



# Assessing the Robustness of ISO with the NOAA Metadata Rubric

Anna Milan

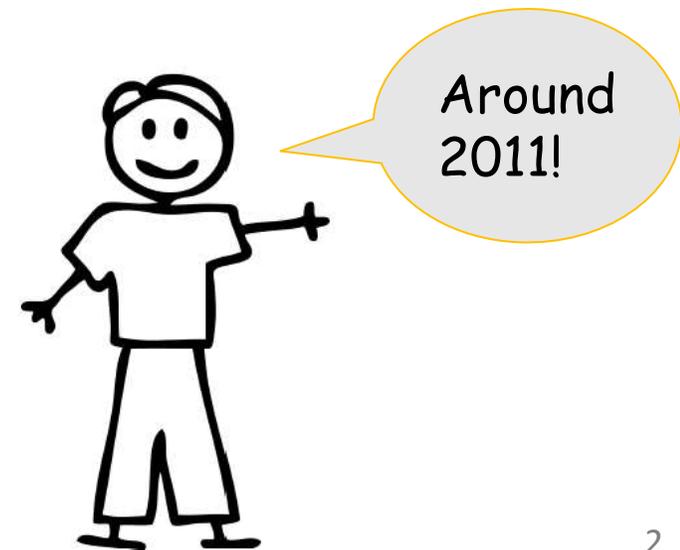
NOAA National Centers for Environmental Information

2017-05-23



# It all started when...

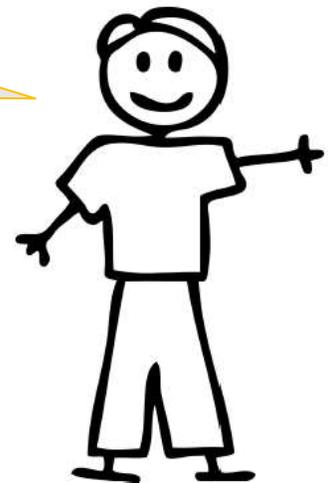
NOAA **required**, per the NOAA Data Documentation Procedural Directive, to document their datasets, models, observations and other earth science resources with the **ISO Standards**.



# However...

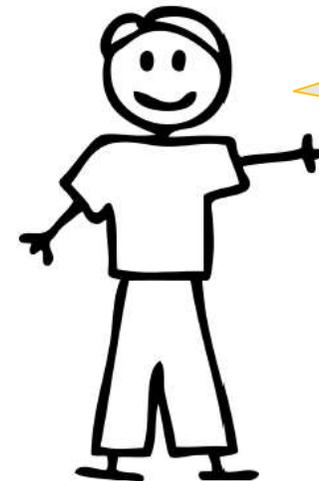
There are many **different ways** to implement the standards and **very little content** is required to create a **valid** metadata record.

How do I know what else is important to include?



# Therefore...

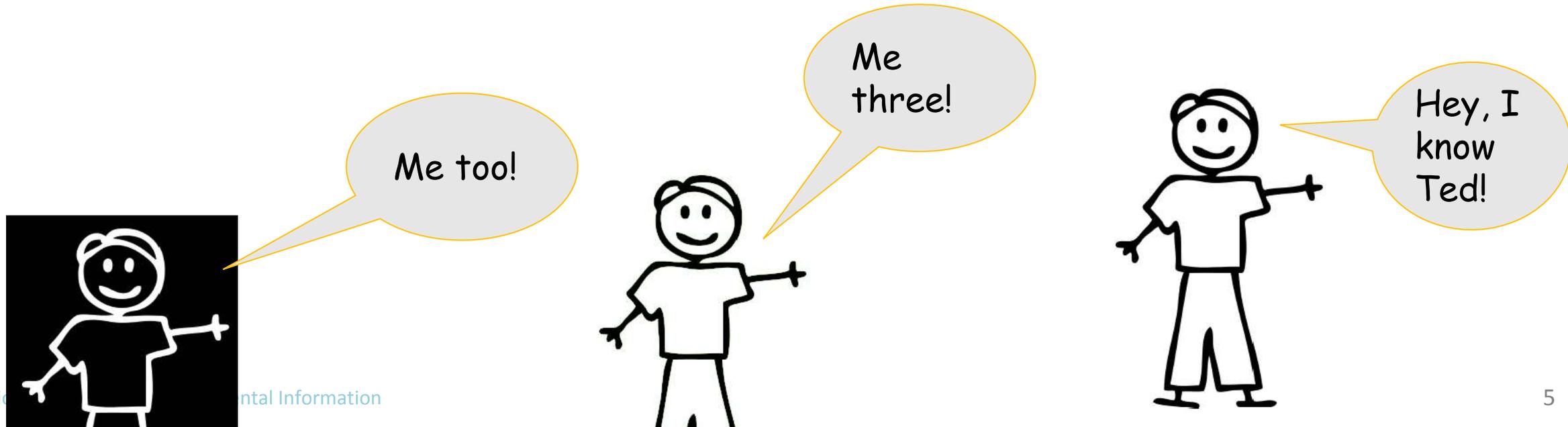
In addition to ISO compliance, the **Completeness Rubric** provides an **extra level of assessment** to help metadata authors provide **more thorough descriptions**.



