



ISO 19115-1 – Questions and Answers

The ISO Standards are evolving.

Metadata has been revised from 19115 to 19115-1

Data Quality metadata has moved from ISO 19115 to ISO 19157.

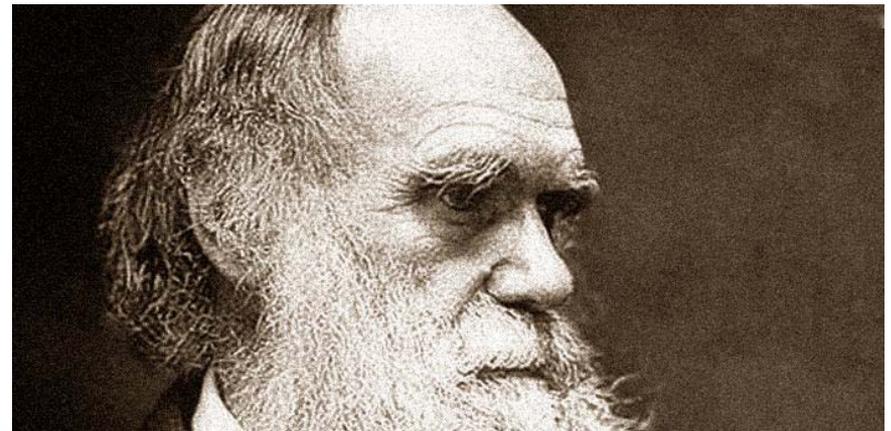
Ted Habermann

Director of Earth Science

The HDF Group

thabermann@hdfgroup.org

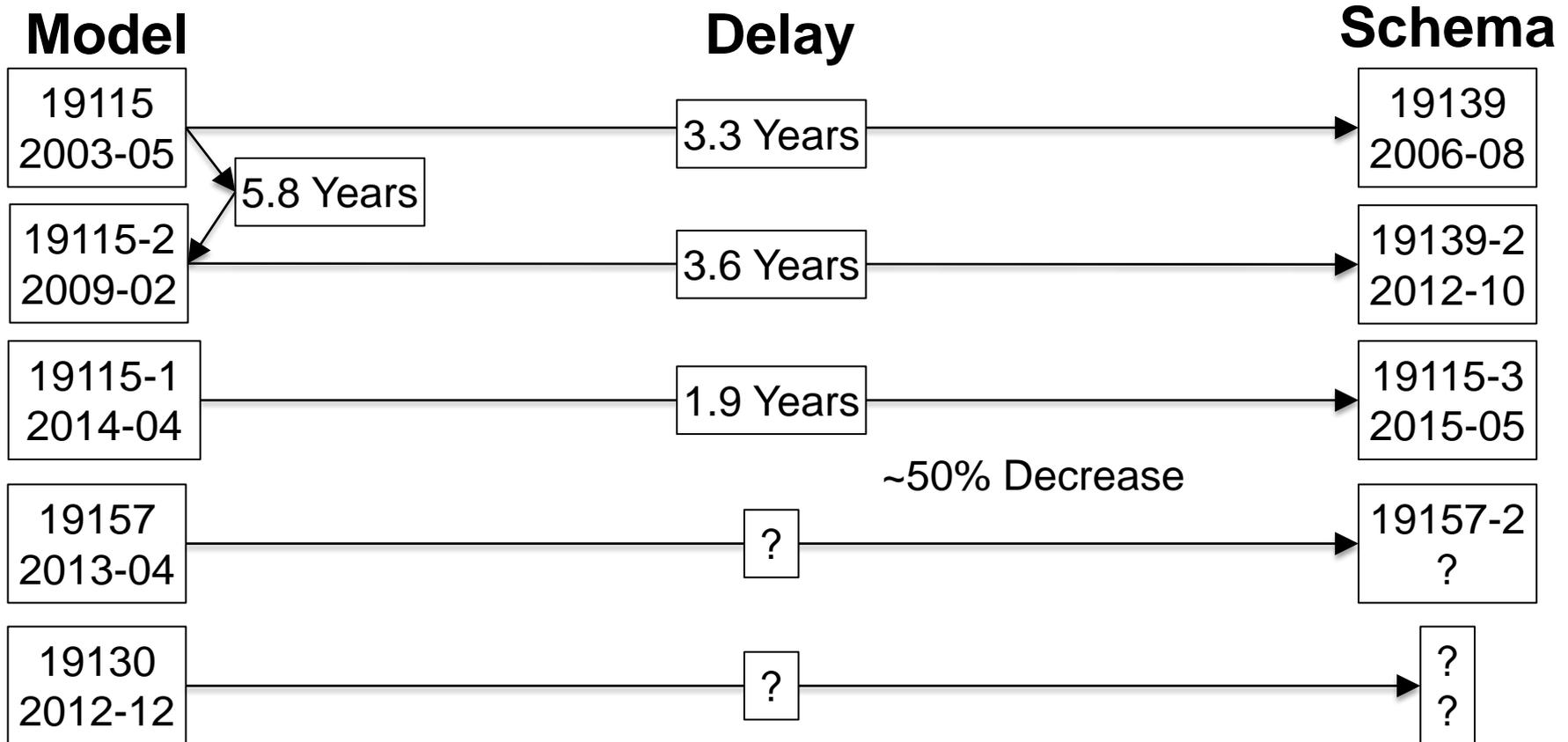
Evolution is good.





Conceptual Models and Implementations

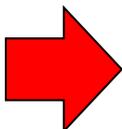
“Conceptual models are useful, but I need XML implementations to get going.”



Now creating XML schemas directly from UML models using software developed in OGC Testbed (ShapeChange). This capability is also being added directly into the tool by the vendor (Sparx Systems).

“I need to unambiguously identify metadata records in multiple repositories”

ISO 19115 identified metadata records using a single character string that often times had to be overloaded to include the information required for unambiguous identification.

05f314b8-bffe-cb8d-418e-744613aa4f01  nice opaque identifier, but who owns the metadata record or the identifier?

ISO 19115-1 brings the advantages of the MD_Identifier (+codeSpace) to the identification of the metadata record itself.

```
<mcc:code>
  <gco:CharacterString>05f314b8-bffe-cb8d-418e-744613aa4f01</gco:CharacterString>
</mcc:code>
<mcc:codeSpace>
  <gco:CharacterString>nz.govt.geodata</gco:CharacterString>
</mcc:codeSpace>
```



Tracking the Metadata Life Cycle

"I need to track when changes in my metadata happen"

ISO 19115 includes a dateStamp with the creation time for the metadata. Many other kinds of times are also important in the life-cycle of metadata.

ISO 19115-1 includes any number of CI_Date objects for the metadata which allows tracking of the metadata throughout its life-cycle. 19115-1 includes many more dateTypes than 19115.

```

<mdb:dateInfo>
  <cit:CI_Date>
    <cit:date>
      <gco:DateTime>2111-11-11T11:11:11</gco:DateTime>
    </cit:date>
    <cit:dateType>
      <cit:CI_DateTypeCode
        codeList="codeListLocation#CI_DateTypeCode"
        codeListValue="lastUpdate">lastUpdate
      </cit:CI_DateTypeCode>
    </cit:dateType>
  </cit:CI_Date>
</mdb:dateInfo>

```

19115
included 3
dateTypes

19115-1
adds 13
new
dateTypes

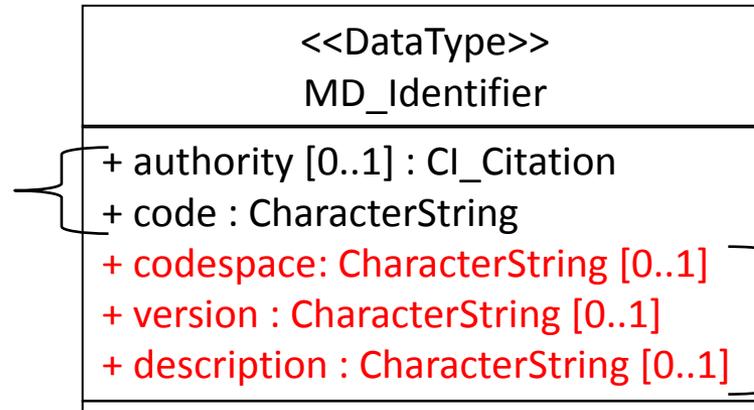
<<CodeList>> CI_DateTypeCode	
+ creation	+ inForce
+ publication	+ adopted
+ revision	+ deprecated
+ expiry	+ superseded
+ lastUpdate	+ validityBegins
+ lastRevision	+ validityExpires
+ nextUpdate	+ released
+ unavailable	+ distribution



Identifiers from Multiple Sources

“My metadata includes identifiers from different sources and namespaces”

19115 identifiers include a code and a citation to the authority of the code. Including a namespace for the identifier is not straightforward.



19115-1 adds a namespace (codespace), a version, and a description.

