International Standards Organization 19115-1 Recommendation for Data Discovery Analysis for the National Center for Atmospheric Research

Table of Contents

Executive Summary

This report presents the results of a National Center for Atmospheric Research (NCAR) metadata completeness evaluation with respect to the International Organization for Standardization 19115-1 recommendation (ISO) for Data Discovery. ISO is an organization formed to help improve consistent identification of data and other resources with the goal of making data more accessible and useful. ISO 19115-1 provides a three-level metadata recommendation that includes mandatory, conditional, and optional elements.

NCAR has many ways of sharing the data they produce and archive. The Data Stewardship Engineering Team (DSET) is responsible for helping NCAR labs share their data efficiently and in a unified manner. Currently there are 9 labs, each with varying types of information collections. There are also a variety of metadata practices employed; some groups use xml standards from external sources, some use XML standards formed at NCAR, and some have their own structured documentation in the form of a database or ASCII headers. This report focuses on the ISO, MODS, DataCite, netCDF, and RDA-CISL holdings at NCAR. We acquired samples of each of these collections in order to explore completeness with respect to the ISO recommendation and as an introduction to the metadata evaluation and improvement tools we are developing. Applying the recommendation to a data center's metadata, regardless of the dialect they utilize can help prepare organizations that are currently hoping to identify the preparedness of their labs for sharing metadata concepts the organization has determined are important.

One important observation is that only the ISO 19115-1 dialect contains all of the concepts required by the ISO 19115-1 recommendations. RDA-CISL is missing two mandatory concepts, one conditional concept, and three optional concepts. DataCite is missing four mandatory concepts, and nine optional concepts. MODS is missing one mandatory concepts, and two optional concepts. netCDF is missing one mandatory concepts, two conditional concepts, and four optional concepts. EOL is missing one conditional concept, and two optional concepts. If it is essential to meet the ISO 19115-1 recommendation, choices must be made regarding expanding, translating or changing the current dialects used by NCAR labs.

Our metadata sample included 2505 records from 7 NCAR labs. Of those, no records included all of the metadata concepts in the ISO-1 recommendation that are contained in the dialect the record is written in. The first two signature groups of the ISO dialect are the most complete records at NCAR labs with respect to the ISO 19115-1 Recommendation.
CGD did not select assets with sharable metadata. Mapping was done with CGD to determine whether the CESM experiments database contained the concepts in the ISO 19115-1 Recommendation and EOL to provide comparisons of unshared metadata to the concepts in the ISO 19115-1 Recommendation. ACOM did not have metadata, or machine readable structured documentation. As such, research was done to find shared vocabularies and ontologies in chemistry to help them document their data. Decisions have to made about how they want to store and create metadata before the assets can be assessed.

NCAR Labs and the Dialects Analyzed

	CGD	DataCite	EOL	ISO	MODS	netCDF	RDA-CISL
Atmospheric Chemistry Observations and Modeling (ACOM)							
Climate & Global Dynamics (CGD)	X						
Computational and Informational Systems Lab (CISL)							X
Earth Observing Lab (EOL)			X	X			
High Altitude Observatory (HAO)		X					
Integrated Information Services (IIS)					X		
Mesoscale and Microscale Meteorology (MMM)		X					
Research Applications Lab (RAL)						X	
Unidata (UCP)		X		X			

ISO 19115-1 Recommendation for Data Discovery – What is it?

The International Organization for Standardization (ISO) is an international non-governmental organization founded to develop consensus-based international standards that support research, governmental and commercial enterprises. The ISO 19115-1 recommendation focuses on providing the means for discovery of geospatial data. The ISO 19115-1 dialect is the newest version of ISO 19115-2 and ISO 19115, dialects which are in common use by governmental, non-governmental and private organizations to describe geospatial based data. About 20% of the NCAR records analyzed are written in these dialects.

In the context of the terminology we use (see Glossary), ISO is an organization that created a recommendation with three levels, mandatory, conditional, and optional described in Table F.1 of the ISO 19115-1 Metadata Specification. Concepts included in all three levels are listed with definitions and XPaths in several dialects on the ISO 19115-1 Recommendation Page. The recommendations are useful for communities looking for expert guidance about metadata concepts that are valuable for data discovery. Applying the ISO recommendation to a data center's metadata, regardless of the dialect they utilize, can help prepare organizations that are interested in improving the discovery of their information collections.

This assessment of a sampling of 7 collections from the 7 NCAR labs with sharable metadata is based on the ISO 19115-1 Data Discovery recommendation. The NCAR assets come from 5 XML dialects, ISO, MODS, DataCite, netCDF, RDA-CISL. The mapping of the EOL Zith database as well as CGD's CESM experiments database are also added to the comparison of dialects and the ISO 19115-1 recommendation to highlight the opportunities labs at NCAR have to develop internal metadata into a shared format.

Recommendation Dialect Comparison – How Does My Dialect Fit?

Recommendations are created in order to address metadata needs perceived by the organizations that create them, e.g. data discovery, use, understanding. It is important to understand the fit, and the misfit, between the recommendation and the dialects.

This section provides information about similarities and differences between the ISO-1 recommendation and the NCAR dialect implementations. We describe the recommendation-dialect fit in the following ways:

