

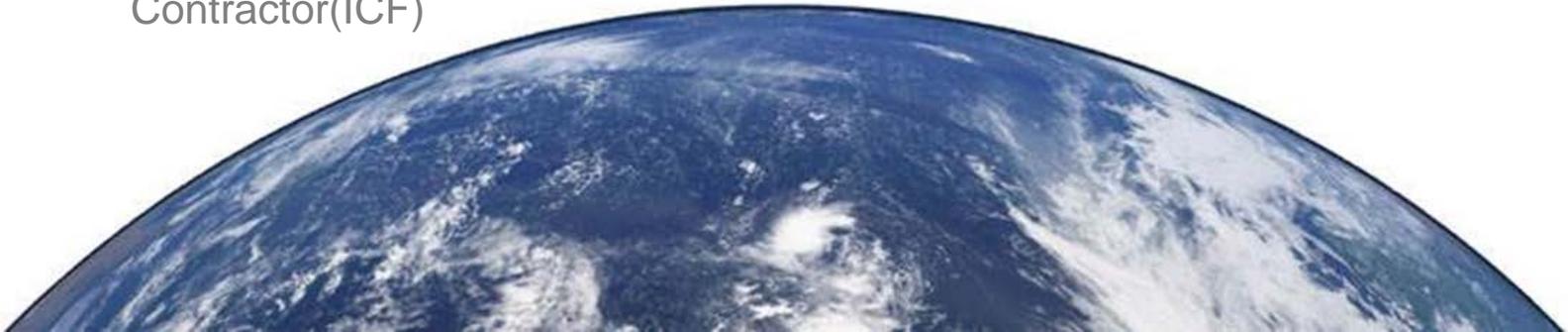
From Data Sharing to Data Management: Data Challenges of Supporting Integrated Environmental Research



Prepared by:
Jennifer Saleem Arrigo^{1,2}

¹ Formerly, Program Manager NOAA CPO/OOMD, Climate Monitoring Program

² Now, Advance Science Lead, US Global Change Research Program, Contractor(ICF)



NOAA EDMC Policies

Environmental Data Management Committee - <https://nosc.noaa.gov/EDMC/>



OSTP
PARR
Policies
Completed

NOAA Administrative Order 212-15
(2010)

**NOAA Environmental Data Management
Framework** *(2012-2013)*

**Data Management
Planning
Directive**
(2011; rev. 2014)

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

Data Access Directive
(2015)

**Archive Appraisal
Procedure**
(2008)

Data Citation Directive
(2015)

**Data Sharing Directive
for NOAA Grantees**
(2012; rev. 2015)

One-line Summaries of NOAA EDM Policies

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

- **NAO 212-15:** data must be visible, accessible, independently understandable, and managed throughout their life-cycle
- **EDM Framework:** States goals, rationale, governance, life-cycle definition
- **DM Planning:** All observing systems must have DM Plan
- **Data Documentation:** Metadata must be created and published
- **Data Access:** All data must be registered in NOAA Data Catalog and accessible on-line
- **Data Sharing by Grantees:** funded data & journal articles must be publicly accessible
- **Archive Appraisal:** NCEI* procedure for accepting data
- **Data Citation:** Requirements for getting NOAA dataset ID from NCEI*

**NCEI = National Centers for Environmental Information (former NCDC, NGDC, NODC, NCDDC)*

FY2017 EDMC Work Plan

- 2017 NOAA EDM Workshop
- *No new Directives in 2017!*
- Focus on assessment & implementation of existing directives
 - Finish EDM Dashboard improvements
 - NOAA EDM Program assessment
 - DM Plan Repository & Stats
 - Establishment of online forms-based DM Plan tool
 - Methods for tracking grantee data sharing
 - If needed, minor updates to existing directives
- Status reports to NOSC & CIO Council



Example: Climate Program Office

NOAA's Climate Program Office (established in October 2005) provides a unique and highly flexible climate research enterprise that focuses on:

- competitive grant programs that advance and extend our research capabilities;
- partnerships with academia, businesses and other agencies to develop and deliver targeted research and data products; and
- knowledge and information to improve public climate literacy and decision-making needed to maintain resilient economies and environmental services

- **>300 active grants**
- **>700 published papers per year**, contributing to our understanding of climate variability and change
- Worked with more than **70 partners** to implement a sustained **Global Ocean Observing System**
 - >1900 floats**
 - >1400 drifters**
 - >60 moorings**





From “Data Sharing” to “Data Management”

Initial Evaluation

“Data production efforts will flow through {Investigator’s Lab} to ensure data are properly organized and documented prior to posting on the lab website “

“All final data generated in this project will be made available to the public... accessible through the {Institutional Library}... has the capacity to store the above materials.”

Project Reporting

DATA DOIs!

