



**NOAA EDM**

# **NOAA Environmental Data Management**

**Update for Unidata SAC 2014-10-08**

Jeff de La Beaujardière, PhD

NOAA Data Management Architect

[jeff.deLaBeaujardiere@noaa.gov](mailto:jeff.deLaBeaujardiere@noaa.gov)

+1 301-713-7175

# NOAA data are unique, valuable, and irreplaceable

- ***Many observing systems:***
  - 10 satellites
  - 3 buoy networks
  - 120+ weather radars
  - 200+ tide gauges
  - 17 ships
  - 10 aircraft
  - 5 supercomputers
  - human observers
  - animal telemetry
- ***Scope is from the bottom of the ocean to the surface of the Sun***

- ⇒ *Wide variety of data collected for different purposes*
- ⇒ *Many formats & dissemination methods*
- ⇒ *Operational & legacy systems*



# Vision for NOAA Data Management

All NOAA environmental data are to be



**Discoverable**



**Accessible**



**Usable**



**Preserved**

**for all types of users and applications.**

## Authorities:

- *NOAA Administrative Order 212-15 (2010)*
- *OMB Open Data Policy (2013)*
- *OSTP Public Access to Research Results memo (2013)*

# EDMC Procedural Directives

<https://www.nosc.noaa.gov/EDMC/PD.all.php>

## **Data Management Planning**

Describe how you will preserve, document and distribute your data.

*(2011) In revision 2014*

## **Archive Approval**

Data Center process for approving archive requests.

*(2008) Reviewed 2014; no change*

## **Data Documentation**

How to apply ISO 19115 metadata for discovery, use & understanding.

*(2011; no change)*

## **Data Sharing by Grantees**

Grantees write data sharing plan, and share data within 2 years.

*(2012; to be revised 2015 per PARR)*

## **Data Citation**

Assign persistent identifiers to datasets and encourage citation.

*(in preparation 2014)*

## **Data Access**

Make data accessible, preferably via on-line services. REcommend specific service and formats for particular classes of data.

*(in preparation 2014)*

## **Ocean Data Acquisitions**

Conversion of NAO 216-101 to EDMC PD

*(1990; in preparation 2015)*

# NOAA Environmental Data Management Framework

## Data Management Framework

Principles				
Governance				
Resources				
Standards				
Architecture				
Assessment				
Data Lifecycle	Data Lifecycle	Data Lifecycle	Data Lifecycle	Data Lifecycle

Purpose:  
To organize, guide and support NOAA environmental data management activities

