

## Session 1 - Intro and Overview

Tuesday, June 25, 1:00 to 2:30 PM (Eastern)

Workshop Chair: Jeff de La Beaujardière, EDMC Chair

Keynote Speaker: Peter Colohan, OSTP

### Keynote: Peter Colohan, OSTP:

- “Data and information are the lifeblood of environmental decision making.”
- The importance of data is starting to be appreciated by a much broader audience.
- Presentation title (emphasis) - NOAA environmental data in the context of multiple data initiatives across the federal government
- Policy environment:
  - 1st. Open Gov Directive from 2009. Require agencies to take specific steps in transparency and move towards open data. Improving accessibility of data to the public is critical.
  - 2nd. Public Access to Research Results and Executive Order on Machine Readable Data
  - Working toward - Access, Use, Discovery of environmental data
  - NOAA a leader in environmental data & integration (referenced Adm. Lautenbacher's leadership in establishing GEO)
  - 3rd. Interagency Group on Earth Observations (to support GEO) is the backdrop for OSTP's work in its National Strategy for Earth Observation. Focus on facilitating data management frameworks across the agencies - is a common framework possible?
  - Big Earth Data Initiative (BEDI) is not “Big Data” (only)
    - NITRD is focused on Big Data
    - USGEO/BEDI main priority is to get high impact data into machine readable formats - with NOAA, USGS & NASA and others. Shared approach toward interoperable data.
    - USGEO hoping to create an environment to help staffers get their job done.
- Open Data Initiative
  - Specific effort catalyze external stakeholders to address specific challenges
  - 6 focus areas:
    - Education, energy, health, global development, public safety finance
    - A separate set of Climate data initiatives per Obama's announcement today
  - Curation of data on data.gov related to specific user challenges
  - Produce tools to support these user needs
  - Demonstrate economic value of government-generated value (per Todd Park, CTO)
  - Resource hubs to acquire data to meet their needs
  - Greater consolidation & integration of data
    - e.g. ECO INFORMA (DOI, EPA)
      - BISON project - biodiversity information
      - National Atlas for?
    - Ocean-related decision making - National Ocean Council
    - Global Change Information System (GCIS)
      - Access point for National Climate Assessment
      - Resilience Toolkit - climate adaptation toolkit based on Climate.gov

- resources
  - Adaptation Clearinghouse
- 4-levels -
  - 1. policy directives
  - 2. open data initiatives
    - Very public
  - 3. Information system initiatives
    - aimed at specific user groups
  - 4. Big Earth Data Initiative
    - designed to make environmental data interoperable
- 1 primary thing to do. BEDI should work together with the high-level policy directive.
- FGDC:
  - This is the maps and GIS world, primarily 2D data.
    - Starting to make a lot of progress with env data in GIS world. This is challenging.
- OSTP is attempting to integrate all of these policies and initiatives so that they are coordinated. There are many dimensions of this challenge.

**Questions for Peter Colohan:**

1. What is the purpose of the upgrade to Data.gov? What are the new features of Data.gov?

(Nancy Grady)

---new features will focus on more harvesting of resources

2. Is there a written description of the organizational soup just described? (Chris Calloway)

---No

3. The BEDI is described in a strategy document and Peter's blog. What the next steps of move strategy to action? (George Percivall)

---will create a data management WG within USGEO

4. When will the BEDI strategic vision documents be available for release? (Ed Kearns) Can links be shared? (Dan Kowal)

---review by agencies this summer (hopefully) as part of the Civil Earth Observation plan

5. Regarding the New Open Data Policy requirements for creating an agency-level enterprise data inventory - what is the difference between this and what NOAA already does to publish our data via data.gov? (Kim Jenkins)

---data.gov curates data to address specific questions

---NOAA data catalogs are curated more generally that can be harvested to answer question

--- Colohan unable to give an answer right now - will get back to us

--- FY 14 budget requests funds for data management specifically for NOAA and a few other agencies

6. AS USGEO is rechartered, will the Architecture and Data Management WG be re-established? (John Lever)

---Yes & No, new name will be Data Management WG

7. What's a flash catalog? (Dan Kowal) Jeff mentioned this during Peter's talk about data.gov connecting to geoportals. He said when a "flash" catalog is established, then there's no longer a need to connect to geoportal.... Maybe I heard it wrong.

---No answer given need to follow up with Dan.

8. I wanted to ask the OSTP guy about "Taxonomies, Ontologies and Web Semantics" - From Harry Tabak after the session. His specific question was:"Is there any initiative under way to develop a taxonomy, ontology, or use web semantics to organize this mass of data." Need to follow up.

**Actions:**

- Need to post links to the documents that Peter referred to during his talk
- 

**Jeff DLB's Talk - NOAA's EDM Activities:**

- Public Access to Research Results - copyright issues being addressed; addresses grantees' requirements for compliance
  - Draft plan by 8/22/13
  - NOAA Committee (Jeff DLB, Neal Kaske co-chairs)
- Open Data Policy (Executive Order 2013-05-09)
  - Business data and Research data
  - Inventory of agency data assets at [agency.gov/data](http://agency.gov/data)
  - Initial deadline is 11/9/2013
  - Use of open standards and machine readable formats
- Big Earth Data Initiative
  - Coming FY2014 - zero or more dollars
  - focused on high-value Earth observations
  - improve Discoverability, Accessibility, Usability of Data
- NOAA Environmental Data Management Committee
  - Authority to write procedural directives
    - Data Management planning PD
      - archiving (what & how)
      - documentation (how to use ISO 19115)
      - sharing (share w/in 2 years)
      - Data Citation (in progress)
      - Data Access (in progress)
  - Reports to CIO Council & NOSC
  - One member/line office
  - Data Management Integration Team (less formal, larger group)
    - E-mail Jeff to get on this e-mail list
- Resources
  - Budget
  - Personnel (better recognition of contributions)
  - Annual Workshop

- DMIT
  - Standards
    - ISO 19115/19139 Metadata
    - Formats and Services (per Open Data Policy)
    - Vocab (CF)
    - Quality Control
    - Security
  - Architecture
    - Avoid purpose-built systems
    - Need a data services layer (compatible formats & vocabs, documentation, data access services, search & discovery services)
  - Cloud
    - Use cloud to access NOAA data
    - Retain 'master' copy of NOAA data in house
    - Use also for internal NOAA activities
  - Assessment
    - EDM Dashboard - summarizes NOAA data management progress
    - Observation System of Record EDM Assessment
  - Data Lifecycle
    - Planning/Production
    - Data Management
    - Usage
    - Citation - DOIs assigned
  - All NOAA Data should be
    - Discoverable
    - Accessible
    - Usable
    - Preserved
- 

#### **National Geospatial Assessment - Kim Jenkins**

- 148 NOAA Systems Sampled
  - custom metadata
  - non-standardized

Jeff mentioned this during Peter's talk about data.gov connecting to geoportals. He said when a "flash" catalog is established, then there's no longer a need to connect to geoportal.... Maybe I heard it wrong. (Dan Kowal)

---

Is there information on how data sharing requirement will be monitored? (Kelly Carignan)  
---the Procedural Directive covers some of this but still working on how to monitor it effectively.

Is NOAA minting their own DOIs? (Mark McInerney)

--- Yes

What is BEDI? (Mark Miller)

--- Big Earth Data Initiative

what about model data? I did not see it mentioned in the data layer (Georgi Kostov)

---They are both a data source and data consumer. Important.

Applications that plot everything tend to be often sluggish and hard to use as they are set up to do too much and can't be specific. Will this issue be addressed when mandating the use of certain map/web services? (Cathy Smith)

---Want to make data visible via mapping services (e.g. WMS). Enabling one or more map layers is important

Where is the link for the Data Archive Procedure? (Lauren Dolinger Few)

- all PDs can be found here: <https://www.nosc.noaa.gov/EDMC/PD.all.php>

What was the link to the edm dashboard? (Georgi Kostov)

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

What authority does the EDMC have to enforce PDs, particularly on NOAA systems? (Carmel Ortiz)

- authority to write
- CIO enforces

Have you implemented a cloud solution, or did your slide just represent a vision? (Nancy Grady)

- Just a vision on Jeff DLB's side, there have been some pilots within NOAA

Is NOAA intending to set standards for geospatial software (ESRI) or set standards for data (OGC)? (Chris Paternostro)

- Hybrid approach - using a specific protocol, using the named standard with some solutions that have been given a green light by the NOAA security community

Can you give me an idea on the granularity wrt TOI, as to what constitutes a dataset? In other words, do you have idea, order of magnitude, how many data sets exist right now? (James Potemra)

---coarsest granularity possible; not sure of number of datasets in existence.

