

Interoperable Documentation

Ted Habermann, NOAA/NESDIS/NGDC
NCAR Earth Observing Laboratory, June 2010

Links:

GEO-IDE Wiki: <https://www.nosc.noaa.gov/dmc/swg/wiki/index.php>

ISO WAF: <http://www.ngdc.noaa.gov/metadata/published/19115/isoMetadataHome.html>

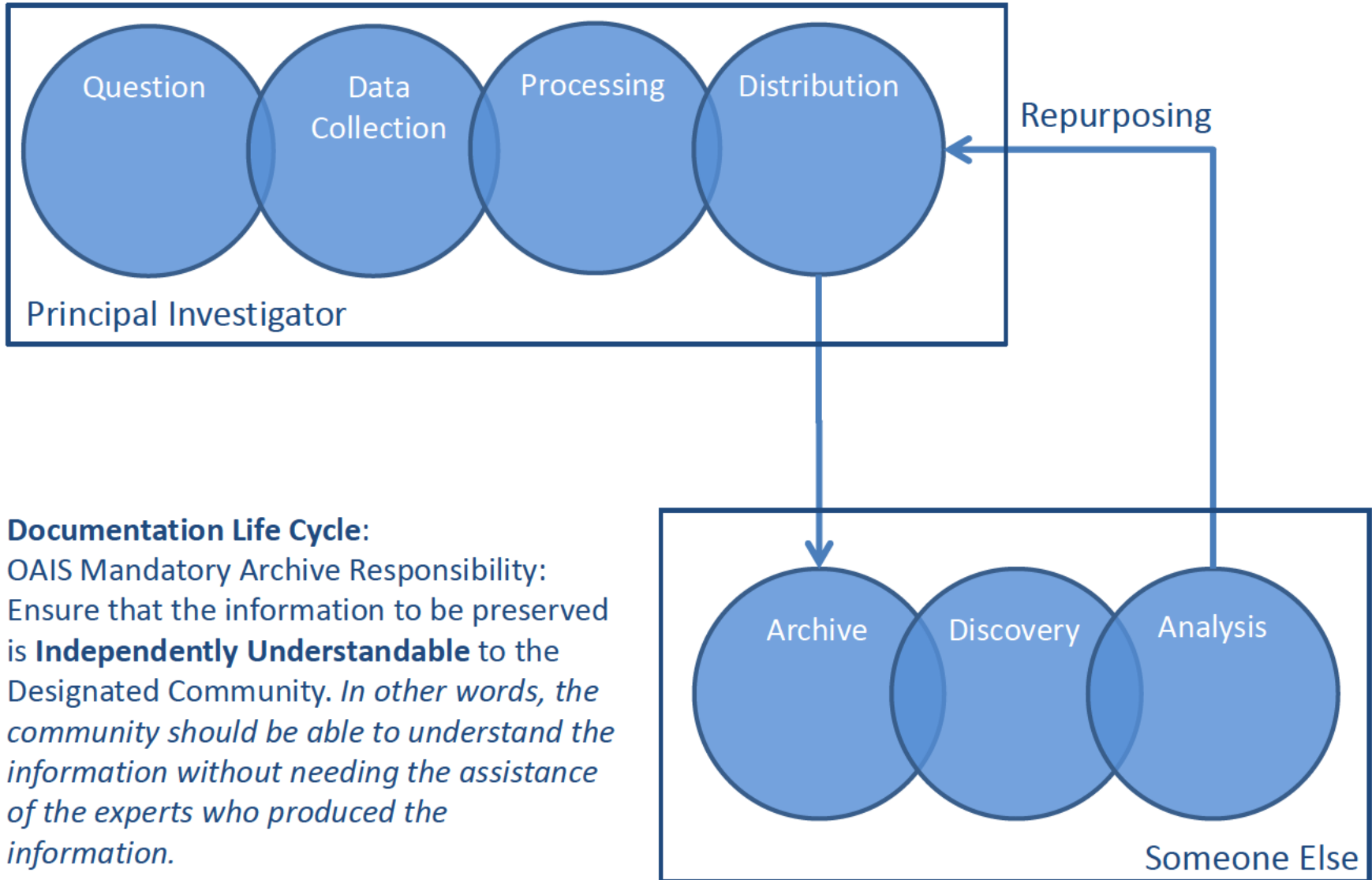
THREDDS Extractor: <http://www.ngdc.noaa.gov/eds/tds/oceanSITESMetadataAssessment.html>

Spirals: https://www.nosc.noaa.gov/dmc/swg/wiki/index.php?title=Creating_Good_Documentation

Questions During Seminar: jweber@unidata.ucar.edu

Questions / Comments / Suggestions: ted.habermann@noaa.gov

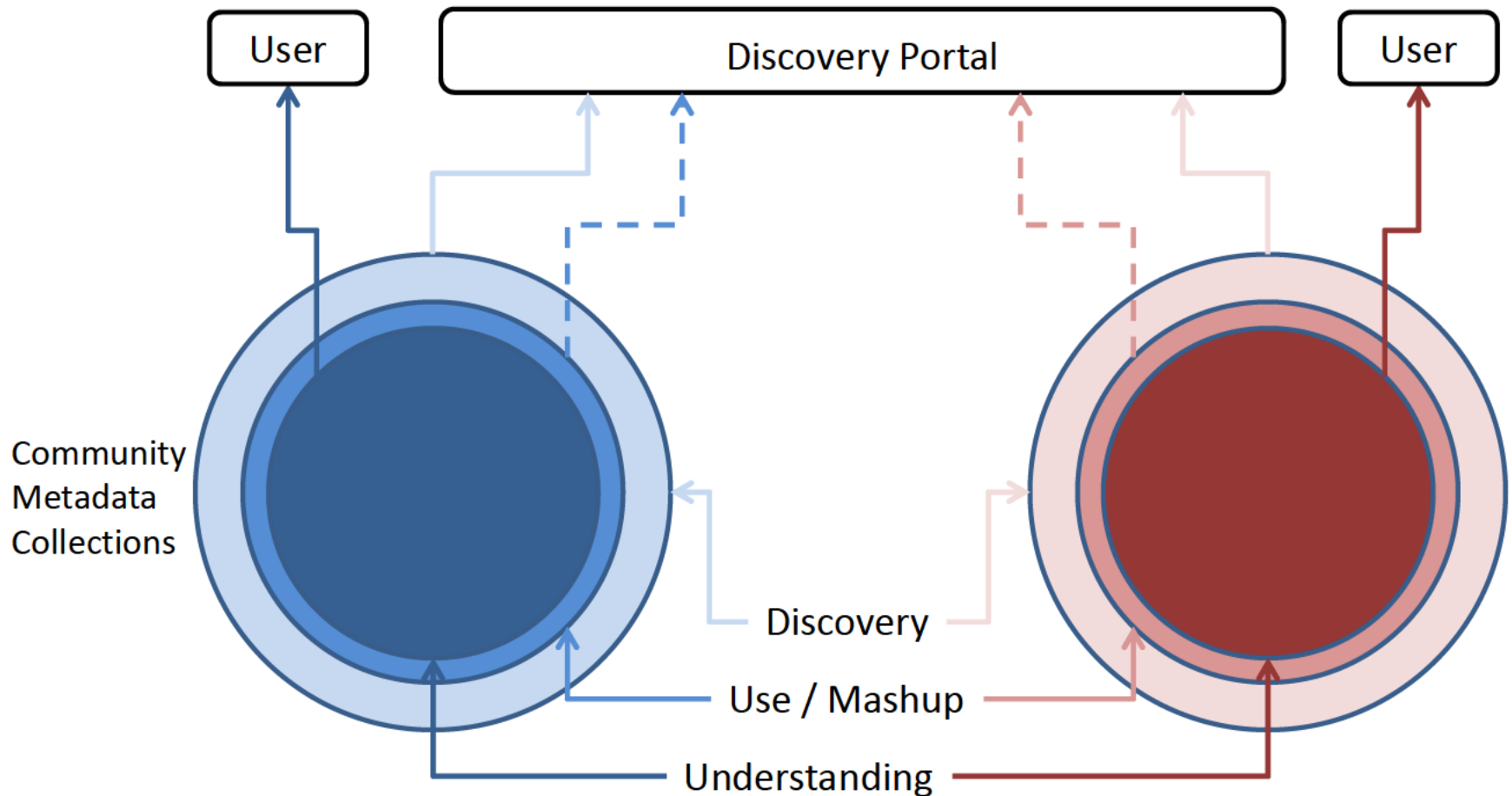
Data Life-Cycle



Documentation Life Cycle:

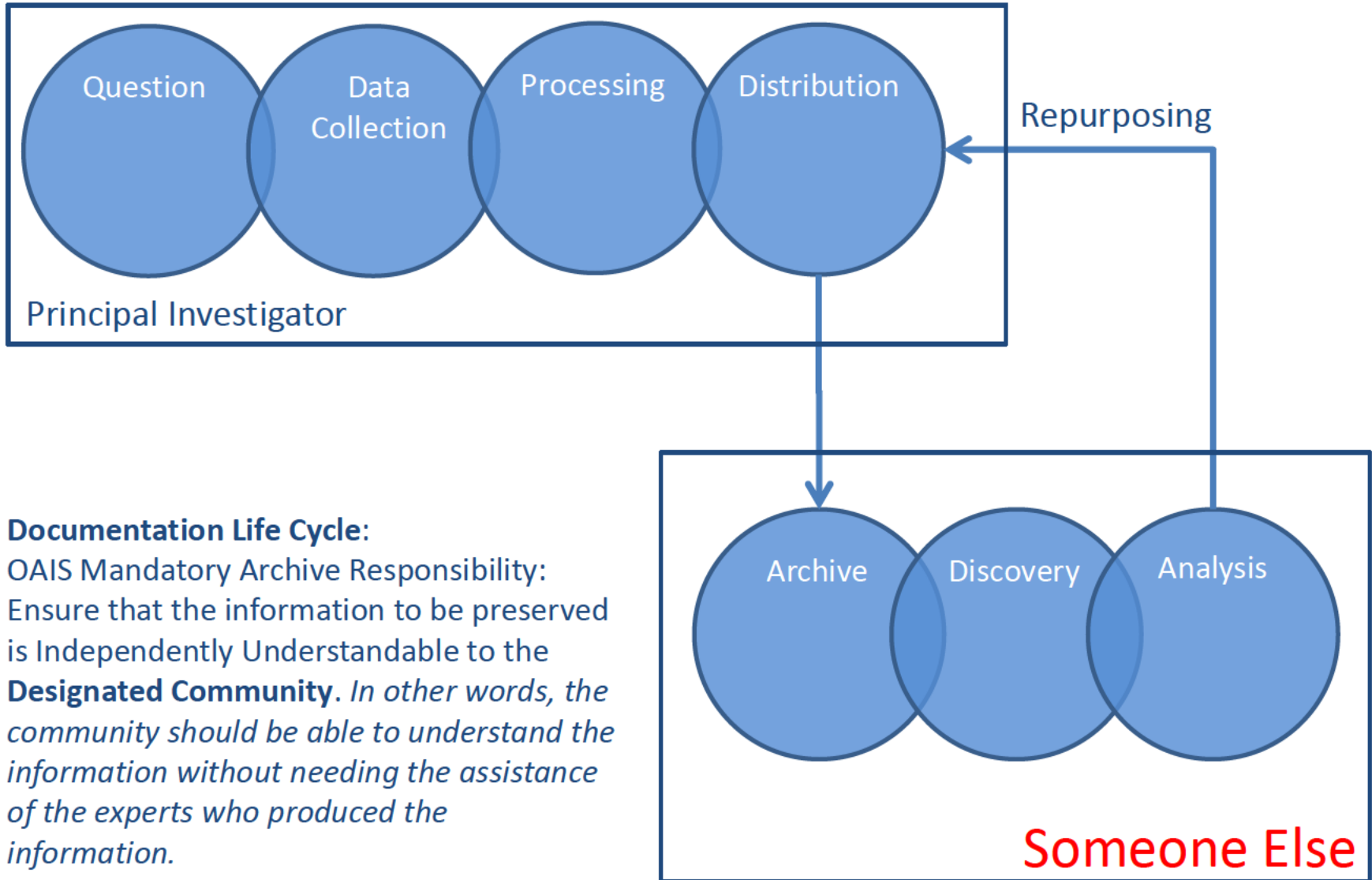
OAIS Mandatory Archive Responsibility:
Ensure that the information to be preserved is **Independently Understandable** to the Designated Community. *In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information.*

Metadata Types and Sharing



More documentation is required for understanding data than discovering or using it.