<https://www.nosc.noaa.gov/EDMC/framework.php>

# White House Policies (2013)

## Public Access to Research Results

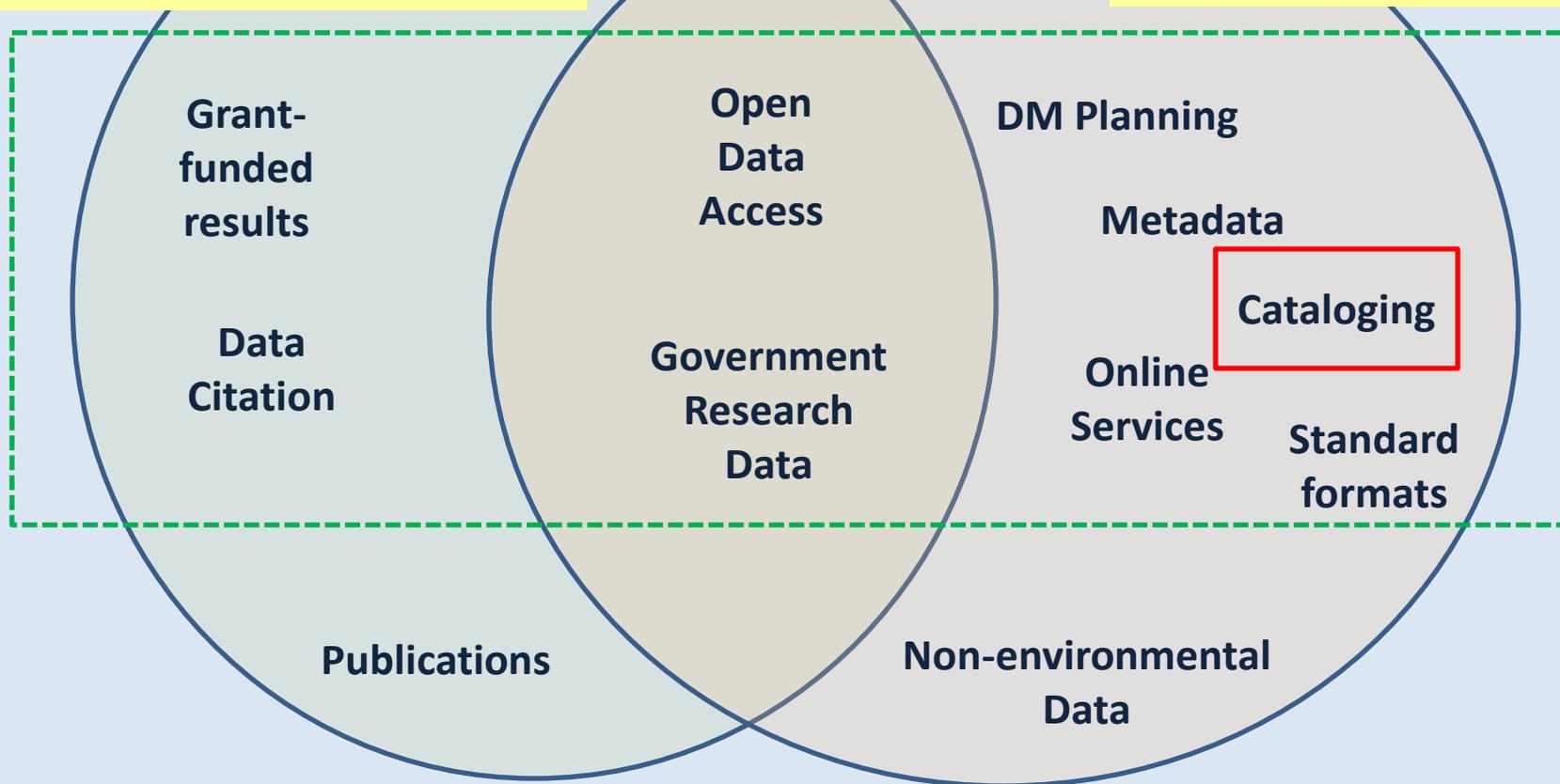
(OSTP memo 2013-02-22)

[http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)

## Open Data Policy

(Exec. Order 2013-05-09)

<http://www.whitehouse.gov/sites/default/files/es/omb/memoranda/2013/m-13-13.pdf>