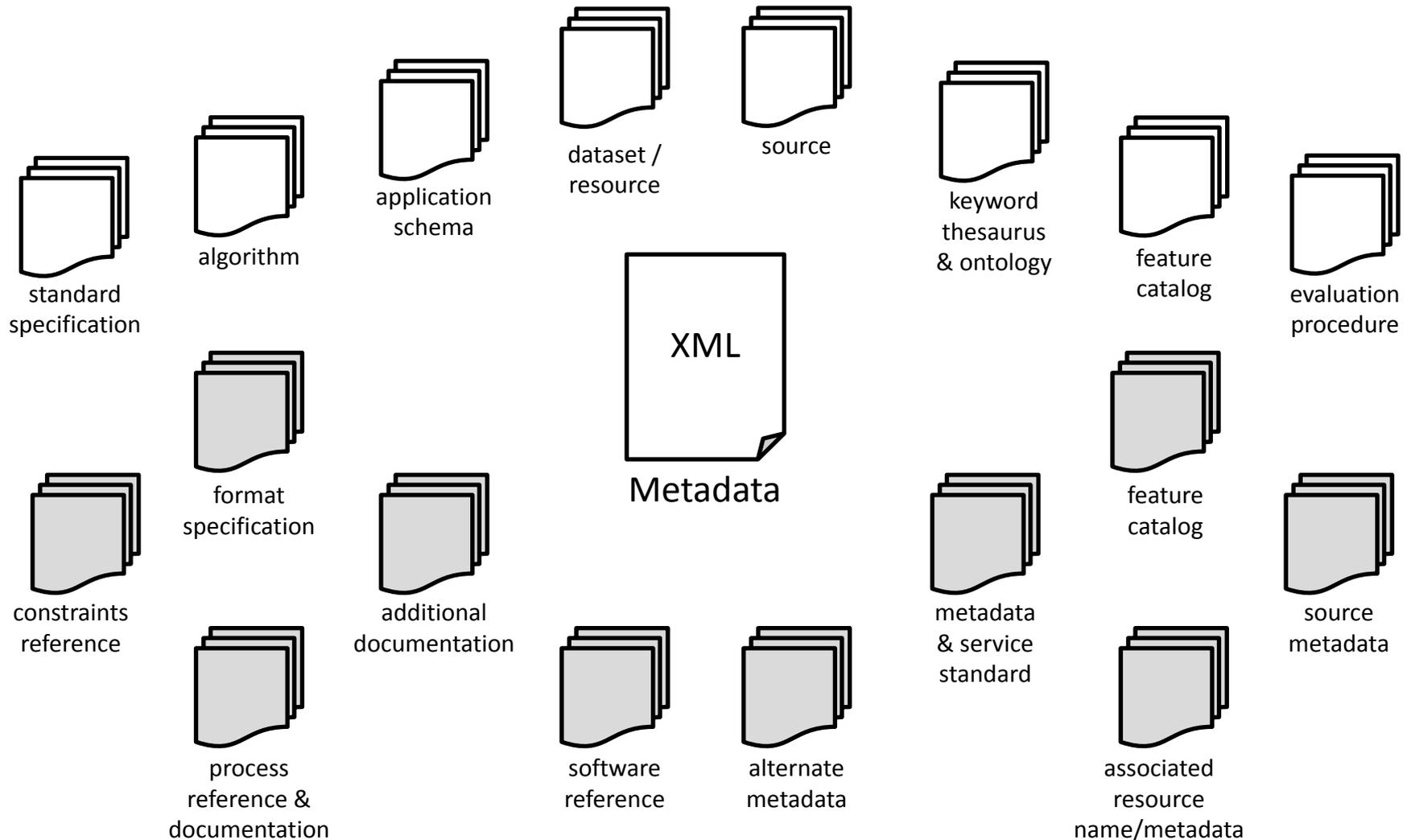
ISO 19115-1 added three important new elements to the MD_Identifier:

1. codespace provides a namespace for the identifier authority
2. version provides a mechanism for including a versioning identifiers
3. description provides a free-text field that can help users understand the identifier.



Connecting Other Documentation

"I have many existing documentation resources that can help users"





Stand Alone Quality Reports

“There are papers and web pages that describe the quality of my data.”

Papers and reports that describe data quality are StandAloneReports. Metadata can include brief descriptions of the results (abstracts) and references to any number of these (citations).

Abstract: The fire training-set may also have been biased against savanna and savanna woodland fires since their detection is more difficult than in humid, forest environments with cool background temperatures [Malingreau, 1990]. There may, therefore, be an under-sampling of fires in these warmer background environments.

Citation: Malingreau J.P, 1990, The contribution of remote sensing to the global monitoring of fires in tropical and subtropical ecosystems. In: *Fire in Tropical Biota*, (J.G. Goldammer , editor), Springer Verlag , Berlin: 337-370.

DQ_StandaloneQualityReportInformation

+ reportReference: CI_Citation
+ abstract : CharacterString



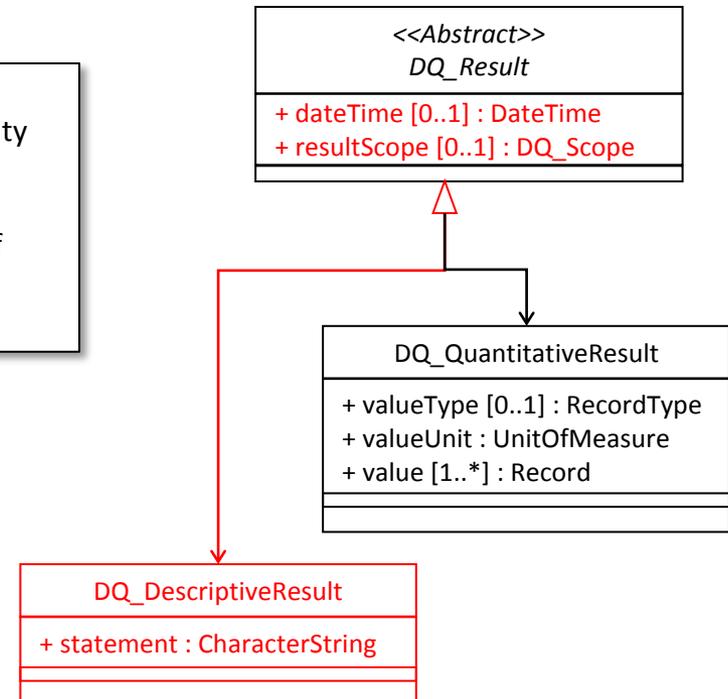
Textual Quality Descriptions

“My metadata includes textual descriptions of quality.”

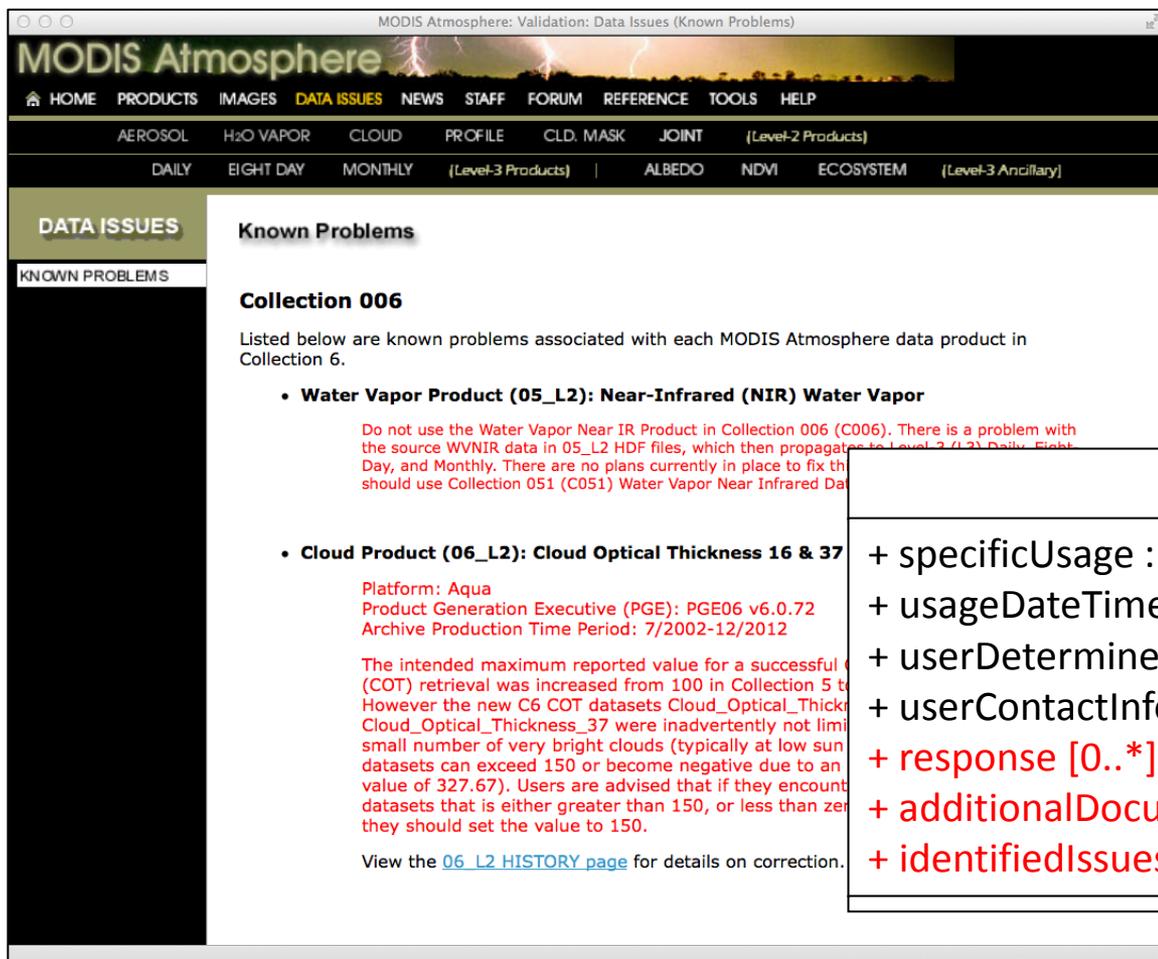
<Quality>
QA performed by CDIAC One of the roles of the Carbon Dioxide Information Analysis Center (CDIAC) is quality assurance (QA) of data. The QA process is an important component of the value-added concept of assuring accurate, usable information for researchers, because data received by CDIAC are rarely in condition for immediate

<Quality>
Note that Data File 12, Report #2, TASK 2 (Auclair et al., 1994a) is a Quality Assurance and Quality Control chapter for the areas of Canada, Alaska, United States (48 states), with range estimates of validation and error, a listing of discussions with experts in the field and a review of the draft of data files.
</Quality>

ISO 19157 adds a resultScope that allows multiple scopes in a single DQ_DataQuality object and includes a new kind of report (DQ_DescriptiveResult) that includes a simple text description of the result of the quality test.



“Users increase our understanding of data quality. We need to keep them in the loop.”



