



**NOAA EDM**

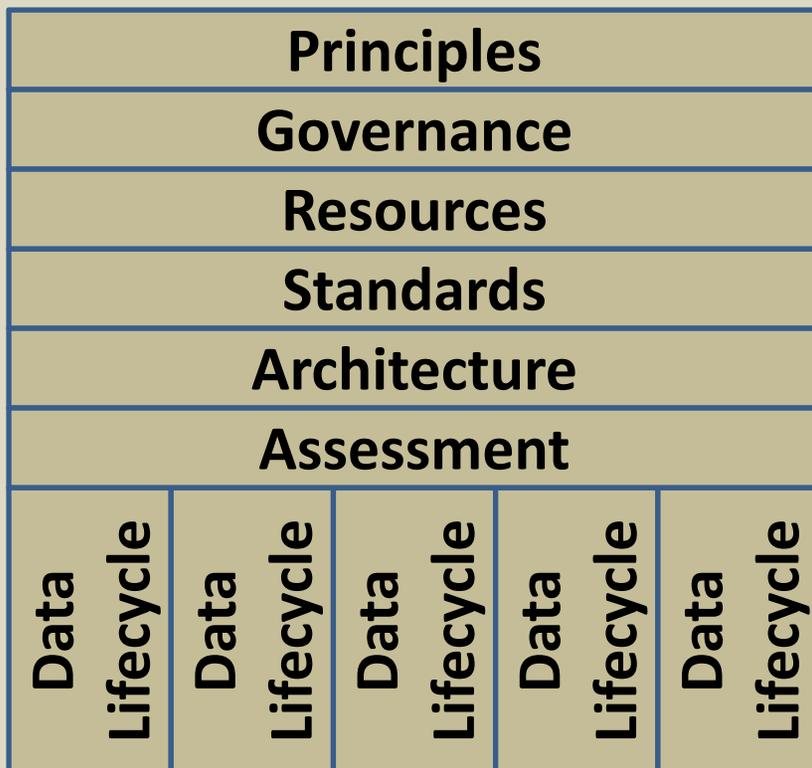
# **NOAA Environmental Data Management**

**Report to EDM Virtual Workshop 2013-06-25**

Jeff de La Beaujardière, PhD  
NOAA Data Management Architect  
[jeff.deLaBeaujardiere@noaa.gov](mailto:jeff.deLaBeaujardiere@noaa.gov)  
+1 301-713-7175

# NOAA EDM Framework

## Data Management Framework



- Purpose: To organize, guide and support NOAA environmental data management activities.
- <https://www.nosc.noaa.gov/EDMC/framework.php>

### ***Contributors:***

*Jeff de La Beaujardière  
(Editor)*

*Anne Ball, NOS CSC*

*Julie Bosch, NCDDC*

*Deirdre Byrne, NODC*

*Leo Carling, OMAO*

*Don Collins, NODC*

*Darien Davis et al., OAR*

*Nancy DeFrancesco*

*Larry Goldberg, NMFS*

*Peter Grimm, CID*

*Ted Habermann, NGDC*

*Karl Hampton, OSPO*

*Steve Hankin, PMEL*

*Scott Hausman, NCDC*

*Michelle Hertzfeld,*

*NESDIS IIAD*

*Christina Horvat, NWS*

*Roy Mendelsohn, NMFS*

*Lewis McCulloch, TPIO*

*Ana Pinheiro Privette,*

*NCDC*

*Nancy Ritchey, NCDC*

*Steve Rutz, NODC*

*Matt Seybold, OSPO*

*Micah Wengren, NOS*

*OCS*

# Data Management Framework

## Data Management Framework

<b>Principles</b>				
<b>Governance</b>				
<b>Resources</b>				
<b>Standards</b>				
<b>Architecture</b>				
<b>Assessment</b>				
<b>Data Lifecycle</b>	<b>Data Lifecycle</b>	<b>Data Lifecycle</b>	<b>Data Lifecycle</b>	<b>Data Lifecycle</b>

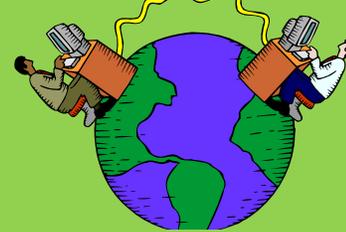
# Vision for NOAA Data Management

All NOAA data will be:

• Discoverable



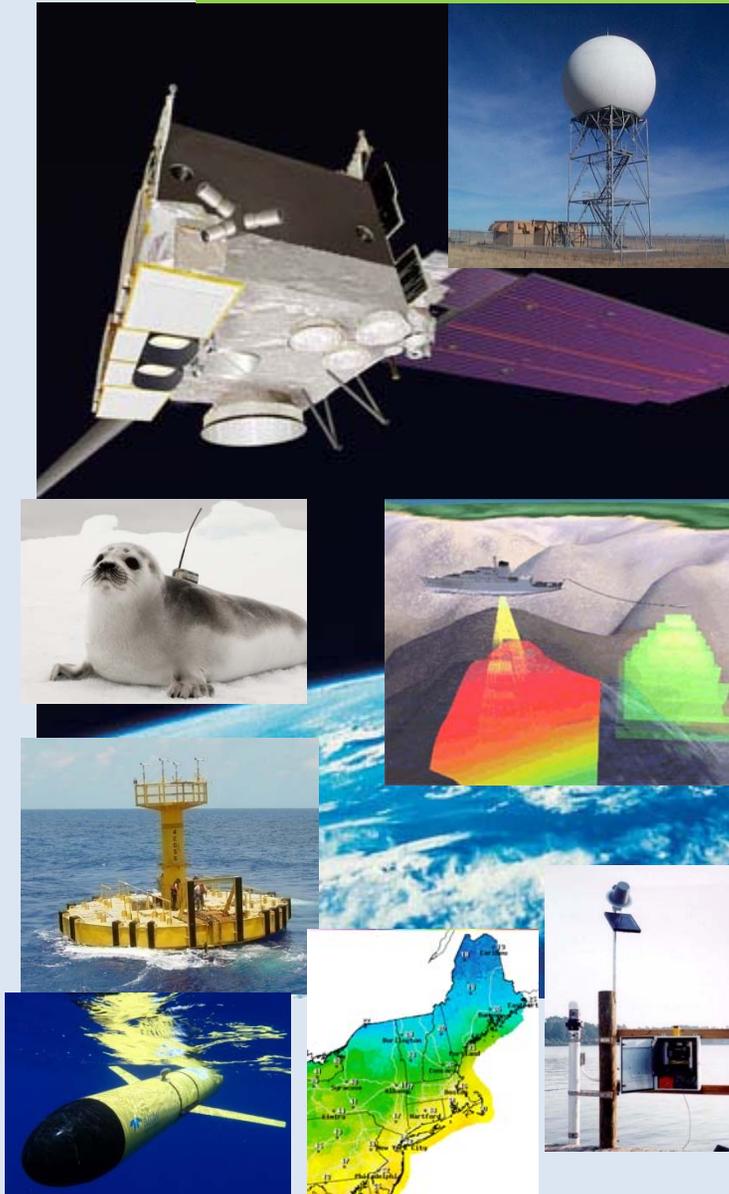
• Accessible



• Usable



• Preserved



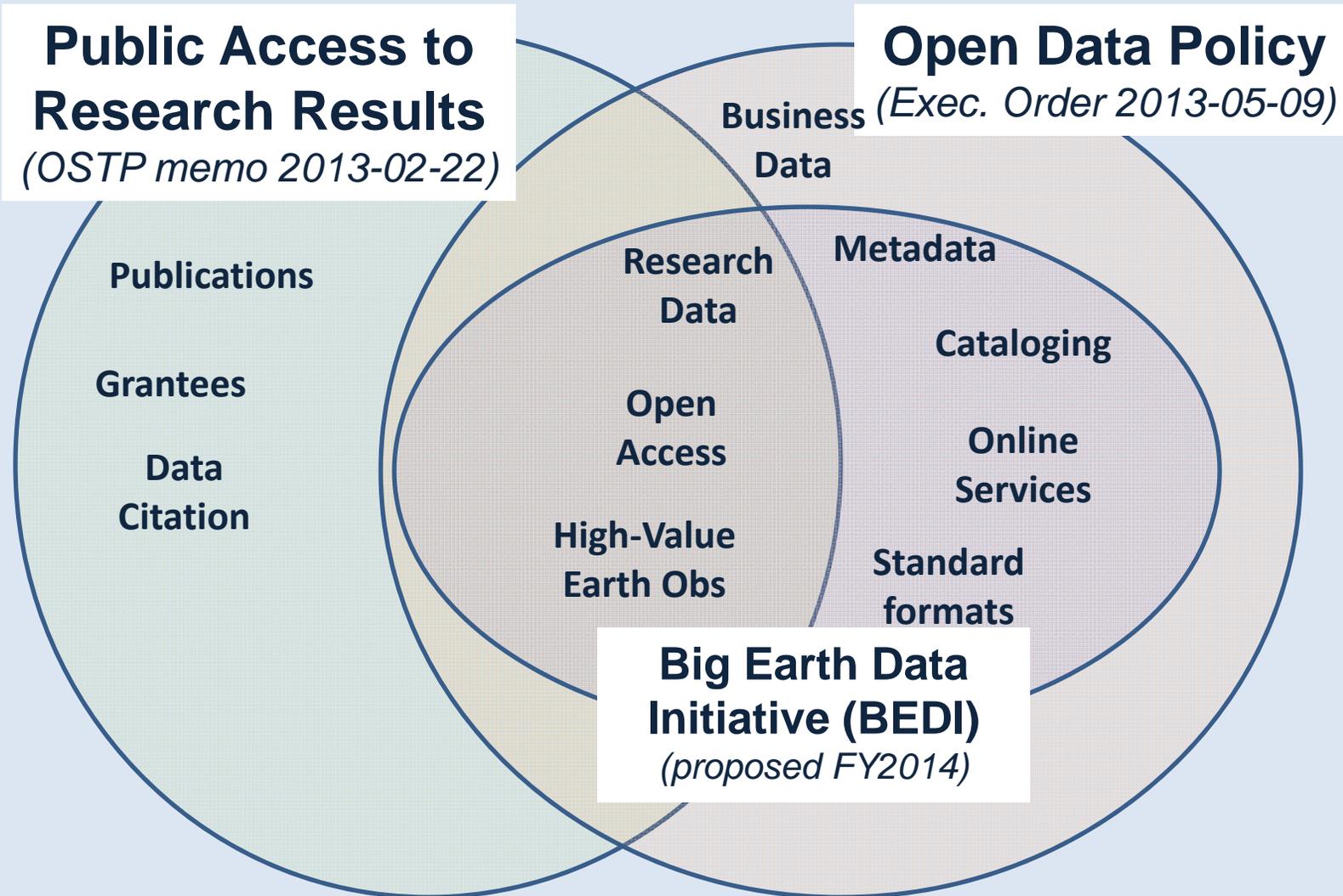
for all types of users and applications.

# Data Management Framework

## Data Management Framework

<b>Principles</b>				
<b>Governance</b>				
<b>Resources</b>				
<b>Standards</b>				
<b>Architecture</b>				
<b>Assessment</b>				
<b>Data Lifecycle</b>	<b>Data Lifecycle</b>	<b>Data Lifecycle</b>	<b>Data Lifecycle</b>	<b>Data Lifecycle</b>

# Recent Federal Initiatives



# Public Access to Research Results (PARR)

- **Publications and Digital Data** must be freely available
  - Embargo period for journal publications
- Memo from White House Office of Science and Technology Policy (OSTP):  
**"Increasing Access to the Results of Federally Funded Scientific Research"**
  - [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)
  - Focus is more on policy than technology
  - Memo issued 2013 Feb 22; draft plans due **2013 Aug 22**
- **NOAA PARR Cmtee established by Research Council** to draft plan
  - *Co-chairs: Jeff DLB (EDMC), Neal Kaske (NOAA Library)*
  - *Members: Danielle Tillman - NRC liaison, Mark Brady - NMFS, Paul Comar - NOS, Cecile Daniels - OMAO, Ingrid Guch - NESDIS, Douglas Perry - OMAO, Allen Shimada - NMFS, Ronald Stouffer - OAR, Harry Tabak - NWS, Marina Timofeyeva - NWS*
- OSTP hosting **interagency meetings**

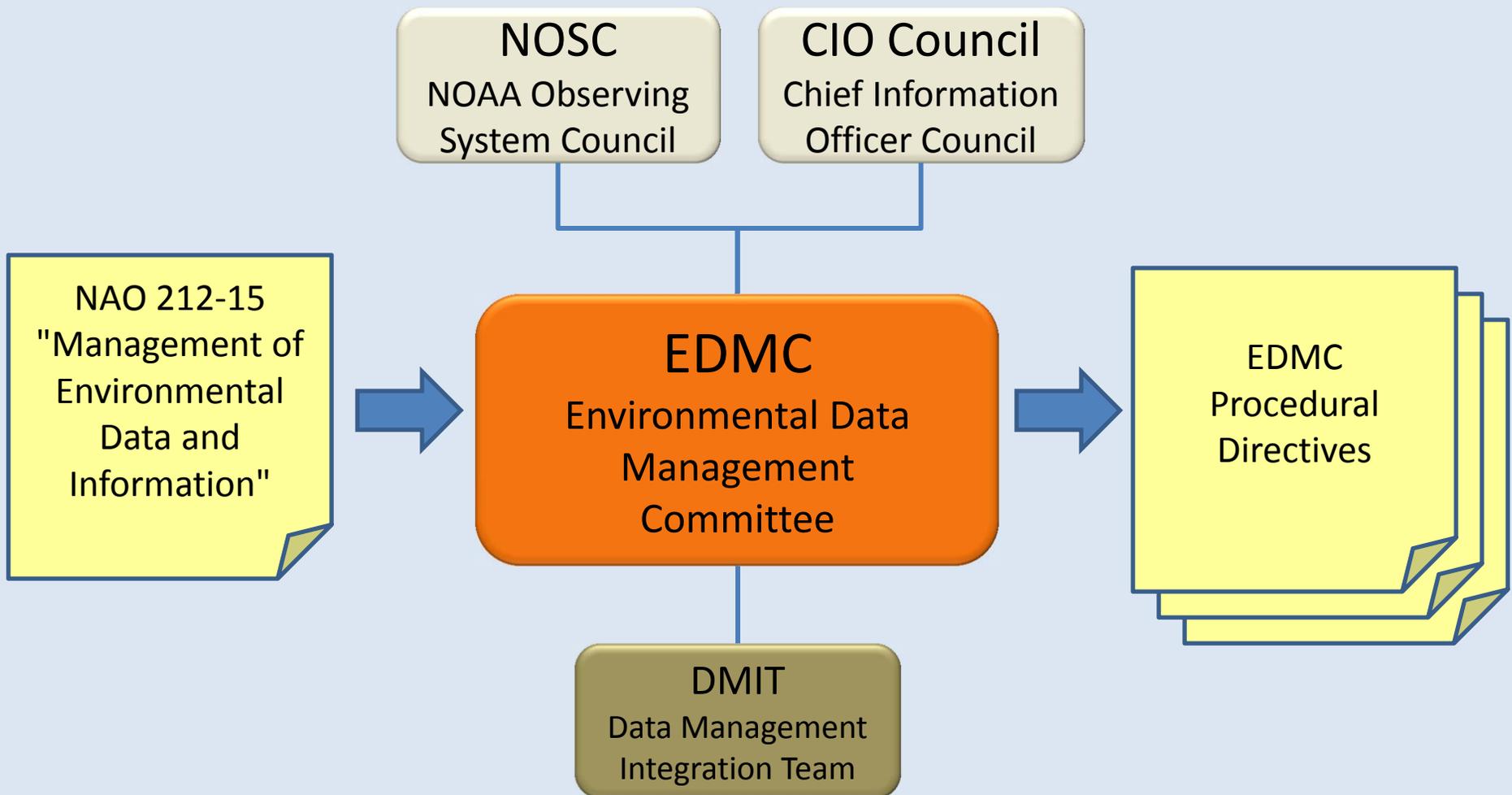
# Open Data Policy (May 9, 2013)

- Requires:
  - Use of **open standards** and **machine-readable formats**
  - Life-cycle **data management planning**
  - **Inventory of agency data** assets at [agency.gov/data](http://agency.gov/data)
- Initial deadlines **November 9, 2013**
- NOAA planning begun w/chairs of EA, EDM, GIS, & Web cmtees
- References:
  - Executive Order -- "**Making Open and Machine Readable the New Default for Government Information**"
    - <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government>
  - OMB Open Data Policy -- Managing Information as an Asset
    - <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>
  - Project Open Data site on GitHub
    - <http://project-open-data.github.io/>

# Big Earth Data Initiative (BEDI)

- proposed multi-agency activity coordinated through US Group on Earth Observations (USGEO)
  - Goal: Improve **discoverability, accessibility, & usability of data**
  - Focus on "**high value**" datasets, e.g. from:
    - OSTP Earth Observations Assessment
    - USGCRP National Climate Assessment
    - NOAA Observing Systems of Record
- NOAA activity:
  - **FY2014 Funding request** in Pres Bud (*need Congress approval*)
  - Preliminary planning with all NOAA LOs remainder of 2013
  - TPIO: Project management
  - EDMC, DMIT: Technical coordination & implementation
  - NOSC, CIO Council: Oversight

# NOAA Environmental Data Management Committee



# EDMC Procedural Directives

(Environmental Data Management Committee)