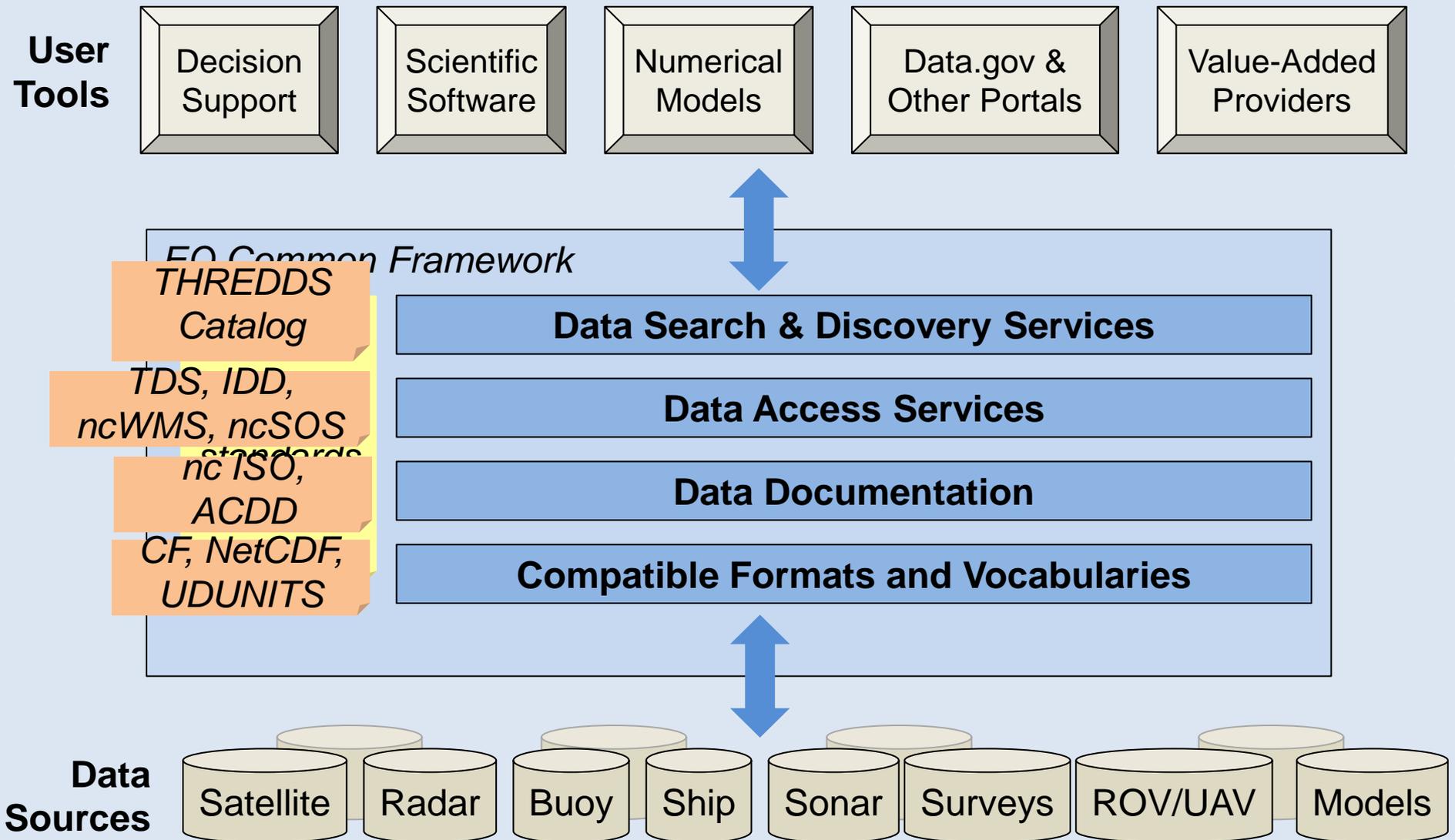
# NOAA PARR Plan Highlights *(draft pending OSTP approval)*

*(per OSTP Public Access to Research Results Memorandum)*

- States NOAA will **continue existing EDM efforts** to ensure data are accessible, usable, and archived.
- Assigns responsibility to **NOAA Programs and PMs** to properly manage data they produce.
- Requires **grant programs** to include summary DM Plan in announcements for data likely to result from grant.
  - Consider whether to archive at NOAA National Data Center
- Requires **grantees** to include project DM Plan in proposals.
- Requires **submission of final manuscripts** to NOAA Central Library.
  - Visible after 1 year embargo
- ***If plan approved, new provisions take effect in:***
  - *FY 2016 for new/current intramural data and publications*
  - *FY 2017 for new extramural publications*
  - *FY 2018 for new extramural data*

# Goal: "Earth Observations Common Framework"

*Any user tool able to connect to any Earth Observation data source*



data.noaa.gov

+ Add Dataset

Established  
Nov 2013

Home / Datasets

**NOTE:** This NOAA Data Catalog is a prototype under active development. Availability and completeness are not guaranteed.

