Quality Elements in ISO 19115 Metadata Design for Proposed SMAP Data Products

Barry Weiss

Jet Propulsion Laboratory,
California Institute of Technology

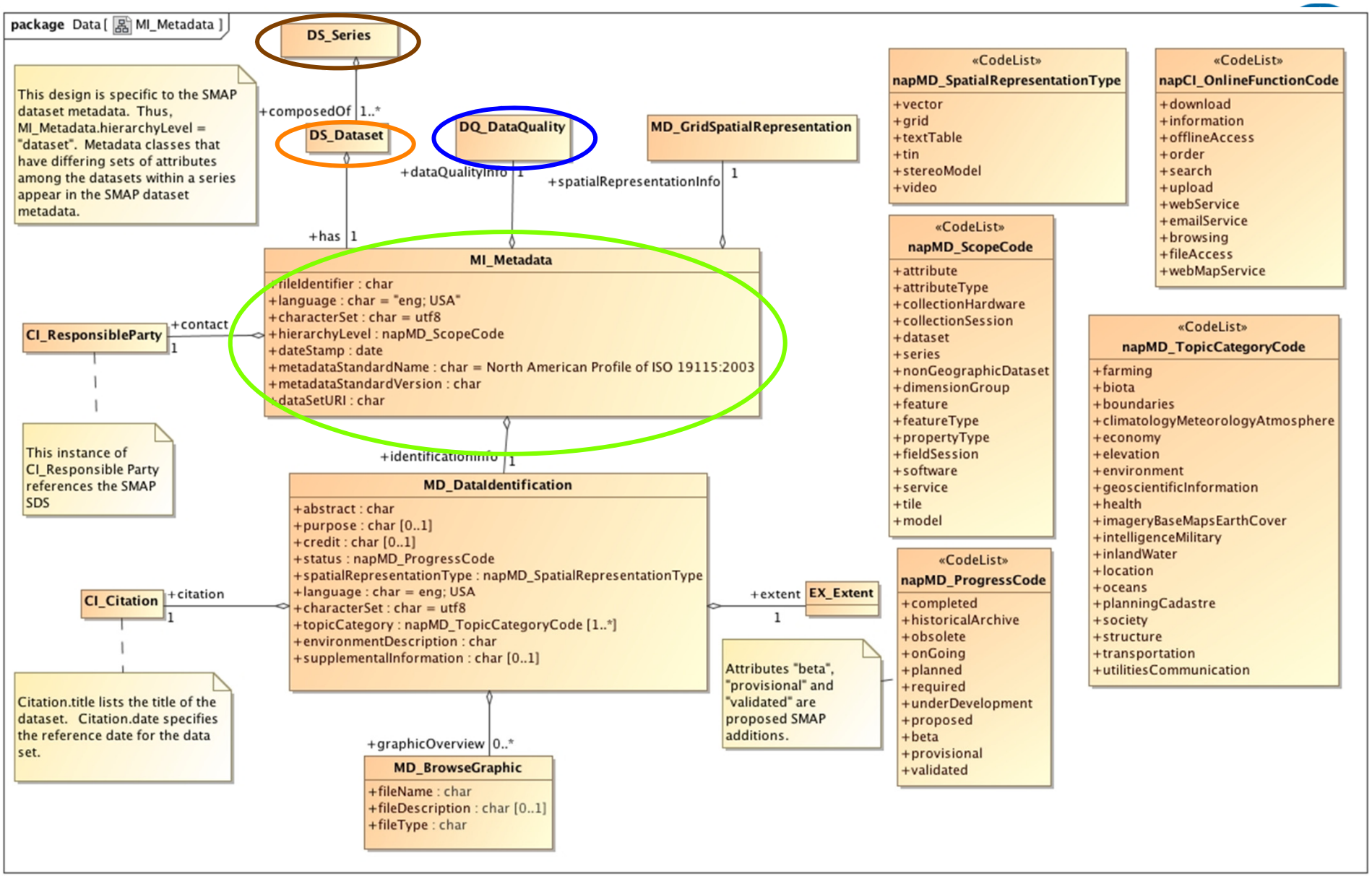
Copyright, 2011
California Institute of Technology
Government Sponsorship Acknowledged

Proposed SMAP Use of ISO

- Specified in Level 1 Requirements with NASA
- SMAP Engineers participated in the Metadata Evolution MENDS for NASA Earth Science Data Missions (MENDS) study
- Concluded that ISO would enhance SMAP Data Products
  - Use of standardized means to display metadata makes products enhances interoperability within and outside of NASA
  - ISO specifies storage of metadata in standard XML – ISO 19139 – standardization of the means to exchange metadata
  - Growth of international standard provides a far larger audience with the ability to read and interpret metadata
- Major caveat for implementation is the availability of resources for:
  - Generation of a SMAP schema
    - How that might fit with a NASA specific flavor of ISO
    - Coordination of the effort with other NASA entities and other missions
  - Generation of tools that
    - Automate generation of ISO metadata and
    - Automate capability to read ISO metadata
    - Enable users to seek and interpret available metadata

SMAP ISO Metadata Preliminary Design

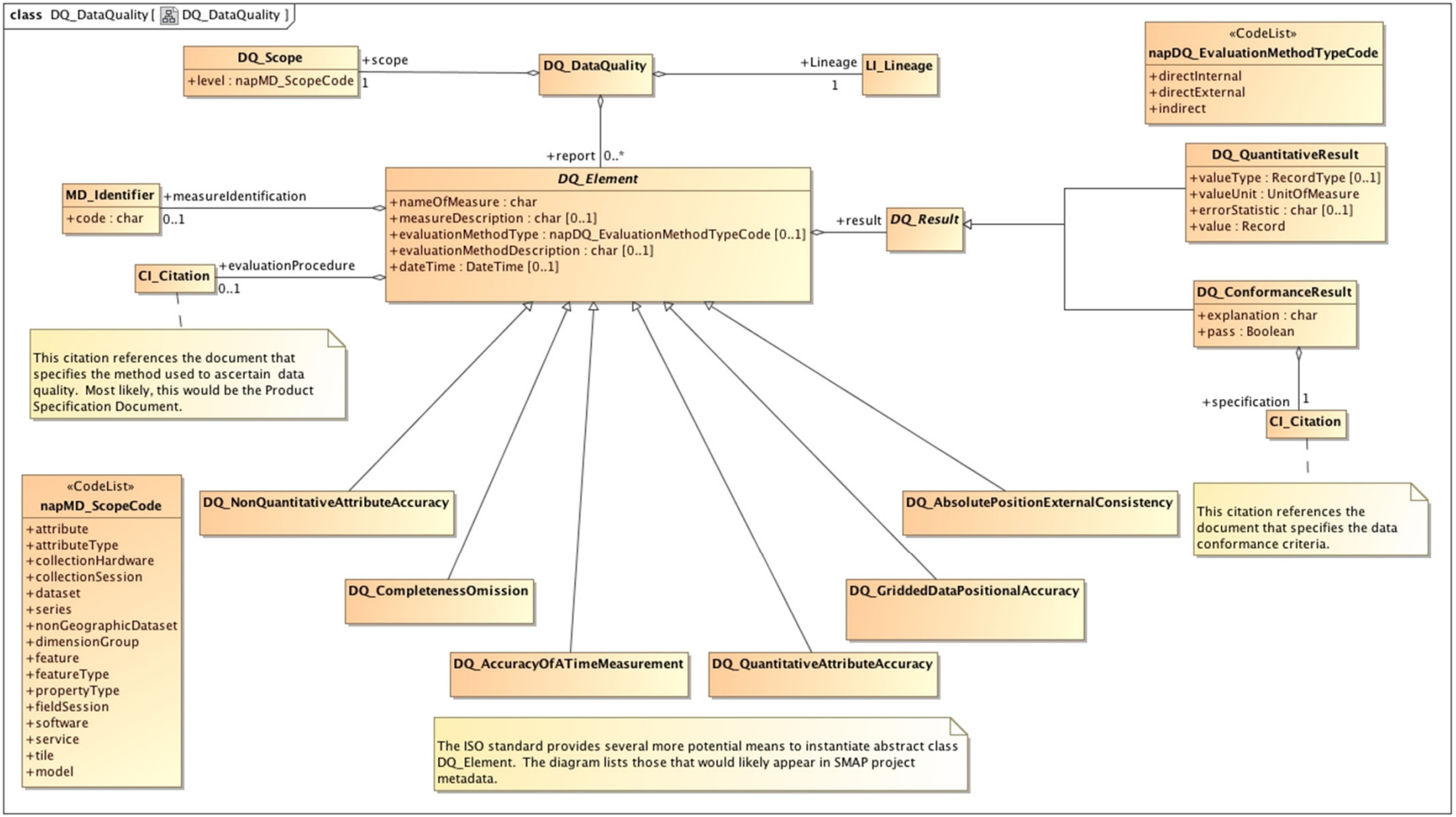
- The ISO design on the following slides is divided into two major structures
  - Dataset metadata - applies to each "granule" in a collection
    - The dataset metadata references the ISO quality classes
    - The design is based on the North American Profile of ISO 19115
      - Differences between the International Standard and the NAP are minor
      - The SMAP design will be modified to conform to the international standard
  - Series metadata - applies to a "collection" of data granules.
    - The SMAP series metadata design does not reference the ISO quality classes.
    - The ISO standard enables referencing quality metadata where single values or sets of values apply to an entire collection
      - Example - a quality metric determined during the cal/val phase of a mission. This value is not impacted by ongoing data processing of individual granules
- Diagrams do not include all classes and all attributes in the ISO standard
  - All classes and attributes that ISO specifies as mandatory are included
  - Design includes those elements that provide SMAP products with metadata that:
    - Provide users with adequate self description of the products
    - Enable transfer of search and discovery elements to appropriate file storage catalogs