## Data Management Planning PD

*Plan, in advance, how you will preserve, document and distribute your data.*

### Archive Procedure

*What to archive, how to submit to archive.*

### Data Documentation

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

### Data Sharing by NOAA Grantees

*State in proposal how you will share data. Share within 2 years.*

### Data Citation

*Assign persistent identifiers to datasets and encourage citation.*

### Data Access

*Establish & improve on-line services for data access*

*in prep.*

# Data Management Framework

## Data Management Framework

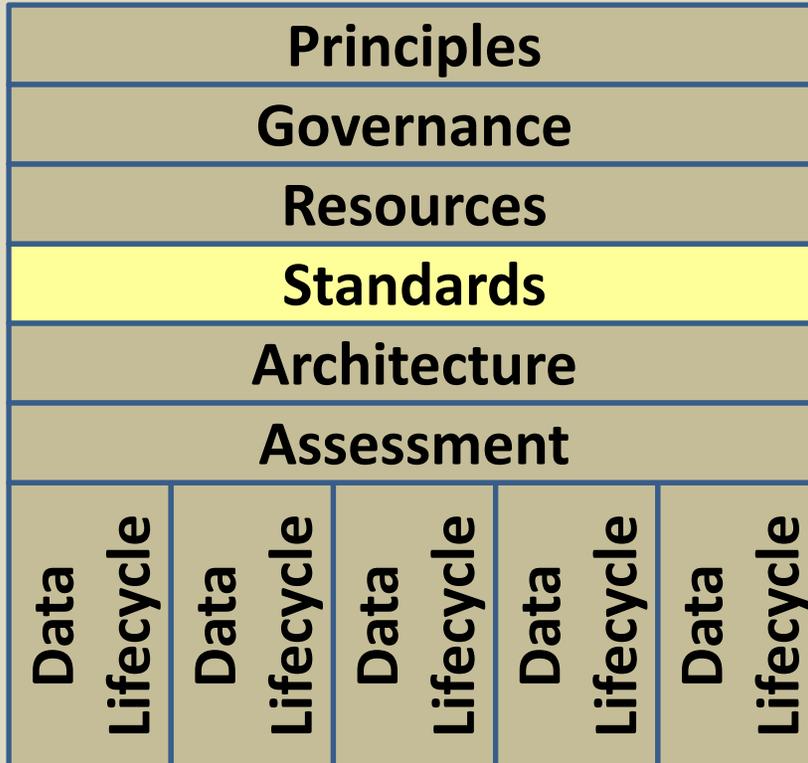
Principles				
Governance				
Resources				
Standards				
Architecture				
Assessment				
Data Lifecycle	Data Lifecycle	Data Lifecycle	Data Lifecycle	Data Lifecycle

## Resources

- Budget
  - Project-specific
  - **FY2014 - BEDI?**
- Personnel
  - Training
  - **Recognition**
  - Authority
- Other Resources
  - **Annual Workshop**
  - **Teams (e.g. DMIT)**
  - **EDM Wiki:**  
<https://geo-ide.noaa.gov/wiki/>

# Data Management Framework

## Data Management Framework



## Standards

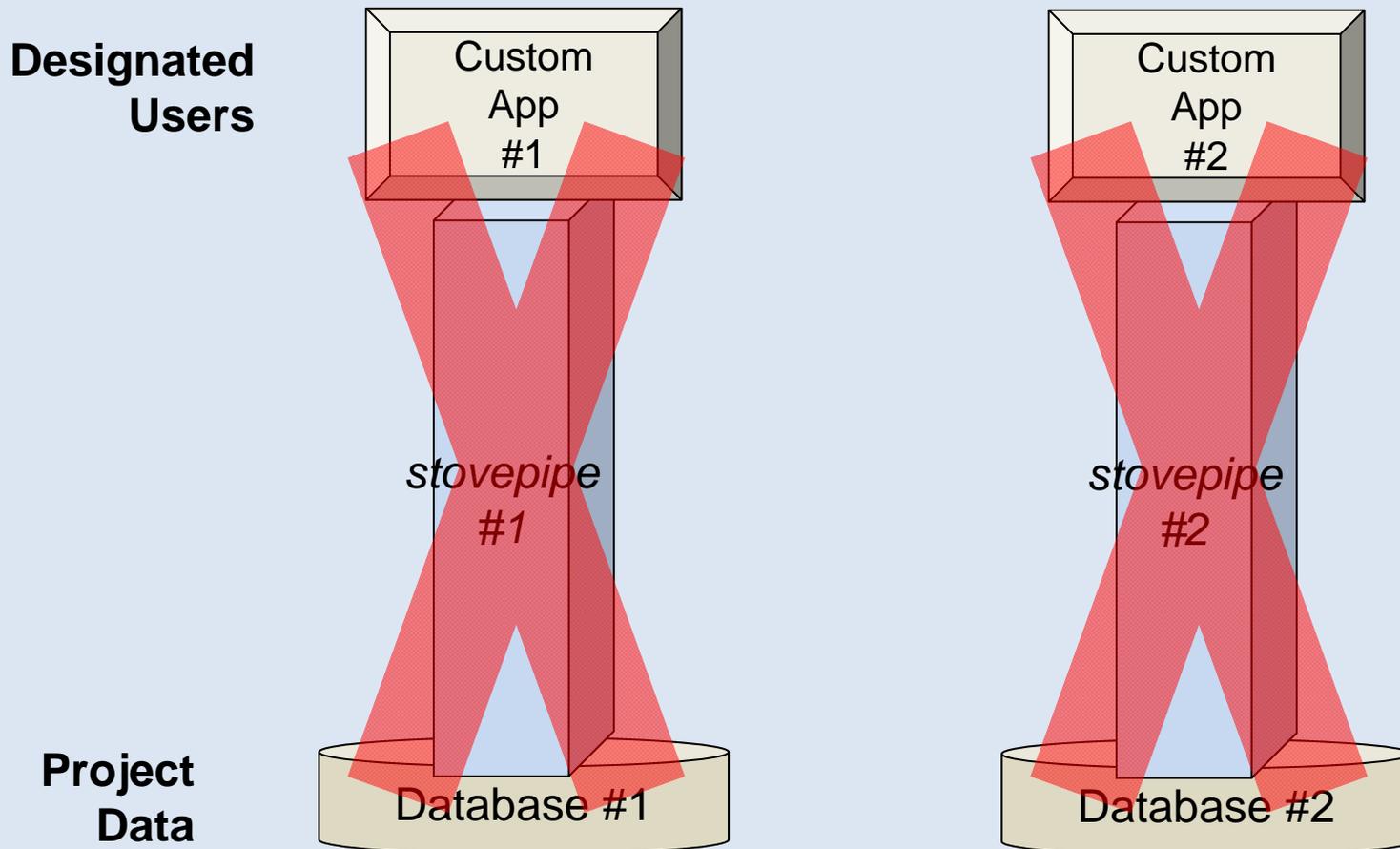
- **Metadata**
  - ISO 19115/19139
- **Formats and services**
  - Open Data Policy mandate
- Vocabularies
- Quality Control
- IT Security
  - NAO 212-13

# Data Management Framework

## Data Management Framework

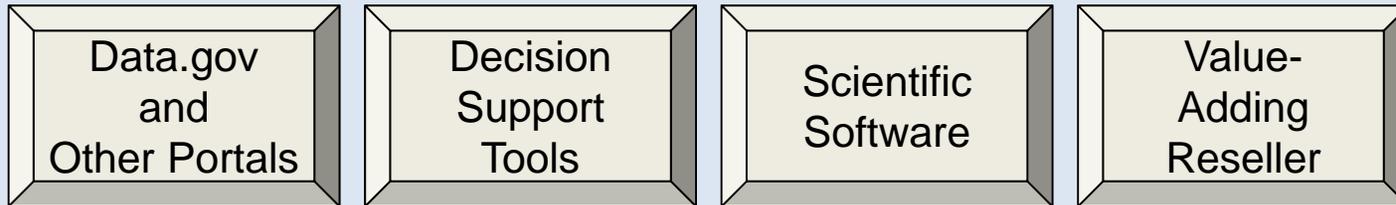
<b>Principles</b>				
<b>Governance</b>				
<b>Resources</b>				
<b>Standards</b>				
<b>Architecture</b>				
<b>Assessment</b>				
<b>Data Lifecycle</b>	<b>Data Lifecycle</b>	<b>Data Lifecycle</b>	<b>Data Lifecycle</b>	<b>Data Lifecycle</b>

