NOAA’s Big Data Project

Dr. Edward J. Kearns, NOAA Chief Data Officer
National Oceanic and Atmospheric Administration
U.S. Department of Commerce

U.S. Committee on the Marine Transportation System (CMTS) Webinar
April 20, 2018
Outline

- Introduction to NOAA Open Data
- What is the Big Data Project (BDP)?
- Findings and lessons-learned from 2015-2018
- Challenges for the BDP related to ocean data
- Goals and future direction of the BDP
Our mission

To understand and predict changes in climate, weather, oceans, and coasts, to share that knowledge and information with others, and to conserve and manage coastal and marine ecosystems and resources.

Science

9

Key focus areas

NOAA’s diverse mission extends from the surface of the sun to the depths of the ocean

Service

12,000

Total number of NOAA personnel worldwide

Dedicated to the understanding and stewardship of the environment.

Stewardship

6,773

NOAA scientists and engineers

Studying and monitoring our evolving planet
NOAA has been doing Data Science and “Big Data” for many years...
How much data do we have?

- ~70,000 open datasets
- 30 PB archived data
- ~200 PB of modeled and other data
- Weather
- Climate
- Ocean
- Coasts
- Fisheries
- Ecosystems
- Charts & Maps
- Almost all are Full and Open Data
- Most are accessible via internet today
Members shall provide on a free and unrestricted basis essential data and products which are necessary for the provision of services in support of the protection of life and property and the well-being of all nations...
Debate about roles & responsibilities among NOAA, commercial weather interests, and academic researchers concerning the provision of data and services drove a study by the National Academies of Science in 2003 called the “Fair Weather” report.

Focused on effective Partnerships and how data should be shared for the benefit of all.

“Recommendation 5. The NWS should make its data and products available in Internet-accessible digital form. Information held in digital databases should be based on widely recognized standards, formats, and metadata descriptions to ensure that data from different observing platforms, databases, and models can be integrated and used by all interested parties in the weather and climate enterprise.”
NOAA's participation in the environmental information enterprise is founded on:

- Mission connection
- Consultation
- Open information dissemination
- Equity
- Recognition of Roles of Others
The Value of the "Blue Economy"

The future of the Ocean Economy...
Total ocean industry value-added expected to double by 2030

USD billion

Overview of industry-specific value-added in 2010 and 2030

- Maritime and coastal tourism
- Offshore oil and gas
- Port activities
- Maritime equipment
- Fish processing
- Offshore wind
- Water transport
- Shipbuilding and repair
- Industrial capture fisheries
- Industrial marine aquaculture

Source: OECD (2016) The Ocean Economy to 2030
While most companies remain privately held, making it difficult to confirm the size of the private weather industry, its market capitalization is estimated at $7B and predicted to grow 10-15 percent annually as it continues to serve consumers, businesses, and governments both in the U.S. and abroad. AccuWeather, for example, has begun to serve consumers and businesses in China, and other companies have worked with governments in developing countries to build out their weather infrastructure. Through acquisitions, large companies like Panasonic, IBM, and Monsanto (which was recently purchased by Bayer) have entered the weather space in the past five years...

https://www.weather.gov/media/about/Final_NWS%20Enterprise%20Analysis%20Report_June%202017.pdf
Obstacles to Effective Use of Open Data

1. Technical – how to distribute so much data?
2. Understanding – how can these technical data be widely understood?

Idea: Let’s try to leverage the value of the data to take down these obstacles...
Increasing Volume and Demand for NOAA Data

NOAA/NCEI's Environmental Data Data Archive

Types of Data Delivered:
- Satellite
- Radar
- Model

Due to increase in satellite and model data

Slide courtesy of NESDIS/NCEI

DEMAND

BUDGETS (optimistic)

COSTS

Slide courtesy of NESDIS/NCEI

Slide courtesy of NESDIS/NCEI
Increasing Volume and Demand for NOAA Data

NOAA/NCEI’s Environmental Data Archive

2016 Total: 28.6 Petabytes

Due to increase in satellite and model data

Slide courtesy of NESDIS/NCEI
NOAA/NCEI's Environmental Data Archive

2016 Total: 28.6 Petabytes

Increasing Volume and Demand for NOAA Data

NOAA/NCEI’s Environmental Data Archive

Types of Data Delivered:
- Satellite
- Radar
- Model
- Ground Based and Ocean Based

Slide courtesy of NESDIS/NCEI
Increasing Volume and Demand for NOAA Data

NOAA/NCEI’s Environmental Data Archive

DEMAND

BUDGETS (optimistic)

COSTS
Industry Builds What is Needed to Support its Users

- Uses
- Decision Support Tools
- Cloud Platforms
- NOAA Data and Expertise

Information Consumers

3rd Parties
A Value-Driven Open Data Ecosystem

**Collaboration Terms:**
- 3 Year Cooperative R&D Agreements (CRADAs)
- No net cost to taxpayer
- Data remain full and open