Pre-decisional – for planning and discussion purposes only

ISO 19115 Quality Components

- DQ_Quality class aggregates a large number of component classes
  - DQ_Scope - defines the scope of the quality in terms of spatial and temporal extent as well as for a single dataset or a specific series of datasets
  - LI_Lineage - enables definition of the full set of input that are used to generate the product as well as the processing steps that take place to generate the product. (See slide 17)
  - DQ_Element - provides the major components of quality measurement
    - An abstract class with a multitude of potential child classes
      - SMAP design selected those employed in Earth Science Data Model
      - ISO enables considerable expansion on these quality measures
    - Measurement identification
      - Code value that specifies the type of quality measure acquired.
      - Description and name of measure
    - Quality evaluation descriptors
      - Method used to evaluate quality
      - Includes a citation that describes the evaluation procedure
    - Result measure - can be a quantitative result or a conformance result
      - Quantitative result contains specified outcome
      - Conformance result is a Boolean with an explanation and a citation



Pre-decisional – for planning and discussion purposes only

ISO 19157 Quality Additions and Changes

- DQ_MeasureReference and DQ_EvaluationMethod classes are aggregate classes in the DQ_Quality class
  - DQ_MeasureReference is designed to reference quality measures from a standardized list of data quality measures
    - ISO 19157 references a set of standard quality measures structured as DQM_Measures
  - DQ_EvaluationMethod can cite both an evaluation procedure and a reference document
- DQ_Result now incorporates a DQ_DescriptiveResult
  - Enables entry of qualitative evaluation
- Categorization of child classes of DQ_Element changed to mesh with DQM_Measures
- Entirely new child classes of DQ_Element
  - DQ_UsabilityElement.
    - Specifies data usability based on user requirements
  - DQ_Metaquality
    - Data confidence measure - trustworthiness of the data result
    - Data representativity measure - degree that the sample used produces a result that is representative of the entire population
    - Data homogeneity measure - expected of tested uniformity of the result

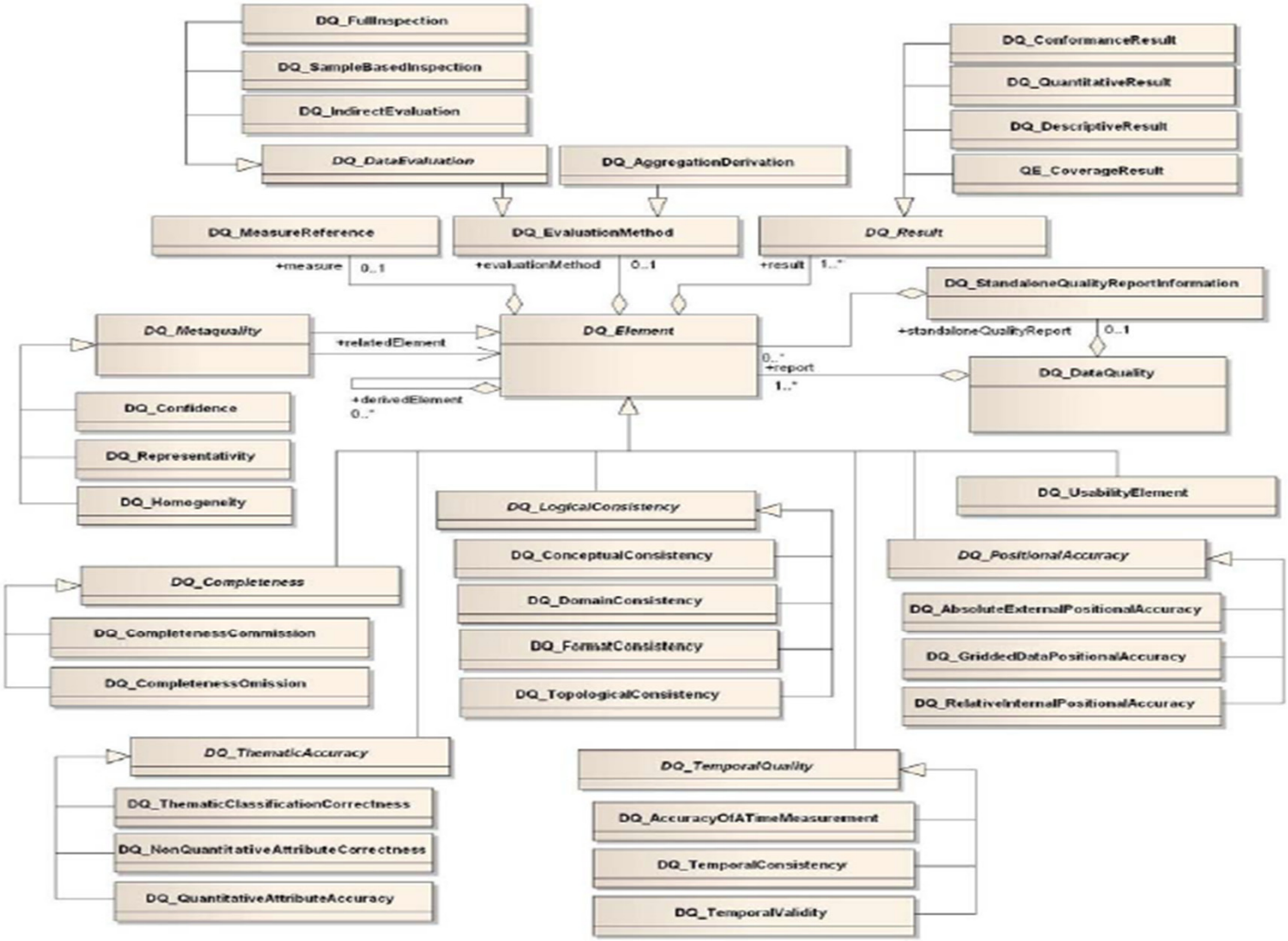
Model ISO 19157 DQM_Measure

Table D.7 — Rate of missing items

Line	Component	Description
1	Name	rate of missing items
2	Alias	–
3	Element name	omission
4	Basic measure	error rate
5	Definition	number of missing items in the dataset in relation to the number of items that should have been present
6	Description	–
7	Parameter	–
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	10% (The dataset has 10% less houses than the universe of discourse)
12	Identifier	7