# Avoid Purpose-Built Systems

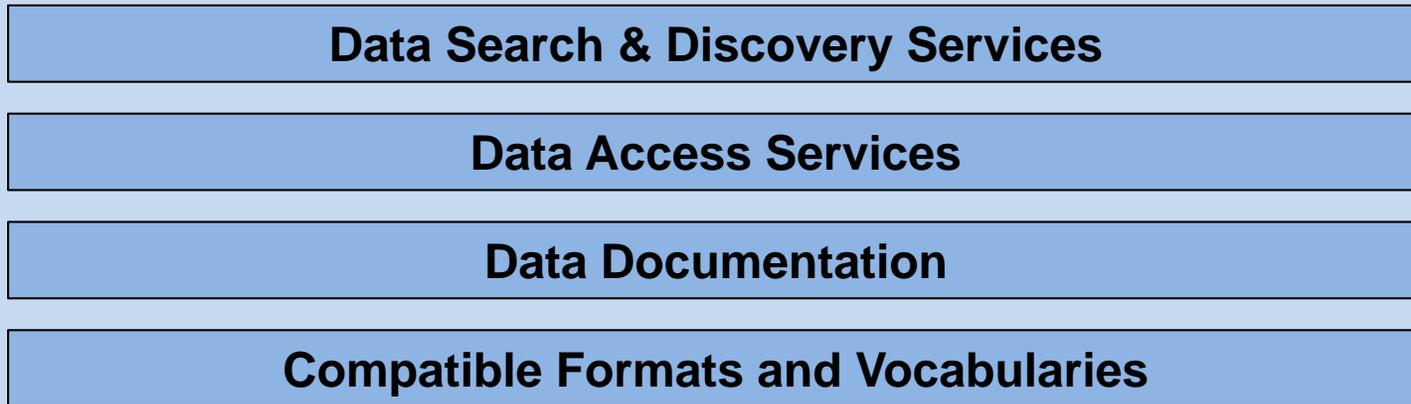


# Data Services Layer

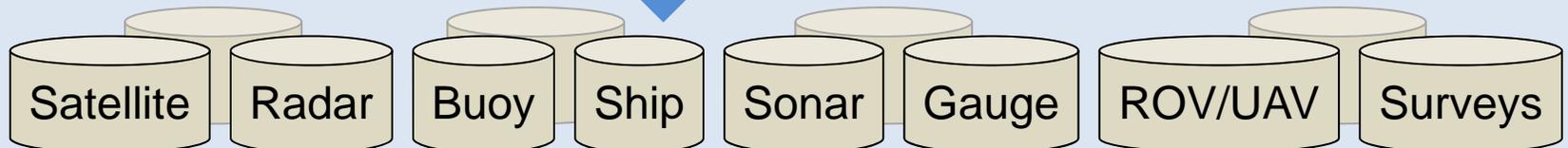
**User Tools**



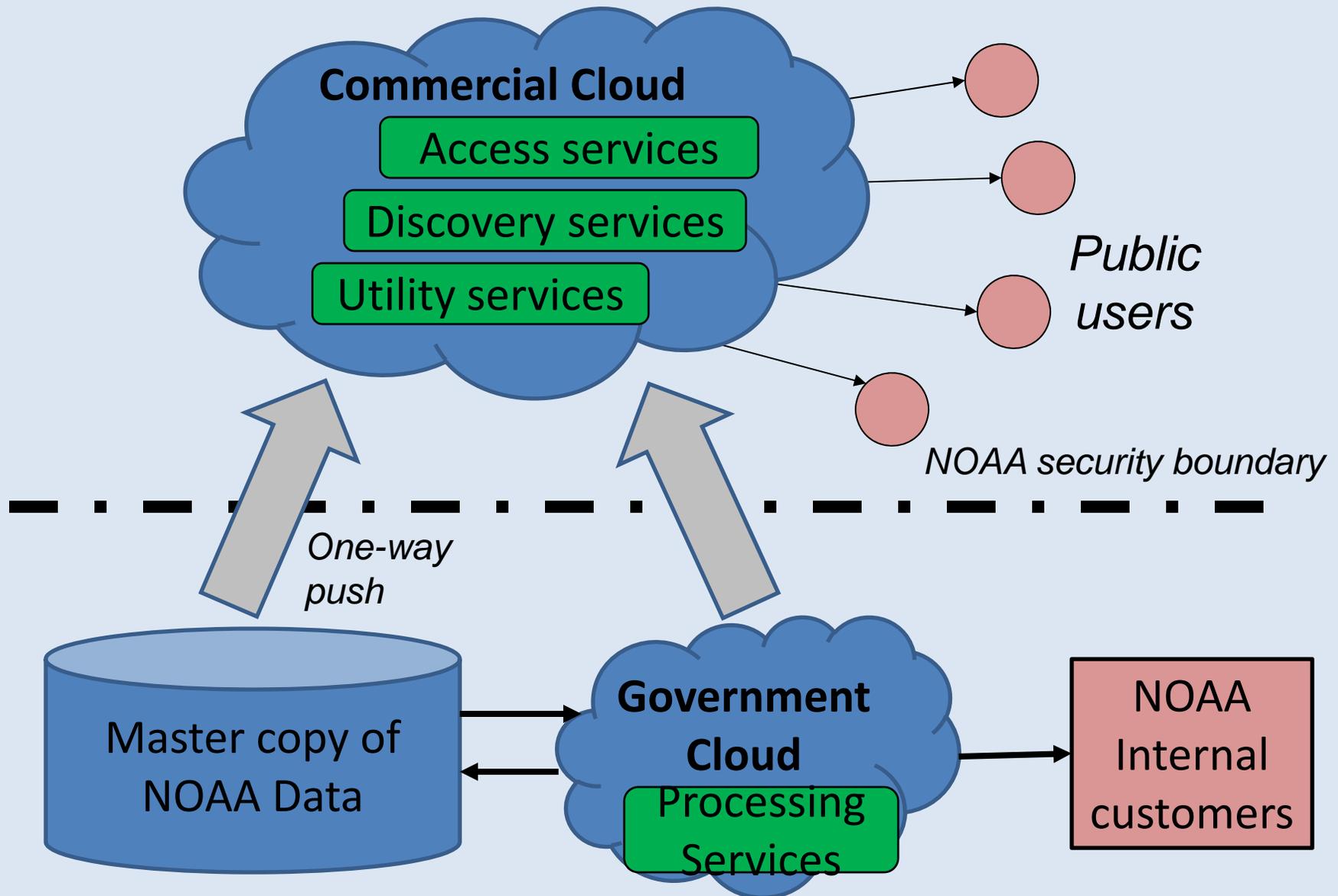
*data services layer*



**Data Sources**



# Potential Cloud Deployment Scenario



# Data Management Framework

## Data Management Framework

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

### Assessment

- EDM Dashboard
- Observing System of Record  
EDM Assessment
- Geospatial Systems  
Assessment



NOAA EDM

NOAA Environmental Data Management

EDM Dashboard: Project Description, The Team, Project Calendar, Tasks & Actions, Documents, Issues and Comments, Sitemap

Recent site activity

EDM Dashboard

edited by Matthew Austin - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Jason Chasse - NOAA Federal

comment from Denise Blake - NOAA Federal

comment from Jason Chasse - NOAA Federal

comment from Jennifer Bosch - NOAA Aftiz

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

comment from Jeff de La Beaugerie - NOAA Federal

comment from Martin Yapur - NOAA Federal

### EDM Dashboard

(Under Development)

NOTE: Explanations of each chart may be found on the [Project Description](#) page.

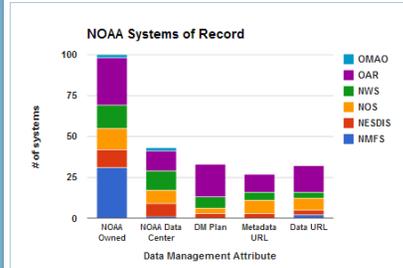
EDM Work Plan Status

Work Plan Item	Due	On Track?	Status	Data
Develop EDM Framework (SAB action)	2013 Mar	DONE	DONE	Version 1.0 posted on EDMC web site. <a href="https://www.nos.noaa.gov/EDMC/">https://www.nos.noaa.gov/EDMC/</a>
Develop External Data Usage Recommended Practice (SAB action)	2013 Mar	DONE	DONE	Version 1.0 posted on EDMC web site. <a href="https://www.nos.noaa.gov/EDMC/">https://www.nos.noaa.gov/EDMC/</a>
Host the 2013 EDM Virtual Workshop (June 25-27, 1-4pm)	2013 Jun	Green	In-person conference with travel not authorized. Planning Virtual Workshop instead.	<a href="http://goo.gl/L26Lr">http://goo.gl/L26Lr</a>
Implement and monitor approved Procedural Directives	ongoing	Green	Reporting mechanism established via Google Drive. Monthly input from LO reps.	<a href="http://goo.gl/5KEe0">http://goo.gl/5KEe0</a>
Write new Data Citation PD	2013 Sep	Yellow	Small team meeting by telecon. Working on dataset requirements and landing page creation mechanism.	<a href="http://goo.gl/ENiok">http://goo.gl/ENiok</a>
Write new Data Access PD	2013 Sep	Yellow	Outline of PD drafted. Increasing priority in light of OSTP PARR memo and Big Earth Data Initiative.	<a href="http://goo.gl/0SXMS">http://goo.gl/0SXMS</a>
Perform EDM Assessment of Observing Systems of Record	2013 May	Yellow	EDMC issued data call 2013-03-14 at NOSC direction. Responses due 2013-04-28. ~16% received so far.	<a href="http://goo.gl/Autzv">http://goo.gl/Autzv</a>
Establish prototype EDM Dashboard	2013 Feb	DONE	DONE	Demo to CID Council Feb 12, NOSC Mar 7. <a href="https://sites.google.com/a/noaa.gov/dashboard/">https://sites.google.com/a/noaa.gov/dashboard/</a>