Norris, J. R., and A. T. Evan. 2015. Cloud Properties from ISCCP and PATMOS-x Corrected for Spurious Variability Related to Changes in Satellite Orbits, Instrument Calibrations, and Other Factors. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. <http://dx.doi.org/10.5065/D62J68XR>.

DATA WEBSITES <https://climatedataguide.ucar.edu/climate-data/walsh-and-chapman-northern-hemisphere-sea-ice>

DATA PROMISES, PROMISES....

“I contributed significantly to the data compilation for [Third Party Effort]”

“We are working on website modifications that will allow the release of the {X} dataset to the public.”



Climate Monitoring Data Sharing Experiences

“Data Set Development”

- *Physical Climate Data and products*
- *Partnerships with NOAA (NCEI, Labs)*
- *Data sets developed or improved with specific user groups in mind*

- *Data Sharing Plans “mirrored” NSF requirements*

- *Public Webpages most common*
 - *Easily get out of date*
 - *End of project = end of accountability*

Field Campaigns

Social Science Data

Physical/Chemical/Biological

Model Intercomparisons

Sustained Observations/ International Networks



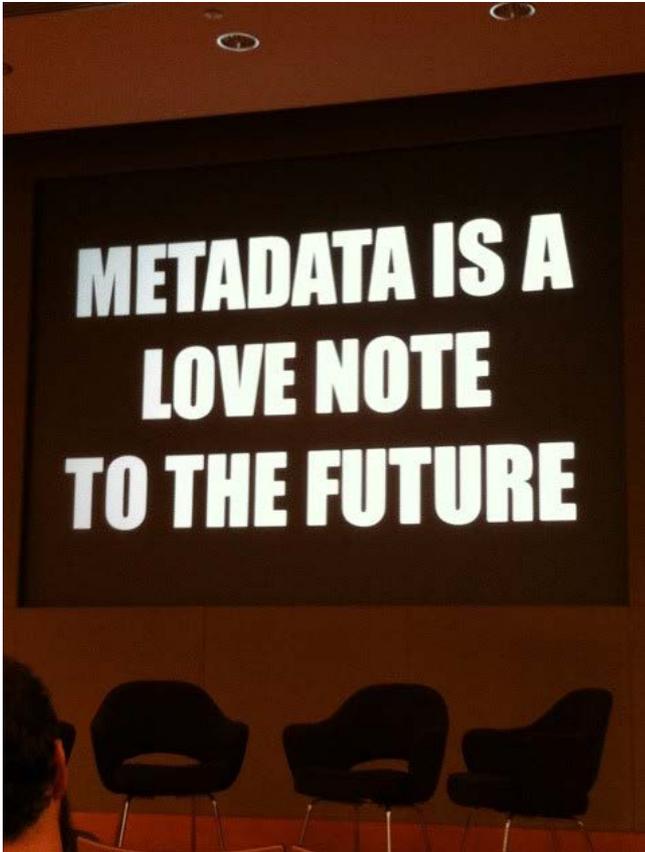
What does “DATA MANAGEMENT” Mean?

- “Environmental data and information collected and/or created under this grant/cooperative agreement will be made **visible**, **accessible** and **independently understandable** to users in a *timely manner (typically no later than two (2) years after the data are collected or created)* free of charge or at minimal cost that is no more than the cost of distribution to the user, except where limited by law, regulation, policy or by security requirements.”
- ***NOAA-funded research results resulting from the work will be “**documented**, **discoverable**, **accessible** and preserved”***



Wait... so what does “Data Management” Mean?

- **Documented METADATA**



[“Metadata is a love note to the future.”](#)
[#nypl_labs](#) on [Twitpic](#) (via [@kissane](#))

Metadata

What are Metadata?

Metadata are documentation for your data set, expressed using a formal syntax. They are used to record all information needed for your data’s future use, and attached to the data set itself, usually as separate files. Here is a list of what metadata might contain:

- a brief or detailed description of the data itself;
- names, labels and descriptions for variables, records and their values;
- explanation of codes and classification schemes used;
- codes of, and reasons for, missing values;
- derived data created after collection, with code, algorithm or command file used to create them;
- weighting and grossing variables created and how they should be used;
- data listing with descriptions for cases, individuals or items studied, for example for logging qualitative interviews;
- descriptions of applications (commercial or open-source) were used to run analyses, and the versions of those applications;
- descriptions of file formats used to store the data;
- documentation of experimental protocols;
- documentation of the code written for statistical and other analyses.

What Kind of Metadata Should I Use?

That depends on your field. Here are some standards and controlled vocabularies (standard terminology for specific fields). If you don’t see your field represented, let’s talk — we may be able to find a standard for you, or help in some other way.

