

# Frontiers in Data Collection, Storage, and Sharing

*Ruth Duerr, Ronin Institute for  
Independent Scholarship*

# Outline

---

- A bit about Ronin and I
- Improvements in
  - Data collection
  - Storage
- Challenges
  - Quality
  - Sharing
    - Access
    - Availability
  - Security



# Ronin Institute Principles

---

The Ronin Institute was founded to facilitate and support high-quality academic scholarship by independent scholars and anyone interested in a non-traditional academic career.

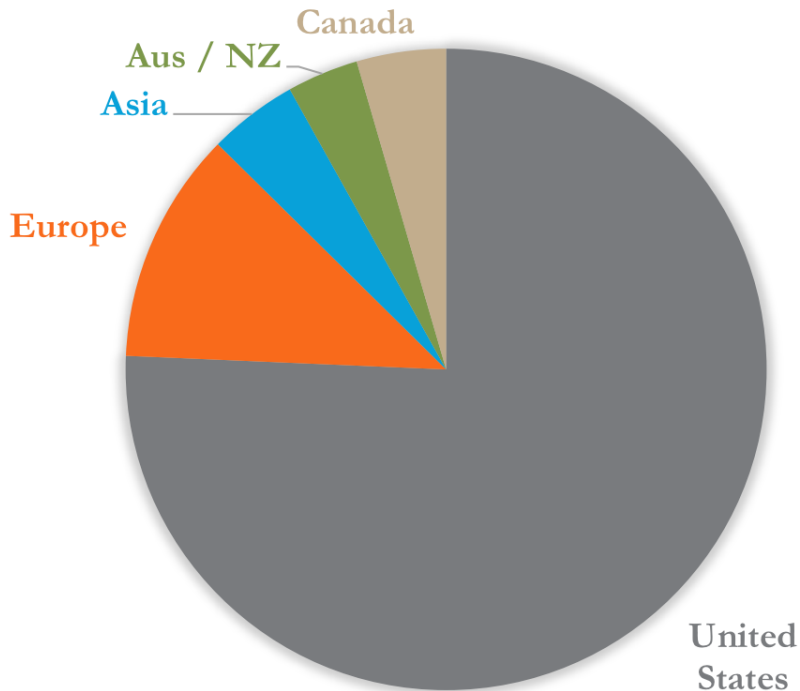
Research Scholar is the only academic rank at the Ronin Institute.

The motto, *scientiam consecemus*, is latin for “Let’s chop up some knowledge”

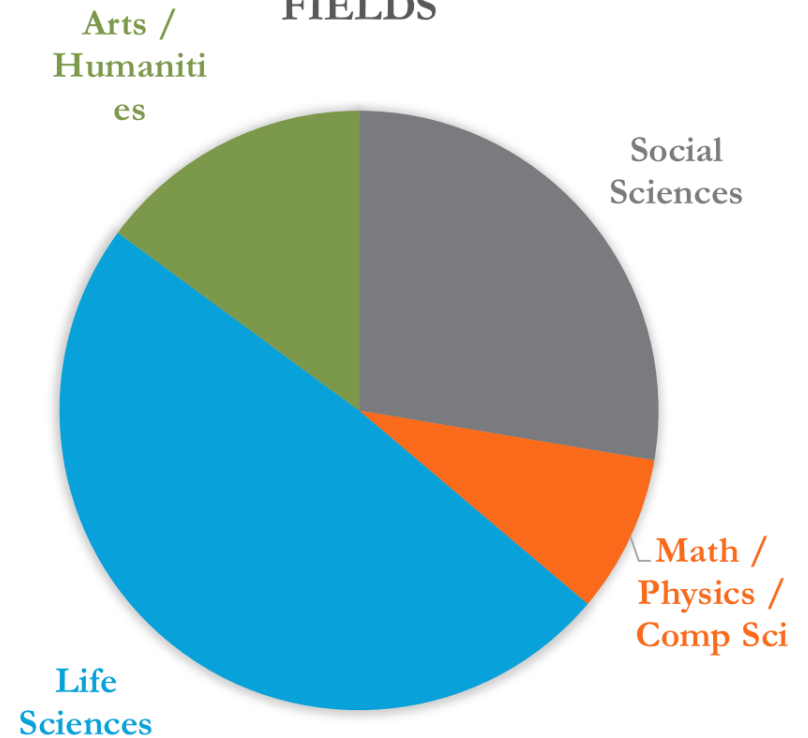
# Ronin Institute Statistics

- The Ronin Institute currently has over 100 Research Scholars representing fields from Philosophy to History to Biology to Physics.