- A recommendation comparison report
- A chart comparing the concepts in the ISO-1 recommendation and the NCAR dialects.
- A graph comparing the maximum number of concepts in each dialect compared to the recommendation.
- A Recommendation/Dialect comparison that lists all concepts in the ISO-1 recommendation and NCAR dialects.
- Tables that describe the concepts in the ISO-1 recommendation that are missing in the NCAR dialects.
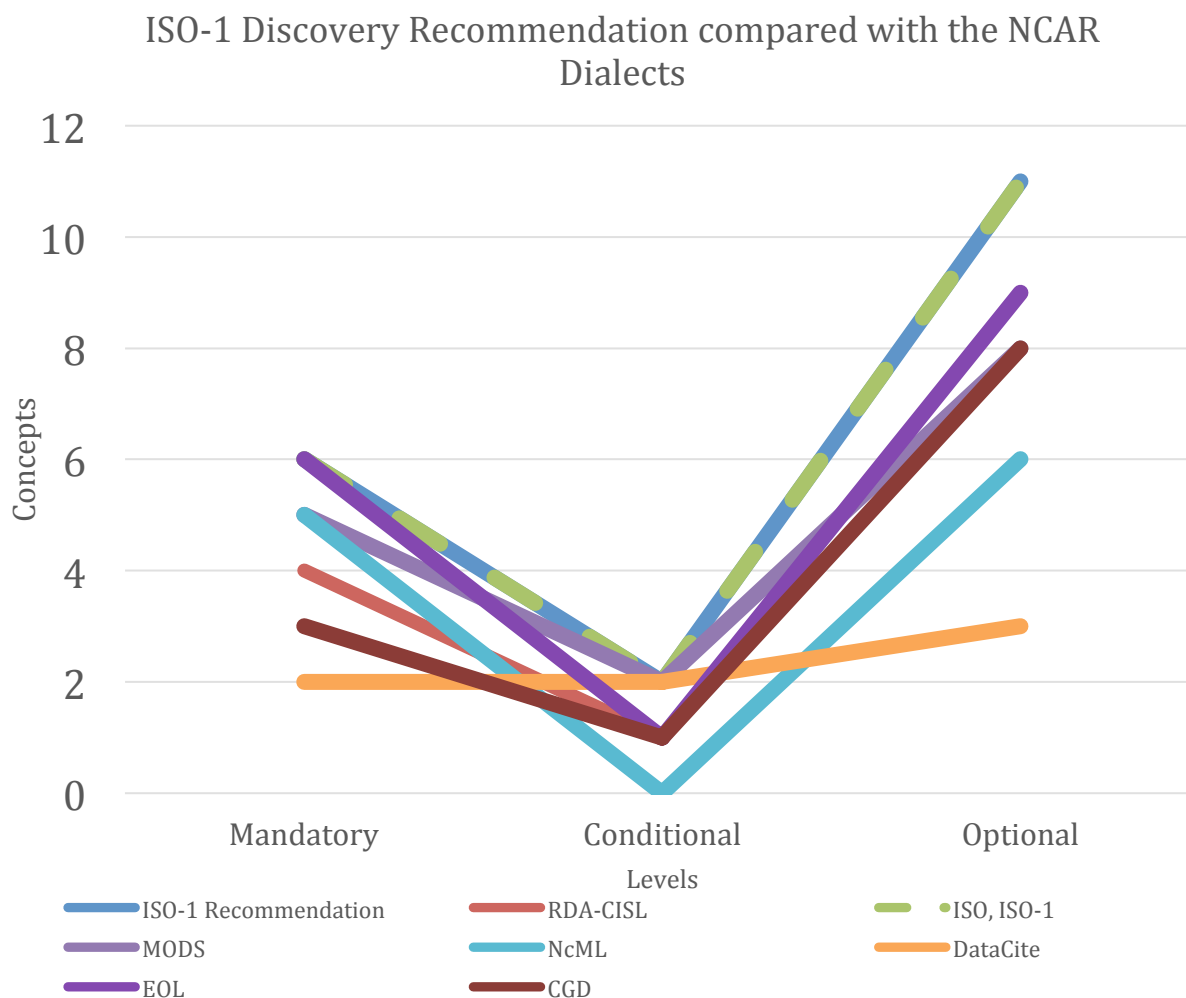
Recommendation Comparison Report

The purpose of the Recommendation Comparison report is to show the concepts that are included in each of the recommendation levels (See Glossary) being compared. A concept is a generalized term for a documentation entity, and a recommendation level is a list of concepts that an organization identifies for achieving a documentation goal. The recommendation levels included in this study are: ISO-1 for Data Discovery Mandatory, ISO-1 for Data Discovery Conditional, and ISO-1 for Data Discovery Optional (see table below).

Concept	Score	Description	Mandatory	Conditional	Optional
Abstract	1	A paragraph describing the resource.	X		
Bounding Box	1	A bounding box for identifying a geographic area of interest	X		
Metadata Contact	1	The organization or person currently responsible for the metadata.	X		
Metadata Identifier	1	A phrase or string which uniquely identifies the metadata file/record.			X
Metadata Use Constraints	1	Information about how the metadata may or may not be used after access is granted to assure the protection of privacy or intellectual property.			X
Modified Date	1	Date on which the metadata record (not the resource) was created or updated within the catalogue.	X		
Resource Contact	1	The organization or person responsible for answering questions about the resource.			X
Resource Creation/Revision Date	1	The date the resource was created			X
Resource Identifier	1	Identifier for the resource described by the metadata			X
Resource Language	1	The language of the resource.		X	
Resource Lineage	1	A description of the source(s) and production process(es) used in producing the resource.			X
Resource on-line Link	1	Online link referencing additional information about the resource.			X
Resource Title	1	A short description of the resource. The title should be descriptive enough so that when a user is presented with a list of titles the general content of the data set can be determined.	X		
Resource Type	1	A resource code identifying the type of resource; e.g. dataset, a collection, an application (See MD_ScopeCode) for which the metadata describes.		X	
Spatial Resolution	1	The nominal scale and/or spatial resolution of the resource.			X
Temporal Extent	1	The temporal extent of the resource			X
Theme Keyword	1	A word or phrase that describes some aspect of a resource. Can be one of several types.			X
Topic Category	1	High level category enumeration used in ISO	X		
Vertical Extent	1	The vertical extent of the resource			X

Recommendation/Dialect Maximum Graph

This graph compares the number of concepts included in the ISO-1 recommendation levels (recommendation maximum) to the maximum number of these concepts supported by the NCAR dialects (dialect maximum). The three levels of the ISO-1 recommendation (mandatory, conditional and optional concepts) include 6, 2, and 11 concepts respectively as indicated by the upper line in the Figure below. This Recommendation Maximum defines the highest completeness scores with respect to these recommendations for any metadata dialect. RDA-CISL is missing two mandatory concepts, one conditional concept, and three optional concepts. DataCite is missing four mandatory concepts, and nine optional concepts. MODS is missing one mandatory concepts, and two optional concepts. netCDF is missing one mandatory concepts, two conditional concepts, and four optional concepts. EOL is missing one conditional concept, and two optional concepts. The lower lines in the Figure shows these dialect maxima. The difference between the Recommendation Maximum (6 2 11) and the Dialect Maximum e.g. RDA-CISL (4 1 8) indicates that there are two mandatory ISO-1 concepts that are missing from the RDA-CISL dialect, as well as one conditional concept and three optional concepts.

ISO-1 Discovery Recommendation compared with the NCAR Dialects

Recommendation/Dialect Comparison Report

These tables show all of the concepts included in the ISO-1 recommendation, and verify their existence in the NCAR dialects with an "X".

