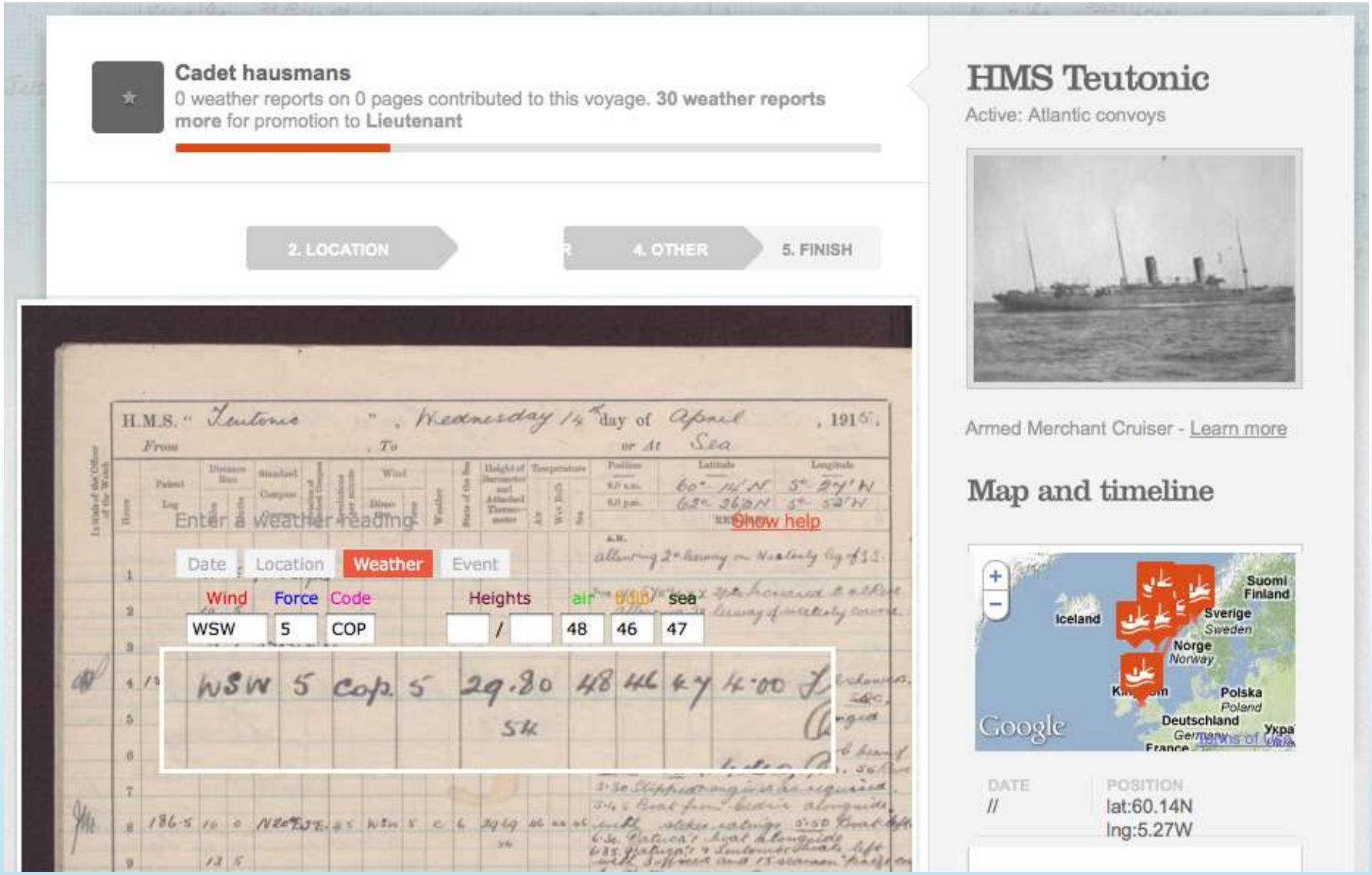
ADLER PLANETARIUM | JOHNS HOPKINS UNIVERSITY | UNIVERSITY OF MINNESOTA | The University of Nottingham | UNIVERSITY OF OXFORD | ROYAL OBSERVATORY GREENWICH

