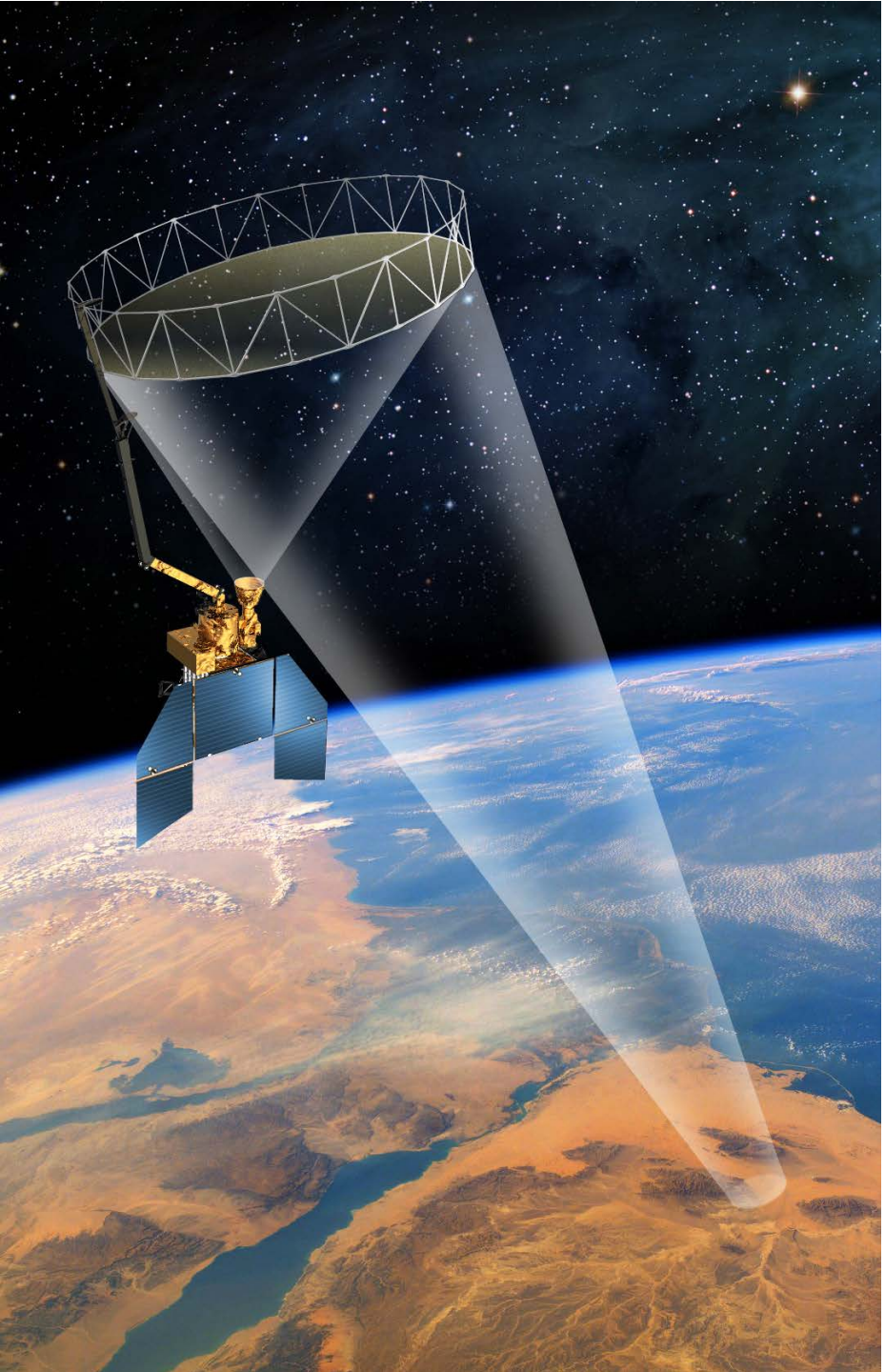


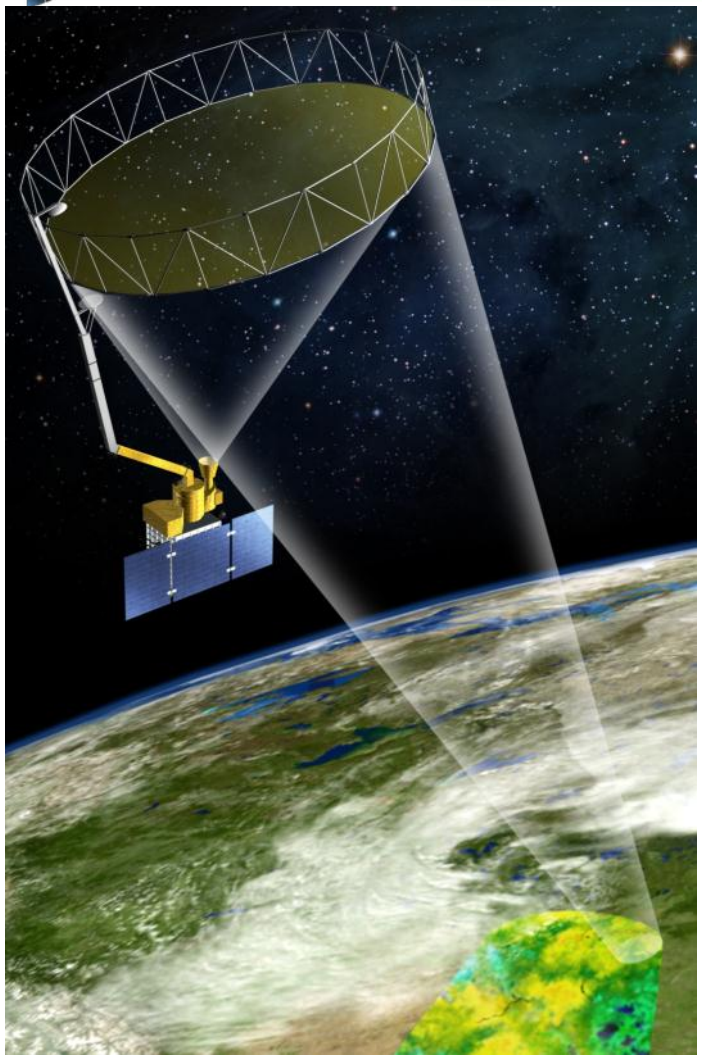
A Practical Application Using ISO Metadata –
Incorporating ISO Metadata into SMAP Data Products

Barry Weiss
Hook Hua
Vance Haemmerle

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA

November 28, 2012

SMAP Mission Overview



http://smap.jpl.nasa.gov/

Primary Science Objectives

- Global, high-resolution mapping of soil moisture and its freeze/thaw state to
  - Link terrestrial water, energy, and carbon-cycle processes
  - Estimate global water and energy fluxes at the land surface
  - Quantify net carbon flux in boreal landscapes
  - Extend weather and climate forecast skill
  - Develop improved flood and drought prediction capability

Mission Implementation

Partners	• JPL (project & payload management, science, spacecraft, radar, mission operations, science processing) • GSFC (science, radiometer, science processing)
Risk	• 7120-81 Category 2; 8705.4 Payload Risk Class C
Launch	• Oct. 2014 on Delta II system
Orbit	• Polar Sun-synchronous; 685 km altitude
Duration	• 3 years
Payload	• L-band SAR (JPL) • L-band radiometer (GSFC) • Shared 6-m rotating (13 to 14.6 rpm) antenna (JPL)

NRC Earth Science Decadal Survey (2007) recommended SMAP as a tier-one mission

Why SMAP and ISO

- SMAP is the first in a series of Decadal Survey Missions
- Level 1 Requirement for ISO metadata.
- Rationale for use of ISO
  - A common international metadata model
  - A common representation of the metadata
  - All NASA Earth Science data services would use the same standard
    - Data producers – missions
    - Data archives and distribution centers – DAACs
    - Clearing houses – ECHO
    - End users
      - For tool development and use