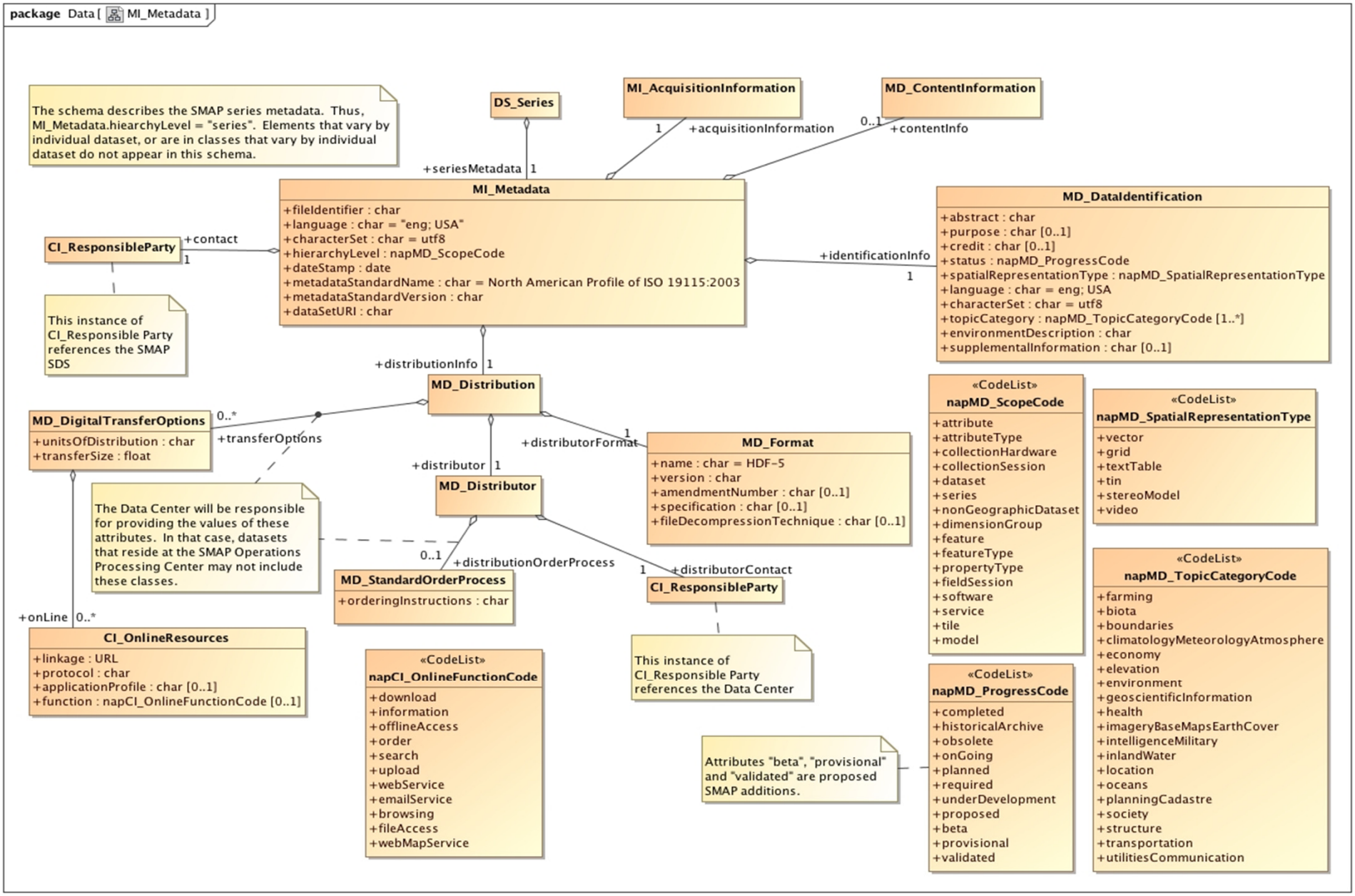
ISO 19157 DQM_Measure referenced from the DQ_Completeness_Omission class

