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

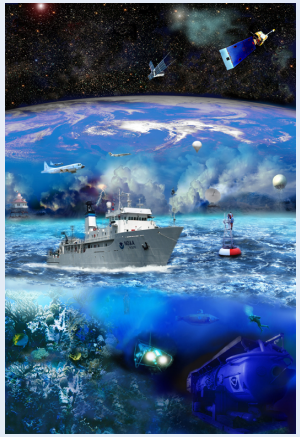
US National Oceanic and Atmospheric Administration

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

- **US Group on Earth Observations:** http://usgeo.gov/
- White House Office of Science & Technology Policy (OSTP) coordination among agencies
- National Civil Earth Obs. Assessment (2012, 2016)
- Issued *National Strategy for Civil Earth Observations* (2013), *National Plan for Civil Earth Observations* (2014), *Common Framework for Earth Observation Data* (2015)
- US representation in international Group on Earth Observations (GEO)

jeff.deLaBeaujardiere@noaa.gov

Role #1: Data Production

NOAA contribution:

- Satellites
- Weather radars
- Buoy networks
- Tide gauges
- Human observers
- Animal telemetry
- Ships
- Aircraft
- <u>Numerical models</u>

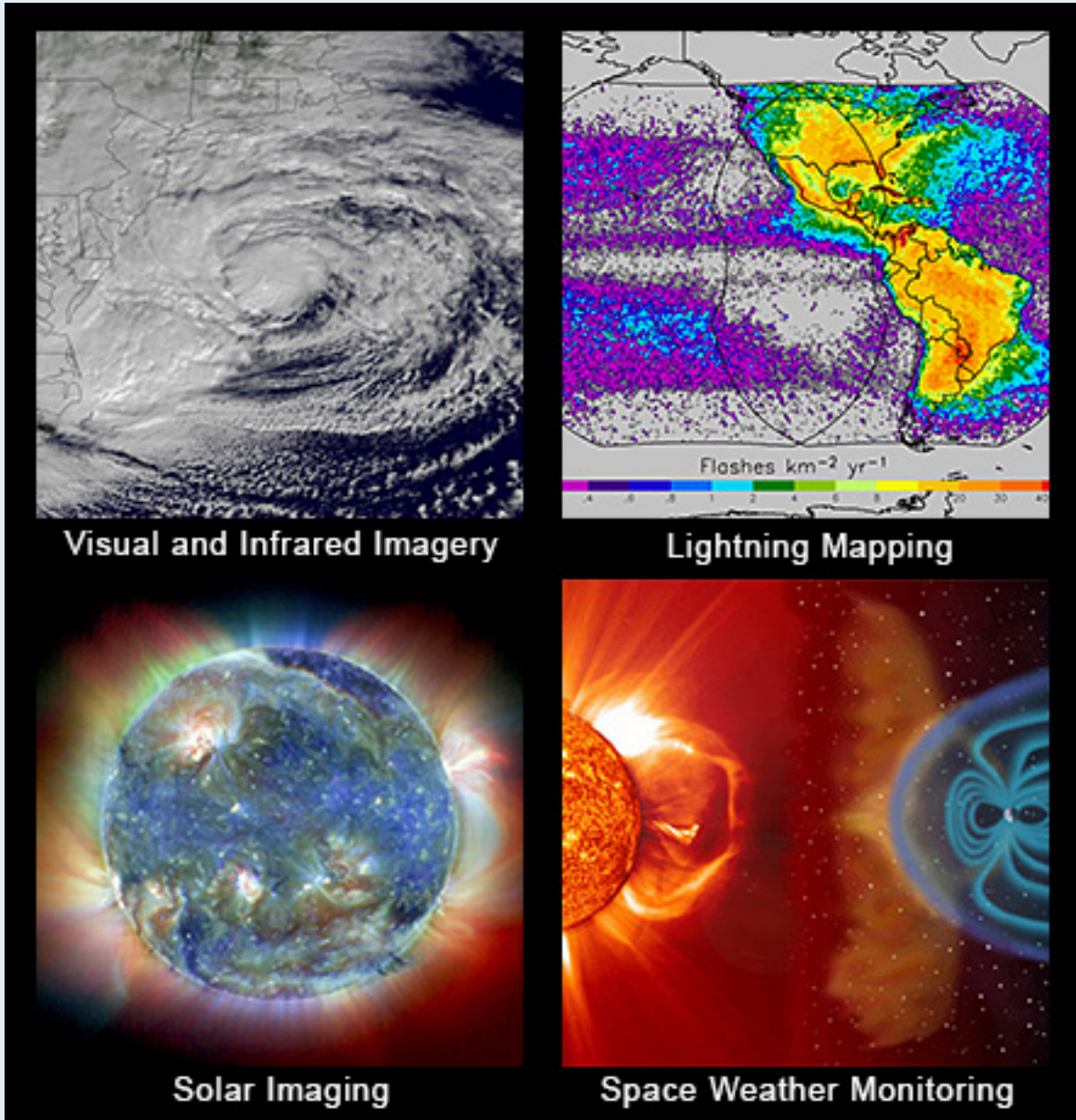
NASA, USGS, EPA, USDA, NSF
also collect (or fund) Big Data

GOES-R: NOAA's next generation of geostationary weather satellites

Geostationary Operational Environmental Satellite (GOES)

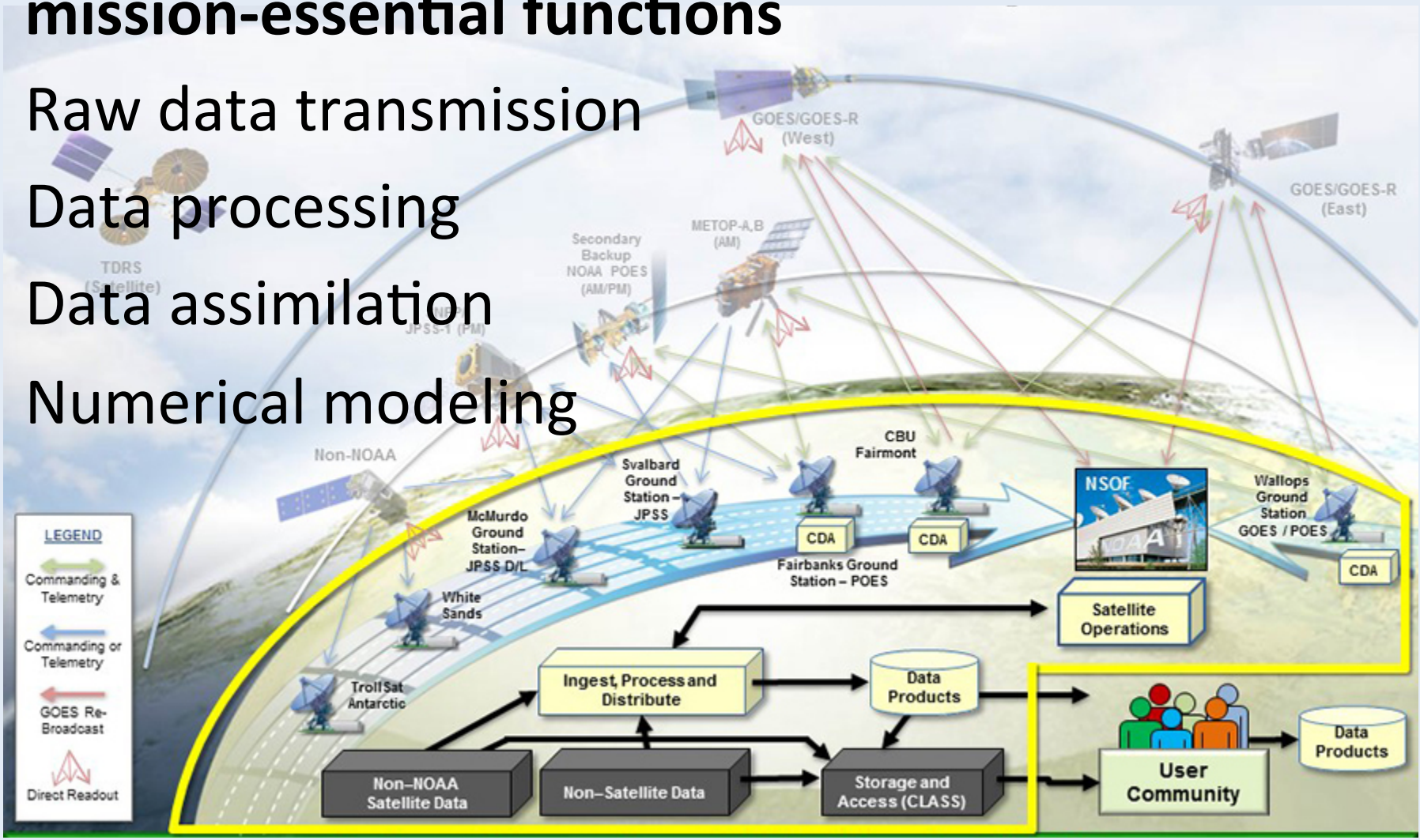
GOES-R launch 2016-11-19 successful: 1st of a 4-satellite program (GOES-R/S/T/U) that will extend operational GOES system through 2036.