The screenshot shows the 'MODIS Atmosphere' website with a navigation menu including HOME, PRODUCTS, IMAGES, DATA ISSUES, NEWS, STAFF, FORUM, REFERENCE, TOOLS, and HELP. Under 'DATA ISSUES', there are sub-menus for AEROSOL, H₂O VAPOR, CLOUD, PROFILE, CLD. MASK, JOINT, and (Level-2 Products). Below these are DAILY, EIGHT DAY, MONTHLY, (Level-3 Products), ALBEDO, NDVI, ECOSYSTEM, and (Level-3 Ancillary). The main content area is titled 'DATA ISSUES' and 'Known Problems'. It lists 'Collection 006' and states: 'Listed below are known problems associated with each MODIS Atmosphere data product in Collection 6.'

- Water Vapor Product (05_L2): Near-Infrared (NIR) Water Vapor**
 Do not use the Water Vapor Near IR Product in Collection 006 (C006). There is a problem with the source WVNIR data in 05_L2 HDF files, which then propagates to Level-2 (L2) Daily, Eight Day, and Monthly. There are no plans currently in place to fix this problem. Users should use Collection 051 (C051) Water Vapor Near Infrared Data.
- Cloud Product (06_L2): Cloud Optical Thickness 16 & 37**
 Platform: Aqua
 Product Generation Executive (PGE): PGE06 v6.0.72
 Archive Production Time Period: 7/2002-12/2012
 The intended maximum reported value for a successful Cloud Optical Thickness (COT) retrieval was increased from 100 in Collection 5 to 327.67 in Collection 6. However the new C6 COT datasets Cloud_Optical_Thickness_16 and Cloud_Optical_Thickness_37 were inadvertently not limited to 327.67. A small number of very bright clouds (typically at low sun elevations) in some datasets can exceed 150 or become negative due to an error in the retrieval algorithm (value of 327.67). Users are advised that if they encounter datasets that is either greater than 150, or less than zero, they should set the value to 150.
 View the [06_L2_HISTORY](#) page for details on correction.

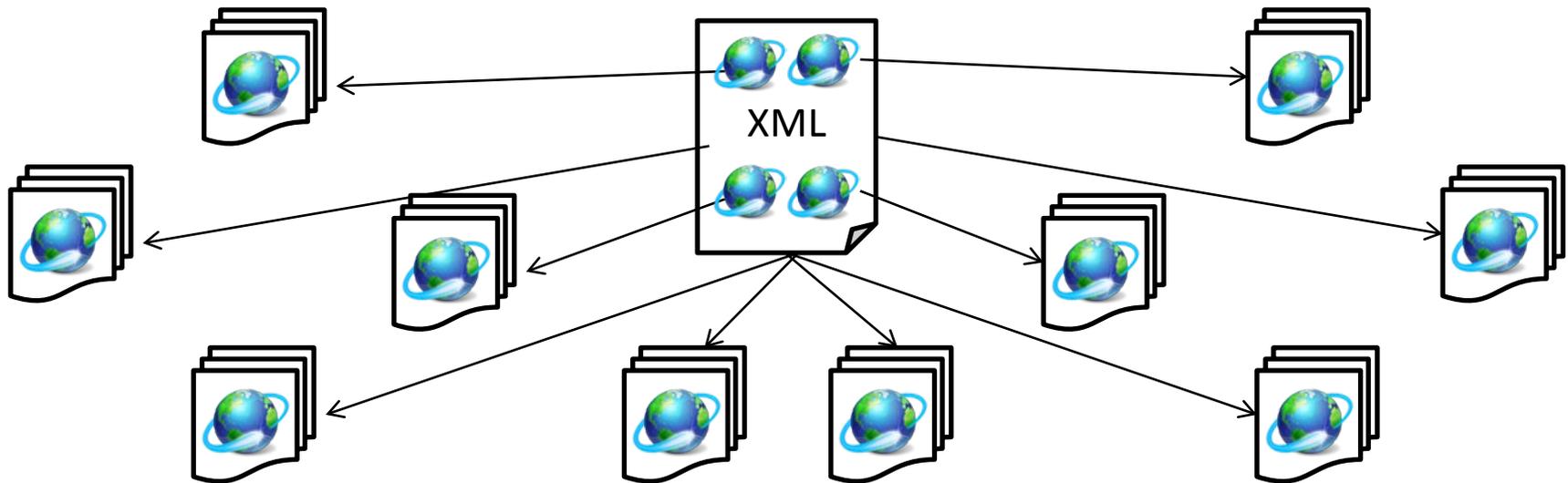
MD_Usage
+ specificUsage : CharacterString + usageDateTime [0..1] : DateTime + userDeterminedLimitations [0..1] : CharacterString + userContactInfo [1..*] : CI_ResponsibleParty + response [0..*] : CharacterString + additionalDocumentation [0..*] : CI_Citation + identifiedIssues [0..1] : CI_Citation

"I have many existing web resources that can help users"

The ISO 19115 CI_Citation worked well for citing books, journal articles, and other physical resources.

ISO 19115-1 added two important new elements to the CI_Citation:

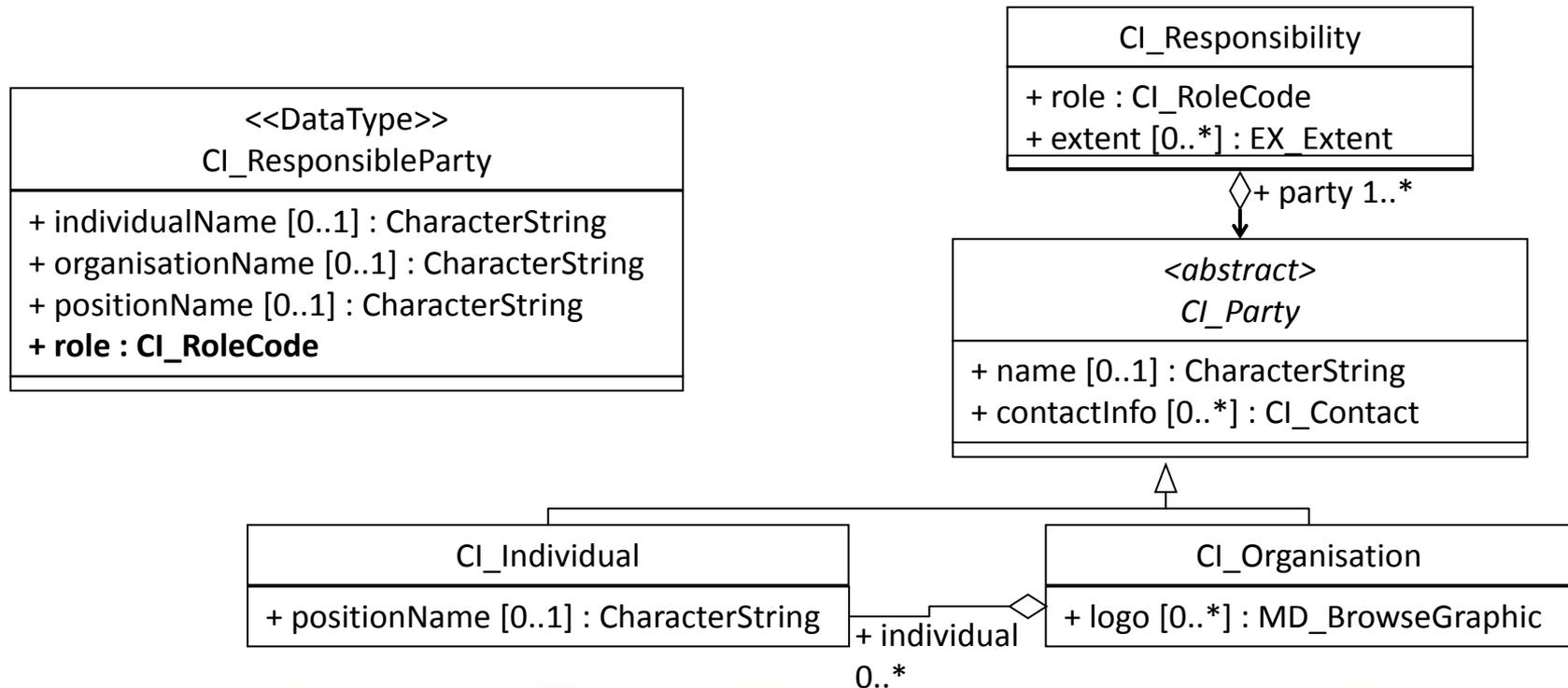
1. onlineResource provides a web address for the cited resource
2. graphic provides a graphic that can be used for display of the cited resource.



“I need consistent information about people and organizations.”

The ISO 19115 CI_ResponsibleParty object included a codeList for roles that people and organizations played. This made it difficult to reuse information in multiple records.

ISO 19115-1 separated the role codeList so that people and organizations can be re-used.





Individuals, Organizations, and Roles

"I have people and organizations in many roles."

19115 included
11 standard
CI_RoleCodes

<<CodeList>> CI_RoleCode	
+ resourceProvider	+ sponsor
+ custodian	+ coAuthor
+ owner	+ collaborator
+ user	+ editor
+ distributor	+ mediator
+ originator	+ rightsHolder
+ pointOfContact	+ contributor
+ principalInvestigator	+ funder
+ processor	+ stakeholder
+ publisher	
+ author	

19115-1 adds
9 new
CI_RoleCodes

“I use a Creative Commons License for my data”

ISO 19115 included limited descriptions of constraints related to the data (useLimitations) or imposed by organizations (legal and security constraints). This made it difficult to describe commonly used open source licenses.

The Marine Community Profile extended 19115 to include Creative Commons License.

ISO 19115-1 included the necessary elements.