EXAMPLE: GALAZXYZOO.ORG

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

EXAMPLE: OLDWEATHER.ORG

EXAMPLE: OLDWEATHER



"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

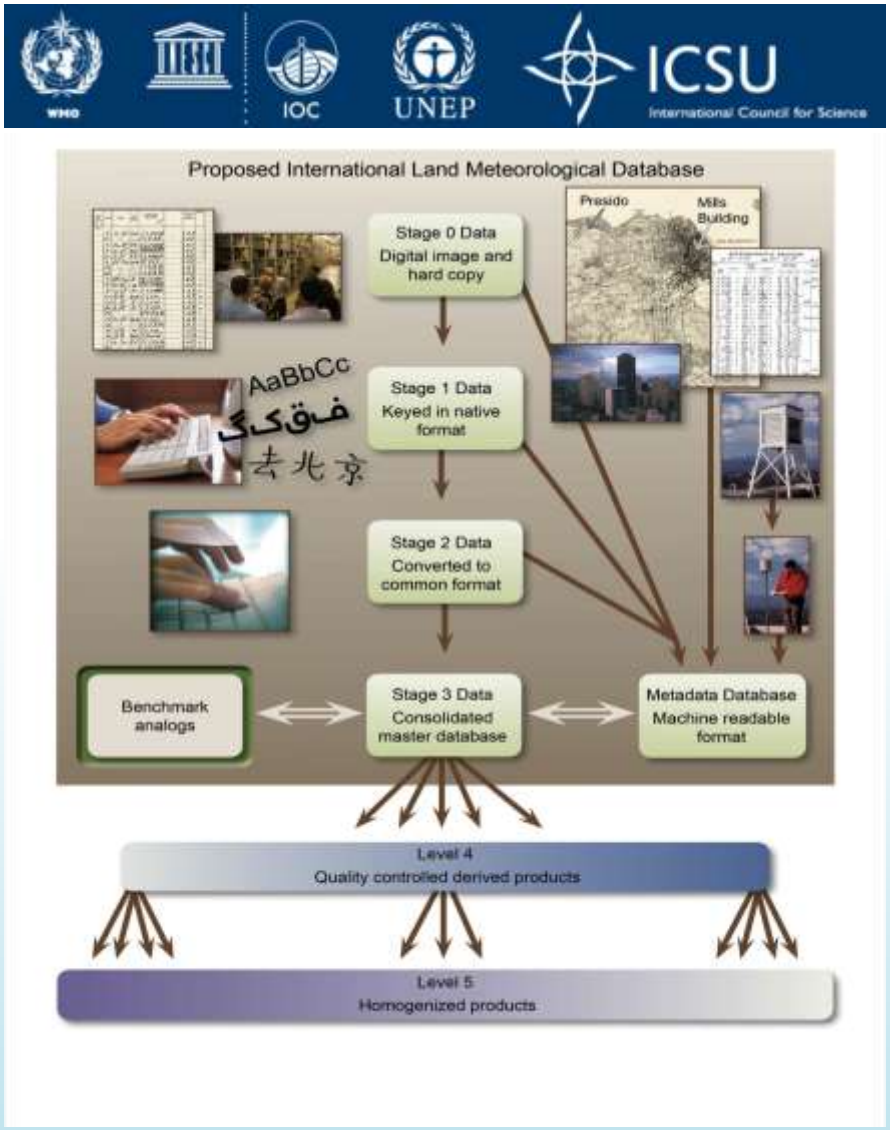
- Establishing a long-term partnership with Citizen Science Alliance (Zooniverse)
  - Effort led by scientists from the Cooperative Institute for Climate and Satellites in North Carolina (CICS-NC)
- In short-term, developing prototype capabilities
  - International exchange of scientific expertise and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

