IMAGERY METADATA DEVELOPMENT BASED ON ISO/TC 211 STANDARDS

Rong Xie[1*] and Ryosuke Shibasaki[2]

Center for Spatial Information Science, The University of Tokyo
IIS CW501, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
[1]Email: xierong@iis.u-tokyo.ac.jp
[2] *Email:* shiba@iis.u-tokyo.ac.jp

ABSTRACT

This paper reviews the present status and major problems of the existing ISO standards related to imagery metadata. An imagery metadata model is proposed to facilitate the development of imagery metadata on the basis of conformance to these standards and combination with other ISO standards related to imagery. The model presents an integrated metadata structure and content description for any imagery data for finding data and data integration. Using the application of satellite data integration in CEOP as an example, satellite imagery metadata is developed, and the resulting satellite metadata list is given.

Keywords: Imagery metadata, Imagery metadata model, Imagery metadata class, Imagery metadata element, ISO standards, UML model

1 INTRODUCTION

Within the geospatial environment and its applications, imagery data are important information sources and products. To apply various imagery data, it is necessary to represent or reconstruct exactly geometric conditions used in observing or acquiring these data. Information about description or reconstruction of the observation conditions is usually represented by imagery metadata.

For imagery metadata development, two primary uses can be identified: (1) finding data and (2) integrating data. In finding data, users try to identify data that may be useful for their own purposes. This usually requires two general categories of information about imagery data: suitability for the user's intended purpose and methods for understanding and interpreting the image. Therefore, metadata are required to provide a minimum amount of information to support users' judgment or evaluation of usability of the data. In integrating data, users spatially and temporally overlay the data so that the spatio-temporal relationships among data and/or pixels can be explicitly reconstructed. For this purpose, metadata are required to include enough information to let users overlay the retrieved data. For example if satellite data are not geometrically and radiometrically corrected, metadata should provide users with enough information to do their own geometric correction and radiometric correction, such as for satellite orbit, sensor attitude, and some calibration parameters, etc.

Standardization on imagery metadata is being conducted by several international organizations such as ISO/TC211 (Geographic information/Geomatics). ISO/TC 211 is now producing a series of extensible metadata specifications to effectively manage data for users' retrieval, including ISO 19115 geographic information – metadata (ISO/TC 211, 2002), ISO 19115 geographic information – metadata – part 2: extension for imagery and gridded data (ISO/TC 211, 2005). However, these standards do not provide all the metadata information needed to describe imagery data or to meet all the requirements of applications. Further, the current metadata standards are still in draft phase.

The main problems are:

(1) To apply imagery data to geographic information, the geolocation information must be added to the imagery metadata, so metadata and image processing are sufficient to permit users to extract directly the information needed to geo-reference the data. Present standards do not provide the information necessary for geolocation of an image.
(2) Although these standards provide metadata schema for geographical data, these metadata are not rich enough for all types of thematic requirements specific to the applications. New metadata items are required for the specific needs in applications.
(3) There is currently no imagery metadata model. An integrated imagery metadata model is required to describe imagery data. It is necessary to apply ISO metadata standards (ISO 19115 and ISO 19115 part 2) and to combine them with information about other ISO standards related to imagery, such as ISO

19130, while resolving issues of overlap. Also it is necessary to consider the extension for cases of professional applications and diverse communities. Thus imagery metadata can be developed to provide users with enough information for the applications of finding data and data integration.

In the paper, an imagery metadata model is proposed that establishes an integrated metadata structure defining the essential components for imagery data based on the ISO existing standards and imagery-specific needs. The rest of the paper is organized as follows. Section 2 reviews the present status of the ISO existing standards related to metadata and imagery. Section 3 proposes imagery metadata model and uses the model for metadata design. An example using the CEOP satellite imagery data is given in section 4. Finally, further discussion and conclusions are presented in section 5.

2 REVIEWS OF THE EXISTING STANDARDS RELATED TO METADATA AND IMAGERY

2.1 ISO 19115 – metadata

The international community approved ISO 19115 in 2003 as a tool to define metadata in the field of geographic information. It defines schemas required for describing geographic information and services and provides information about identification, extent, quality, spatial and temporal schemas, spatial reference, and distribution of geographic data. The document includes metadata UML models, core metadata, metadata package, data dictionary, descriptions of extensions, and profiles.