**Use**

**Cloud**

**Open Data**

**Expertise**
One-way transfer out of federal systems. Only a trusted user inside security boundary.

Distributing a single copy of data can support all users.
Big Data Project
Collaborators’ Data Offerings

- AWS
  - https://aws.amazon.com/noaa-big-data/
  - https://registry.opendata.aws/

- Google Cloud Platform
  - https://explorer.earthengine.google.com/#index

- IBM - PaaS
  - https://noaa-crada.mybluemix.net/

- Microsoft (No public-facing services to date)

- Open Commons Consortium
  - http://edc.occ-data.org/
AWS and Weather Radar

Entire NEXRAD 88D Weather Radar Archive transferred to AWS, Google and OCC in Oct 2015 (~ 300TB, 20M files)

Following AWS service release:
- Increased usage (2.3 times), 50% reduction on NOAA servers.
- New uses – bird migration, mayfly studies
- 80% of NOAA NEXRAD data orders are now served by AWS.
  - (Ansari et al, 2017 BAMS)
**New Datasets**

**National Water Model:**
- 23-year reanalysis
- Real-time forecasts

**Provisional GOES-16:**
- Now: Imagery
- L1b ABI Products
- L2 products
- Lightning Mapper (GLM)

National Water Center: [http://water.noaa.gov/tools/nwm-image-viewer](http://water.noaa.gov/tools/nwm-image-viewer)

NOAA NESDIS: [https://www.nesdis.noaa.gov/content/flashy-first-images-arrive-noaa%E2%80%99s-goes-16-lightning-mapper](https://www.nesdis.noaa.gov/content/flashy-first-images-arrive-noaa%E2%80%99s-goes-16-lightning-mapper)
Cooperative Institute for Climate and Satellites – North Carolina

• Providing GOES-16 data from NOAA/NESDIS Ground System (PDA) to Collaborators.

1 source, 5 validated feeds to the CRADA Collaborators

• Timing - As fast as they appear at NOAA/NESDIS distribution point
• Latency - Single hop through CICS-NC systems, w/checksums
• Impact - Minimal load on NOAA’s operational systems and networks

Observed additional latencies from CICS-NC transfer from NOAA Ground System to BDP Collaborator platforms

• Maximum additional latency: 2 to 3 min (full disk ABI, Band 2)
• **Typical Range of additional latency: 4.5 to 5.5 seconds**
GOES-16 AWS Accession Volume by Age

Age is the time from observation. Accession is anytime the data are read.

Courtesy of Otis Brown and Jonathan Brannock, CICS-NC, NC State Univ
GOES-16 AWS Accession Locations

Courtesy of Otis Brown and Jonathan Brannock, CICS-NC, NC State Univ
Note: April 2018 is only a half-month, from 4/1-4/15

Courtesy of Otis Brown and Jonathan Brannock, CICS-NC, NC State Univ
From a single Collaborator’s platform, 10x the stored volume is being accessed by the public.

*Courtesy of Otis Brown and Jonathan Brannock, CICS-NC, NC State Univ*
1.2 PBs of climate and weather data accessed through Google BigQuery, in 4 months

30-100x of NOAA deliveries in that time

Google Cloud Platform

- GOES-16 (began July 2017)
- National Water Model data
- NEXRAD Level 2

---

https://cloud.google.com/bigquery/public-data/noaa-ghcn
NOAA Climate Data in Google Earth Engine

NOAA CDR PATMOSX V53: Cloud Properties, Reflectance and Temperature - open in workspace

NOAA - Every day from 1979 to 2018

This dataset provides high quality Climate Data Record (CDR) of multiple cloud properties along with Advanced Very High Resolution Radiometer (AVHRR) brightness temperatures derived from the AVHRR.

NOAA CDR AVHRR Leaf Area Index and FAPAR Version 4 - open in workspace

NOAA - Every day from 1981 to 2018

The NOAA Climate Data Record (CDR) of AVHRR Leaf Area Index (LAI) and Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) dataset contains derived LAI and FAPAR values from the AVHRR.

NOAA CDR AVHRR Surface Reflectance Version 4 - open in workspace

NOAA - Every day from 1981 to 2018

The NOAA Climate Data Record (CDR) of AVHRR Surface Reflectance contains gridded daily surface reflectance and brightness temperatures derived from the AVHRR.

NOAA CDR AVHRR Aerosol Optical Thickness: Daily AOT - open in workspace

NOAA - Every day from 1981 to 2017

The NOAA Climate Data Record (CDR) of Aerosol Optical Thickness (AOT) is a collection of global daily 0.1 degree derived data from the PATMOS-x AVHRR level-2.