Mandatory Level

Concepts	Score	Description	RDA-CISL	ISO	MODS	netCDF	DCITE	CGD	EOL
Abstract	7	A paragraph describing the resource.	X	X	X	X	X	X	X
Bounding Box	5	A bounding box for identifying a geographic area of interest	X	X	X	X			X
Metadata Contact	5	The organization or person currently responsible for the metadata.		X	X	X		X	X
Modified Date	4	Date on which the metadata record (not the resource) was created or updated within the catalogue.		X	X	X			X
Resource Title	7	A short description of the resource. The title should be descriptive enough so that when a user is presented with a list of titles the general content of the data set can be determined.	X	X	X	X	X	X	X
Topic Category	3	High level category enumeration used in ISO	X	X					X

Conditional Level

Concepts	Score	Description	RDA-CISL	ISO	MODS	netCDF	DCITE	CGD	EOL
Resource Language	4	The language of the resource.		X	X		X		X
Resource Type	5	A resource code identifying the type of resource; e.g. dataset, a collection, an application (See MD_ScopeCode) for which the metadata describes.	X	X	X		X	X	

Optional Level

Items	Score	Description	RDA-CISL	ISO	MODS	netCDF	DCITE	CGD	EOL
Metadata Identifier	4	A phrase or string which uniquely identifies the metadata file/record.		X	X			X	X
Metadata Use Constraints	1	Information about how the metadata may or may not be used after access is granted to assure the protection of privacy or intellectual property.		X					
Resource Contact	6	The organization or person responsible for answering questions about the re-	X	X	X	X		X	X

		source.							
Resource Creation/Revision Date	7	The date the resource was created	X	X	X	X	X	X	X
Resource Identifier	6	Identifier for the resource described by the metadata	X	X	X	X	X		X
Resource Lineage	1	A description of the source(s) and production process(es) used in producing the resource.		X				X	
Resource on-line Link	5	Online link referencing additional information about the resource.	X	X	X			X	X
Spatial Resolution	6	The nominal scale and/or spatial resolution of the resource.	X	X	X	X		X	X
Temporal Extent	6	The temporal extent of the resource	X	X	X	X		X	X
Theme Keyword	7	A word or phrase that describes some aspect of a resource. Can be one of several types.	X	X	X	X	X	X	X
Vertical Extent	4	The vertical extent of the resource	X	X		X			X

ISO-1 Concepts missing from NCAR Dialects

The Tables below provide lists of the ISO-1 recommendation concepts that are missing from the NCAR dialects for each of the three levels. The ISO and ISO-1 dialects contain all concepts in the recommendation. If a level is missing from the dialect's section, it is complete; there are no missing concepts in that level.

RDA-CISL
Missing Mandatory Concepts

Concept	Description
Modified Date	Date on which the metadata record (not the resource) was created or updated within the catalogue.
Metadata Contact	The organization or person currently responsible for the metadata.

Missing Recommended Concepts

Concept	Description
Resource Language	The language of the resource.

Missing Optional Concepts

Concept	Description
Metadata Identifier	A phrase or string which uniquely identifies the metadata file/record.
Metadata Use Constraints	Information about how the data may or may not be used after access is granted to assure the protection of privacy or intellectual property. This includes any special restrictions, legal prerequisites, terms and conditions, and/or limitations on using the data set. Data providers may request acknowledgement of the data from users and claim no responsibility for quality and completeness of data.
Resource Lineage	A description of the source(s) and production process(es) used in producing the resource.

MODS
Missing Mandatory Concepts

Concept	Description
Topic Category	High level category enumeration used in ISO

Missing Optional Concepts

Concept	Description
Metadata Use Constraints	Information about how the data may or may not be used after access is granted to assure the protection of privacy or intellectual property. This includes any special restrictions, legal prerequisites, terms and conditions, and/or limitations on using the data set. Data providers may request acknowledgement of the data from users and claim no responsibility for quality and completeness of data.
Resource Lineage	A description of the source(s) and production process(es) used in producing the resource.
Vertical Extent	The vertical extent of the resource

netCDF
Missing Mandatory Concepts

Concept	Description
Topic Category	High level category enumeration used in ISO

Missing Conditional Concepts

Concept	Description
Resource Language	The language of the resource.
Resource Type	A resource code identifying the type of resource; e.g. dataset, a collection, an application for which the metadata describes.

Missing Optional Concepts

Concept	Description
Metadata Identifier	A phrase or string which uniquely identifies the metadata file/record.
Metadata Use Constraints	Information about how the metadata may or may not be used after access is granted to assure the protection of privacy or intellectual property.
Resource Lineage	A description of the source(s) and production process(es) used in producing the resource.
Resource on-line Link	Online link referencing additional information about the resource.

DataCite
Missing Mandatory Concepts

Concept	Description
Bounding Box	A bounding box for identifying a geographic area of interest
Metadata Contact	The organization or person currently responsible for the metadata.
Modified Date	Date on which the metadata record (not the resource) was created or updated within the catalogue.
Topic Category	High level category enumeration used in ISO

Missing Optional Concepts

Concept	Description
Metadata Identifier	A phrase or string which uniquely identifies the metadata file/record.
Metadata Use Constraints	Information about how the metadata may or may not be used after access is granted to assure the protection of privacy or intellectual property.
Resource Contact	The organization or person responsible for answering questions about the resource.
Resource Lineage	A description of the source(s) and production process(es) used in producing the resource.
Resource on-line Link	Online link referencing additional information about the resource.
Spatial Resolution	The nominal scale and/or spatial resolution of the resource.
Temporal Extent	The temporal extent of the resource
Vertical Extent	The vertical extent of the resource

EOL
Missing Conditional Concepts

Concept	Description
Resource Type	A resource code identifying the type of resource; e.g. dataset, a collection, an application for which the metadata describes.

Missing Optional Concepts

Concept	Description
Metadata Use Constraints	Information about how the metadata may or may not be used after access is granted to assure the protection of privacy or intellectual property.
Resource Lineage	A description of the source(s) and production process(es) used in producing the resource.

CGD
Missing Mandatory Concepts

Concept	Description
Bounding Box	A bounding box for identifying a geographic area of interest
Modified Date	Date on which the metadata record (not the resource) was created or updated within the catalogue.
Topic Category	High level category enumeration used in ISO

Missing Conditional Concepts

Concept	Description
Resource Language	The language of the resource.

Missing Optional Concepts

Concept	Description
Resource Identifier	Identifier for the resource described by the metadata
Metadata Use Constraints	Information about how the metadata may or may not be used after access is granted to assure the protection of privacy or intellectual property.
Resource Contact	The organization or person responsible for answering questions about the resource.
Resource Lineage	A description of the source(s) and production process(es) used in producing the resource.
Resource on-line Link	Online link referencing additional information about the resource.
Spatial Resolution	The nominal scale and/or spatial resolution of the resource.
Temporal Extent	The temporal extent of the resource
Vertical Extent	The vertical extent of the resource

Metadata Analysis – How Complete Are My Metadata?

This section presents the results of an analysis of the completeness of a collection of metadata records in a dialect or a set of dialects with respect to the recommendation(s) applied. A collection is a group of metadata records, commonly organized by data center, organization or project and often stored in a database or web accessible folder. Collections are composed of metadata records of the same dialect.

