



NOAA Data Management Policies & Implementation

**GEO Plenary Side Event: Best Practices in Data
Management Policy & Implementation
2017-10-23**

Jeff de La Beaujardière, PhD

National Oceanic and Atmospheric Administration

NOAA Data Management Architect

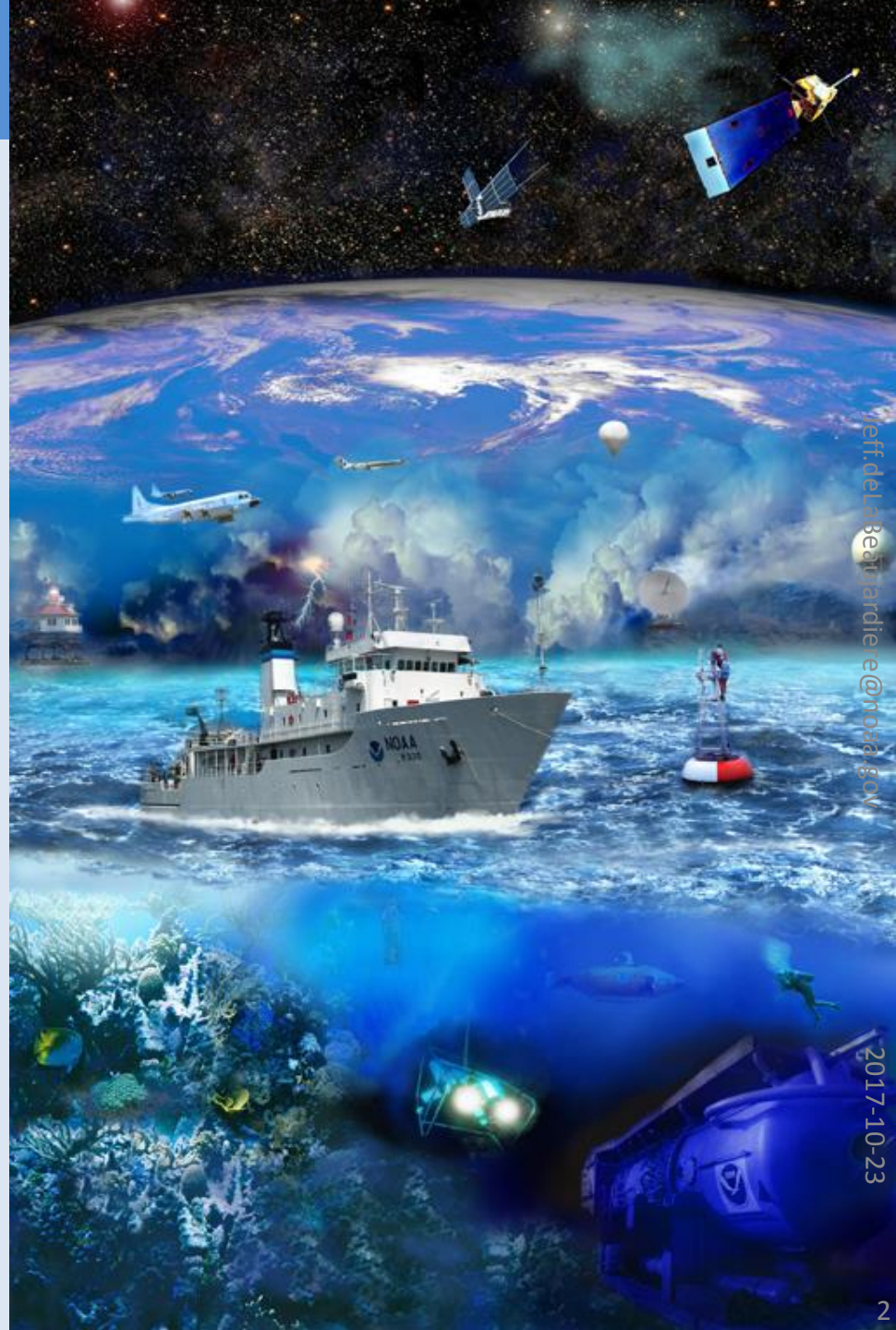
jeff.deLaBeaujardiere@noaa.gov

NOAA has "Big Data"

(Volume, Variety, Velocity, ...)

- **Satellites**
- **Weather radars**
- **Ocean bathymetry**
- **Buoy networks**
- **Tide gauges**
- **Ships**
- **Aircraft**
- **Autonomous vehicles**
- **Human observers**
- **Numerical models**

These data are unique, valuable, irreplaceable, and collected at public expense



jeff.delaBeaulardie@noaa.gov

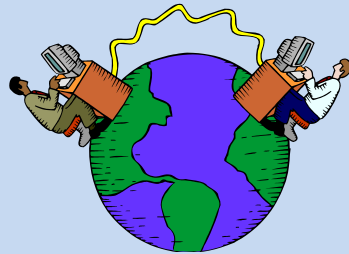
2017-10-23

Vision for NOAA Data Management

All NOAA environmental data shall be



Discoverable



Accessible



Usable



Preserved

for all types of users and applications.