RESEARCH SCHOLAR LOCATIONS



RESEARCH SCHOLAR PRIMARY FIELDS



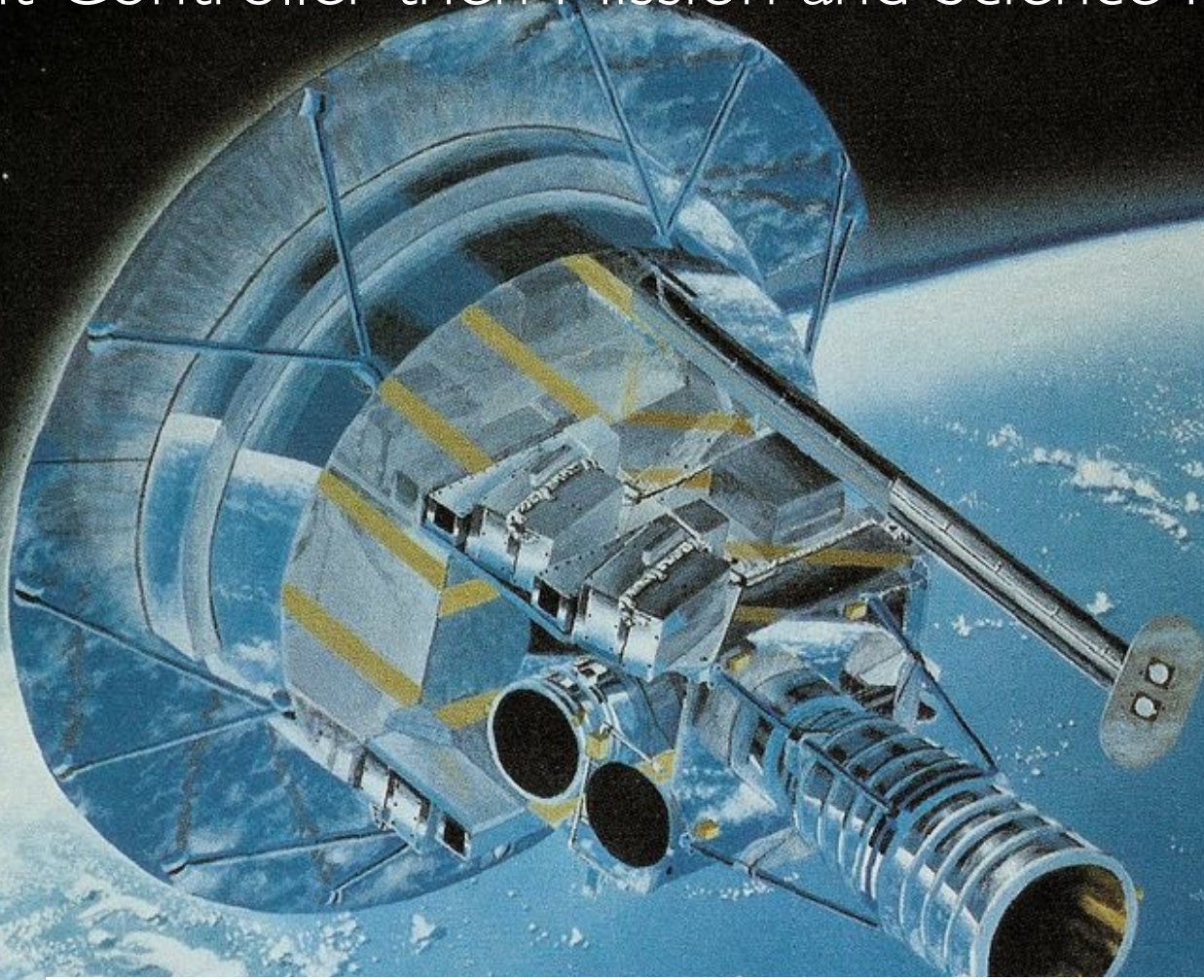
# Observing on Kitt Peak





# Solar Mesosphere Explorer

## Flight Controller then Mission and Science Planner



### Scientific Objectives of the Solar Mesosphere Explorer Mission

*Abstract* – The 1981–82 Solar Mesosphere Explorer (SME) mission is described. The SME experiment will provide a comprehensive study of mesospheric ozone and the processes which form and destroy it. Five instruments will be carried on the spinning spacecraft to measure the



# UARS-SOLSTICE

Solar Stellar Irradiance Comparison Experiment

## U A R S

Upper Atmosphere Research Satellite

*Understanding Our Atmosphere*





# The Alaska SAR Facility, Fairbanks



# The Alaska Satellite Facility, Fairbanks - now





# NSIDC affiliations and sponsorship

Cooperative Institute  
for Research in  
Environmental  
Sciences



World Data  
System



Main sponsors:



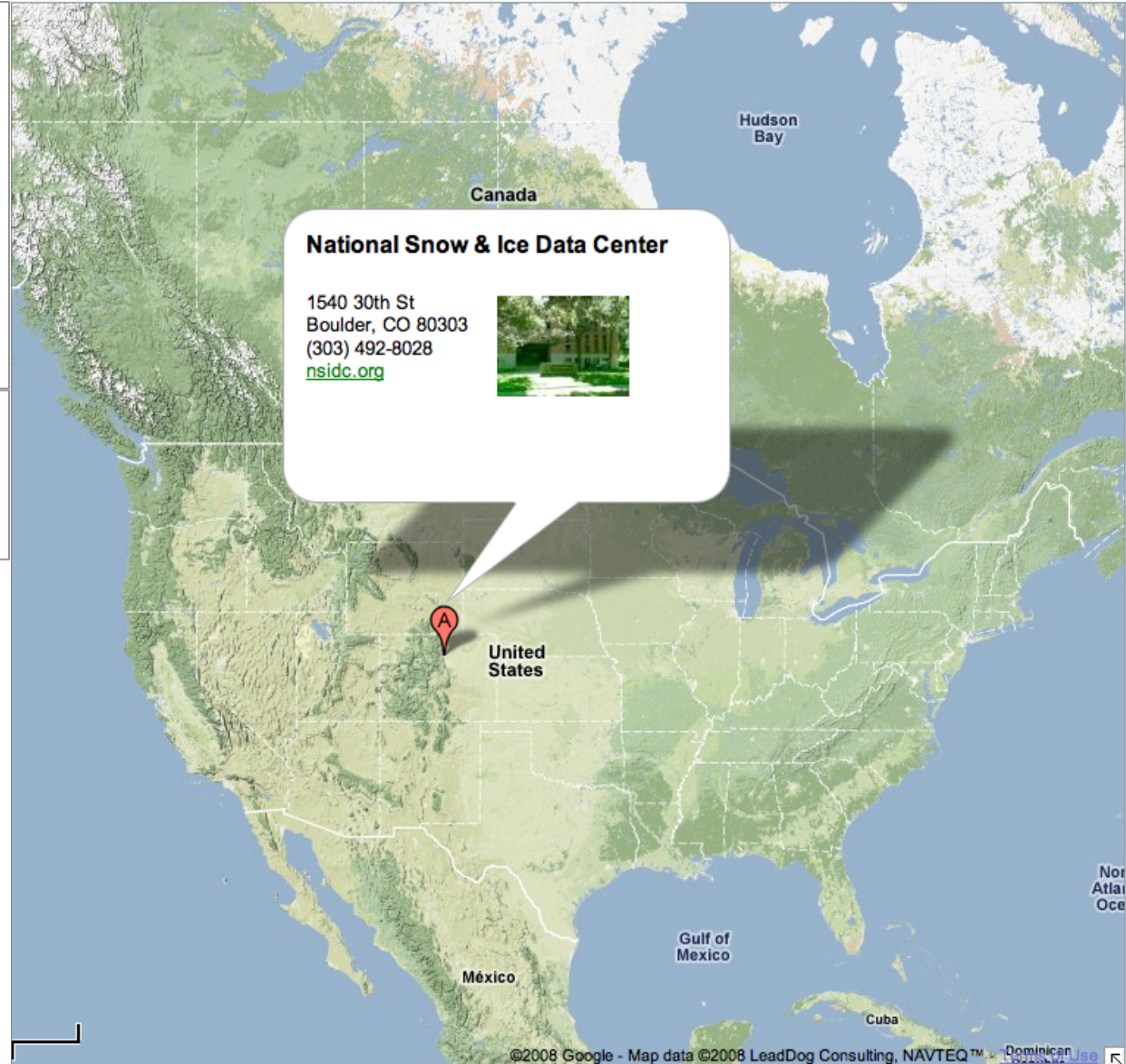
NASA



National Science  
Foundation



National Oceanographic  
and Atmospheric  
Administration

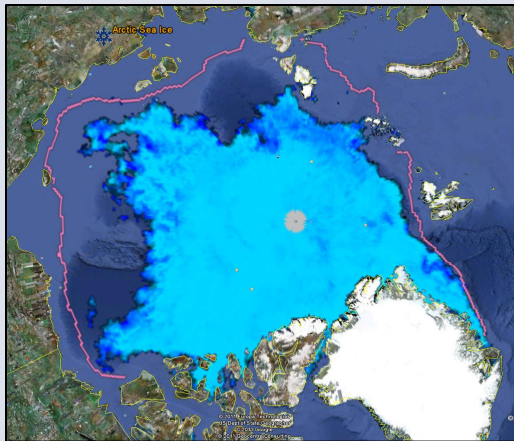


# NSIDC Distributed Active Archive Center

*Data from NASA's past and current Earth Observing System (EOS) satellites and other satellite and field measurement programs.*