MD_Commons
+ useLimitation [0..*] : CharacterString
+ jurisdictionLink : URL
+ licenseLink : URL
+ imageLink : URL
+ licenseName : CharacterString
+ attributionConstraints : CharacterString

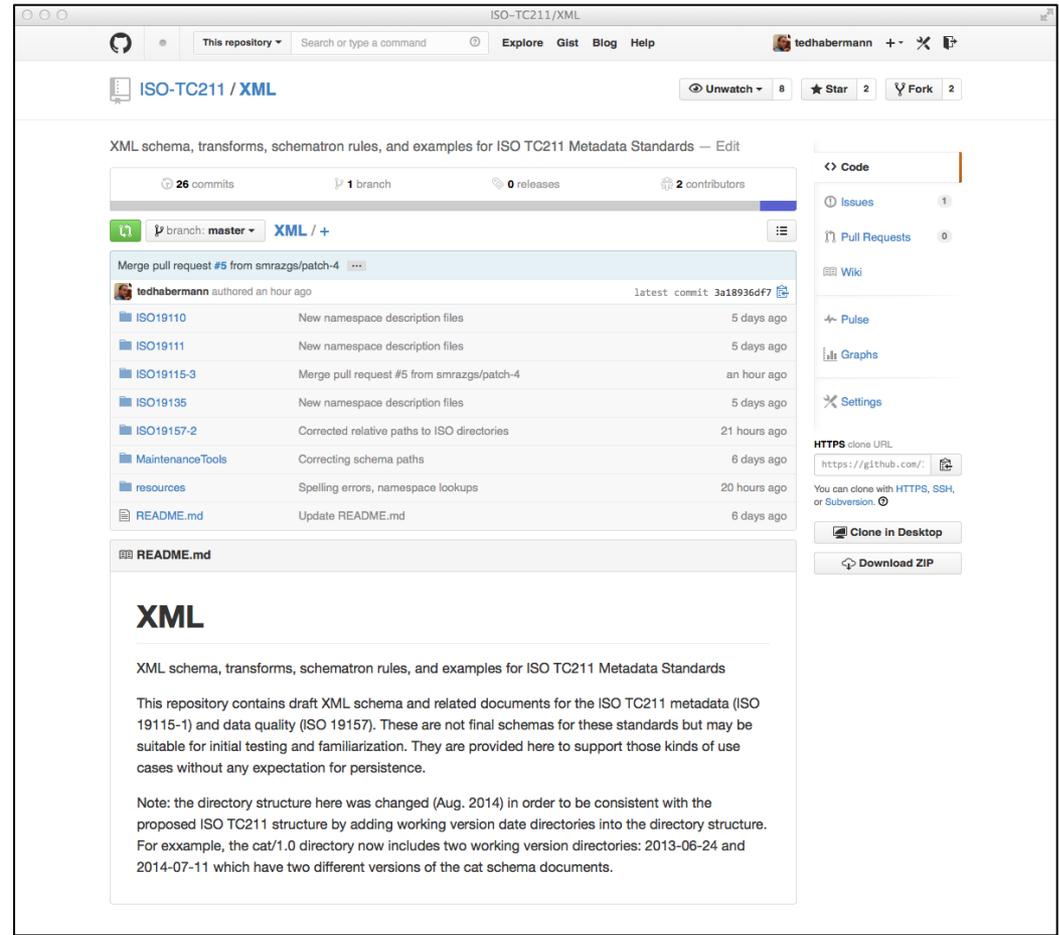
MD_Constraints
+ useLimitation [0..*] : CharacterString
+ constraintApplicationScope [0..1]: MD_Scope
+ graphic [0..*] : MD_BrowseGraphic
+ reference [0..*] : CI_Citation
+ MD_Releasability [0..1] : MD_Releasability
+ responsibleParty [0..*] : CI_Responsibility

jurisdictionLink : URL -----> responsibleParty
 licenseLink : URL -----> reference
 imageLink : URL -----> graphic
 licenseName : CharacterString -----> reference
 attributionConstraints : CharacterString -----> reference

“How do I become familiar with and help test the new implementations?”

The XML schemas and other resources and RDF/OWL implementations are available at:

<https://github.com/ISO-TC211>



ISO-TC211 / XML

XML schema, transforms, schematron rules, and examples for ISO TC211 Metadata Standards — Edit

26 commits 1 branch 0 releases 2 contributors

branch: master XML / +

Merge pull request #5 from smrazgs/patch-4

tedhabermann authored an hour ago latest commit 3a18936df7

Commit	Description	Time
ISO19110	New namespace description files	5 days ago
ISO19111	New namespace description files	5 days ago
ISO19115-3	Merge pull request #5 from smrazgs/patch-4	an hour ago
ISO19135	New namespace description files	5 days ago
ISO19157-2	Corrected relative paths to ISO directories	21 hours ago
MaintenanceTools	Correcting schema paths	6 days ago
resources	Spelling errors, namespace lookups	20 hours ago
README.md	Update README.md	6 days ago

README.md

XML

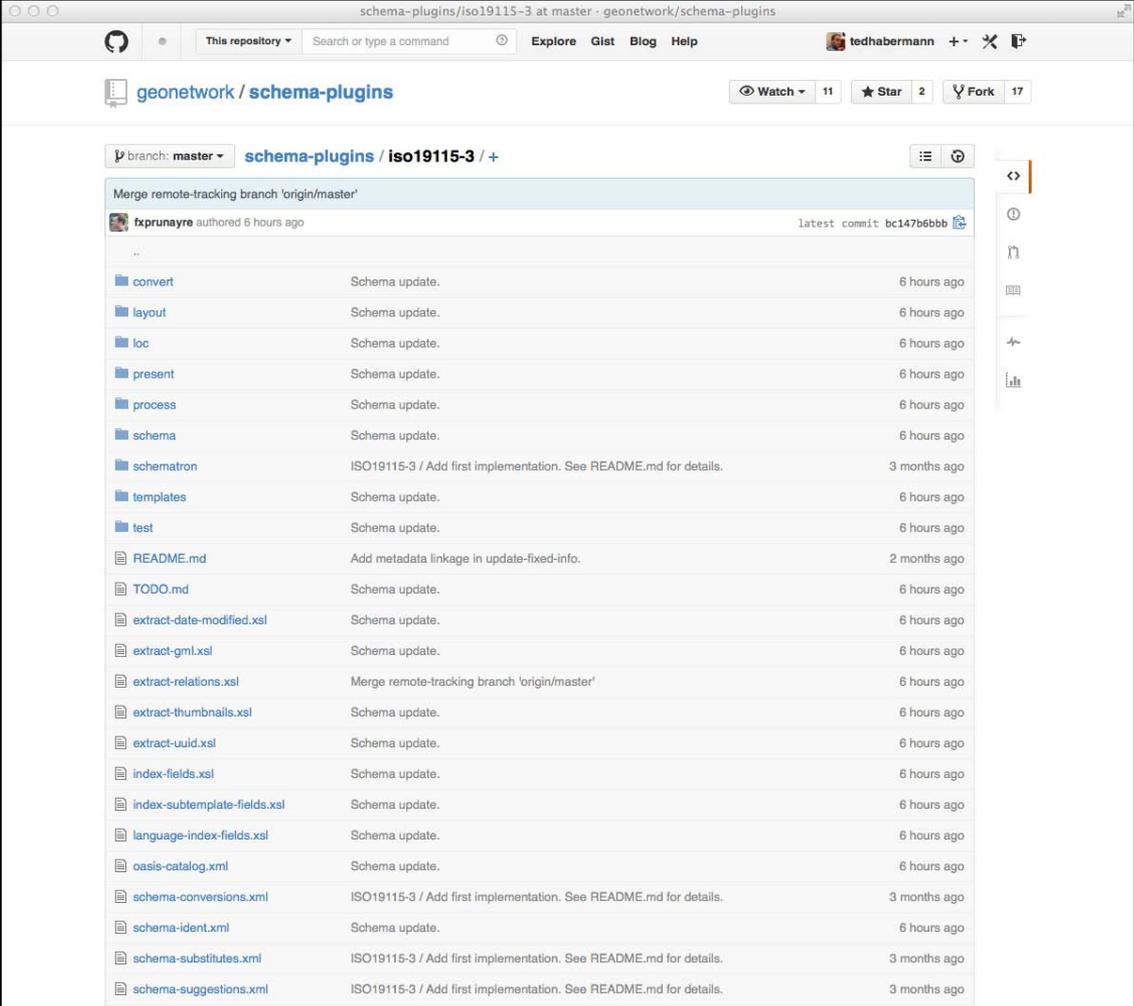
XML schema, transforms, schematron rules, and examples for ISO TC211 Metadata Standards

This repository contains draft XML schema and related documents for the ISO TC211 metadata (ISO 19115-1) and data quality (ISO 19157). These are not final schemas for these standards but may be suitable for initial testing and familiarization. They are provided here to support those kinds of use cases without any expectation for persistence.

Note: the directory structure here was changed (Aug. 2014) in order to be consistent with the proposed ISO TC211 structure by adding working version date directories into the directory structure. For example, the cat/1.0 directory now includes two working version directories: 2013-06-24 and 2014-07-11 which have two different versions of the cat schema documents.

"I use GeoNetwork to manage metadata. How can I try 19115-1?"

GeoNetwork 19115-3 resources are available at:
<https://github.com/geonetwork/schema-plugins/tree/master/iso19115-3#iso-19115-3-schema-plugin>



The screenshot shows the GitHub repository page for `geonetwork/schema-plugins`. The current branch is `master`, and the selected branch is `iso19115-3`. The page displays a list of files and folders, including:

- `convert`: Schema update. 6 hours ago
- `layout`: Schema update. 6 hours ago
- `loc`: Schema update. 6 hours ago
- `present`: Schema update. 6 hours ago
- `process`: Schema update. 6 hours ago
- `schema`: Schema update. 6 hours ago
- `schematron`: ISO19115-3 / Add first implementation. See README.md for details. 3 months ago
- `templates`: Schema update. 6 hours ago
- `test`: Schema update. 6 hours ago
- `README.md`: Add metadata linkage in update-fixed-info. 2 months ago
- `TODO.md`: Schema update. 6 hours ago
- `extract-date-modified.xsl`: Schema update. 6 hours ago
- `extract-gml.xsl`: Schema update. 6 hours ago
- `extract-relations.xsl`: Merge remote-tracking branch 'origin/master'. 6 hours ago
- `extract-thumbnails.xsl`: Schema update. 6 hours ago
- `extract-uuid.xsl`: Schema update. 6 hours ago
- `index-fields.xsl`: Schema update. 6 hours ago
- `index-subtemplate-fields.xsl`: Schema update. 6 hours ago
- `language-index-fields.xsl`: Schema update. 6 hours ago
- `oasis-catalog.xml`: Schema update. 6 hours ago
- `schema-conversions.xml`: ISO19115-3 / Add first implementation. See README.md for details. 3 months ago
- `schema-ident.xml`: Schema update. 6 hours ago
- `schema-substitutes.xml`: ISO19115-3 / Add first implementation. See README.md for details. 3 months ago
- `schema-suggestions.xml`: ISO19115-3 / Add first implementation. See README.md for details. 3 months ago

“Can I migrate my existing metadata to 19115-1?”