NOAA Data Policies

<https://nosc.noaa.gov/EDMC/>

**NOAA Administrative Order 212-15:
Management of Environmental Data**
(2010)

NOAA Environmental Data Management Framework
(2012-2013)

**Data Management Planning
Directive**
(2011; rev. 2015)

**Data Documentation
Directive**
(2011; rev. 2016)

Data Access Directive
(2015)

Archive Appraisal Procedure
(2008)

Data Citation Directive
(2015)

**Data Sharing Directive
for NOAA Grantees**
(2012; rev. 2016)

NOAA Data Catalog (est. 2013)



data.noaa.gov

/ Datasets

Organizations

National Oceanic an... (65734)

Groups

There are no Groups that match this search

Tags

DOC/NOAA/NESDIS/NCE... (46157)

oceanography (24639)

DOC/NOAA/NESDIS/NOD... (24589)

EARTH SCIENCE > OCE... (22120)

EARTH SCIENCE > Oce... (21540)

DOC/NOAA/NESDIS/NGD... (21107)

In Situ Ocean-based... (20448)

CONTINENT > NORTH A... (18149)

EARTH SCIENCE > Oce... (17954)

Hydrographic Survey...

Search datasets...

65,734 datasets found

Order by:

*Collaborators: Chris MacDermaid,
NOAA Catalog WG*

NOAA JPSS Visible Infrared Imaging Radiometer Suite (VIIRS) Sensor Data Recorder...

Sensor Data Records (SDRs), or Level 1b data, from the Visible Infrared Imaging Radiometer Suite (VIIRS) are the calibrated and geolocated radiance and reflectance data produced...

[HTML](#)

Gravity Data for the Vernal, U

The gravity station data (4,778 records) from the International Gravity Field (IGF) academia) using a variety of methods...

[HTML](#)

Ekman Upwelling, METOP A

NOAA CoastWatch distributes near real-time Ekman upwelling data. This data is derived from the SeaWiFS satellite...

[HTML](#)

[WCS](#)

[WMS](#)

Fluorescence, Aqua MODIS,

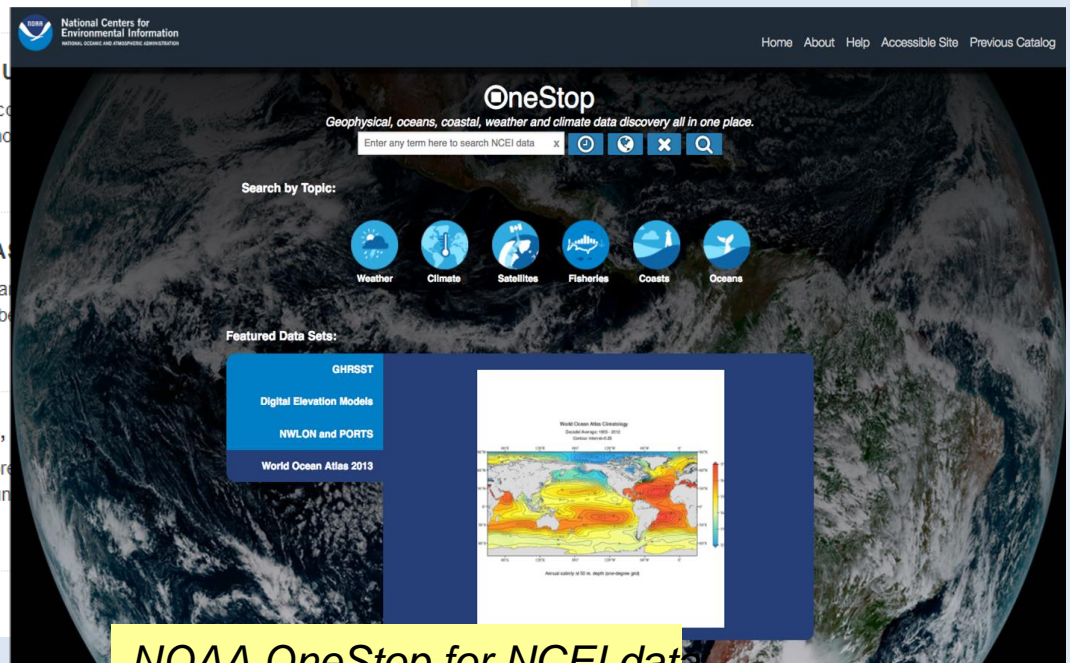
MODIS measures chlorophyll fluorescence from space. When phytoplankton are under stress, they emit fluorescence...

[HTML](#)

[WCS](#)

[WMS](#)

ICODAS 1-degree Standard




*NOAA OneStop for NCEI data
(Ken Casey, OneStop team)*

Data Access Support

- ERDDAP
 - Hosts & serves gridded & tabular data

Bob Simons, Roy Mendelssohn (NMFS)

**ERDDAP**
Easier access to scientific data

Brought to you by [NOAA NMFS](#) [SWFSC](#) [ERD](#)

ERDDAP

ERDDAP is a data server that gives you a simple, consistent way to download subsets of gridded and tabular scientific datasets in common file formats and make graphs and maps. This particular ERDDAP installation has oceanographic data (for example, data from satellites and buoys).

Easier Access to Scientific Data

Our focus is on making it easier for you to get scientific data.

Different scientific communities have developed different types of data servers, for example, OPeNDAP, WCS, SOS, OBIS, and countless custom web pages with forms. Each is great on its own. Without ERDDAP, it is difficult to get data from different types of servers:

- Different data servers make you format your data request in different ways.
- Different data servers return data in different formats, usually not the common file format that you want.
- Different datasets use different formats for time data, so the results are hard to compare.


ERDDAP unifies the different types of data servers so you have a consistent way to get the data you want, in the format you want.

- ERDDAP acts as a middleman between you and various remote data servers.