## Passive Microwave

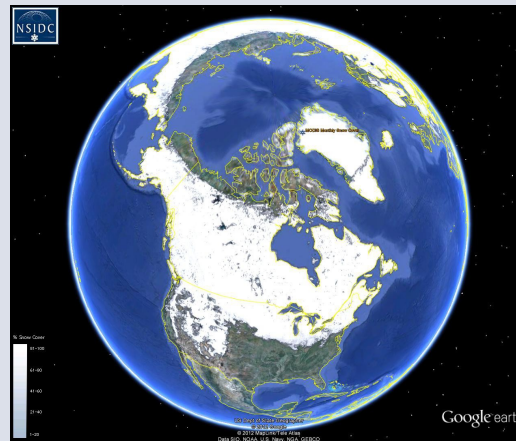
- AMSR-E (Aqua)
- AMSR (ADEOS II)
- SMMR (Nimbus 7)
- SSM/I, SSMIS (DMSP series)



AMSR-E 12.5 km Sea Ice Concentration

## VIS/IR Moderate Resolution

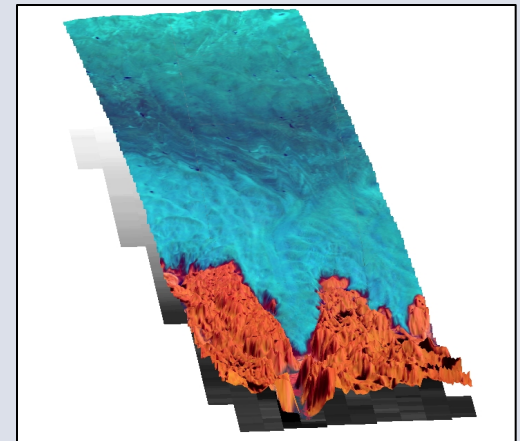
- MODIS (Terra/Aqua) snow and ice products
- AVHRR polar data (NOAA series)



MODIS Monthly Global Snow Cover

## Satellite & Airborne Altimetry

- IceSAT I/GLAS altimetry and atmospheric lidar data
- Digital Elevation Models (DEMs)
- IceBridge



IceBridge ATM Qfit Data on Landsat Image







## Roger G. Barry Archives and Resource Center



### Why

- An essential resource for understanding past climate and science
- Manage physical materials for present and future researchers
- Make archived materials discoverable and accessible



### What

- 100,000+ glacier photographs, from the 1800s to the present
- 6,800 sea ice charts from the pre-satellite era
- Records and data from field expeditions from the 1850s to the 1970s
- Rare books on polar research and observations



# ELOKA - Mission Statement



To provide data management and support services to facilitate the collection, preservation, exchange, and use of local [Indigenous] observations and knowledge of the Arctic.



<http://eloka-arctic.org/>

# ELOKA - Philosophy



Local and Indigenous knowledge and scientific expertise are complementary and reinforcing ways of understanding the Arctic system. Collecting, documenting, preserving, and sharing knowledge is a cooperative endeavor, and ELOKA is dedicated to fostering understanding and shared knowledge between northern communities and community members, scientists, educators, policy makers, and the general public. ELOKA operates on the principle that all knowledge should be treated ethically, and intellectual property rights should be respected.



# Building infrastructure for data curation

## PRESERVE

Collect and take care of research data.

## SHARE

Reveal data's potential across many disciplines.

## DISCOVER

Promote re-use and new combinations of data.

Key components of Data Conservancy's software:

SOFTWARE NOW AVAILABLE!



DC PACKAGING SPECIFICATION



PACKAGE INGEST SERVICE



LINKED DATA GRAPHS



API EXTENSION ARCHITECTURE

# *Data Collection & Storage Improvements*

---

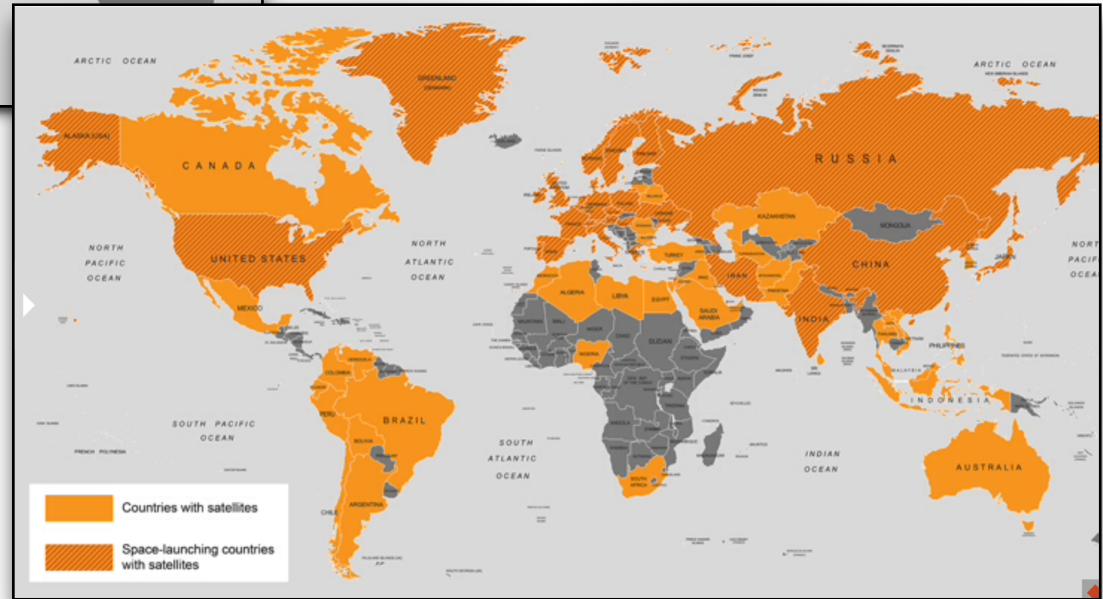
- Access to space
- Miniaturization of components
- Rapid development processes
- Openness and transparency
- Internet, Social Media, and the Cloud



# *Improvements - Access to Space*



Courtesy Union of Concerned Scientists





# *Improvements - Miniaturization of Components*

---



- Microphones
- Speakers
- Cameras
- GPS
- Wireless
- Cell phone
- Motion sensors
- Accelerometers
- Clocks
- Computers

# Improvements - Rapid Development Processes

- Heritage from W. Edward Deming and his 14 points for management?