I want to  
get at least  
90% score!

# History

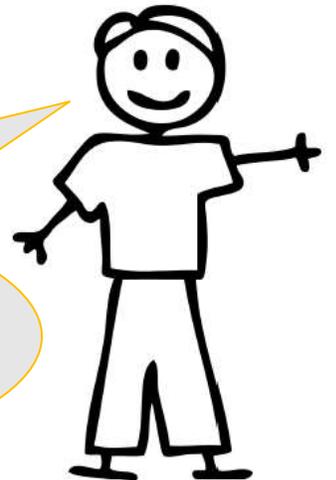
The **rules** for Completeness Rubric V2 were determined by the **NOAA Metadata Working Group** and evolved from the first NOAA Rubric developed by Ted Habermann.



# Standards Supported by Rubric V2

- ISO 19115:2003 Geographic information -- Metadata
- ISO 19115-2:2009 Geographic information -- Metadata -- Part 2: Extensions for imagery and gridded data
- ISO 19139 Geographic information -- Metadata XML schema implementation

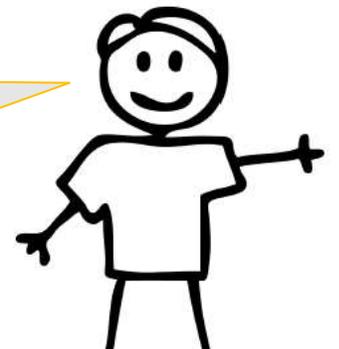
Just say: "The ISO 19\*\*\*  
Suite of Metadata Standards"



# About Completeness Rubric V2

There are **10 Categories**. Each Category looks for the **existence** of certain elements. It doesn't know if the content is meaningful, but it does check a couple codelist values to determine the required categories and some of the required elements.

The rubric supports different types of resources.



# Always Required Categories

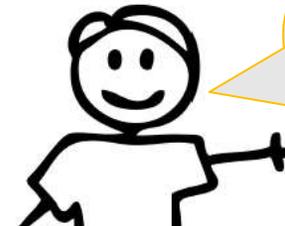
Category	Definition
<b>Identification</b>	Needed for basic discovery of the resource.
<b>Coverage</b>	The extent of the resource, such as temporal range of content, geographic bounds of content and general place names.
<b>History</b>	How the resource was collected or processed. Acquisition subcategory is recommended for raw observations. Lineage subcategory is recommended for processed resources.
<b>Connections</b>	Checks that all URLs have a Function Code and either a Name or a Description.
<b>Metadata</b>	Information about the metadata record.



This slide needs more room...

# Mostly Required Categories

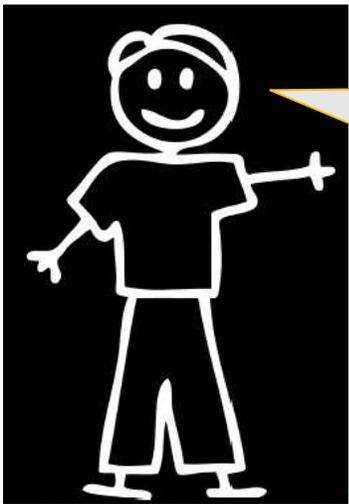
Category	Definition
<b>Access</b>	How to obtain the data, including formats, access points, distribution contacts and disclaimer statements. Not required when Resource Hierarchy Level='fieldSession' or Status = 'planned'.
<b>Content</b>	Identifies the parameters, variables or features of the resource and can be documented with at least one of the two subcategories: Attributes or Features. Not Required when Resource Hierarchy Level = 'fieldSession'.
<b>Quality</b>	Reports on how complete or accurate a resource is. Not required when Resource Hierarchy Level = 'fieldSession'.