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



Graphic: NOAA NESDIS [Office of Satellite Ground Services]

Role #3: Scientific Validity

Domain expertise

Sensor development

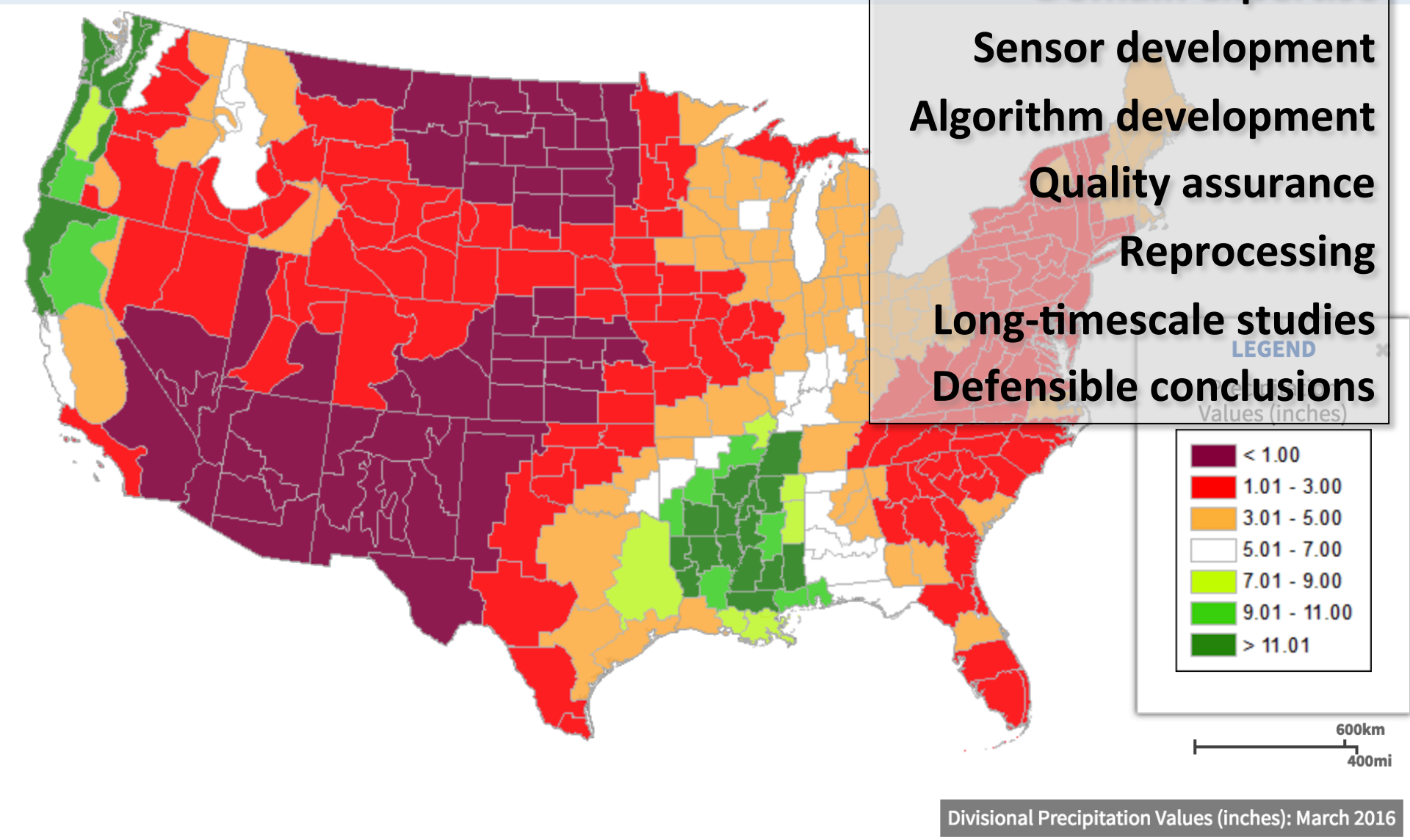
Algorithm development

Quality assurance

Reprocessing

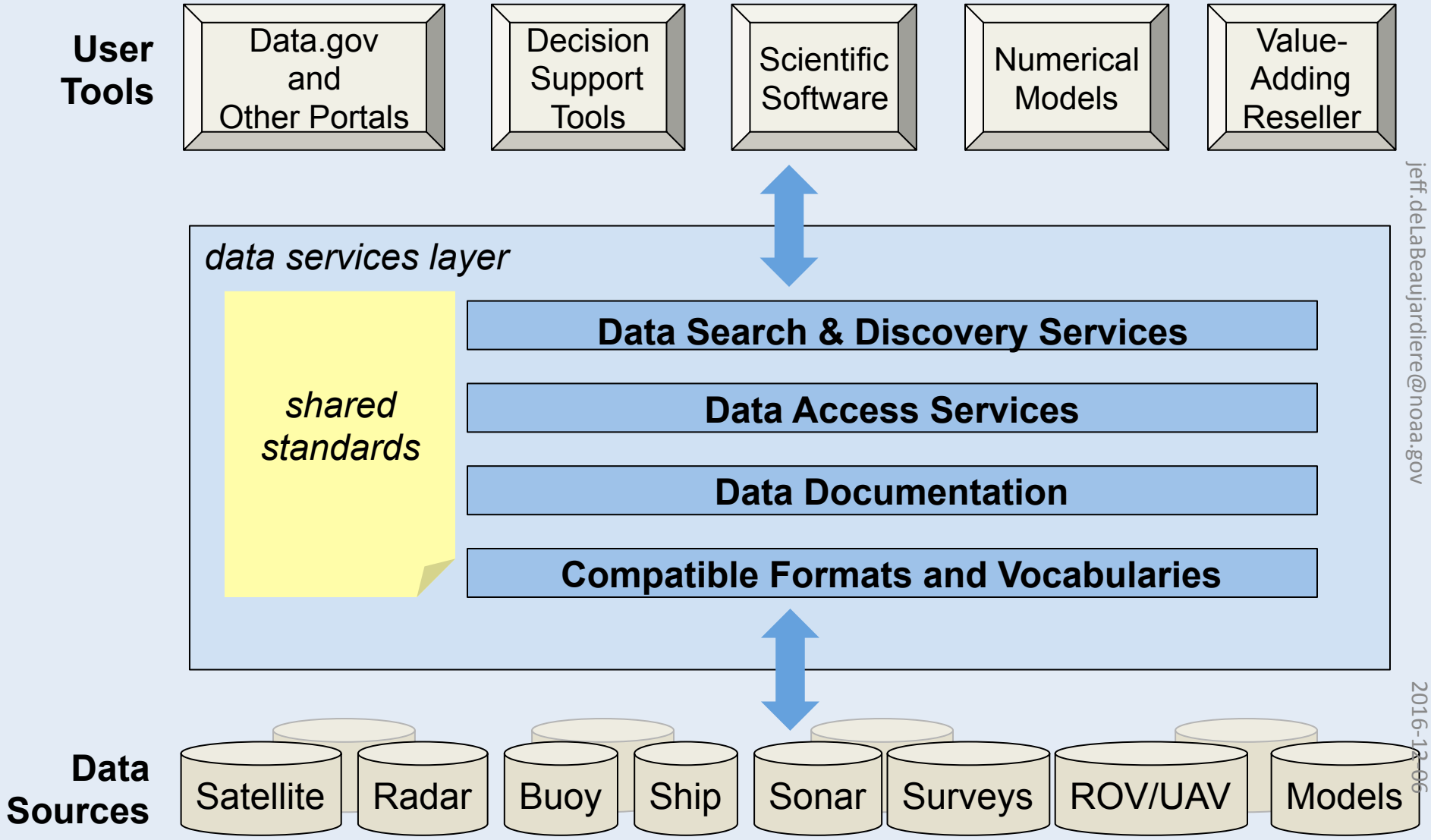
Long-timescale studies

Defensible conclusions



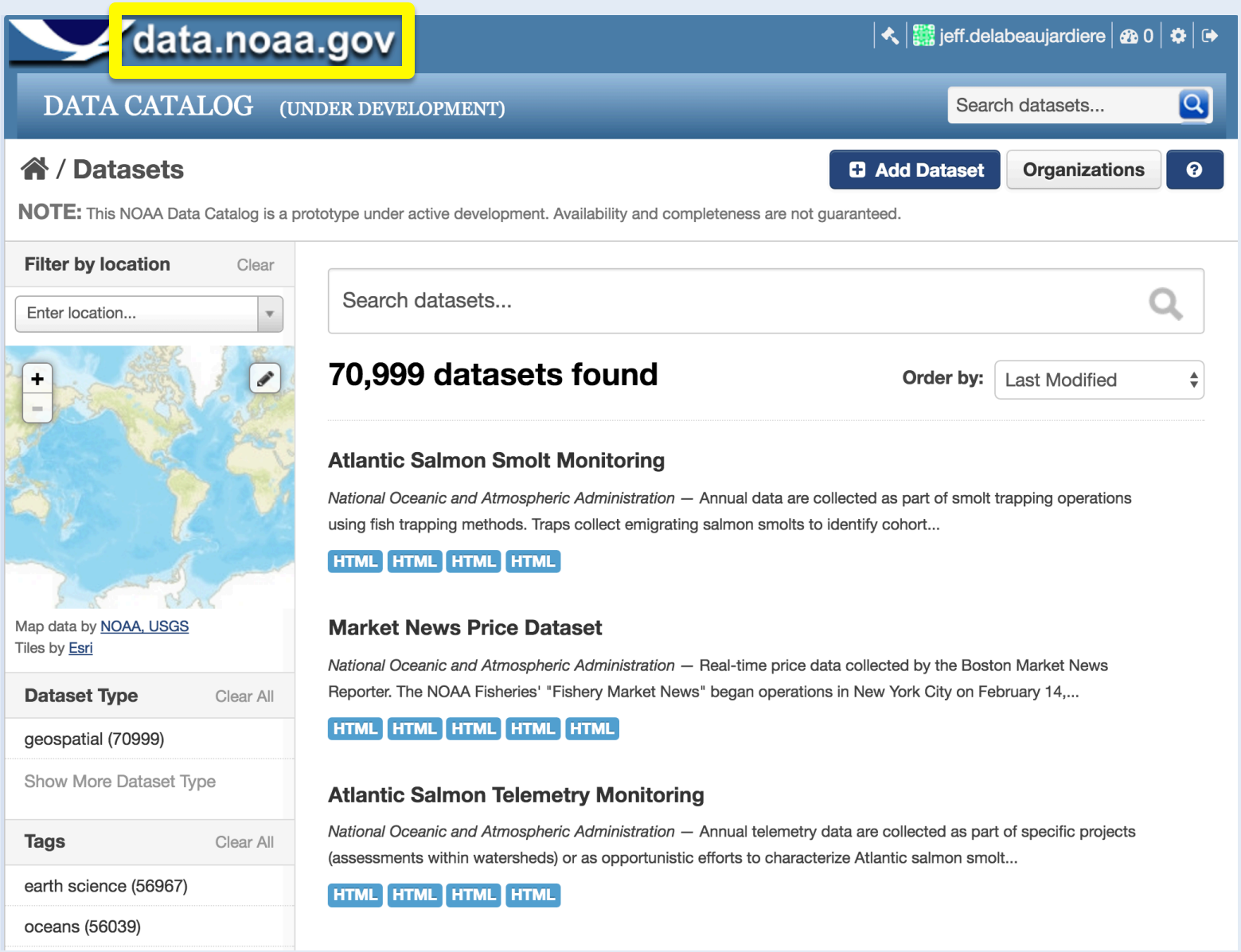
Graphic: NOAA NCEI [Climate at a Glance]