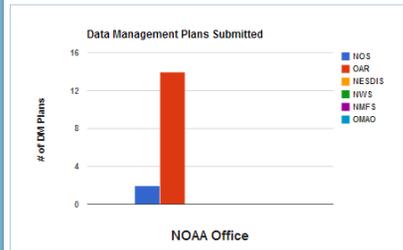
EDMC

Open EDMC Work Plan Status

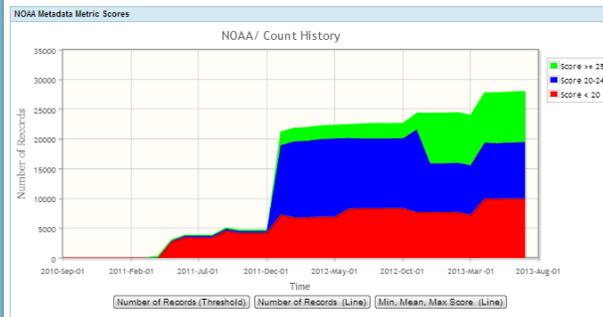
### NOAA System of Record EDM Assessment



Preliminary numbers--assessment is in progress. Data source for this chart: [NOAA SoR EDM Assessment \(as of 2013-05-21\)](#)



Data source for this chart: [NOAA Data Management Plan Repository](#)



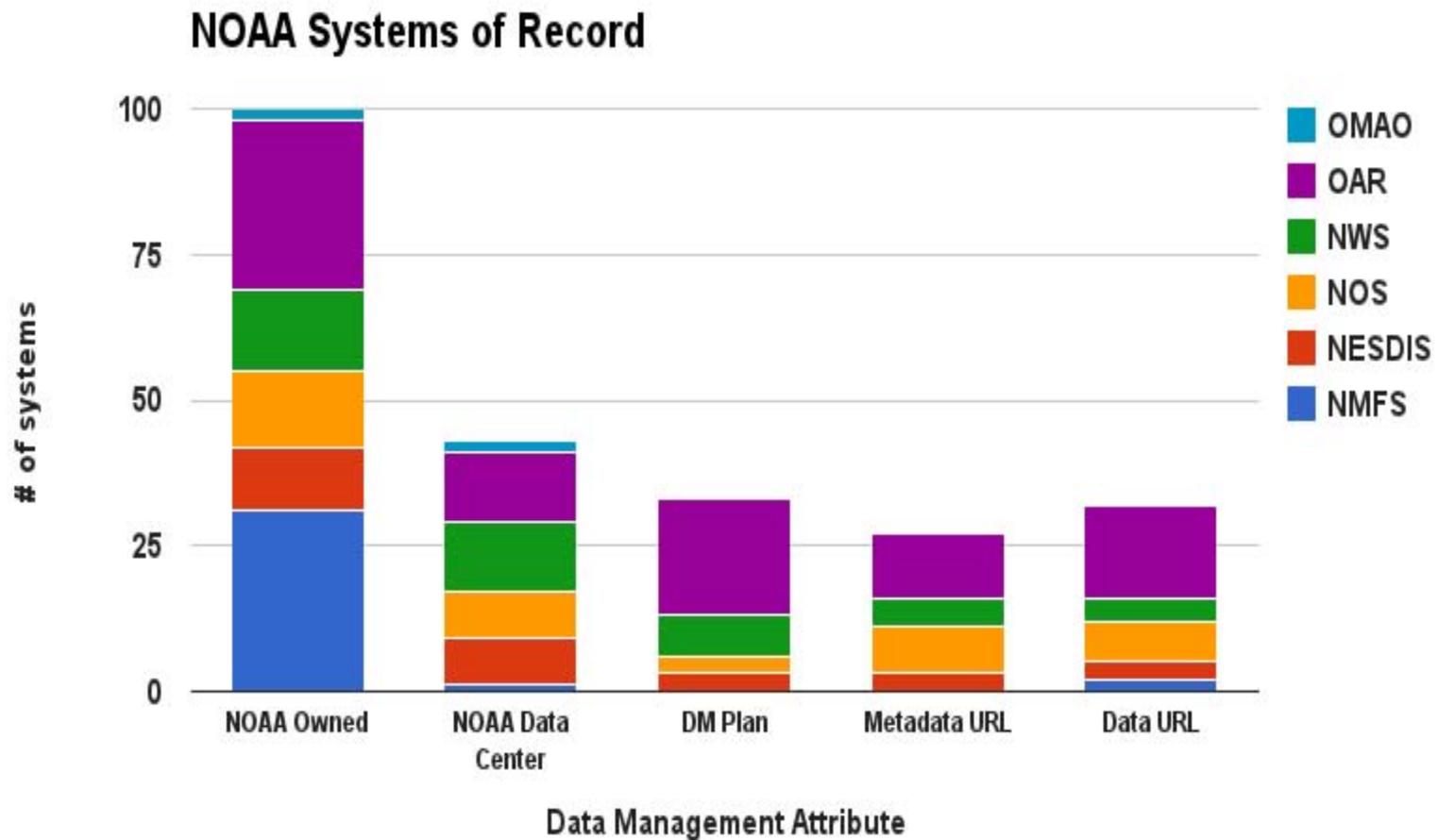
Data source for this chart: NOAA Metadata Metric Scores

# EDM Dashboard

<http://sites.google.com/a/noaa.gov/edm-dashboard/>

(internal access only)

# Obs. System of Record EDM Assessment



# Data Management Framework

## Data Management Framework

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

# Data Lifecycle

*Planning and  
Production  
Activities*

Requirements Definition  
Planning  
Development  
Deployment  
Operations

*Data  
Management  
Activities*

**Collection**  
**Processing**  
**Quality Control**  
**Documentation**  
**Cataloging**  
**Dissemination**  
**Preservation**  
**Stewardship**  
**Usage Tracking**  
**Final Disposition**

*Usage  
Activities*

Discovery  
Reception  
Understanding  
Analysis  
Value-Added Products  
User Feedback  
Citation  
Tagging  
Gap Assessment

## Data Lifecycle Activities

# Data Lifecycle

*Planning and Production Activities*

Requirements Definition  
Planning  
Development  
Deployment  
Operations

## Applicability of EDMC Directives

DM Planning

*Data Management Activities*

Collection  
Processing  
Quality Control  
Documentation  
Cataloging  
Dissemination  
Preservation  
Stewardship  
Usage Tracking  
Final Disposition