Sample metadata were obtained from RAL, Unidata, CISL, EOL, HAO, MMM, and IIS after meetings with labs to determine the state of the metadata for the assets the lab wanted to have analyzed. CGD and ACOM had a database and ASCII headers respectively. The samples from collections are variable in size from 4 to 1300 records. This section presents the results of an analysis of the completeness of these metadata collections with respect to the ISO-1 recommendation. Completeness is measured by determining how many concepts are missing from the metadata records for each of the ISO-1 recommendation levels.

These results are presented as counts of records with identical completeness scores with respect to the recommendation(s). The completeness scores are given in terms of the **number of elements that are missing** from a record, so **low scores are good**. When a recommendation includes multiple levels (e.g. Mandatory, Recommended, and Optional), the scores are given as a series of numbers, one for each level. These are termed signatures (see Glossary). Typically, many records are missing the same concepts and, therefore, have identical signatures. The signature 2 3 1 indicates a metadata record that has been tested for three levels and is missing 2 mandatory, 3 recommended, and 1 optional concepts. This record is less complete than a record with a signature of 1 1 1 and more complete than a record with a signature of 3 4 3.

Completeness Results for NCAR Metadata Dialects

The differences between the completeness scores in the following chart reflect concepts that are present in the more complete records and missing from the less complete ones. By organizing records based on their scores across all three levels we create a set of signatures for the collection, which allows us to identify groups of records that typically contain the same concepts.

We report completeness for the three ISO-1 recommendation levels. The order of the levels is mandatory, conditional, and optional. A score of "0 0 0" indicates that the record is as complete as possible with respect to the ISO-1 recommendations. There are few complete records in the collections analyzed in this report.

Additionally, the concepts missing in each signature group are given. Concepts that do not appear in any records but are contained in the given dialect are listed and described as "unused concepts". All "missing concept" signatures also include the unused concepts.
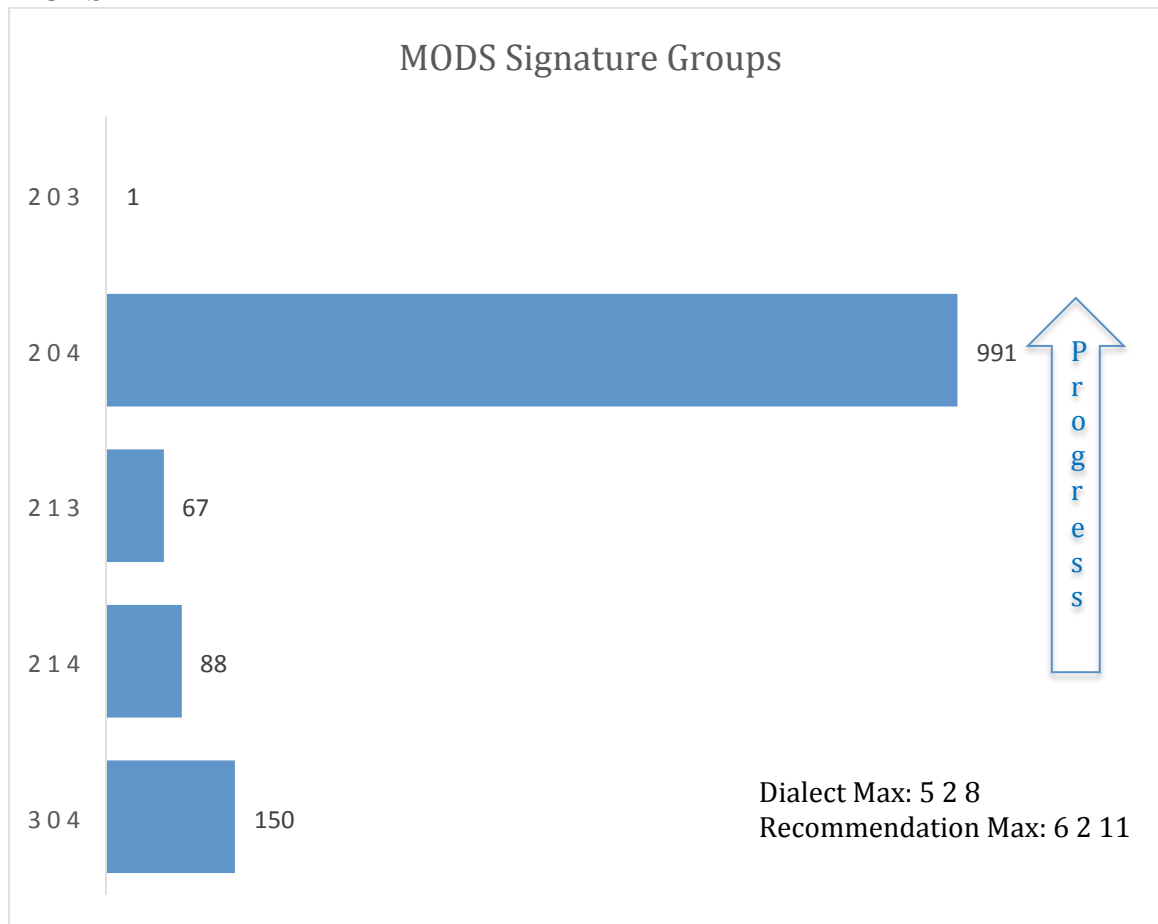
RDA-CISL

RDA-CISL Signature Groups

Signature	Value
0 0 0	2
0 0 1	133
0 0 2	140
0 0 3	1
1 0 1	3
1 0 2	196
1 0 3	177
1 0 5	2
1 1 3	1

Progress

Dialect Max: 4 1 8
Recommendation Max: 6 2 11

Incomplete Concepts by Signature

Score	Concepts
0 0 0	
0 0 1	Vertical Extent
0 0 2	Vertical Extent, Resource on-line Link
0 0 3	Spatial Resolution, Vertical Extent, Resource on-line Link
1 0 1	Bounding Box, Spatial Resolution
1 0 2	Bounding Box, Spatial Resolution, Vertical Extent, Resource on-line Link
1 0 3	Bounding Box, Spatial Resolution, Vertical Extent, Resource on-line Link
1 0 5	Bounding Box, Spatial Resolution, Resource Creation/Revision Date, Temporal Extent, Vertical Extent, Resource on-line Link
1 1 3	Bounding Box, Spatial Resolution, Vertical Extent, Resource on-line Link, Resource Type

MODS

MODS Signature Groups

Score	Value
2 0 3	1
2 0 4	991
2 1 3	67
2 1 4	88
3 0 4	150

Progress

Dialect Max: 5 2 8
Recommendation Max: 6 2 11

Unused Concepts

Concept	Description
Modified Date	Date on which the metadata record (not the resource) was created or updated within the catalogue.
Bounding Box	A bounding box for identifying a geographic area of interest
Spatial Resolution	The nominal scale and/or spatial resolution of the resource.
Temporal Extent	The temporal extent of the resource
Metadata Identifier	A phrase or string which uniquely identifies the metadata file/record.