CITIZEN SCIENCE ALLIANCE

Project 1: Data Rescue for Surface Temperature Databank

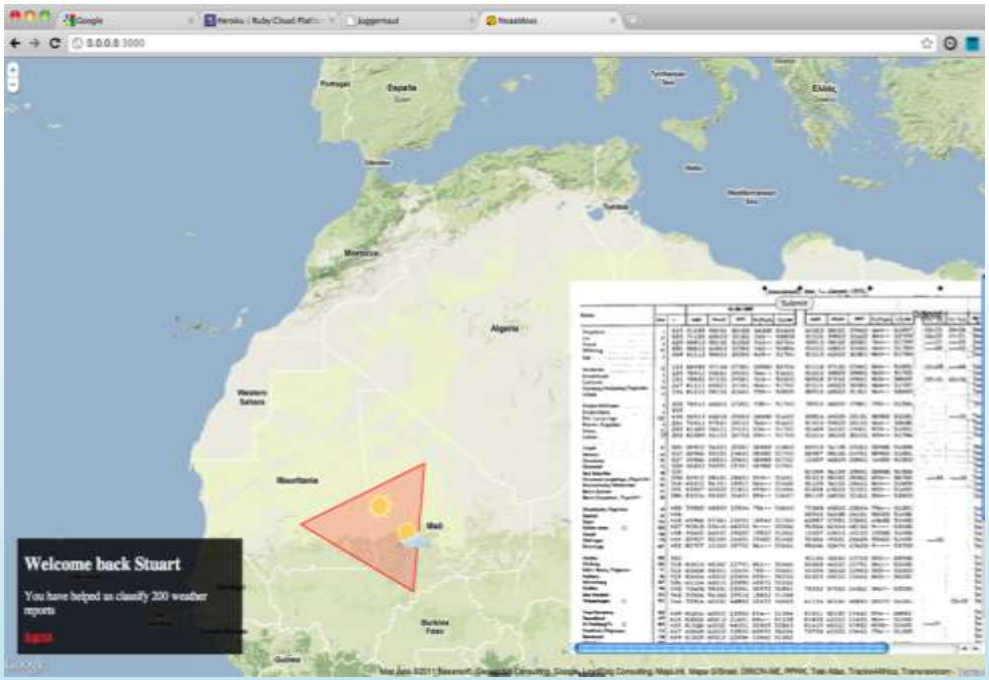
GLOBAL DATABANK

- Society expects openness and transparency in the understanding of the (un)certainty 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.surfacetempera-tures.org



Data Digitization through Crowdsourcing

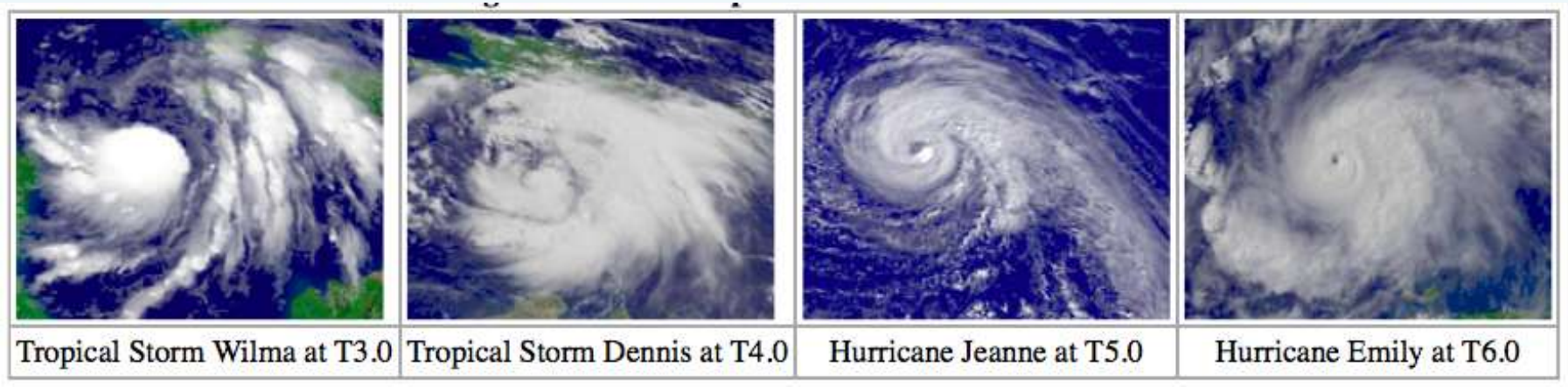
- Proposed method similar to oldWeather.org
  - Multiple redundant keying of historical images
  - Similar to ReCAPTCHA, use human to refine OCR results
- Digitized data placed in The Databank
  - Retain the redundant values as valuable metadata
  - Full provenance and version tracking – anyone can go 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