ISO 19115 is designed to be the general metadata standard applicable to all data sets with geographic information. It identifies a set of core metadata derived from the many metadata elements it defines. It also specifies the conditions that metadata elements may be mandatory, conditional, or optional. Although there is some service metadata in ISO 19115, particularly in the area of identification, much of the service metadata is defined in ISO 19119 (ISO/TC 211, 2001). The metadata in 19115 provides only limited information about the spatial and temporal schemas: the extents of both and the spatial representation information.

2.2 ISO 19115 part 2 – metadata extension for imagery and gridded data

At the time when ISO 19115 was drafted, it provided some provisions of metadata models for imagery and gridded data, but it did not provide all the metadata elements needed to describe imagery. ISO 19115 part 2 extends the metadata defined in ISO 19115 and identifies additional metadata required to describe imagery and gridded data, such as data quality, spatial representation, content, and acquisition information. It provides information about the properties of the measuring equipment used to acquire data, geometry of the measuring process employed by the equipment, and production process used to digitize the raw data. A committee draft (CD) for ISO 19115 part 2 has been published.

2.3 ISO 19130 – sensor data model for imagery and gridded data

To apply imagery data to geographic information, the geolocation information is very important. The ISO 19115 and ISO 19115 part 2 do not provide the information required for geolocation of an image. ISO 19130 specifies the information required to support geolocation and sensor properties if georeferenced imagery. It defines how the sensor measurements and the geolocation information are logically associated.

The georeferencing information in ISO 19130 is a subset of the georeferencing description of ISO 19115 part 2, and, that area of ISO 19130 should be associated with ISO 19115 part 2. In order to develop a full set of imagery metadata, it is necessary to combine the relevant parts of ISO metadata standards (ISO 19115, ISO 19115 part 2) with geolocation information or sensor properties from ISO imagery standard (ISO 19130) or both.

3 IMAGERY METADATA MODEL

Imagery metadata is a description of the content, quality, condition, and characteristics of imagery datasets. Figure 1 presents a proposed structure and content for imagery metadata. A full set of imagery metadata contains one or more imagery metadata classes and elements. A metadata class is defined as a set of metadata elements describing the same aspects of data. A logically primitive item of metadata is called metadata element, which is defined as a discrete unit of metadata. The layout of imagery metadata class is described in Figure 1(a), including 8 main metadata classes as follows: metadata set information, identification information, data quality

information, spatial representation information, geolocation information, reference system information, content information, and specific information.

Metadata elements can be organized into three types of components as shown in Figure 1(b): core elements, extended elements, and specific elements. Core elements define the basic minimum number of metadata elements that must be reported for all applications of a geographical dataset. The extended elements define those element extensions necessary for proper characterization of an imagery dataset. Specific elements include professional elements or organizational elements. Professional elements define conditional or optional metadata for specific professional (discipline) needs. Organizational elements define conditional or optional metadata for the specific needs of different research communities.

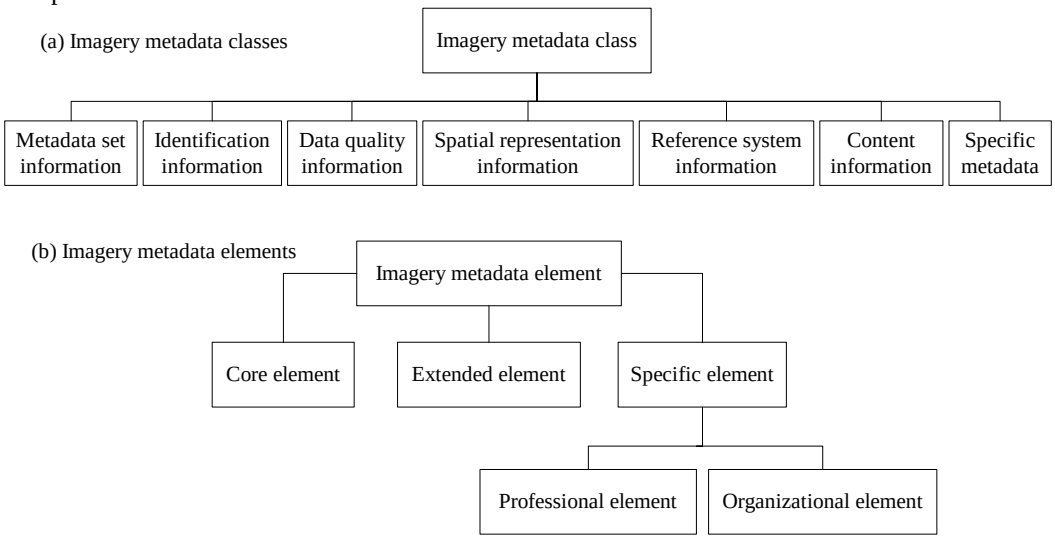

Figure 1. Imagery metadata structure and content

3.1 Core metadata elements

Even though metadata standards define an extensive set of metadata elements, typically only a subset of the full number of elements is used. It is essential to define the minimum set of metadata elements, called core metadata elements, needed for all application datasets, without which the application data set is not well described. ISO 19115 defines 5 core metadata classes with 13 core metadata elements. These are the mandatory metadata components that should be included in our metadata design. Figure 2 shows an UML model of these elements and their relationship. The definition of each imagery metadata element follows.