Data Documentation

Data Sharing by Grantees

*Data Access*

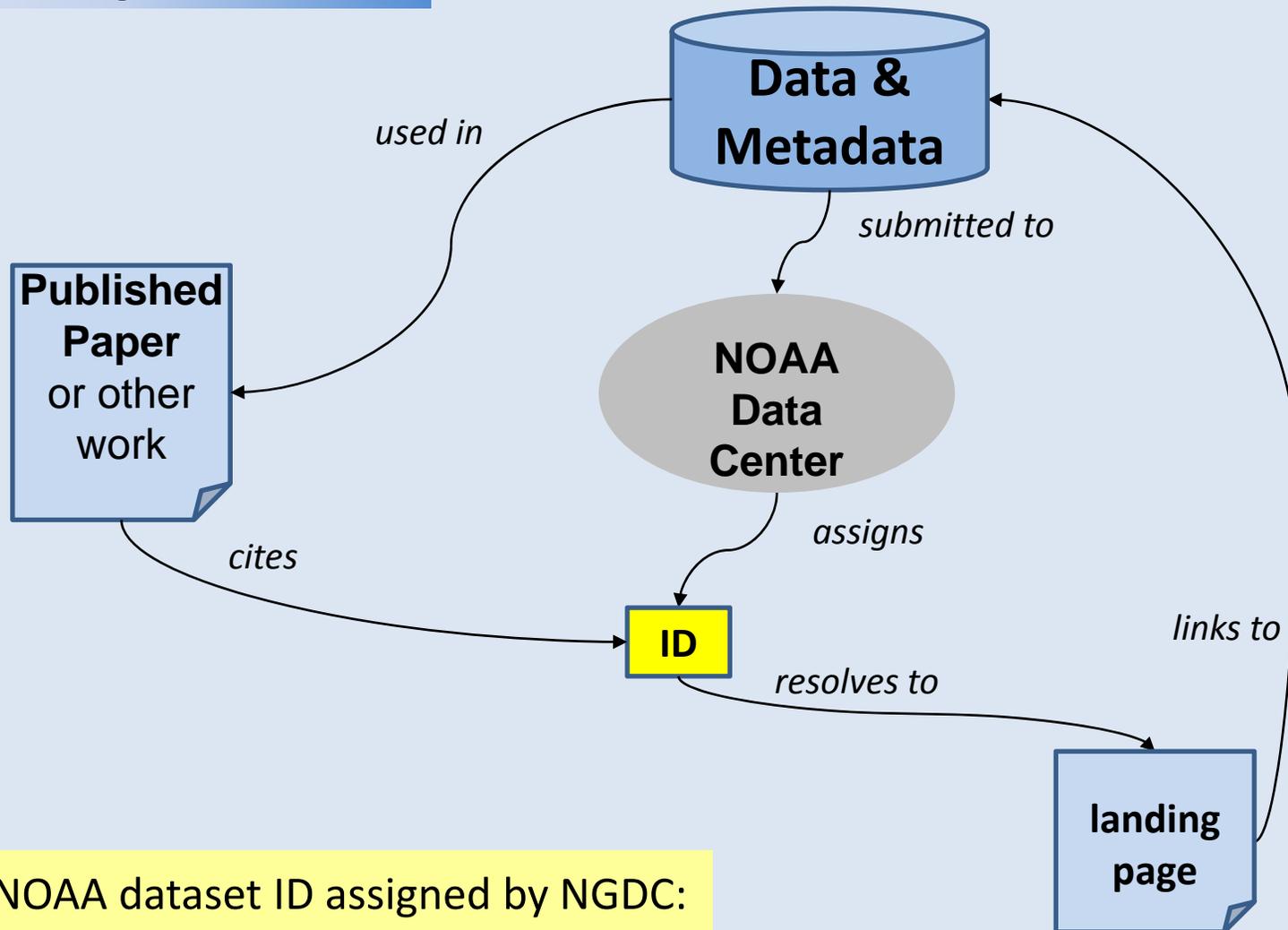
Archive Procedure

*Usage Activities*

Discovery  
Reception  
Understanding  
Analysis  
Value-Added Products  
User Feedback  
Citation  
Tagging  
Gap Assessment

*Data Citation*

# Dataset Identifier Project



First NOAA dataset ID assigned by NGDC:  
10.7289/V5154F01  
(<http://dx.doi.org/10.7289/V5154F01>)

# Timeline

OSTP  
PARR  
memo

Open  
Data  
Policy

PARR  
Plan  
due

noaa.gov/data  
inventory  
due

BEDI execution?

NOAA-wide BEDI pre-planning

NOAA  
DOI  
license

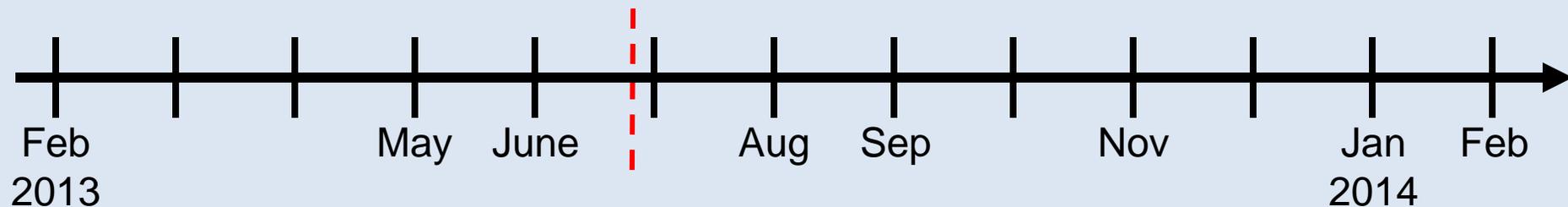
NOAA  
FY2014  
budget

1st  
NOAA  
DOI

Virtual  
EDMW

EDMC  
FY2014  
plan

NOAA  
DOI  
license  
renewal



# Summary

NOAA data should be:

- Discoverable



- Accessible



- Usable



- Preserved



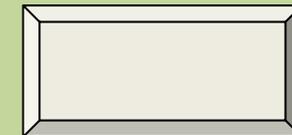
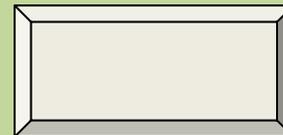
Public Access to  
Research Results

Open Data Policy

Big Earth Data  
Initiative (BEDI)