Proposed ISO 19157 Components

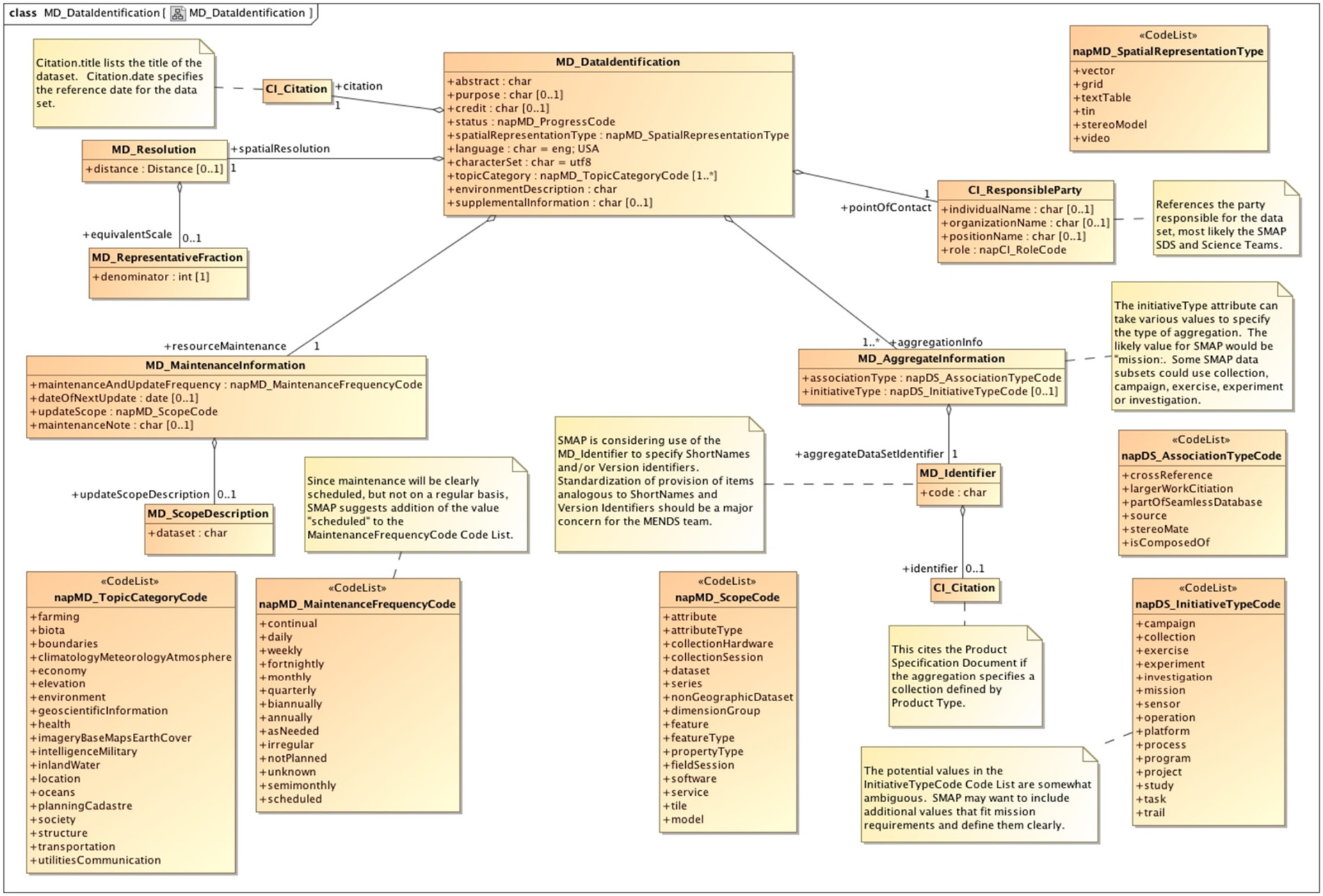


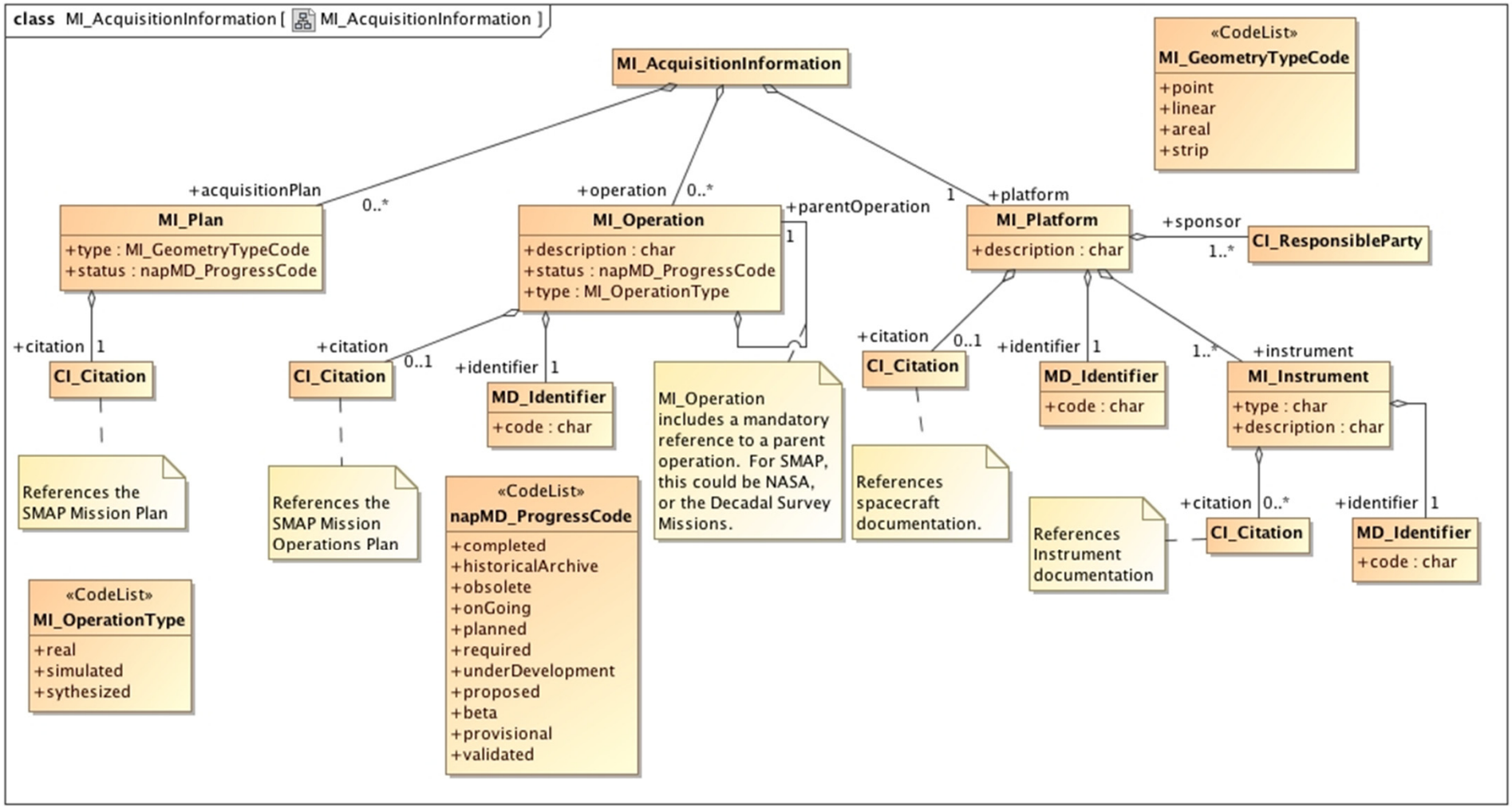
Proposed SMAP Series Metadata Design

Pre-decisional – for planning and discussion purposes only

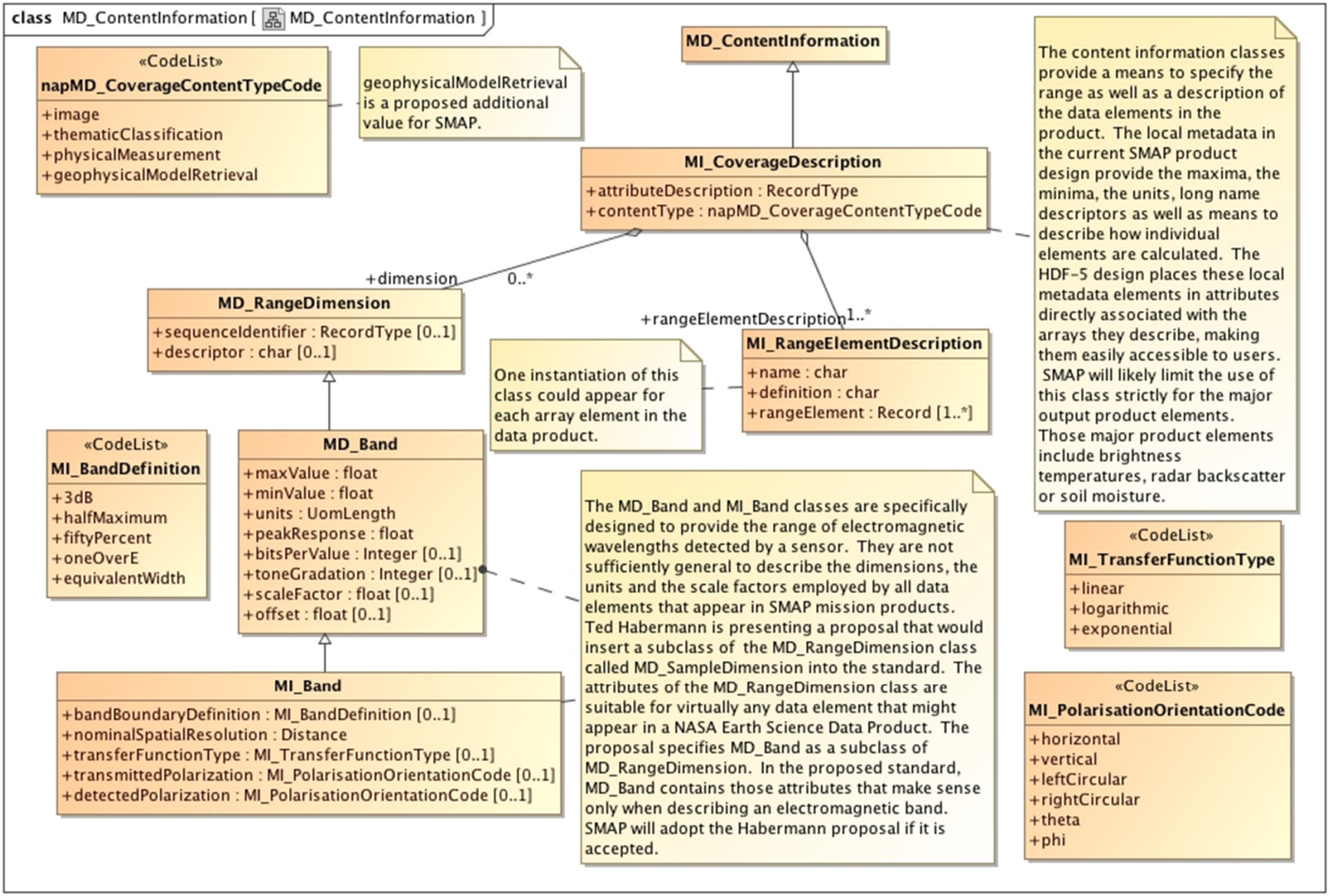


Pre-decisional – for planning and discussion purposes only

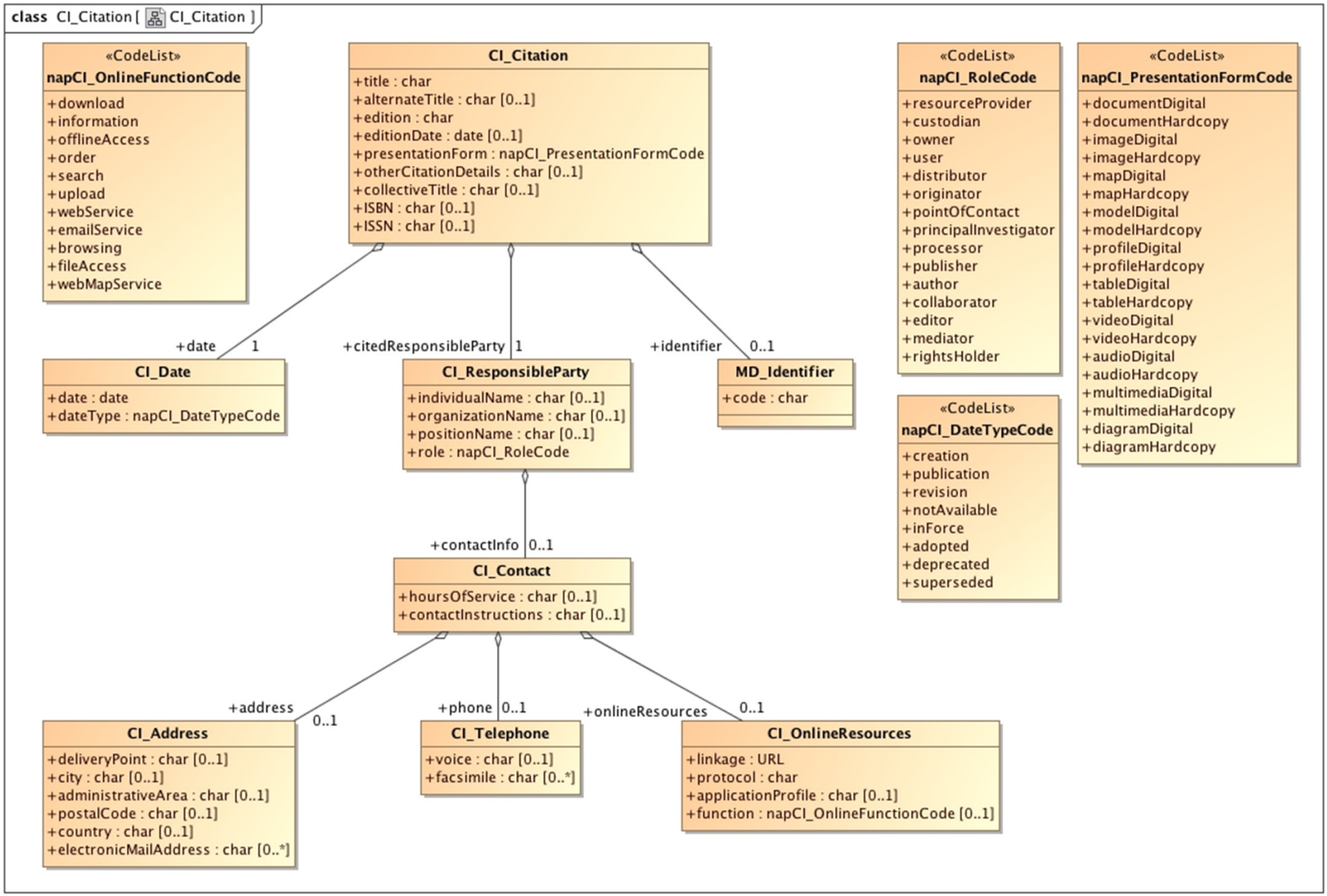

class MD_DataIdentification [MD_DataIdentification]
Citation.title lists the title of the dataset. Citation.date specifies the reference date for the data set.
CI_Citation
+citation
1
MD_DataIdentification
+abstract : char
+purpose : char [0..1]
+credit : char [0..1]
+status : napMD_ProgressCode
+spatialRepresentationType : napMD_SpatialRepresentationType
+language : char = eng; USA
+characterSet : char = utf8
+topicCategory : napMD_TopicCategoryCode [1..*]
+environmentDescription : char
+supplementalInformation : char [0..1]
«CodeList»
napMD_SpatialRepresentationType
+vector
+grid
+textTable
+tin
+stereoModel
+video
MD_Resolution
+distance : Distance [0..1]
+spatialResolution
1
+equivalentScale
0..1
MD_RepresentativeFraction
+denominator : int [1]
1
+pointOfContact
CI_ResponsibleParty
+individualName : char [0..1]
+organizationName : char [0..1]
+positionName : char [0..1]
+role : napCI_RoleCode
References the party responsible for the data set, most likely the SMAP SDS and Science Teams.
+resourceMaintenance
1
MD_MaintenanceInformation
+maintenanceAndUpdateFrequency : napMD_MaintenanceFrequencyCode
+dateOfNextUpdate : date [0..1]
+updateScope : napMD_ScopeCode
+maintenanceNote : char [0..1]
1..*
+aggregationInfo
MD_AggregateInformation
+associationType : napDS_AssociationTypeCode
+initiativeType : napDS_InitiativeTypeCode [0..1]
The initiativeType attribute can take various values to specify the type of aggregation. The likely value for SMAP would be "mission:. Some SMAP data subsets could use collection, campaign, exercise, experiment or investigation.
+updateScopeDescription
0..1
MD_ScopeDescription