Role #4: Public Access & Usability



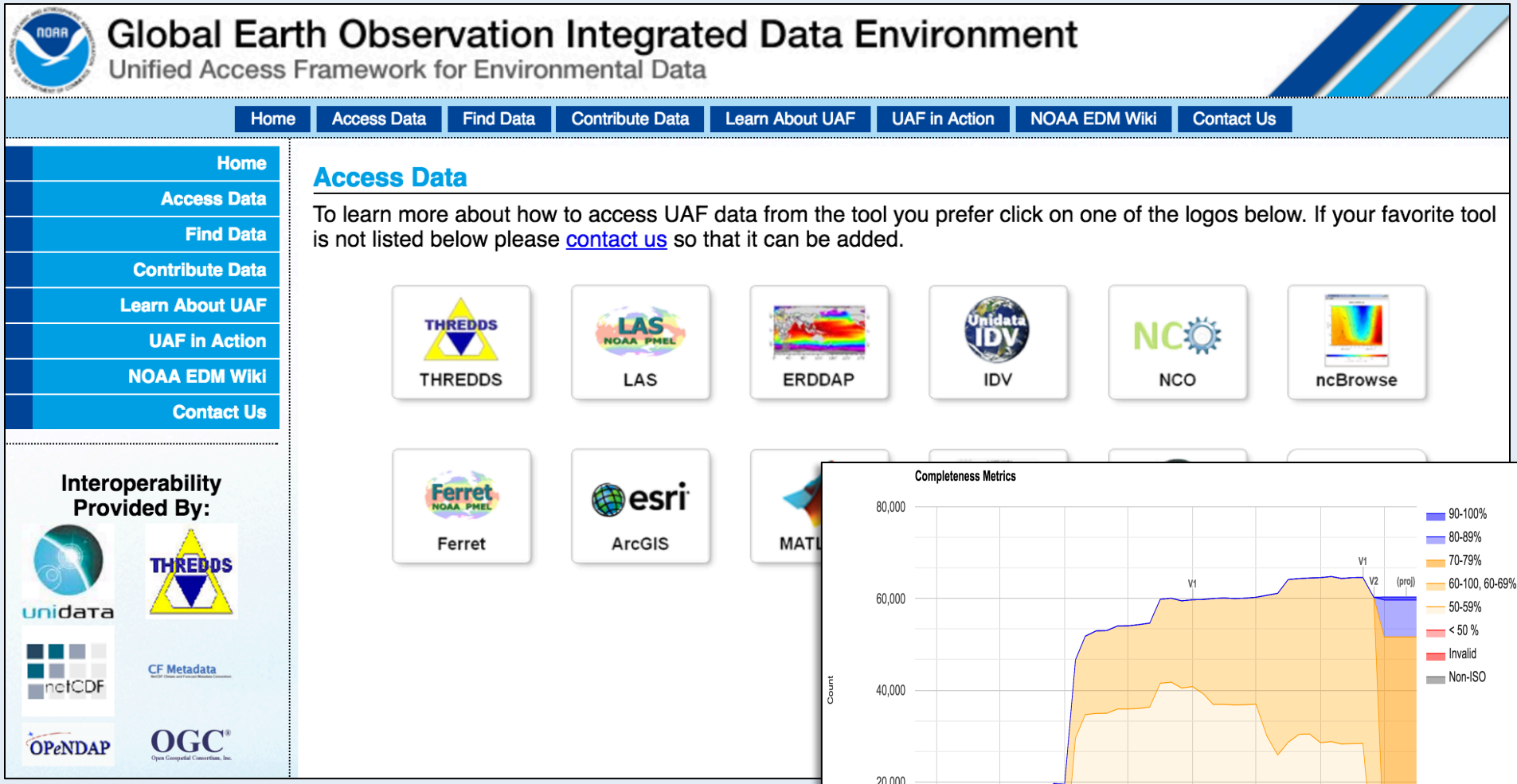
jeff.deLaBeaujardiere@noaa.gov

NOAA Data Catalog

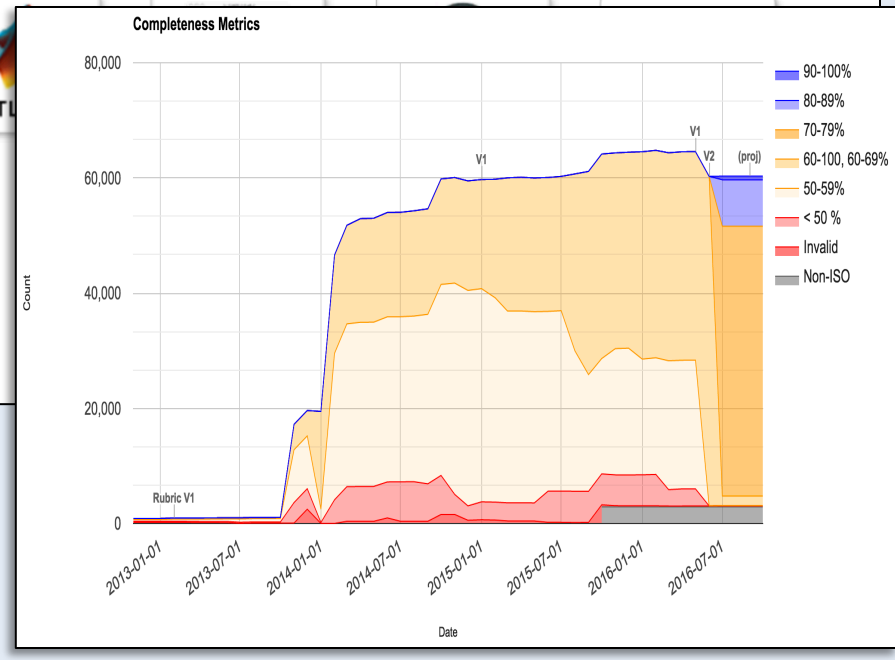


Collaborators: NOAA Catalog WG

Unified Access Framework; Metadata Metrics

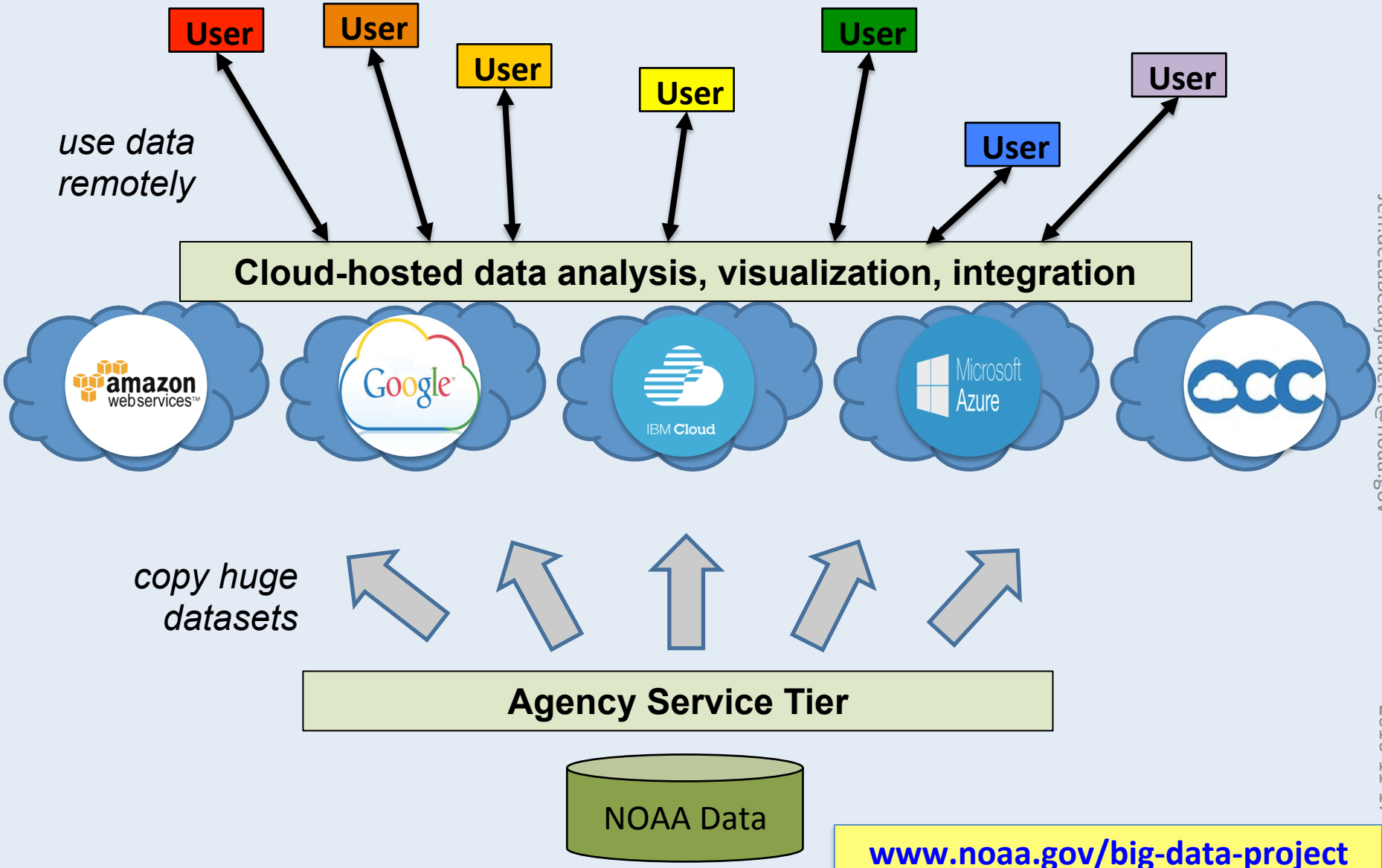


Collaborators: UAF project



Collaborators: EMMA project

NOAA Big Data Project (BDP)



www.noaa.gov/big-data-project

Jeff.deLaBeaujardiere@noaa.gov

BDP CRADA Collaborators

(CRADA: Cooperative Research & Development Agreement)

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.