ISO Basic Concepts

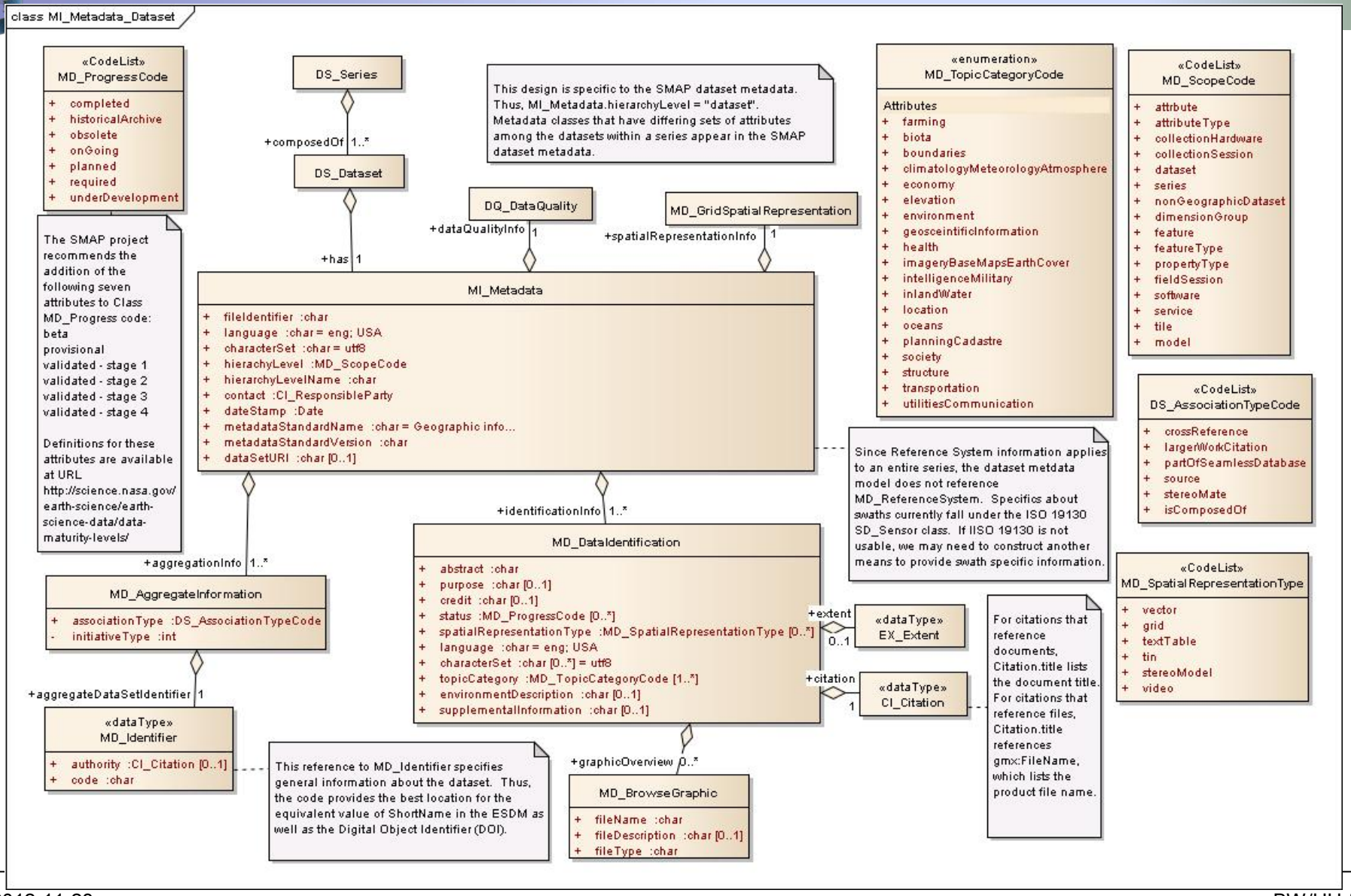
- Series metadata
  - Applies to a large set of data products
  - Analogous to “collection metadata”.
- Dataset metadata
  - Applies to the metadata that describe a single instance of a data product
  - Analogous to “granule metadata”
  - For SMAP: dataset metadata describes the entire file content
  - Metadata that describe individual elements appear in associated HDF5 attributes
    - Local Metadata – SMAP uses CF names to define these metadata elements.
- Codelists – Enumerated list with acceptable values defined in a schema
- Profile: Community agreement to make particular elements mandatory, or have more succinct definitions
- Extension: Explicit modifications to the model to add new elements or subclasses

- Class names begin with two characters followed by an underscore
  - CI_Citation, EX_Extent, LE_Source

ISO Geographic Standards

- ISO standards decouple the model from the encoding method
- Abstract Models
  - ISO 19115:2003
    - Original model - Used extensively in the SMAP model
  - ISO 19115-2:2009
    - Extensions for imagery and gridded data – Used extensively in the SMAP model
  - ISO 19115-1 Part 1: Fundamentals
    - Additional extensions - under review for adoption in 2013
  - ISO 19130:2010
    - Imagery sensor models for geopositioning – extracted one class as an extension
  - ISO 19157 Geographic information – Data quality
    - Under review for adoption in 2014
- Encoding rules to derive an XML schema for the ISO UML models
  - ISO 19118:2005, Geographic information — Encoding
- XSD Schema encoding
  - ISO 19139:2007 Geographic information – Metadata – XML schema implementation

ISO Model - UML



SMAP Approach to Dataset and Series Metadata

- Dataset metadata represented two ways
  - ISO 19139 compliant XML
    - Separate file for delivery to Data Center *with data product*
    - Included in the product for user extraction
  - HDF5 group/attribute structure within each product

- Series metadata represented the same two ways
  - ISO 19139 compliant XML
    - Separate file for delivery to Data Center *for each new release of any SMAP data product type*
    - Included in the product for user extraction
  - HDF5 group/attribute structure within each product

Group/Attribute Metadata Structure

- Products that employ the Earth Science Data Model (ESDM) place the product level metadata in a single HDF5 Metadata group
- SMAP adopted multiple sub-groups under the HDF5 Metadata group
- The HDF group/attribute structure provides a representation layer that is more user friendly
  - Groups represent major ISO classes
  - Attributes in the groups map to attributes in the ISO classes
  - In some instances, the design employs modified names of HDF5 groups or attributes to ease user comprehension of the model.
  - Group/attribute structure also reduces deeply nested layers within the HDF5 representation.
  - Design enables full conversion to compliant ISO 19139

- Product contains a large number of LI_Lineage/LE_Source classes. HDF5 group names reflect the product described in the group. Group names include:
  - Attitude, Ephemeris, Antenna Pointing, SCLK
- Renamed MI_Identifier used to specify **DOI** as *identifier_product_doi*.

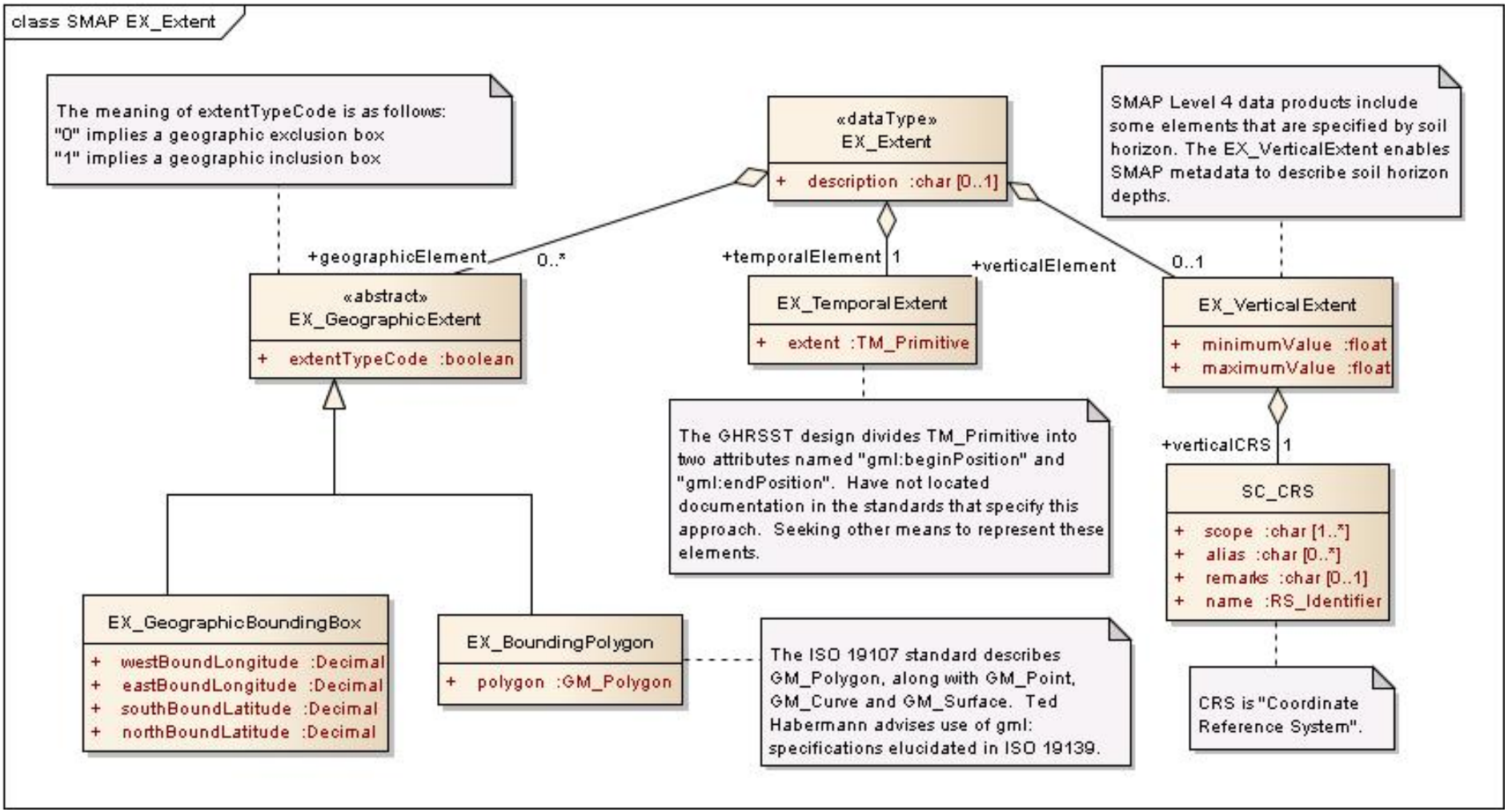
SMAP Rationale

- Both the ISO 19139 XML and the HDF group/attribute structure ultimately reflect the ISO model
  - Exclusive use of the ISO 19139 XML serialization requires the development of tools that enable science users to find the metadata they seek
    - Concern is use of metadata for research reference and algorithms
    - Use of data search applications is not as serious
    - Tools do not yet exist – mission can not rely on future availability of tools in time for SMAP launch
  - Finding metadata requires knowledge of
    - The ISO standard
    - The locations in the ISO standard where NASA stores particular data entries

Devising the ISO Model

- Design overall model based on ISO standard
  - UML diagrams that describe each of the major classes
- List the required metadata elements for each of SMAP's 15 data products
- Generate spreadsheets that map metadata elements to ISO equivalents or near equivalents
- Model correspondence to ESDM is not one-to-one
  - Some instances require different thinking to provide necessary information
  - Example:
    - ESDM provides fields that specify beginning and ending of gaps
    - ISO provides class which specifies begin and end times
      - Can employ more than one instance of the class in a single file