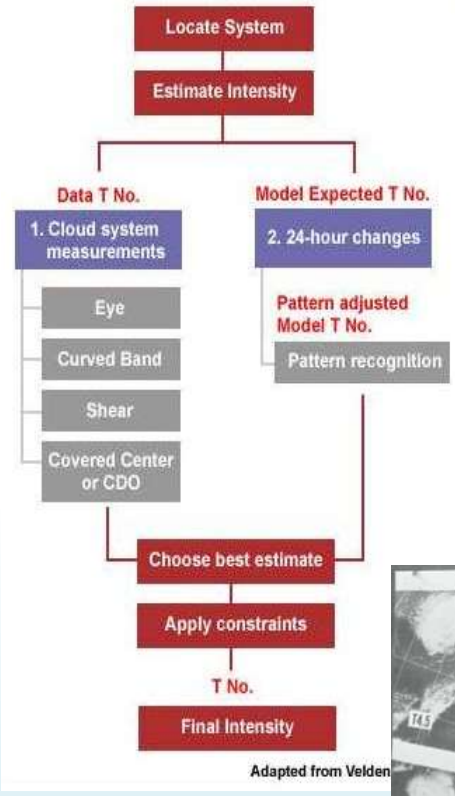
Tropical Cyclone Reanalysis

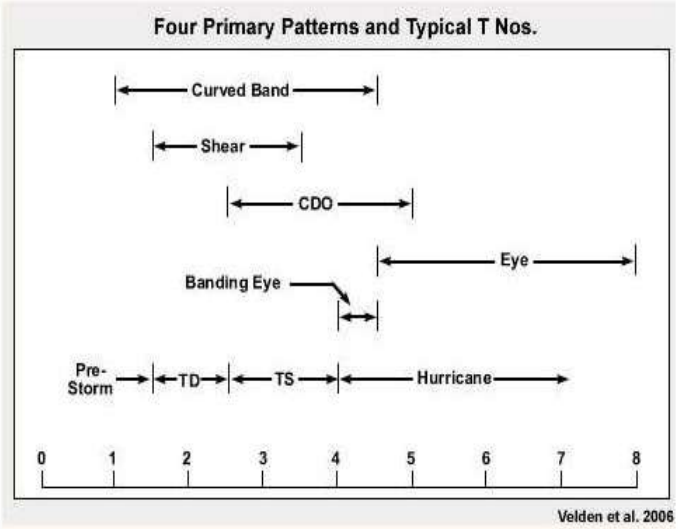
- Historical intensity records of tropical storms are based on regional methods, leading to basin-to-basin differences
  - Methods have also changed with time and even differ by forecaster
- Since the late-1970s we have archived satellite images
- Use the 'crowd' to create a consistently analyzed historical 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

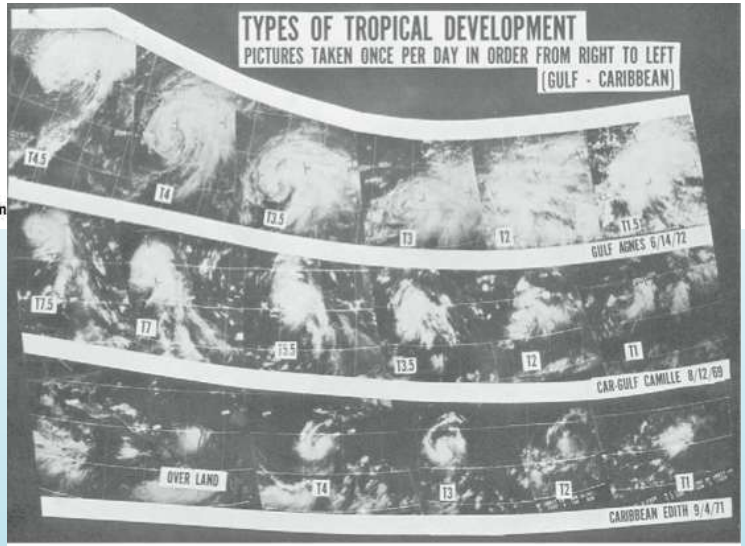


DVORAK TECHNIQUE

- 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





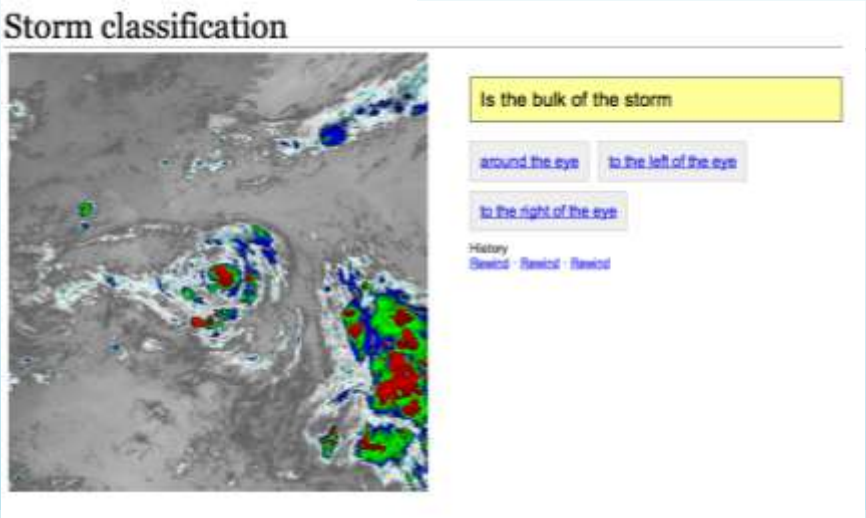


HURRICANEZOO.ORG

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



Summary

- Critical environmental data has been digitally "rescued" for long-term preservation, but essentially remains "lost" to scientific inquiry
- Croudsourcing 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

www.ncdc.noaa.gov