Start Using ERDDAP: Search for Interesting Datasets

- [View a List of All 1,338 Datasets](#)
- **Do a Full Text Search for Datasets**
- **Search for Datasets by Category**

Datasets can be categorized in different ways by the values of various metadata attributes. Click on an attribute ([cdm_data_type](#), [institution](#), [loos_category](#), [keywords](#), [long_name](#), [standard_name](#), [variableName](#)) to see a list of categories (values) for that attribute. Then, you can click on a category to see a list of relevant datasets.
- **Search for Datasets with Advanced Search**

**Global Earth Observation Integrated Data Environment**
Unified Access Framework for Environmental Data

Home Access Data Find Data Contribute Data Learn About UAF UAF in Action NOAA EDM Wiki Contact Us

Home
Access Data
Find Data
Contribute Data
Learn About UAF
UAF in Action
NOAA EDM Wiki
Contact Us

What is the Unified Access Framework (UAF)?

UAF is a NOAA-wide effort to make environmental datasets easy to find and use. It is an important contribution to realizing the vision of [NOAA's Global Earth Observation - Integrated Data Environment \(GEO-IDE\) Initiative](#).

UAF follows the design philosophy: *Don't solve problems - Copy success.*

In its first year it has been exploiting the opportunities for gridded data access and interoperability that have been created by wide-spread use of a suite of open [standards and technologies](#).

Highlighted Data Set and Data Provider:

Group for High-Resolution Sea Surface Temperature (GHR SST) data sets are now available as part of the UAF project courtesy of the National Oceanographic Data Center (NODC).







GHR SST provides high-resolution (<10km) sea surface temperature products to the oceanographic, meteorological, climate and general scientific community.

[Learn more about NODC](#) [Start exploring GHR SST data](#)

What is GEO-IDE?

[NOAA's Global Earth Observation - Integrated Data Environment \(GEO-IDE\)](#) is framework for engaging NOAA's environmental data management communities through sharing expertise, experience, and information.

Interoperability Provided By:



Unified Access Framework (UAF)

– Serves netCDF data on THREDDS & ERDDAP servers

*Kevin O'Brien,
Eugene Burger
(OAR)*

leffdel@Reaunardiere@noaa.gov

2017-09-27

Dataset Identifier Project

DOI
(Digital
Object
Identifier)

landing
page

**Data &
Metadata**

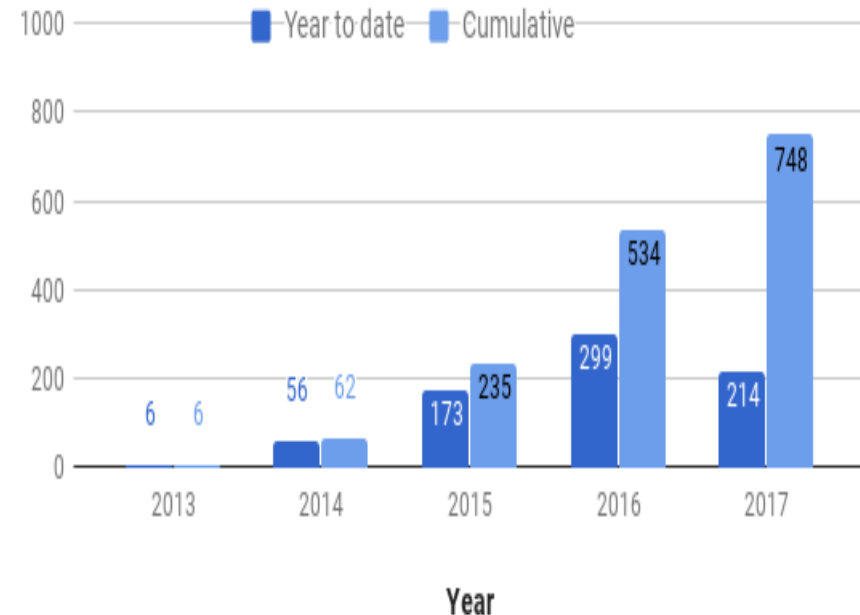
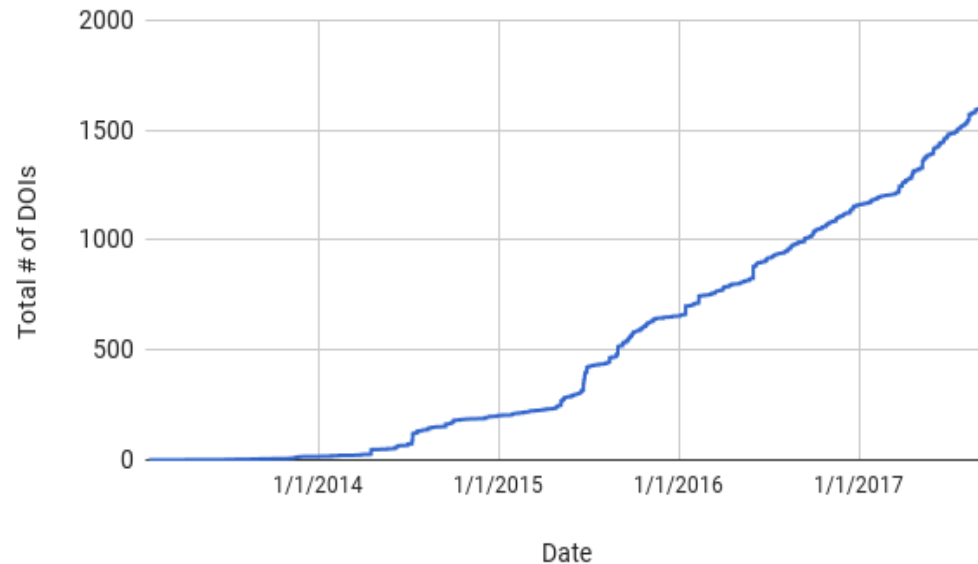
DOI benefits:

- Permanent, citable ID.
- International standard (ISO 26324).
- Recognition by publishers.