NODC stewards roughly 100000 accessions at present (Deirdre Byrnes)

---coarsest

---per Ken Casey, 25000 DOIs in the near term

---

### Non-substantive notes

\*\*Technical adjustments for Days 2 & 3

- Stress again the usage of Chat (logistical) v. Questions (substantiveGoToMeeting controls are different on iPad per Ken Casey
  - Polls don't work on iPad
- poll time is too long (1 minute would be appropriate)

- best practice during webinar is to have slides
  - need to clarify within Explorer the interface is different - a chat & a question feature combined - best practice - use Firefox or Chrome?
  - have a “emergency call” number for urgent questions? trouble connecting - someone’s cell phone?
  - need URL to slides at start of every presentation
- 

## Q & A for Session 1 - Intro and Overview

1. Questions for Peter Colohan:
  - a. Questions were answered using the GoToMeeting Webinar feature.
2. Questions for Jeff DLB:
  - a. Kelly Carignan: Is there information on how data sharing requirement will be monitored?
  - b. Georgi Kostov: what about model data? I did not see it mentioned in the data layer

## Session 2 - Catalog and Search

Tuesday, June 25, 3:00 to 4:30 PM (Eastern)

Session Chair: Kim Jenkins

### Current State (Kim Jenkins)

- Catalogs & Search - Key Goals
  - Users should be able to find the data they need when they are looking for it
  - Users should be able to use their preferred tools
  - Use standard formats & metadata
  - Create & Register Once, Used & Visible by Many
  - NOAA leadership can see improvements in discovery & access
- Current State
  - No complete inventory or one-stop shop
  - Many portals, potentially overlapping - Portal Palooza
  - Various catalog implementations (Geoportal Server, 1+/data center)
  - Limited integration with commercial search engines

### Ignite Talks

#### Data.gov and the GeoPlatform: Piecing Together the Puzzle - Adam Bode and Randy Warren, NOS/CSC

- The Hunt for Data doesn't give info on quality, source, timeliness
- Data.gov - active Ocean Community - 120+ NOAA data/tools
- NOAA vision is to allow tools to use data through services
- Past - catalogs stood alone - Today - all integrated thru CKAN harvesting
- National Geospatial Platform - NOAA GeoPlatform, ArcGIS Platform & National GeoPlatform
  - transitioning NOAA GeoPlatform to National GeoPlatform
  - Uses CKAN & ArcGIS
  - NOAA will be a community on the National GeoPlatform (coming in August 2013)

#### The Cryptkeeper Speaks! Overcoming the Scary in Data Discovery - Christine White, Esri

- Disparate Systems
- Overarching Portal
- Different protocols, formats
- Data & services brokering to the rescue! (Geoportal Server)
- Harvest between portals
- Supports federated search between portals
- Web services - API gives status of a given service thru FGDC
- Huge volume of data
- Data inundation
- Faceted search integration - Geoportal integrated with Apache Solr
- Hosting - Esri Managed Services - Cloud GIS solutions
- New version of Geoportal to be released imminently

## Discovery and Access of International Satellite Data Collections Using the CEOS WGISS

### Integrated Catalog (CWIC) - Ken McDonald, NESDIS/OSD/TPIO

- Discovery & access to EO satellite collection
- CEOS - global space agencies
- WGISS - focuses on standards & collaborative development
- GEO - GEOSS architecture - CEOS is satellite arm of GEO
- CWIC is a virtual clearinghouse working with CEOS agencies
- CWIC has no European partners - European developing their own federated search capabilities (FedEO)
- CWIC-Start Client
- CWIC/FedEO - common open search interface
- More information - <http://wgiss.ceos.org/cwic>

### WMO Information System and Global Information System Centre Washington - Walter Smith?

- WMO - Part of the UN - 191 member states & territories - NWS is US designee
- WMO Congress (2007) directed to implement the WIS - a global information system for the management for weather, water and climate data
- WIS Structure - 3 types of centers - GISCs, DPCs,
- Comprehensive DAR catalog
- WMO will adopt ISO 19115 metadata standard

### Target State

- 2013 Presidential EO \* Open Data Policy requirements
- Enterprise data inventory for public & non-public data
- Use common core metadata to document the data
- Automatic registration
- Help NOAA improve how it develops/maintains multiple catalogs, including federating
- Utilize Open Data Policy initiatives on GitHub

### Potential Next Steps

- Harmonizing metadata - using Google?
- Establish a Master NOAA Catalog?

---

## Q & A for Session 2 - Catalog and Search

1. *Question for Walter from Kevin O'Brien:* Are there any plans for WMO to require improved or more complete ISO metadata to be included with data being sent through GTS?

Walter Smith - WMO initiative to place metadata in the catalog is new. (1.2 version WMO Core Metadata)

2. *Question from David Fulker:* Given that a key factor in Google's success is its complete independence of catalogs and metadata (standardized or otherwise), is any of NOAA's info-discovery strategy focused on exploring alternative approaches, perhaps the union of crawling and content-summarization services?

Jeff DLB - NOAA needs to support structured catalog as well as search engines. Need to look at schema.org approach - provide structure to search engines as they search resources

Response from Ken C: Not true! We expose all our metadata, in HTML form, to Google and other bots.

Response from Bob S: Google reads and "understands" ISO 19115/19139 .xml files which it finds. It plucks out values from certain tags intelligently.

Response from Ken C: Bob, can you point us to more info about Google's ISO 19115/19139 abilities? I guess I could google it! ;-)

*Jeff DLB:* Potential next step: explore/research how can we leverage potential commercial options and search engines.

3. *Question for Adam Bode from Anna Milan:* How does geoplatform recognize map services from data.gov? Does it query urls in metadata?

Adam Bode - Queries URLs in metadata & makes distinctions in formats. CKAN reading both data.gov & geoplatform. Content being migrated to national geoplatform.

Kim Jenkins - If metadata for services are currently (or have been previously published) to Data.gov, Geo.data.gov, the NOAA GeoPlatform, etc. you don't need to register them again to make them accessible by the GeoPlatform.

4. *Question for Christine White from Anna Milan:* Can ArcGIS portals monitor a variety services? E.g thredds?

Christine White - Service Checker leverages Service Checker API. Lists services that are supported. From Christine: <http://registry.fgdc.gov/statuschecker/>

5. *Question from Dan Kowal:* Is the message that our individual instances of geoportal with their capability to federate, is that not adequate enough?

Jeff DLB - If the geoportal catalogs are able to federate, it \*may\* satisfy the requirement to have a data catalog (but it might not). The answer is not clear - more can be done to see if this is a possible approach.

6. *Question for Adam Bode/Randy Warren:* Does geoplatform have an interface for providers to manage holdings and descriptions?

There is an admin dashboard for providers to manage their holdings.

7. *Question from Cathy Smith:* Not all data is geographically based. For example, climate indices such as the PNA. Would these geocatalogs be able to handle non geographically based data?

CKAN can handle non-geographically-based data, eg Data.gov might have Excel spreadsheets.

More from Cathy Smith: or other indices such as the global temperature trend. I am just asking about non geographic based data.

Anna Milan: Some work has been done on this and NOAA is currently working on an approach for dealing with non-geographic data with ISO metadata standards; We should be including these into data.gov (via the CKAN metadata admin interface).

8. *Question from Dan Kowal to Ken McDonald:* In Ken's slides, he showed how CWIC is tying into CLASS, how about through the data centers if we have CSW available?

Ken Casey sez: NODC will be providing to CWIC the ocean holdings we steward using CLASS. We just started working on Jason-2 GDRs this week, for example. Already do all the GHRST

products (but those are archived locally at NODC)

9. *Question from Sarah O'Conner:* Related to question 5 above. Not all NOAA data is captured through the Geoportals. Has there been any work to tie NMFS InPort into the Geoportal federated searches?

The Geoportals reside at the data centers and not much fisheries data is making its way into the data centers. As such, InPort & Geoportal are not tied.

Ken Casey sez: Jeff said not much fisheries data is making it into the Data Centers... that isn't quite true. We have over 1200 accessions matching the search term "NMFS" in the NODC archive: <http://data.nodc.noaa.gov/geoportal/rest/find/document?searchText=NMFS&start=1&max=25&contentOption=intersecting&f=searchPage&expand=True>

I would agree that there is MUCH MUCH more to get into the archive, however.

#### **Catalog and Search Verbal Q&A:**

10. *Question from Kevin O'Brien:* General question... If someone has a catalog is there a place that we can register them so that they can be discovered?

- Jeff DLB: there should be a master NOAA catalog. There isn't such a place yet.
- Kevin O: could be a good next step in this area
  - Ken Casey: What about the EDM Wiki page on Geoportal Server? Many are already listed here: [https://geo-ide.noaa.gov/wiki/index.php?title=ESRI\\_Geoportal](https://geo-ide.noaa.gov/wiki/index.php?title=ESRI_Geoportal)
  - Others note it might be worth investigating further - a Next Step

11. *Question from Jaci Mize:* have we thought about a "landing page" that points maybe by topic/organization to existing noaa catalogs...we all have various standards, implementation of standards and branding we want to preserve?

Kim: If we have existing multiple requirements for this, let's kill multiple ones at same time if possible (i.e. noaa.gov/data, noaa.gov/data/catalogs, noaa.gov/data/environmental, etc.)