- [Astronomy Visualization Metadata Standard](#)
- [Content Standard for Digital Geospatial Metadata](#)
- [Darwin Core](#)
- [Dublin Core](#)
- [Ecological Metadata Language](#)
- [Data Documentation Initiative \(DDI\)](#)
- [Swingle Plant Anatomy Collection data dictionary](#)
- [Altova Schema library](#)
- [Seeing Standards: A Visualization of the Metadata Universe](#) (humanities: information on over a hundred cultural heritage metadata standards)
- [Text Encoding Initiative](#)

What Else Should I Document?

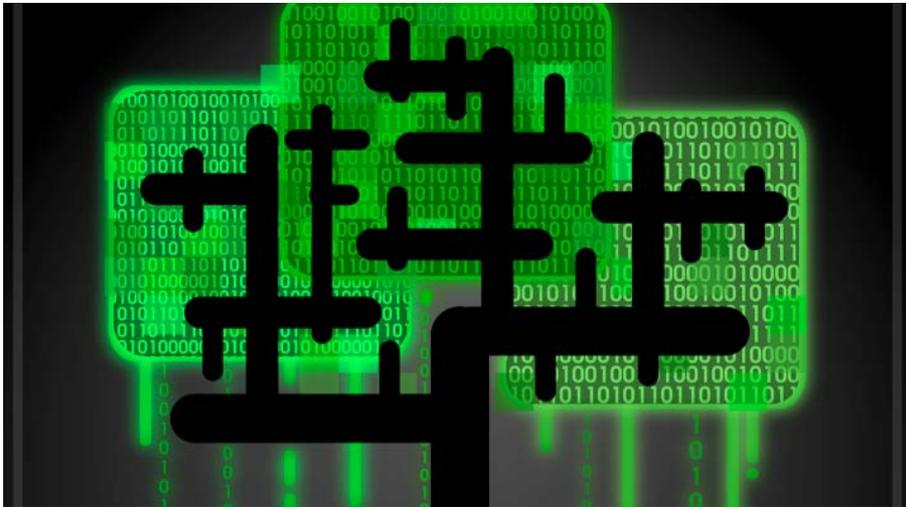
This really depends on your project, but here are some ideas to get you started. We will be happy to help you come up with a final list of documentation needed for your dataset.



Love your data? Love your Metadata

Metadata is the key to:

- Other users understanding your data
- Fitness for purpose
- *Attribution* once your data are “out in the wild”



“Publicly-available” doesn’t mean for a short period of time only. It doesn’t mean for a select group of people. Or with barriers to entry like being charged to access it. It means, quite literally, *publicly* available. To anyone, anywhere, anytime.

- (1) do you know what is being done for metadata now? Does your network have an agreed upon standard? Is it adequate? Is it evolving
- (2) What are your ideas on how we (within OOMD) would like to see data/networks we fund to appear/ be attributed as in “third party” portals and catalogs?



Wait... so what does “Data Management” Mean?

- **Discoverable ...**
- **Accessible ...**



Question: Who will determine if my data are visible, accessible and independently understandable?

Answer: The person generating the data will have first responsibility for determining this. Common data quality standards in your scientific discipline may help you decide if the data are understandable. **Ultimately, others who use your data will know whether they are visible, accessible and understandable to them.** If there are concerns with data access or understandability, they can be reported to NOAA, who will do an independent check.



Wait... so what does "Data Management" Mean?

- *Preserved?*

National Centers for Envir x

https://www.ncei.noaa.gov

NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Home About Contact

National Centers for Environmental Information (NCEI)

NOAA's National Centers for Environmental Information (NCEI) are responsible for hosting and providing public access to one of the most significant archives for environmental data on Earth with over 20 petabytes of comprehensive atmospheric, coastal, oceanic, and geophysical data. [Read more about NCEI](#) >

Center for Weather and Climate

NCEI's Center for Weather and Climate acquires, preserves, monitors, and assesses the Nation's treasure of climate and historical weather data and information.

Weather and Climate

Center for Coasts, Oceans, and Geophysics

NCEI's Center for Coasts, Oceans, and Geophysics acquires, preserves, monitors, and assesses the Nation's treasure of coastal, oceanographic, and geophysical data and information.

Coasts Oceans Geophysics



So what should I be doing?

- ***Take an active role in data management***
- ***NOAA PIs – EDMC directives, many labs have procedures in place to meet***
 - *Consider EDMW next year*
- ***Consider the reasons for data management***
 - *Who are your stakeholders/ users? Who could they be?*
 - *Do they find your data accessible and independently understandable?*
 - *How could you help?*
- ***Think about your data “in the wild”***
 - *Where are your data going?*
 - *Do you take a look for them?*
 - *E.g. JCOMMOPS, OSMC*
 - *How do you want your data represented?*