Jeff.deLaBeaujard

NOAA DOI Citations in Google Scholar

NOAA DOIs Issued



Data Management is not the goal

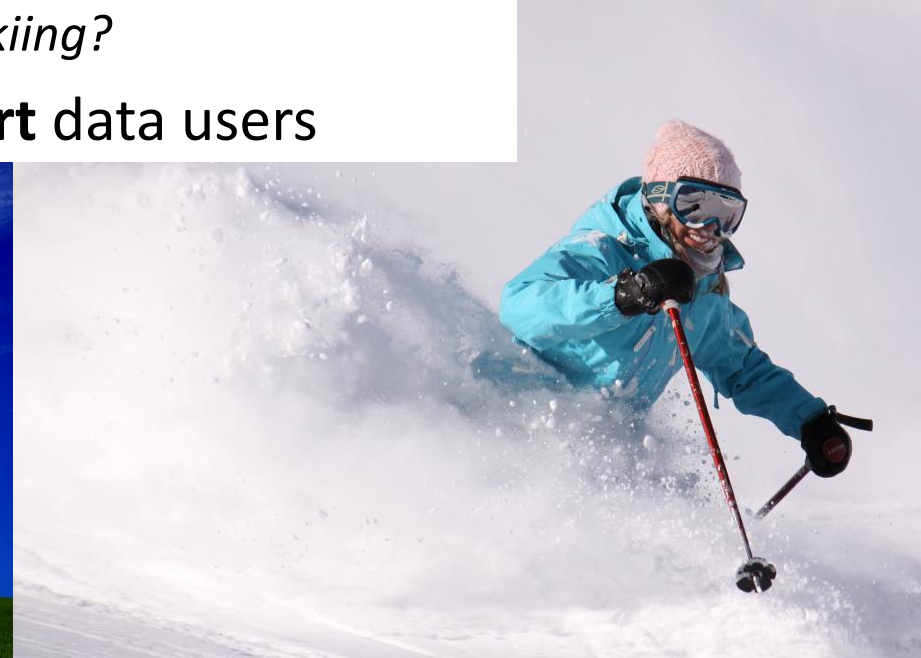
We don't want to just "manage" data.
We want to use and reuse data, and
extract maximum value from it.

Users need answers, not huge datasets (... or 100s of tiny datasets)

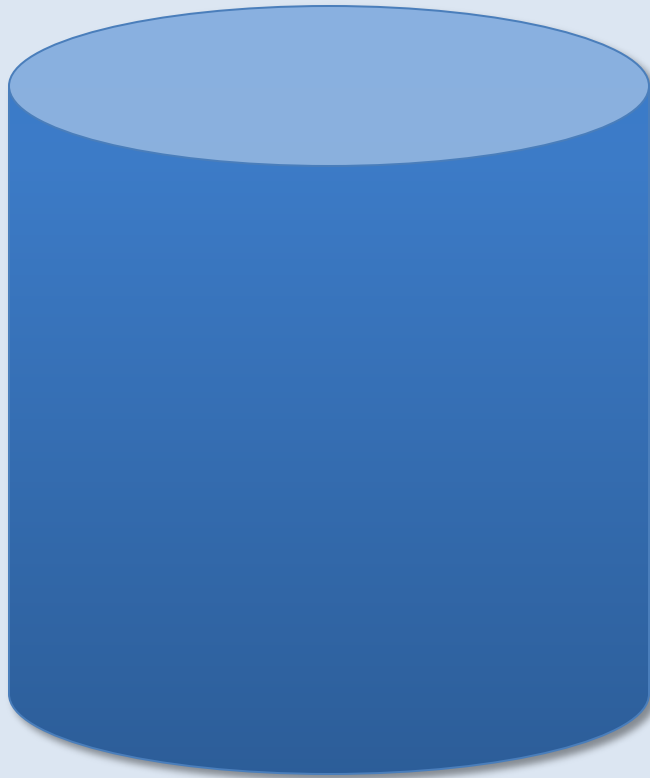


Data to Decisions:

- Distill huge & complex data to **~1 bit**:
plant crop? evacuate?
build wind farm? go skiing?
- Support **non-expert** data users