### Session 3 - Data Access

Wednesday, June 26, 1:00 to 2:30 PM (Eastern)

Session Chair: Kevin O'Brien

#### Current State (Kevin O'Brien)

- Wide variety of data generated by NOAA & its partners
- Current state:
  - not all data available online
  - not all data available via web services (no subsetting)
  - variety of access methods (both 'standard' and custom)
  - multiple points of access for some datasets
  - with exception of archives, groups have to serve up their own data (how & what software)

#### Ignite Talks

##### *Doing More with Less - Bob Simons, NMFS/SWFSC*

- NOAA data used by thousands, therefore, impact potentially large
- Efficiencies -
  - Reusable Software
  - Free and Open Source Software (FOSS)
  - one example, ERDDAP - free data server/broker; Google scanned; supports gridded, in situ, other data
  - If you need additional functionality - it can be requested & then used by you & others
  - Turn a big job into a small job

##### *Designing and Prototyping the Cloud-Enabled Hyrax Server - James Gallagher, OPeNDAP*

- OPeNDAP in the Cloud
- Data sub-sampling engine - reduces file download size
- OPeNDAP solution - Data Store (HTTP GET & HEAD requests) - simple, virtual hierarchical organization of data
  - wanted to avoid vendor lock-in - developed in-house XML catalogs
- Result - OPeNDAP server at modest cost

##### *Relevant advances in Sensor Web Enablement and Semantic Mediation - Nadine Alameh, OGC*

- What in the world was I thinking to cover both sensor & semantic web in one talk?!
- Sensor Web Enablement standards fairly mature, with many implementations (e.g. NASA SensorWeb, NOAA IOOS)
- Sensor Planning Service Standard implemented by ESA
- EEA receives environmental data from its 32 member countries via the SWE standard
- Implementations - compliance testing thru OGC
- Billions of sensors - need to take advantage of peer-to-peer mediation, semantic mediation for discovery
- Trying to enable cross-community interoperability

##### *NOAA's HPCC OGC Service Hosting Environment - Micah Wengren, NOS/OCS*

- Open Source based project
- Who Should Use This? - small offices with limited budgets, no IT resources, @noaa.gov email
- <http://services.ogc.noaa.gov>
- GeoServer Capabilities
  - scriptable data upload
  - flexibility of SQL view layers
  - robust styling capability
  - experimental OGR data store
  - ability to work with raster data sets
  - Scripting & WPS
  - Upcoming - nedCDF plug in
  - Add-ons - WMS animator

*NWS Integrated Dissemination Program - Mark Miller, NWS (speaking on behalf of Lou Cano)*

- NOAA Approach to Data Dissemination
  - satellite, web, on demand, subscription, push/pull, point-to-point, middleware, radio
- Customer base varied
  - federal, state, local governments, academic, commercial
- Different standards used
- NOAA Integrated Dissemination Program focuses on a integrated enterprise-wide dissemination service
  - Phase I (near term) - framework for NWS services
  - Phase II - integration of NWS & NOAA services
  - Phase III - NOAA-wide dissemination capabilities

**Target State** (Kevin O'Brien)

- All NOAA datasets need to be available online
- Need to be made available via consistent data access service types
  - consistent & standardized
  - support machine-to-machine exchange
  - no bias toward a client app
- Shared Hosting Option
- Authoritative source v. republication
- Pre-configured software stacks for self-hosting

**Next Steps/Actions**

- from the UAF (under DMIT) Working Group
  - Improve UAF holdings, org and reliability thru UAF Master THREDDS catalog
  - Outreach to NOAA data centers to bring data into UAF
  - Produce best practices for data providers who want to serve data via OPeNDAP
  - Create tool to 'clean' remote catalogs
  - Provide rubric to clearly display quality of data catalogs (currently limited THREDDS data server catalogs)
  - Utilize ERDDAP, especially for serving collections of in situ data

- Possible next steps:
  - Work with Luis Cano to establish standard access methods that will be incorporated in IDP, so that EDM doesn't devolve to a myriad of silos.
  - Might look at NASA's [Standards Process Group](#) (SPG), for that software stack. Or ideas for how to select the software.
  - Improving the link to access data from the catalog/search results
    - rubric for metadata check data access link, with stats/analytics (possibly in the dashboard?)

---

## Q & A for Session 3 - Data Access

1. *Question asked during Micah's talk:* Does GeoServer work with netCDF 4?
  - a. *Response from Micah:* I'll have to check on netCDF 4 compatibility of GeoServer. It is a newly developed plugin that isn't yet a GeoServer community module. It was developed for Navy oceanographic data.
  - b. *Update from Micah on 6/27:* NetCDF versions 3,4, and 5 are supported, the plugin uses UCAR 4.2 libraries under the hood, so whichever versions are compatible with these libraries will work.
  
1. *Another question for Micah:* What is "OGR?"
  - a. OGR <http://www.gdal.org/ogr/>
  - b. Another link: [http://www.osgeo.org/gdal\\_ogr](http://www.osgeo.org/gdal_ogr)
  - c. OGR is part of the GDAL library - it is for point data
  
1. *Question asked during Mark's talk:* Is this the "NOAA" Integrated Dissemination Program or the "NWS" IDP? Looks like only data from a single LO.
  - a.
  
1. *Question asked during Mark's talk:* Also, is it only meant to serve real-time, operational data/customers?
  - a. Mark is not sure.
  
1. *Question for Mark Miller from David Fulker:* thinking that significant parts of IDP might need to employ "push" and to verify receipt (by subscribers), is your emphasis on Web services complemented by something better matched to those kinds of needs?
  - a. most definitely, customers have varying needs
  
1. *Question asked during Target State Discussion:* Please define "authoritative source."
  - a. the source that claims ownership of the data
  
1. *Question asked during Target State Discussion:* Can we avoid the use of the word "ecosystem" to refer to software development activities? NOAA and other environmental orgs already study real ecosystems in nature and have an existing definition of that term. Plus, ecosystem includes predator/prey relationships.
  - a. From Ken C: "Ecosystem" is used broadly in the world of Big Data. Even if we avoided the

use of phrases likening the data management world to ecosystems in NOAA, it would still be used widely outside NOAA.

1. Can NOAA Data Access services and catalog requirements be included in the NOAA Integrated Dissemination Project (Mark Miller Ignition presentation)? For example, will OPeNDAP and ERDDAP be included in IDP?
  - a. looking at all aspects and are open to integrating other access methods
  
1. A general question: In defense of stovepipes, I do not always see the functionality of touted services as matching the functionality of stovepipes or issues in using those services addressed or even acknowledged.
  - a. *Comment from Deirdre Byrne:* Stovepipes are cheap to implement and very expensive to maintain. Not a good long-term solution. They typically don't support interoperable or standard APIs.
  - b. *Comment from Scott Cross:* Re: stovepipes: if the questioner means crafting a custom end-user experience in interacting with data, then there is definitely value in preserving that kind of thinking. However that can be done on top of the kinds of integrated architectures we are talking about here. (Bob Simons adds: Yes! And using reusable software to offer basic data access (quick and easy) responds to the needs of users who want a standards-based way to access the data **and** usually makes it much easier to build custom solutions.)
  
1. What is the status of server-side query by data value via OPeNDAP or ERDDAP services? For example, return Salinity along an isopycnal surface.
  - a. OGC Met/Ocean DWG has developed Best Practice for data on computed surfaces. Applies to WMS but could be generalized. (-Jeff DLB)
  - b. Bob Simons says: For gridded data, ERDDAP extends the DAP standard to support queries base on domain values, e.g., if sst is sst[time][latitude][longitude] then a standard DAP query like sst[14][22:76][106:144] can also be requested in time/lat/lon units as sst[(2000-04-03)][(20):(40)][(-135):(-115)].  
For tabular data, ERDDAP supports sequence queries. Please see my answer to question #11 below. But if you are asking DAP/ERDDAP to do calculations (interpolations) to figure out the depth (at each point) of the isopycnal surface and return that to you, then sorry, no. ERDDAP offers a few, simple server side functions, but not an extensive system for server side calculations. Roy Mendelssohn has been interested in adding extensive server-side functions but hasn't yet seen a good approach.
  
1. *Question for Bob's talk:* what do you mean that THREDDS is not suited for in-situ data?
  - a. See <http://coastwatch.pfeg.noaa.gov/erddap/images/erddapTalk/TablesAndGrids.html>
  
1. *Question for James Gallagher:* Isn't storing large datasets in the cloud very expensive compared to storing it locally?
  - a. *Comment from Deirdre:* Cloud storage is relatively cheap. Downloading is expensive. That's why using a subsampling service is so key. We are implementing a NESDIS OpenStack pilot cloud this summer.

- i. I'd be happy to provide some slides detailing our analysis of Amazon costs to the questioner, if we know who that is. (The questioner is Bob Simons, and I guess "cheap" is relative. The current cost to store data on Amazon S3 is roughly \$80/TB per month! For that money, I can buy (just once!) almost 2 TB worth of drives to store the data locally and redundantly, or buy 1TB within a RAID. We are a small shop and have ~50 TB of data. How much do the data centers have? I bet the monthly price to store that on S3 is "significant". Note that Amazon doesn't guarantee it won't lose the data (and even if they did, what would they do?, give you \$80 for that data that is now lost forever?), so you still have to archive the data locally. And if the data is local, I don't have to pay Amazon bandwidth fees to re-serve it. Yes, I would like to see your calculations related to this.)
          - b. Comment from Mark Miller: cloud usage and implementation is a cost-benefit consideration. Depends on what the requirement, CONOPS and usage.
          - c. *Comment from Ken C:* And the download costs are why allowing people to associate their virtual machines with the data stored in the cloud is important too, along with allowing people to do OPeNDAP-style downloads.
1. *Question on Target State from Dave Fulker:* to what extent might the notion of data "access" need to be expanded to include a (growing) set of pre-retrieval functions provided as an integral or layered part of data acquisition services?
2. *Question for Gallagher from Jeff DLB:* Can you say a few more words about the open-source solution you mentioned as an Amazon S3 alternative? Could/should NWS IDP use it on an internal government cloud?
  - a. OpenStack most visible open source. Has its own object store system and supports S3. Question is can/should NOAA run it's own OpenStack? It could be great thing for NOAA to have their own cloud services but not develop the software.
2. What are the most important OGC standards development activities to pay attention to?
  - a. Geo package
  - b. LifeForce for use on mobile
  - c. SWE profile - profiled for different domains
  - d. OGC meteorology/oceanography - time dependent and elevation dependent surfaces
2. Possible Next Steps: Work with Luis Cano to establish standard access methods that will be incorporated in IDP, so that EDM doesn't devolve to a myriad of silos. (NOTE: moved this to next steps in Notes doc)
  - a. Comment from Harry T.: IDP has been adopted by NOAA CIO.
  - b. Comment from James G: Might look at NASA's SPG for that software stack. Or ideas for how to select the software.