Filter by location Clear  
Enter location...



Map data CC-BY-SA by [OpenStreetMap](#)  
Tiles by [MapQuest](#)

Search datasets...

**55,513 datasets found**

Order by: Last Modified

Harvests from  
existing  
metadata  
collections

Inclusion  
mandatory  
per US Open  
Data Policy

**PRESSURE - WATER and Other Data from COLUMBUS CALIFORNIA  
Other Platforms from 19940525 to 19941224 (NODC Accession 9500**

*National Oceanic and Atmospheric Administration —*

[HTML](#) [HTML](#) [HTML](#) [HTML](#)

**Delayed XBT Data assembled by US NOAA Atlantic Oceanographic  
Meteorological Laboratory (AOML) for the Global Temperature-Salin  
Profile Program (GTSP) and submitted the month of January 2004 (  
Accession 0001336)**

*National Oceanic and Atmospheric Administration — XBT data were collected from MULTIPLE PLATFORMS from a  
World-Wide distribution. Data were collected by the Atlantic Oceanographic and Meteorological*

[HTML](#) [HTML](#) [HTML](#) [HTML](#)

On Amazon  
Federal  
GeoCloud

Starting to  
compute  
statistics (e.g.,  
# with data  
access URL)

Dataset Type Clear All

geospatial (55512)

Show More Dataset Type

Tags Clear All

earth science (24316)

oceans (23481)

**Oceanographic and surface meteorological data collected from sta  
by Everglades National Park (ENP) and assembled by Southeast Coa  
Ocean Observing Regional Association (SECOORA) in the Coastal W  
Florida and North Atlantic Ocean from 2014-02-13 to 2014-08-31 (NODC**