Challenges



**Data
Volume**



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

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

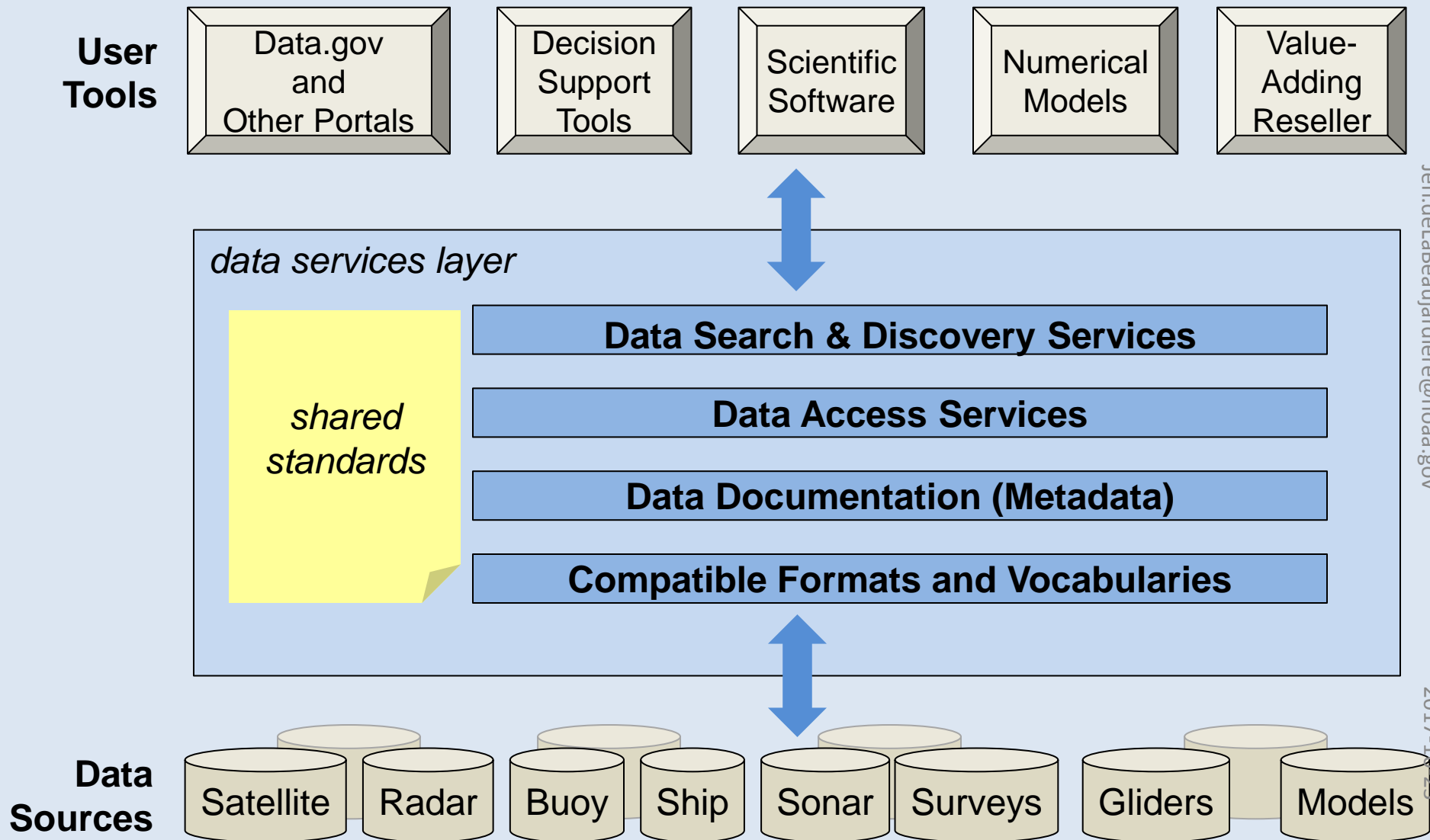
$$\mathbf{E} = -\mathbf{V} \times \mathbf{B} + \frac{1}{en} \frac{(1 - Zm_e/m_i)}{(1 + Zm_e/m_i)} \mathbf{J} \times \mathbf{B} + \eta \mathbf{J}$$

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

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

**Data
Complexity**

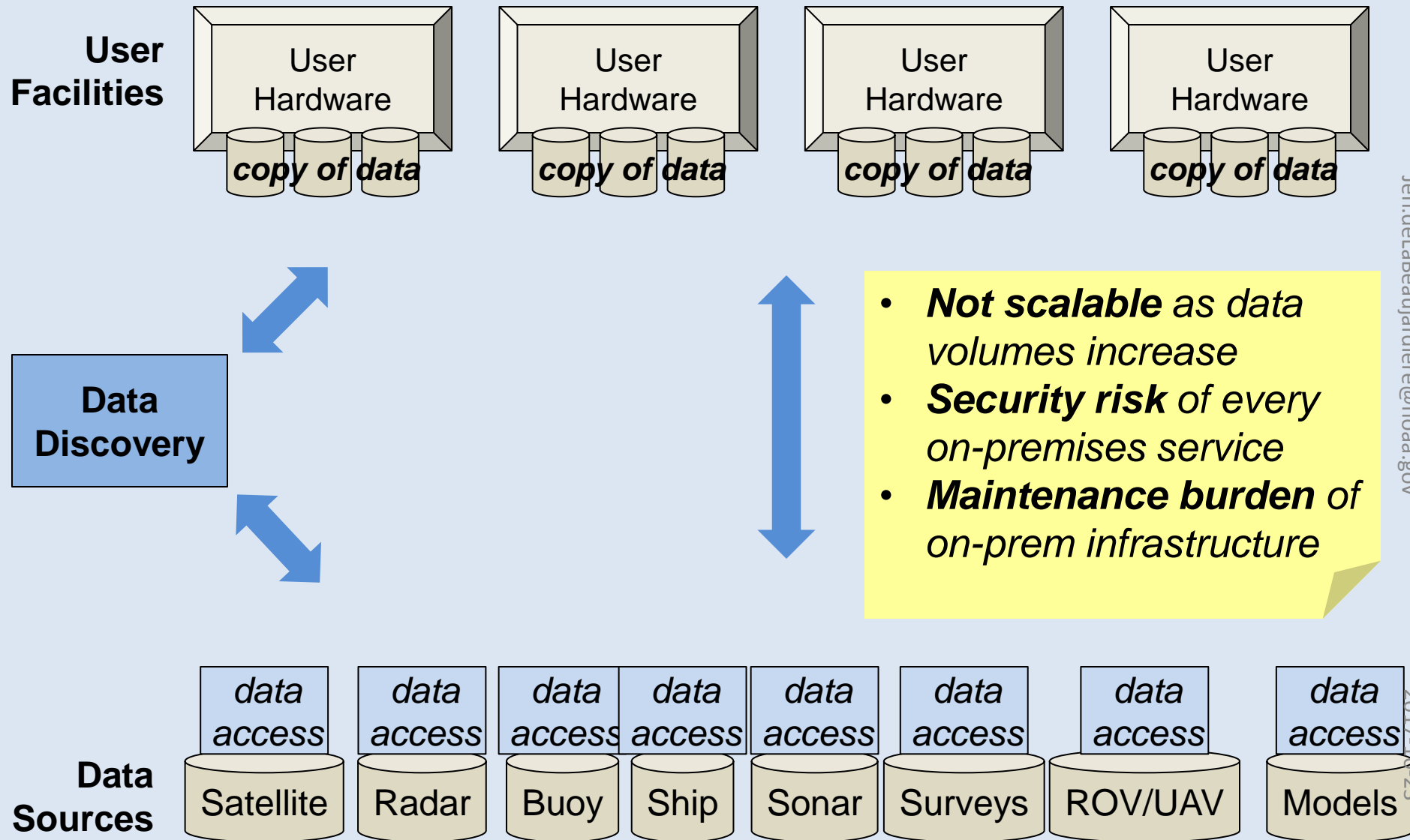
Traditional Data Services Approach - theory