NOAA CDR AVHRR Normalized Difference Vegetation Index: Daily NDVI - open in workspace

NOAA - Every day from 1981 to 2018

The NOAA Climate Data Record (CDR) of AVHRR Normalized Difference Vegetation Index (NDVI) contains gridded daily NDVI derived from the NOAA AVHRR Surface Reflectance.

NOAA CDR GRIDSAT-B1: Geostationary IR Channel Brightness Temperatures - open in workspace

NOAA - Every day from 1980 to 2018

This dataset provides scientific researchers a high quality Climate Data Record (CDR) of global infrared measurements from geostationary satellites.

NOAA CDR Ocean Heat Fluxes V2 - open in workspace

NOAA - Every day from 1988 to 2017

The Ocean Heat Fluxes dataset is part of the NOAA Ocean Surface Bundle (OSB) and provides high quality Climate Data Record (CDR) of the air/sea interface heat flux.

NOAA CDR Near-surface Atmospheric Properties V2 - open in workspace

NOAA - Every day from 1988 to 2017

The Near-surface Atmospheric Properties dataset is part of the NOAA Ocean Surface Bundle.
OCC’s Environmental Data Commons

http://edc.occ-data.org/

The OCC Environmental Data Commons

Repository for environmental public data sets of scientific interest, hosted as part of the Open Science Data Cloud Ecosystem

GOES 16

NEXRAD

Tools | Notebooks
A matter of miles

How the slightest shift kept Hurricane Irma from turning into an even worse disaster

By NATHANIEL LASH and NEIL BEDO
Times Staff Writers
Sept. 20, 2017

Scroll down
IBM PaaS
Earth Systems Data Portal

Get Started >> REST API Download
ACCESS DATA VIA REST API

Downloading RAP Data Files Via REST API

The following documentation utilizes cURL to download a NOAA RAP weather file. The parameters, shown below in blue, must be configured by the user.

curl --request GET \
--url https://api.us.apiconnect.ibmcloud.com/jmsalaiusibmcom-poc/noaa-data/rap/{yyyymmdd}/{tt00}/{ff} \
--user "username:password" \
--output "local file directory/rap_130_#1_#2_#3.grb2"

https://noaa-crada.mybluemix.net/
Ensuring Data Authenticity

- Users trust NOAA data they access from NOAA sites

- What about outside NOAA?
  - Collaborators’ sites?

- Example: Fake Irma forecast

Nearly 40,000 shares on this fake Irma forecast on Facebook. There is no actual threat to any land in next 5 days. pic.twitter.com/hwmuE5UwKn

10:11 AM - Sep 1, 2017

Everyone needs to pay attention to Hurricane Irma. She predicted to come through Mexico hit us and everything in between up to Houston. She's already a Category 2 and hasn't even got into warm water yet.
Ocean Data Challenges for BDP

- Ocean data have proven to be more difficult to move to Collaborators than weather and climate data.
- Why? Different market.
- Establishing an understanding of the data.
- Approach the problem differently, as a whole?
- Move multiple datasets to address many questions.
- “Ocean data platform” concept – focus on community.
Publish NOAA Datasets, such as:

- Coastal observations
  - Water levels, currents, tides, winds, waves, air gap, bathymetry
- Model outputs
  - Coastal regional models, ocean models
- Fisheries data
  - Catch, trawl data
- Vessel/traffic data
- Acoustics, ‘Omics data, eDNA
- Satellite data

Encourage others to bring Skills: Will users bring their code and tools and work collaboratively on a common platform?
AIS Data

- **USCG collects and maintains data**
  - Roughly 1 – 2TBs per year

- **Currently available through MarineCadastre.gov**
  - Ongoing collaboration between NOAA and BOEM
  - AIS data from 2009-2014 currently available, with 2015-2017 to be available this summer

- **Smaller vessel data (VMS system) available through NOAA**

- **Currently discussing ways ahead with both BDP Collaborators and NOAA Marine Cadastre personnel**
Future Direction for the BDP

- NOAA is concluding the CRADA / Experimental Phase of the project in Apr 2019.
  - All BDP Collaborators are still engaged.
  - Gathering requirements from inside NOAA.

- What is the proper business model to sustain the Partnerships & public data access on the commercial cloud?
- What is the optimal service model?
Summary

- **NOAA** has a proud history of government open data leadership and partnerships with industry and academia.

- NOAA is collaborating with industry through the Big Data Project CRADA to learn how to make NOAA’s open data accessible and widely usable at no net cost to the US taxpayer.
  - Increased access levels for NOAA open data
  - Higher Levels of Service to the data consumer
  - Reduced loads on NOAA access systems

- NOAA is seeking:
  - Sustainable partnership model to enable enhanced, cloud-based access to all of NOAA’s open data.
Questions?

ed.kearns@noaa.gov

#NOAABigData

http://www.noaa.gov/big-data-project