Public Sector Role in Big Data for the Environment

Congress on Harnessing Big Data for the Environment
Renewable Natural Resources Foundation
2016-12-06

Jeff de La Beaujardière, PhD
US National Oceanic and Atmospheric Administration
NOAA Data Management Architect

jeff.deLaBeaujardiere@noaa.gov
http://orcid.org/0000-0002-1001-9210
Overview

7 public-sector roles
(focus on NOAA)

NOAA Big Data Project

Challenges and Wishes
Role #0: Coordination

- White House Office of Science & Technology Policy (OSTP) coordination among agencies
- National Civil Earth Obs. Assessment (2012, 2016)
- US representation in international Group on Earth Observations (GEO)
Role #1: Data Production

NOAA contribution:

• Satellites
• Weather radars
• Buoy networks
• Tide gauges
• Human observers
• Animal telemetry
• Ships
• Aircraft
• Numerical models

NASA, USGS, EPA, USDA, NSF also collect (or fund) Big Data
GOES-R: NOAA's next generation of geostationary weather satellites

Geostationary Operational Environmental Satellite (GOES)

GOES-R launch 2016-11-19 successful: 1st of a 4-satellite program (GOES-R/S/T/U) that will extend operational GOES system through 2036.

Advanced imaging with increased spatial resolution and faster coverage for more accurate forecasts, real-time mapping of lightning activity, and improved monitoring of solar activity.
Role #2: Operational Reliability

- 24/7/365 operations with minimum latency for mission-essential functions
- Raw data transmission
- Data processing
- Data assimilation
- Numerical modeling
Role #3: Scientific Validity

Domain expertise
Sensor development
Algorithm development
Quality assurance
Reprocessing
Long-timescale studies
Defensible conclusions

Graphic: NOAA NCEI Climate at a Glance
Role #4: Public Access & Usability

User Tools
- Data.gov and Other Portals
- Decision Support Tools
- Scientific Software
- Numerical Models
- Value-Adding Reseller

Data Services Layer
- Data Search & Discovery Services
- Data Access Services
- Data Documentation
- Compatible Formats and Vocabularies

Data Sources
- Satellite
- Radar
- Buoy
- Ship
- Sonar
- Surveys
- ROV/UV
- Models

Data Documentation
Compatible Formats and Vocabularies
Shared Standards

Data.gov and Other Portals
Decision Support Tools
Scientific Software
Numerical Models
Value-Adding Reseller

jeff.deLaBeaujardiere@noaa.gov
NOAA Data Catalog

70,999 datasets found

Atlantic Salmon Smolt Monitoring

National Oceanic and Atmospheric Administration — Annual data are collected as part of smolt trapping operations using fish trapping methods. Traps collect emigrating salmon smolts to identify cohort...

Market News Price Dataset

National Oceanic and Atmospheric Administration — Real-time price data collected by the Boston Market News Reporter. The NOAA Fisheries' "Fishery Market News" began operations in New York City on February 14,...

Atlantic Salmon Telemetry Monitoring

National Oceanic and Atmospheric Administration — Annual telemetry data are collected as part of specific projects (assessments within watersheds) or as opportunistic efforts to characterize Atlantic salmon smolt...

Collaborators: NOAA Catalog WG
Unified Access Framework; Metadata Metrics

Collaborators: UAF project

Collaborators: EMMA project
NOAA Big Data Project (BDP)

Cloud-hosted data analysis, visualization, integration

Use data remotely

Copy huge datasets

Agency Service Tier

NOAA Data

www.noaa.gov/big-data-project
BDP CRADA Collaborators are Cloud providers serving as focal points for possible new value-added products.

Members of industry, research, and academia may join Collaborators.

All NOAA data made available equally to all the Collaborators, who choose what data to store in their Cloud based on potential Use Cases.

Collaborators may charge only for computing time and egress charges, not the original data.

Business partners may charge for value-added products and services.
1st BDP Dataset: NEXRAD L2

- NEXRAD = Next-generation Radar
  - Level 2 = reflectivity data from 150+ stations
  - 1991-present; ~850 TB uncompressed
Access to NEXRAD Level II Data


Figure from S, Ansari et al, 2016
Future Datasets for BDP

- Multi-Radar/Multi-Sensor
- Geostationary Satellite
- Numerical Models
- Fisheries catch data

Future Datasets for BDP
Role #5: Long-Term Preservation

- Ensure original data remain available 75+ years
- Perform updates, format migrations, etc as needed

Graphic courtesy Steve Del Greco & Ken Casey, NOAA/ NCEI
Dataset Identifiers & Citations for Archival Data

**NOAA DOIs Issued**

- Total # of DOIs
- 0 - 1000

**NOAA DOI Citations in Google Scholar**

- Year to date
- Cumulative
- # of hits
- 2013: 6
- 2014: 56, 62
- 2015: 173, 235
- 2016: 235, 470

Collaborators: NOAA Data Citation WG
MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren
Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

1. Policy Principles

The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.

Scientific research supported by the Federal Government catalyzes innovative breakthroughs that drive our economy. The results of that research become the grist for new insights and are assets for progress in areas such as health, energy, the environment, agriculture, and national security.

Access to digital data sets resulting from federally funded research allows companies to focus
Public Sector Roles in Big Data for the Environment

0) Coordination
1) Data production
2) Operational reliability
3) Scientific validity
4) Public access and usability
5) Long-term preservation
6) Research funding
Users need answers, not data

Data to Decisions:
- Distill huge & complex data to ~1 bit: take umbrella? permit building? mitigate flooding? build seawall?
- Support non-expert data users
Challenges

Data Volume

Data Complexity

\[ \frac{\partial n}{\partial t} + \nabla \cdot n \mathbf{V}_e = 0 \]

\[ \rho \left( \frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla \right) \mathbf{V} = \mathbf{J} \times \mathbf{B} - \nabla p - \nabla \cdot \Pi \]

\[ E = -\mathbf{V} \times \mathbf{B} + \frac{1}{en} \frac{1 - Z_{m_e} / m_i}{1 + Z_{m_e} / m_i} \mathbf{J} \times \mathbf{B} + n \mathbf{J} \]

\[ + \frac{1}{\varepsilon_0 \omega_p^2} \left[ \frac{\partial \mathbf{J}}{\partial t} + \nabla \cdot (\mathbf{J} \mathbf{V} + \mathbf{VJ}) + \sum_{\alpha=1,e} \frac{q_\alpha}{m_\alpha} (\nabla p_\alpha + \nabla \cdot \Pi_\alpha) \right] \]

\[ \frac{3}{2} \left( \frac{\partial}{\partial t} + \mathbf{V}_\alpha \cdot \nabla \right) p_\alpha = -\frac{5}{2} p_\alpha \mathbf{V} \cdot \mathbf{V}_\alpha - \mathbf{V} \cdot q_\alpha - \Pi_\alpha : \nabla \mathbf{V}_\alpha + Q_\alpha, \quad \alpha = i,e \]
Wish #1: Fully Leverage the Cloud

Operational Customers (e.g., NWS)

Archive

Cloud

Other Users
Wish #2: Enable Better Decision-making

- Earth Observations
- Model Outputs
- Ancillary Data

Decision Support Tools

Policy & Business Decisions

Public/Private collaboration opportunities:
- Derived information products
- Data integration
- Location-specific analysis
- Use Case-specific analysis
- Statistics & Trends
- What-if scenarios

non-scientist users
Thank you!

Jeff de La Beaujardière

jeff.deLaBeaujardiere@noaa.gov
http://orcid.org/0000-0002-1001-9210