Rubric doesn't require planned data to have data download URLs.

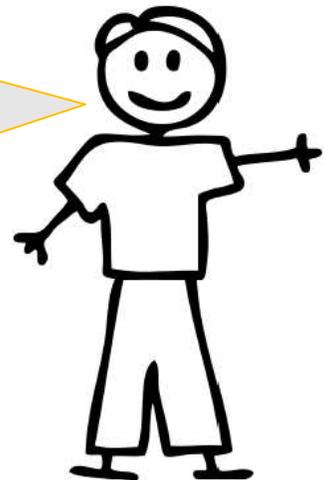
# Highly Recommended Categories

Category	Definition
<b>Associated Resources</b>	Citations to related papers, sister resources, projects or programs.
<b>Attribution</b>	Information to cite the resource and provide proper credit, such as creator names, publisher names and a Digital Object Identifier (DOI).



This info seems really important! How come it is not required?

Sometimes, we just don't have this level of detail...



# Understanding the Rubric Report

## Completeness Rubric for: Enhanced Magnetic Model 2015

Completeness Score: 90% + 16.5



Resource Hierarchy Level: 'model' — Status: 'completed'

CATEGORY	SCORE + EXTRA CREDIT	
Identification	100% + 4	Required
Access	100% + 3	Conditional - not required when Resource Hierarchy Level = 'fieldSession' or Status = 'planned'.
Coverage	100% + 2	Required
Content	100% + 0	Conditional - not required when Resource Hierarchy Level = 'fieldSession'.
History	100% + 0	Conditional - not required when Resource Hierarchy Level = 'fieldSession'.
Quality	0% + 0	Conditional - not required when Resource Hierarchy Level = 'fieldSession'.
Connections	67% + .5	Required
Metadata	100% + 1	Required
Associated Resource	+ 1	Highly Recommended
Attribution	+ 5	Highly Recommended

Total Score + Extra Credit Points

Hierarchy and Status Codelists

Category Scores + Extra Credit Points

There's more below the fold

About

ABOUT COMPLETENESS RUBRIC

MORE INFORMATION

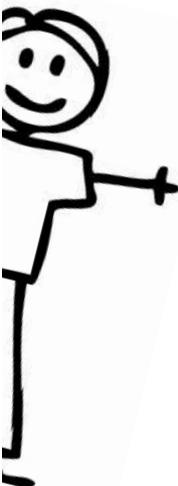
CONTACT

Please register any bugs, typos, or suggestions with this [Rubric Feedback Form](#).

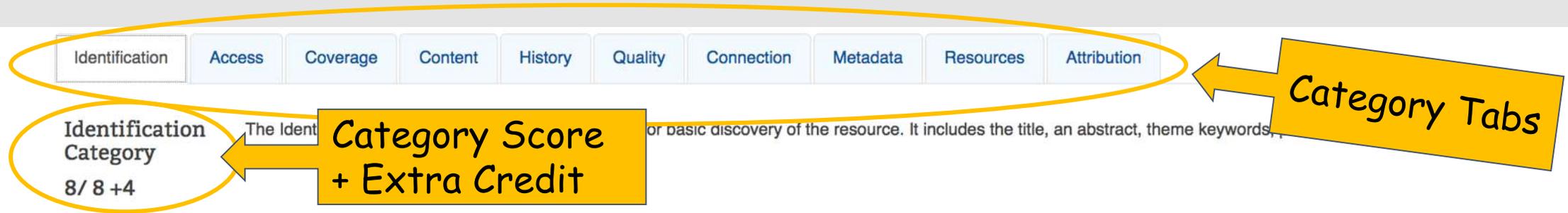
View history and status of issues at [Rubric Feedback Manager](#)

You may contact [NOAA Enterprise Email List](#) if you'd like to connect with our community. We'd love to have your involvement!