Incomplete Concepts by Signature

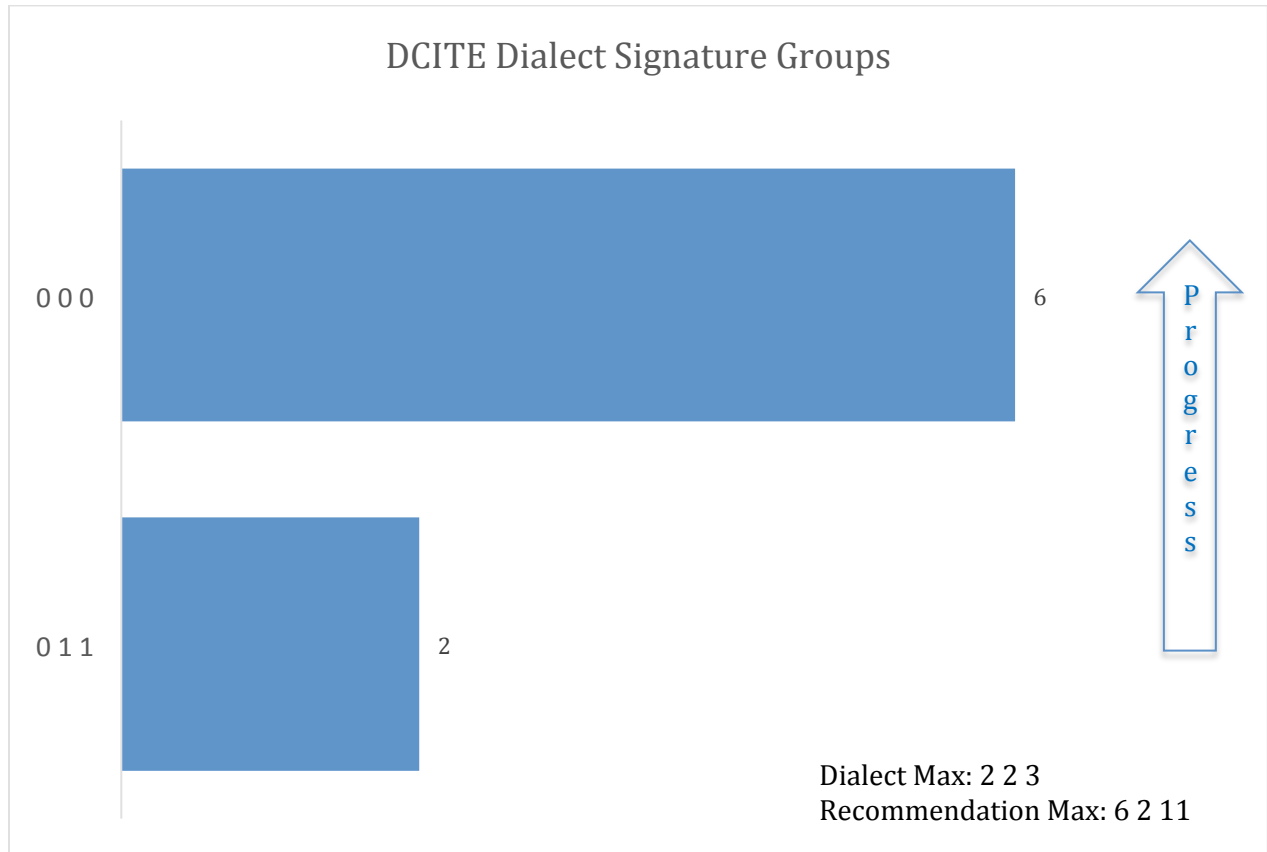| Score	Concepts
2 0 3	The unused concepts described above.
2 0 4	Resource Creation/Revision Date
2 1 3	Resource Language
2 1 4	Resource Creation/Revision Date, Resource Language
3 0 4	Resource Creation/Revision Date, Abstract

ISO

ISO Signature Groups
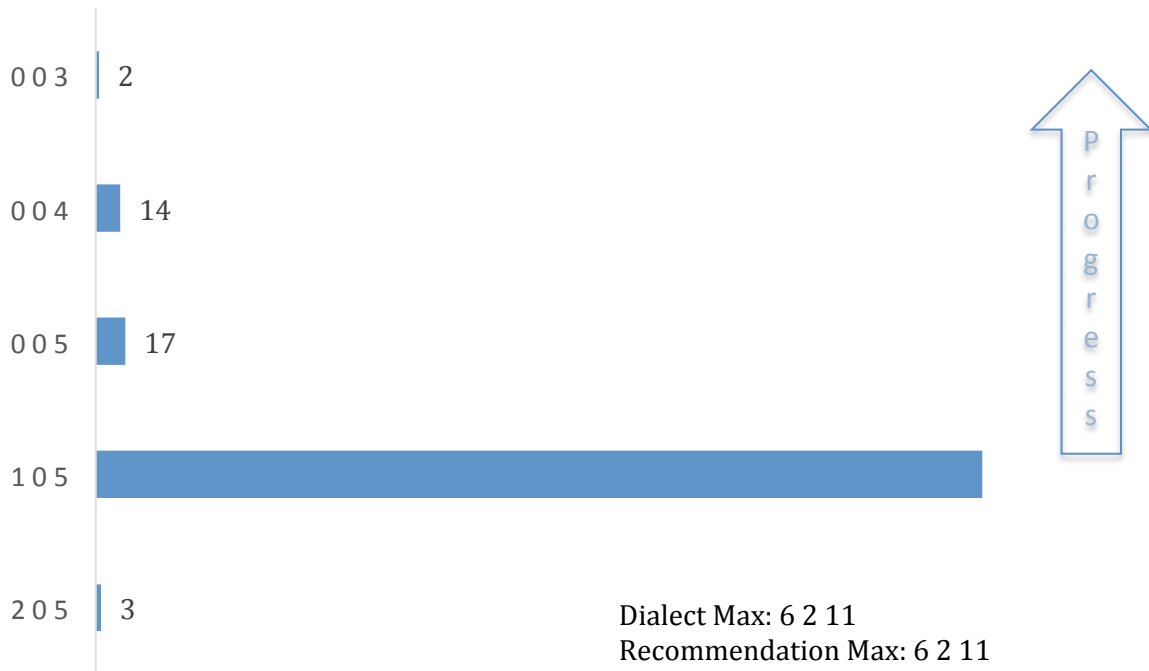
Dialect Max: 6 2 11
Recommendation Max: 6 2 11

Unused Concepts

Concept	Description
Metadata Use Constraints	Information about how the metadata may or may not be used after access is granted to assure the protection of privacy or intellectual property.
Spatial Resolution	The nominal scale and/or spatial resolution of the resource.