SMAP Implementation of EX_Extent Class



EX_Extent Class in ISO 19139


```
<gmd:extent>
        <gmd:EX_Extent>
          <gmd:description>
            <gco:CharacterString>The SMAP spacecraft downlinks SAR data that were acquired under following data conditions: 1) While acquiring forward looking and aft looking data over all land surfaces except Antarctica when the satellite is viewing the Earth at approximately 6 AM local time, 2) While acquiring forward looking data over all land surfaces North of 45 degrees North latitude when the satellite is viewing the Earth at approximately 6 PM local time, and 3) While acquiring forward looking data over coastal ocean regions when the satellite is viewing the Earth at approximately 6 AM local time. Coastal ocean regions are defined as ocean areas within 1000 km of land.</gco:CharacterString>
          </gmd:description>
          <gmd:geographicElement>
            <gmd:EX_GeographicBoundingBox id="swathBoundingBox">
              <gmd:extentTypeCode>
                <gco:Boolean>1</gco:Boolean>
              </gmd:extentTypeCode>
              <gmd:westBoundLongitude>
                <gco:Decimal>0.4685729</gco:Decimal>
              </gmd:westBoundLongitude>
              <gmd:eastBoundLongitude>
                <gco:Decimal>0.4689502</gco:Decimal>
              </gmd:eastBoundLongitude>
              <gmd:southBoundLatitude>
                <gco:Decimal>0.3194116</gco:Decimal>
              </gmd:southBoundLatitude>
              <gmd:northBoundLatitude>
                <gco:Decimal>0.3197749</gco:Decimal>
              </gmd:northBoundLatitude>
            </gmd:EX_GeographicBoundingBox>
          </gmd:geographicElement>
          <gmd:temporalElement>
            <gmd:EX_TemporalExtent>
              <gmd:extent>
                <gml:TimePeriod gml:id="swathTemporalExtent">
                  <gml:beginPosition>2015-05-30T16:01:00.000Z</gml:beginPosition>
                  <gml:endPosition>2015-05-30T16:01:06.003Z</gml:endPosition>
                </gml:TimePeriod>
              </gmd:extent>
            </gmd:EX_TemporalExtent>
          </gmd:temporalElement>
        </gmd:EX_Extent>
</gmd:extent>
```

EX_Extent Class Equivalent in Group/Attribute Structure

Metadata structure within the HDF5 Product

/Metadata

- /DatasetIdentification
- /DataQuality
- /Extent
  - /westBoundLongitude
  - /eastBoundLongitude
  - /northBoundLatitiude
  - /southBoundLatitude
  - /rangeBeginDateTime
  - /rangeEndDateTime
- /Lineage
- /AcquisitionInformation

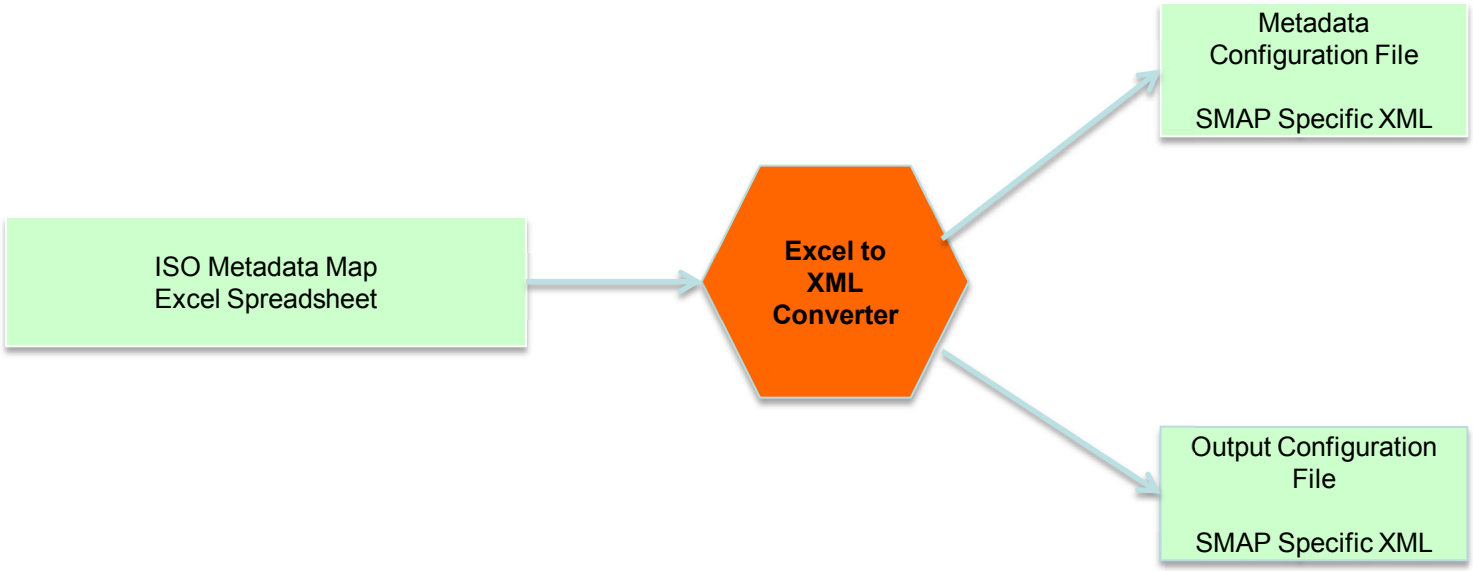
Mapping to the ISO Model

The example below is a simplification of the mapping exercise

Note that each element is mapped from a major ISO subclass to the ESDM and the HDF5 Group/Attribute structure

ISO Subclass Path	Data Type	Equivalent Element in ESDM	HDF5 Group/Attribute Location
EX_Extent/decription	char		Metadata/Extent/description
EX_Extent/geographicElement/EX_GeographicalBoundingBox/extentTypeCode	Boolean		Not needed in Group/Attribute - must transfer to ISO 19139
EX_Extent/geographicElement/EX_GeographicalBoundingBox/eastBoundingLongitude	Decimal	WestBoundingCoordinate	Metadata/Extent/westBoundLongitude
EX_Extent/geographicElement/EX_GeographicalBoundingBox/northBoundingLatitude	Decimal	NorthBoundingCoordinate	Metadata/Extent/eastBoundLongitude
EX_Extent/geographicElement/EX_GeographicalBoundingBox/southBoundingLatitude	Decimal	SouthBoundingCoordinate	Metadata/Extent/southBoundLatitude
EX_Extent/geographicElement/EX_GeographicalBoundingBox/westBoundingLongitude	Decimal	EastBoundingCoordinate	Metadata/Extent/northBoundLatitude
EX_Extent/temporalElement/EX_TemporalExtent/extent/timePeriod/beginPosition	DateTime	RangeBeginningDateTime	Metadata/Extent/rangeBeginningDateTime
EX_Extent/temporalElement/EX_TemporalExtent/extent/timePeriod/endPosition	DateTime	RangeEndingDateTime	Metadata/Extent/rangeEndingDateTime

Run Preparation



- ISO Metadata Map provides design as well as starting point for processing
- Metadata configuration file dictates the source of all metadata values. Executable then knows where to locate value for each metadata element.
- Output configuration file dictates the precise HDF5 content and structure of each SMAP data product.
- Both configuration files are in a SMAP specific flavor of XML

Series Metadata

- Series metadata are constant over a collection of granules
- SMAP Data Architect curates the series metadata for each data product type
  - Model is ISO 19115 compliant with a few SMAP extensions
  - Encoding is ISO 19139 compliant
  - One file represents a specific SMAP data product for each build
- The SMAP Science Data System delivers the curated series metadata to ESDIS with each build
  - This delivery enables ingestion of data products at the Data Centers
- SMAP software automatically incorporates the entire series metadata into a single HDF5 attribute in each data product granule

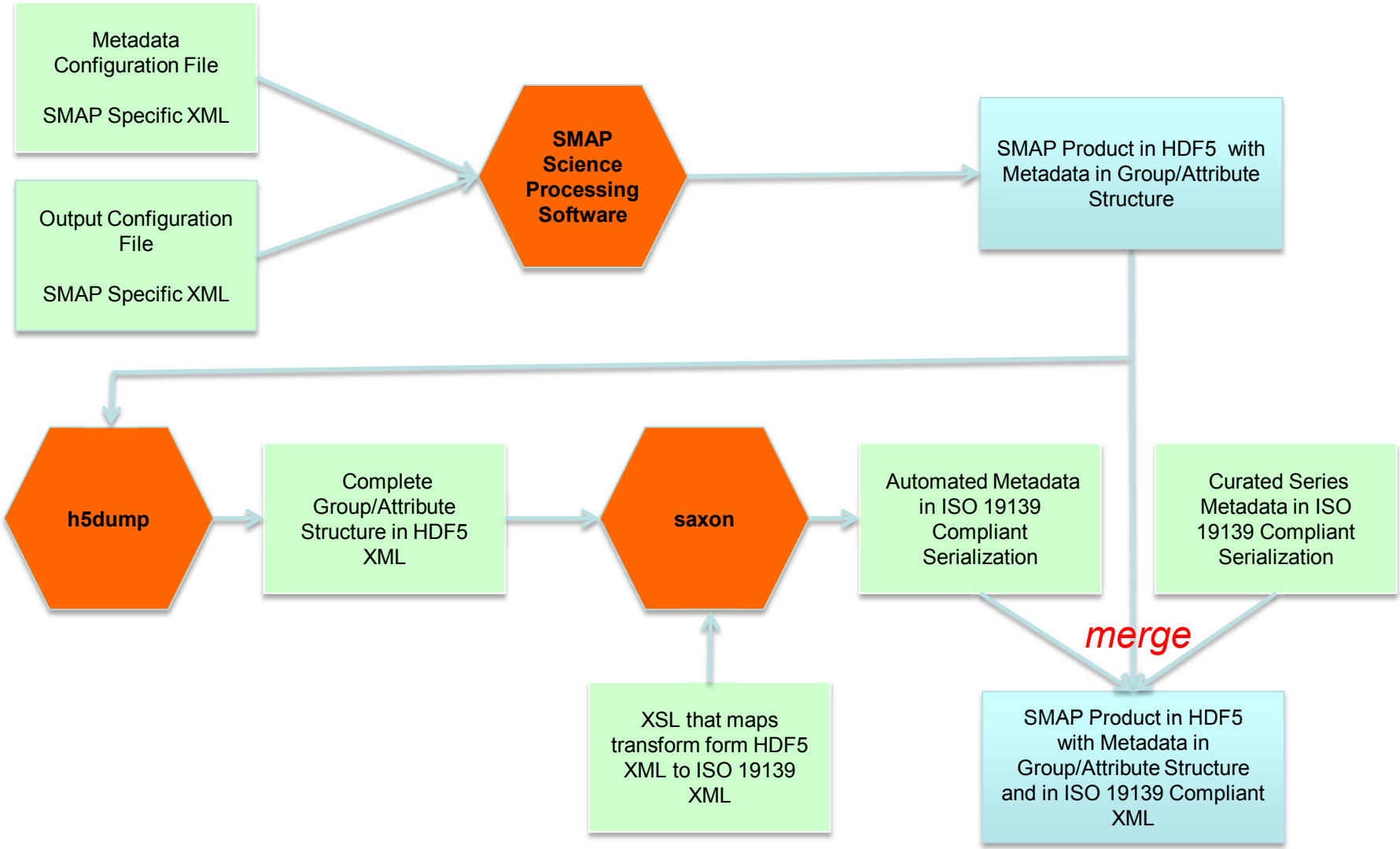
Dataset Metadata

- Dataset Metadata vary from instance to instance
  - Requires an automated implementation method
- Major challenge of SMAP implementation:
  - Automate the generation of dataset metadata
    - Metadata values vary from one granule to the next
    - Provide these values in group/attribute form
    - Provide same values using ISO 19139 compliant XML serialization
- Each data product generation executive generates dataset metadata in HDF5