The screenshot shows an XML Editor window with two XML files open. The left file is 'Just_MI_MetadataPaths.xml' and the right file is 'Just_MI_MetadataPaths.xml'. A red box highlights the text 'Transform Available for Testing' overlaid on the XML content. The XML content includes various metadata elements such as <gco:CharacterString>, <gmd:language>, <gmd:characterSet>, <gmd:MD_CharacterSetCode>, <gmd:contact>, <gmd:dateStamps>, <gmd:spatialRepresentationInfo>, <gmi:MI_Georectified>, <gmd:cellGeometry>, <gmd:transformationParameterAvailability>, <gmd:checkPointAvailability>, <gmd:cornerPoints>, <gml:Point>, <gml:pos>, <gmd:pointInPixel>, <gmd:MD_PixelOrientationCode>, <gmd:pointInPixel>, <gmi:checkPoint>, <gmi:MI_GCP>, <gml:geographicCoordinates>, <gmd:MI_GCP>, <gmi:checkPoint>, <gmi:MI_Georectified>, <gmd:spatialRepresentationInfo>, <gmd:spatialRepresentationInfo>, <gmd:identificationInfo>, <gmd:contentInfo>, <gmd:attributeDescription>, <gmd:RecordType>, <gmd:attributeDescription>, <gmd:contentType>, <gmd:MD_CoverageContentTypeCode>, <gmd:dimension>, <gmd:MI_Band>, <gmd:sequenceIdentifier>, <mdb:metadataIdentifier>, <mc:MD_Identifier>, <gco:CharacterString>, <mc:code>, <gco:CharacterString>, <mc:codeSpace>, <gco:CharacterString>, <mc:codeSpace>, <mdb:metadataIdentifier>, <mdb:defaultLocale>, <lan:PT_Locale>, <lan:language>, <lan:LanguageCode>, <lan:Language>, <lan:characterEncoding>, <lan:MD_CharacterSetCode>, <lan:characterEncoding>, <lan:PT_Locale>, <mdb:defaultLocale>, <mdb:contact>, <mdb:dateInfo>, <cit:CI_Date>, <cit:date>, <gco:DateTime>, <cit:dates>, <cit:dateType>, <cit:CI_DateTypeCode>, <cit:dateType>, <cit:CI_Date>, <mdb:dateInfo>, <msr:spatialRepresentationInfo>, <msr:MI_Georectified>, <msr:numberOfDimensions>, <msr:cellGeometry>, <msr:transformationParameterAvailability>, <msr:checkPointAvailability>, <msr:cornerPoints>, <msr:pointInPixel>, <msr:MD_PixelOrientationCode>, <msr:pointInPixel>.

<https://github.com/ISO-TC211/XML/blob/master/resources/transforms/19115to19115-1.xsl>

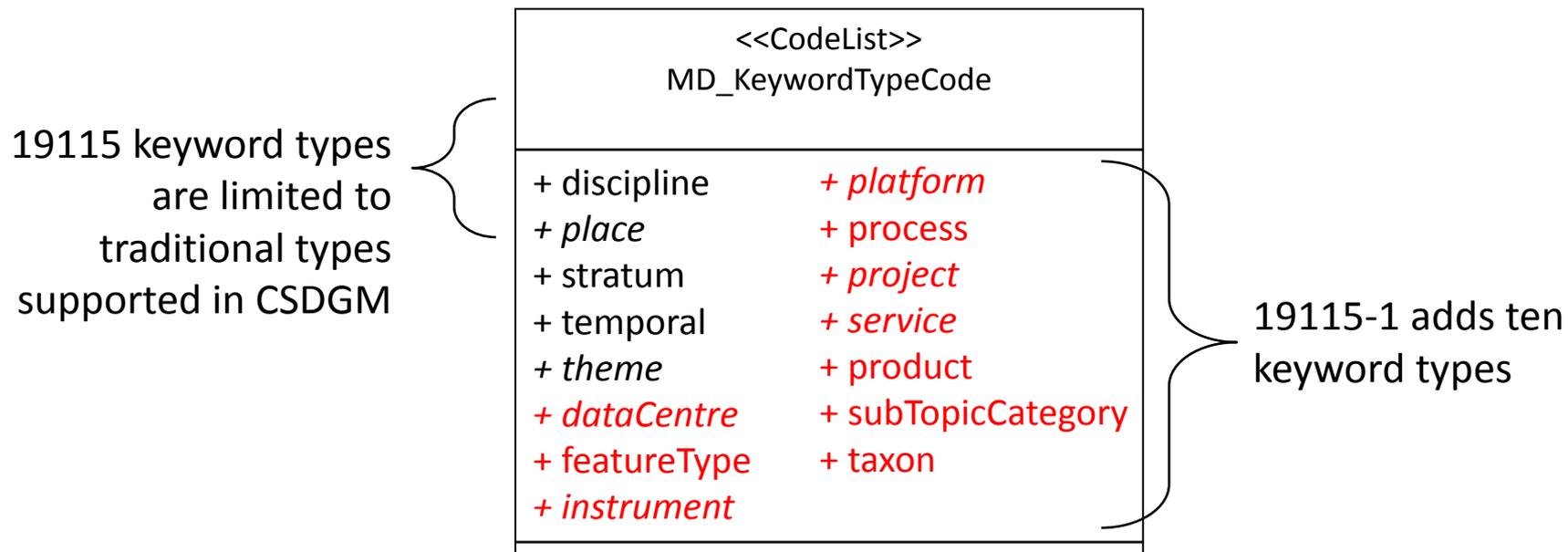


Questions?

tedhabermann@hdfgroup.org

Keyword Types

"I use many shared vocabularies for consistency across collections."



Keywords are the largest single component of many metadata collections, regardless of the dialect. Using common types to classify these keywords is critical for consistent discovery, particularly using faceted searches. Shared vocabularies that include these new types are important contributors to consistent, interoperable metadata. The keyword types in *italics* are supported by NASA Global Change Master Directory (GCMD) and used in many existing metadata collections.



Connecting to Collections

“My datasets are parts of larger collections”

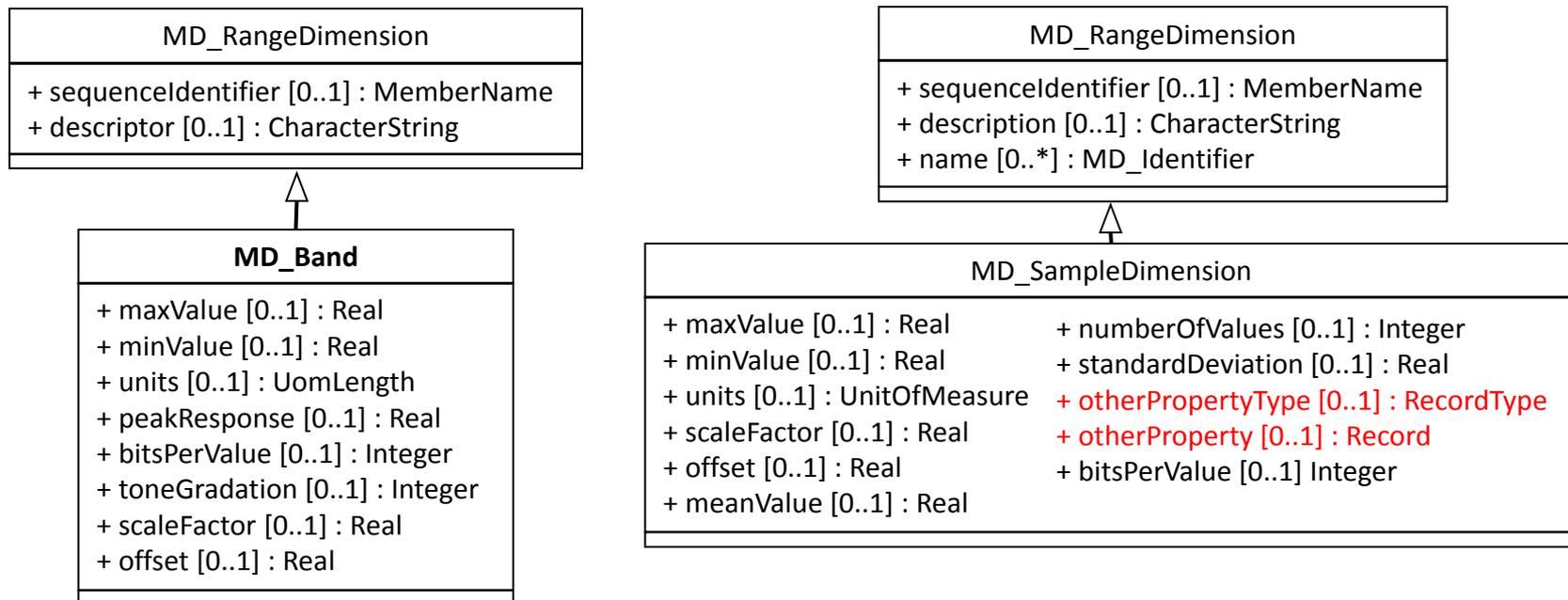
MD_Metadata
...
+ parentIdentifier [0..1] : CharacterString
...

MD_Metadata
...
+ parentMetadata [0..1] : CI_Citation
...

The concept of parent/child relationships between metadata for collections and for items in collections has been supported in many metadata dialects. ISO 19115 included a CharacterString as a parentIdentifier.

ISO 19115-1 brings the advantages of a complete CI_Citation for specification of the parent metadata. That CI_Citation includes any number of MD_Identifier that provide unambiguous identification of the parent metadata.