jeff.deLaBeaujardiere@noaa.gov

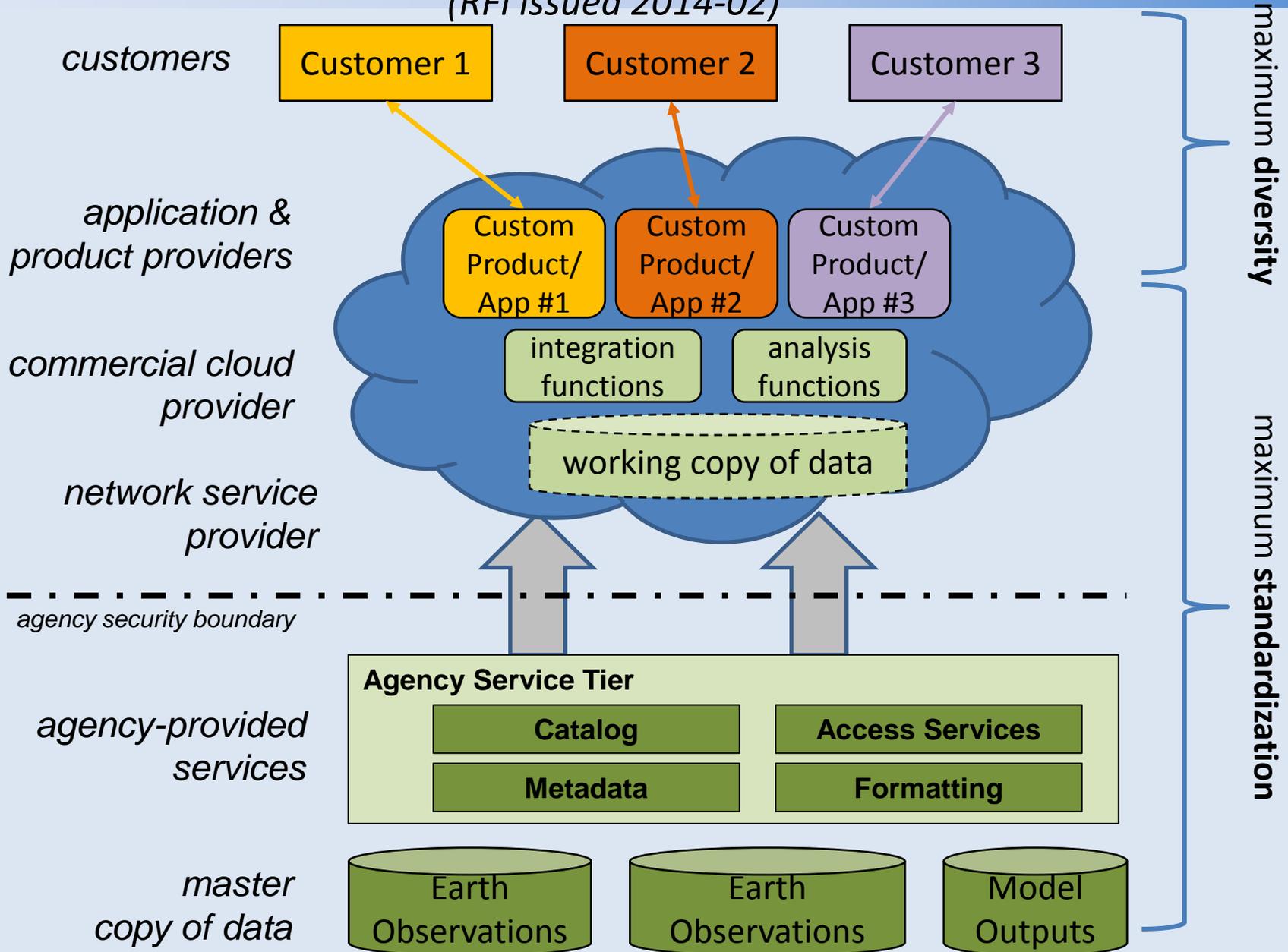
2014-09-09

# Data Accessibility Activities

- **NWS Integrated Dissemination Program (IDP) Enterprise Geospatial Services**
  - Operational hosting at NCEP for NWS & NOAA
- **Unified Access Framework**
  - THREDDS Data Servers
  - CF conventions for *in situ* data
  - OpenDAT/Unidata Linked Servers (OPULS) grant
- **IOOS THREDDS & Sensor Observation Services**
- **Data Center Cloud Pilot**
- **Big Data Partnership RFI ( $\neq$  *BEDI*)**
  - Copy of NOAA data in Cloud with computing capability
  - 1<sup>st</sup> RFI issued Feb 2014; 2<sup>nd</sup> RFI issued
  - Industry Day 2014-10-17
  - See [https://www.fbo.gov/index?s=opportunity&mode=form&id=cdbfd2f6b096dfe93aecae44b67fcc40&tab=core&\\_cview=1](https://www.fbo.gov/index?s=opportunity&mode=form&id=cdbfd2f6b096dfe93aecae44b67fcc40&tab=core&_cview=1)

# Conceptual Model of NOAA Big Data Partnership

(RFI issued 2014-02)