Automating the Generation of ISO Metadata

- SMAP employs a standard framework in all executables that generate data products
  - One component of the framework writes the output into HDF5
- The framework HDF5 component enables automation of group/attribute metadata generation
  - Developers receive metadata definition from systems engineers in a mapping spreadsheet
  - Developers convert the spreadsheet content to configuration files
  - SMAP framework automates generation of metadata in the group/attribute structure
- **At a minimum, group/attribute approach only requires the standard HDF5 library**

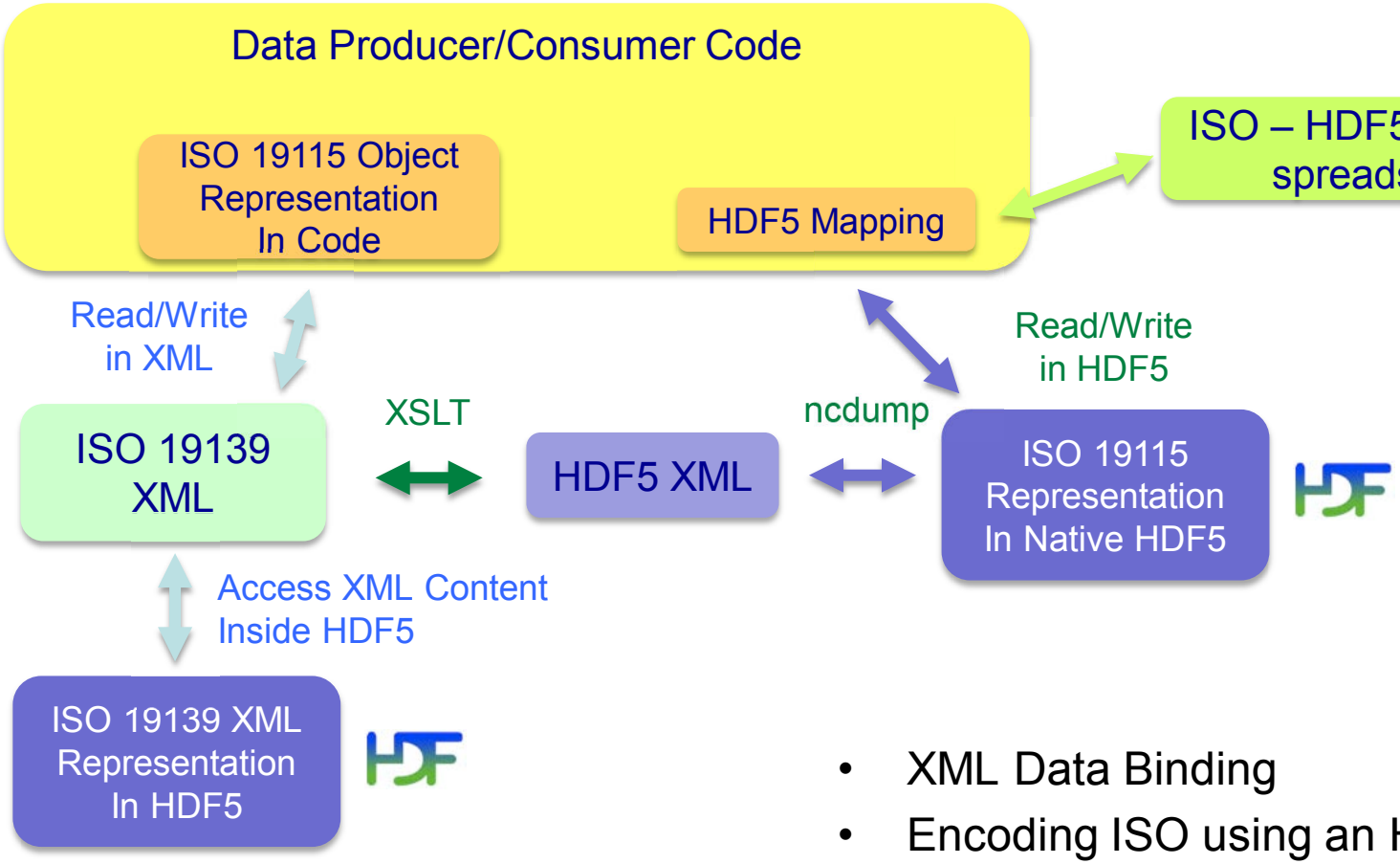
SMAP Tool Chain Flow



Automating the Generation of Dataset Metadata

- Each data product contains complete set of metadata in HDF5 group/attributes
- Developed an XSLT that maps the HDF5 group/attribute metadata into a representation that complies with ISO 19139 XML encoding
- Near the completion of each executable run, the SMAP software:
  - Dumps the HDF5 group/attribute metadata into HDF5 XML
  - Executes the Saxon XSLT engine to convert HDF5 XML to ISO 19139 XML
  - Incorporates the ISO 19139 compliant dataset metadata into an HDF5 attribute in the output data product granule
  - Incorporates the curated ISO 19139 series metadata into a separate HDF5 attribute
- The SMAP mission delivers the ISO dataset 19139 compliant metadata to the Data Centers in two forms
  - Embedded in the data product metadata for the user community
  - In a collocated file for Data Center ingestion
    - The separate file does not travel with the product

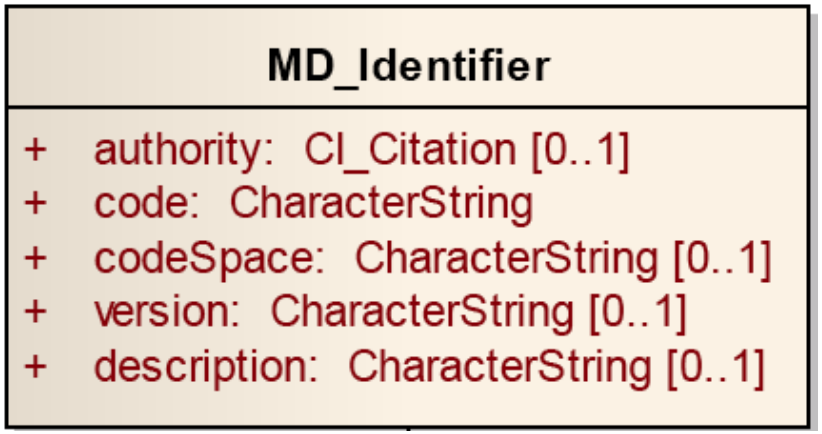
Tools for ISO Metadata Producers



- XML Data Binding
- Encoding ISO using an HDF5 API
- Automate mapping from design spreadsheets

Representing DOIs and UUIDs

- Like the ShortName, DOIs and UUIDs are identifiers
  - DOIs as collection/series identifier
  - UUIDs as granule/dataset identifier
- All of these elements appear in the MI_Identifier class.
  - SMAP extended the MD_Identifier Class so that it matches ISO 19115-1 specification



Additional SMAP Extensions to the ISO Model

- Need to specify orbit mechanical information
  - Lifted SD_OrbitMeasuredLocation class from ISO 19130
  - Aggregated SD_OrbitMeasuredLocation class under MI_Platform
- Need to specify various algorithm parameters that vary from one product to the next
  - Use of eos:EOS_Processing, which extends LE_Processing by adding additionalAttributeType and additionalAttributeValue.
  - Implemented an extended class named LI_AdditionalAttributes
  - Aggregated LI_AdditionalAttributes to class LE_Processing
  - Habermann/White suggest a Record/RecordType approach
- Most classes enable multiple citations
  - Citations can reference documents, files, algorithms or other entities
  - The ISO class LE_Algorithm includes only one citation
    - SMAP needs two references in the class:
    - One reference to provide a version for the algorithm
    - A second reference for the Algorithm Theoretical Basis Document