amazon web services

Google Cloud Platform

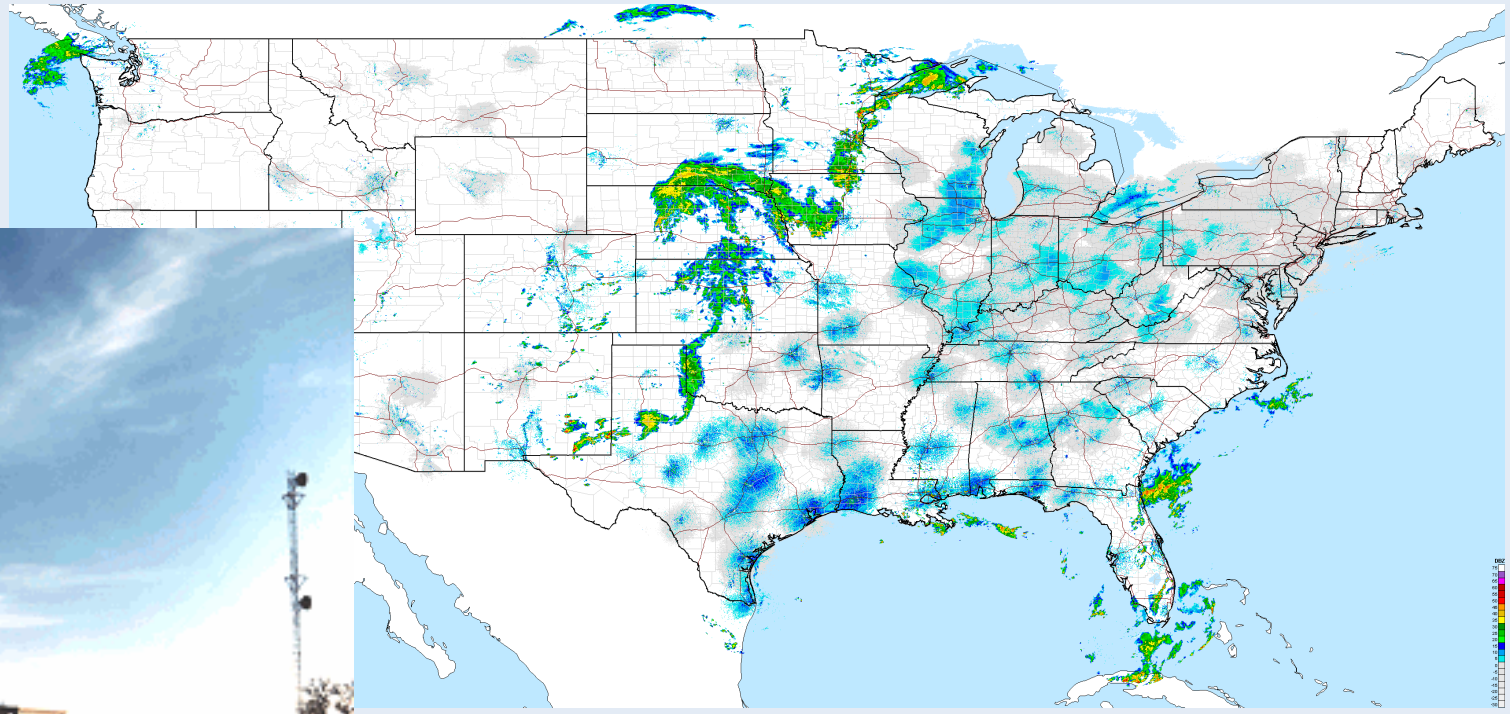
IBM

Microsoft

OPEN COMMONS CONSORTIUM

1st BDP Dataset: NEXRAD L2

- NEXRAD = Next-generation Radar
  - Level 2 = reflectivity data from 150+ stations
  - 1991-present; ˜850 TB uncompressed



jeff.deLaBeaujardiere@noaa.gov

