NOAA *OneStop* Data and Metadata Improvement Tiers:

A Comprehensive Guide for Data Managers

Published by the *OneStop* Metadata Team
2021-05-27; v2.1

Revision History

Version	Description	Revised Sections	Date
2.0	Initial Publication	ALL	2018-09-06
2.1	Link maintenance	Readiness Tiers, References, Appendix	2021-05-27

Purpose

The purpose of this guidance is to define a system of *OneStop* readiness criteria that data managers and providers can use to increase the discoverability, accessibility, and understanding of the data and metadata in the NOAA *OneStop* portal. These improvements will also provide a richer user experience that leverages all of *OneStop's* advanced capabilities.

Scope

This guidance applies initially to all of NCEI's data holdings and any data manager or provider can utilize it to improve their data and metadata. In the future, the scope will expand to include all of NOAA's data.

Definitions and Acronyms

- **Collection** - A collection is a grouping of environmental data or products that share common characteristics, is represented by a single metadata record, and consists of one or more granules. Collections are also referred to as datasets.
- **Collection Metadata Enterprise Tool (CoMET)** - A collection metadata tool that aspires to become the de facto metadata creation/editing tool for NCEI, other NOAA line offices, and external partners.
- **Data Group** - A superset of collections.
- **Data Stewardship Maturity Matrix (DSMM)** - A unified framework for measuring stewardship maturity of environmental datasets.
- **Granule** - A granule is the smallest aggregation of data that can be independently managed (described, inventoried, and retrieved) in the *OneStop* system. Granules are often referred to as inventory or file level metadata though granules can actually be pieces of files. Granules cannot exist without being associated with a collection. Granules may have a different metadata model than the parent collection in order to support values for additional attributes as needed for those granules.
- **ISO-Lite** - A streamlined version of standard ISO granule metadata developed specifically for *OneStop* that allows quicker processing/creation of granule metadata.
- **Relevancy Ranking** - The process of sorting the search results so that the files which are most likely to be relevant to your query are shown at the top.

OneStop Metadata Readiness Tiers

Minimum Entry Requirement

For data and metadata to be discoverable in the *OneStop* portal you only need to have a valid ISO 19115-2 compliant metadata record in accordance with NOAA's policies described in the Data Documentation Procedural Directive[1]. However, the standard is extremely flexible and compliance can be achieved with the bare minimum of information. To move beyond the minimum level of compliance and support NOAA's mission of improving discoverability, accessibility, usability and understanding of NOAA's earth science data, five options are described below and organized into three tiers of readiness that correspond to a common athletic award structure: bronze, silver and gold.

Bronze Tier

The first and lowest level of readiness, "Bronze" meets the minimal requirements for *OneStop* compliance. To obtain bronze, your metadata must be fully compliant with the ISO 19115-2 metadata standard and have a completeness v2 rubric score of at least 90%. A score of 90% ensures thorough documentation for maximum understanding, including but not limited to lineage, quality, related resources, and parameters. Furthermore, your metadata must use NASA Global Change Master Directory (GCMD) keywords when applicable and in accordance with the NOAA Keyword Implementation Guidance. Achieving the Bronze Tier of readiness optimizes the discoverability of collections indexed by *OneStop*, but does not support relevance ranking and granule search capabilities. The Silver Tier describes the additional requirements for metadata, data access and data interoperability.

Bronze Readiness

ISO Compliant Collection Metadata Required

- ISO 19115-2 Compliant Collection Metadata Records
  - Every collection level record in the data group shall have a completeness score of at least 90% following the V2 Completeness Rubric
  - URL to a browse graphic thumbnail
  - Cannot reference another collection in the gmd:parentIdentifier field
  - GCMD Science Keywords
  - GCMD Organization Keywords
  - GCMD Place Keywords
  - GCMD Platform Keywords (if applicable)
  - GCMD Instrument Keywords (if applicable)
  - GCMD Service Keywords (if applicable)
  - GCMD Project Keywords (if applicable)
  - GCMD Keyword thesaurus titles following NOAA's best practice

[1] https://ngdc.noaa.gov/wiki/index.php?title=Data_Documentation_PD

Silver Tier

The second tier of *OneStop* readiness is "silver." Silver allows data managers the greatest flexibility in balancing the level of effort to make the improvements versus OneStop capabilities needed to meet the needs of the user community. For example, if granule search capability is most important, then the data manager can pursue that course, by providing granule metadata following the OneStop ISO-Lite template. However, if having datasets appear closer to the top of the search results is a higher priority, then a data manager may want to pursue the Data Stewardship Maturity Matrix (DSMM) assessment which allows their metadata to take advantage of the Elasticsearch advanced relevance ranking algorithm. The DSMM assessment can be carried out either by using the DSMM template or by answering the DSMM questionnaire in the Collection Metadata Enterprise Tool (CoMET). Use of the DSMM questionnaire in CoMET is preferred. Finally, a data manager may choose to provide users with data that conforms to the user community needs and that is easier to analyze by converting the dataset format into one that is machine readable and interoperable (e.g., NetCDF, HDF, GeoTIFF, etc.)