Extensions of the ISO Model

- Potential use of Record/Record Types to include Product Specific Metadata

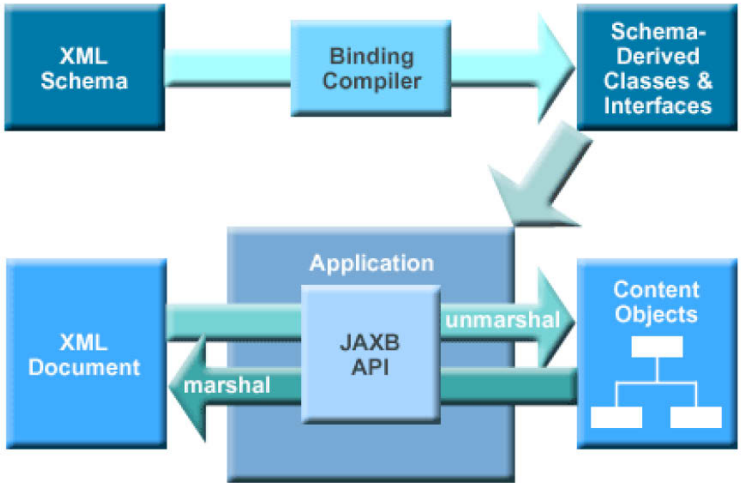
```
<gmd:lineage>
  <gmd:LI_Lineage>
    <gmd:processStep>
      <gmi:LE_ProcessStep>
        <gmd:description>
        <gmd:dateTime>
        <gmd:processor>
        <gmi:processingInformation>
          <!-- extends LE_Processing by adding
additionalAttributeType and additionalAttributeValue -->
          <eos:EOS_Processing>
            <gmi:identifier>
            <gmi:softwareReference>
            <gmi:procedureDescription gco:nilReason="unknown"/>
            <gmi:documentation>
            <gmi:runTimeParameters>
            <gmi:algorithm>
            <eos:otherPropertyType>
              <gco:RecordType
xlink:href="http://www.echo.nasa.gov/ingest/schemas/operations/Coll
ection.xsd#xpointer(//element[@name='AdditionalAttribute'])">Echo
Additional Attribute</gco:RecordType>
            </eos:otherPropertyType>
            <eos:otherProperty>
              <gco:Record
xmlns:echo="http://www.echo.nasa.gov/ingest/schemas/operations">
                <echo:AdditionalAttributes>

                </echo:AdditionalAttributes>
              </gco:Record>
            </eos:otherProperty>
          </eos:EOS_Processing>
        </gmi:processingInformation>
      </gmi:LE_ProcessStep>
    </gmd:processStep>
    …
  </gmd:LI_Lineage>
</gmd:lineage>
```

```
<echo:AdditionalAttribute>
  <echo:Name>RFIThreshold</echo:Name>
  <echo:DataType>float</echo:DataType>
  <echo:Description/>
  <echo:MeasurementResolution/>
  <echo:ParameterRangeBegin/>
  <echo:ParameterRangeEnd/>
  <echo:ParameterUnitsOfMeasure>degrees</echo:ParameterUnitsOfMeasure>
  <echo:ParameterValueAccuracy/>
  <echo:ValueAccuracyExplanation/>
  <echo:Value>5</echo:Value>
</echo:AdditionalAttribute>
<echo:AdditionalAttribute>
  <echo:Name>TimeVariableEpoch</echo:Name>
  <echo:DataType>float</echo:DataType>
  <echo:Description/>
  <echo:MeasurementResolution/>
  <echo:ParameterRangeBegin/>
  <echo:ParameterRangeEnd/>
  <echo:ParameterUnitsOfMeasure>degrees</echo:ParameterUnitsOfMeasure>
  <echo:ParameterValueAccuracy/>
  <echo:ValueAccuracyExplanation/>
  <echo:Value/>
</echo:AdditionalAttribute>
<echo:AdditionalAttribute>
  <echo:Name>epochJulianDate</echo:Name>
  <echo:DataType>float</echo:DataType>
  <echo:Description/>
  <echo:MeasurementResolution/>
  <echo:ParameterRangeBegin/>
  <echo:ParameterRangeEnd/>
  <echo:ParameterUnitsOfMeasure/>
  <echo:ParameterValueAccuracy/>
  <echo:ValueAccuracyExplanation/>
  <echo:Value>2.45154e+06</echo:Value>
</echo:AdditionalAttribute>
<echo:AdditionalAttribute>
  <echo:Name>epochUTCDateTime</echo:Name>
  <echo:DataType>dateTime</echo:DataType>
  <echo:Description/>
  <echo:MeasurementResolution/>
  <echo:ParameterRangeBegin/>
  <echo:ParameterRangeEnd/>
  <echo:ParameterUnitsOfMeasure/>
  <echo:ParameterValueAccuracy/>
  <echo:ValueAccuracyExplanation/>
  <echo:Value>2000-01-01T11:58:55.816</echo:Value>
</echo:AdditionalAttribute>
```

Coupling of type and value needs change

Validation of ISO 19139 XML Instances

- XSD based validations
  - XML data binding tools can validate XML instances against the schema
  - Tests against the value type and the grammar
  - Adoption of extended schemas requires community agreement



- Schematron validation
  - Employs a rule based
  - Enhances specification of constraints
    - Expressed using an XPath
    - Deployable as XSLT code
    - Includes conditional constraints

```
<rule context="image">
 ...
  <assert test="count(width) = count(height)">
    Width and Height elements should be
balanced
  </assert>
</rule>

<rule context="width">
  <assert test="preceding::height or
following::height">
    A width should be accompanied by a height
  </assert>
</rule>
```

Limitations of HDF5 1.8 for ISO

- Representing lists of objects
  - No current support for multiple cardinality of groups
    - Inhibits a group/attribute structure with multiple instances of LE_Source
  - Can employ arrays of user defined types (UDTs)
    - Does not appear as a list of objects
    - Complicates nesting of object arrays
  - Could name subgroups appended with indices
    - Examples include LE_Source/1, LE_Source_1, LE_Source.1
- h5dump to XML provides limited support for UDTs
  - Not fully structured.
  - XSLT transforms cannot easily map into 19139 XML
- Upcoming updates to HDF5 XML may considerably ease this approach

Recommendations for the NASA Flavor

- Division between Series and Dataset categories
  - Some classes such as MI_AcquisitionInformation should be constant for an entire collection, thus Series metadata
    - Attributes within the class break the mold
    - Chose to place the entire class and sub-class under the Dataset category
- Establish standards to define and employ Record/RecordTypes
- Use of the ECHO schema will require changes
  - Decouple the type and value in echo:AdditionalAttribute
  - Enable referencing of remotely defined types
- Adaptation of codelists and enumerations for NASA use
- Units of measure
  - Need to adopt a convention for representation of units
- Namespace for ISO 19139 XML

Concerns Going Forward

- A NASA standard does not yet exist
  - SMAP is the first NASA mission to generate ISO metadata
    - Some relevant fields are not present in the ISO standards
    - SMAP extended ISO to provide necessary elements
  - NASA needs to adopt a standard flavor
    - Ensures that common values appear in common locations across all NASA Earth Data Products
    - This effort has begun
      - SMAP's ISO adoption provides an initial example for other new missions
      - John Moses's effort on DOIs,
      - Ted Habermann's work with ECHO
- SMAP products contain a subset of the ISO standard
  - SMAP only employs ISO 19115 and ISO 19115-2
    - To ease initial implementation, many optional attributes in these standards that offer the richness of ISO are not in the SMAP model
  - SMAP implementation does not include valuable upcoming features in ISO 19115-1 and ISO 19157

Concerns Going Forward

- Lack of effective schema checking tools
  - SMAP necessarily extended the ISO standard
  - Presence of extensions requires that modification to the standard schema
  - SMAP can reuse and extend existing validation tools
- Generation of ISO tools
  - Tools will anticipate specific attributes or some format that SMAP did not use
  - Performance with tools may generate a perception that the SMAP ISO model is deficient
- NASA missions are reluctant to adopt cutting edge information technologies

Incorporation of ISO into Remaining SMAP Products

- Ideally, this effort should be covered under the SMAP budget
  - Once the infrastructure is set, implementation should function smoothly
  - SMAP plans include time to generate metadata in each of these products
- Issues that might confront the SMAP Mission
  - SMAP has 15 major data products.
    - These products may have metadata requirements that don't fit the current model
  - Tasks required for metadata implementation.
    - Are these standard implementation or additional effort?
      - Generation of XSLT for each product
      - Modification of the metadata validation to handle unforeseen extensions
      - Curation of an initial set of series metadata for each product
  - Adaptation to NASA standard as it evolves
    - Can SMAP get away with freezing its metadata model?
    - What is a reasonable policy considering the adoption of ISO over time?
  - Adaptation to new technologies and tools as they become available

Lessons Learned

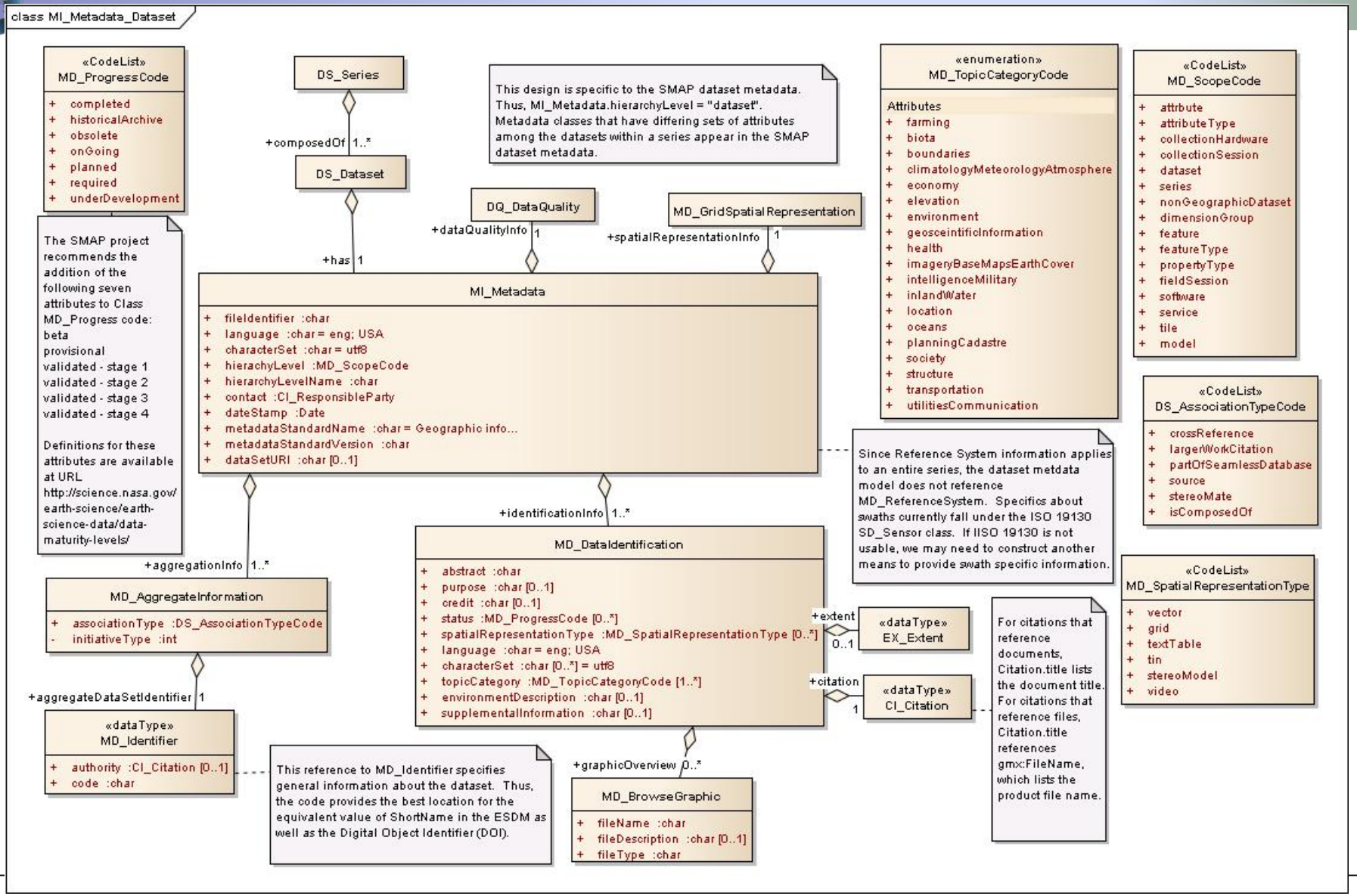
- Software development and applications
  - Product developers and users need time to ease into ISO
- Usability
  - The ISO model is deeply nested
  - Implicit structure of the ISO model requires users to drill deeper to locate specific elements they need
- Simplicity
  - Development teams need to keep support of ISO as simple as possible.
- Maturity of tools
  - ISO schemas for new models need to be made available earlier
  - Tool developers should leverage existing software and infrastructures
- Need to establish a NASA flavor
  - Usage conventions still need to be settled