## Agile Manifesto Values

- Individuals and interactions over processes and tools
- Working software over comprehensive documentation
- Customer collaboration over contract negotiation
- Responding to change over following a plan



## Lean and Agile Adoption with the Laloux Culture Model

from **Agile For All** PRO 2 years ago | more

# *Improvements - Openness and Transparency*

---

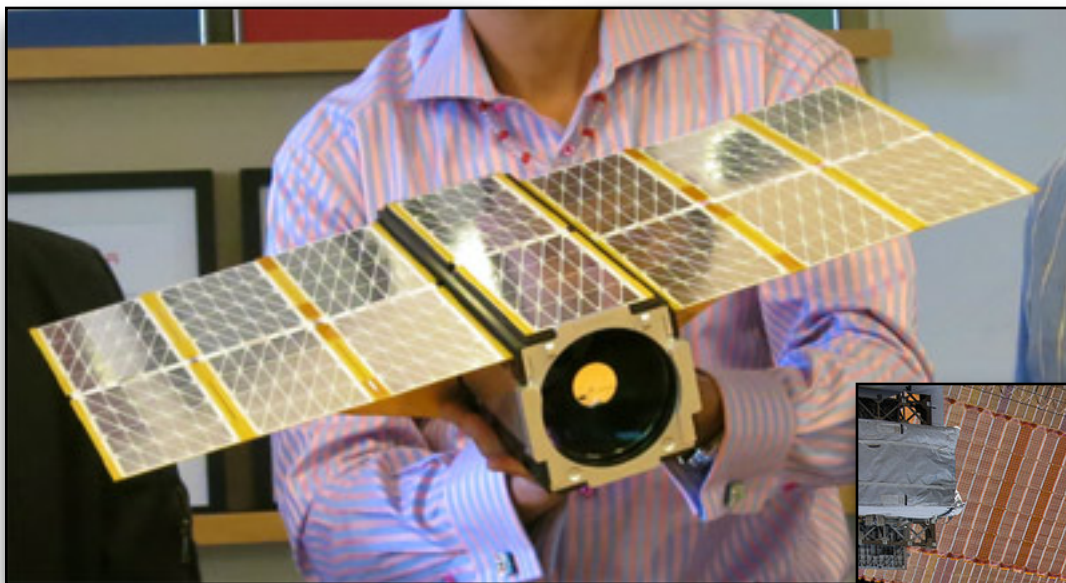
- Science started as an open and transparent enterprise; but lost its way with the advent of digital technology
- Increasing calls for it to return to that norm
  - Ethos of science requires it
  - Current government mandate for publicly funded data to be available
  - Data and software citation guidelines
  - Journals rejecting papers if the data and software aren't available
- Increasing calls for other areas to adopt these same principles (e.g., open source software, open government)

# *Improvements - Internet, Social Media, and the Cloud*

---







Copyright Steve Jurvetson

**SEE CHANGE. CHANGE THE  
WORLD.**

Founded in 2010 by a team of ex-NASA scientists, Planet is driven by a mission to image the entire Earth every day, and make global change visible, accessible, and actionable.

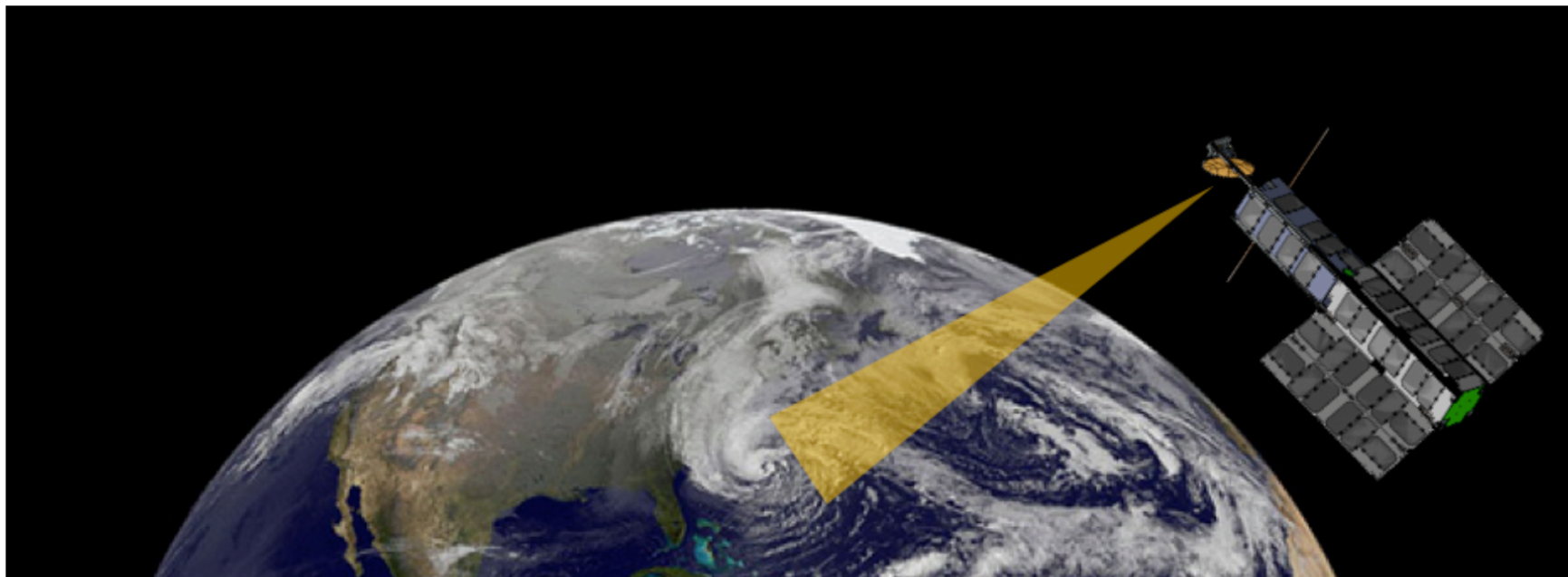
## USING SPACE TO HELP LIFE ON EARTH

Making global change visible,  
accessible, and actionable.