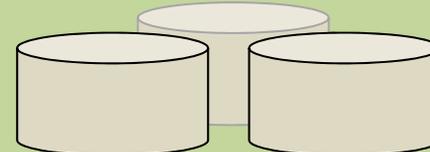
EDMC  
Procedural  
Directives

User  
Tools



*data services layer*

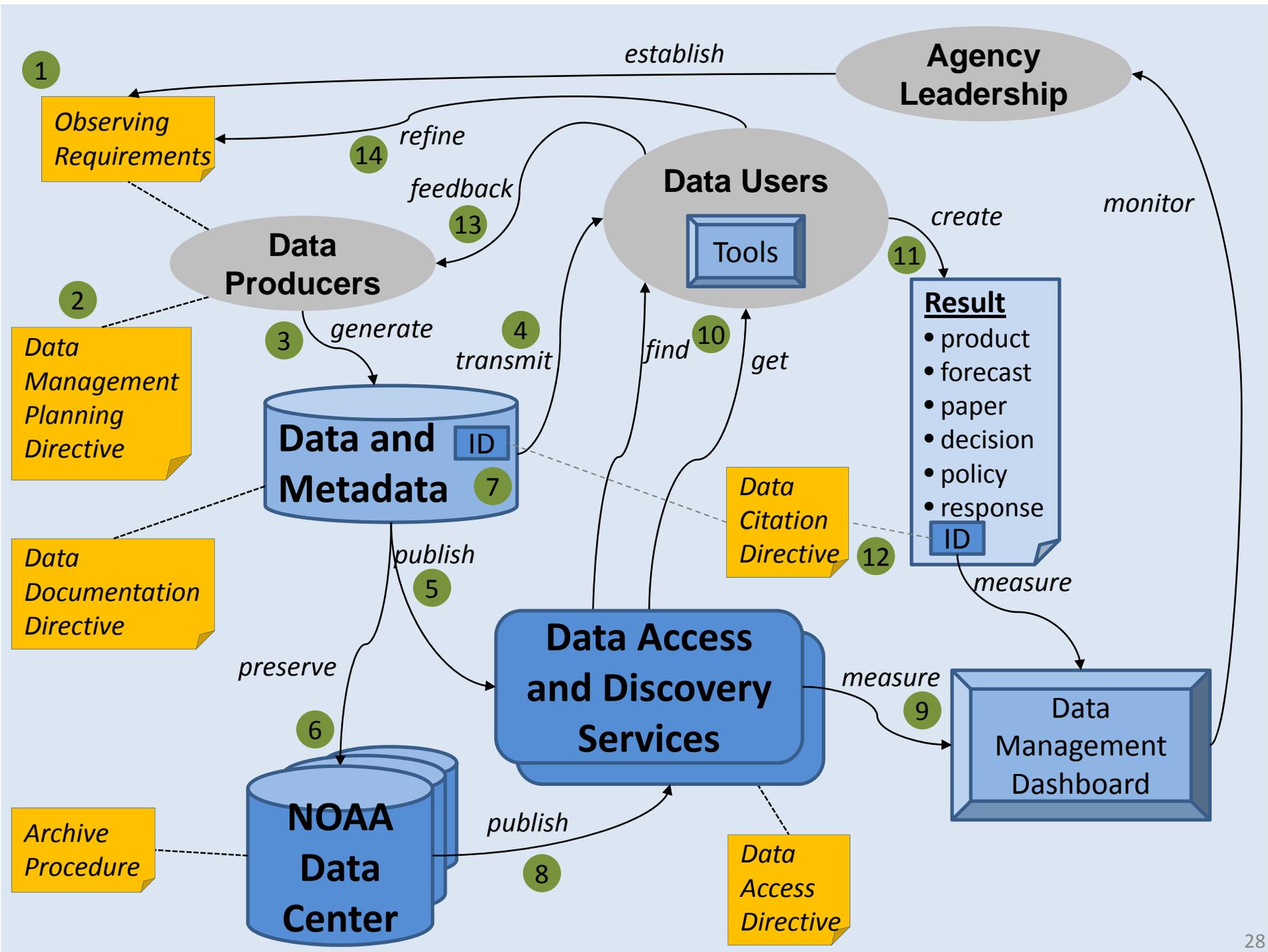
Data  
Sources





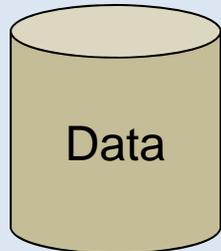
**NOAA EDM**

# Backup Slides

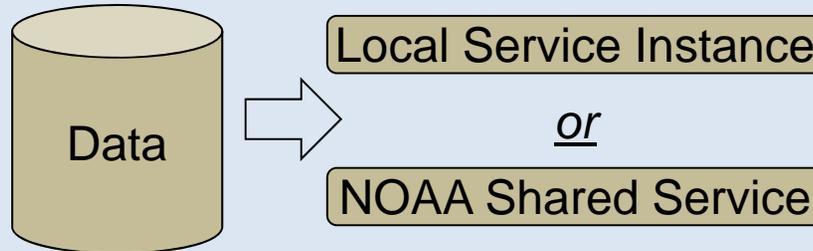


# Data Provider Roles

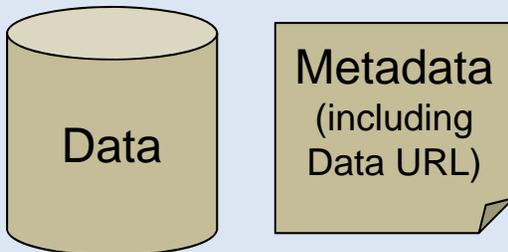
## 1: Obtain Data.



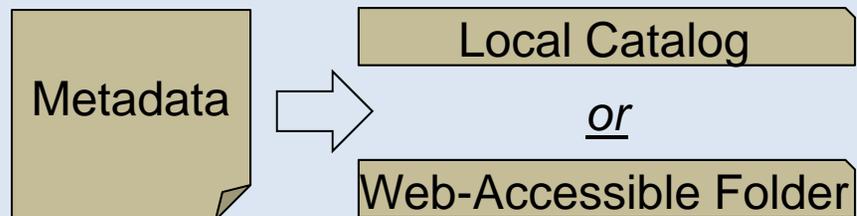
## 2: Make your data accessible on-line.



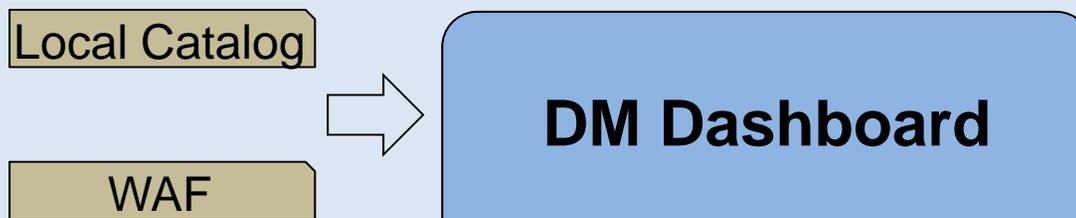
## 3: Document your Data well.



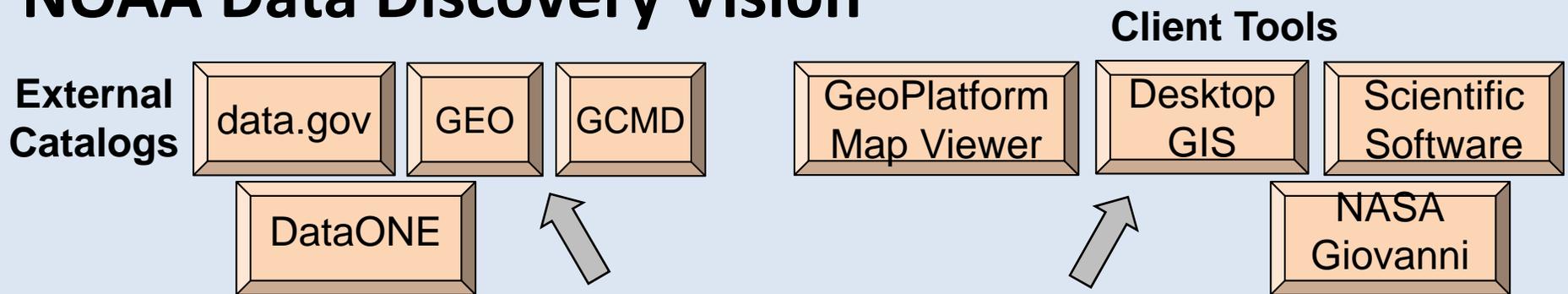
## 4: Make your metadata accessible.



## 5: Tell us where to find your metadata



# NOAA Data Discovery Vision



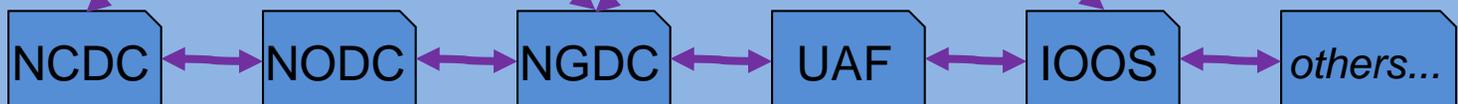
## NOAA data cataloging

Find NOAA data by:

- Location
- Time
- Phenomenon
- Other characteristics

*without knowing NOAA organization*

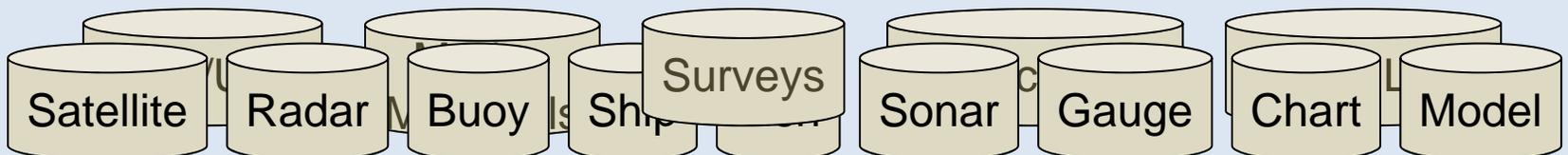
**Local Catalogs & Metadata Folders**



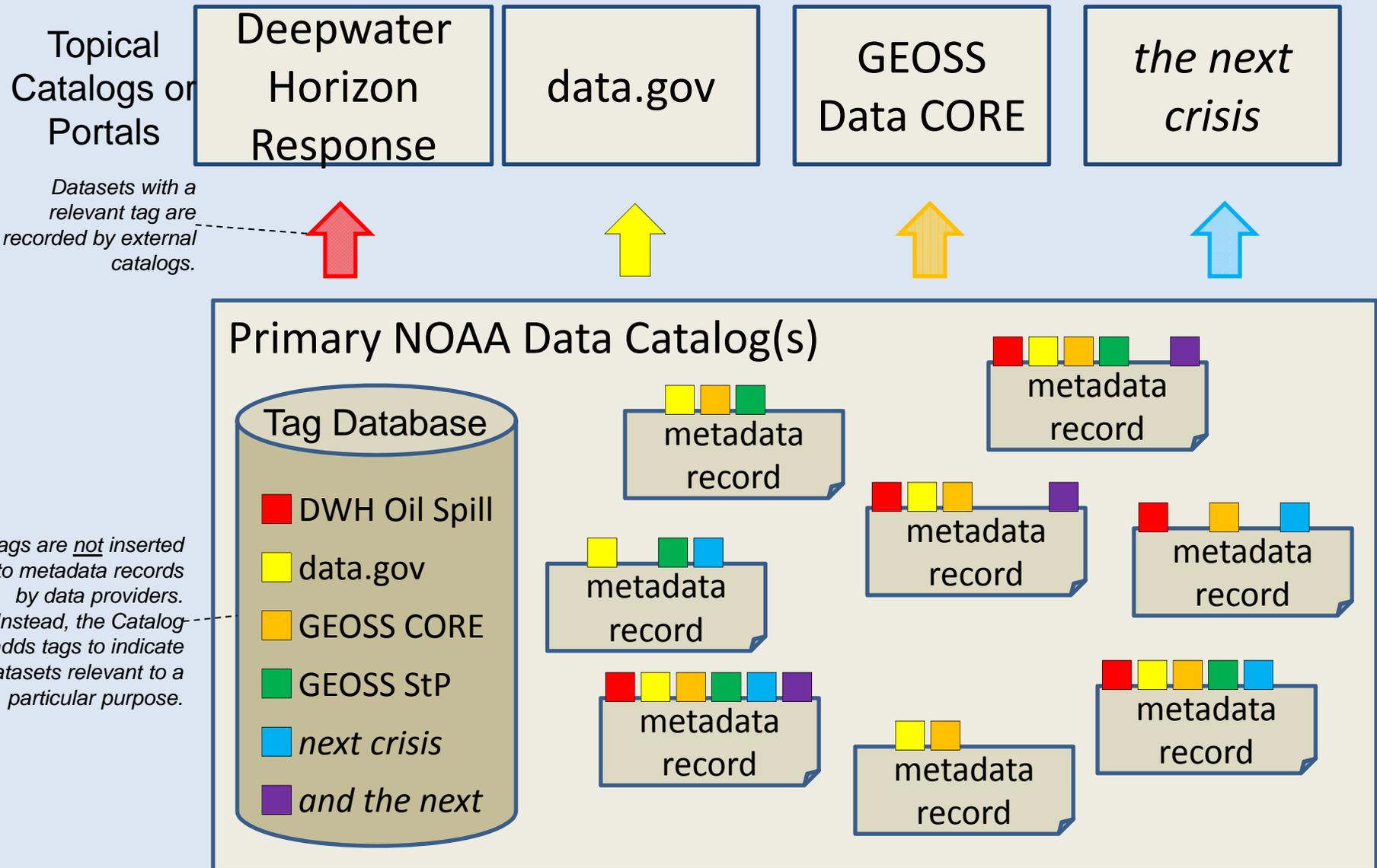
**Data Access Services**



**Data Sources**

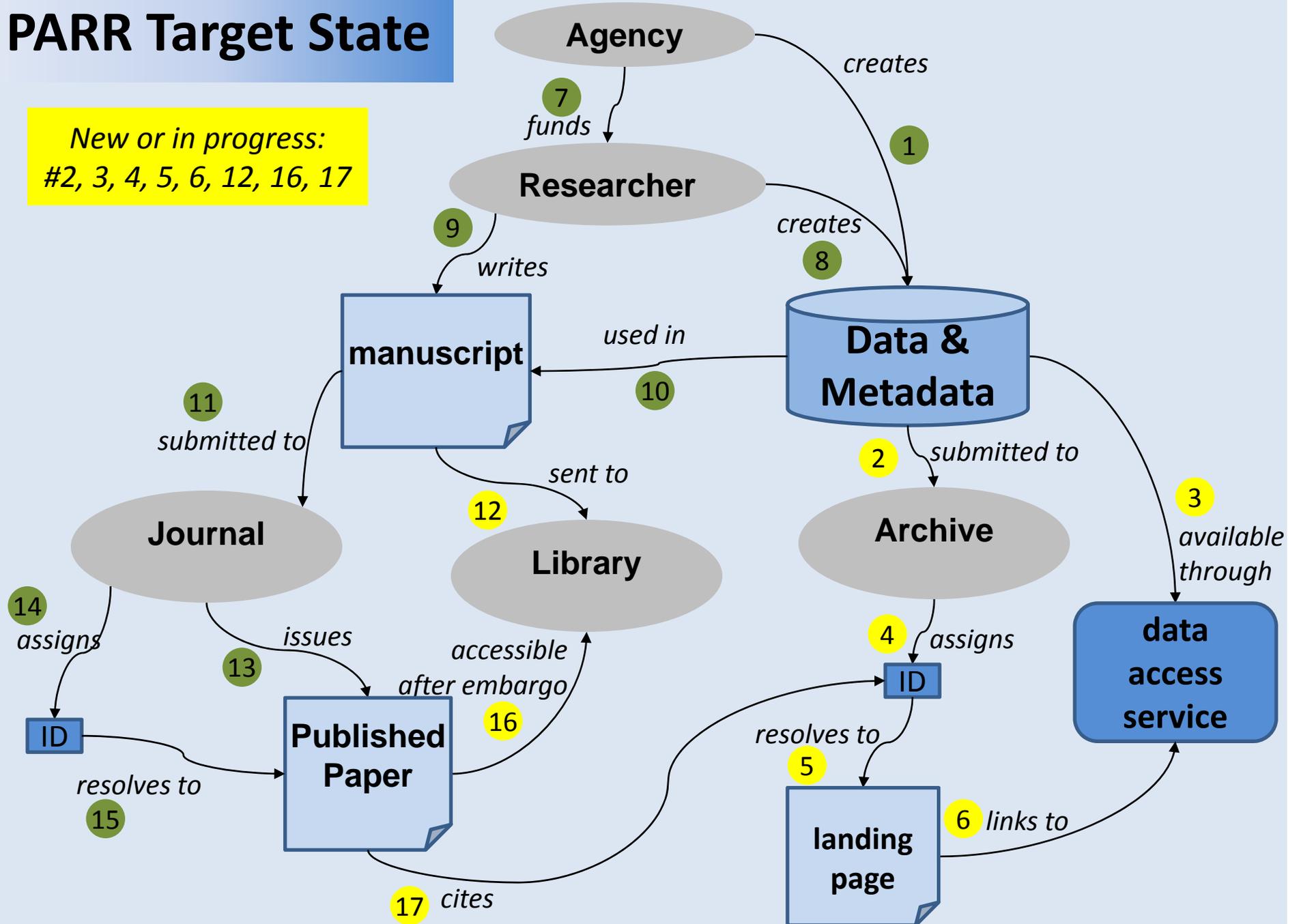


# Tagging Concept



# PARR Target State

New or in progress:  
#2, 3, 4, 5, 6, 12, 16, 17



# Obstacles to improving NOAA EDM

- Changes cost money.
- Savings from standardization and reuse are not visible until later, or may accrue to later adopters.
- Standards are complicated, too general, or too specific to a different purpose.
- NOAA datasets and programs are numerous, heterogenous, mission-specific and independent.
- Operational systems cannot be readily changed.
- System developers start from scratch rather than first seeking existing solutions.
- Good DM practices are not part of employee job descriptions or recognition methods.
- People who approve new projects are not always aware of existing approaches and directives.



**NOAA EDM**

# Session 2-5 Slides

## Catalog & Search - Current state:

- No complete inventory
- No single point of entry
- limited integration with commercial search engines
- Numerous Portals with overlapping scope & manual population
- Various catalog implementations
  - Geoportal Server at each Data Center
  - UAF Catalog
  - NWS WIS GISC DAR

## Catalog & Search - Target state:

- Agency data inventory per Presidential Executive Order
- Single point of entry to federated search API
- Automatic registration with data.gov, GCMD, GEOSS, GCIS, thematic portals
- Better integration with commercial search engines

## Catalog & Search - Potential next steps:

- improve/augment Data Center catalogs
- federation of existing catalogs
- agency-wide data inventory ([noaa.gov/data](http://noaa.gov/data))
- [schema.org](http://schema.org)
- CS/W
- OpenSearch

## Data Access - current state:

- variety of data access methods offered
  - DAP, OGC WMS, WFS, WCS, SOS, Esri, FTP, web links, CLASS API, custom services
- differing implementations of same service type
- multiple access points for some datasets, with authoritative source unclear
- some datasets not accessible online at all
- no permanent shared hosting capability
  - except archival data from National Data Centers
- no CIO-approved open-source software stack

## Data Access - target state:

- all NOAA datasets accessible online via services
  - per Presidential mandates and NOAA policy
  - NOAA data made available to the widest community possible
- small number of data access service types
  - services as consistent and standardized as possible
  - services that support machine to machine exchange
  - services not biased towards any particular client application
- shared hosting option

## Data Access - potential next steps

- Improve UAF holdings, organization and/or reliability for relevant data types
  - UAF outreach and work with Data Centers - could be expanded to monthly workshops or training sessions, etc.
  - UAF to produce best practices for data providers who want to serve data via TDS
  - UAF to create TDS catalog rubric to describe quality of TDS catalogs
  - UAF to utilize ERDDAP capabilities to server in situ data/DSG feature types
- Improved links to access data from catalog &

## Data Usability - Current State

- Inconsistent approach to metadata
  - missing, incomplete
  - unstructured [text, HTML, PDF]
  - variety of standards - FGDC CSDGM standard, ISO 19115/19139, NMFS InPort
- Partial deployment of metadata metrics
  - not all metadata records are known & scored
- Variety of output formats
- Limited use of controlled vocabularies
- limited browse imagery and on-demand visualization

## Data Usability - Target State

- Comprehensive metadata based on international standards
- All available metadata scored
- All data available in at least one broadly-supported format
- Common or controlled vocabularies used
- Web Map Service endpoint for all relevant datasets

## Data Usability - potential next steps

- Agree on small # of target formats
- Complete & updated scoring of metadata
- Begin establishing Web Map Services
- Agree on vocabularies

# Preservation and Citation - Current State

- not all data archived
  - ship data not getting to archive
  - PI/project data at local offices
- datasets not citable
- algorithms/code not preserved
- need to save all intermediate products -- cannot reprocess

## Preservation and Citation - Target State

- routine & complete archiving
- persistent identifiers, with landing pages and citation guidelines
- all algorithms/code preserved & document
- on-demand reprocessing capability

## Preservation and Citation - potential next steps

- Continue Data Citation project
- Simplify archive submission