Incomplete Concepts by Signature

| Score	Concepts
0 0 3	The unused concepts described above plus Resource Lineage
0 0 4	Resource Creation/Revision Date, Vertical Extent
0 0 5	Resource on-line Link, Vertical Extent, Resource Lineage
1 0 5	Vertical Extent, Resource Lineage, Resource on-line Link, Topic Category
2 0 5	Vertical Extent, Resource Lineage, Resource on-line Link, Topic Category, Abstract

DataCite

DCITE Dialect Signature Groups

0 0 0   6

0 1 1   2

Progress

Dialect Max: 2 2 3
Recommendation Max: 6 2 11

Incomplete Concepts by Signature

Score	Missing Concepts
0 0 0	
0 1 1	Resource Language, Resource Creation/Revision Date

netCDF

netCDF Signature Groups

0 0 2   4

Dialect Max: 5 0 6
Recommendation Max: 6 2 11
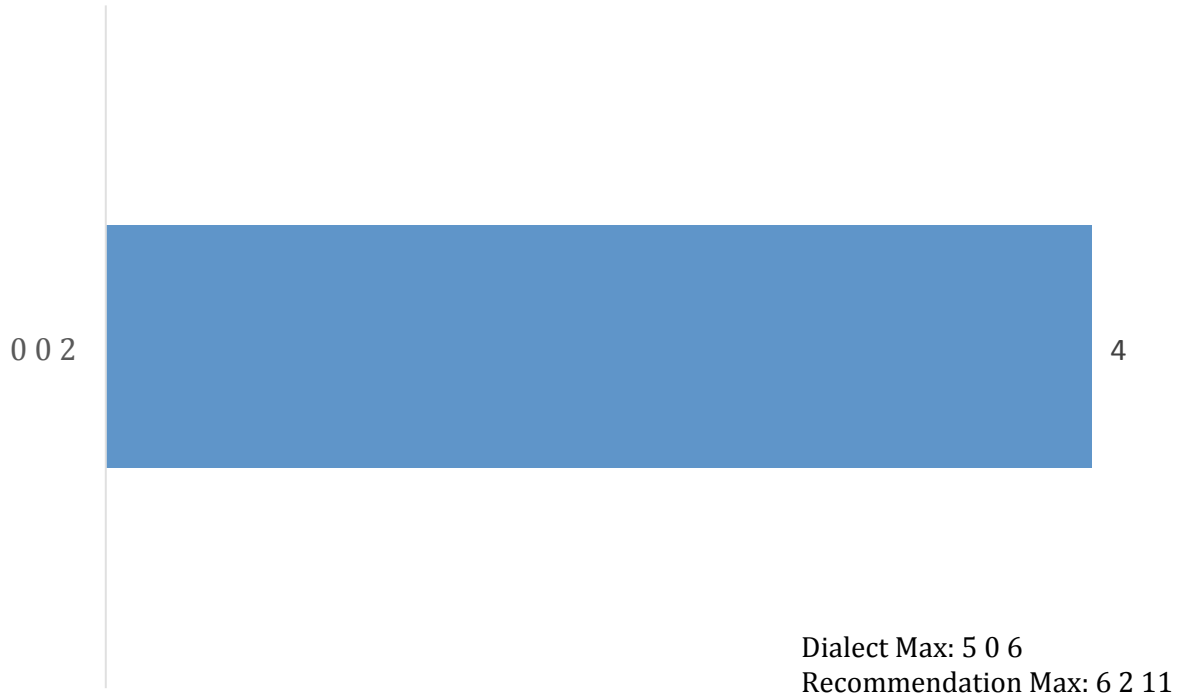
Unused Concepts

Concept	Description
Spatial Resolution	The nominal scale and/or spatial resolution of the resource.
Vertical Extent	The vertical extent of the resource

Specific Guidance – How to Improve the Metadata

The analysis above identifies specific concepts that are missing from NCAR metadata records, but are included in their respective dialects. This section provides specific guidance on how to write metadata for those concepts in a variety of dialects. A positive and straightforward first step is to assess what some NCAR records already include and implement them collection wide. The information is presented in three ways

- A table to describe the dialects usage at NCAR in relation to the recommendation.
- An incomplete concepts chart for each dialect
- Guidance links

The table below is comprised of rows for each ISO-1 recommendation concept and columns for each dialect. Cells are filled with a color or a percentage. The percentage is how many records in the sample set contain that concept. Green represents 100%. Yellow represents 0%, a concept that the dialect contains but is not contained in any record in the sample set for that dialect. Red represents a concept missing from the dialect. The table is intended to show not only how complete a dialect is for the ISO-1 recommendation, as well as how complete the records are with respect to the dialect maxima. An important use of the table is also determining if the currently used dialect is the best for the purposes of labs at NCAR.

We provide charts for each dialect showing how many records are missing a concept and what level the concept belongs to. These charts are intended to help identify the most important and most attainable goals to maximize results for each improvement iteration. If the concept is missing from 95% of records but only from the Optional level of the ISO-1 recommendation it should be lower priority than a Mandatory concept missing in only 50% of records. The Mandatory level is red, the Conditional level is green, and the Optional level is blue.
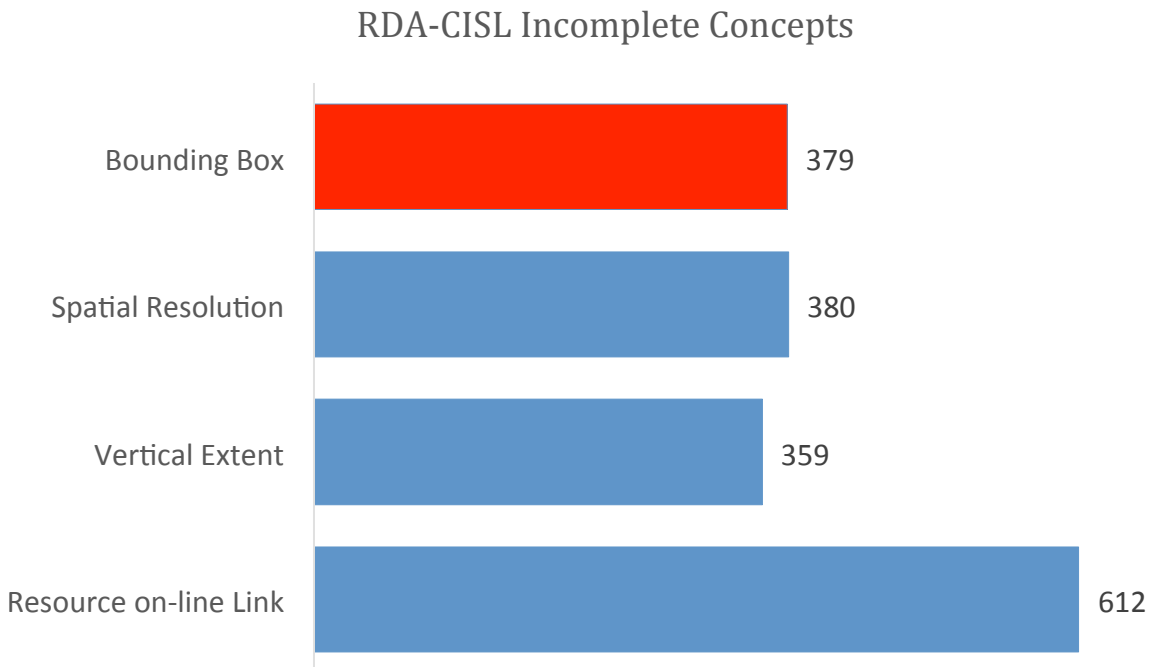
The guidance links resolve to pages on the Earth Science Information Partners wiki. These pages describe the concept as well as dialect specific XPaths to describe how the concept can be contained in a record. They are given for the concepts that are missing from some records and concepts contained in the dialect but unused in the collection. These links also contain xml samples of how the concept is shared using DIF, ECHO, ISO, and CSDGM dialects. Not every concept has a hyperlink because all of the guidance pages are not yet created.