## Session 4 - Data Usability - Metadata, Standards, Visualizations

Wednesday, June 26, 3:00 to 4:30 PM (Eastern)

Session Chair: Anna Milan

### Current State (Christina Lief)

- Data Centers struggling to provide & update metadata for data sets, products
- Metadata not always complete or in correct format
- Metadata needed to facilitate discovery/access & is crucial to data integrity (esp. WRT climate data products)
- OSTP mandate for full public access, in a format that emphasizes interoperability
- Challenges:
  - find standards that support all data producers
  - producers do not typically generate metadata for their data sets/products
  - metadata needs to contain info on how product was generated for reproducibility, provenance, algorithms, QC
  - need metadata catalog & IT infrastructure to support it
  - need to share metadata more broadly
- NOAA Needs:
  - common standards
  - common data formats
  - common vocabulary
  - training
  - data visualization
  - cross-org coordination

### Ignite Talks

#### *The ISO 191\*\* Series: Putting It All Together - Jaci Mize*

- 191\*\* series represents both extensions and individual standards
- ISO is migrating from MD to MI
  - MI Metadata includes Acquisition Info for gridded data
- Service Identification (19119) can be substituted for Data Identification
- Citation URL can link to 19110 Feature Catalog
- 19111 Spatial Reference system info

#### *EMMA - Metadata components, metrics & templates - Anna Milan*

- Enterprise Metadata Management Architecture (EMMA)
- Documentation Directive states that NOAA is to use ISO 19115-2
- Common Problems:
  - Standard is too complex - human readable views
  - I don't understand the order... - ISO Explorer on EDM wiki
  - How an I improve the metadata - Completeness tests
  - Editing the same section in 100s of records - Docucomp
  - I don't include URLs b/c they always break - URL checker
  - How come my metadata is not getting published - validation reports

- I already document my data in a different standard -translations to ISO
- I need persistent identifiers... - DOIs included in metadata & recommended citations/ landing pages
- As a chief, I want to see a complete list of our data holdings & their completeness - Diagnostic and Metric Visualization
- Need more metadata from across NOAA to fulfill requests from NOAA integration office

*Human data sources: Integrating multiple data streams - Mark Brady*

- Unique properties of NOAA;s commercial fishing data
  - humans are primary observers
- data entry errors, intentional misrepresentation
- multiple data streams
  - naming schemes different
  - different levels of reporting
- Fuzzy matching (introduces errors)
- Choose one stream or another for analysis (give up analytical capabilities)
- Integrated reporting System
  - auto generated Trip ID/Vessel ID
  - vessel can report on subtrip
  - vessel reports dealer & pending sale
  - system notes transaction with dealer (internet dating for fishing)
- QA replaces QC

*Data and workflow provenance for traceability, reproducibility, and scalability in Integrated Ecosystem Assessments - Stace Beaulieu*

- Provenance metadata - Use Case: NE Shelf Large Marine Ecosystem Status Report
  - cannot access the data or metadata in the report
- Software design to track provenance - PROV Data Model
- An entity can be a single data product or multiple data products
- iPython Notebook - natively code in MatLab or R - can be shared, output as script, pdf or webpage

**Target State**

- Comprehensive metadata using international standards
- All available metadata scored in EMMA
- Common or controlled vocabularies
- All data available in at least one broadly-supported format
- Service end points for data access (WMS?)
- Integration and reuse of existing tools & methods for metadata development/publication
- Metadata as a means to an end, not an end in and of itself

**Next Steps**

- Get your metadata into EMMA
- Establish & id more web map services
- Learn how to document with the standards
- Support OSTP initiative for noaa.gov/data catalog by Nov.
- Agree on common vocabularies

- Automate individual steps in NOAA's business processes - verifiable using software

#### Discussion

Looking for more benefits of metadata - NEEDS more compelling proofs  
Recommendation: proper use of data it needs proper metadata

JDLB - best person to *create metadata* is the person who finds the least benefit

*Fulker - provo tool has a focus on usability, the presence of metadata made the thing work. iPython notebook enables \_\_\_ and motivates whatever it takes to \_\_\_\_.*

Christina Lief: To find a link to the ATRAC metadata form go to <http://gosis.org> On the left hand menu look for the ' Metadata Creation Form' entry. This will take you to information about the form and a link to it. <https://www.ncdc.noaa.gov/atrac/index.html>

---

## Q & A for Session 4 - Data Usability - Metadata, Standards, Visualizations

### Questions and Comments asked during the Current State discussion:

1. Doesn't the EDMC Documentation Directive already provide common metadata standards?
  - a. The documentation directive wants
1. *Comment from Ken C:* Christina - you said we are working on a "common metadata standard for NOAA". I think that is not quite the right way to state what is going on across the three Data Centers... I think we are figuring out how to commonly apply the ISO 191\* family of standards. We are not developing a new standard. Right?
  - a. Developing common core collection-level, discovery level profile/best practices? Established at NCDC, but extended to other data centers and data.gov
  - b. Best practice - template - NOT a new standard

### Questions and Comments asked during the Ignite Talks:

1. *Question for Jaci:* How do we ensure that metadata elements in ISO are consistently populated by all NOAA data producers? Metadata is not useful if everyone defines what it is differently.
  - a. Docucomp and components are best way to be consistent. Jeff: Components important when already populated. EMMA completeness score may help.
  - b. Schematron - consistent check for content
1. *Related question from Sarah O'Connor:* Common core NOAA metadata standard?? This is the first I've heard of it. I thought this effort was Data Center specific. What efforts are you doing to gain input from all of NOAA? To vet what you are developing among the entire NOAA community.
  - a. Working with Data Centers first and then reaching out to line offices. A first draft will come out sometime this summer.

1. *Question for Christina:* Given all of the ISO elements presented by Jaci, does the NCDC TurboTax approach take into account all of these elements or is it a subset? Reason I ask, is given the heterogeneity of data NGDC manages, we've taken different approaches for applying ISO depending upon the dataset. We're wondering about the applicability of the Turbo Tax Form: what general use case does it satisfy, what are its limitations?
  - a. Form is from ATRAC. In v 2.5 the form can be completed without archiving at NCDC including other agencies
  - b. *Comment from Jaci:* from what I have seen it does not fully support all of the options -- subset. But it has a good base.
  - c. Anna - Identification, Distribution, Quality; approach to iterate. doesn't include service, content info
  
1. *Question for both Anna and Christine:* Are the NCDC and NGDC metadata efforts complimentary or are they separate? It seems like they are two separate systems.
  - a. Working together to develop common standards
  - b. NEIO = National Environmental Information Office (Pre-decisional, NOAA Use Only)
  - c. Anna: Currently, no interoperability between EMMA and ATRAC
  
1. *Question for Mark:* Are research vessels currently using the system for integrated reporting in order to define standardized trip ids? If so, can any vessel use this system?
  - a. Integrated reporting not being used anywhere. The biggest need is in the fishing area.
  
1. *Question for Stace:* How does PROV model map to lineage portion of ISO metadata standard?
  - a. A lot of the info can be included, but still not enough of PROV in ISO
  - b. Will include mapping of PROV-O and ISO lineage on the NOAA EDM wiki
  - c. It may also be possible to link to a PROV record from the lineage section of ISO. See, e.g., [https://earthdata.nasa.gov/wiki/main/index.php/PROV-ES\\_Earth\\_Science\\_extension\\_to\\_W3C\\_PROV](https://earthdata.nasa.gov/wiki/main/index.php/PROV-ES_Earth_Science_extension_to_W3C_PROV) (-Jeff DLB)
  - d. *Comment from Curt Tilmes:* Someone asked about aligning W3C PROV and PROV-O with ISO lineage. There will be a session on an Earth Science extension to PROV (PROV-ES) at the ESIP Federation Summer meeting. We've been discussing that issue a bit.
  
1. Do any of the speakers have recommendations for educational programs to pick up more expertise in metadata?
  - a. *Comment from Ken C:* Sign up with NODC/NCDDC's Jaci Mize!
  - b. From Jaci: <http://www.ncddc.noaa.gov/metadata-standards/metadata-training/>
  - c. EDMC Wiki
  
1. *Question for Stace:* So are the links behind the "data" and "code" actually using the Prov-o via a web service to obtain both? Wonder how that's orchestrated in the document you shared.
  - a. Prov-o backend is being developed
  
1. In Anna's presentation she showed some metrics (graphs) and seemed to make the point that metadata were critical for data use. Are there any metrics (graphs or otherwise) that can demonstrate how the use of a particular data set increased after metadata (standard or otherwise) was included?