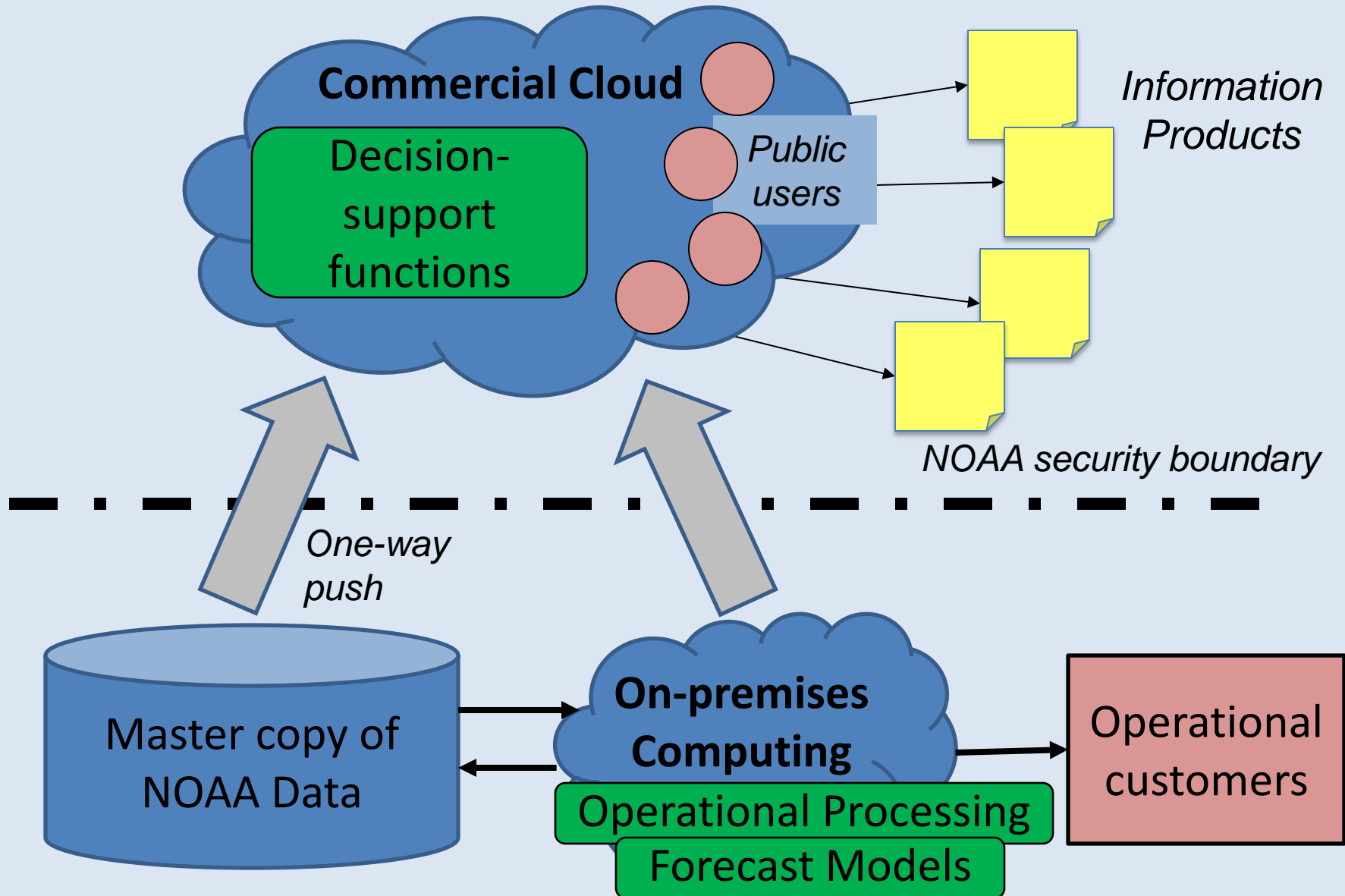
Traditional Data Services Approach - reality