## PolarCube: An Advanced Radiometer 3U CubeSat



### PolarCube Mission Statement

PolarCube will perform tropospheric temperature profiling to complement current weather models applicable to severe mesoscale weather phenomena. The technical applications for PolarCube include; Mesoscale weather research, storm cell observations, hurricane warm core behavior, terrestrial weather detection with 15-20km spatial resolution, and study of sea ice fractions in the Polar Regions. The secondary focus is an educational goal to integrate and fly a passive radiometer while working closely with science and engineering experts.

 Open Science Framework

BrowseSupportQ

Sign UpSign In

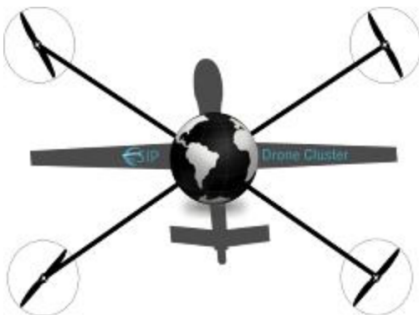
ESIP DronesFilesWikiAnalyticsRegistrationsForks

Home

Getting started

Why Drones for Science

+ Component Wiki Pages



**Welcome to the ESIP Drone Cluster.** We are an open community pursuing the use of unmanned aerial systems in Earth Science.

**Getting in contact:** [The ESIP Drone CLuster wiki](#), contains details on our monthly telecons, how to contact us, and biannual ESIP meeting activities.

**Who we are:** We are working on several ongoing projects including, developing resource catalogs, testing hardware and software, and a range of collaborations seeking to make drones easy to use, accessible, and reliable tools that enables innovative Earth Science to happen everywhere. If you're curious about why see our "[Why drones for science](#)" page, and for suggestions on getting started see our "[getting started](#)" and "[best practices](#)" pages.

OSF project resources:

- [Drone Resource Catalogs](#)
- [Case Study: Hacker drone for CO2 monitoring \(Funding Friday project\)](#)
- [ESIP developed Software for Science Drones](#)

## Drones in Science

---

**Purpose:** This wiki attempts to keep track of known science projects utilising drones. The intention is simply to help facilitate collaboration, spark research ideas, and help us keep track of possible needs and requirements as we seek to enable the use of this nascent technology.

- **Cryosphere Sciences**
- **Volcanology & Geochemistry**
- **Atmospheric Science**
- **Earth Surface Processes**
- **Hydrology**
- **Natural Hazards**
- **Biogeosciences** (Wildlife Management, Vegetation Monitoring)
- **Ocean Sciences**
- **Human Dimensions** (Archeology, Infrastructure, Environmental)

---

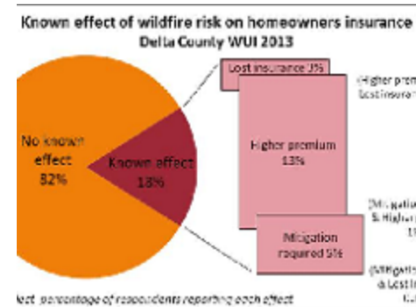
\***Drone Activity Spreadsheet**

---

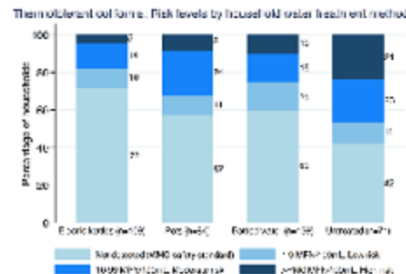
## Stories of Scientists and Communities Working Together to Make an Impact



**Community Health Clubs Tackle Climate Change Impacts in Vietnam**



**Fighting Wildfire Assumptions With Data**



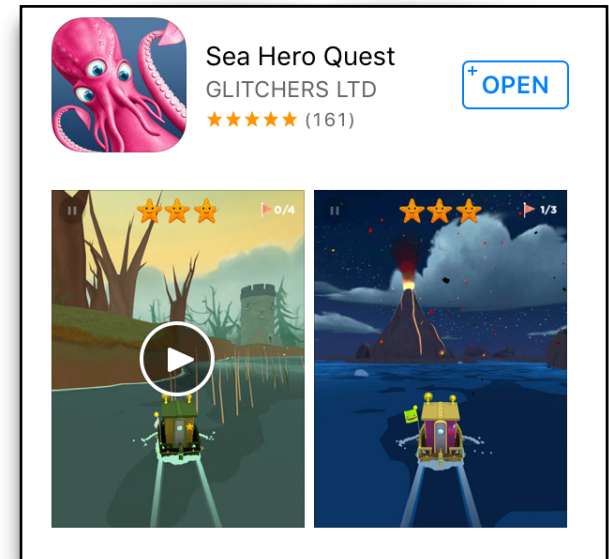
**Collaborating to Improve Home Water Treatment in Rural China**



**Taking Health into Their Own Hands: A Wyoming Town Responds to Natural Gas Operations**



# *Citizen Science, Crowd Sourcing, Gamification*



# THE ZOONIVERSE

is people powered research

1 2 3



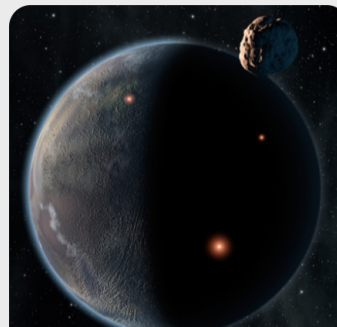
SNAPSHOT SERENGETI



BAT DETECTIVE



CYCLONE CENTER



PLANET HUNTERS



OLD WEATHER



GALAXY ZOO

# SIZONet



Local Observations  
Seasonal Ice Zone Observing Network  
(SIZONet)

Exchange for Local Observations and Knowledge of  
the Arctic



Home About Research Methods Public Information Data Add observation Contacts

## CURRENT SEARCH

Keyword(s):

Date(s):

Location(s):

All

Observer(s):









All

NEW SEARCH ►

## Search Results

Showing observations 1 - 20 of 2385

Sort by:

Date	Observer	Village	Additional elements
2012-02-13	Joe Leavitt	Barrow	
2010-12-04	Paul Apangalook	Gambell	   
2010-12-03	Joe Leavitt	Barrow	   

Home Search Observations About Research methods Contacts Add observation Logged in as ELOKA Team Log out

Observation ID:

SHINAL40508\_1

Recorder: Mette Kaufman

Observer: Curtis Nayokpuk

(Shishmaref)

Date: 2014-05-08

New Search ►

## Observation Details

[Back to search results](#)

[Edit Observation](#)

[Previous](#) [Next](#)

### Transcript

PHOTO OBSERVATION: Shorefast ice is stable and hunters are launching to hunt Bearded Seals. Shorefast "winter" ice averages 3ft, and the launch site is 3 miles from town. ( source: Sea Ice for Walrus Outlook (SIWO) <http://www.arcus.org/search-program/siwo> )

### Multimedia



8 May 2014 - Piled ice



8 May 2014 - Hunters cross the sea ice outside of



Local Observations  
Seasonal Ice Zone Observing Network  
(SIZONet)

Home About Research Methods Public Information Data Add observation Contacts

Logged in as Joe Leavitt Log out

General observation information

Weather detail

Ice detail

Wildlife

Activity detail

Photos/Video

Observation ID: GAMAP101204

Recorder: Paul Apangalook

Observer:

Observation Location:

Observation date

2010-12-04

(yyyy-mm-dd)

Observation time

:  (hh:mm)

Conditions:

Precipitation:

Skies:

Wind speed:

Visibility:

Change wind dir:

Wind direction:

Ice fog: ☐

Air temperature is approximately  °C

Wind speed is approximately  m/sec

Visibility is approximately  km

The SIZONet Observation Database is a collaborative between [SIZONet](#) and [ELOKA](#). This Web site is hosted by the [National Snow and Ice Data Center](#).