Backup

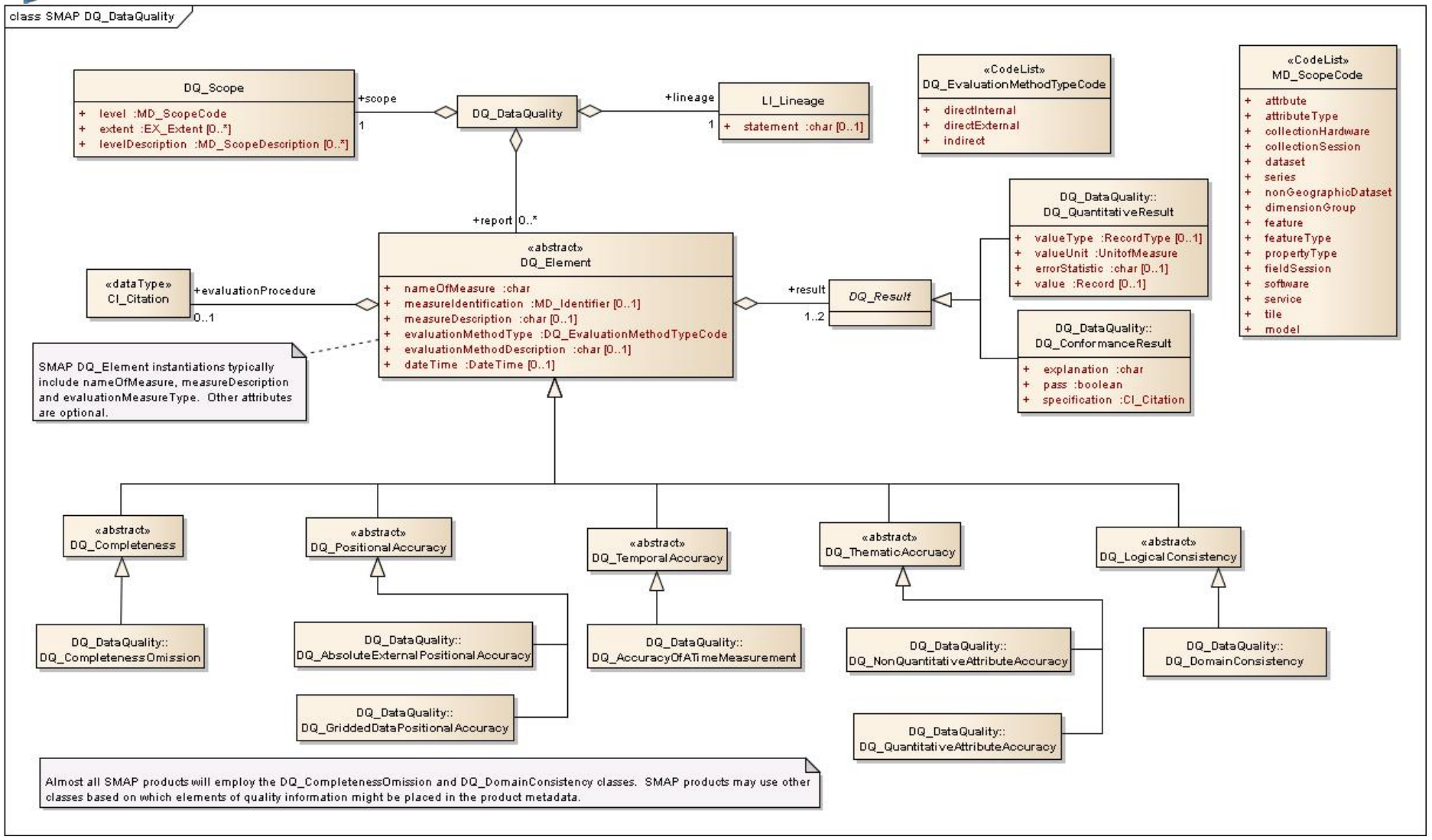
SMAP Data Products

Data Product Short Name	Description	Grid Resolution	Granule Extent
L1A_Radar	Parsed Radar Instrument Telemetry		Half Orbit
L1A_Radiometer	Parsed Radiometer Instrument Telemetry		Half Orbit
L1B_S0_LoRes	Low Resolution Radar σ_o in Time Order	5x30 km (10 slices)	Half Orbit
L1C_S0_HiRes	High Resolution Radar σ_o on Swath Grid	1 km	Half Orbit
L1B_TB	Radiometer T_B in Time Order	39x47 km	Half Orbit
L1C_TB	Radiometer T_B	36 km	Half Orbit
L2_SM_A	Radar Soil Moisture (includes Freeze-Thaw)	3 km	Half Orbit
L2_SM_P	Radiometer Soil Moisture	36 km	Half Orbit
L2_SM_AP	Active-Passive Soil Moisture	9 km	Half Orbit
L3_FT_A	Daily Global Composite Freeze/Thaw State	3 km	North of 45° N
L3_SM_A	Daily Global Composite Radar Soil Moisture	3 km	Global
L3_SM_P	Daily Global Composite Radiometer Soil Moisture	36 km	Global
L3_SM_AP	Daily Global Composite Active-Passive Soil Moisture	9 km	Global
L4_SM	Surface & Root Zone Soil Moisture	9 km	Global
L4_C	Carbon Net Ecosystem Exchange	9 km	North of 45° N