Designated Communities - Users



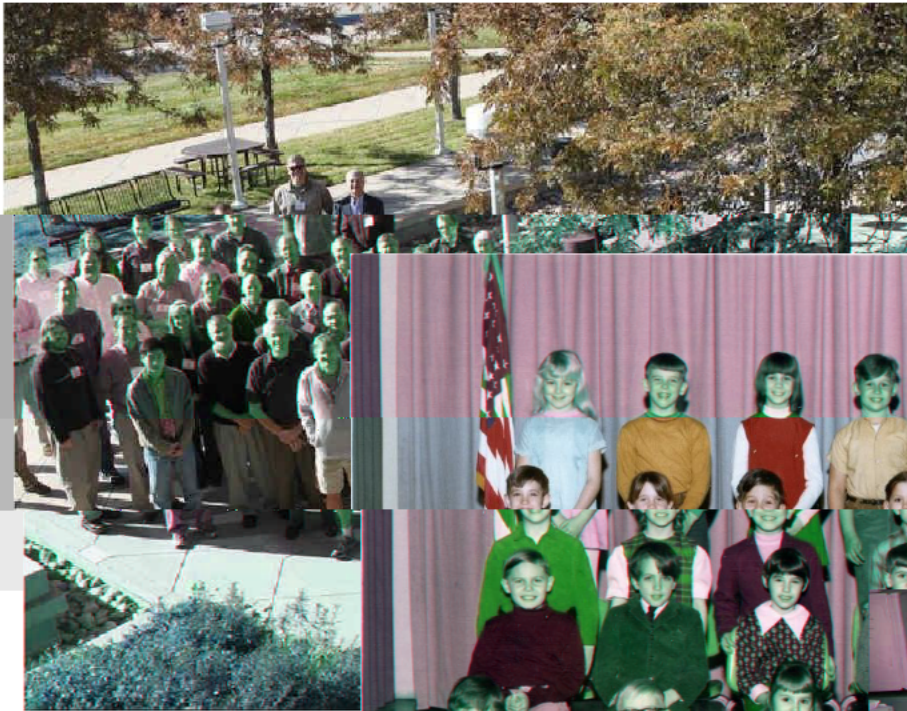
Documentation Life Cycle:

OAIS Mandatory Archive Responsibility:

Ensure that the information to be preserved is Independently Understandable to the **Designated Community**. *In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information.*

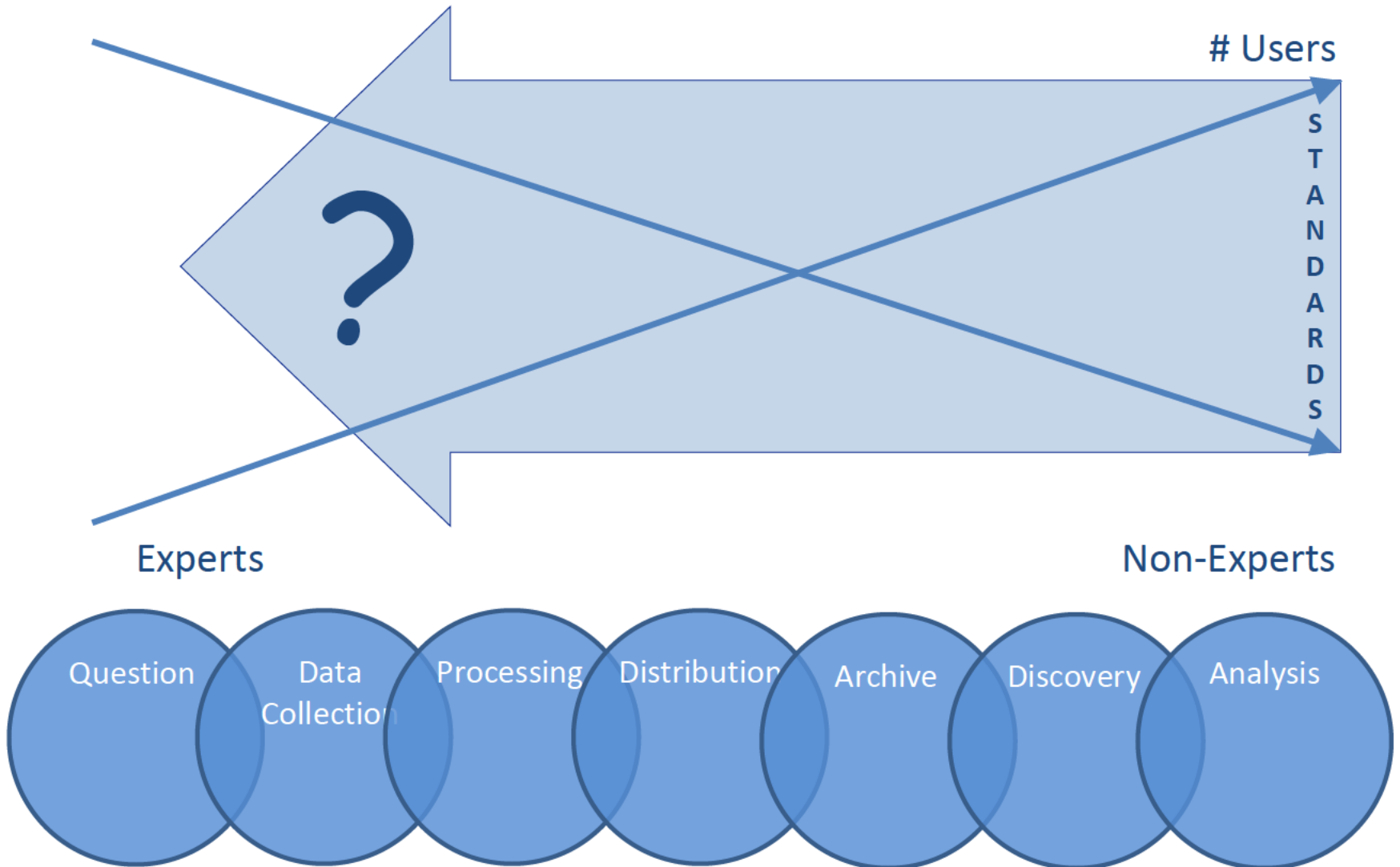
Someone Else

Designated Communities - Users

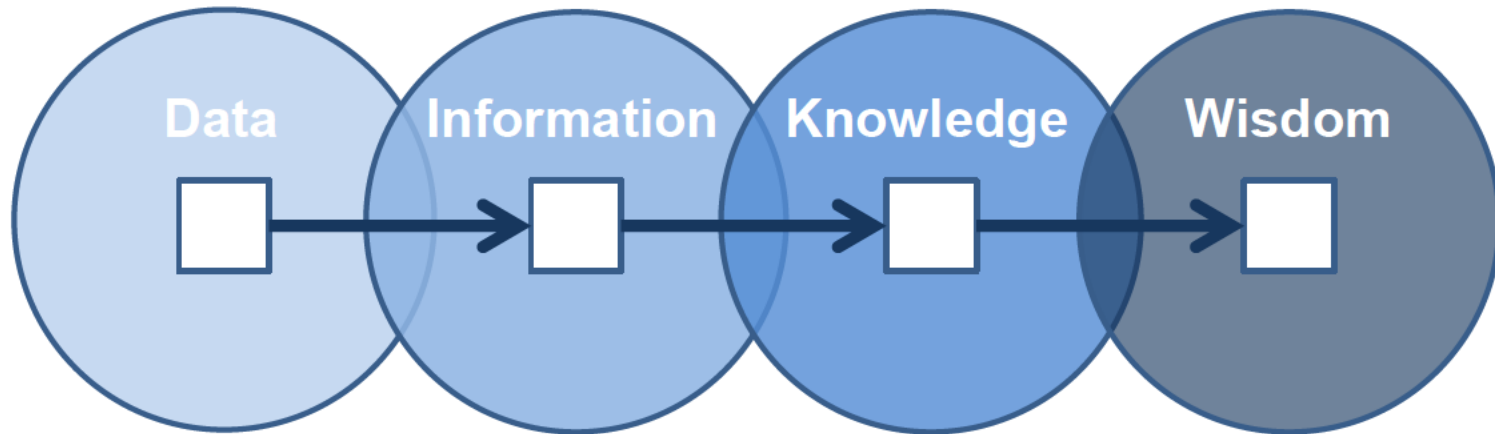


Data preservation is communicating with the future

Designated Communities - Users



Interoperable Documentation



Many concepts have been well developed and successfully implemented to achieve (or at least improve) data interoperability

Can the same concepts facilitate interoperable information?

Data to Information Concept Mapping:

Evolution

Variables and Properties

Multiple Dialects

Persistence vs. Transport

Hierarchical Organizations

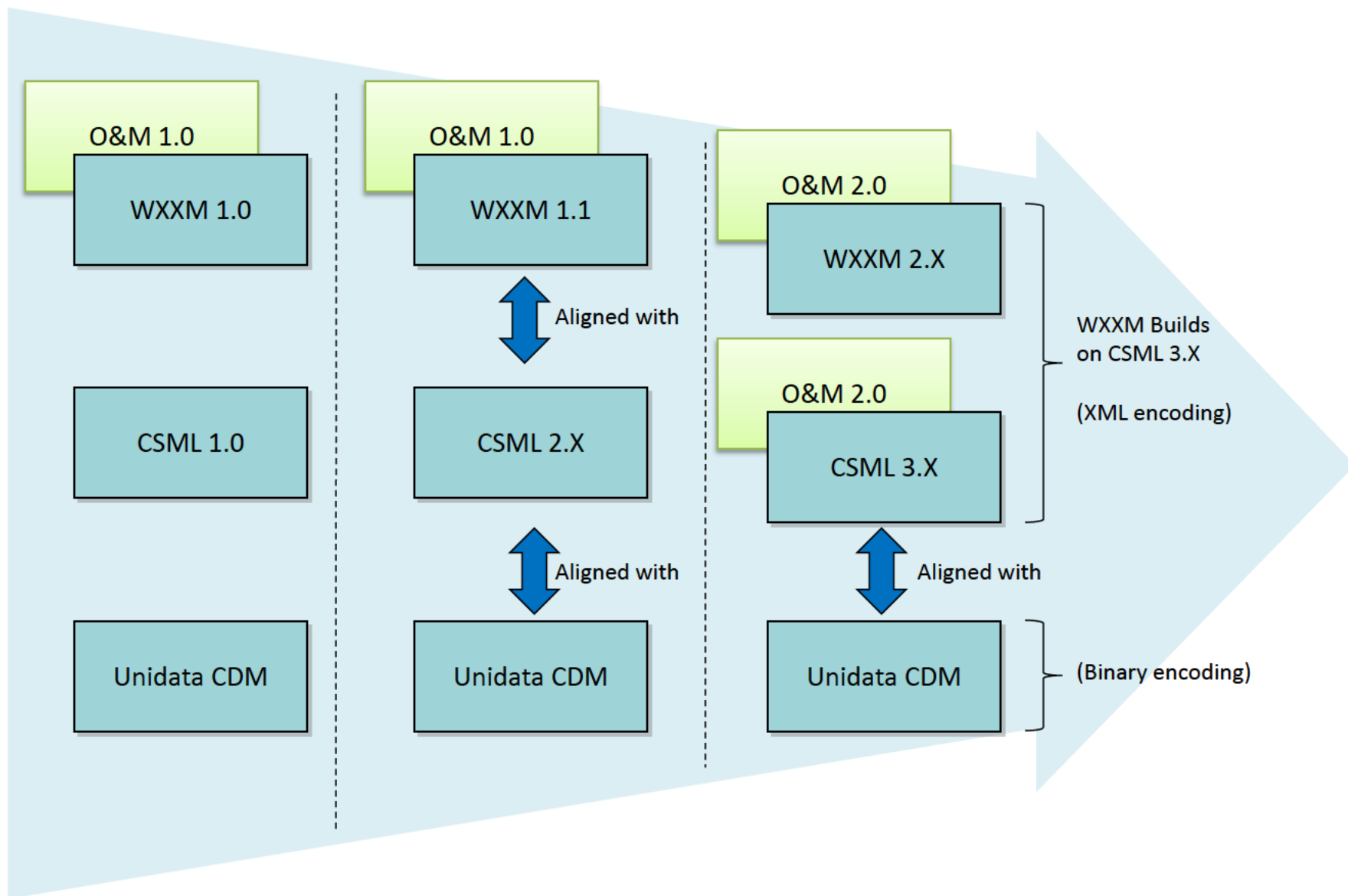
Standards and Conventions

Spiral Development

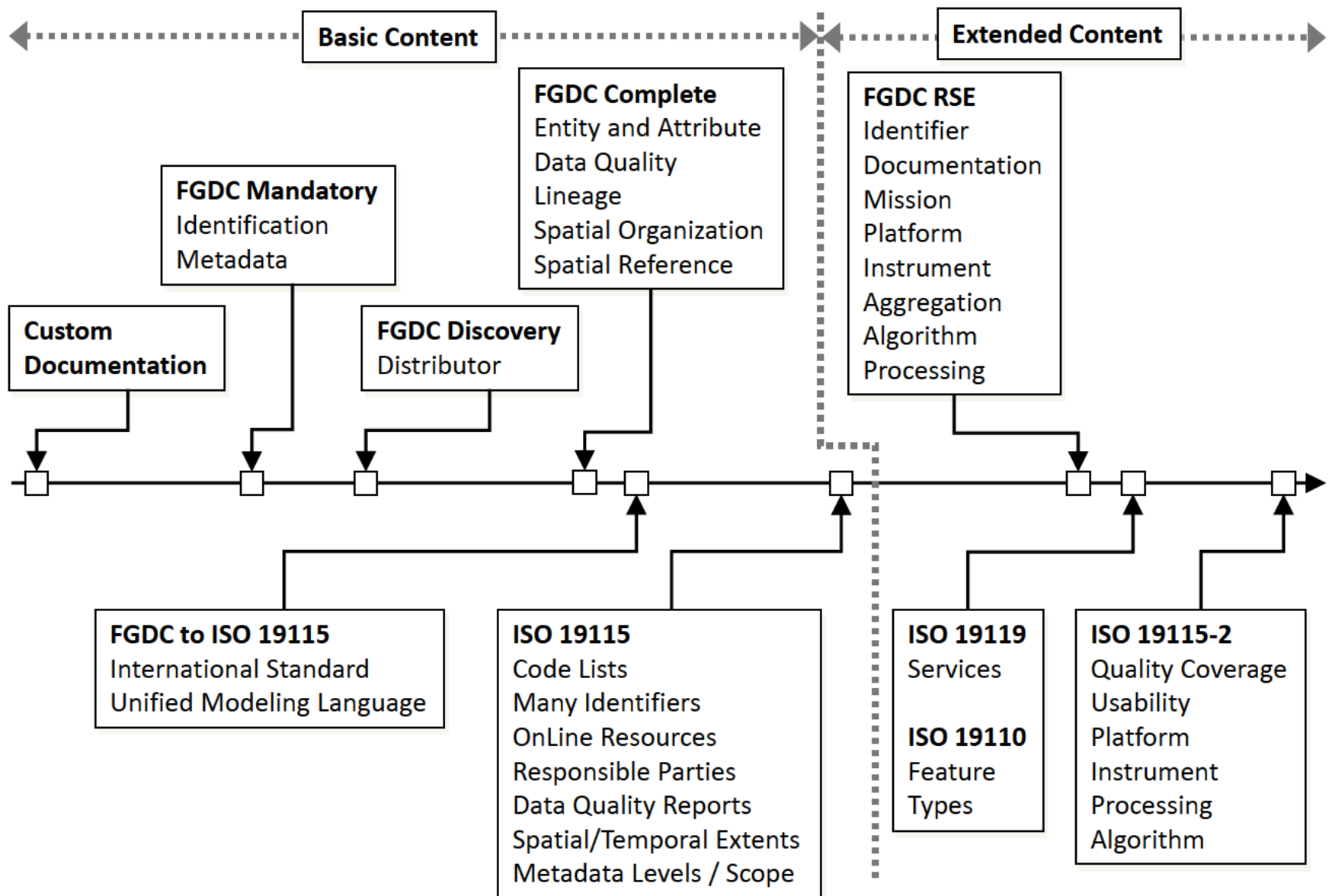
Spatial/Temporal Data

Training

Evolution: Data



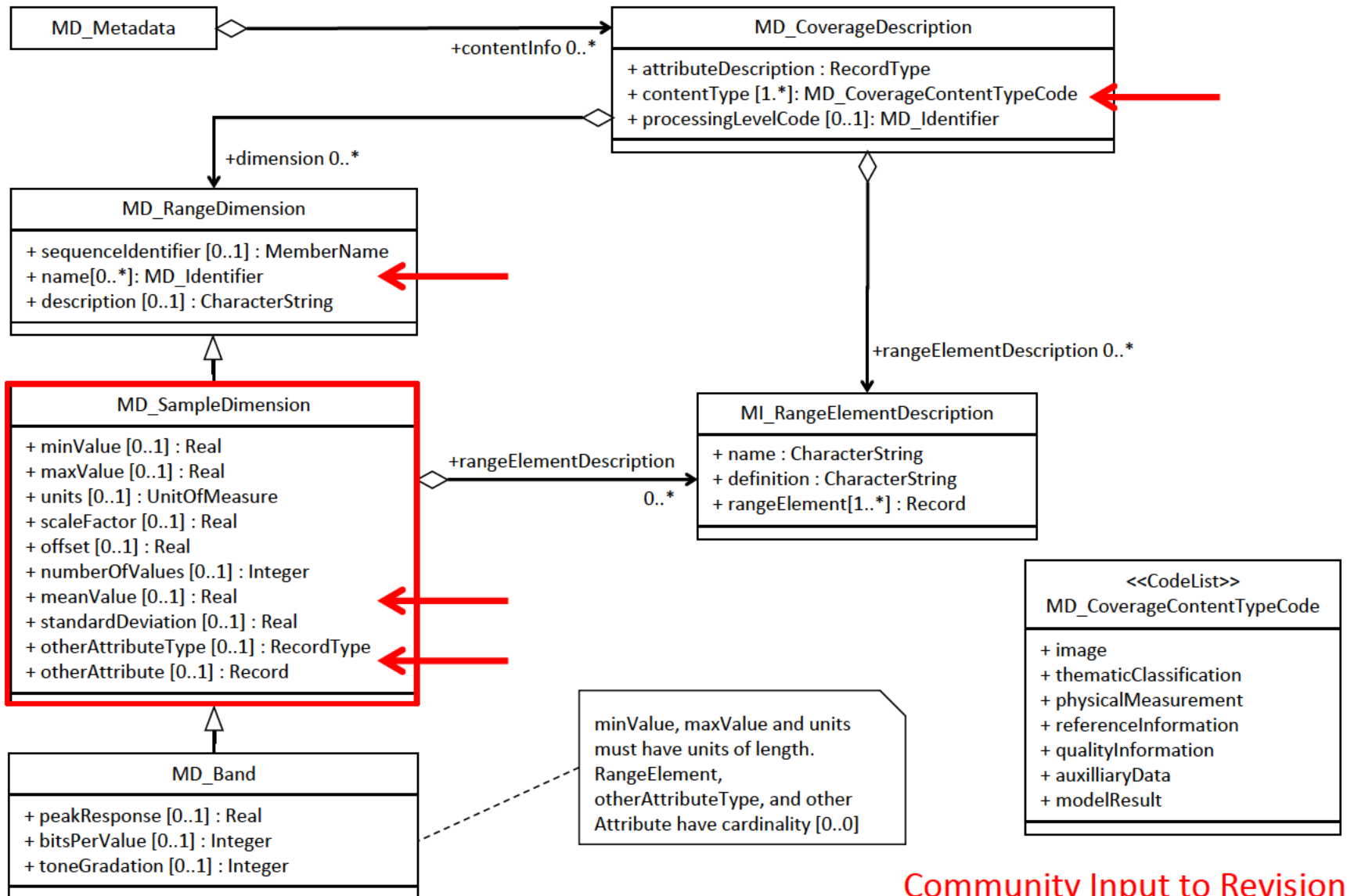
Evolution: Documentation Standards



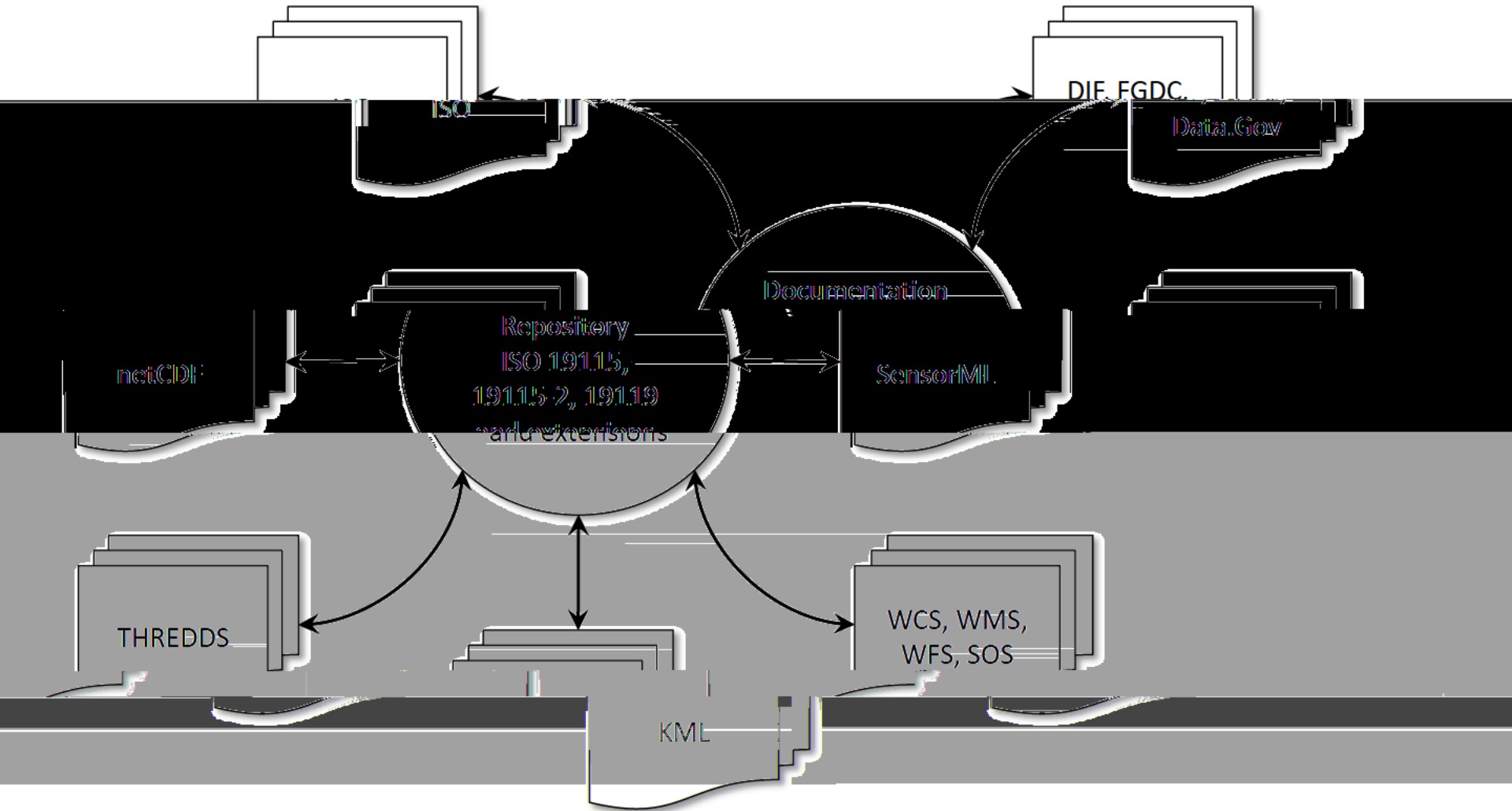
Variables and Properties - Data

```
<variable name="MHchla" type="float" shape="time altitude lat lon">
  <attribute name="_CoordinateAxes" value="time altitude lat lon "/>
  <attribute name="_FillValue" value="-9999999.0 " type="float"/>
  <attribute name="actual_range" value="0.01 63.997 " type="float"/>
  <attribute name="coordsys" value="geographic"/>
  <attribute name="fraction_digits" value="2 " type="int"/>
  <attribute name="long_name" value="Chlorophyll-a, Aqua MODIS, NPP, 0.05
    degrees, Global, Science Quality"/>
  <attribute name="missing_value" value="-9999999.0 " type="float"/>
  <attribute name="numberOfObservations" value="9664503 " type="int"/>
  <attribute name="percentCoverage" value="0.2589298000257202 "
    type="double"/>
  <attribute name="standard_name"
    value="concentration_of_chlorophyll_in_sea_water"/>
  <attribute name="units" value="mg m-3"/>
</variable>
```