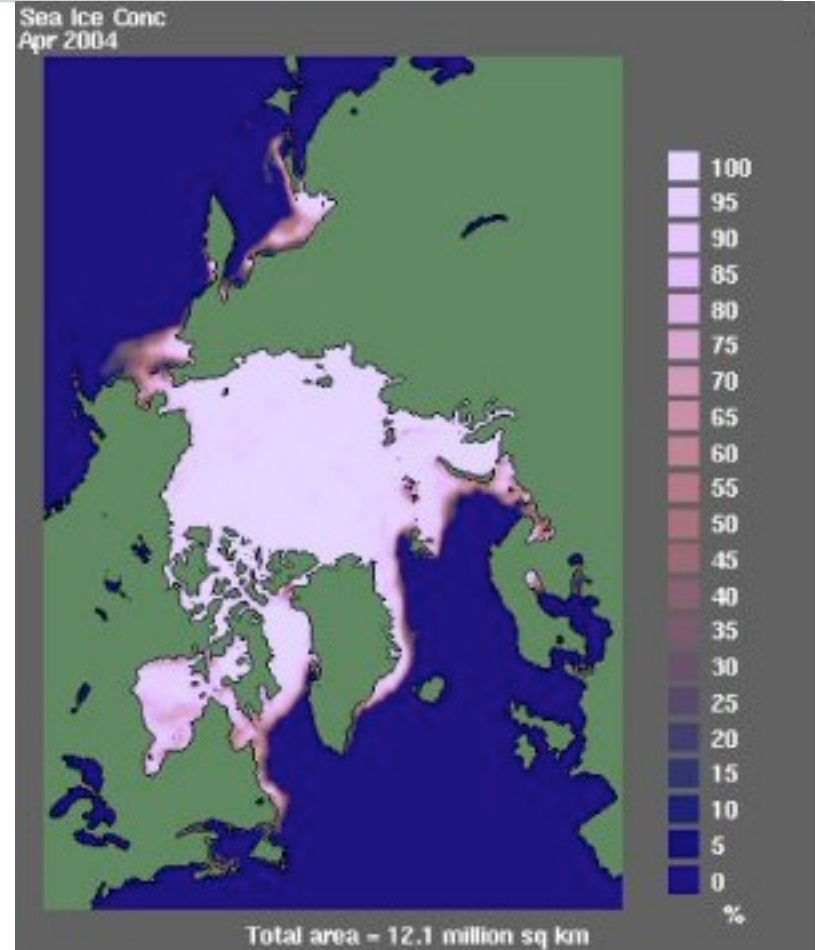
# *Challenges - Quality*

---

- Statistical characteristics
  - Resolution, accuracy, cadence, noise, etc.
- Analytic potential
  - Preservation readiness
    - Organization
    - Documentation
    - Formats
  - Number of potential user communities
  - Fitness for purpose

# The Remote Sensing Record

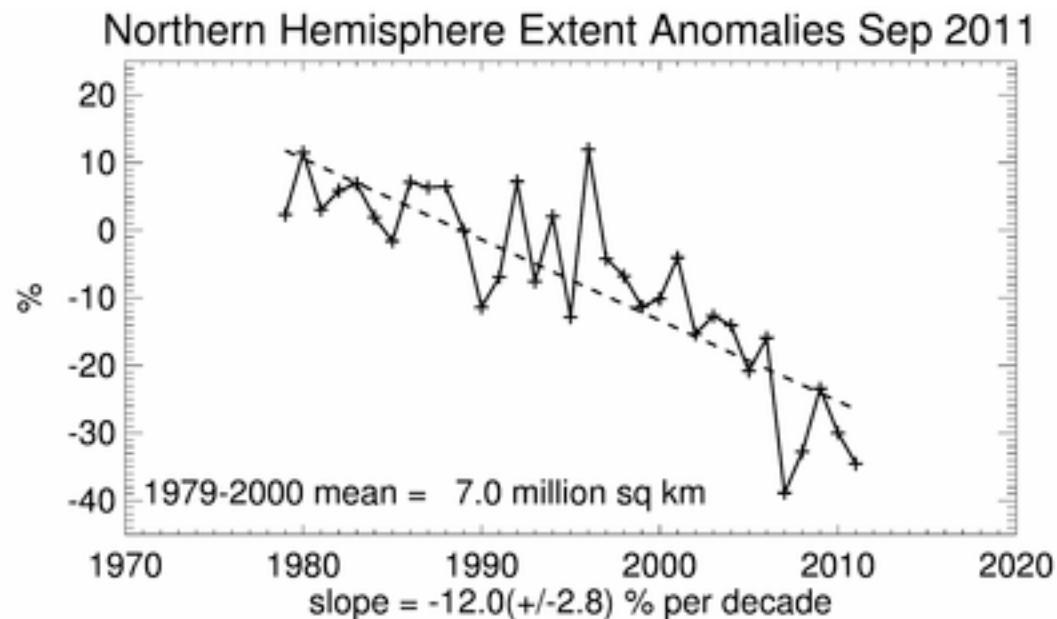
- Satellite-based Passive Microwave sensors have been measuring sea ice since 1972
- Consistent collection of data started in 1978 with the SMMR series of instruments
- Why passive microwave?
  - Distinguishing sea ice from ocean is straightforward
  - Passive microwave works through clouds and in the dark
- Initial user base was cryospheric scientists



Arctic sea ice concentration in April 2004, calculated from data measured by the Special Sensor Microwave/Imager (SSM/I) on the Defense Meteorological Satellite Program (DMSP) satellite. The image is centered over the North Pole, with continents shown in green. - Image courtesy of Florence Fetterer and Ken Knowles, National Snow and Ice Data Center, University of Colorado, Boulder, CO.