NCAR Labs Usage of ISO-1 Data Discovery Recommendation Concepts (% complete)

	DCITE	ISO	MODS	RDA-CISL	netCDF
Total Number of Records	**8**	**541**	**1297**	**655**	**4**
Resource Title			96%		
Abstract		99%	85%		
Modified Date					
Metadata Contact					
Bounding Box				42%	
Topic Category		6%			
Resource Language	60%		85%		
Resource Type			96%		
Metadata Identifier			96%		
Theme Keyword			96%		
Resource Creation/Revision Date	60%	97%	5%		
Resource Identifier			96%		
Resource Contact					
Spatial Resolution				42%	
Temporal Extent					
Vertical Extent		0.3%		45%	
Resource Lineage		3%			
Resource on-line Link		3%		7%	
Metadata Use Constraints					

RDA-CISL Evaluation

A mapping of the native dialect xml representation of the Research Data Archive to the ISO-1 metadata concepts was created.
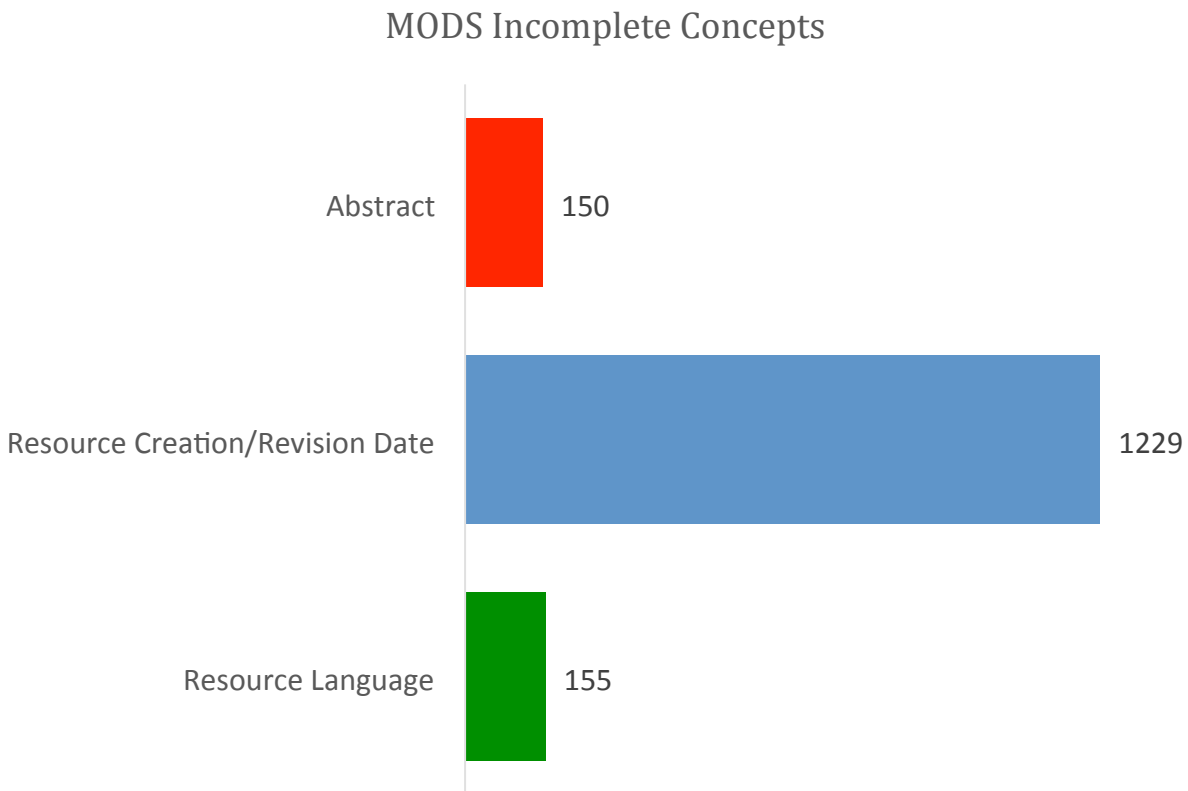
RDA-CISL Incomplete Concepts

Concept	Value
Bounding Box	379
Spatial Resolution	380
Vertical Extent	359
Resource on-line Link	612

Metadata Improvement

The concepts in the table below are not contained in every record (incomplete). All of the concepts listed below can be represented in Research Data Archive native dialect records. Click on the concept below to access online guidance for writing the concept in a variety of dialects.

Guidance Links

Incomplete	Bounding Box	Spatial Resolution	Vertical Extent	Resource on-line Link

MODS Evaluation

The MODS User Guidelines version 3 was used to map the dialect to the concepts found in the DataCite recommendation. The user guide comes from the Library of Congress and can be found here. There were 48 records that used a namespace from NSDL. These records were removed from the analysis. This is the largest collection analyzed.