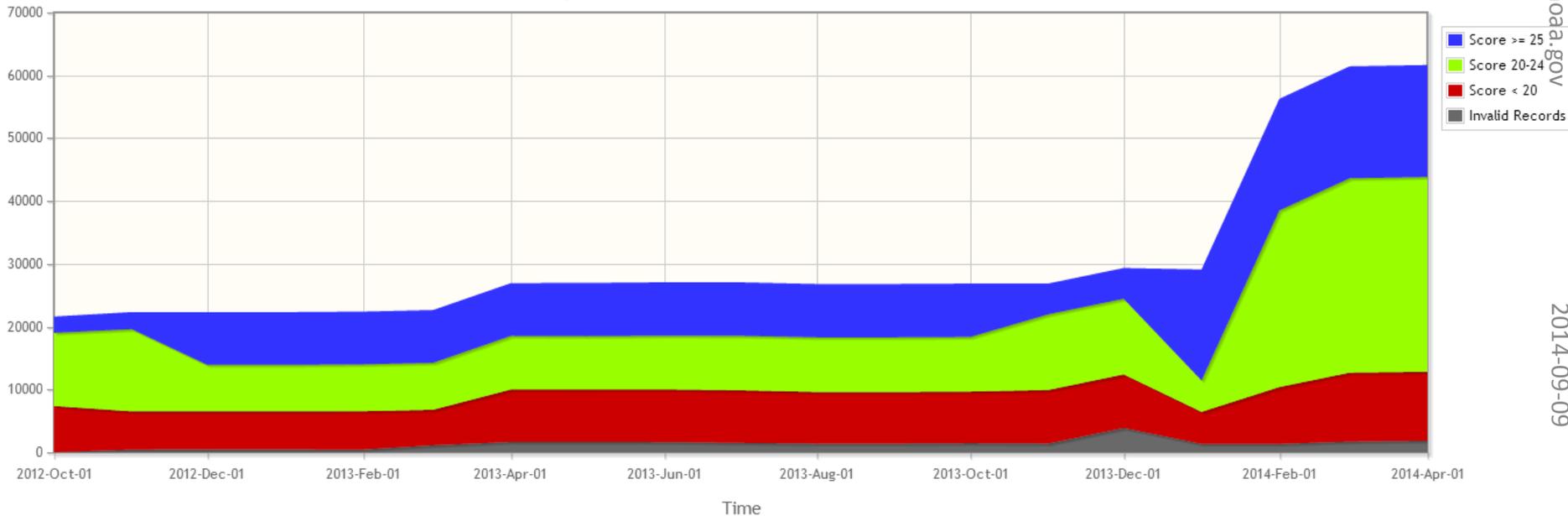
# Data Usability Activities

- **Metadata training webinars**
- **Conversion to ISO metadata standard at NODC**
- **ISO export of NMFS InPort Metadata**
- **ATRAC metadata editor at NCDC**
- **Metadata metrics & diagnostics at NGDC**