Report Bugs or Request New Features



# Understanding the Rubric Report



Identification Access Coverage Content History Quality Connection Metadata Resources Attribution

**Category Score + Extra Credit**

Identification Category 8/ 8 +4

Category Tabs

LABEL (COUNT)	RULE	SCORE	METADATA CONTENT	GUIDANCE: XPATH AND ISO EXPLORER LINKS
Resource Hierarchy Level (1)	Required	1	model	//gmi:MI_Metadata/gmd:hierarchyLevel/gmd:MD_ScopeCode/@codeListValue
Enhanced Magnetic	Recommended	EC	doi:10.7289/V56971HV	//gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:title
Resource ID (1)	Recommended	EC	doi:10.7289/V56971HV	//gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:identifier/gmd:MD_Identifier/gmd:code
Abstract (1)	Required	1	The Enhanced Magnetic Model (EMM) extends to degree and order 720, resolving magnetic anomalies down to 56 km wavelength. The higher resolution of the EMM results in significantly improved pointing accuracy than the World Magnetic Model (WMM), which (...etc)	//gmd:MD_DataIdentification/gmd:abstract
Purpose (1)	Required	1	Magnetic anomaly maps provide insight into the subsurface structure and composition of the Earth's crust. Over continental areas, magnetic anomalies illuminate geologic, tectonic, and geothermal evolution of crust and lithosphere. In the world's oceans (...etc)	//gmd:MD_DataIdentification/gmd:purpose
Resource Date (1)	Required	1	2015-05-15 ~ publication	//gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:date[gmd:CI_Date or gmd:CI_Date/gmd:date/@gco:nilReason]
Status (1)	Required	1	completed	//gmd:MD_DataIdentification/gmd:status//gmd:MD_ProgressCode/@codeListValue



# Rubric Guidance Tips



## GUIDANCE: XPATH AND ISO EXPLORER LINKS

/gmi:MI\_Metadata/gmd:hierarchyLevel/gmd:MD\_@C  
odeListValue

//gmd:MD\_DataIdentification/g  
on/gmd:CI\_Citation/gmd:ti  
tle

//gmd:MD\_DataIdentification/gmd:citation/gmd:CI\_Citation/gmd:i  
dentifier/gmd:MD\_Identifier/gmd:code

//gmd:MD\_DataIdentification/gmd:abstract

Links to ISO Explorer wiki pages

Copy Xpath and find  
location in XML record

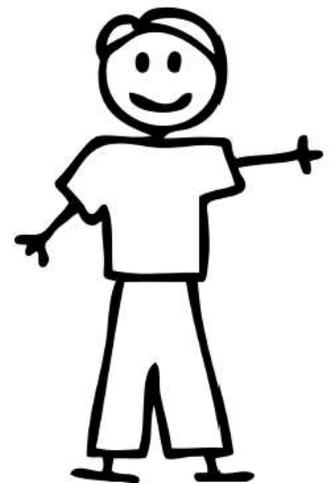
## MD DataIdentification

Elements			Definition and Recommended Practice
1	citation	1	Citation for the resource, includes name, publication date, identifiers, originators and publishers.
2	abstract	1	Brief narrative summary of the resource contents. Include summary information on general content and features; resource applications: GIS, CAD, image, database; geographic coverage: county/city name; time period of content: begin and end date or single date; and special data characteristics or limitations.
3	purpose	0..1	Summary of intentions for which the resource was developed.