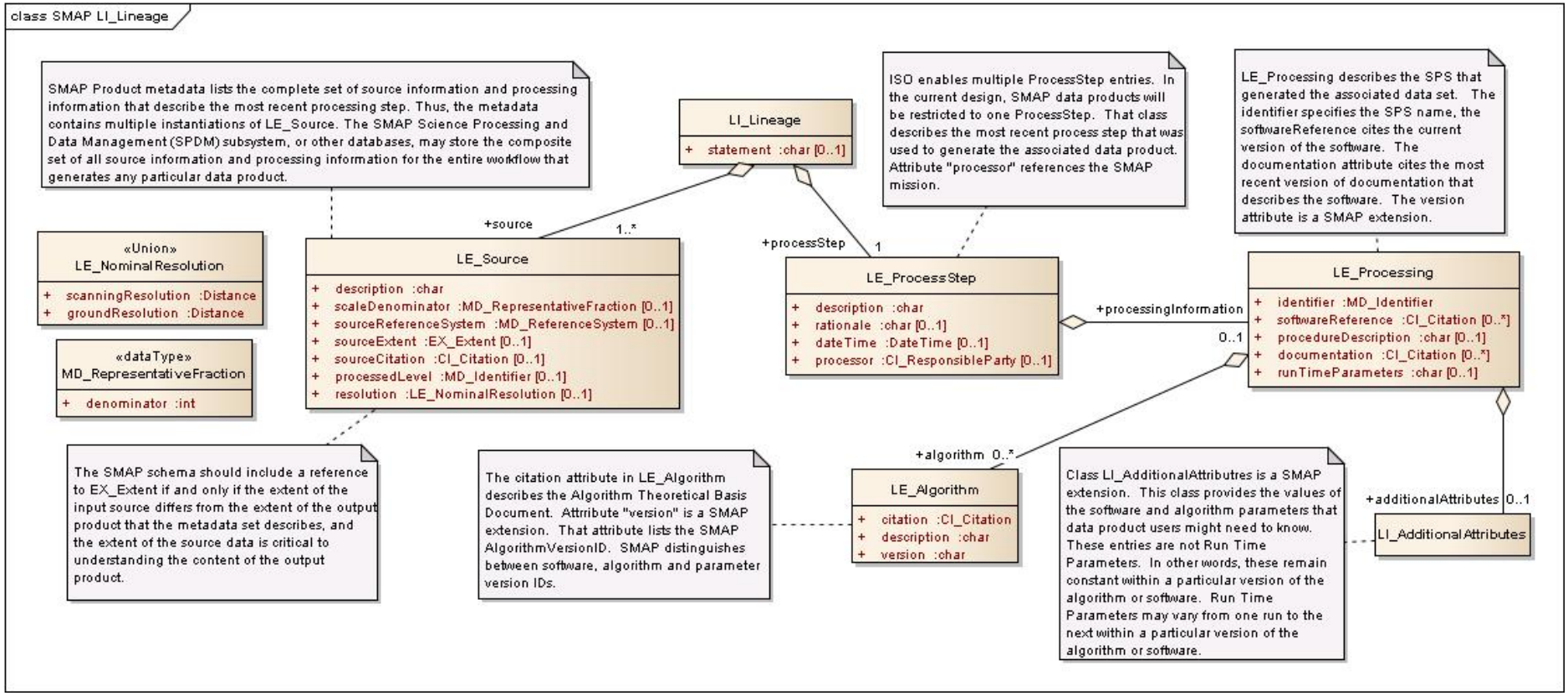
SMAP ISO Model – Dataset – MI_Metadata



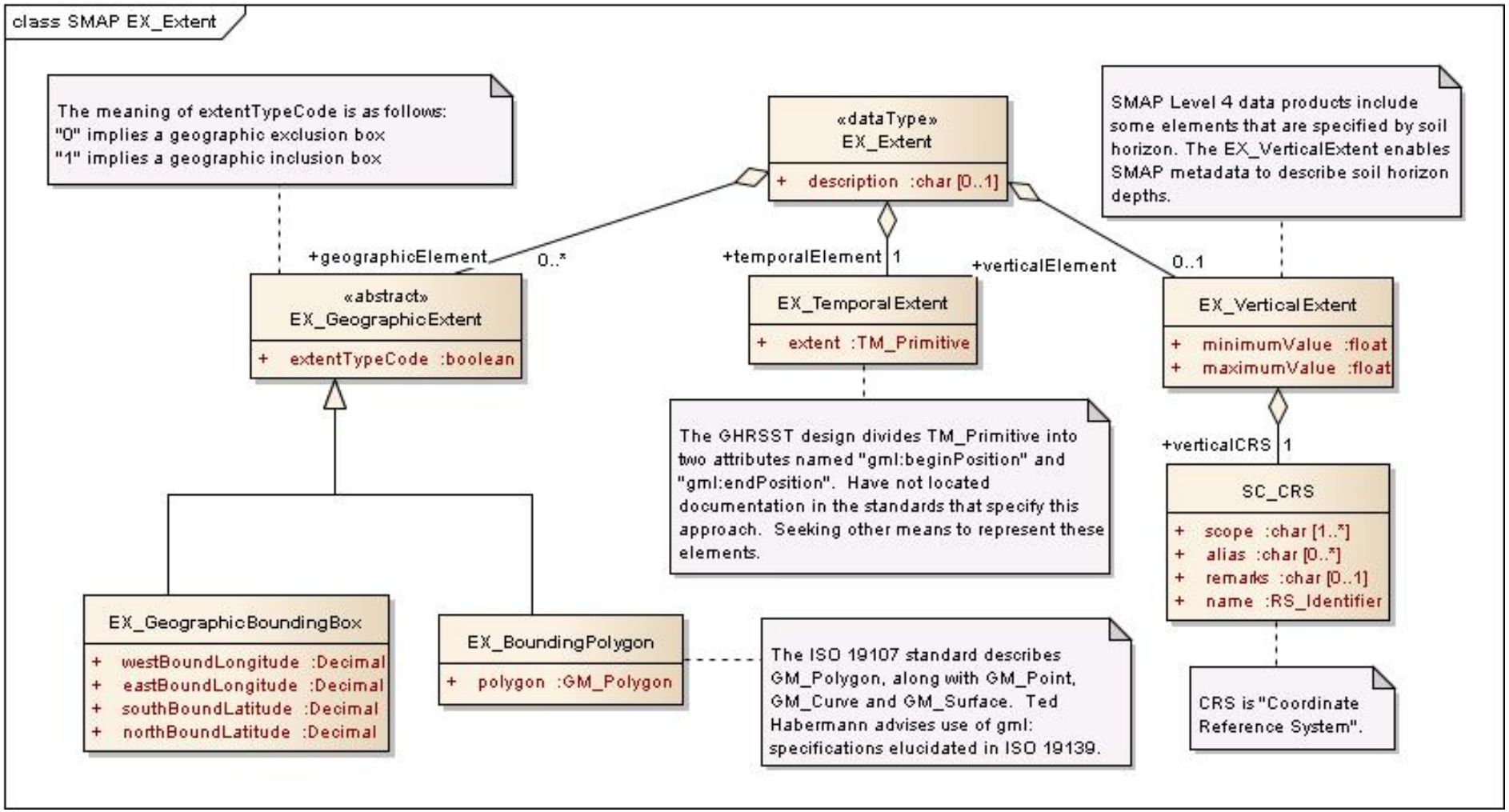
SMAP ISO Model – Dataset – DQ_Quality



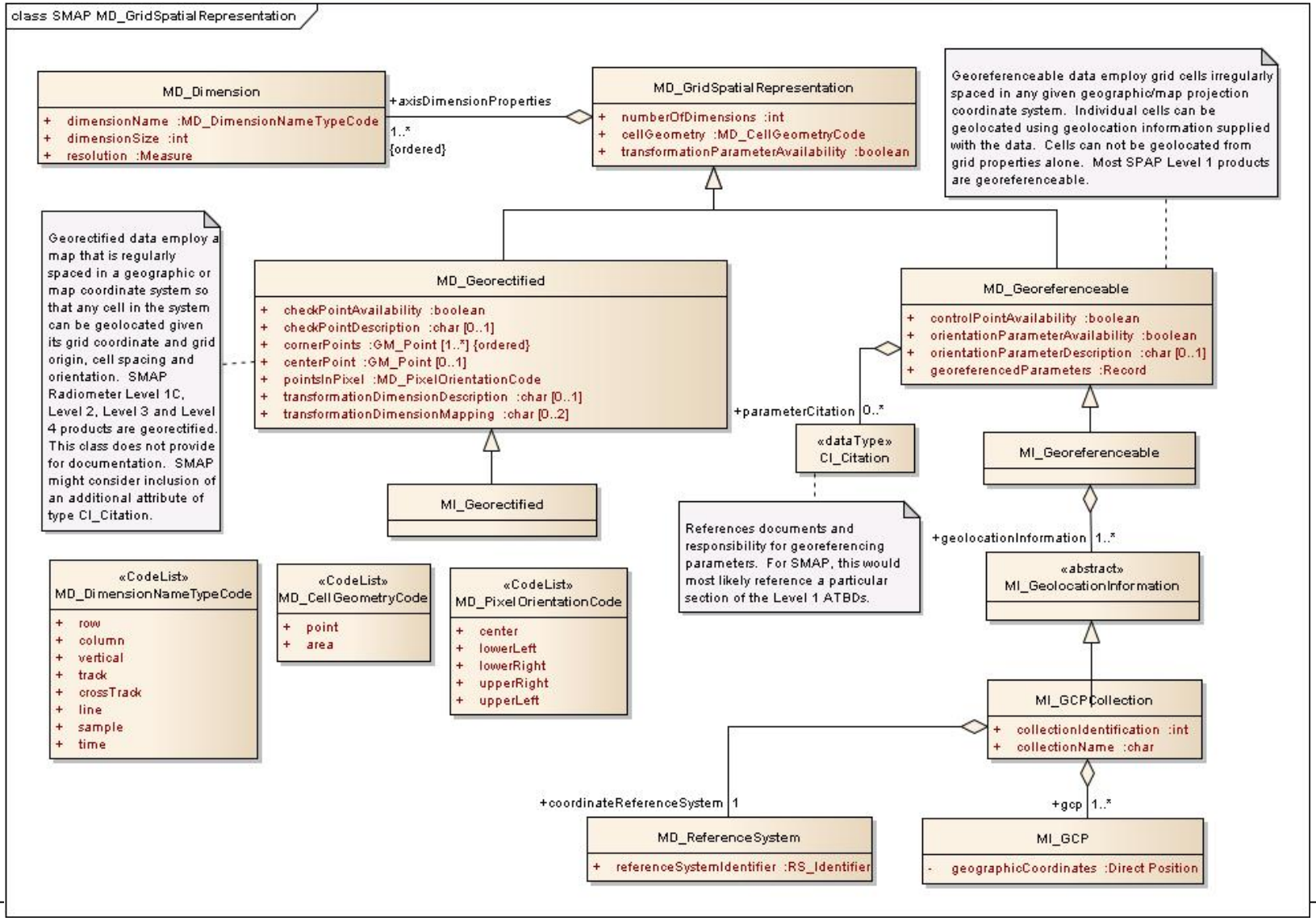
SMAP ISO Model – Dataset – LI_Lineage



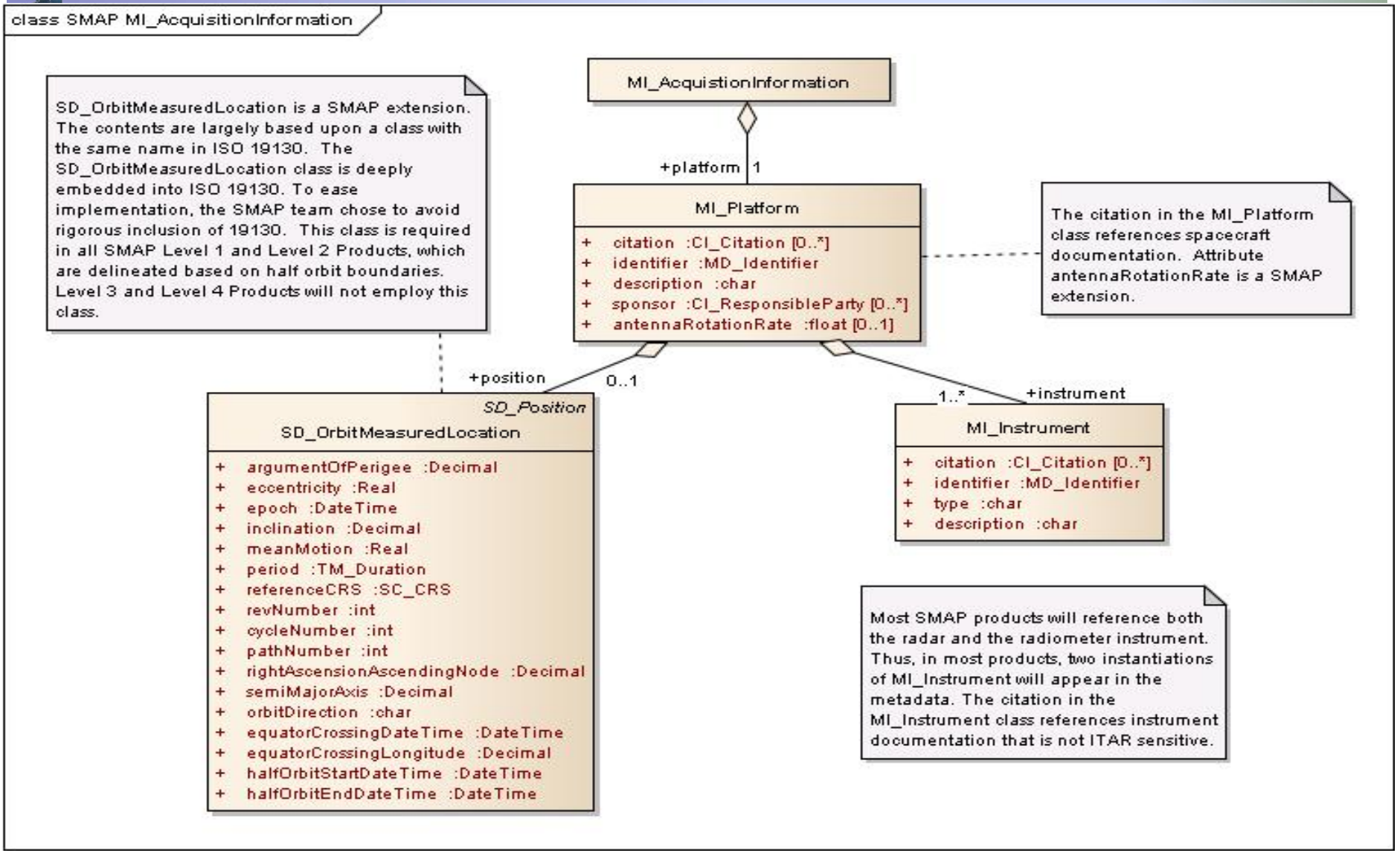
SMAP ISO Model – Dataset – Ex_Extent