+dataset : char
Since maintenance will be clearly scheduled, but not on a regular basis, SMAP suggests addition of the value "scheduled" to the MaintenanceFrequencyCode Code List.
SMAP is considering use of the MD_Identifier to specify ShortNames and/or Version identifiers. Standardization of provision of items analogous to ShortNames and Version Identifiers should be a major concern for the MENDS team.
+aggregateDataSetIdentifier
1
MD_Identifier
+code : char
«CodeList»
napDS_AssociationTypeCode
+crossReference
+largerWorkCitiation
+partOfSeamlessDatabase
+source
+stereoMate
+isComposedOf
«CodeList»
napMD_TopicCategoryCode
+farming
+biota
+boundaries
+climatologyMeteorologyAtmosphere
+economy
+elevation
+environment
+geoscientificInformation
+health
+imageryBaseMapsEarthCover
+intelligenceMilitary
+inlandWater
+location
+oceans
+planningCadastre
+society
+structure
+transportation
+utilitiesCommunication
«CodeList»
napMD_MaintenanceFrequencyCode
+continual
+daily
+weekly
+fortnightly
+monthly
+quarterly
+biannually
+annually
+asNeeded
+irregular
+notPlanned
+unknown
+semimonthly
+scheduled
«CodeList»
napMD_ScopeCode
+attribute
+attributeType
+collectionHardware
+collectionSession
+dataset
+series
+nonGeographicDataset
+dimensionGroup
+feature
+featureType
+propertyType
+fieldSession
+software
+service
+tile
+model
+identifier
0..1
CI_Citation
This cites the Product Specification Document if the aggregation specifies a collection defined by Product Type.
«CodeList»
napDS_InitiativeTypeCode
+campaign
+collection
+exercise
+experiment
+investigation
+mission
+sensor
+operation
+platform
+process
+program
+project
+study
+task
+trail
The potential values in the InitiativeTypeCode Code List are somewhat ambiguous. SMAP may want to include additional values that fit mission requirements and define them clearly.



Pre-decisional – for planning and discussion purposes only



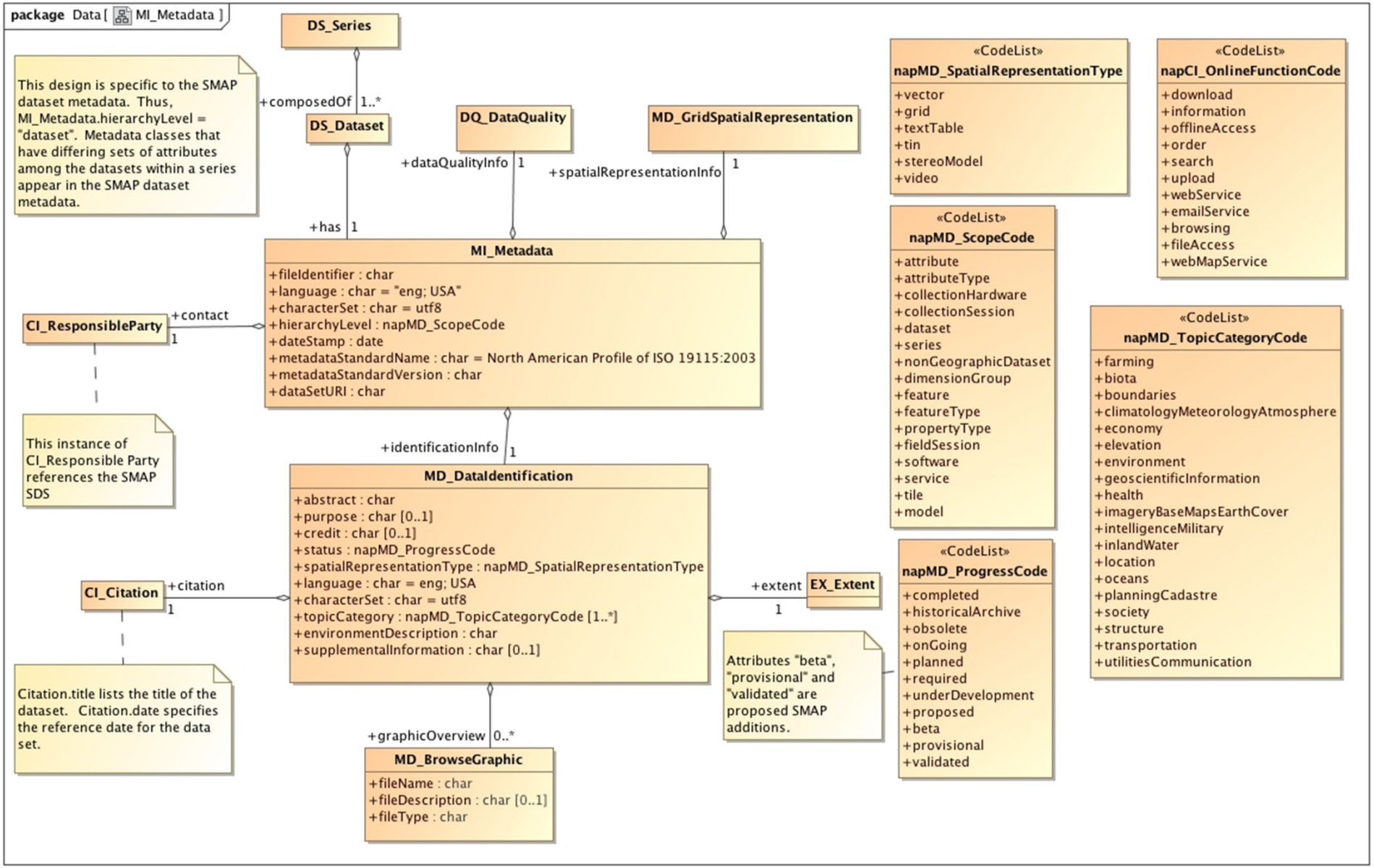
Pre-decisional – for planning and discussion purposes only



