NOAA/NCEI's Challenges in Meeting New Open Data

Nancy A. Ritchey, NOAA's National Centers for Environmental Information
Nancy.Ritchey@noaa.gov

May 5, 2020

NOAA Satellite and Information Service | National Centers for Environmental Information

Introduction

- NOAA's National Centers for Environmental Information (NCEI) stewards one of the world's largest and most diverse collections of environmental data in digital and physical data formats and media.

- NCEI strives to develop processes, guidance and tools to facilitate the creation and curation of independently understandable and Open data, and now also FAIR data.

> *Tackling FAIR and Open data requires forethought, consistently utilized and evolving best practices, user community engagement, and elbow grease!*

Challenges to Open and FAIR

- Requirements for Open and FAIR data are evolving:
  - throughout the data lifecycle,
  - as the user community changes,
  - with transition to new technology such as cloud, and
  - in accordance with data licensing

- Increased interdisciplinary research and the use of artificial intelligence and deep learning leads to Data being utilized far beyond the original purpose and requires additional effort to be compliant with new initiatives.

- Open and FAIR are on a continuum. Consistently measuring and documenting compliance is not standardized.

FAIR

- FAIR Principles provide guidelines to improve the findability, accessibility, interoperability, and reuse of digital assets.

- The ultimate goal is to optimize the reuse of data which requires well-described data, metadata, software, infrastructure to support data replication and/or combinations in different settings.
  - R1. Meta(data) are richly described with a plurality of accurate and relevant attributes
    - R1.1. (Meta)data are released with a clear and accessible data usage license
    - R1.2. (Meta)data are associated with detailed provenance
    - R1.3. (Meta)data meet domain-relevant community standards

NCEI Enables FAIR and Open

- ## Data are easily discoverable.
  - Data are searchable in Google
  - Metadata are discoverable in multiple venues
  - Community collaboration supports development
- ## Data are independently understandable and citable.
  - Documentation is standardized, complete, and easily understood by many disciplines
  - Data and metadata are machine-readable
  - Data quality is consistently assessed and documented
  - Data have persistent identifiers
- ## Data are easily integrated by non-experts.
  - Data are in usable formats
  - Clear documentation is provided
  - Tools for analysis or visualization enhance usage
- ## Services are provided to promote usage.
  - Customer support is offered
  - Activities/events engage current and new users

Measuring Open and FAIR

Completed an initial crosswalk of FAIR Principles with language and concepts into the Data Stewardship Maturity Matrix (DSMM) and adapt where needed.

DSMM Defines Measureable, Five-Level Progressive Practices in **Nine** *Quasi-Independent Key Components including Accessibility and Usability*

A Unified Framework for **Measuring Stewardship Practices** Applied to Individual Digital Earth Sciences Data Products

http://tinyurl.com/DSMMintro

Maturity Scale / Key Component	Level 1 - Ad Hoc Not Managed	Level 2 - Minimal Managed Limited	Level 3 - Intermediate Managed Defined, Partially Implemented	Level 4 - Advanced Managed Well-Defined, Fully Implemented	Level 5 - Optimal Level 4 + Measured, Controlled, Audit
Preservability	The state of dataset being preservable				
Accessibility	The state of dataset being publicly searchable and accessible				
Usability	The state of data product being easy to understand and use				
Production Sustainability	The state of data production being sustainable and extendable				
Data Quality Assurance	The state of data product quality being assured/screened				
Data Quality Control /Monitoring	The state of data product quality being controlled and monitored				
Data Quality Assessment	The state of data product quality being assessed				
Transparency /Traceability	The state of data product being transparent, trackable, and traceable				
Data Integrity	The state of data integrity being verifiable				

Documenting Open and FAIR

- Metadata:
    - Adopted user community standards (ISO 19115), include attributes that reflect FAIR principles.
    - Provide open access to metadata
    - DSMM results included in metadata record
- Data management/curation/stewardship practices:
    - Providing open access to Data Management practices, processes, and policies
    - Curate our practices to ensure they continue to meet user needs
- Legalities:
    - In the process of adopting open data licenses and we will include in metadata
    - Working with Partners to ensure their licenses provide open use
    - Providing Partner use agreements to ensure the data can be curated and shared openly
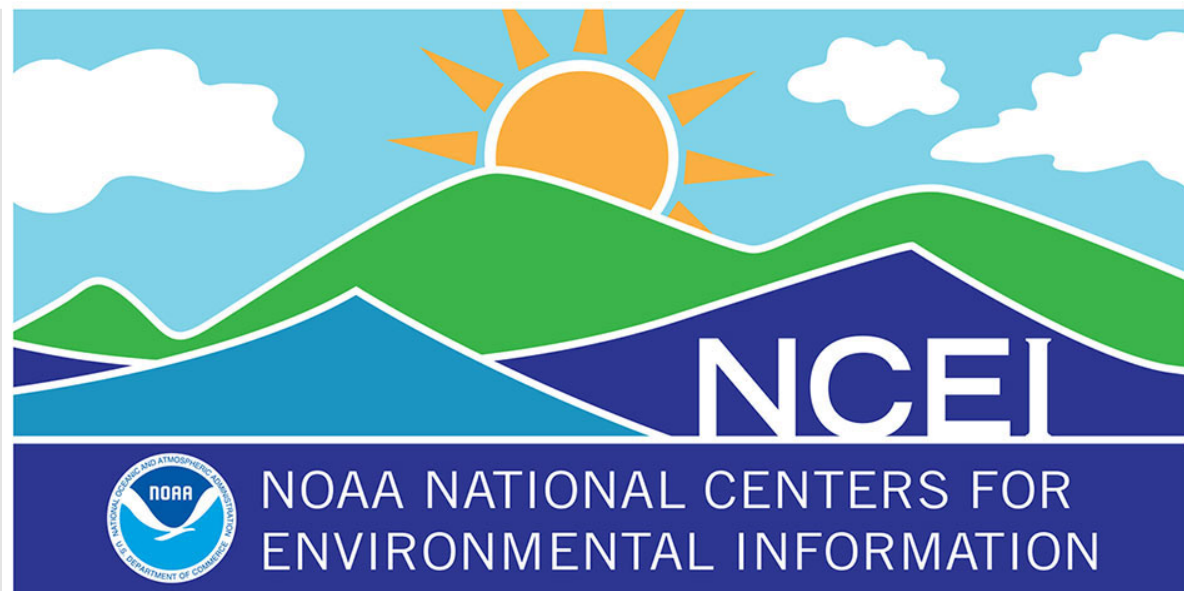
Final Thoughts

Open and FAIR data are on a continuum

Measuring and documenting compliance is a point in time

Tackling FAIR and Open data requires forethought, consistently utilized and evolving best practices, user community engagement, and elbow grease!

www.ncei.noaa.gov
www.climate.gov

NCEI Climate Facebook: http://www.facebook.com/NOAANCEIclimate
NCEI Ocean & Geophysics Facebook: http://www.facebook.com/NOAANCEIoceangeo
NCEI Climate Twitter (@NOAANCEIclimate): http://www.twitter.com/NOAANCEIclimate
NCEI Ocean & Geophysics Twitter (@NOAANCEIocngeo): http://www.twitter.com/NOAANCEIocngeo