Access to NEXRAD Level II Data

AWS: October 2015 https://s3.amazonaws.com/noaa-nexrad-level2 **(full record)**
OCC: June 2016 http://occ-data.org/NOAANEXRAD/ **(2015+)**

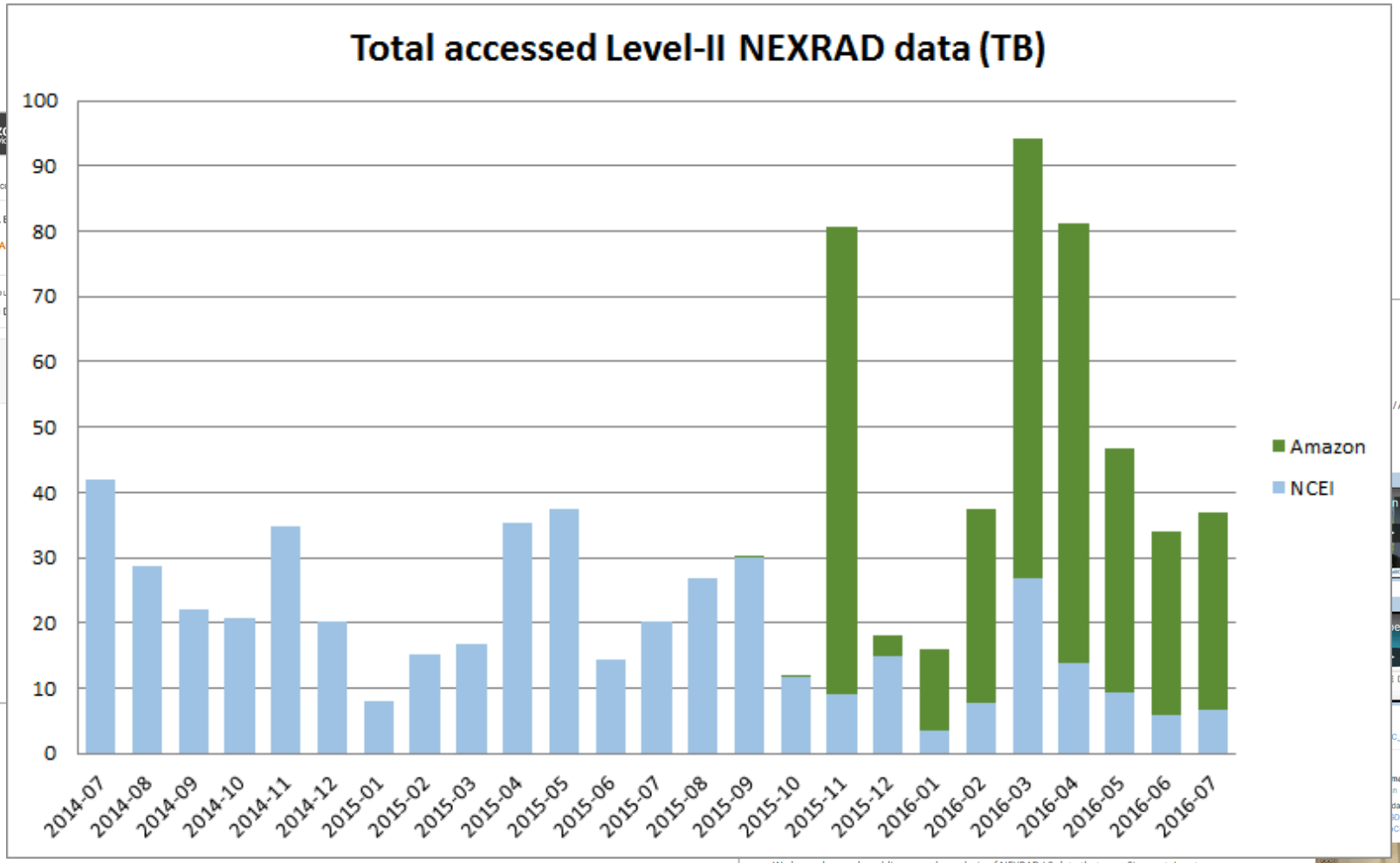
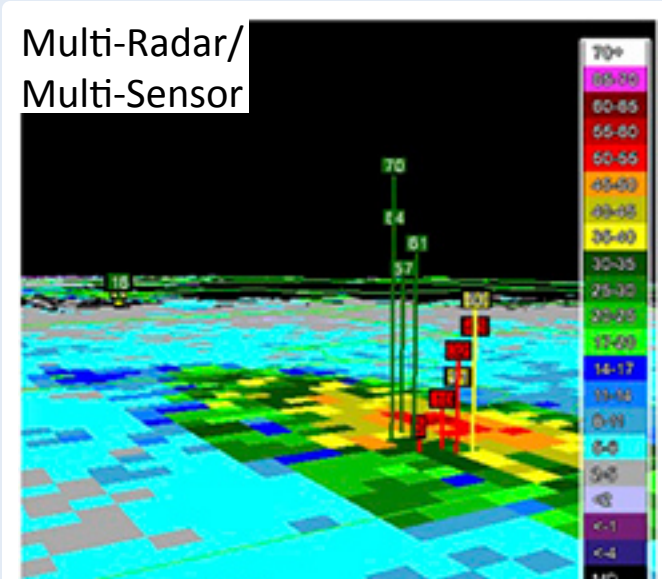
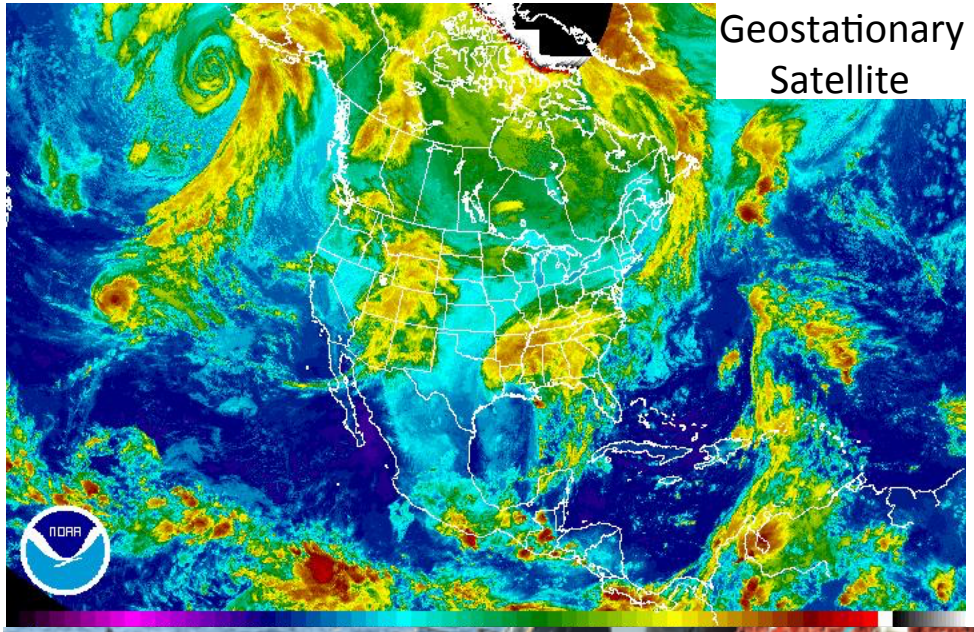
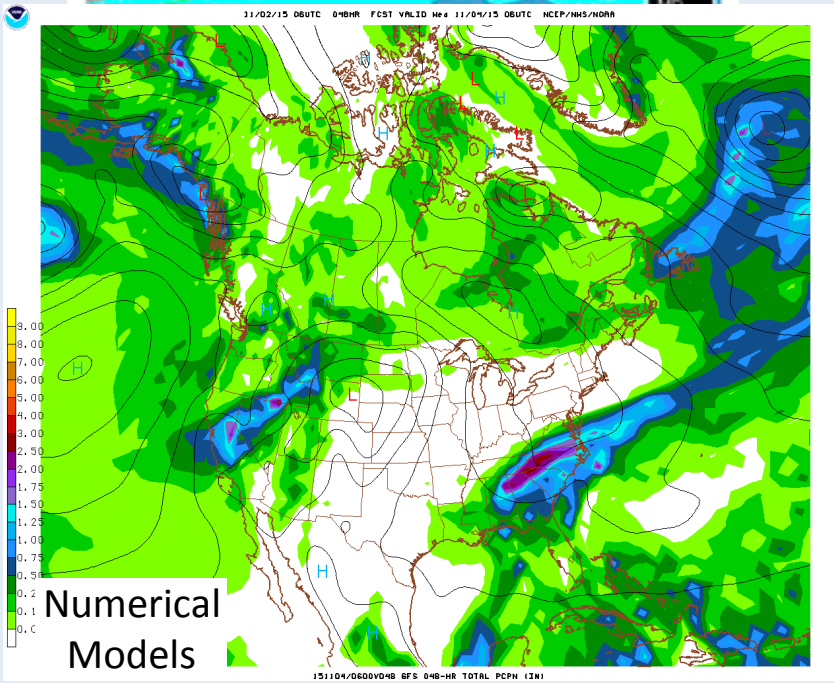