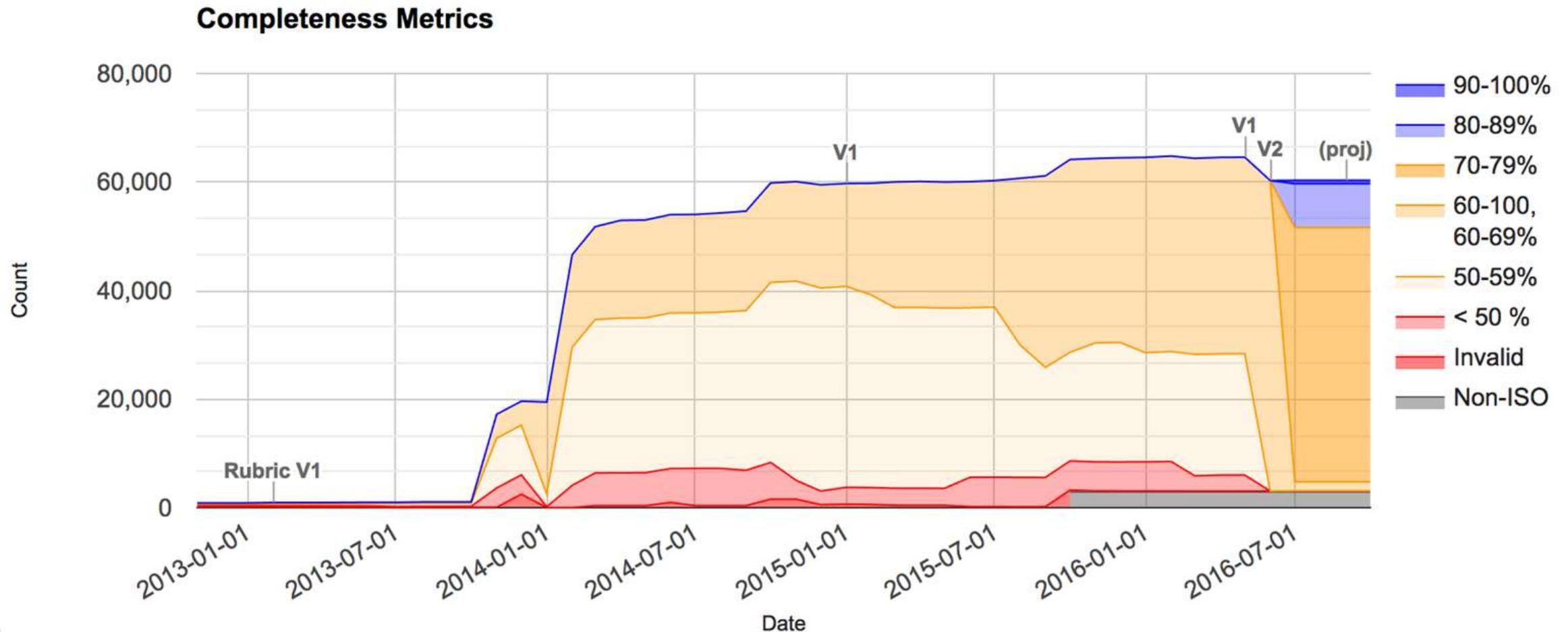
# Use the Rubric

- While developing templates
- While developing a solution that will produce high quality metadata
- While editing individual records
- Right before publishing records for discovery
- To understand the completeness of metadata over time

But HOW?