## *Trending and audience*

- Audience stayed pretty stable for roughly 20 years
- That changed once the science community started reporting statistically significant trends in sea ice extent (in the early 2000's)



- At that point climatologists became interested in the data



# *Development of the Sea Ice Index*

---

- But did NSIDC really have the data the climatologists needed?
- Not really as they need:
  - Monthly averages
  - Climatologies
  - Anomaly maps and trends
- Thus the Sea Ice Index (2002) data set was born with support from NOAA NESDIS
- But the graphics and image display part of the Sea Ice Index probably would not have been created if Ken Knowles hadn't kept grumbling about the “last 10%” problem



# Sea Ice Index

Arctic

Antarctic

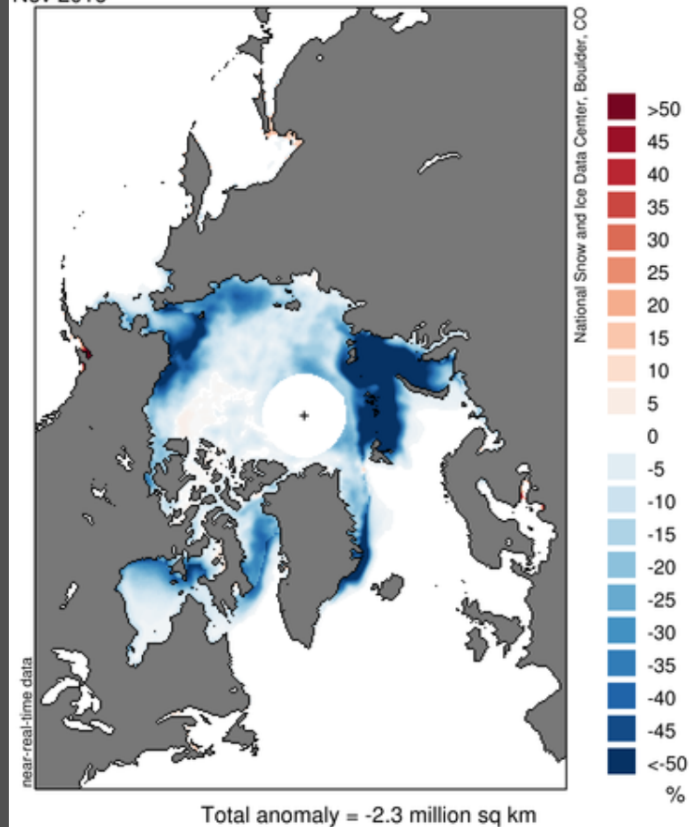
**6 July 2016: Sea Ice Index Now Version 2.** [Read More...](#)[Animate Monthly Images](#)[Compare Anomalies](#)[Compare Trends](#)[Documentation](#)[FAQ](#)[Data Archive](#)

Monthly

Daily

Blue Marble View

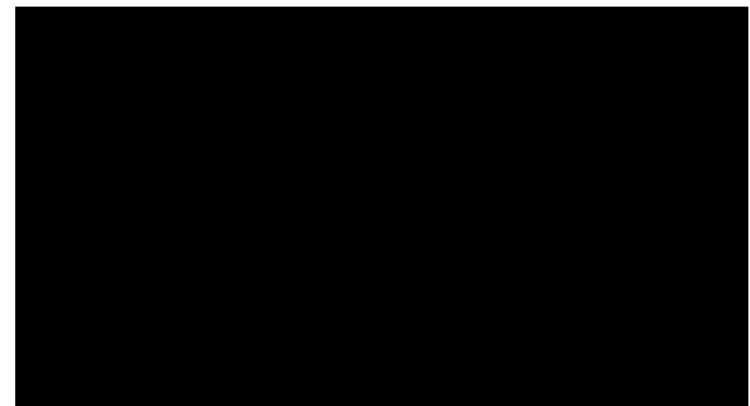
Sea Ice Concentration Anomalies  
Nov 2016



The monthly Sea Ice Index provides a quick look at Arctic-wide changes in sea ice. It is a source for consistently processed ice extent and concentration images and data values since 1979. Monthly images show sea ice extent with an outline of the 30-year (1981-2010) median extent for that month (magenta line). Other monthly images show sea ice concentration and anomalies and trends in concentration.

[More about the monthly images »](#)

## Monthly Sea Ice Extent Anomaly Graph



About  
this  
graph



Extent

Concentration

Anomalies

Trends

## *What happened next?*

---

- 2007 sea ice minimum was extreme
- Media attention was also extreme – NSIDC fielded 150 media contacts in just a few weeks, blurb above the fold on the front page of the New York Times, private briefing for Al Gore, etc.
- That led to the general public becoming one of NSIDC's most vocal and consistent audiences
- This also led to a rise in requests by the science community for access to our data
- Arctic Sea Ice News and Analysis (ASINA) site developed to prevent User Service Office from being overwhelmed



## Daily Image Update

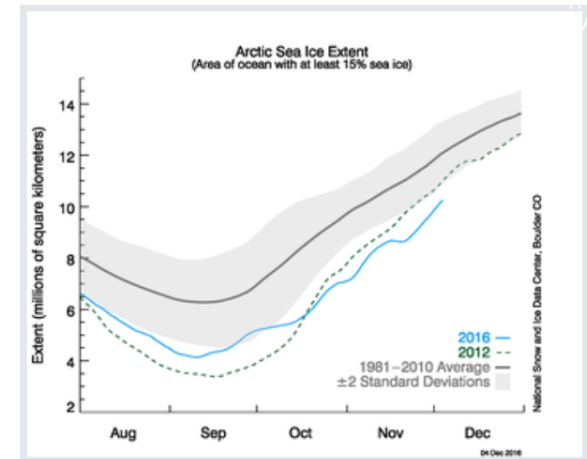
Read scientific analysis on Arctic sea ice conditions. We provide an update during the first week of each month, or more frequently as conditions warrant.

### ALSO SEE:

Icelights: Answers to your burning questions about ice and climate.



Click for high-resolution image. —Credit: National Snow and Ice Data Center



Antarctic daily images

November  
2, 2016

## Sluggish ice growth in the Arctic

← Previous Article



Search



After a quick initial freeze-up during the second half of September, ice growth slowed substantially during

## *And then....*

---

- Because of ASINA the audience for our data expanded into a community without a lot of background science knowledge but who are “loyal, tenacious, and perceptive”
- They ask a lot of questions and often many of them ask the same or similar questions in a short period of time
- We didn’t have the resources to address these one at a time; so IceLights was born

# Icelights....

11:39 AM82%


← → ↺

nsidc.org/cryosphere/icelights/

↑

≡

7


National Snow & Ice Data Center


[DATA](#)

[RESEARCH](#)

[NEWS](#)


[ABOUT](#)

[Web pages](#) 



## Icelights: Your Burning Questions About Ice & Climate


[Icelights Home](#) [About Icelights](#) [Crash Course](#) [Archives by Topic](#)



### Is the Arctic Ocean a carbon sink?

We've all heard it: Arctic sea ice is melting. Sea ice is thinner year to year and there is less of it. In 2007, scientists observed a nearly 50 percent loss of summer ice as compared to 1980. With such a dramatic shift, what else is taking place in the Arctic Ocean?