Figure from S. Ansari et al, 2016

Future Datasets for BDP





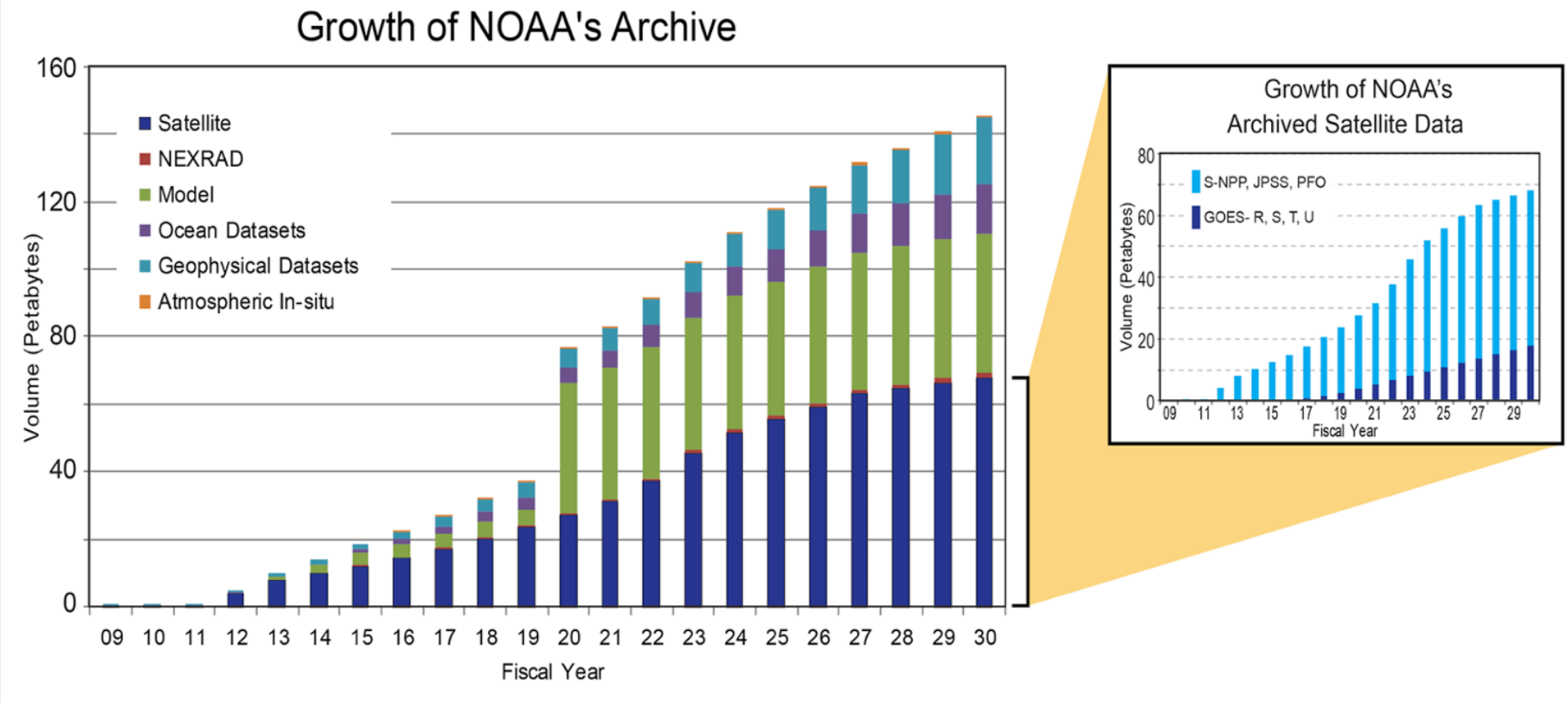


jeff.deLaBeaujardiere@noaa.gov

Role #5: Long-Term Preservation

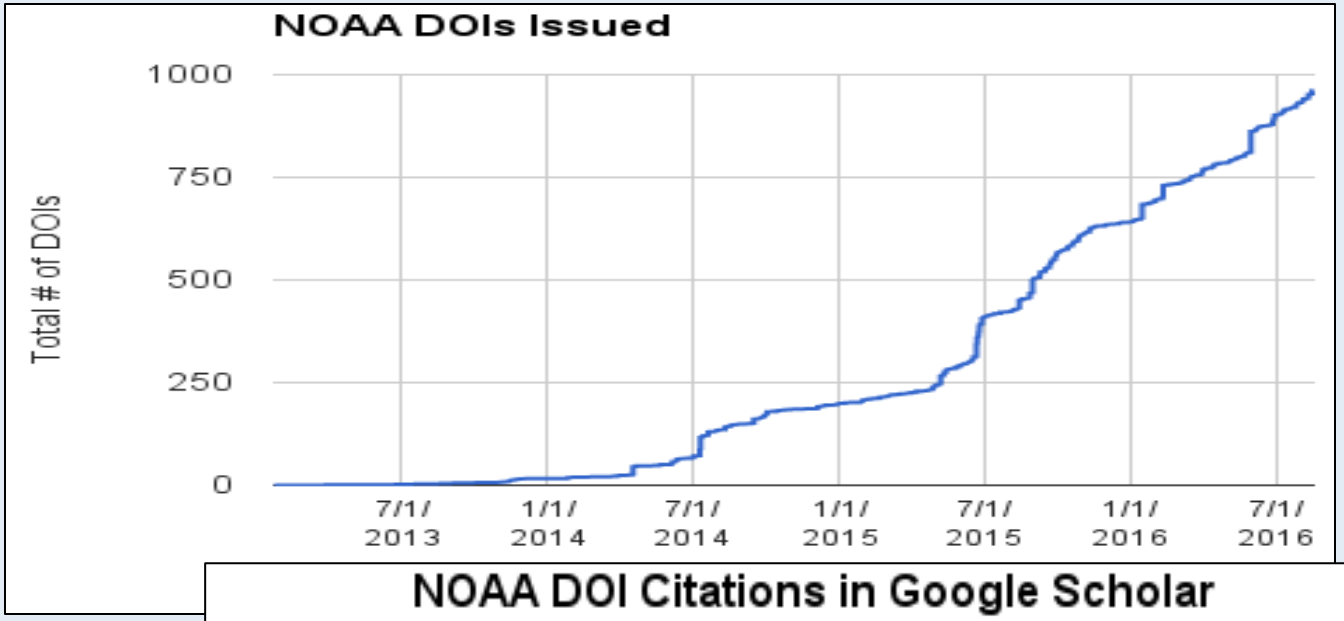
- Ensure original data remain available 75+ years
- Perform updates, format migrations, etc as needed