“My products need specific metadata that are not included in the general model”

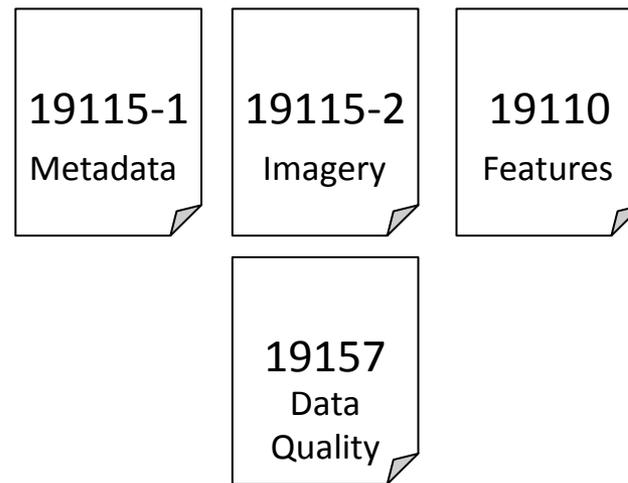
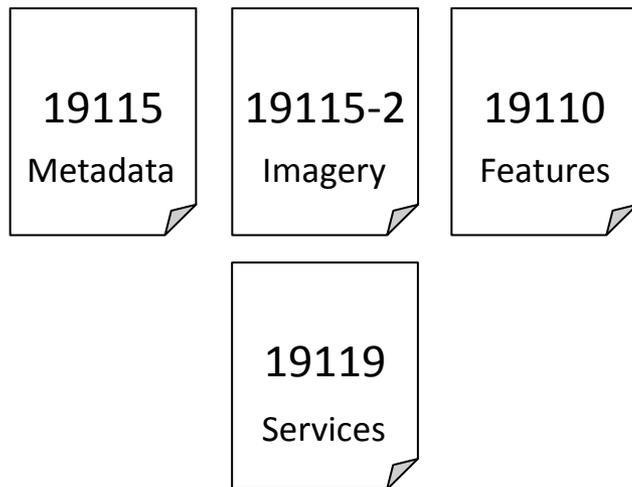


The ISO 19115 MD_Band object included a fixed set of image-specific properties.

ISO 19115-1 introduces the MD_SampleDimension which includes the capability to add product specific attributes using standard ISO objects.

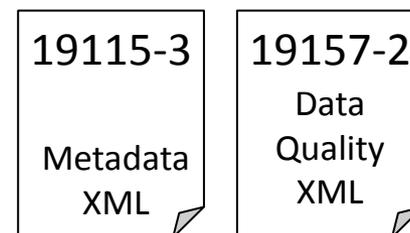
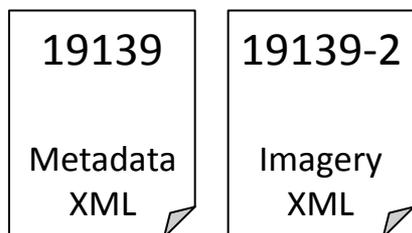
"I am confused by all of these numbers!"

Conceptual Models (UML)



Then Now

Implementations (XML)

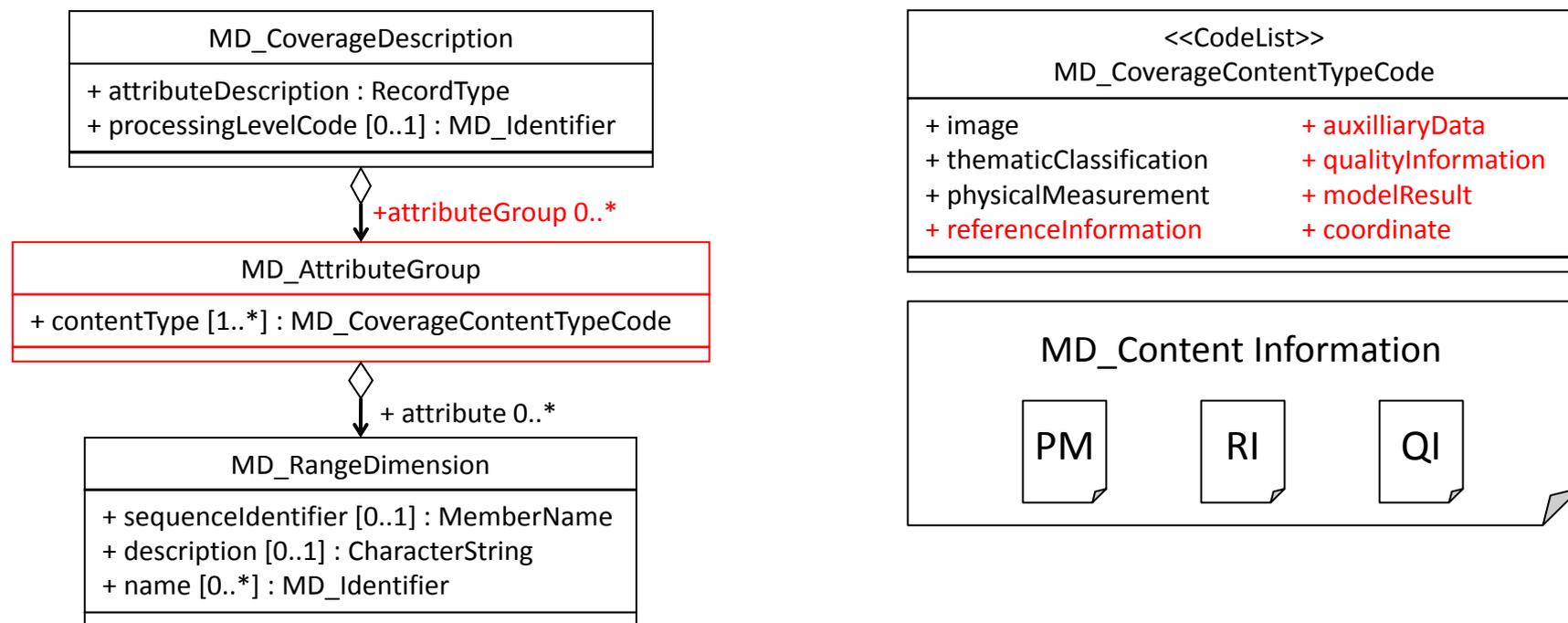


HDF Organizing Parameters

“My datasets include measured parameters, reference and quality information”

The ISO 19115 allowed only one type of information in each contentInfo section.

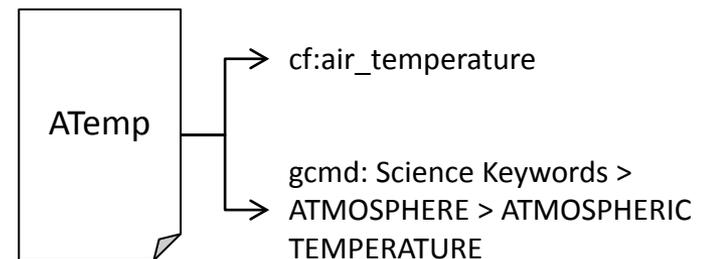
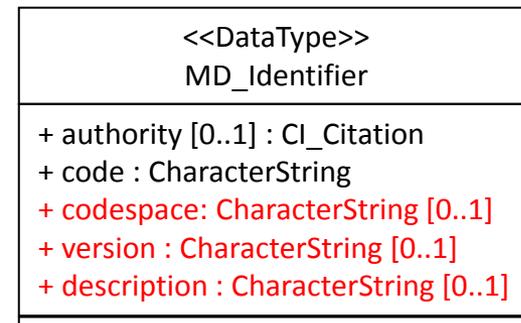
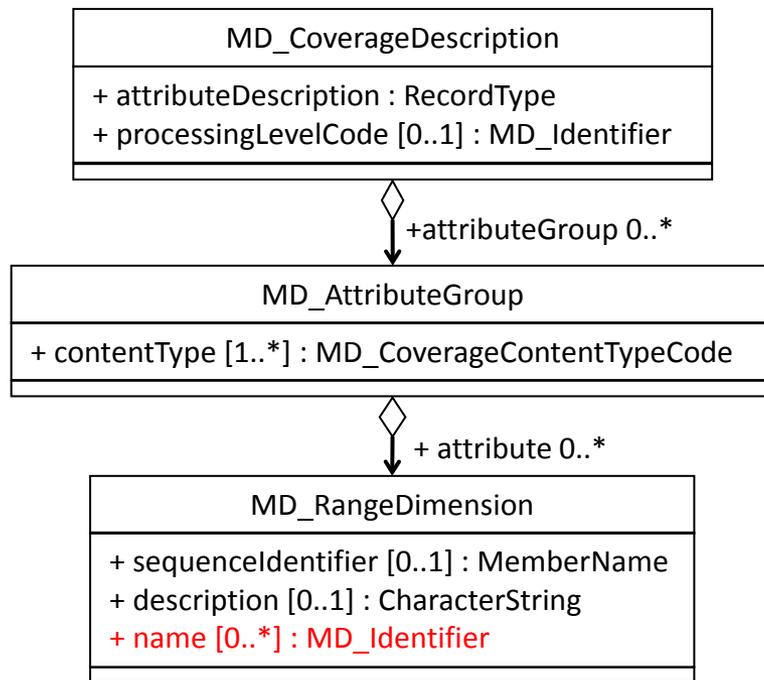
ISO 19115-1 adds the capability to group similar coverages and introduces more coverage types.