- a. When we assign DOI's and establish landing pages they may show some usage statistics
  - b. *Comment from Don Collins:* the NOAA library has experience in doing bibliographic citation analysis, so a 'before-after' analysis might show some changes in data usage.
  - c. *Comment from Dan K.* Use Google Analytics?
1. *Question for Anna from Ken C:* WMS won't be suitable for all NOAA enviro data... not all enviro data can be mapped (for example, time series statistics for a global dataset, space weather data, etc.)
  - a. Aware of that use case.
  - b. More generally, can we provide visualize/browse services to improve usability (e.g., to get an idea of the data before actually requested download or subset)?
  - c. Ken C: The ability to provide WMS service endpoints for all data is not within available resources, by a LONG SHOT. The archives are full of inherently mappable, but not well standardized data. To convert it into a format that WMS will work with is non trivial, and would involve extensive human involvement and therefore WAY MORE \$\$\$ than we have.
1. Are these "standard vocabularies" for keywords? What metadata components exactly do these vocabularies standardize?
  - a. *Comment from Jaci:* vocabs are often community specific but good rule of thumb is to use GCMD science keywords as baseline community specific vocabs include OER specific keywords, itis, etc
1. I'm not hearing anything about vocabularies. What is the vocabulary standard target?
  - a. *Comment from Jaci:* Vocabs are often community specific but good rule of thumb is to use GCMD science keywords as baseline. Community specific vocabs include OER specific keywords, itis, etc.
  - b. No standard vocabs really defined/chosen yet. Even the agency name is non-standard (NOAA, US DOC/NOAA, National Oceanic ... etc)
  - c. CF standard names
  - d. *Comment from Don Collins:* Important to keep in mind that GCMD science valids are mostly thematic, not specific or detailed variable identifiers, so GCMD is good for common thematic groupings but less good for entity-attribute like details.
1. How will we avoid "common core" metadata profile from becoming the \*only\* metadata people provide? Need to encourage higher metadata completeness scores--not just discovery-level metadata
  - a. Showing and displaying metrics provide feedback
  - b. Can we explicitly describe this as "spiral #1" in the language of the EDMC Data Documentation procedural directive?
  - c. Ken C: Spirals are nice... but... the reality of dealing with a multitude of datasets coming at you means that spirals can be challenging to achieve
1. *Question for Anna:* What is the connection between EMMA and Data.Gov?
  - a. Ken C: They looked into each others' eyes and it was love at first site. (pardon the pun)
1. *Observations by Dave Fulker on the Target State:* I think the iPython Notebook approach reported by Stace Beaulieu is so exemplary as to be treated as a "target." In particular, it unequivocally

treats (standard) metadata not as an end but as a means to an end (namely, reproducible and scientifically descriptive documents intended for human readers). I highlight two benefits:

- a. 1) The approach establishes a very direct and immediately testable link between metadata and usability rather than building on assumptions that standard metadata yields usability.
  - b. 2) The approach appears to enable better discovery via common search engines, and it might even set the stage for inferential approaches to dataset discovery.
  - c. *Comment from Curt Tilmes:* Yes, "National Climate Assessment" through the Global Change Information System (GCIS)
  - d. *Comment from Stace:* she is correct - others in the same RPI group are working with PROV-O and GCIS National Climate Assessment report
1. Jim Potemra - benefits that can justify metadata?
- a. Comment from James G: One thing we've had considerable experience with are datasets that have COARDS/CF versus those that don't. There is a huge difference in usability when COARDS/CF is part of the dataset. Note that COARDS is very simple and CF is pretty closely focused on `_use_`. So, CF is really different from ISO19115...
  - b. One problem is that the best person to author the metadata is usually the person who needs it the least, because s/he already knows the information. -Jeff DLB
  - c. *Comment from Brian Westra:* In some respect, as metadata and the surrounding systems become better, to some degree it will be less visible (similar to other parts of the infrastructure). Discovery, context, provenance, permissions, all are good reasons.
  - d. *Comment from Jaci:* I've found that the best author of metadata is often computer generated (only if lots of information about the data is available but often we capture MUCH more about the data than human put into their metadata). humans can get lazy....or pulled in too many direction ;)
  - e. Analogous question: what benefits justify adding comments to the software you write?
  - f. NOAA should use metadata as much as possible in its own business processes. E.g., make it the case that producing good metadata means that some calls for information (e.g., in response to questions by leadership or congress) are automated rather than manual. -Jeff DLB
1. *Comment from Chris Calloway:* Easy way to install IPython notebook (which can be difficult to install all the parts separately): <https://www.enthought.com/products/canopy/>
1. What is the rubric score for an ATRAC record?
- a.
1. What is URL of ATRAC?
- a. <https://www.ncdc.noaa.gov/atrac/index.html>

snapshot of atrac editor:

# 2013 NOAA EDM Virtual Workshop Consolidated Session Notes

Modified: 4/12/2013

Form input is managed in two ways: 1) click "Save" to record your input and return to the form at a later time, or 2) click "Submit" to validate your input and submit the form to the data center for review. Form input can be modified as needed after it has been saved or submitted.

**Metadata** Identification Keywords Distribution Lineage Quality

- \* Metadata ID:** The unique string that identifies this metadata record. The Metadata ID is normally assigned by the data center responsible for the dataset.
- Identify a Point of Contact responsible for this metadata record.**  
Last Name:   
First Name:   
**\* Organization:**   
Position:   
Telephone:   
Email:   
Contact Instructions:   
**\* Role:**
- \* Metadata Maintenance and Update Frequency:**  
  
Scheduled date of next metadata update:

NOAA

## Session 5 - Data Preservation and Citation

Thursday, June 27, 1:00 to 2:30 PM (Eastern)

Session Chair: Nancy Ritchey

### Current State, Nancy Ritchey

- Data Preservation is the Hub of Data Management
  - enable catalogs
  - requires metadata & vocabularies
  - all enable usage thru standardized data formats & tools
- Preservation & Citation - Current State
  - preservation an afterthought - not part of the planning; not funded well
  - not clear who is responsible for preservation
  - not all critical data being archived
  - data standards are not established across the agency
  - datasets are not citable

*NR: data stewardship is the responsibility of data providers, esp for large campaigns*

### Ignite Talks

Digital Object Identifiers - Ruth Duerr, NSIDC & co-Chair, ESIP Federation Data Stewardship Committee

- Open Access paper on DOIs from work done in ESIP (Federation of Earth Science Information Partners)
- Data Citation Guidelines for Data Providers &
- DOIs for datasets & collections
- UUIs for data items
- Why not? handles, ARKS, PURLS, XRIS, LSIDS - publishers like DOIs
- Crossref and EZID provide DOIs
  - CrossRef - publications
  - EZID -
  - ESIP testbed tested DOI usage
- Many publication models
  - Is your data like a book? - give it all a DOI; 2nd edition, new DOI
  - Serial publication?
  - Encyclopedia? - each article should have its own DOI/citation
- Contact: [wiki.esipfed.org/index.php/Preservation\\_and\\_Stewardship](http://wiki.esipfed.org/index.php/Preservation_and_Stewardship); [esip-preserve@lists.esipfed.org](mailto:esip-preserve@lists.esipfed.org)

*RD: if you are not ready to publish and archive, then don't assign a DOI, use something else, like ARKs contact [esip/ruth](mailto:esip/ruth) to provide more use cases and help improve the guidance*

NOAA Data Citation Project - Heather McCullough, NESDIS/NGDC

- NOAA Data Citation Policy - Why?
  - Procedural Directive to assign persistent identifiers
  - Credit providers

- Serve users
- Traceability
- Data Citation Index - linked to Thomson Reuters
- Using EZID - creates, stores simple metadata & directs to landing pages
- Use ISO metadata
- DOI directs to Landing Page of the dataset, highlighting the dataset's metadata
- Cite As (recommended) - in the ? section of the metadata
- What's Next - 10 Datasets; Add Cite as statements for all datasets; Education; train metadata authors
- What Level to assign?
  - granular, collection
- Transfer of ownership? - new DOI?
- Continue to test DOI assignment; on-going discussions - feedback welcome!

Pier to Preservation - Steve Rutz, NESDIS/NODC

- Ships - Data - Google
- Ships are like satellites - collect lots of data
- NOAA R2R - created shipboard data stewardship procedures
- Ship comes to port with data - uploads data to NODC -NODC moves to ingest directory for processing - NODC digest checks for metadata completeness - NODC assigns identifier & publishes data
- NOAA R2R - deployed Esri's Geoportal Web Server App discovery layer
- Provides standard FTP; HTTP access & OPeNDAP

Steve Rutz - winner of favorite Ignite Talk for this session

How would you describe your role at NOAA?

Data producer/observing system specialist - 30%

Data services or metadata creator - 60%

Data user/scientist - 30%

Data steward/archivist - 40%

Other - 35%

### Target State

- Preservation & funding is instituted across NOAA and begins during planning
- Standards established: metadata, formats
- Routine archival of fully documented data & information (e.g. software used to create data product, provenance, tools)
- Persistent identifiers, with landing pages and citation guidelines

### Potential Next Steps

- Plan for preservation, including funding
- Improve communication between data producers & archives
- Promote tools to assist with preservation (e.g. archive requests; metadata creation, others?)
- Establish standards: metadata, formats
- Institute data identifiers within the archives

- Work with journals to require identifiers for data used in publications

Discussion of next steps

- Need a model that addresses the DOI ownership question (ESIP & elsewhere should study this)

---

## Q & A for Session 5 - Data Preservation and Citation

Questions asked during Ignite Talks:

1. *Asked during Heather's talk:* What is the status for obtaining DOI's for NOAA Publications?
  - a. *From Jeff:* Status is a bit TBD. If it's a published journal article by NOAA author, journal will assign. If tech document worthy of assigning a DOI, then work with NOAA central library. Discussed during public access research results.
  - b. Comment from Linda Pikula: that is the OMB Directive 2013 for public access to data
1. *Asked during Heather's talk:* How can labs or Science Centers can assign a DOI to their products?
  - a. From Nancy: DOIs are for archived and fully documented data, so archive (NOAA Data Center) assigns when the data are archived (or archiving has started in the case of ongoing data collection).
  - b. From Ruth: EZID recommends no DOI until it is archived. EZID is promoting ARKs b/c they can be used with ARK. DOI is forever. ARK is ephemeral, but still locator.
  - c. At present, only the 3 NOAA Data Centers can issue DOIs or ARKs, so in any case you need to get in contact with one. We can help.
    - i. Re: comment added by Jeff DLB to question 2(?): Is it correct that only the data centers would assign an ARK? I thought the data centers would only be responsible for DOIs.
1. *Question for Heather:* Is your group at NGDC also working with the group who produced the recent "Ocean Data Publication Cookbook" (UNESCO Manual and Guide #64, 2013):  
[http://www.iode.org/index.php?option=com\\_content&view=article&id=110&Itemid=129](http://www.iode.org/index.php?option=com_content&view=article&id=110&Itemid=129)
  - a. *Heather:* Involved and aware. Have also been looking at other guides such as ESIP.
1. *Can Ruth explain what an ARK is?*
  - a. *From Ruth:* Archival Resource Key - another locator, click to find data. ARK can be assigned for every item in a dataset using pre/postfix. Not a part of datacite world, so not included in Thompson-Reuters Index.
1. Comment: Fishery journal already has DOI's. Fisheries produce a lot of documents, there is any way we can be part of the DOI's discussion?
  - a. *From Jeff:* You can be part of discussion. Discussion has been a series of telecons. \*Focused on dataset identifiers
  - b. Response to Q5 from Don Collins - Should the NOAA Library be leading a parallel discussion about NOAA publications DOIs?

1. *Would DOIs be consistent across NOAA or will have some indication of individual data center, i.e., NCDC, NODC, or NGDC?*
  - a. *From Ruth: It is a best practice that DOIs are opaque with no indication that it comes from NOAA. Over very long-term if there is an organizational change, then would need to redo all DOIs.*
  - b. *Heather: Prefix will still be a # for NOAA.*
  - c. *Anna: indication of data center is not transparent in DOI string, but metadata supporting the DOI identity creator/provider and publisher/archive.*
  
1. *For Ruth, if the DOIs are opaque, how would we get metrics from DataCite on the number of records published from the Archive.*
  - a. *They are still connected to their datasets.*
  - b. *Comment from Curt Tilmes - You can update metadata, you can never update the DOI itself -- that's why it's opaque.*
  - c. *Comment from Ken C - We have no embedded semantics in our DOIs. SO we follow the ESIP best practice.*
  
1. *If a publication has a DOI is the data DOI complementary to the publication DOI? For example, the publication presumably reports on scientific analysis of data, and the referenced data has a DOI. So the data can be cited the ways publications are already cited. Is this a correct understanding?*
  - a. *Ruth: Yes - reference list will cite both publications and data used.*
  
1. *Question for Ruth from Phil Jones: You said to only assign DOIs to datasets that will be archived forever. Is it acceptable to assign DOIs to datasets which may be eventually disposed from the archive but will have metadata that are persistent?*
  - a. *Ruth: Less than optimal, but is supported as long as there is a landing page.*
  
1. *Question/comment for Heather and Ruth: Should DOI's be organized based upon the data collection, or how the data will be most commonly used? For example, a collection of real time ocean data will have data from a variety of platforms. In addition to the complete collection, would DOIs be minted for the individual platforms types? Or perhaps it's better to mint DOIs based upon the way people are likely to actually USE the data. For example, mint DOIs to correlate with the various feature types represented, such as profiles, or trajectories (as defined by CF chap 9 DSGs).*
  - a. *Ruth: Think about what your data looks like, because people need to cite the part of the data they used. Decisions for science community.*
  - b. *Heather: Need to interact more with data managers and users. Not a black and white solution.*
  - c. *Jeff DLB: The NOAA Data Citation pilot project does not yet have answers to all these questions, and we are trying to figure them out as we go. Recommendations will be documented in the (pending) EDMC Data Citation Procedural Directive.*
  - d. *Nancy: A lot of archive staff don't have science library backgrounds. Should look more to the library community for organizing and accessing data*
  - e. *Ruth: Please keep in touch with the ESIP Community. They are looking for other use cases*

1. *Question for Ruth:* Is there ESIP guidance for transfer of DOI ownership?
  - a. There is EZID guidance that is very straightforward.
  - b. Comment from Ken C: I think there is a subtlety involved with the transfer of ownership issue which makes the EZID system not capable of supporting our use case. Greg Janee from EZID told us that there is a concept of co-ownership, but that the group minting the DOI is the owner and that is the only way to become an owner. SO, for example, if CDIAC creates a DOI, they can not give ownership of it to us.
  - c. More from Ken C: Here is more on DOI ownership: My question to EZID: A detailed question I have is - once co-owners of a DOI are identified, can either of them remove the other as a co-owner so the DOI ultimately becomes owned only by one organization? I'm concerned about cases where someone sends us data to archive, and wants to give us full and sole control of that DOI. If they or we can remove them as co-owner, then problem solved. But if not, then it opens up possible problems were the now-defunct co-owner accidentally messes things up, like changing the URL to the landing page by accident.
    - i. And here is the EZID response: Short answer: no. Currently, an owner can add/remove co-owners, and co-owners can modify the identifier's target URL and metadata, but a co-owner cannot add/remove the owner or other co-owners. So, an identifier co-owner plays a secondary role to the owner, much as an aircraft co-pilot is secondary to the pilot. Or, if you're familiar with SQL, a co-owner has read/write privileges, but only the owner has the privilege to grant privileges.
    - ii. However, you've presented an interesting use case. I don't believe that EZID's current ownership model is the One and Only way to do things. If a slightly different model would better fit people's use cases, I'd be open to changing it. Of course, we'd have to consider those use cases in which the secondary status of co-owners is preferable. -Greg
  
1. *Question for Nancy:* She mentioned that standards are not established across the organization. What standards is she referring to? Isn't that the objective of the Procedural Directives?
  - a. Using standards in a broad term.
    - i. comment from Anna Milan: goal is to develop **common recommended practices** across NOAA within the standards.
  - b. *Comment from Bob Simons:* Please mention ACDD. It is now an ESIP-based set of global attributes. This metadata is used by nclSO in THREDDS and by ERDDAP to automatically generate ISO 19115/19139 metadata files. It is a great starting point for metadata that all datasets should have. Please join the ACDD ESIP group to help develop ACDD. (Attribute Conventions for Data Discovery)
  - c. Comment from Curt Tilmes: Core includes provenance, not just discovery...
  
1. What role do you see for the librarian in this DOI process? Amended from Joan Segal: What role do you see for the NOAA Library Network of librarians in Addition to question?
  - a. from Anna: experts with experience that can help guide this effort
  - b. From Don Collins: Parallel discussion about how we determine and apply DOI to datasets. NOAA librarians to have another conversation about DOI guidance for reports like ship

reports

- c. *Comment From Anna Fiolek, Metadata Librarian:* Thank you Linda and Don for your comments and proposing that the NOAA Central Library (NCL) become a central place for assigning the DOIs for NOAA publications, including technical reports, cruise reports, etc. The NOAA Central Library, according to NAO 250-17, is a repository library for all NOAA publications in all formats, including printed and online. At the time metadata is created, a DOI number can be assigned/generated and embedded into the metadata.
2. FYI - There are 8 NOAA librarians on this call according to Linda Pikula
    - a. Joan - Librarians great asset to bring in for DOI construction and citation.
  1. How do NOAA librarians who aren't currently involved in the process become involved?
    - a. ACTION = need to get back to Joan Segal on this question.
  1. Can/will the EZID license also be available to the Library for minting DOIs for NOAA Publications?
    - a. Jeff DLB: Let's talk about it. If it is appropriate to assign DOIs to NOAA pubs that aren't getting DOIs from a Journal, then yes, those DOIs should use the same license.
  1. Scientists will not want to have people reference DOI's for their papers because their H score will suffer if people cite DOI's and not publications. Can this be addressed?
    - a. Ruth: Yes. Thompson-Reuters will include datasets. Transition phase, but hopefully, credit for data will be as valuable as papers.
    - b. Comment from Linda Pikula: I recommend that librarian's become involved in discussion with Thomson Reuters in their new Data Citation Software to help set parameters
  1. Should NOAA Researchers be encouraged to obtain an ORCID or ResearcherID?
    - a. Ruth: Yes. Likely to become more important over time. ESIP group is working on identifiers for other things...
    - b. Comment from Curt Tilmes: Both ORCID and ResearcherID! They are linked -- both organizations are working together
    - c. Comment from Deirdre Byrne: ORCID ID's are important to disambiguate objects. Not just people, but people from, say, similarly-named geographic features.
    - d. Comment from Ge Peng: Yes. More and more scientific data journals now require a DOI for publishing papers on datasets.

*Questions asked during Target State Discussion:*

1. *Comment Phil Jones:* For the next steps, I would add that we should establish an enterprise process to implement standards, including creating data in sustainable formats such as netCDF w/CF Conventions.
  - a. *Comment from Phil Jones:* The point of my comment is have a process to implement the standards, not a process to establish standards.
  - b. *From Don Collins:* Re: comment 1 from Phil: How well does netCDF with CF conventions support audio and video data and other types of data (e.g., genomic and other experimental

data)? We need to be considering all types of data that NOAA collects and generates when we discuss recommending or requiring standard formats, not just the types we are currently familiar with managing.