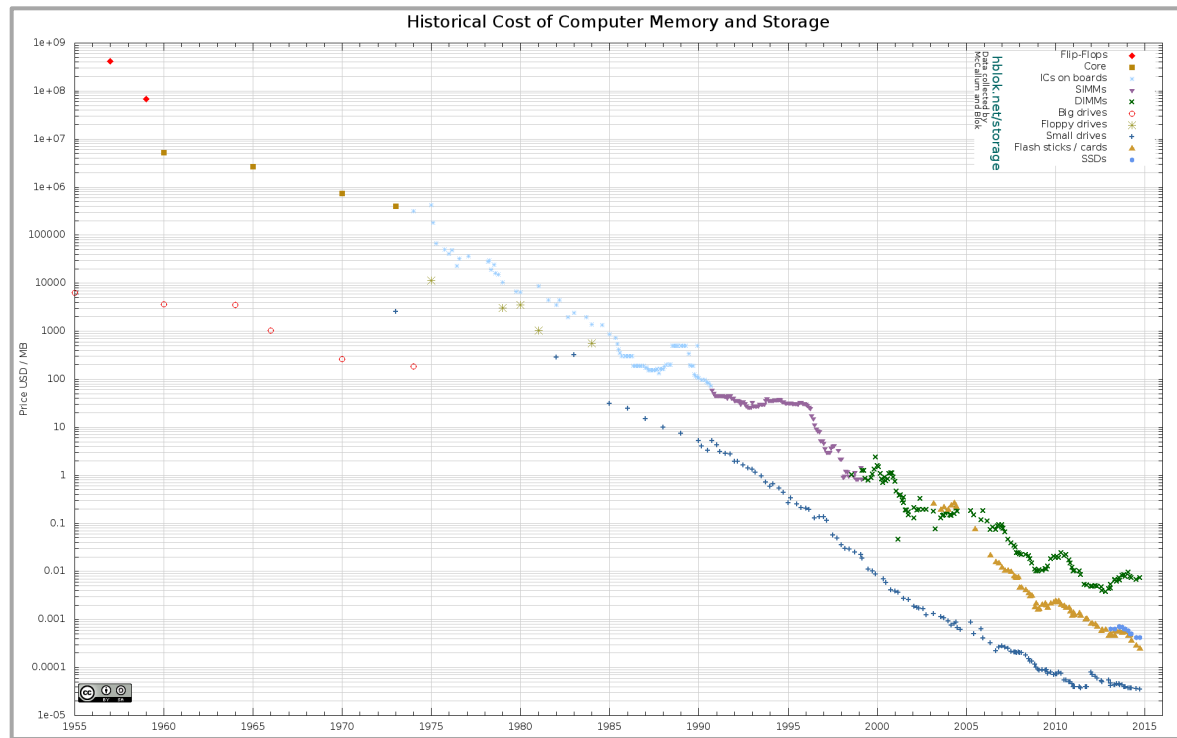
[Read more...](#)





# Challenges - Access & Availability

- Data volumes are typically growing by a factor of 10 every 5 years (Forbes, June 2012)
- Moore's law says that storage costs drop 50% every 18 months



from Harvard at <http://hblok.net/blog/posts/2013/02/13/historical-cost-of-computer-memory-and-storage/>

*But...*

The screenshot shows the top portion of a web article. At the top is a dark red header with the 'InfoWorld' logo on the left and a magnifying glass icon on the right. Below the header, the article title 'Data creation outstrips storage for the first time' is displayed in large, bold, black font. Underneath the title is a summary paragraph: 'A new IDC report shows that data is being created at a faster rate than previously thought, and yearly data production is exceeding available storage space'. Below the text is a row of social media sharing icons: Twitter, Facebook, LinkedIn, Google+, Reddit, StumbleUpon, Email, and Print. To the right of these icons is a red speech bubble icon. At the bottom of the article preview, it says 'By Jon Brodtkin' and 'Network World | Mar 12, 2008'. The date 'Mar 12, 2008' is enclosed in a pink rounded rectangle, with a pink arrow pointing from the text '2008!' to it.

**InfoWorld**

## Data creation outstrips storage for the first time

A new IDC report shows that data is being created at a faster rate than previously thought, and yearly data production is exceeding available storage space

By Jon Brodtkin  
Network World | Mar 12, 2008

2008!

*Soo...*

*NOT EVERYTHING CAN BE  
SAVED!!!*



# *Challenges - Access & Availability*

---

- Many disciplines do not have data repositories
- Funding mechanisms for repositories have not been figured out
- The question of who does the work to make data accessible to audiences outside the original community is open
- Research Libraries are only now becoming familiar with data and are under financial stress as well
- Researchers in many fields are
  - resistant to sharing their data with those outside their project
  - resistant to providing data to domain repositories
  - resistant to the need to document their data adequately so that others may use it



# *Challenges - Access & Availability*

---

- Social media data is generally considered to be ephemeral in value
- Many projects have no long term data management planning
- Even federal funding for data repositories is subject to priorities
- If the data is available in a repository somewhere can you even find it?
- "Public good" seems to be a concept that doesn't work in the US

# *Solutions? - Access & Availability*

---

- Lots of copies keeps stuff safe
  - World Data System started in 1958, to have 3 copies of data spread around the world
  - Internet Archive is funding a copy of itself in Canada
- Towns didn't have public libraries until Dale Carnegie, where is the equivalent for data?
- Do we really think crowdsourcing will solve this problem?

# *Challenges - Access & Availability*

---

- Ethics
  - IRB is not enough
  - US law based on rights of the individual
  - But what about:
    - Rights of communities?
    - Rights of families?
    - Rights of special populations?
    - Genetic rights?

# Challenges - Security

---

- Hackers are sophisticated and security experts are expensive
  - Both Landsat-7 and the NASA Terra EOS satellites were hacked way back in 2007 and 2008!!!
    - 2011 Report to Congress of the U.S.-Chinese Economic and Security Review Commission
  - Most companies that take credit cards prefer to insure against data theft rather than prevent it
  - The general wisdom in the CS world is that the only way to secure software is to make it open source (many eyes) and ensure there are no back doors.
- Data and all communication paths needs strong encryption
- Ongoing data integrity checks are mandatory



# Questions?