MODS Incomplete Concepts

Concept	Value
Abstract	150
Resource Creation/Revision Date	1229
Resource Language	155

Metadata Improvement

The concepts in the table below are either not contained in every record (incomplete), or in any record (unused). All of the concepts listed below can be represented in MODS dialect records. Click on the concept below to access online guidance for writing the concept in a variety of dialects.
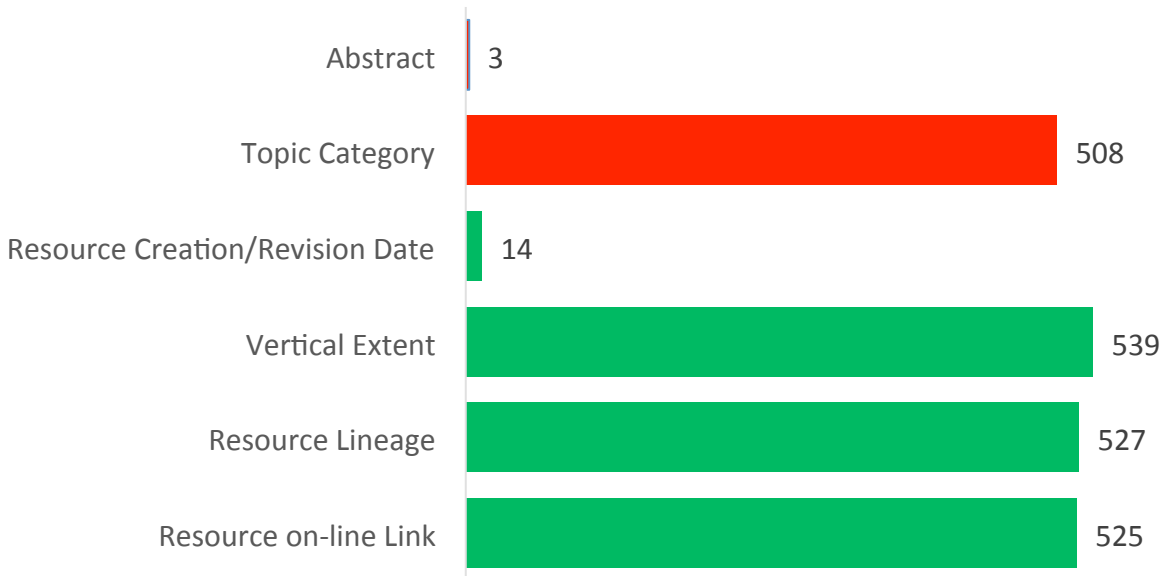
Guidance Links

Incomplete	Abstract	Resource Creation / Revision Date	Resource Language		
Unused	Modified Date	Bounding Box	Spatial Resolution	Temporal Extent	Metadata Identifier

ISO Evaluation
The ISO records came from collections at EOL and Unidata as well as the DCERC ISO sample set of EOL records.

ISO Incomplete Concepts

Concept	Value
Abstract	3
Topic Category	508
Resource Creation/Revision Date	14
Vertical Extent	539
Resource Lineage	527
Resource on-line Link	525

Metadata Improvement
The concepts in the table below are either not contained in every record (incomplete), or in any record (unused). All of the concepts listed below can be represented in ISO dialect records. Click on the concept below to access online guidance for writing the concept in a variety of dialects.

Guidance Links

Incomplete	Resource on-line Link	Vertical Extent	Resource Lineage	Resource Creation/Revision Date	Topic Category	Abstract
Unused	Spatial Resolution	Metadata Use Constraints				

DataCite Evaluation
The DataCite sample set is the most diverse, coming from MMM, HAO, UniData, and CISL.
While the sample size is quite small there are many differences as to what is included in a record.

DCITE Dialect Incomplete Concepts
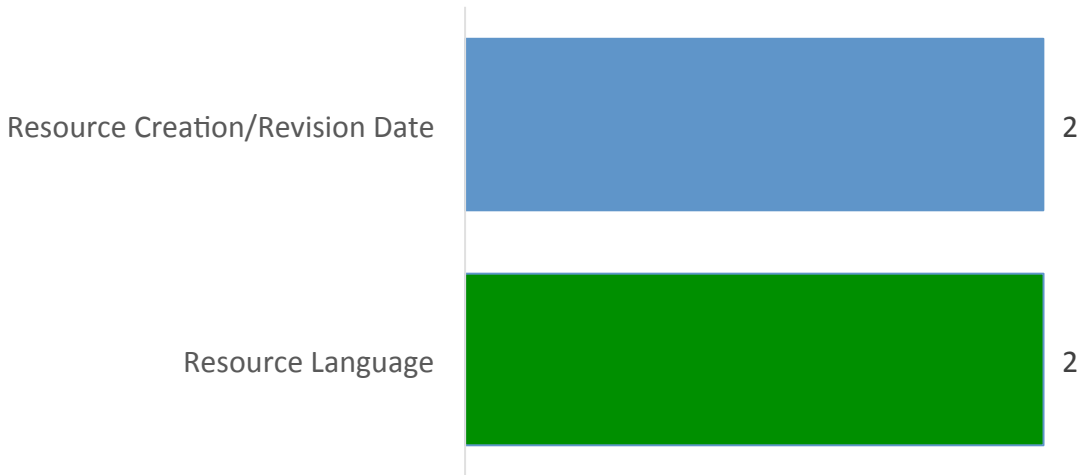
Metadata Improvement
The concepts in the table below are not contained in every record (missing). All of the concepts listed below can be represented in ISO dialect records. Click on the concept below to access online guidance for writing the concept in a variety of dialects.

Guidance Links

Incomplete	Resource Language	Resource Creation /Revision Date

netCDF Evaluation

The NcML files in the sample set were extracted from RAL NetCDF files. There are no incomplete concepts.

Metadata Improvement

The concepts in the table below are not contained in any record (unused) while they do exist in the NcML dialect. Click on the concept below to access online guidance for writing the concept in a variety of dialects.

Guidance Links

Unused	Spatial Resolution	Vertical Extent

Glossary

Collection: A group of metadata records commonly organized by a data facility, organization or project and often stored in a database or web accessible folder.

Concept: General term for describing a documentation entity. Concepts can occur in many dialects where they are typically represented (in XML) by an element.

Dialect: A particular form of the documentation language that is specific to a community.

Dialect Maximum: The maximum number of concepts from a particular recommendation that are included in a particular recommendation. Note: the dialect maximum is always less than or equal to the recommendation maximum.

Element: An item providing a value for a concept, typically in an XML representation. Elements depend on dialects. They are the instantiation of a concept in a dialect.

Level: Recommendations may have different degrees of necessity associated with a concept's occurrence in a record. These subsets of concepts within a recommendation are called levels.

Recommendation: A set of concepts that an organization identifies as valuable for achieving a documentation goal.

Recommendation Maximum: The number of concepts included in a particular recommendation. Note that the recommendation maximum is the maximum completeness score available for a metadata record being evaluated with respect to that recommendation. The recommendation maxima are always greater than or equal to all dialect maxima for that recommendation.

Signature: A series of numbers that give the number of concepts/elements missing from a metadata record (or a group of metadata records) in a series of levels. Signatures with low numbers indicate fewer missing elements and a signature made up completely of 0's indicates a record or group of records that is complete with respect to a particular recommendation/dialect combination. A signature of 2 3 indicates that 2 elements are missing from the first level and 3 are missing from the second. The sum of the numbers in a signature is the total number of elements missing from a record or group of records.