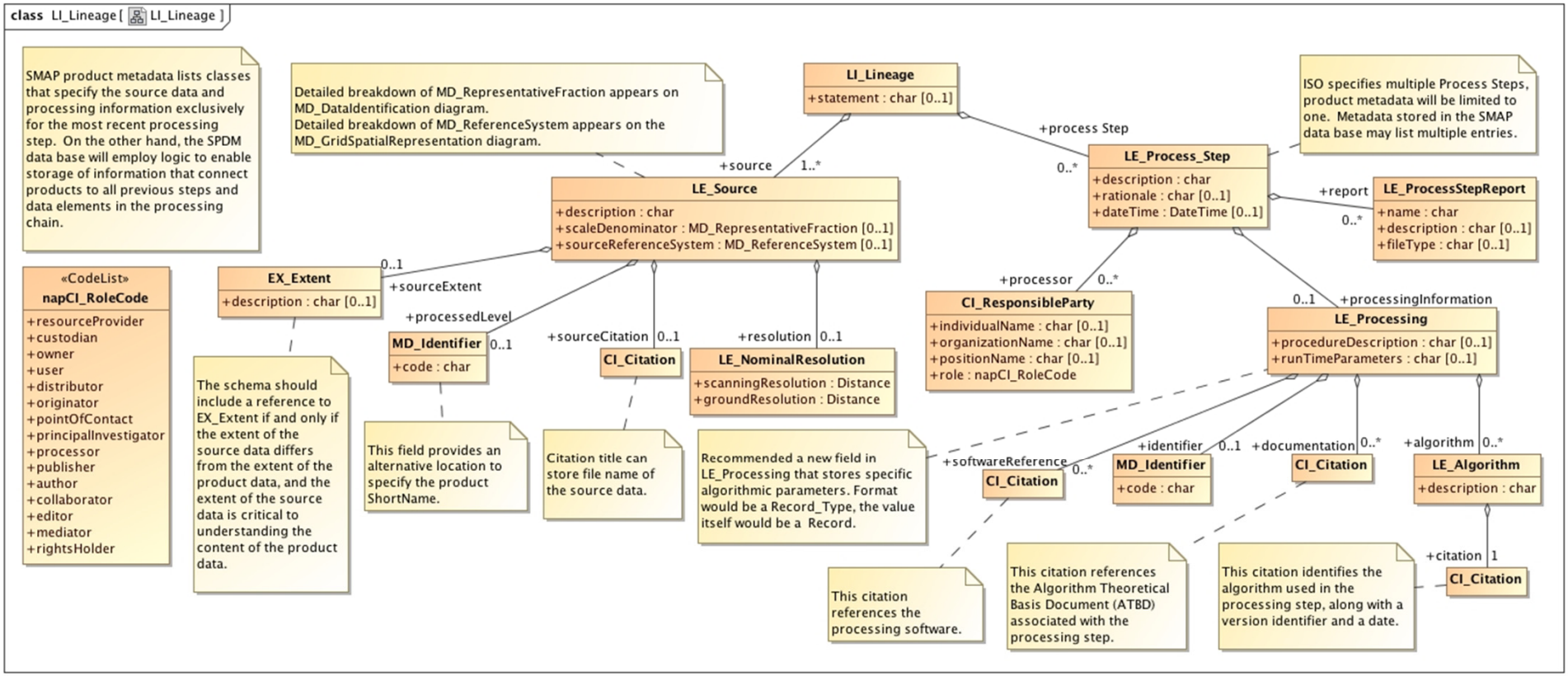
Pre-decisional – for planning and discussion purposes only

Proposed SMAP Dataset Metadata Design

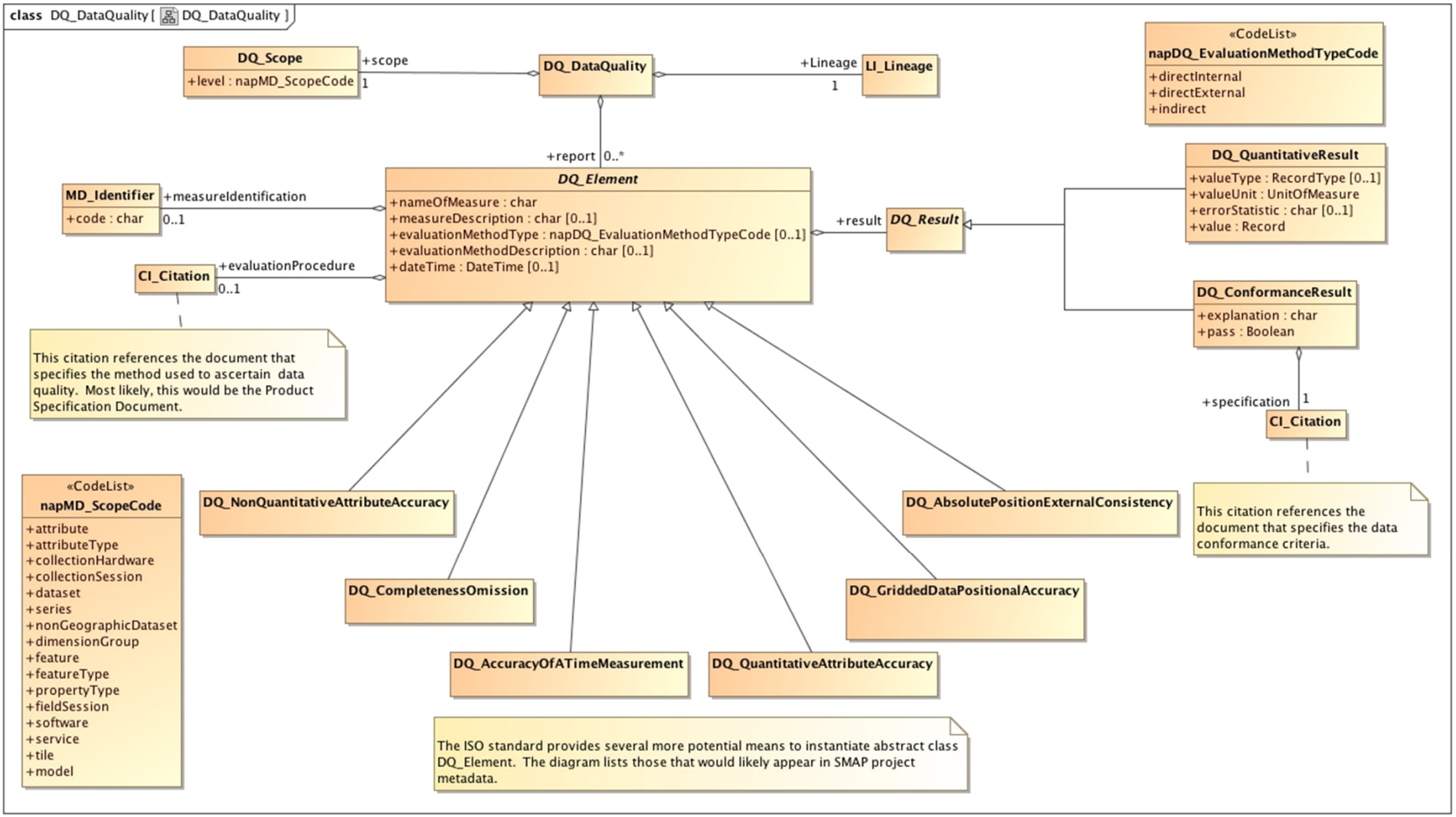
Pre-decisional – for planning and discussion purposes only



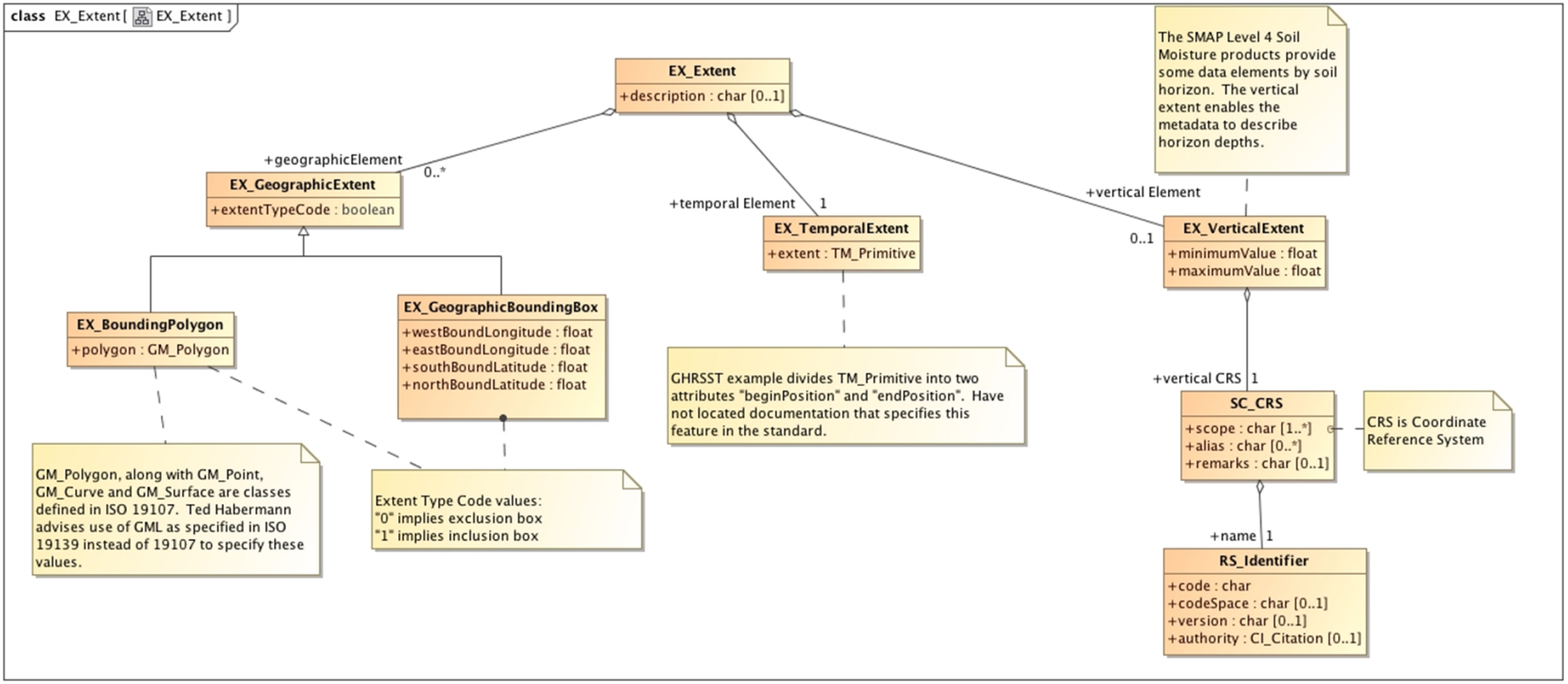
Pre-decisional – for planning and discussion purposes only