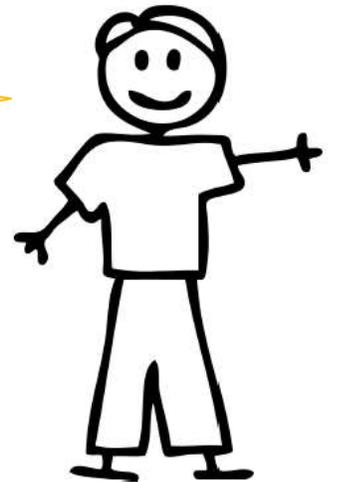
# NOAA Completeness Metrics Over Time



# How to Apply the Rubric

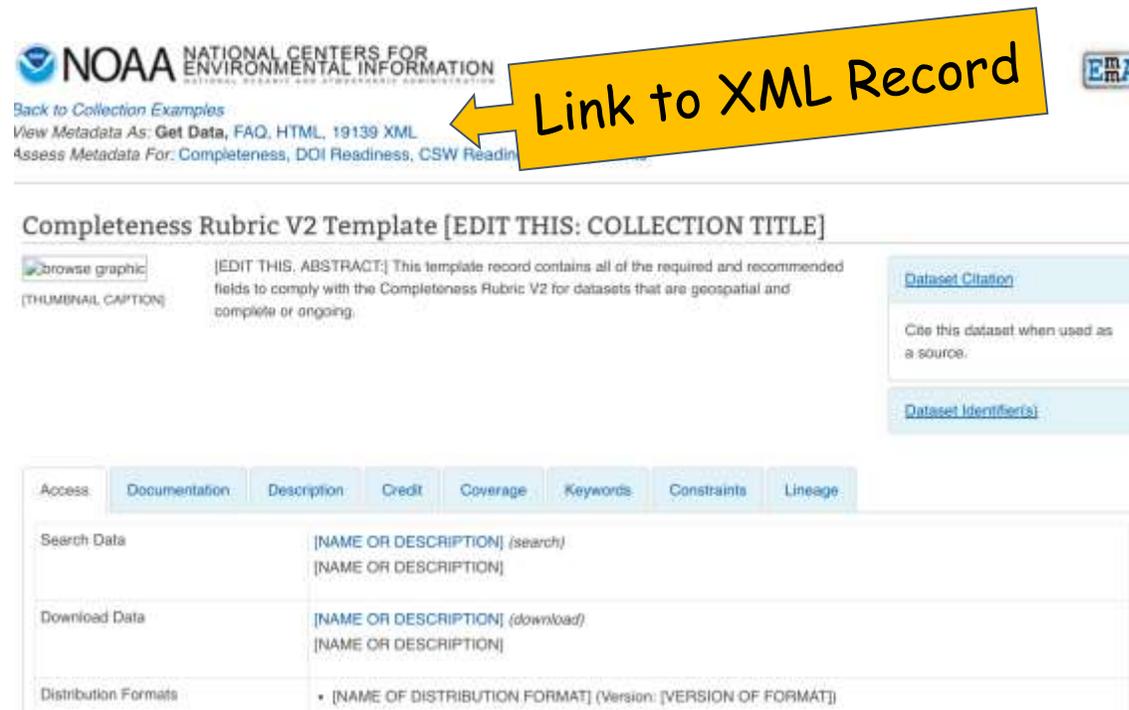
- Via the [Docucomp Record Services](#)
- Run [XSL](#) transformation in an XML editor
- Register a collection of metadata with [EMMA](#)\* for bulk assessments

This isn't helping.  
HOW how?



# When Starting from Scratch....

Use the Rubric V2 [Template](#) in the Examples WAF and bring into your favorite XML Editing environment.



NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION

Back to Collection Examples  
View Metadata As: [Get Data](#), [FAQ](#), [HTML](#), [19139 XML](#)  
Assess Metadata For: [Completeness](#), [DOI Readiness](#), [CSW Readiness](#)

**Link to XML Record**

Completeness Rubric V2 Template [EDIT THIS: COLLECTION TITLE]

[browse graphic](#)  
(THUMBNAIL CAPTION)

[EDIT THIS. ABSTRACT:] This template record contains all of the required and recommended fields to comply with the Completeness Rubric V2 for datasets that are geospatial and complete or ongoing.

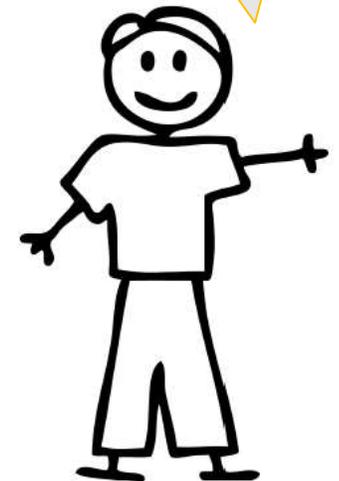
[Dataset Citation](#)  
Cite this dataset when used as a source.

[Dataset Identifier\(s\)](#)

Access: [Documentation](#) [Description](#) [Credit](#) [Coverage](#) [Keywords](#) [Constraints](#) [Lineage](#)

Search Data	[NAME OR DESCRIPTION] (search) [NAME OR DESCRIPTION]
Download Data	[NAME OR DESCRIPTION] (download) [NAME OR DESCRIPTION]
Distribution Formats	• [NAME OF DISTRIBUTION FORMAT] (Version: [VERSION OF FORMAT])

I like  
Oxygen  
XML  
EDitor



# Docucomp Record Services

## Record Services

A suite of tools for creating, validating, resolving and improving ISO 19115-2 metadata records.