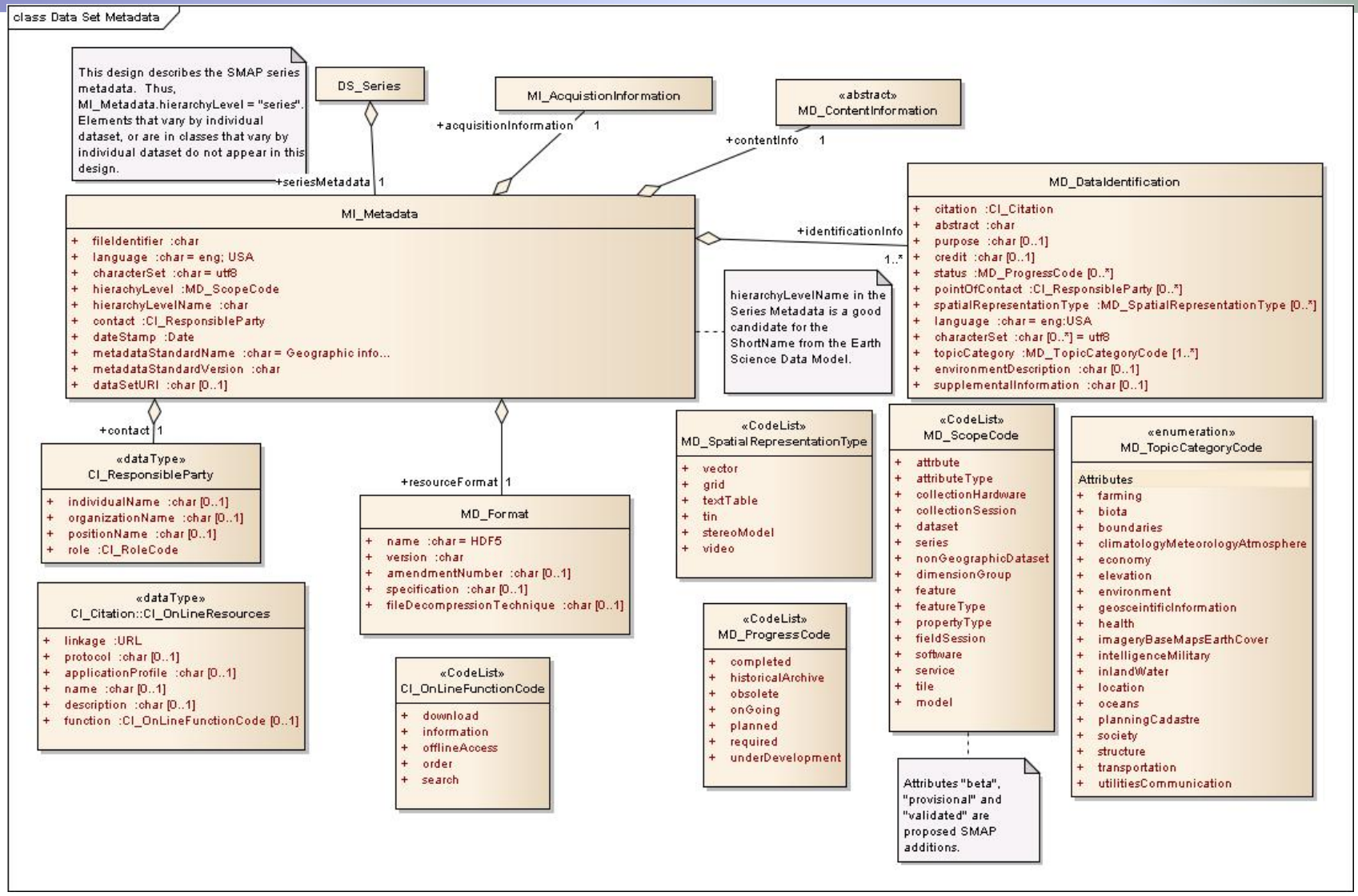
SMAP ISO Model – Dataset – MD_GridSpatialRepresentation



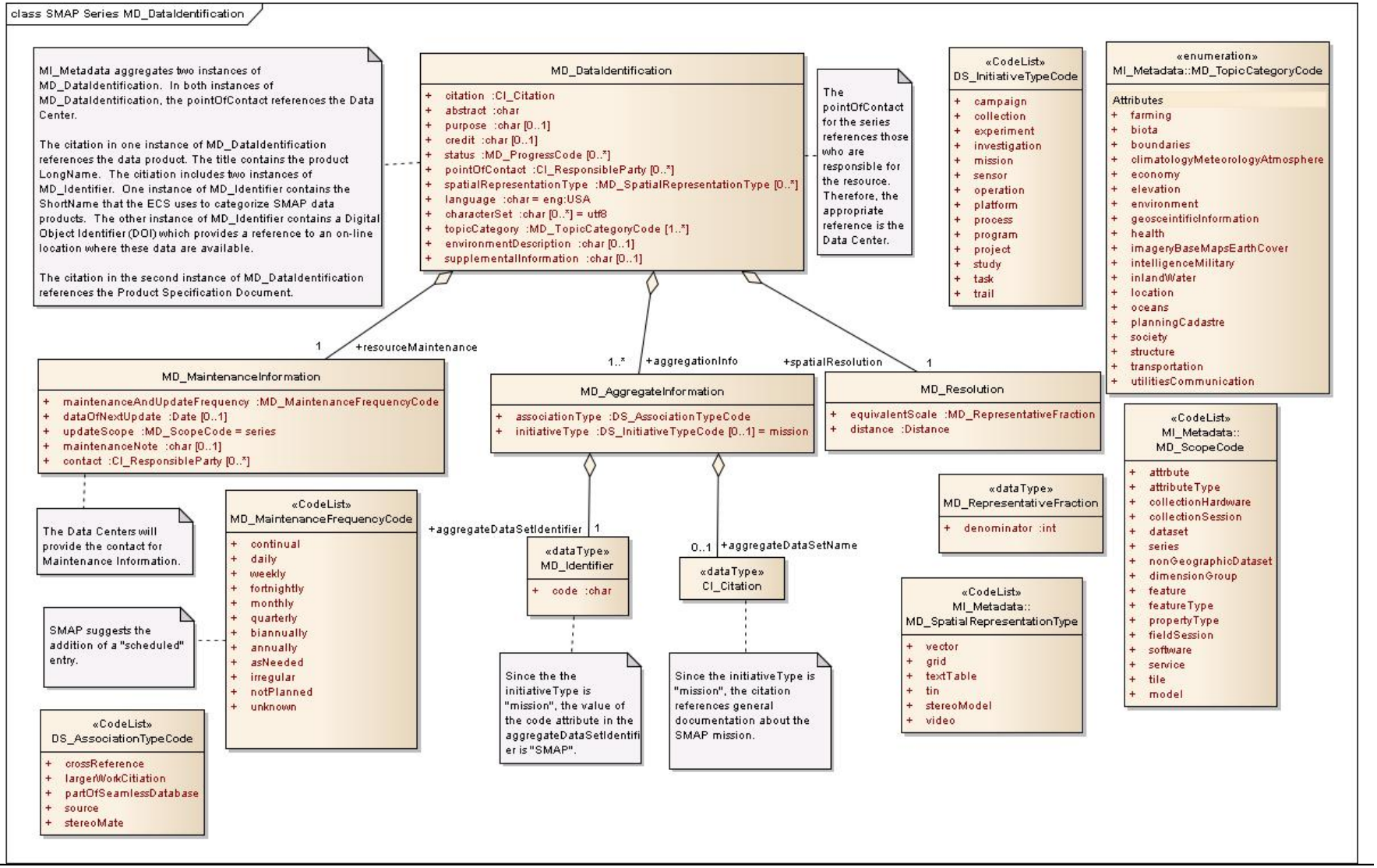
SMAP ISO Model – Data – MI_AcquisitionInformation



SMAP ISO Model – Series – MI_Metadata



SMAP ISO Model – Series – MD_DataIdentification



SMAP_ISO_Model – CI_Citation