- c. *Reply to Don from Phil:* Agreed. There are multiple file formats suitable for long-term data preservation. NetCDF is one example. Again, the point of my comment is that NOAA needs a way to ensure that these recommended data standards are actually implemented, not just established.
1. DOI's cost money. Does NOAA get a special discount, for free being a government entity?
    - a. Jeff: No discount, paid same fee as everyone else. Two choices - EZID and DOE. Cost was the same. Could switch between two at any point. EZID API was a bit better for bulk assignment. EZID also accepted credit cards, whereas DOE required MOU.
  1. Where to send data?
    - a. From Don Collins: To send data to NODC, send email to [NODC.DataOfficer@noaa.gov](mailto:NODC.DataOfficer@noaa.gov). Like NODC, we would notify NGDC or NCDC if NODC received a Request to Archive data that was outside the scope of NODC holdings.
    - b. From Steve Rutz: I recommend having a single email address with a distribution list that includes the three NNDCs.
      - i. Comment from Don Collins: I agree with Steve's and Anna's recommendations for a decision tree and centralized email to a distribution list.
    - c. From Nancy: Advanced Tracking and Resource tool for Archive Collections (ATRAC) <https://www.ncdc.noaa.gov/atrac/index.html>

## Session 6 - Wrap-up, Action items, Workshop evaluation

Tuesday, June 25, 1:00 to 2:30 PM (Eastern)

Keynote Speaker: Christopher Lenhardt, DAARWG Chair

Workshop Chair: Jeff de La Beaujardière, EDMC Chair

### Keynote: Big Data is Coming for You! - Chris Lenhardt, DAARWG Chair

#### About RENCI

- Research triangle universities, industries & State of NC
- Data-Cyberinfrastructure/High performance computing shop
- Address complicated multidisciplinary problems & research
- Focus on collaboration - CI, data experts, domain experts
- Environmental, Genomics, Coastal modeling, Climate change, Health
- iRods hosted at Renci
- Big Data -National Consortium for Data Science
- one project example - Storm Surge Forecasting (uses NOAA NCEP & NHC data)

#### Introduction: On the Nature of BIG

- Change in notion of 'big' over time - where does this lead?
- Looking to future of science - how it is being conducted today & how data managers need to support it

#### Big Challenges

- World population growth
  - resources, standard of living, pollution
- CO<sub>2</sub> 400 ppm reached in May 2013
  - Oil availability will remain high for a long time
- Soil Crisis
  - available land for agricultural production of food
  - crop diversity diminishing
- Climate Change
  - extremes - droughts, sea level rise, etc.

#### Big Science Questions

- Water - too much , too little
- Soil
- 

#### Big Data

- Compute-Centric to Data-Centric
- Both large & Small data sets in the mix - the Long Tail
- What do we mean by data?
  - how it's represented & transformed to different contexts
  - ubiquity of sensors
    - lead to swamping us with data

#### Other Big Trends

- Crowd Sourcing - private satellites; environmental health (AirNOW - many sources of data); phenology (ebird - bird migration)
- Open Source
- Open Data
- Open Science
- Data Policy
  - budgets for data center - ORNL 1997 budget was almost same as it was almost 20 years later
  - Executive Order on Open Data
- Alternatives to RDMS
  - HADOOP
  - Middleware (datagrids, iRODS); in storage arrays
- Software Defined Networks (on demand)

#### Implications for Data Managers

- Big Challenges & Science Questions
  - inter-, transdisciplinary approaches
  - non-traditional users
  - linking - data across scales; models & obs
  - interagency coordination/collaboration
- Big Data Management
  - Metadata challenges - is it feasible to develop extended metadata models for all data?
  - RDMS on their way out?
  - Role of data centers needs to evolve

#### Role of Environmental Data Managers

- Central to the solution addressing the challenges
- Work with program managers to get out ahead - don't be reactive
- Rethink approach to problem solving

#### Questions for Chris Lenhardt:

What do you mean by leaving RDBMS technologies behind, specifics?

Hadoop is focused on transactional data. Are the analogous science data assets that could be handled as Hadoop handles transactional data?

Some data sources that are so large that they cannot be effectively archived. What approaches can be used for archiving them?

Is all the data worth keeping? Focus on anomalies or on the changing background?

How do you reconcile discipline specific DM practices in domain science with the transdisciplinary questions that domain science is trying to solve?

One example, bringing together domain scientists with developers to work together (agile development) to develop software that meets the needs. Operationalize hypotheses v edge cases for the sw engineers and hydrologists

Can you expand on your definition of a data center and how should they change?

Need continued dialogue. We will have data centers but the traditional 'data center' (e.g. server rooms) might be looking at federating data grids that bring together the data, software, tools and models. The data centers will be more service oriented.

Do you know of any social science research that speaks to the incentives that might get people into compliance?

One example going on is EarthCube, which has engaged social scientists to find out what the barriers are. Part of it might be training (graduate students); data publication.

*"barriers are ... (?).. red herrings"*

---

### Workshop evaluation, lessons learned, future sessions

Overall, how well did the virtual format work? (80% of 98 participants voted)

Poor - 0%

Fair -10%

Good - 42%

Very Good - 35%

Excellent - 14%

- Comments from Lewis:
  - If you combine this into three categories you have almost 50% saying that the conference was either Very Good or Excellent.
  - Only 10% said that the conference was "less than Good."
  - Amazing that no one said Poor!

How often should we hold virtual EDM workshops? (75% of 90 participants voted)

Never, we should only do in person workshops in the future - 6%

Annually - 32%

Twice a year - 51%

Quarterly - 10%

- Comments from Lewis:
  - Clear majority here for bi-annual virtual workshop. And over 80% favored 1 or 2 virtual workshops per year.
  - We should consider a follow-up evaluation survey with more options to get more details on this and other questions.

This is okay if NOAA is permitted to participate in AGU/ESSI & ESIP activities.

Ken Casey asked: What if we want to hold annual meetings in person, and virtual workshops twice a year???

Don Collins: Quarterly/Semi-Annual virtual meeting have it targeted on a single topical area. There was a lot

of discussion DOIs. In 3-6 months, where are we now with DOI? what are the problems? Are the librarians involved?

Mark Brady: 5-minute talk format is challenging - more coaching might be needed to not overload slides & speak too quickly. Audio quality needs to be good.

Dave Fulker: Multiple editors in a document can be confusing to those viewing.

Jeff DLB: TitanPad was used last year, probably discouraged by CIO now.

Nancy Ritchey: Information exchange was impressive through the many channels used (chat, questions, google doc, audio). Shout out to the team! <clap, clap> :D

Jeff DLB: Had 2 (overlapping) teams working on this for both logistics and for program planning.

Deirdre Byrne: I found the 15-minute talks much more enlightening/informative than the ingite talks. I would vote for that format more often if we can accommodate it.

Anna Milan: Is there a mechanism to disseminate information from the EDM to a broader audience? (dashboard, metrics?) - Answer: not really, somewhat scattershot

- Kim Jenkins: non-tech may shy away from DMIT; add half hour to quarterly EDMC call and report on directives

---

## Summary and further discussion of action items and next steps

### Next Steps - General:

- Find ways to make impact/relevance/roles regarding federal and NOAA policies & directives
- Find ways to promulgate info about EDMC & DMIT activities more broadly

### Next Steps? - Discovery:

- DMIT members: List of existing catalogs at defined place (under noaa.gov/data?)
- CIO/EA/EDMC/DMIT: Determine how to structure /data catalog
- CIO/EA/EDMC: Determine how to populate /data catalog
- Explore what Google can really do with ISO metadata

### Next Steps? - Access:

- Jeff DLB, Tony Lavoie, DMIT members: Help NWS IDP decide exactly what software stacks to run
- Select one or more for pre-approved Authority to Operate
- NGDC EMMA team: Collect info/stats on available data access links and methods
- From Kevin O'brien e-mail:

- Continue outreach to NOAA data centers to bring data into UAF
- Produce best practices for data providers who want to serve data via OPeNDAP/THREDDS
- Improve ability to actually access and utilize datasets directly from discovery software
  - For example, when finding results in Geoportal, ability to provide direct links to the data, or to visualization tools that can immediately start using the particular dataset

**Next Steps? - Usability:**

- NGDC/Anna Milan: Improve and augment metadata rubric score system with more content
- Consider whether WHOI iPython Notebook could be offered/used at NOAA. Schedule longer talk by Stace Beaulieu at DMIT
- Anna Milan et al.: mapping b/t Prov-O and ISO project with Stace
  - Stace: thanks, I will be at ESIP and can also participate in one of your upcoming DMIT workshops
- NCDC: Improve ATRAC - import existing ISO XML; Determine whether/how to make it a central resource for NOAA.
- *Agree on small # of target formats*
- *Establish more Web Map Services:*
  - Comment from Bob S: - Isn't it better covered by serving more data via THREDDS and/or ERDDAP, which both make as many datasets as possible available via WMS? And by using THREDDS and/or ERDDAP, the actual data (not just an image) is made available.
- *Agree on controlled vocabularies*

**Next Steps? - Data Preservation & Citation:**

- Dataset Identifier team: Institute data identifiers within the archives, add NOAA librarians to the team
- EDMC: Finish & issue Data Citation PD
- Work with journals to require identifiers for data used in publications, according to Ruth Duerr this is already in the works.
- Establish clear entry point for someone who wants to start archiving.
- Continue NOAA Metadata telecons regarding ATRAC, Send2NODC, EMMA, and other tools and approaches.
- Plan for preservation, including funding
- Improve communication between data producers & archives, goes back to single point entry into archival process
- Promote tools to assist with preservation (e.g. archive requests; metadata creation, others?)
  - Comments from Ken C:
    - I wanted to point out that later this summer, we will be releasing our Send2NODC system for NOAA-wide review and comment.
    - Send2NODC will allow our data producers to document and upload data to NODC for archiving.
    - It is an actual data and metadata submission system.
    - Yes, those conversations are going on actively as part of NEIO planning.
- Establish recommended profiles & practices: metadata, formats and attributes within the formats

---

## Q & A for Session 6 - Wrap-up, Action items, Workshop evaluation

1. *From Dave Fulker:* More specific to alternatives to RDBS that are promising
  - a. simple data structures and repeat info, but b/c higher compute power it is more efficient
  
1. *From Mark Brady:* Standard methodologies for dealing with large volumes of acoustic or other high volume collection methods?
  - a. Compression mechanisms or map-reduce framework
  - b. *Follow up question from Deirdre B for Mark:* How big are the marine acoustic data?
  