jeff.delalaBeaujardiere@noaa.gov

2014-09-09

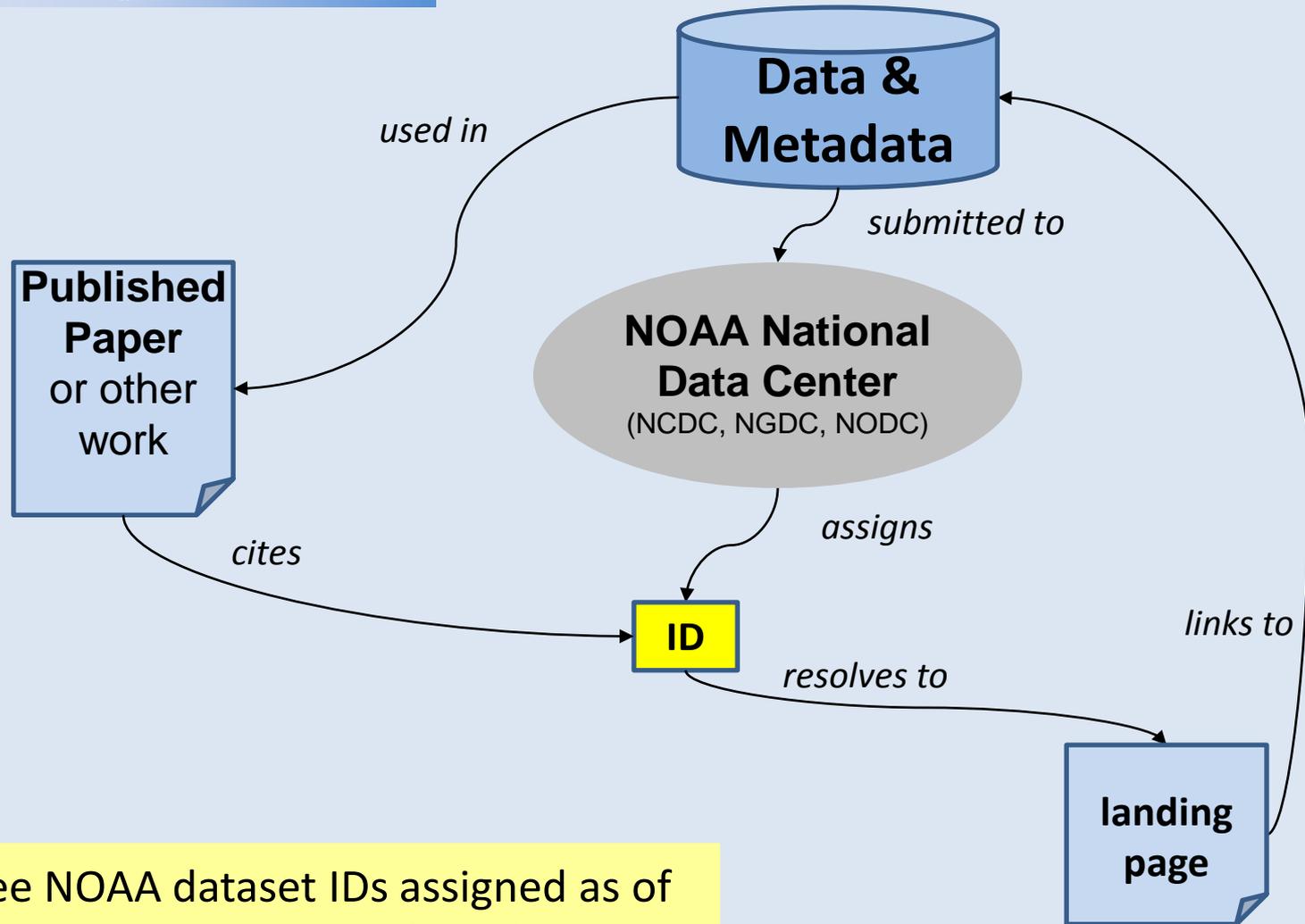
Completeness Score Threshold NOAA/



# Data Preservation Activities

- Ongoing preservation & stewardship activities at NOAA National Data Centers
  - Over 1.5 PB ingested in FY2014
  - CLASS interface improvements
    - Common Submission Interface (CSI)
    - Machine-to-Machine Interface (M2M)
  - National Data Buoy Center using NetCDF groups for monthly archive packages
- Project: Assignment of permanent dataset IDs to archival datasets