Silver Readiness ISO Compliant Collection Metadata and 2 of the 4 Discretionary Items Required
Required

• ISO 19115-2 Compliant Collection Metadata Records
  ○ Every collection level record in the data group shall have a completeness score of at least 90% following the V2 Completeness Rubric
  ○ URL to a browse graphic thumbnail
  ○ Cannot reference another collection in the gmd:parentIdentifier field
  ○ GCMD Science Keywords
  ○ GCMD Place Keywords
  ○ GCMD Platform Keywords (if applicable)
  ○ GCMD Instrument Keywords (if applicable)
  ○ GCMD Service Keywords (if applicable)
  ○ GCMD Project Keywords (if applicable)
  ○ GCMD Keyword thesaurus titles following NOAA's best practice

Discretionary (choose two)
• Data are online with direct download options   ○ Examples include, but are not limited to HTTPS/FTPS, THREDDs, ERDDAP, OPeNDAP, Amazon S3, etc.

• *OneStop* ISO-Lite compliant granule metadata records
  ○ See the granule template for details.
  ○ Average completeness score for all granules in each data group shall be at least 90% or higher and contain the following *OneStop* attributes:

- An exact string match to the file identifier or DOI of the collection metadata in the "parentIdentifier" field.
- The term “Granule” in the “hierarchyLevelName” field.
- Only unique content that identifies the granule distinctly from the collection.
- Enough information to easily distinguish differences across a set of similar granules.
- URLs to every instance of the granule data file.
- Every URL shall have a protocol field associated.

- Each collection shall have a DSMM Assessment
  - Start by reading the quick start guide which provides background on the DSMM assessment.
  - Begin the assessment by filling out the DSMM template. Beginners’ Guide to DSMM document provides additional guidance to assist you in filling out the template.

- Data are in a standardized, non-proprietary machine-readable format
  - Examples include but are not limited to: GRIB, HDF, NetCDF, GeoTIFF, CSV, etc.
  - Highly recommend use of NetCDF following the NCEI NetCDF Template v2.0 or HDF.

Gold Tier

The third and final level of readiness is, “gold.” Every *OneStop* requirement is necessary to achieve gold. This level requires the highest level of compliance however it affords the greatest discoverability within *OneStop* and ensures the accessed data are machine readable. It takes advantage of every *OneStop* capability including the Elasticsearch relevance ranking system. The fifteen pilot data groups improved to meet the gold standard are listed below as they appear in the *OneStop* search results. They can be used as examples for OneStop gold standard compliance. For raw XML examples of ISOLite granules and collections, please see the links provided in the appendix.

1. CEOS/Group for High Resolution SST (GHRSST) (satellite data)
2. CO-OPS National Water Level Observation Network (NWLON) and Physical Oceanographic Real-Time System (PORTS) (observational data)
3. Digital Elevation Models (topographic model data)
4. Global Historical Climatology Network - Daily, v3 (observational data)
5. GOES-16 L1b and L2 ABI (real-time satellite data)
6. Gridded Satellite GOES (legacy satellite data)
7. Hazard Images Database (imagery)
8. NDBC Coastal-Marine Automated Network (C-MAN) and Moored Weather Buoys (observational data)
9. NEXRAD L2 (land-based wx radar data)
10. NOAA Climate Data Records (CDRs) (decadal time frame satellite data)
11. OCS Hydrographic Survey Data (bathymetry data)
12. Quality-Controlled Underway Oceanographic and Meteorological Data (COAPS-SAMOS) (observational data)
13. S-NPP/JPSS derived products L2/EDR including products from NDE (satellite data)
14. World Ocean Atlas 2013 (observational data)
15. World Ocean Database (observational data)

Gold Readiness All Required - Maximum Discoverability in *OneStop*
• ISO 19115-2/1 Compliant Collection Metadata Records
• Data are online with direct download options
• *OneStop* ISO-Lite compliant metadata records (for collections with granules)
• Each collection shall have a DSMM Assessment
• Data are in a standardized, non-proprietary machine-readable format

OneStop Metadata Indexing Workflow

In order to have your collection metadata indexed for *OneStop* you must ensure that your metadata is placed in a public web accessible folder (WAF), from which the *OneStop* tools can then harvest metadata. Currently, indexing is a manual process that is completed on an irregular basis. The goal is to have this process performed automatically and at regular intervals similarly to how the NCEI Consolidated Geoportal and the NOAA Data Catalog currently function. The workflow for providing granule metadata is under development as of this document version. Once those details are finalized, instructions will be provided here.