“My group uses local parameter names but we need standard names to share”

The ISO 19115 sequenceIdentifier only allowed one local name for parameters

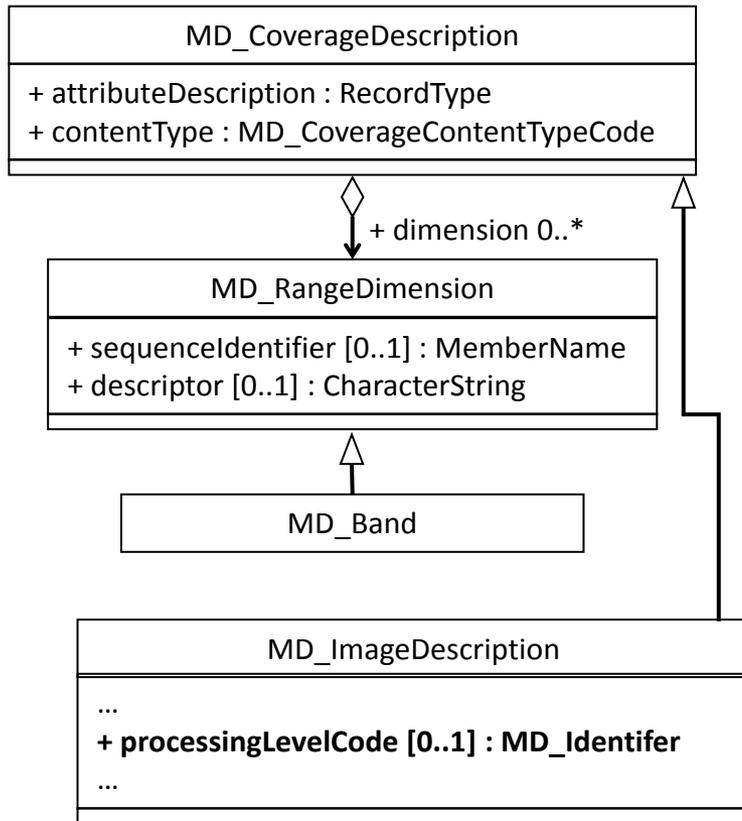
ISO 19115-1 adds the capability to add multiple names for parameters and to identify the sources for those names.





Describe Coverage Processing Levels

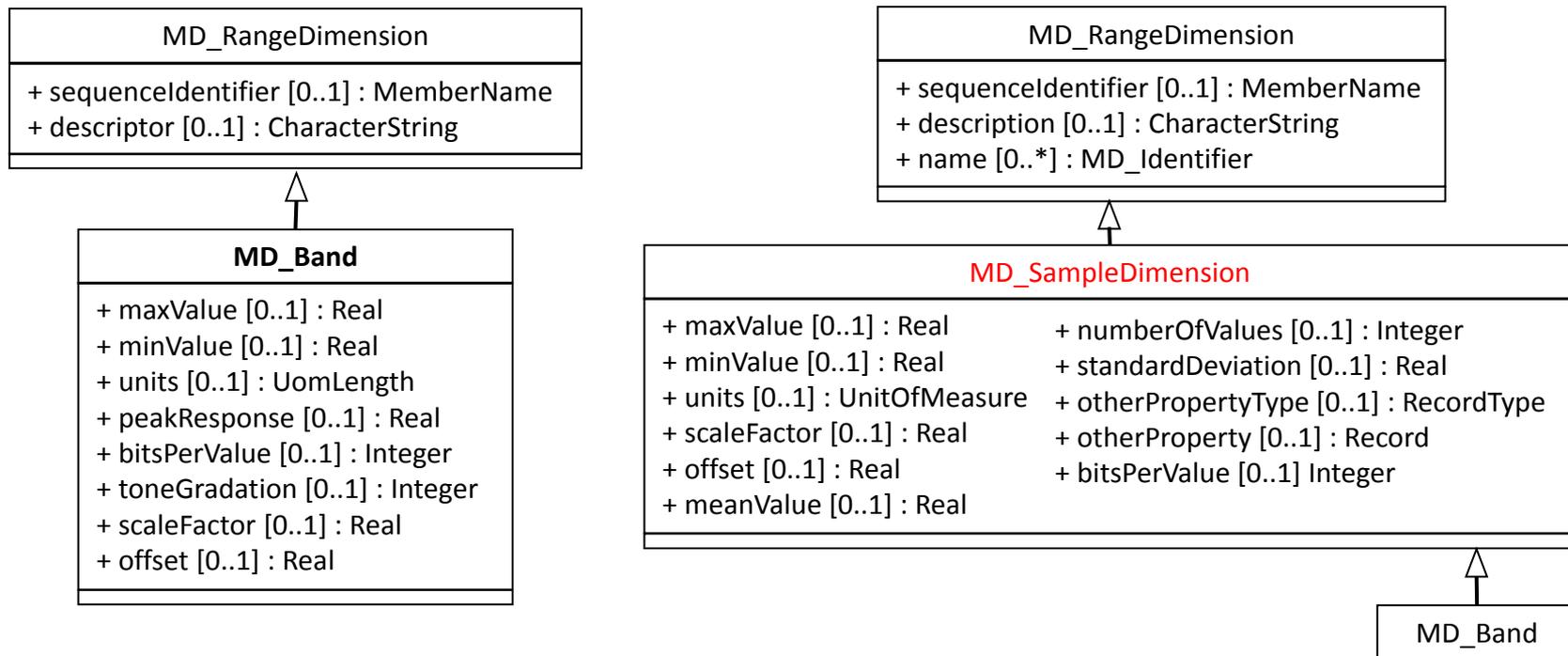
“My data have multiple bands and processing levels”



ISO 19115 includes a processingLevelCode as part of the MD_ImageDescription object. This means that coverages that use MD_RangeDimension (or MD_Band) objects can not use the processingInformation codes.

ISO 19115-1 moves the processingLevelCode into the MD_CoverageDescription object allowing it to be used for either type of coverage.

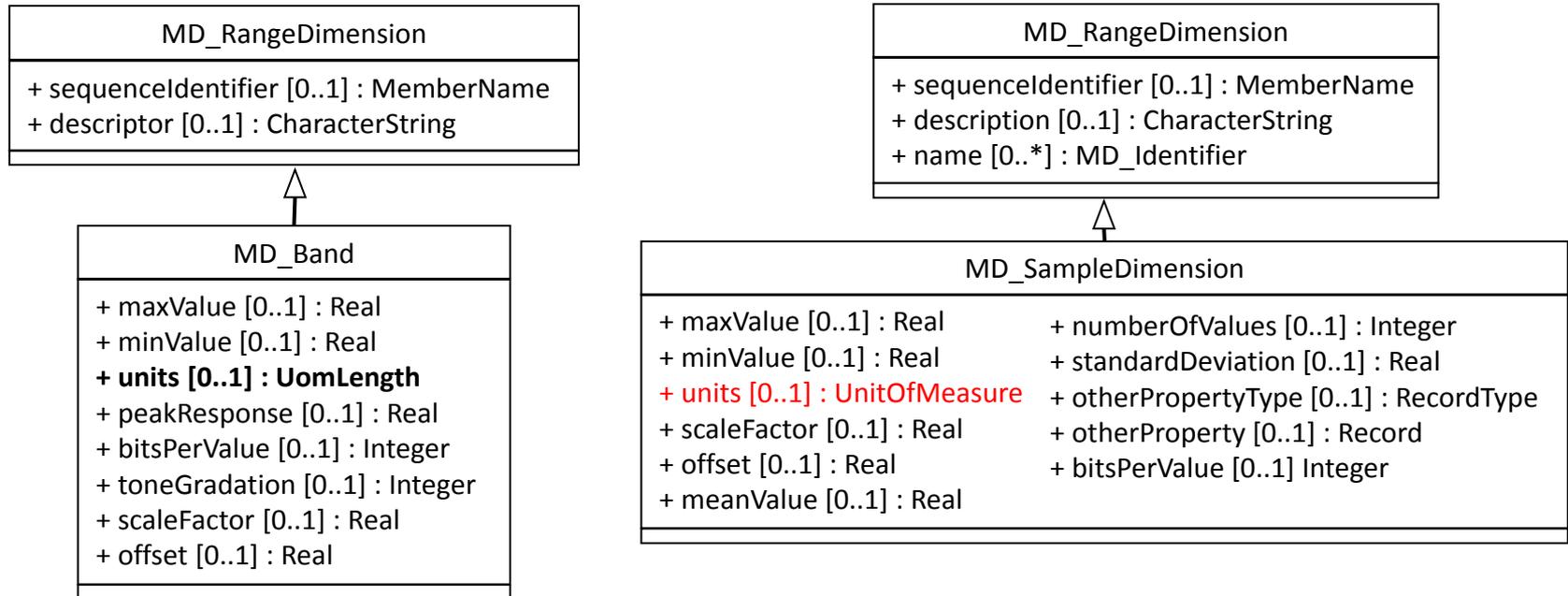
“My data are higher level products that need summary statistics”



The ISO 19115 MD_Band object is designed to describe low-level (Level 1) data as collected from an instrument.

ISO 19115-1 introduces the MD_SampleDimension which includes many general summary statistics for each band. MD_Band still exists as a specialized case.

“Users need to know the units of my products so that they can use them correctly”