Jeff.deLaBeaujardiere@noaa.gov

2017-10-23

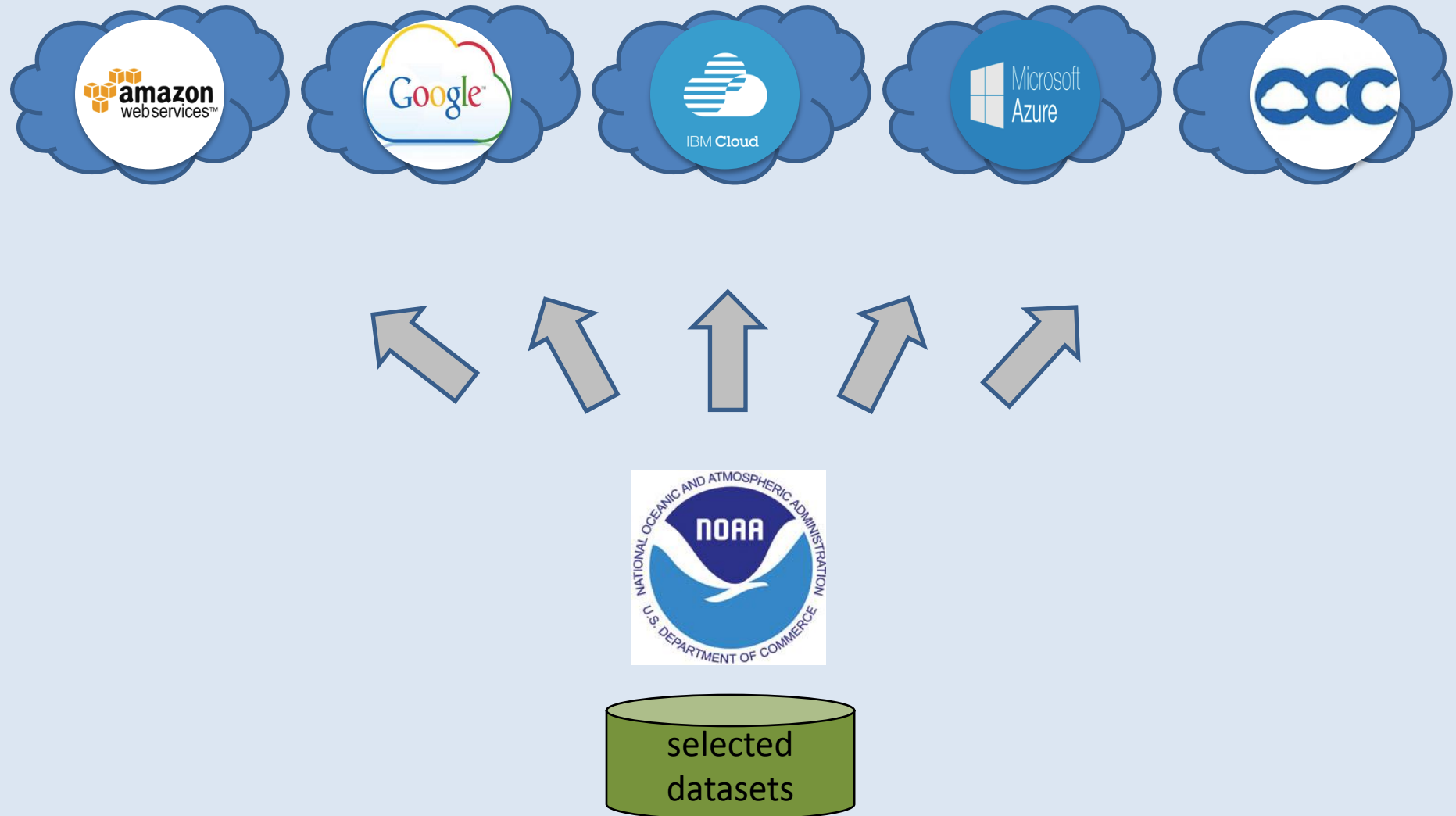


Notional Cloud Deployment Scenario

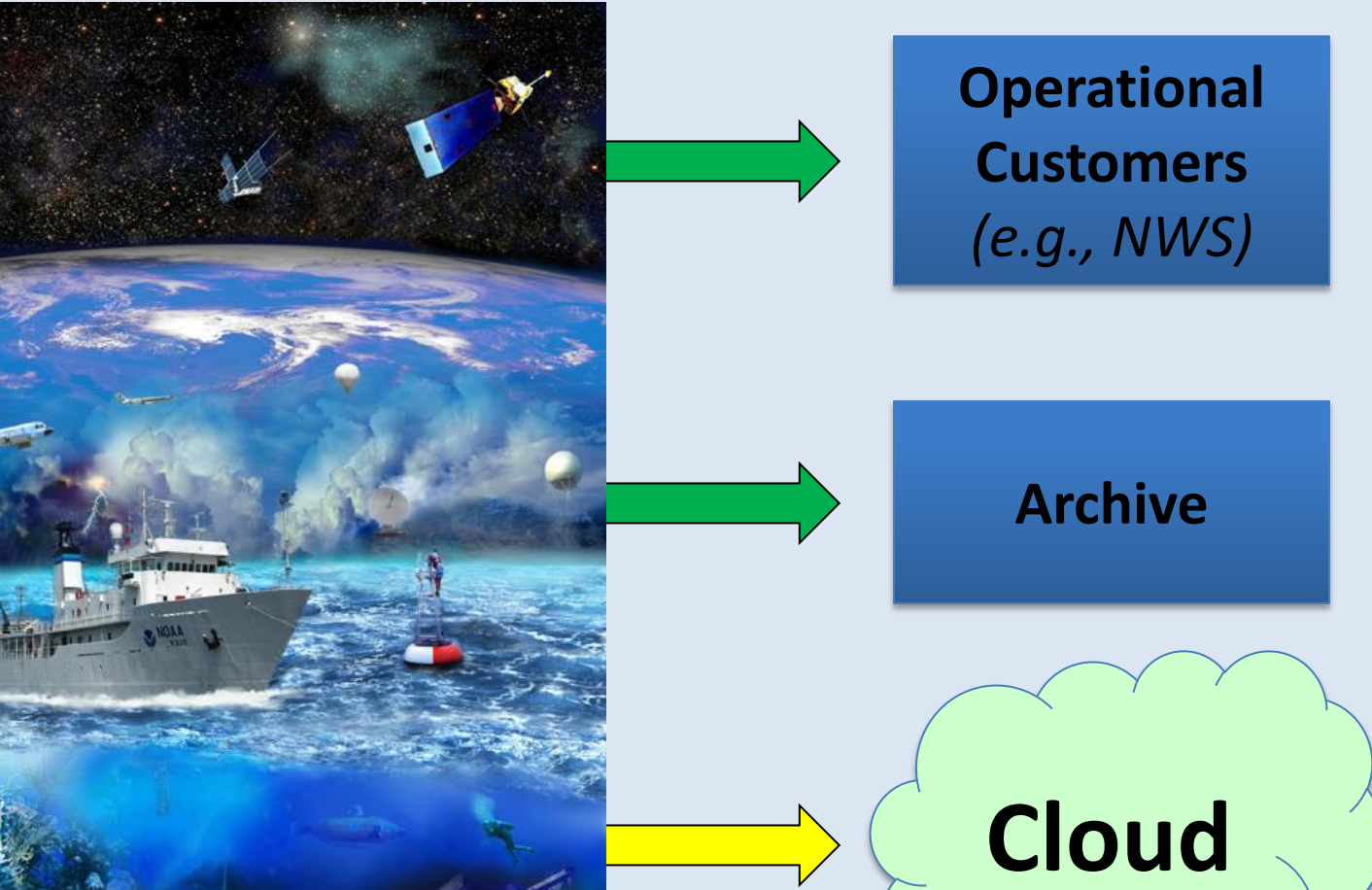


NOAA Big Data Project (R&D)

www.noaa.gov/big-data-project



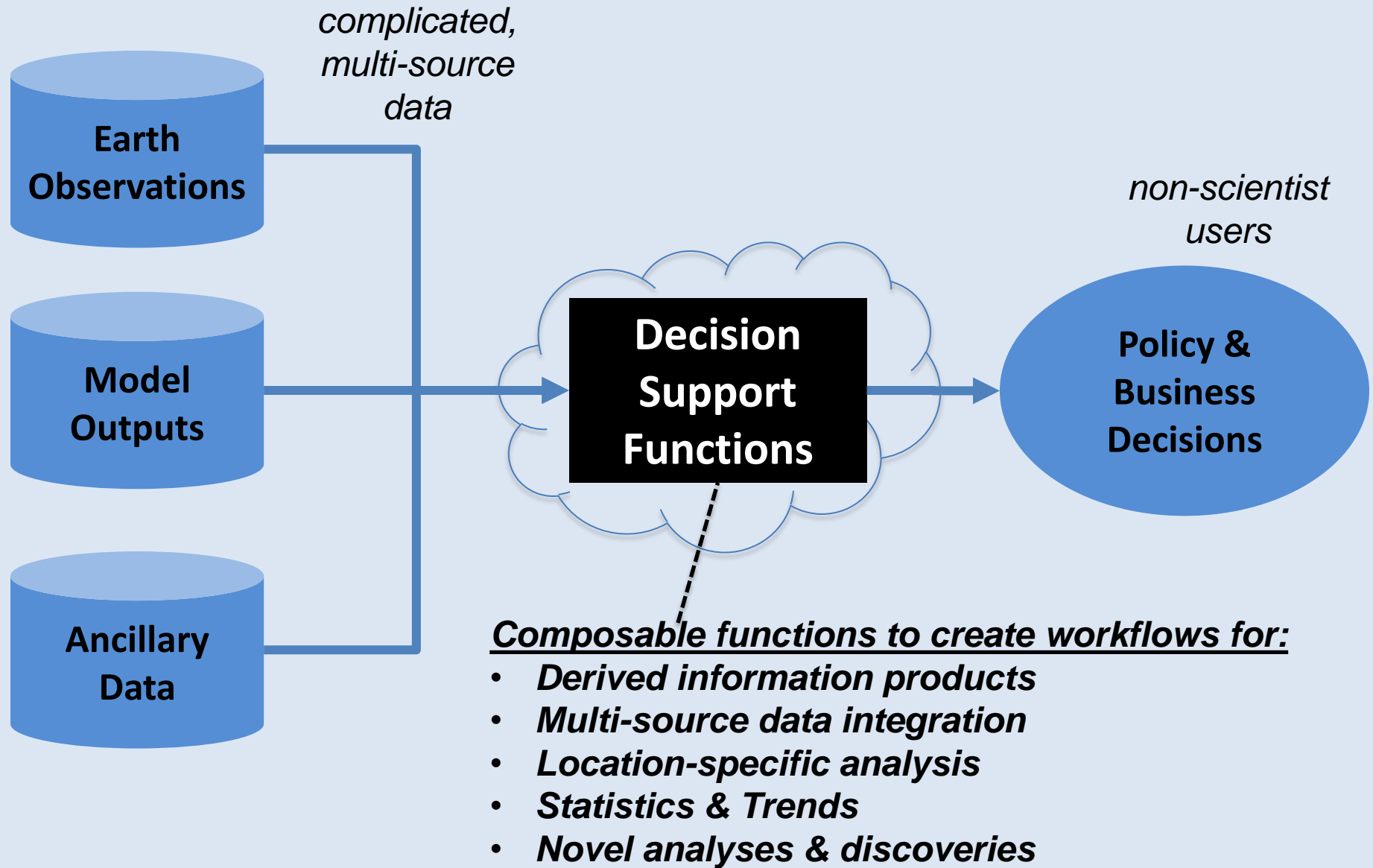
Wish #1: Full Use of the Cloud



Challenges:

- Egress costs vs free data
- Uncertain/unbounded costs
- Re-architecting for performance vs fork-lifting existing apps
- IT security policy mismatch

Wish #2: Composable Functions for Decision Support



Questions?

Jeff de La Beaujardière, PhD

jeff.deLaBeaujardiere@noaa.gov