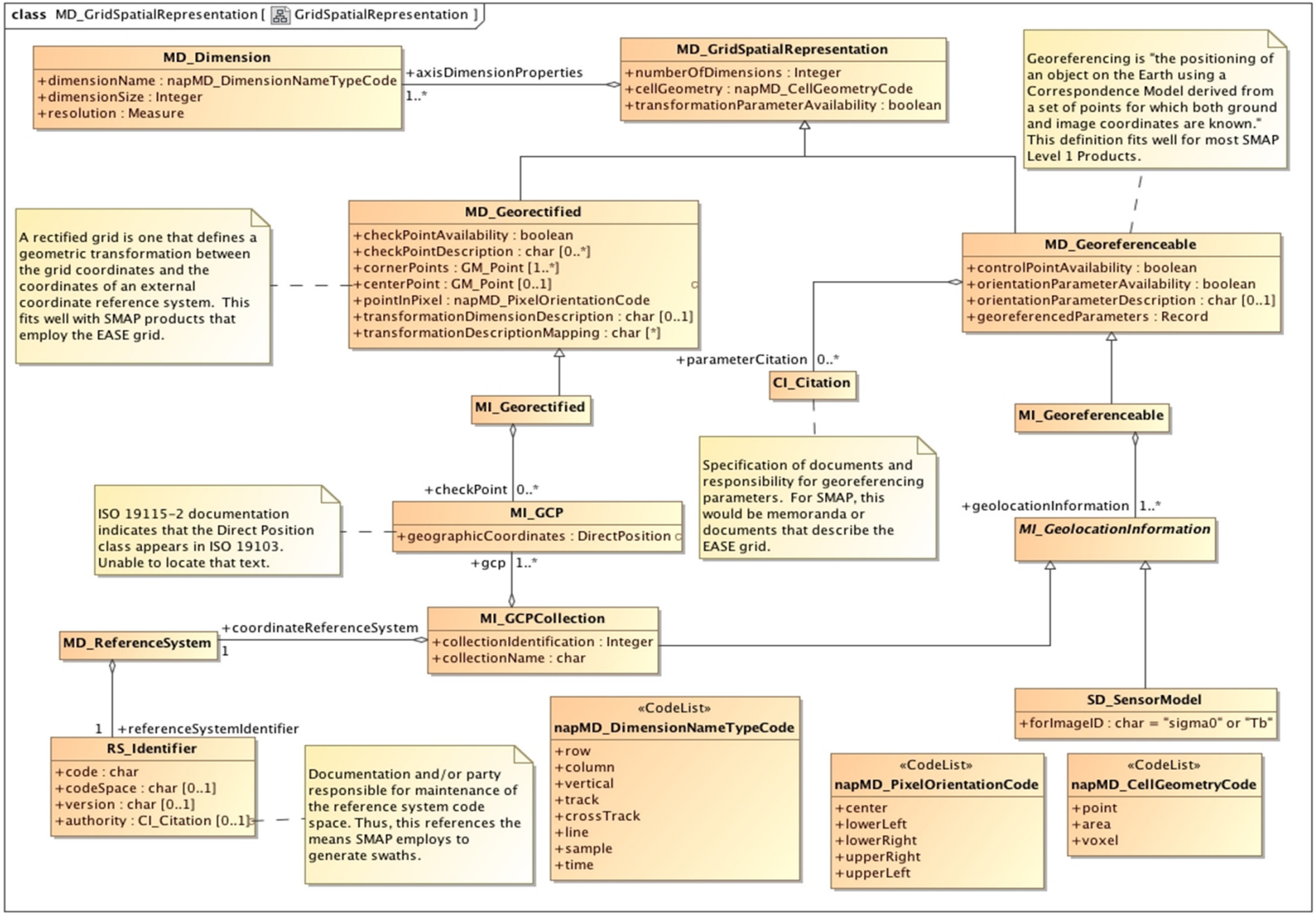
Pre-decisional – for planning and discussion purposes only



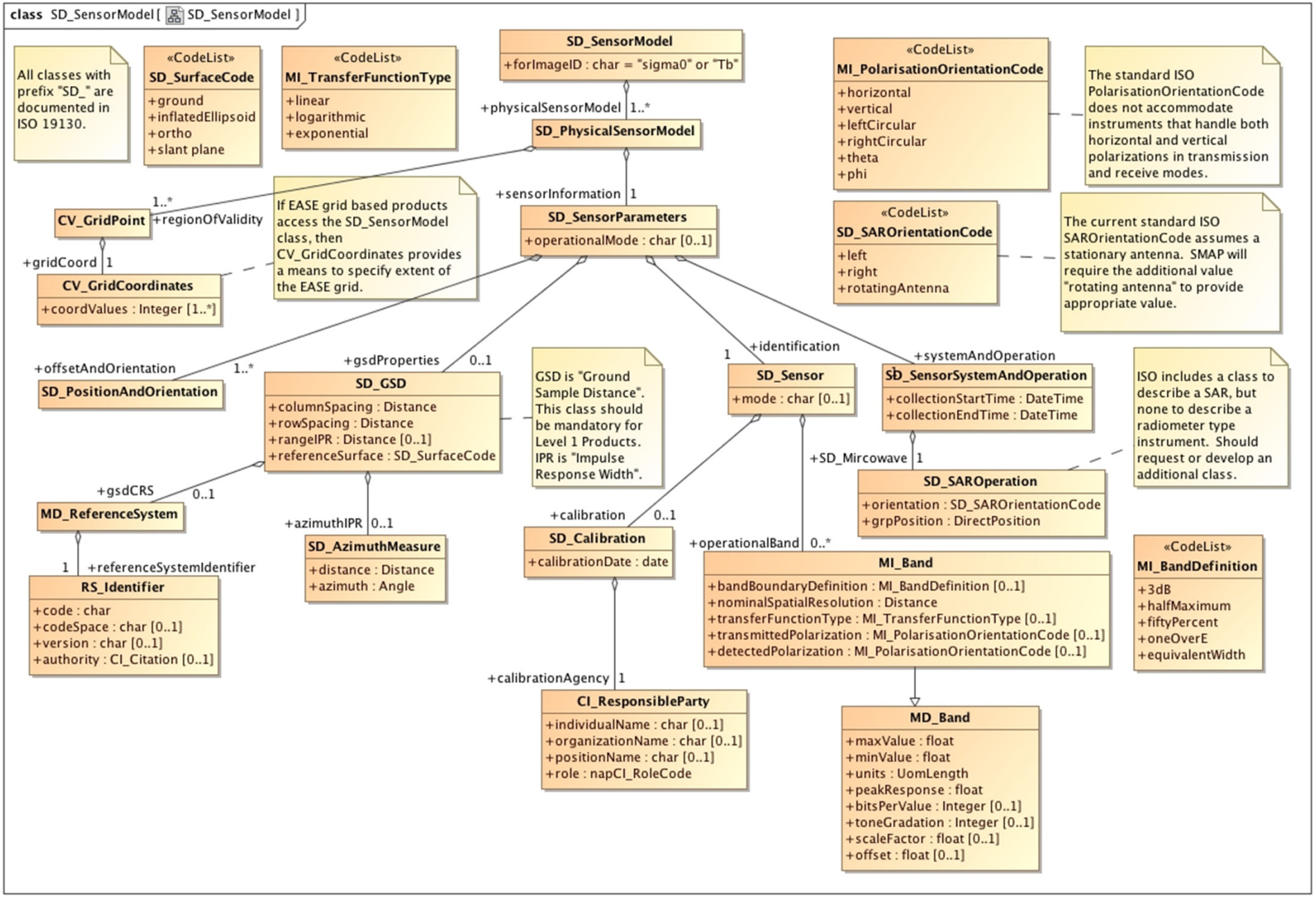
Pre-decisional – for planning and discussion purposes only