2. *From Jonathan Blythe:* we have several disciplines that we work with. Interested in disc. specific data management and how do you reconcile this with working with scientists who are generally “hunkered down” in their areas of expertise?
  - a. NSF-funded project for Science Software Institute premise that just like data management, we’d like for scientist to be able to manage their code more effectively. The challenges if you combine software and domain scientist - they talk past each other. Group has developed agile open source method that utilizes boundary objects to collaborate.
  
1. *From Anna Milan:* Can you expand on your definition of data centers?
  - a. Think that the idea that all your data sits with you at a “data center” is out of date. Data may need to come from lots of different sources.
  
1. *Question from Deirdre Byrne:* How do you incentivize or socialize data management so that it is something that the domain scientist actually wants to do, as opposed to grudging or non-compliance? We struggle with that even within NOAA and certainly between NOAA and external scientists.
  - a. In EarthCube, actively engaged social scientists to understand needs, barriers; Training in grad schools; NOAA and NSF data management plan requirements

### Questions on Wrap-up and Evaluation Session:

1. Comment from Deirdre Byrne: I think the VM only worked because many of us know each other already. Also if meetings are shorter because targeted, it would be easier to attend the whole thing (As in - 4-6 hours on one topic.)
2. Comment from Phil Jones: Half day virtual meeting was a nice format--it made it much more reasonable to attend the entire event.
3. Comment from Curt Tilmes: I prefer 100% in person or 100% virtual compared to 90% in person with 10% on the phone. It is easy to feel left out when you are on the phone and everyone else is interacting in person.
4. Comment from Deirdre Byrne: I think multiple editors to a Google doc works okay as long as everyone has direct view access (not via webex) to the Google doc. We have done this

successfully with a dozen or so contributors in real-time ....

5. Question from Jacqueline Rauenzahn: In future virtual workshops, would you please consider providing audio access/connection via the laptop, as well as via the phone?
6. Comment from Deirdre Byrne: I found the 15-minute talks much more enlightening/informative than the ignite talks. I would vote for that format more often if we can accommodate it.

*Question & Comments raised during next steps discussion:*

1. *Comment from Bob S:* I'm willing to help with the project to "explore what Google can really do with ISO metadata".
2. *Comment from Stace:* I will be at ESIP and can also participate in one of your upcoming DMIT workshops
3. *Comments from Jaci on creation of new tools for Data Usability:* there are other tools...including online guis for creation that support the entire standard or at least most of it...why invest in more tools
4. Comment from Bob S on the next step "Establish more Web Map Services"?
  - a. Isn't it better covered by serving more data via THREDDS and/or ERDDAP, which both make as many datasets as possible available via WMS? And by using THREDDS and/or ERDDAP, the actual data (not just an image) is made available.

## Next Steps (consolidated from all sessions)

Raw text from Session Notes...

---

### Session 2: Catalog and Search

#### Potential Next Steps

- Harmonizing metadata - using Google?
  - Establish a Master NOAA Catalog?
- 

### Session 3: Data Access

#### Next Steps/Actions

- from the UAF (under DMIT) Working Group
  - Improve UAF holdings, org and reliability thru UAF Master THREDDS catalog
  - Outreach to NOAA data centers to bring data into UAF
  - Produce best practices for data providers who want to serve data via OPeNDAP
  - Create tool to 'clean' remote catalogs
  - Provide rubric to clearly display quality of data catalogs (currently limited THREDDS data server catalogs)
  - Utilize ERDDAP, especially for serving collections of in situ data

#### Possible Next Steps

- Work with Luis Cano to establish standard access methods that will be incorporated in IDP, so that EDM doesn't devolve to a myriad of silos.
  - Might look at NASA's [Standards Process Group](#) (SPG), for that software stack. Or ideas for how to select the software.
  - Improving the link to access data from the catalog/search results
    - rubric for metadata check data access link, with stats/analytics (possibly in the dashboard?)
- 

### Session 4: Data Usability - Metadata, Standards, Visualizations

#### Next Steps

- Get your metadata into EMMA
  - Establish & id more web map services
  - Learn how to document with the standards
  - Support OSTP initiative for noaa.gov/data catalog by Nov.
  - Agree on common vocabularies
  - Automate individual steps in NOAA's business processes - verifiable using software
-

### Session 5: Potential Next Steps

- Plan for preservation, including funding
- Improve communication between data producers & archives
- Promote tools to assist with preservation (e.g. archive requests; metadata creation, others?)
- Establish standards: metadata, formats
- Institute data identifiers within the archives
- Work with journals to require identifiers for data used in publications

#### Session 5 Discussion of next steps

- Need a model that addresses the DOI ownership question (ESIP & elsewhere should study this)
1. *Comment Phil Jones:* For the next steps, I would add that we should establish an enterprise process to implement standards, including creating data in sustainable formats such as netCDF w/CF Conventions.
    - a. *Comment from Phil Jones:* The point of my comment is have a process to implement the standards, not a process to establish standards.
    - b. *From Don Collins:* Re: comment 1 from Phil: How well does netCDF with CF conventions support audio and video data and other types of data (e.g., genomic and other experimental data)? We need to be considering all types of data that NOAA collects and generates when we discuss recommending or requiring standard formats, not just the types we are currently familiar with managing.
    - c. *Reply to Don from Phil:* Agreed. There are multiple file formats suitable for long-term data preservation. NetCDF is one example. Again, the point of my comment is that NOAA needs a way to ensure that these recommended data standards are actually implemented, not just established.

---

### Original attempt to consolidate Next Steps from each session (repeated from Session 6 notes)...

#### Next Steps - General:

- Find ways to make impact/relevance/roles regarding federal and NOAA policies & directives
- Find ways to promulgate info about EDMC & DMIT activities more broadly

#### Next Steps? Discovery:

- DMIT members: List of existing catalogs at defined place (under noaa.gov/data?)
- CIO/EA/EDMC/DMIT: Determine how to structure /data catalog
- CIO/EA/EDMC: Determine how to populate /data catalog
- Explore what Google can really do with ISO metadata

#### Next Steps? Access:

- Jeff DLB, Tony Lavoie, DMIT members: Help NWS IDP decide exactly what software stacks to run
- Select one or more for pre-approved Authority to Operate
- NGDC EMMA team: Collect info/stats on available data access links and methods

- From Kevin O'brien e-mail:
  - Continue outreach to NOAA data centers to bring data into UAF
  - Produce best practices for data providers who want to serve data via OPeNDAP/THREDDS
  - Improve ability to actually access and utilize datasets directly from discovery software
    - For example, when finding results in Geportal, ability to provide direct links to the data, or to visualization tools that can immediately start using the particular dataset

#### Next Steps? Usability:

- NGDC/Anna Milan: Improve and augment metadata rubric score system with more content
- Consider whether WHOI iPython Notebook could be offered/used at NOAA. Schedule longer talk by Stace Beaulieu at DMIT
- Anna Milan et al.: mapping b/t Prov-O and ISO project with Stace
  - Stace: thanks, I will be at ESIP and can also participate in one of your upcoming DMIT workshops
- NCDC: Improve ATRAC - import existing ISO XML; Determine whether/how to make it a central resource for NOAA.
- *Agree on small # of target formats*
- *Establish more Web Map Services:*
  - Comment from Bob S: - Isn't it better covered by serving more data via THREDDS and/or ERDDAP, which both make as many datasets as possible available via WMS? And by using THREDDS and/or ERDDAP, the actual data (not just an image) is made available.
- *Agree on controlled vocabularies*

#### Next Steps? Data Preservation & Citation:

- Dataset Identifier team: Institute data identifiers within the archives, add NOAA librarians to the team
- EDMC: Finish & issue Data Citation PD
- Work with journals to require identifiers for data used in publications, according to Ruth Duerr this is already in the works.
- Establish clear entry point for someone who wants to start archiving.
- Continue NOAA Metadata telecons regarding ATRAC, Send2NODC, EMMA, and other tools and approaches.
- Plan for preservation, including funding
- Improve communication between data producers & archives, goes back to single point entry into archival process
- Promote tools to assist with preservation (e.g. archive requests; metadata creation, others?)
  - Comments from Ken C:
    - I wanted to point out that later this summer, we will be releasing our Send2NODC system for NOAA-wide review and comment.
    - Send2NODC will allow our data producers to document and upload data to NODC for archiving.
    - It is an actual data and metadata submission system.
    - Yes, those conversations are going on actively as part of NEIO planning.

- Establish recommended profiles & practices: metadata, formats and attributes within the formats