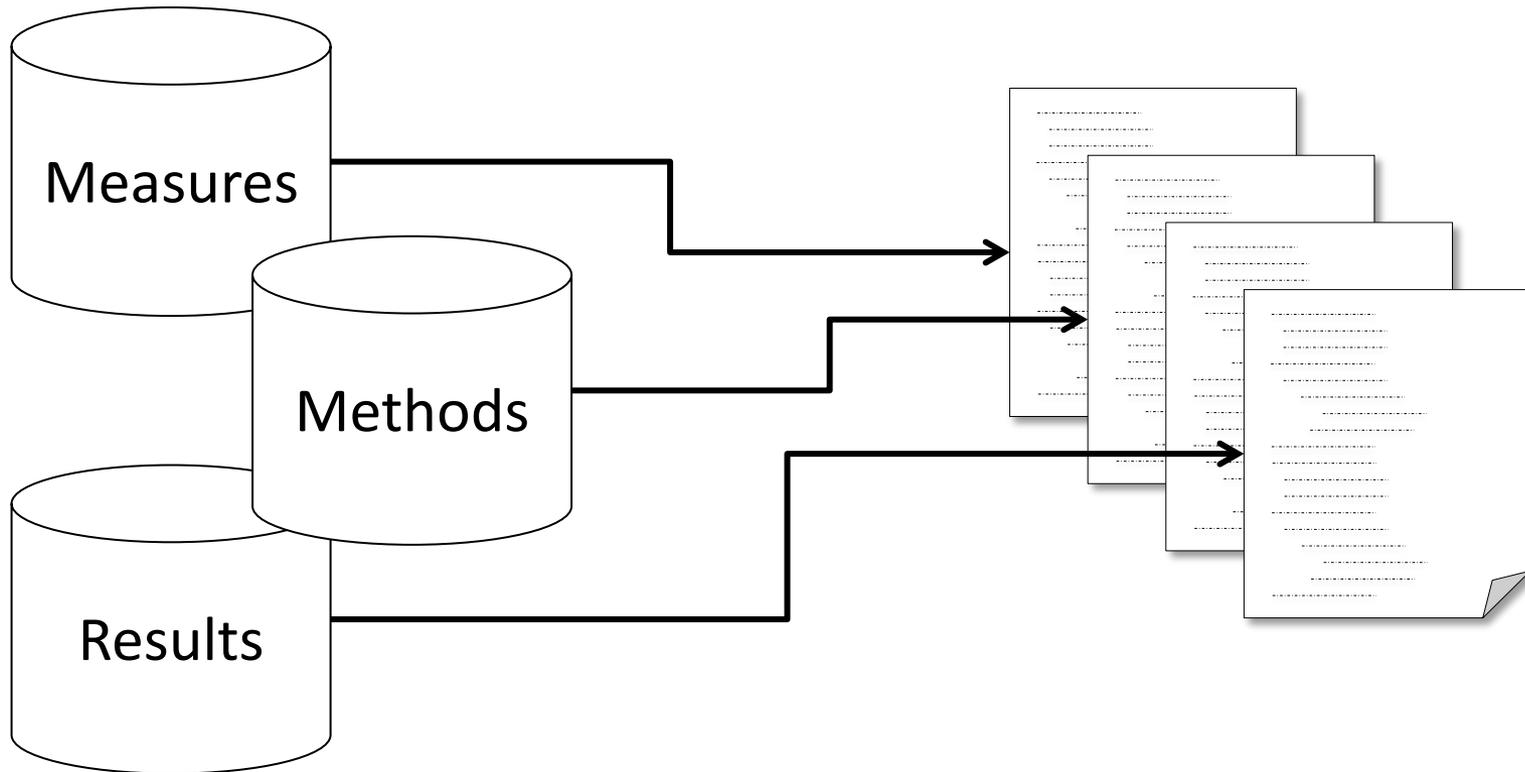


The ISO 19115 MD_Band units are defined as the units used to define the minimum and maximum wavelength for the band. They are units of length. They were not related to the data.

ISO 19115-1 MD_SampleDimension units is defined as the units of the data in the coverage. They can be any unit of measure. They are related to the data.

“My data quality information exists in databases or web services.”

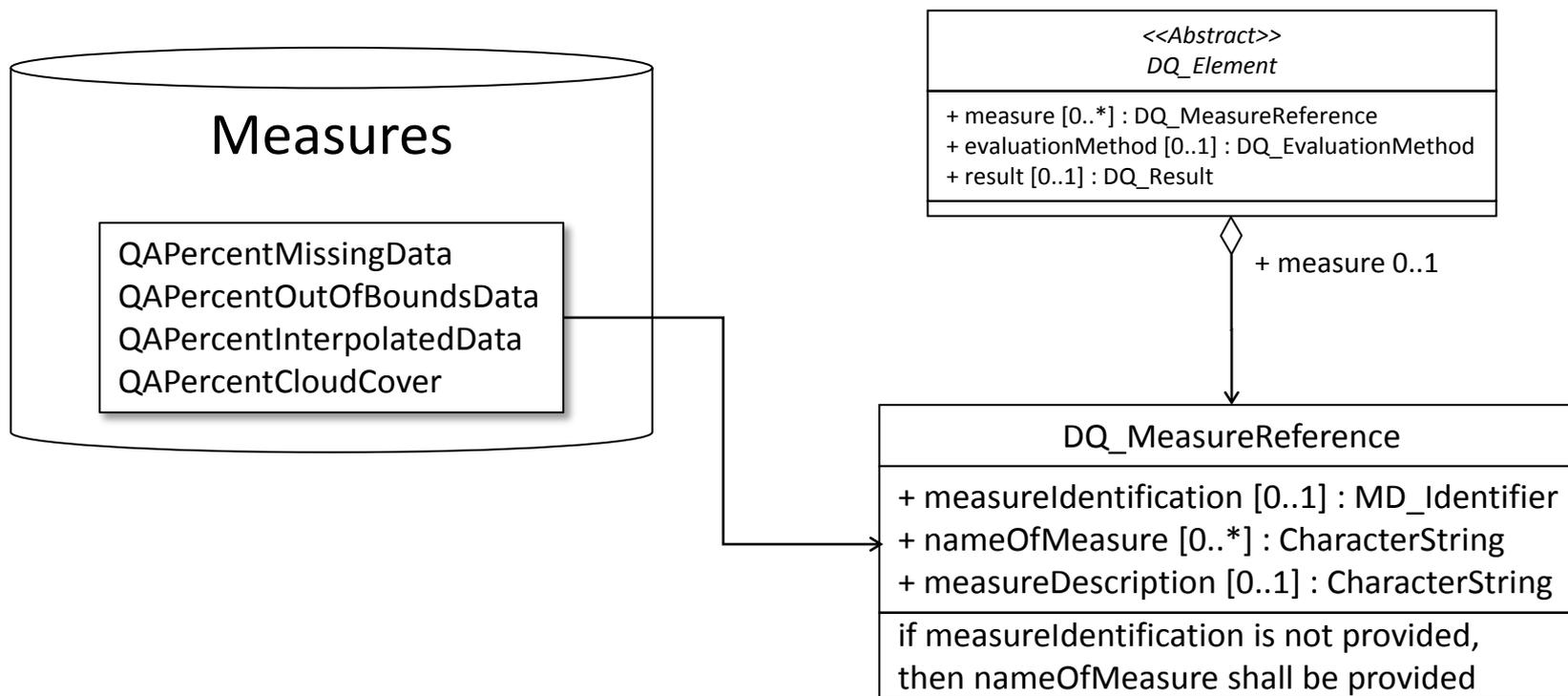
Major elements of the 19157 conceptual model are separate components that can be independently connected to the metadata and reused in multiple records.





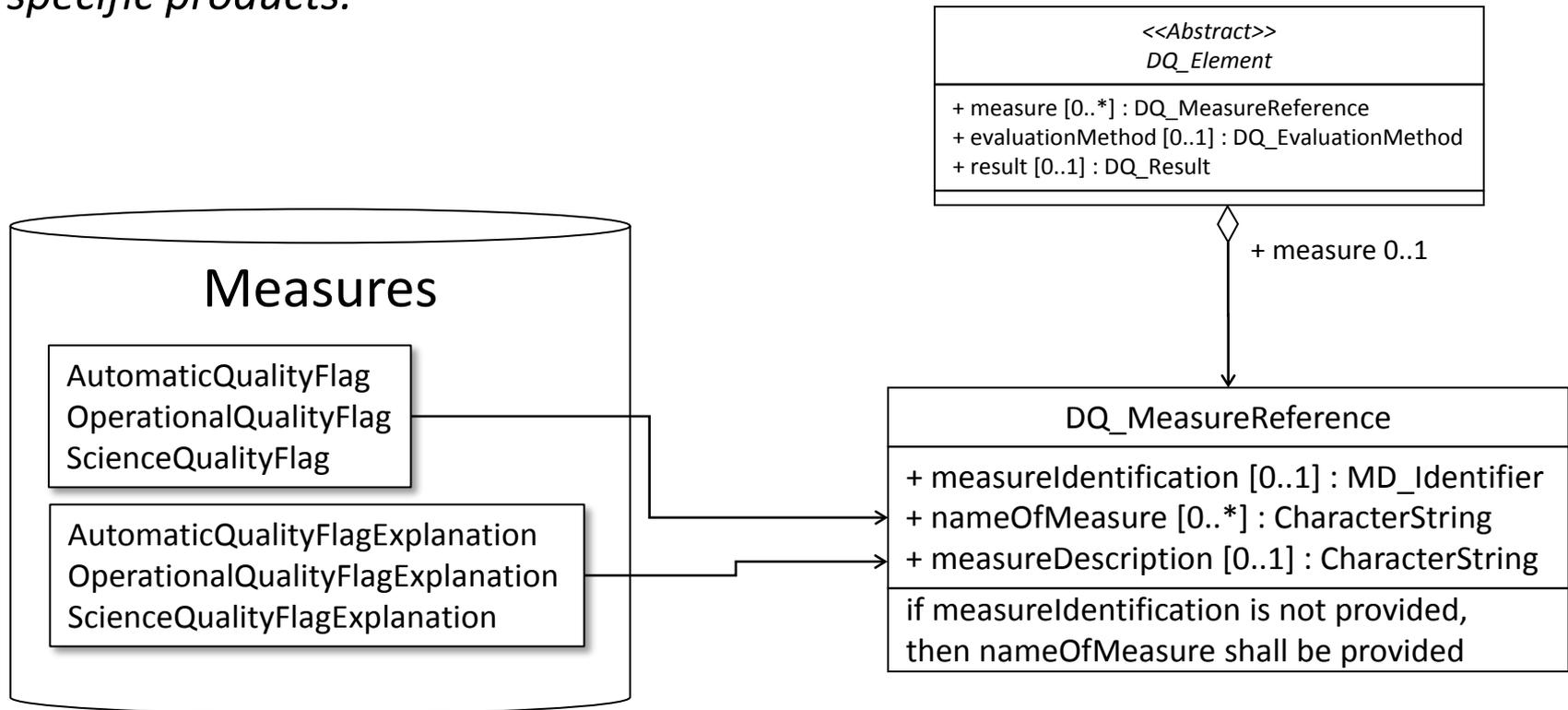
Standard Data Quality Measures

“We use standard quality measures for all products.”



Data quality measures that are the same across many products can be referenced from a measure database using a name or identifier.

“We use classes of quality measures that need implementation details for specific products.”



ISO 19157 data quality measure references identify measures in several ways and provides a brief description of the measure.



Acknowledgements



This work was partially supported by contract number NNG10HP02C from NASA.

Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect the views of NASA or The HDF Group.