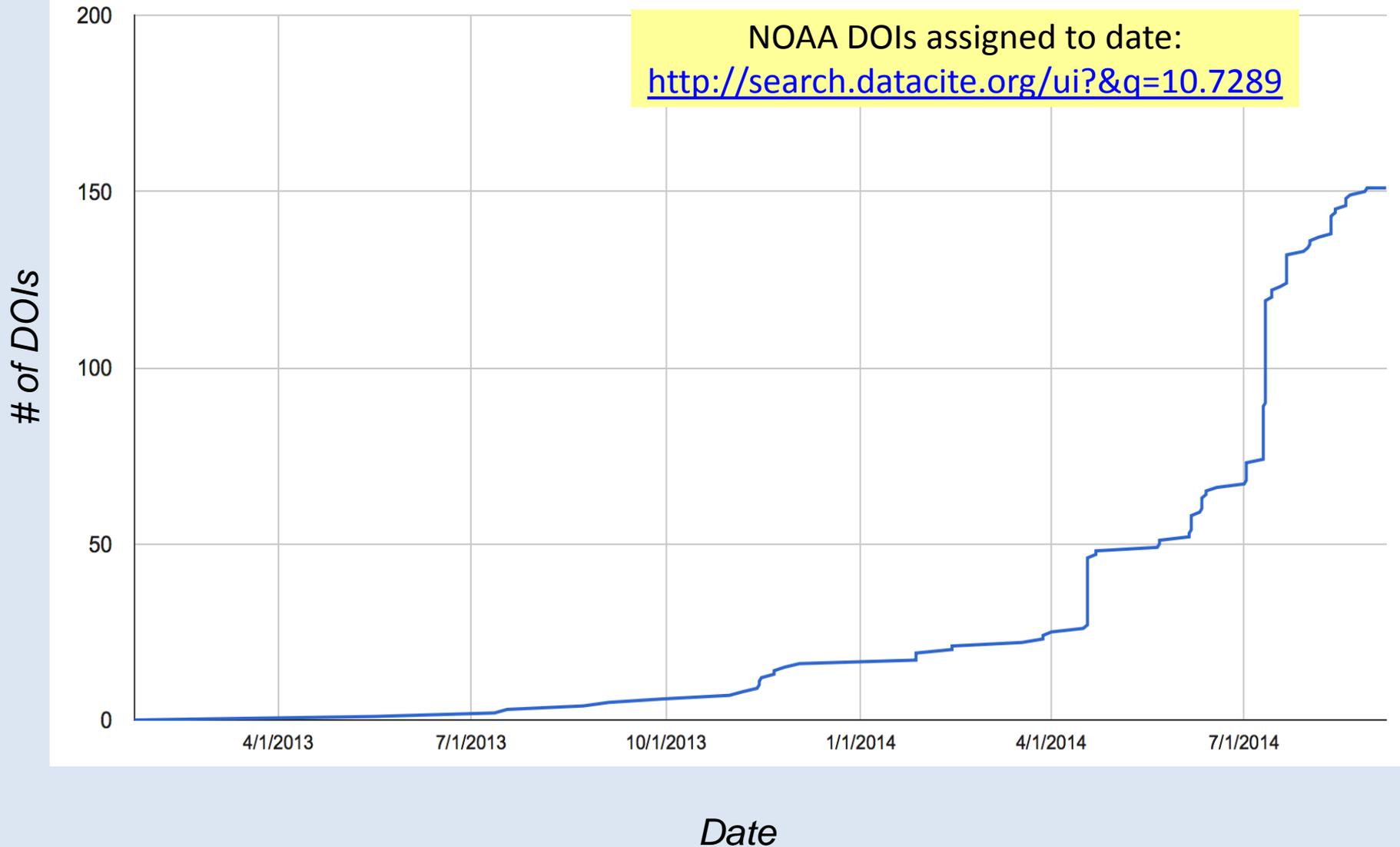
# Dataset Identifier Project



Three NOAA dataset IDs assigned as of July 2013. Target: 20-30 by Feb 2014.

List: <http://goo.gl/KGr0Wy>

# NOAA DOIs Assigned (*data & pubs*)



# Big Earth Data Initiative (BEDI)

- \$2M FY2015 funding request
- Improve **discoverability, accessibility, & usability of data**
- Focus on "**high value**" datasets, e.g. from:
  - OSTP Earth Observations Assessment
  - NOAA Observing Systems of Record
  - USGCRP National Climate Assessment
- **Inter-agency activity** coordinated through US Group on Earth Observations (USGEO) Data Management Working Group
  - Starting more detailed discussion on specific services & approaches

# Closing Wishes

(Jeff's opinion, not NOAA statement)

- Comprehensive view of NOAA usage of Unidata technologies – difficult to get piecemeal from the inside
- Bullet-proof software for operational use
  - Reliable and high-performance
  - Easy installation
  - IT security certification & CIO pre-approval
- Scalable distributions for Cloud use
  - *Glad to see Cloud in 5-year strategy*
  - Use cases include serving Cloud-hosted data, analytical tools running on the Cloud, and efficient transmission of data into Cloud from provider facilities
- *Philosophical question: how can we better leverage, and reduce compartmentalization between, federally-funded activities including NSF/Unidata, NSF/EarthCube, NOAA activities, etc?*