References

- **DSMM Documentation**
  - DSMM Introduction (http://tinyurl.com/DSMMIntro)
  - Quick Start Guide (https://doi.org/10.6084/m9.figshare.5203936.v2)
  - DSMM for Dummies (https://docs.google.com/document/d/1ZenUR-yfaw5peKgbA3jTKvbwkL_swkBNhLqV6hzekiA/edit?usp=sharing)
- **Rubrics**
  - NOAA Completeness Rubric v2 (https://ngdc.noaa.gov/wiki/index.php?title=Completeness_Rubric)
- **Templates**
  - Granule Template (https://sites.google.com/a/noaa.gov/one-stop/metadata-team/metadata-guide)
  - NCEI NetCDF Template v2.0 (https://www.nodc.noaa.gov/data/formats/netcdf)
- **NCEI Keyword Implementation Guidance** (https://docs.google.com/document/d/1m3HufIo_1bH9bBwJRm7YXl84rRhzjaApDYWRIlTvSLk/edit)
- **Tools**
  - CoMET (https://data.noaa.gov/cedit/)
- **NOAA OneStop** (https://data.noaa.gov/onestop/)

- Group for High Resolution Sea Surface Temperature (GHRSST) (https://data.noaa.gov/onestop/#/collections?q=title%3AGHRSST&f=projects:NOAA%20OneStop%20Project)
- CO-OPS National Water Level Observation Network (NWLON) and Physical Oceanographic Real-Time System (PORTS) (https://data.noaa.gov/onestop/#/collections?q=title%3ACO-OPS&f=projects:NOAA%20OneStop%20Project)
- Digital Elevation Models (https://data.noaa.gov/onestop/#/collections?q=title%3A%22elevation%22&f=projects:NOAA%20OneStop%20Project)
- Global Historical Climatology Network Daily v3.0 (https://data.noaa.gov/onestop/#/collections?q=%22Global%20Historical%20Climatology%20Network%20-%20Daily%22&f=projects:NOAA%20OneStop%20Project)
- GOES-16 ABI L1b Radiances and L2 CMIP (https://data.noaa.gov/onestop/#/collections?q=GOES&f=projects:NOAA%20OneStop%20Project;instruments:ABI%20%3E%20Advanced%20Baseline%20Imager)
- Legacy GridSat GOES (https://data.noaa.gov/onestop/#/collections?q=title%3A%22GridSAT-GOES%22&f=projects:NOAA%20OneStop%20Project)
- Hazard Images Database (https://data.noaa.gov/onestop/#/collections?q=Hazards%20AND%20Images&f=projects:NOAA%20OneStop%20Project)
- NDBC Coastal-Marine Automated Network (C-MAN) and Moored Weather Buoys (https://data.noaa.gov/onestop/#/collections?q=title%3A%22C-MAN%22&f=projects:NOAA%20OneStop%20Project)
- NEXRAD L2 (https://data.noaa.gov/onestop/#/collections?q=%22NEXRAD%22&f=projects:NOAA%20OneStop%20Project)
- NOAA Climate Data Records (CDRs) (https://data.noaa.gov/onestop/#/collections?q=%22NOAA%20Climate%20Data%20Record%22&f=projects:NOAA%20OneStop%20Project)
- OCS Hydrographic Survey Data (https://data.noaa.gov/onestop/#/collections?q=title%3A%22NOS%20Hydrographic%20Survey%22)
- Quality-Controlled Underway Oceanographic and Meteorological Data (COAPS-SAMOS) (https://data.noaa.gov/onestop/#/collections?q=title%3A%22SAMOS%22&f=projects:NOAA%20OneStop%20Project)
- S-NPP/JPSS derived products L2/EDR including products from NDE (https://data.noaa.gov/onestop/#/collections?q=title%3AJPSS&f=projects:NOAA%20OneStop%20Project)
- World Ocean Atlas 2013 (https://data.noaa.gov/onestop/#/collections?q=title%3A%22World%20Ocean%20Atlas%22&f=projects:NOAA%20OneStop%20Project)

    - World Ocean Database (https://data.noaa.gov/onestop/collections?q=World%20Ocean%20Database)
- OSGeo List of Protocol Identifiers (https://github.com/OSGeo/Cat-Interop/blob/master/LinkPropertyLookupTable.csv)

Feedback

We welcome feedback. Please utilize this Google Form which will automatically record your email so that we can contact you for clarification and resolution.

Appendix

Collection Metadata Examples (raw XML)

NOAA GOES-R Series Advanced Baseline Imager (ABI) Level 1b Radiances
https://www.ncei.noaa.gov/metadata/geoportal/rest/metadata/item/gov.noaa.ncdc:C01501/xml

World Ocean Atlas 2013
https://www.ncei.noaa.gov/metadata/geoportal/rest/metadata/item/gov.noaa.nodc:0114815/xml

NOS Hydrographic Surveys Collection
https://www.ncei.noaa.gov/metadata/geoportal/rest/metadata/item/gov.noaa.ngdc%3ANOS_Hydrographic_Survey_Collection/xml

Granule Metadata Examples (raw XML)

NOAA GOES-East G16 GLM Lightning Detection from 2018-04-14T03:54:20.0Z to 2018-04-14T03:54:40.0Z
https://sites.google.com/a/noaa.gov/one-stop/metadata-team/metadata-guide/GOES-16-GLM-2018-104OR_GLM-L2-LCFA_G16_s20181040351400_e20181040352000_c20181040352016.C01527.meta.xml?attredirects=0