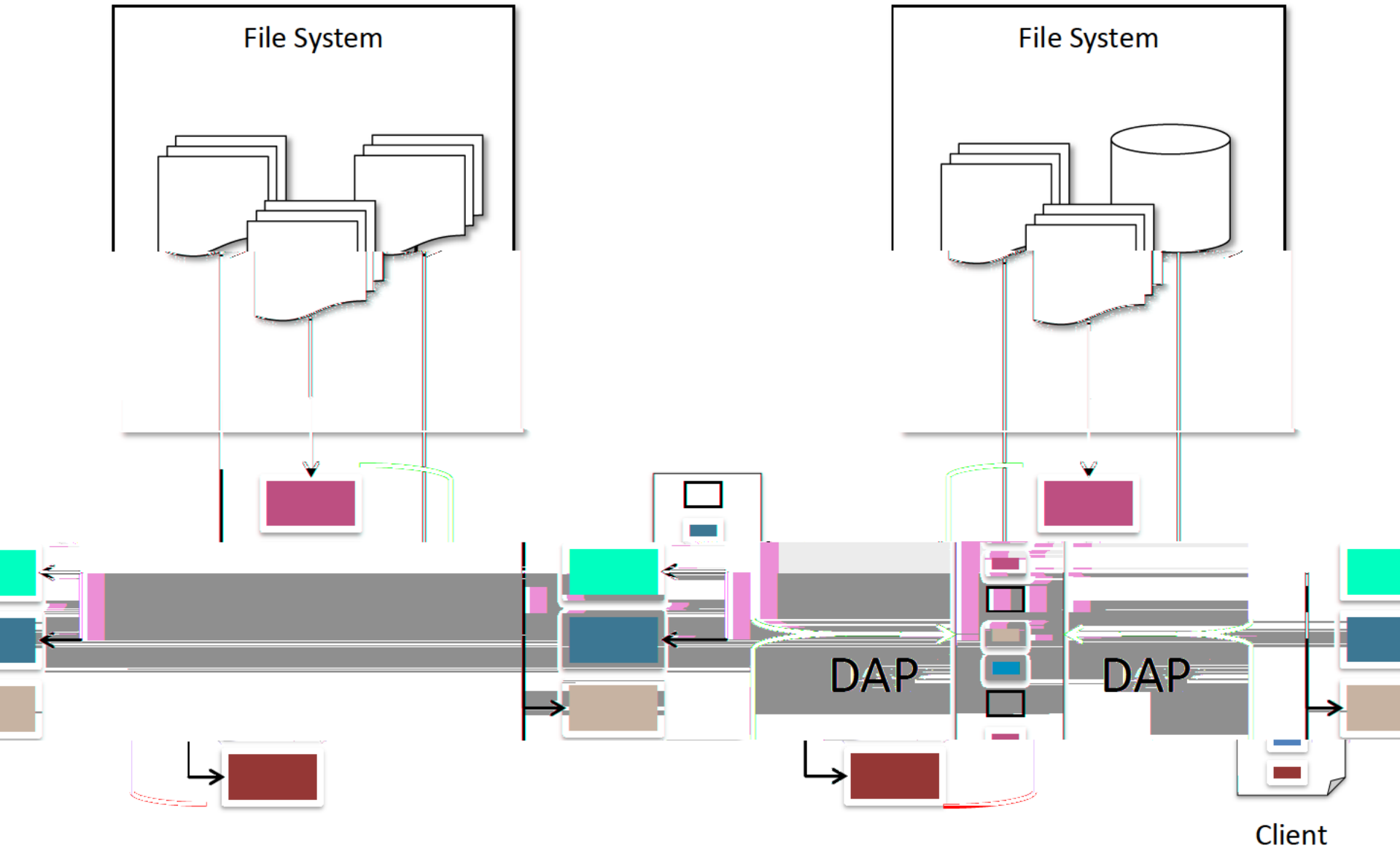
Variables and Properties - Documentation