For help, see [Enterprise Metadata Tools](#)

Input Type:  File:

Input XML Metadata Record

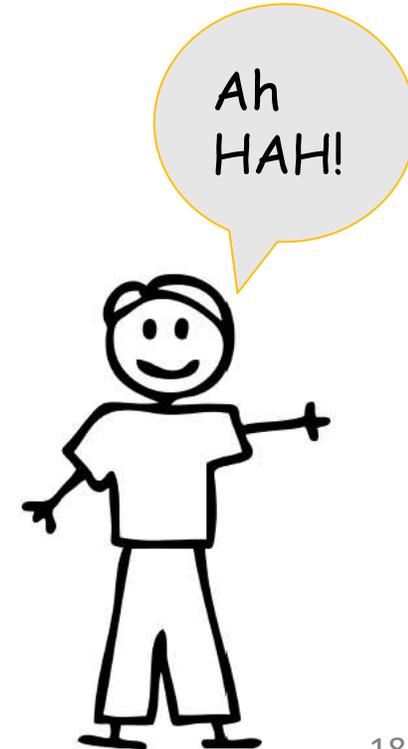
Url:

### Choose Service:

- Check XML format      Check for well formed XML. Input: Any XML file.
- Resolve      Resolve URL references to components. Input: 19139 schema valid XML file
  
- ISO Validate      Validate metadata with 19139 Schema. Input: 19139 XML file
- Schematron      Check metadata with supplementary Schematron rules. Input: 19139 schema valid XML file
  
- Link Check (slow)      Check metadata with supplementary Schematron rules. Input: 19139 schema valid XML file
- ISO To Rubric V2      Generate IMPROVED completeness assessment report. Input: 19139 schema valid XML file
- ISO To Rubric V1      Generate completeness assessment report. Input: 19139 schema valid XML file

Select ISO to Rubric V2

Hit Submit Button



# Evaluate the Results, Edit, Repeat

Completeness Rubric for: Completeness Rubric V2 Template [EDIT THE COLLECTION TITLE]

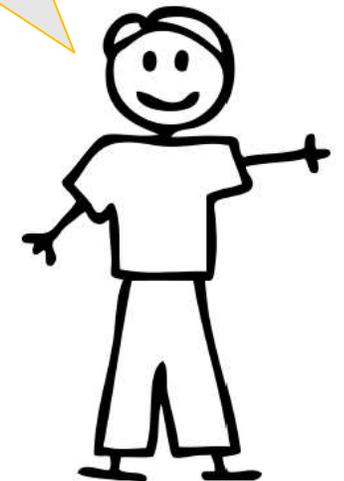
Completeness Score: 100% + 32.9



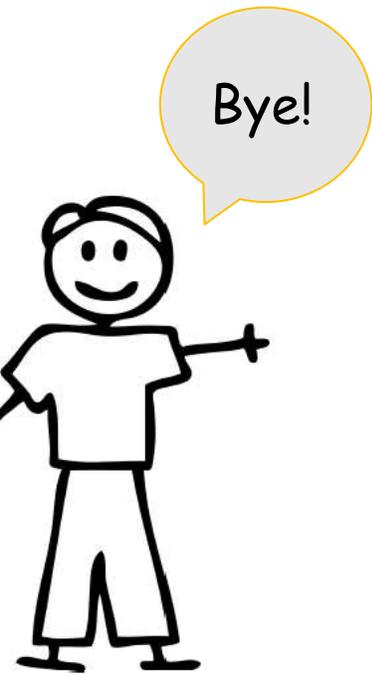
Resource Hierarchy Level: 'series' — Status: 'ongoing'

CATEGORY	SCORE + EXTRA CREDIT	RUBRIC REQUIREMENTS
Identification	100% + 3	Required
Access	100% + 3	Conditional - not required when Resource Hierarchy Level = 'fieldSession' or Status = 'planned'.
Coverage	100% + 4	Required
Content	100% + 4	Conditional - not required when Resource Hierarchy Level = 'fieldSession'.
History	100% + 5	Required
Quality	100% + 1	Conditional - not required when Resource Hierarchy Level = 'fieldSession'.
Connections	100% + .9	Required
Metadata	100% + 1	Required

100%! I'm done, right?



# Thanks!



[enterprise.metadata@noaa.gov](mailto:enterprise.metadata@noaa.gov)



# References

- ISO Explorer on the NOAA Environmental Data Management Wiki
  - [https://geo-ide.noaa.gov/wiki/index.php?title=Category:ISO Explorer](https://geo-ide.noaa.gov/wiki/index.php?title=Category:ISO_Explorer)
- Docucomp
  - <https://www.ngdc.noaa.gov/docucomp>
- Examples WAF
  - <https://www.ngdc.noaa.gov/metadata/published/Examples/iso>
- Enterprise Metadata Management Architecture (EMMA)
  - <https://www.ngdc.noaa.gov/metadata/emma>

See you again soon?