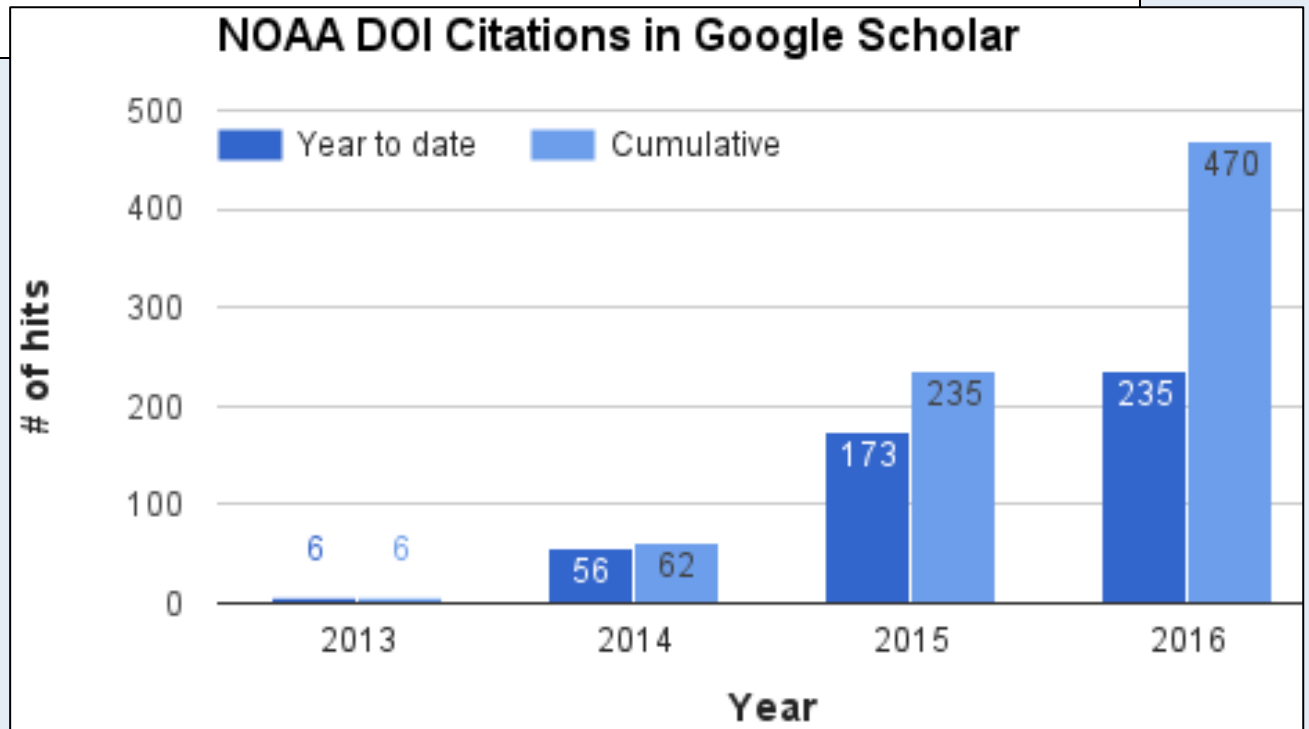
jeff.deL



Graphic courtesy Steve Del Greco & Ken Casey, NOAA/ NCEI

Dataset Identifiers & Citations for Archival Data





Collaborators: NOAA Data Citation WG

Jeff.deLaBeaujardiere@noaa.gov

Role #6: Research Funding

$

Researchers

Data

make data accessible

Publications

deposit manuscripts

NOAA Institutional Repository

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

February 22, 2013

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

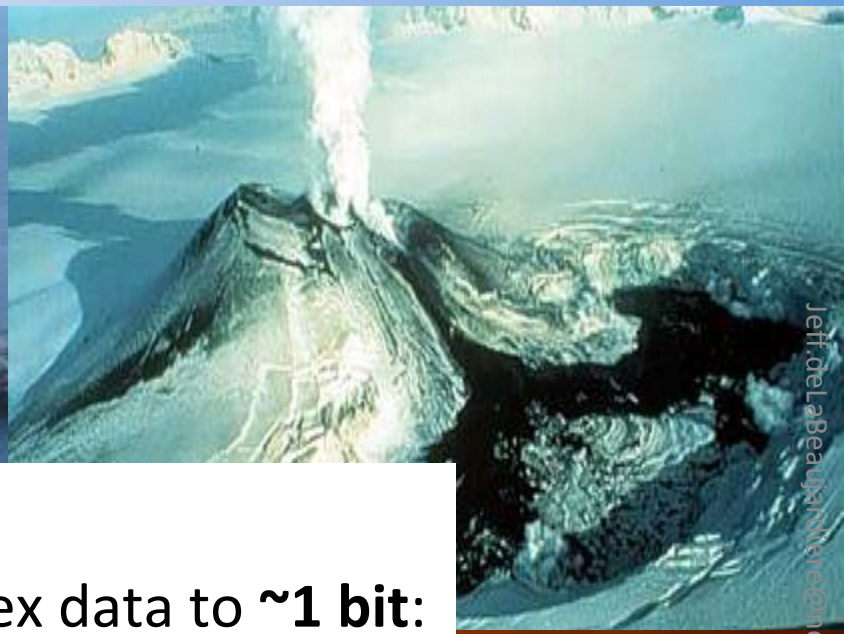
jeff.deLaBeaujardiere@noaa.gov

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

jeff.deLaBeaujardiere@noaa.gov

Users need answers, not data

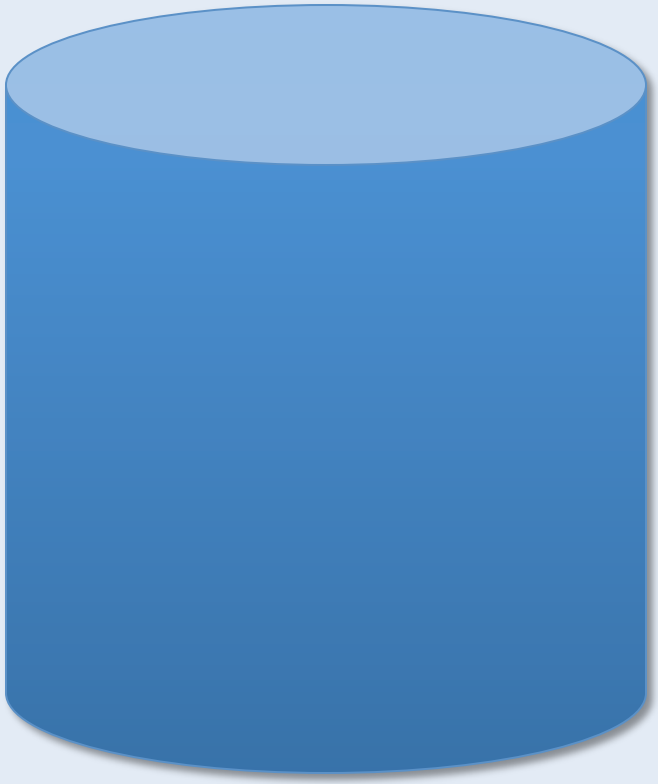


Data to Decisions:

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

Jeff.deLaBeaujardiere@noaa.gov

Challenges



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

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

$$\mathbf{E} = -\mathbf{V}\times\mathbf{B} + \frac{1}{en}\frac{(1 - Zm_e/m_i)}{(1 + Zm_e/m_i)}\mathbf{J}\times\mathbf{B} + \eta\mathbf{J}$$