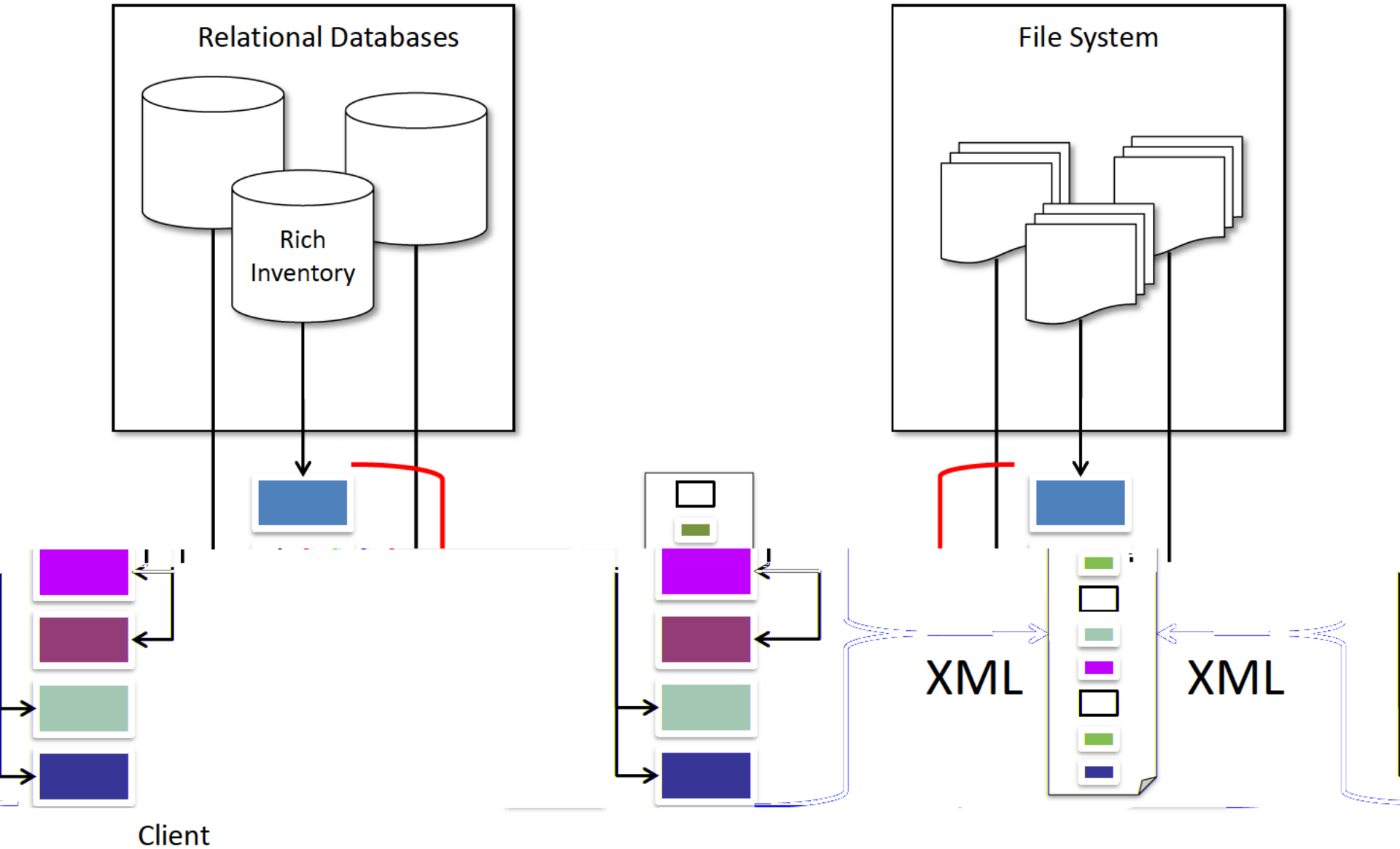
Documentation in Multiple Dialects



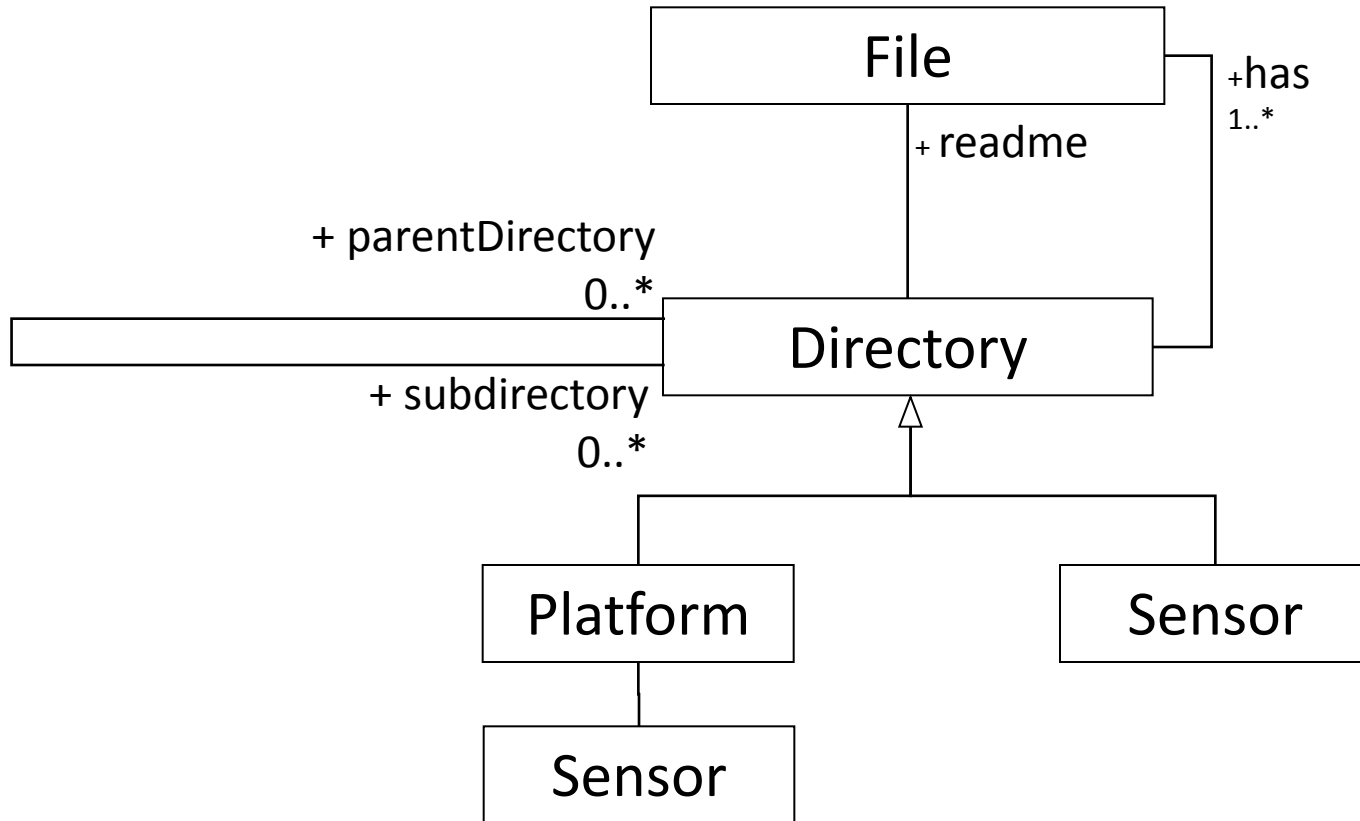
Persistence vs. Transport - OPeNDAP



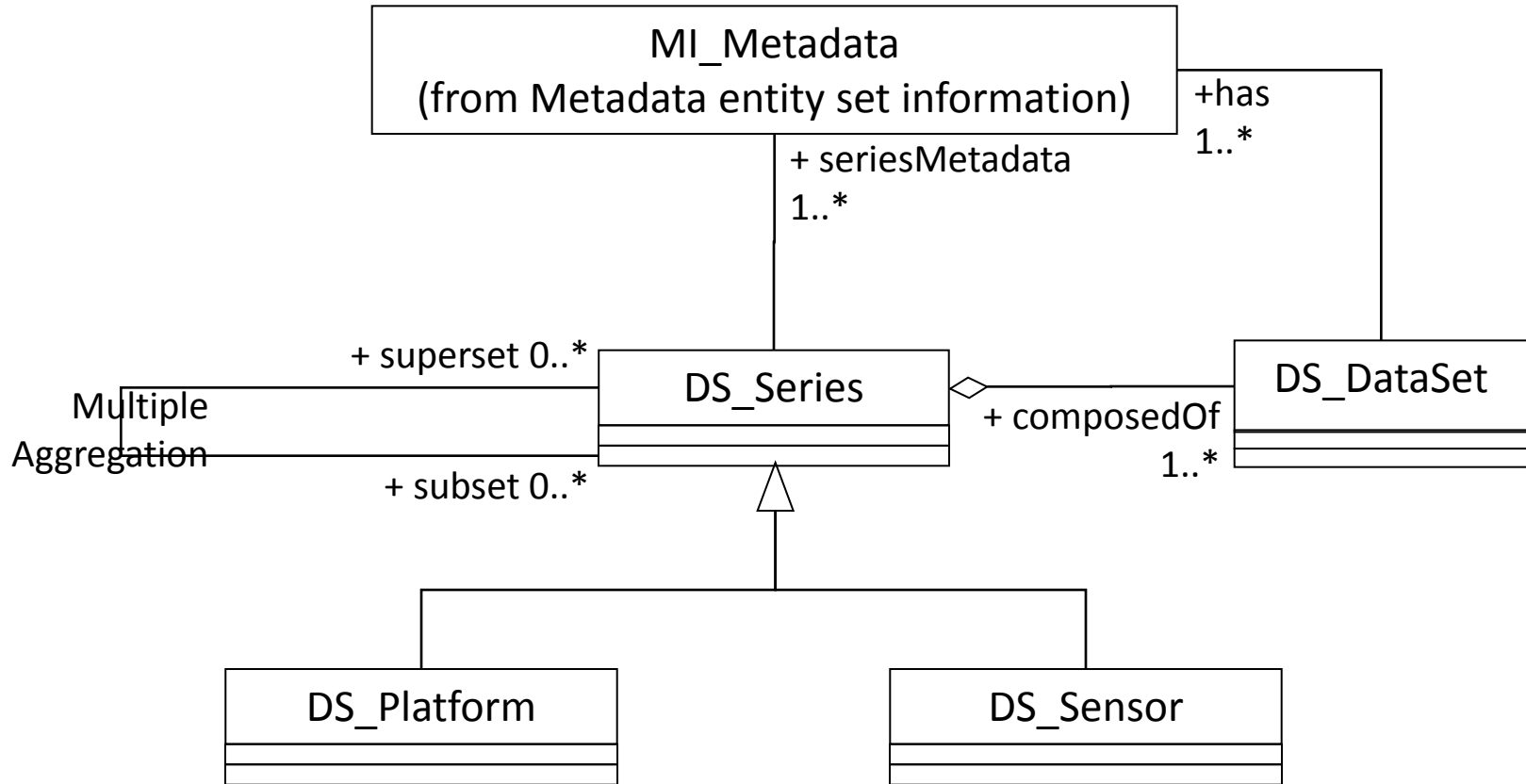
Persistence vs. Transport - Documentation



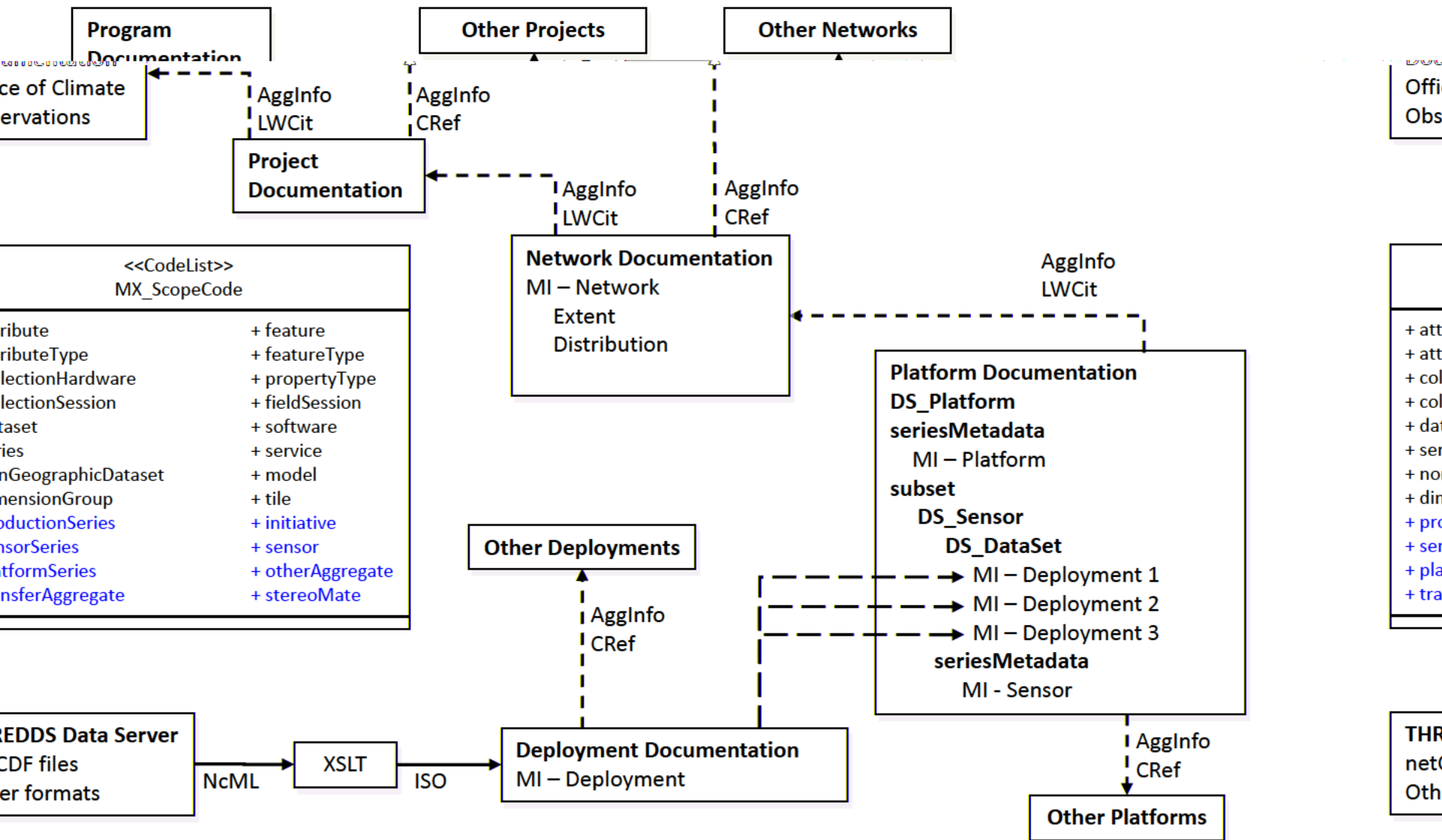
Hierarchical Organization: Data



Hierarchical Organization: Documentation



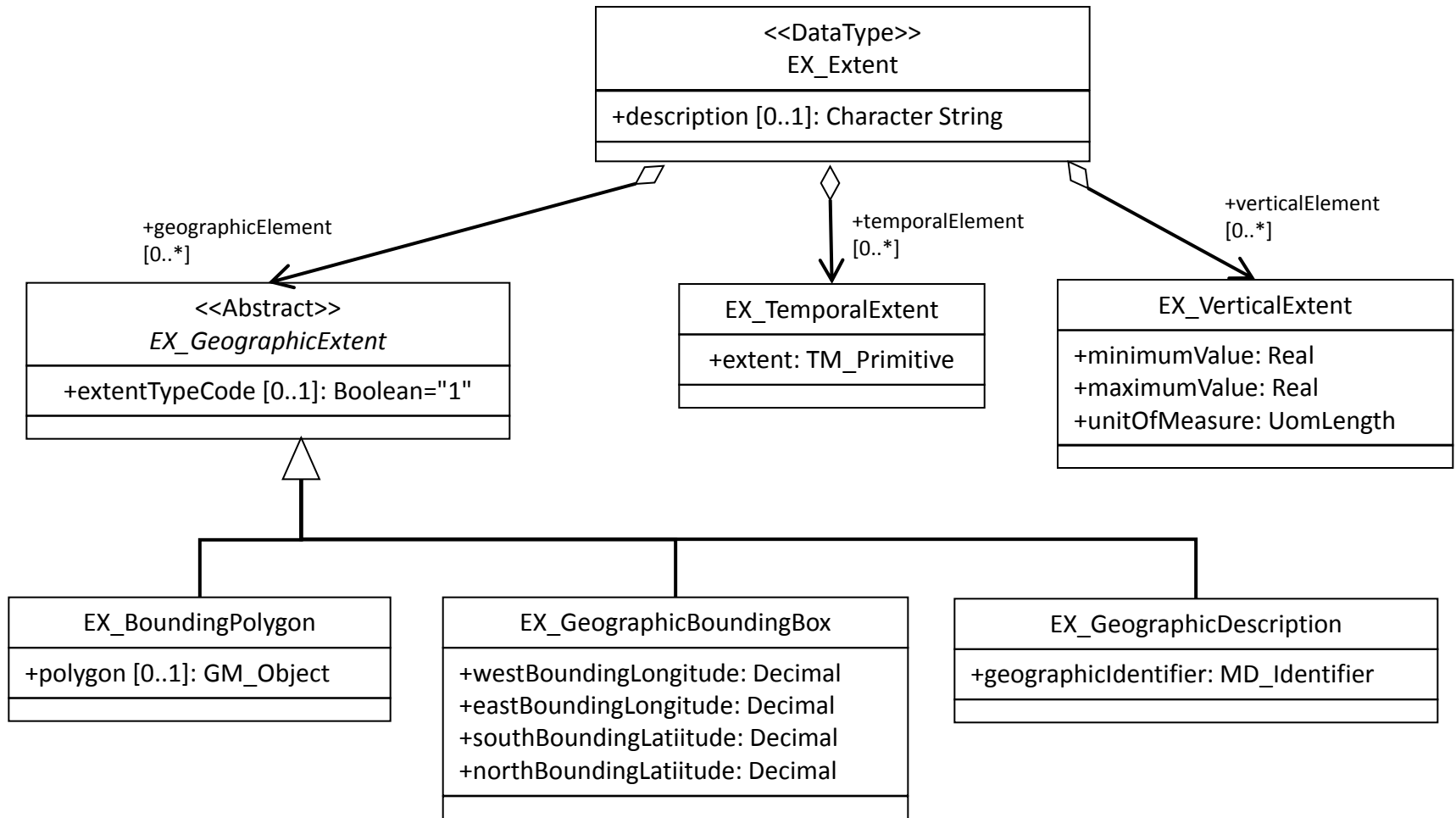
Hierarchical Organization: InSitu Documentation



Standards / Conventions / Best Practices

The image shows a stack of Mozilla Firefox browser windows. The top window is titled 'CF Metadata - CF Metadata - Mozilla Firefox'. Below it is 'Dataset Discovery NetCDF Attribute Convention - Mozilla Firefox'. The third window is 'Main Page - GEO-IDE Guidelines and Best Practices Wiki - Mozilla Firefox', which has a red circle around the 'Categories' link in its navigation menu. The bottom-most window is 'Category:ISO 19115 - GEO-IDE Guidelines and Best Practices Wiki - Mozilla Firefox', displaying a list of pages under the 'ISO 19115' category. The website content includes the NOAA logo and the text 'National Geophysical Data Center'. The 'Category:ISO 19115' page lists various metadata standards and resources, such as 'Buoy Metadata and Aggregation', 'ISO 19115 Identification Information', and 'ISO 19139 Identifiers'.