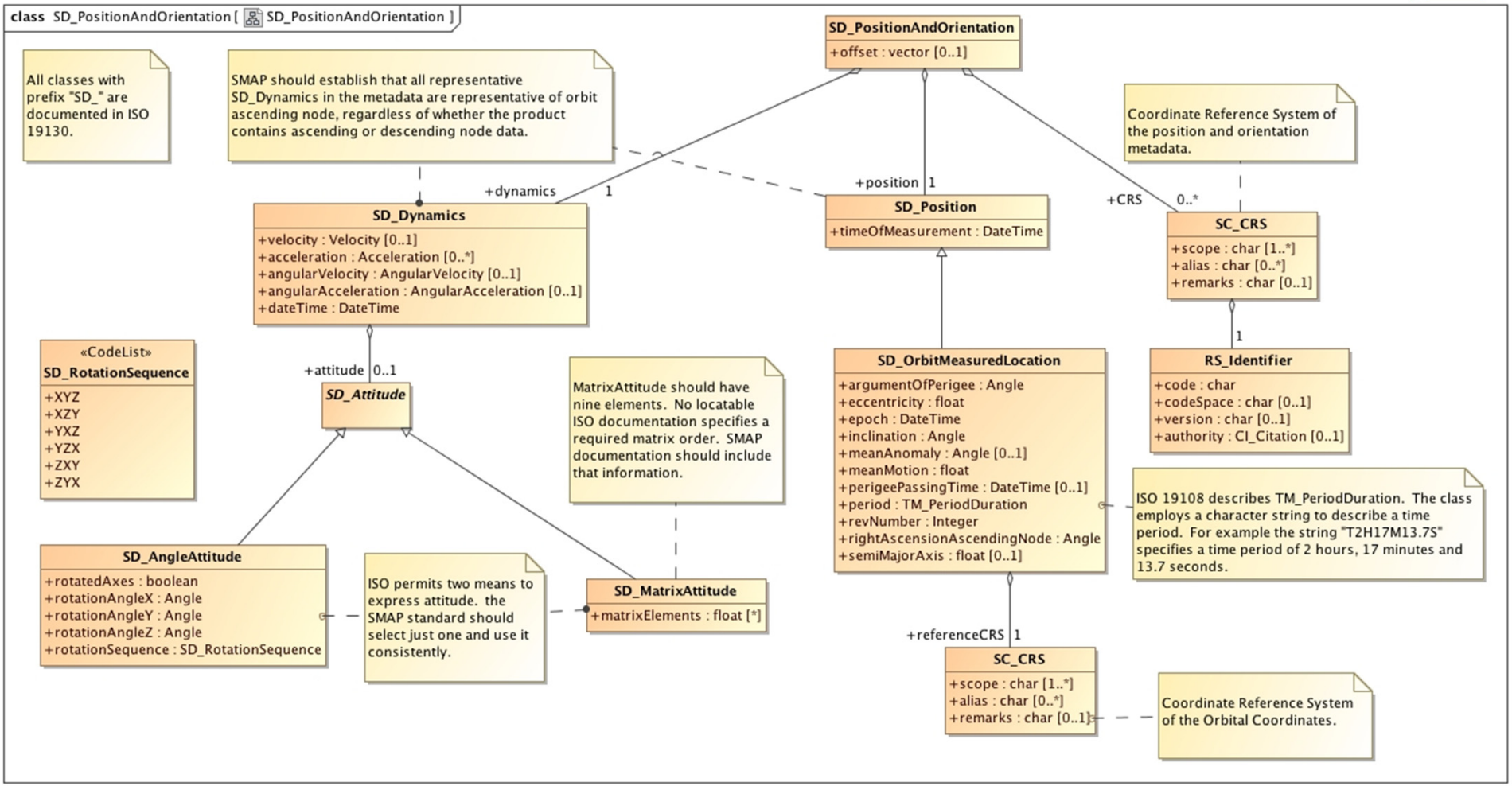
Pre-decisional – for planning and discussion purposes only



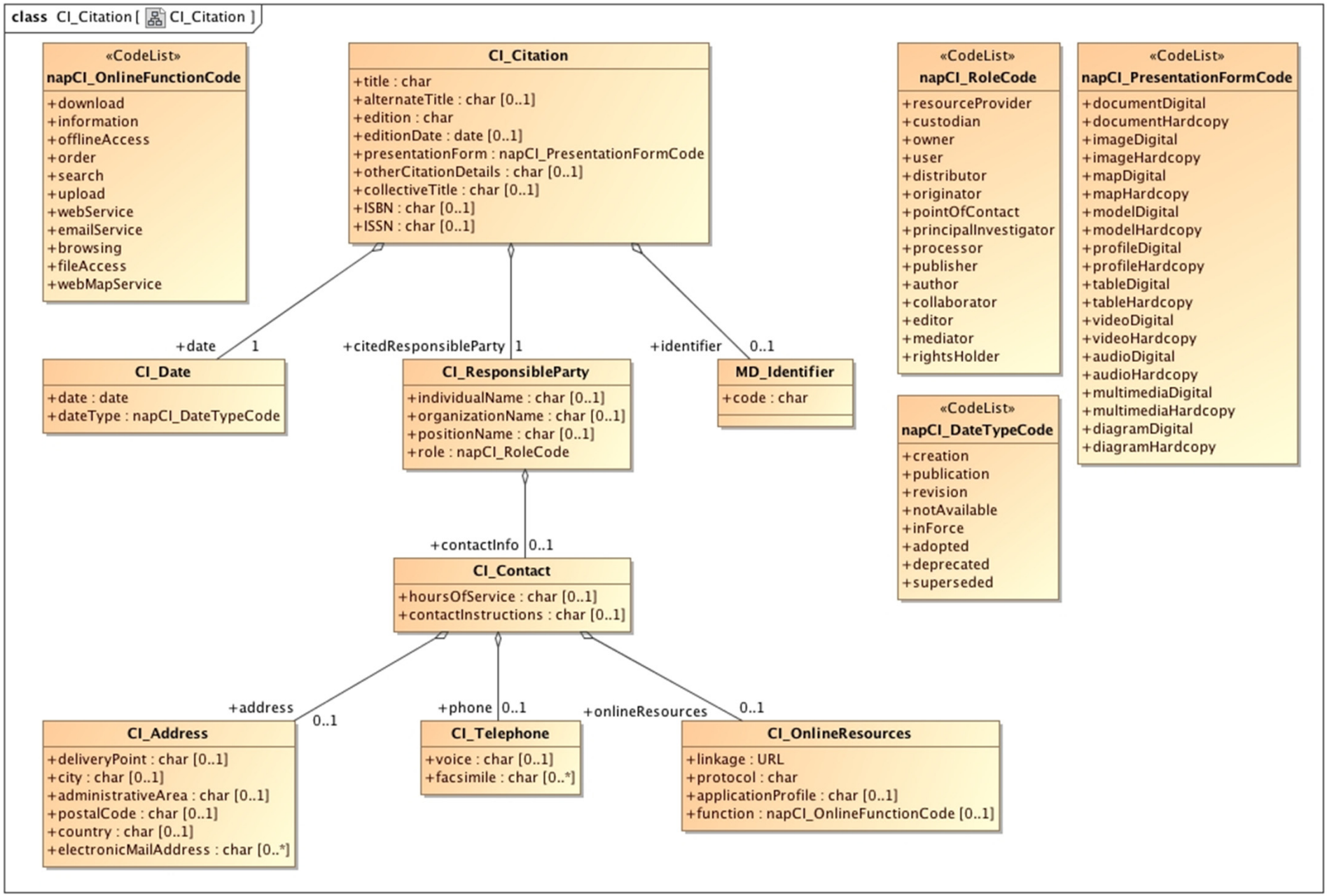
Pre-decisional – for planning and discussion purposes only



Pre-decisional – for planning and discussion purposes only



Pre-decisional – for planning and discussion purposes only



Pre-decisional – for planning and discussion purposes only