$$+ \frac{1}{\varepsilon_0\omega_p^2}\left[\frac{\partial\mathbf{J}}{\partial t} + \nabla\cdot(\mathbf{JV} + \mathbf{VJ}) + \sum_{\alpha=i,e}\frac{q_\alpha}{m_\alpha}(\nabla p_\alpha + \nabla\cdot\Pi_\alpha)\right]$$

$$\frac{3}{2}\left(\frac{\partial}{\partial t} + \mathbf{V}_\alpha\cdot\nabla\right)p_\alpha = -\frac{5}{2}p_\alpha\nabla\cdot\mathbf{V}_\alpha - \nabla\cdot\mathbf{q}_\alpha - \Pi_\alpha : \nabla\mathbf{V}_\alpha + Q_\alpha, \quad \alpha = i,e$$

Data Volume

Data Complexity

Jeff.deLaBeaujardiere@noaa.gov

Wish #1: Fully Leverage the Cloud

Operational Customers
(e.g., NWS)

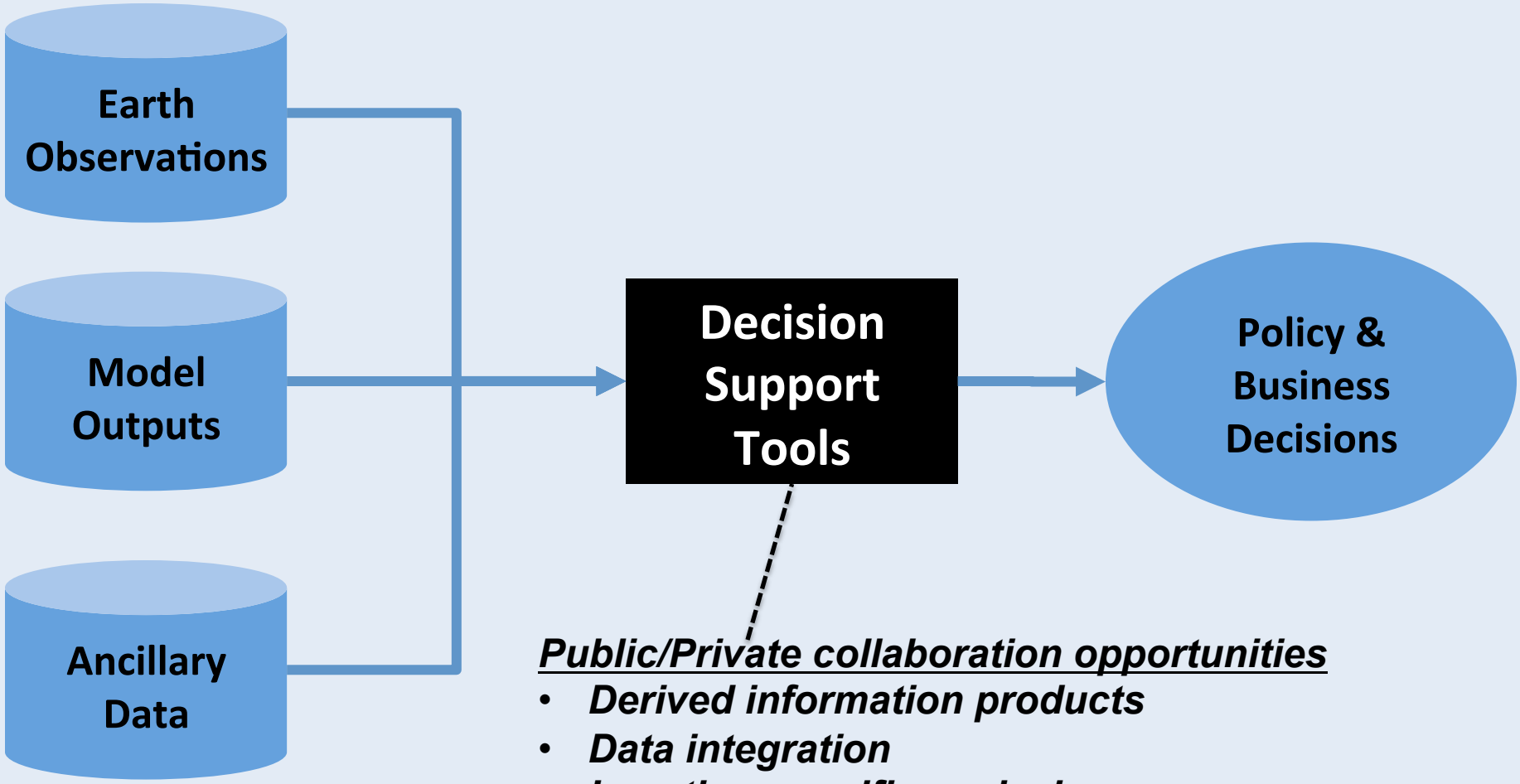
Archive

Cloud

Other Users

Jeff.deLaBeaujardiere@noaa.gov

Wish #2: Enable Better Decision-making



Public/Private collaboration opportunities

- ***Derived information products***
- ***Data integration***
- ***Location-specific analysis***
- ***Use Case-specific analysis***
- ***Statistics & Trends***
- ***What-if scenarios***

complicated science data

non-scientist users

Jeff.deLaBeaujardiere@noaa.gov

Thank you!

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