Spatial/Temporal Data

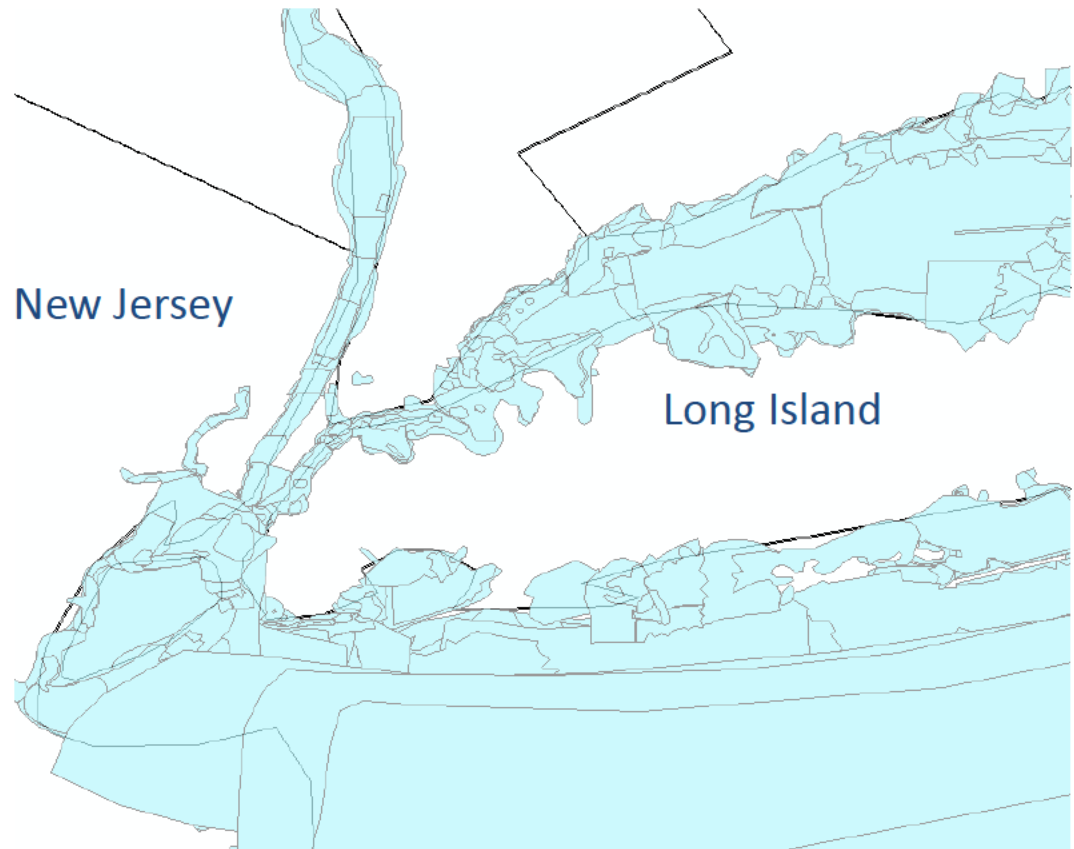


Ex_Extents can be used to describe datasets, sources, and quality reports.

Spatial/Temporal Data

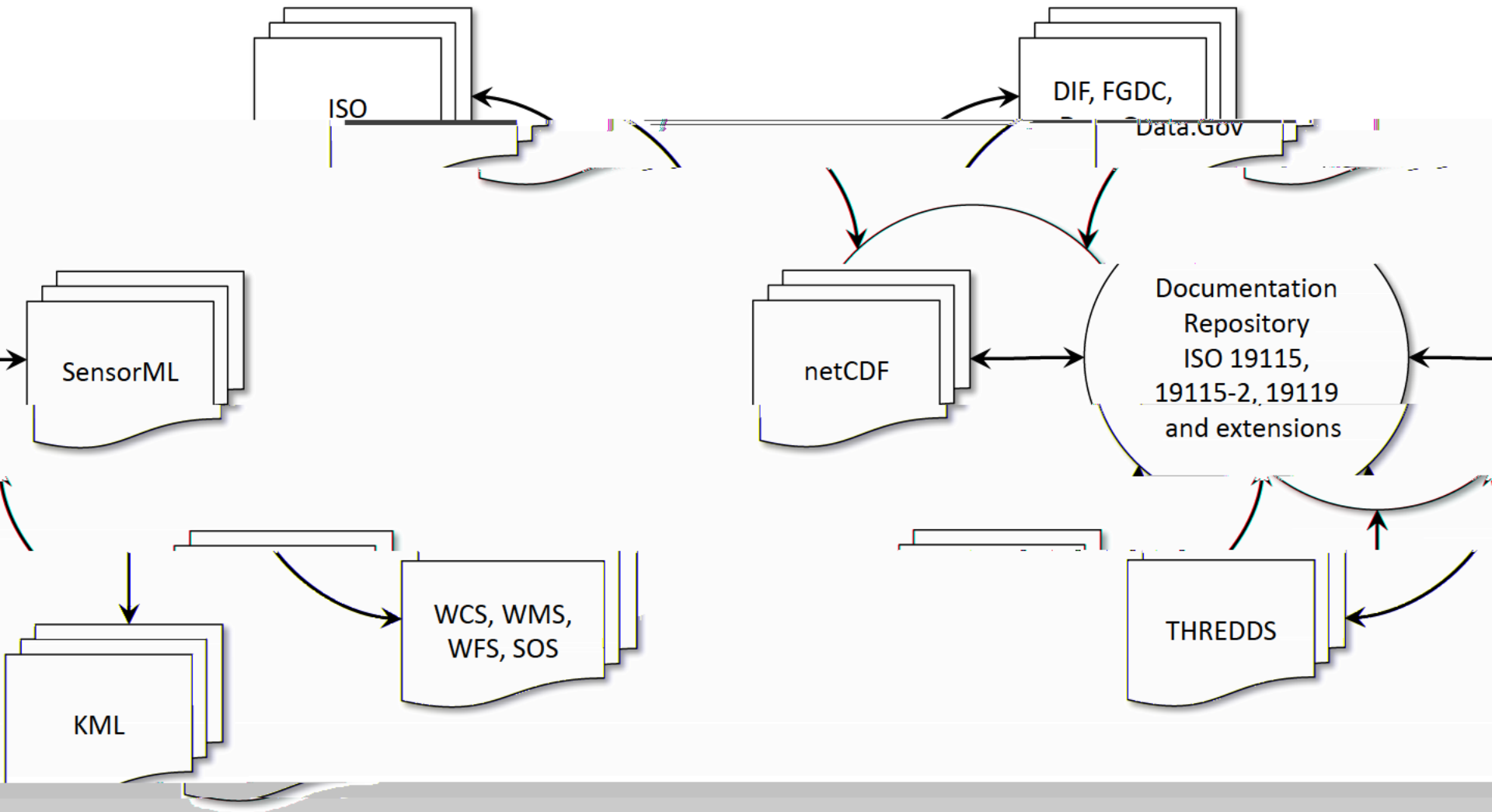
The National Ocean Service Hydrographic Survey dataset includes many sources collected at different locations and times. These extents can be described clearly using the ISO objects discussed in this presentation.

The metadata evolves from a single extent to a complete spatial / temporal dataset.

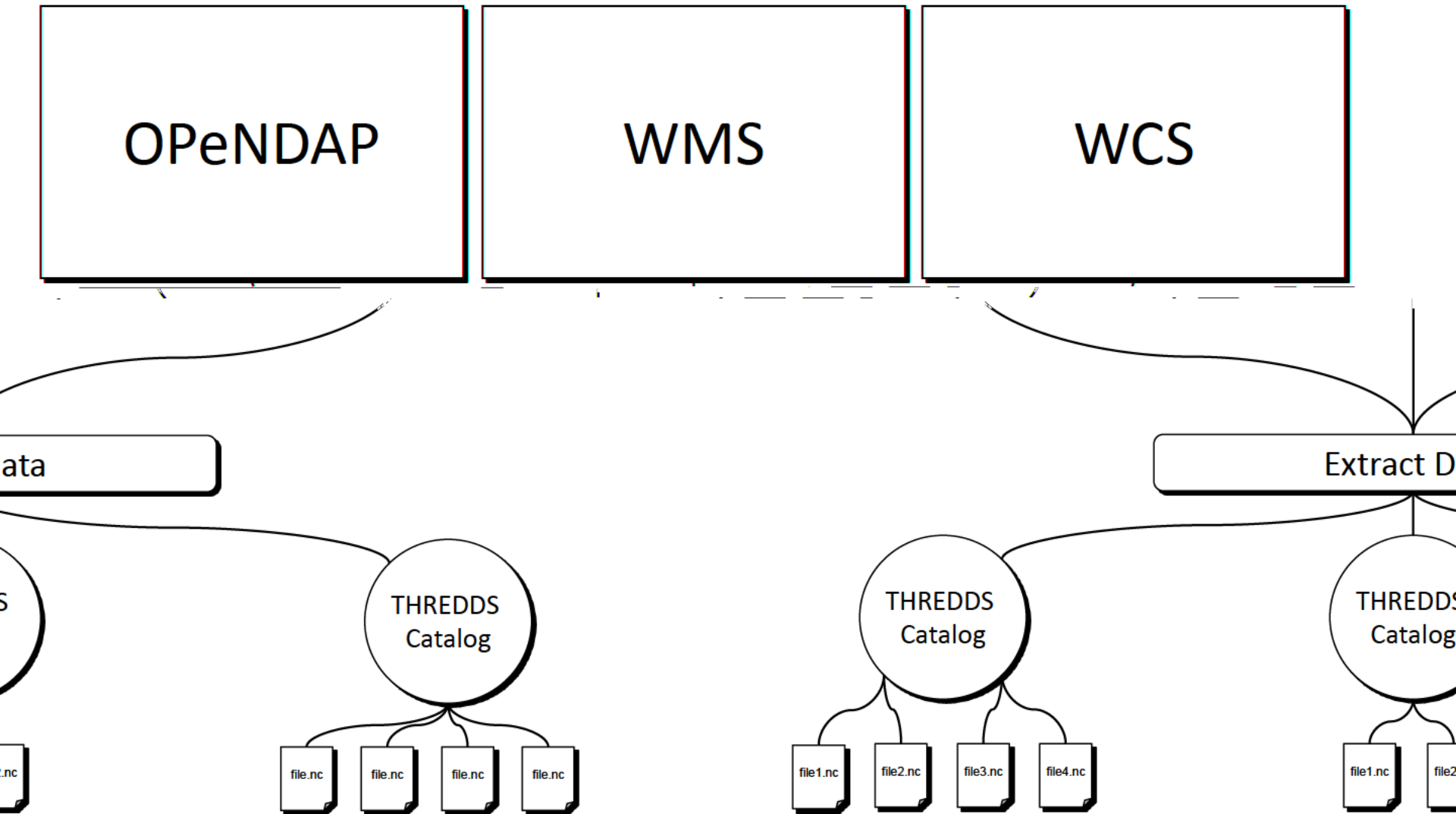


1900-1925 1926-1950 1951-1975 1976-2000 2001-2007

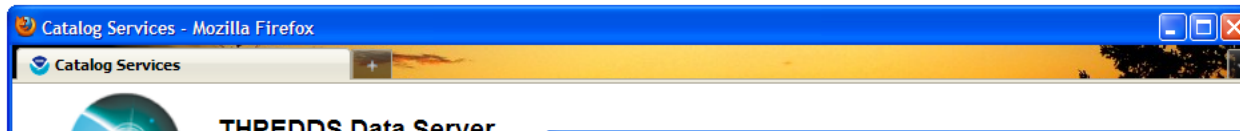
Multiple Dialects Revisited



Multiple Dialects: THREDDS Data Server



Multiple Dialects: Web Accessible Folders



Catalog <http://sdf.ndbc.noaa.gov>

Dataset: THREDDS-IDD WCS
Mexico 1km resolution HF Radar

- Data format: NetCDF
- Data type: Grid
- Naming Authority: unidata.ucar.edu:
- ID: hfradar_usegc_1km

Access:

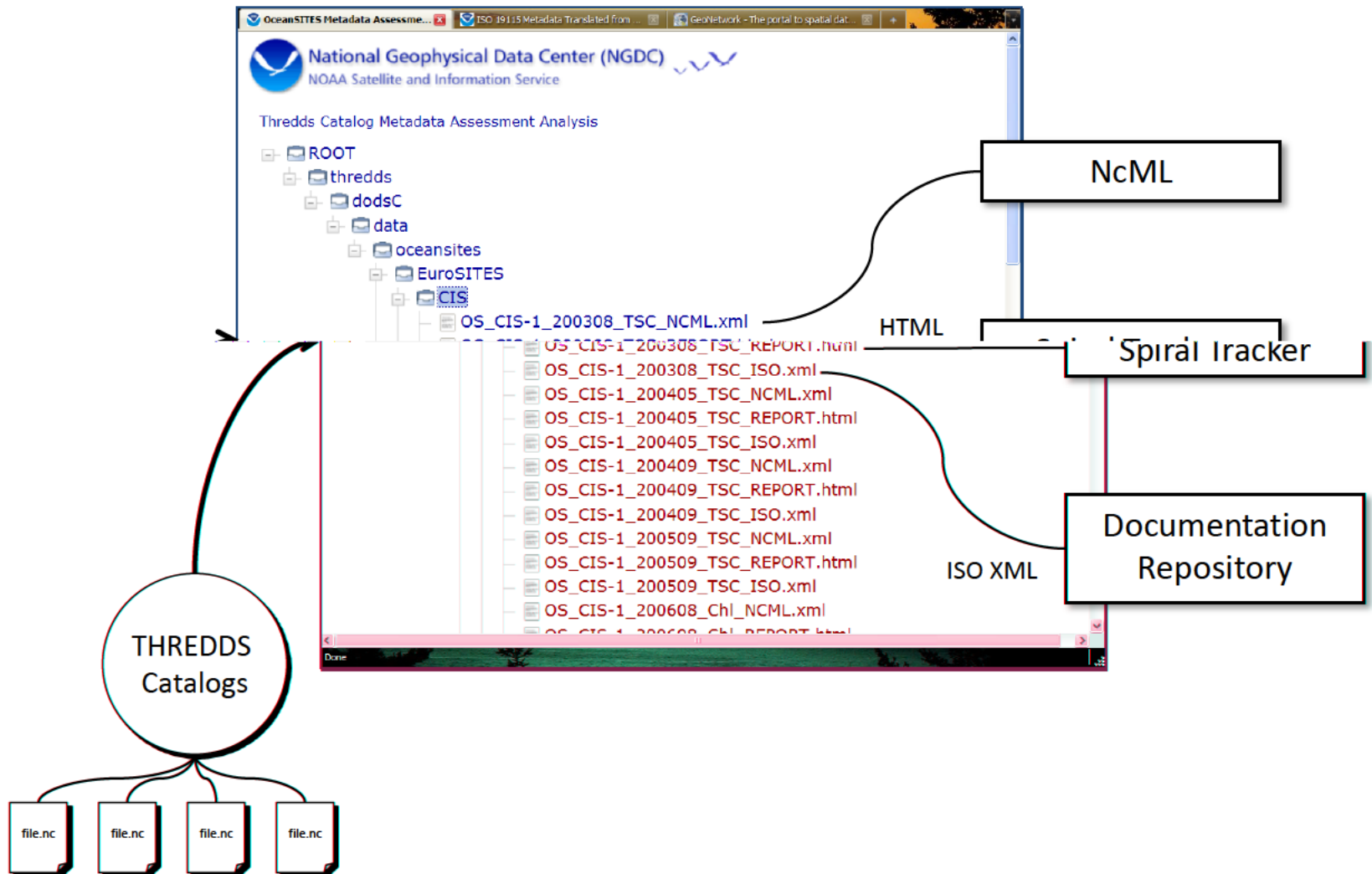
1. WCS: <http://sdf.ndbc.noaa.gov/thredds/wcs/hfradar>
2. OPENDAP: <http://sdf.ndbc.noaa.gov/thredds/doc>

Viewers:

- Integrated Data Viewer (IDV) (webstart)
- NetCDF-Java Tools (webstart)

#	Title	File Identifier	View
1	☆☆☆☆		FAQ, Rubric, KML, Text, XML, Discovery FGDC
2	1-deg x 1-deg Terrestrial Mean Free-Air Anomalies ☆☆☆☆	gov.noaa.ngdc:G01414	FAQ, Rubric, KML, Text, XML, Discovery FGDC
3	2.5-min gravity grid of N. America ☆☆☆☆	gov.noaa.ngdc:G01420	FAQ, Rubric, KML, Text, XML, Discovery FGDC
4	2.5-min Isostatic Gravity Grid for the United States ☆☆☆☆	gov.noaa.ngdc:G01447	FAQ, Rubric, KML, Text, XML, Discovery FGDC
5	30-min x 30-min Terrestrial Mean Free-Air Anomalies ☆☆☆☆	gov.noaa.ngdc:G01415	FAQ, Rubric, KML, Text, XML, Discovery FGDC
6	4th International Conference on Integrating GIS and Environmental Modeling ☆☆☆☆	gov.noaa.ngdc:G02160	FAQ, Rubric, KML, Text, XML, Discovery FGDC
7	5 Minute Gridded World Elevations - ETOPO5 Database ☆☆☆☆	gov.noaa.ngdc:G00931	FAQ, Rubric, KML, Text, XML, Discovery FGDC
8	AA Geomagnetic Activity Indices From Two Antipodal Observatories in Australia and England ☆☆☆☆	gov.noaa.ngdc:G00588	FAQ, Rubric, KML, Text, XML, Discovery FGDC
9	Age, spreading rates, and spreading asymmetry of the world's ocean crust ☆☆☆☆	gov.noaa.ngdc:G04150	FAQ, Rubric, KML, Text, XML, Discovery FGDC
10	Airborne Magnetic Trackline and Survey Data (Vector and Scalar Observations) ☆☆☆☆	gov.noaa.ngdc:G01150	FAQ, Rubric, KML, Text, XML, Discovery FGDC
11	Air Force Geophysics Laboratory Magnetometer Network ☆☆☆☆	gov.noaa.ngdc:G09997	FAQ, Rubric, KML, Text, XML, Discovery FGDC
12	Airglow Data ☆☆☆☆	gov.noaa.ngdc:G00603	FAQ, Rubric, KML, Text, XML, Discovery FGDC
13	Alaska1(ak1_iso) Gravity Data ☆☆☆☆	gov.noaa.ngdc:G01452	FAQ, Rubric, KML, Text, XML, Discovery FGDC
14	Alaska1(ak1_wpn) Gravity Data ☆☆☆☆	gov.noaa.ngdc:G01453	FAQ, Rubric, KML, Text, XML, Discovery FGDC
15	Alaska East-West Deflections (DEFLEC96) ☆☆☆☆	gov.noaa.ngdc:G01421	FAQ, Rubric, KML, Text, XML, Discovery FGDC
16	Alaska Geoid Heights (GEOID96) ☆☆☆☆	gov.noaa.ngdc:G01423	FAQ, Rubric, KML, Text, XML, Discovery FGDC
17	Alaska Gravity Data per 2 x 4 min Cell (96) ☆☆☆☆	gov.noaa.ngdc:G01425	FAQ, Rubric, KML, Text, XML, Discovery FGDC
18	Alaska North-South Deflections (DEFLEC96) ☆☆☆☆	gov.noaa.ngdc:G01422	FAQ, Rubric, KML, Text, XML, Discovery FGDC
19	Alaska Terrain Corrected Free Air Anomalies (96) ☆☆☆☆	gov.noaa.ngdc:G01424	FAQ, Rubric, KML, Text, XML, Discovery FGDC
20	Andes 1997 Gravity Data ☆☆☆☆	gov.noaa.ngdc:G01455	FAQ, Rubric, KML, Text, XML, Discovery FGDC
21	An Index (PC) Aimed at Monitoring the (P)olar (C)ap for Magnetic Activity ☆☆☆☆	gov.noaa.ngdc:G01118	FAQ, Rubric, KML, Text, XML, Discovery FGDC
22	ANWR and Alaska Peninsula Gravity Data ☆☆☆☆	gov.noaa.ngdc:G01457	FAQ, Rubric, KML, Text, XML, Discovery FGDC
23	Auroral Data ☆☆☆☆	gov.noaa.ngdc:G00594	FAQ, Rubric, KML, Text, XML, Discovery FGDC
24	Auroral Electrojet (AE, AL, AO, AU) - A Global Measure of Auroral Zone Magnetic Activity ☆☆☆☆	gov.noaa.ngdc:G00584	FAQ, Rubric, KML, Text, XML, Discovery FGDC
25	Auroral Electrojet Index Designed to Provide a Global Measure, Hourly Intervals, of Auroral Zone Magnetic Activity ☆☆☆☆	gov.noaa.ngdc:G00115	FAQ, Rubric, KML, Text, XML, Discovery FGDC
26	Auroral Electrojet Index Designed to Provide a Global Measure, 1-minute Intervals, of Auroral Zone Magnetic Activity ☆☆☆☆	gov.noaa.ngdc:G00114	FAQ, Rubric, KML, Text, XML, Discovery FGDC
27	Auroral Electrojet Indices Designed to Provide a Global Measure, 2.5-Minute Intervals, of Auroral Zone Magnetic Activity ☆☆☆☆	gov.noaa.ngdc:G00113	FAQ, Rubric, KML, Text, XML, Discovery FGDC
28	Bachelet Calculated UV-B Irradiance for Southern and Eastern Asia ☆☆☆☆	gov.noaa.ngdc:G01998	FAQ, Rubric, KML, Text, XML, Discovery FGDC
29	Bailey Ecoregions of the Continents (reprojected) ☆☆☆☆	gov.noaa.ngdc:G01990	FAQ, Rubric, KML, Text, XML, Discovery FGDC
30	Basin and Range Province, Western US, USGS Guide #1 ☆☆☆☆	gov.noaa.ngdc:G01204	FAQ, Rubric, KML, Text, XML, Discovery FGDC

Multiple Dialects: Documentation Extraction



Documentation in Three Dialects

The screenshot displays three overlapping browser windows. The top window shows a message: "This XML file does not appear to have any style information associated with it. The document tree is shown below." The middle window shows XML code for a netcdf location: `<netcdf location="dods://dods.ndbc.n1_200308_TSC.nc">` followed by dimension and attribute tags. The bottom window shows a metadata assessment report for "OceanSITES CIS in-situ data" with a total score of 16/43. The report includes a table of characteristics:

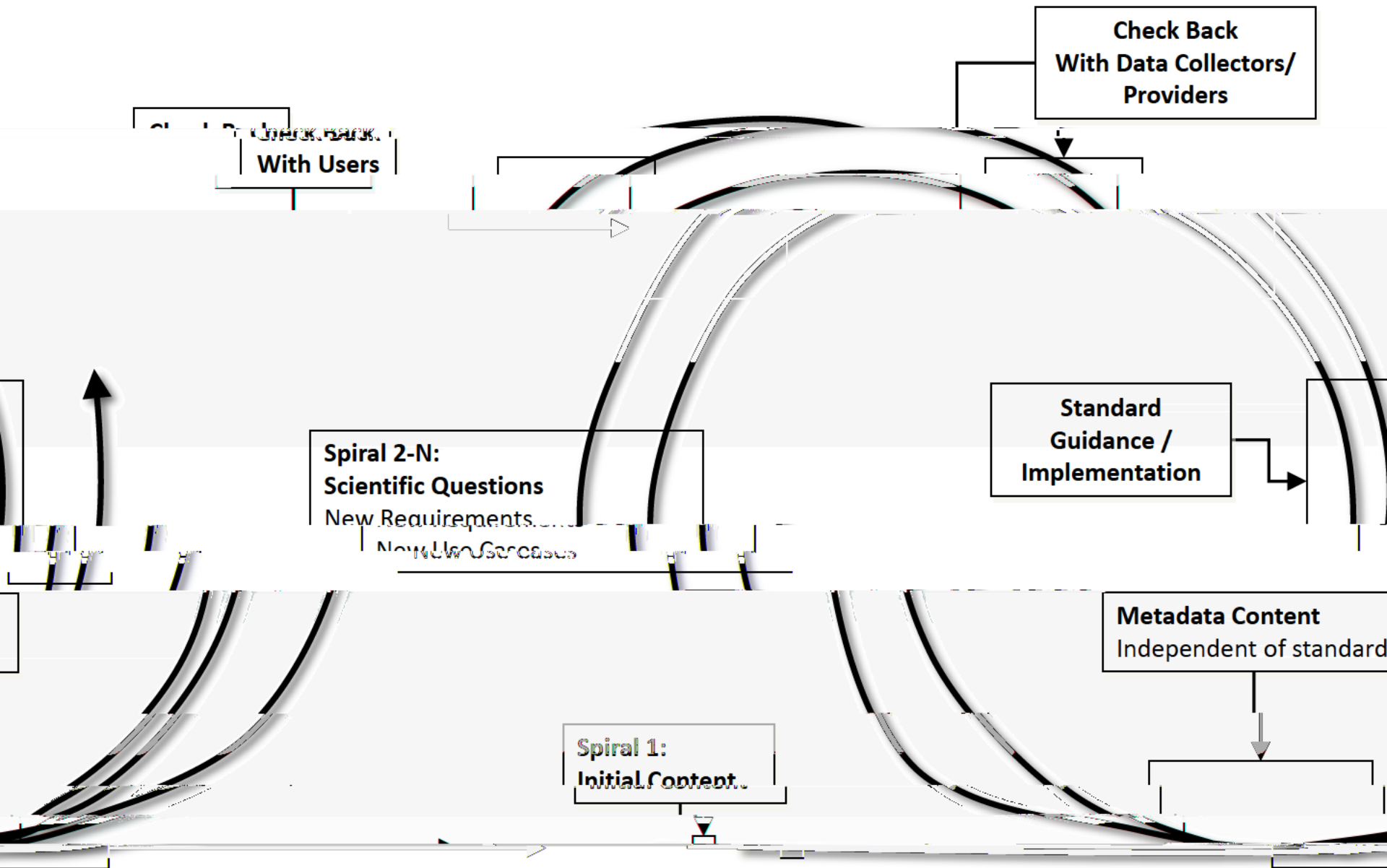
	Spiral	None	1-33%	34-66%	67-99%	All
Total				X		
Identification						X
Text Search				X		
Extent Search						X
Other Extent Information		X				
Creator			X			
Contributor		X				
Publisher		X				

NcML

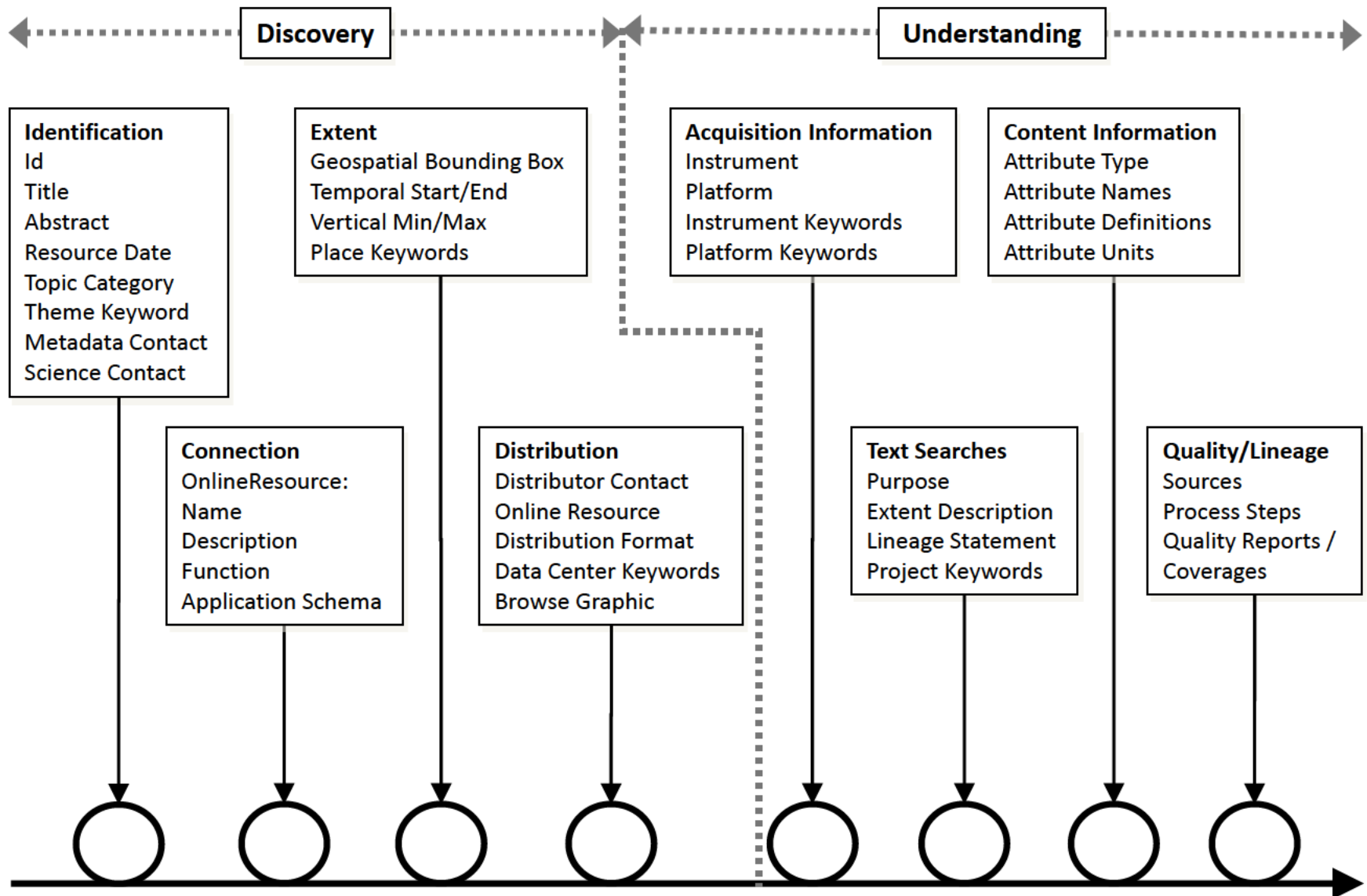
Documentation Repository

Spiral Tracker

Spiral Development / Training



Spiral Development / Training: Potential Spirals



Spiral Development / Training: Rubrics

The screenshot shows two overlapping browser windows. The background window displays the 'NetCDF Attribute Convention for Dataset Discovery Report' with a total score of 28/43. The foreground window displays the 'ISO 19115 Report' for 'Aerosol Optical Thickness (100 KM)' with a total spiral score of 27/41. The ISO report includes a detailed table of spiral scores across various categories.

NetCDF Attribute Convention for Dataset Discovery Report

The Unidata Attribute Convention for Data Discovery is a standard for describing files using standard metadata searches. This report identifies ISO metadata elements described in spirals of documentation development described in [Creating Good Documentation](#). Together these spirals build a strong foundation for high-quality documentation. The ISO Standard includes a number of options for building on that foundation by addressing specific scientific needs. See [Use Cases to CRUD](#) for some examples.

The elements are listed by name and are followed by M, C, or O if they are Mandatory, Conditional or Optional. They are followed by UDD (attribute name) if they are included in the [NetCDF Attribute Convention for Dataset Discovery](#).

The ISO 19115 Standard recommends [Core Elements](#) for inclusion in metadata. This tool tests also conformance with those recommendations.

The Rubric at the top of the report summarizes the results. Each spiral is represented by a row in the rubric. The columns show the % of the elements in that spiral that exist in the record. Click the spiral name for more details.

This report is produced using this [stylesheet](#). Please contact [Ted Habermann](#) if you have questions or suggestions.

Title: Chlorophyll-a, Aqua MODIS

Total Score: 28/43

General File Characteristics

Number of Global Attributes 60
 Number of Variables 5
 Number of Variable Attributes 45
 Number of Standard Names 5

Spiral	None
Total	
Identification	
Text Search	
Extent Search	
Other Extent information	
Creator	
Contributor	
Publisher	X
Other Attributes	

Identification Information

Done

ISO 19115 Report

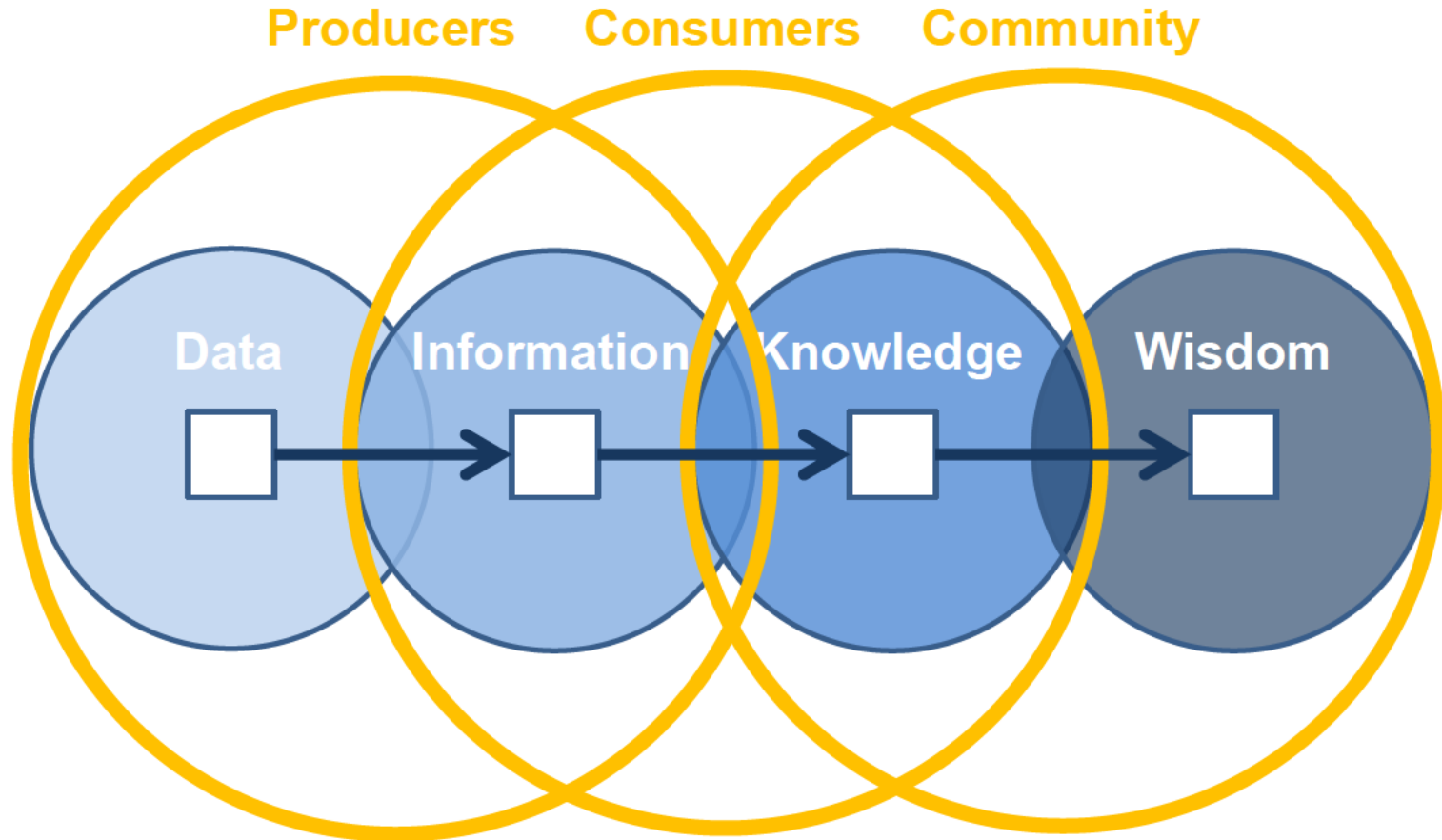
Title: Aerosol Optical Thickness (100 KM)

Total Spiral Score: 27/41

Spiral	None ☆☆☆☆	1-33% ★☆☆☆	34-66% ★★☆☆	67-99% ★★★☆	All ★★★★★
Total Spiral					
Identification					
Extent					
Connection					
Distribution					
Description					
Content					
Lineage					
Acquisition Information					

Done

Data and Information: End-to-End Process



Data to Information Concept Mapping

Variables and Properties
Multiple Dialects
Persistence vs. Transport

Standards and Conventions
Spiral Development
Spatial/Temporal Data Systems

Hierarchical Organizations
Training
Evolution

Questions?