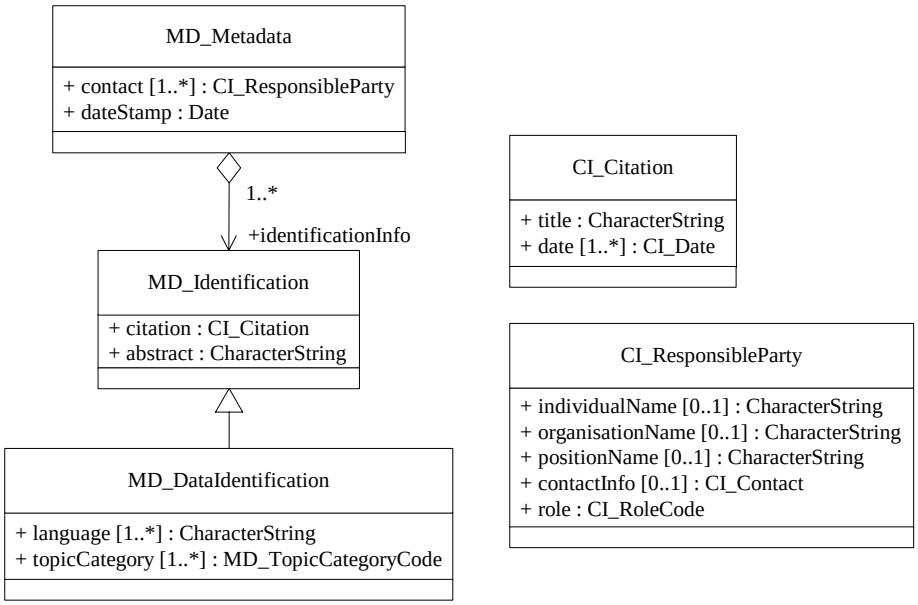

Figure 2. UML model of core metadata elements (from ISO 19115)

- *contact*::MD_Metadata – party responsible for the metadata information
- *dateStamp*::MD_Metadata – date when the metadata was created
- *citation*::MD_Identification – citation data for the resources
- *abstract*::MD_Identification – brief summary of the content of the resources
- *language*::MD_DataIdentification – language(s) used within the dataset
- *topicCategory*::MD_DataIdentification – main themes of the dataset
- *title*::CI_Citation – name by which the cited resource is known
- *date*::CI_Citation – reference date for the cited resource
- *individualName*::CI_ResponsibleParty – name of the responsible person
- *organisationName*::CI_ResponsibleParty – name of the responsible organization
- *positionName*::CI_ResponsibleParty – role or position of the responsible person
- *contactInfo*::CI_ResponsibleParty – address of the responsible party
- *role*::CI_ResponsibleParty – function performed by the responsible party

3.2 Specific metadata elements

The specific metadata elements identify metadata specifically required for the selected use cases. They include professional metadata elements or organizational metadata elements.

(1) Professional metadata elements
One of the significant enhancements of metadata is the ability to extend for specific professional needs. Professional metadata elements define a set of metadata classes and their elements that are used by a specific discipline or profession. The extended elements are outside the ISO standards and must meet rules as follows.

- Each metadata element is unique and not already defined in the model.
- Each metadata class is unique and not previously used.

(2) Organizational metadata elements
To meet specialized requirements of nations, regions, and organizations, organizational metadata elements can be defined by a dataset producer or a user community. The extended elements are also outside the ISO standards but are needed by the dataset producers or communities. Such elements also must meet rules as the same of rules of professional metadata elements.

CEOP (Coordinated Enhanced Observation Period) is currently establishing an integrated global observing system of water and energy cycle for scientific and social needs. Taking an example of CEOP application of satellite data integration, Figure 3 shows the UML model of specific metadata elements related to CEOP metadata, which include EOP3 and 4 resampling data, CEOP distribution information, and imagery content information. The definition of each metadata element is described as follows.

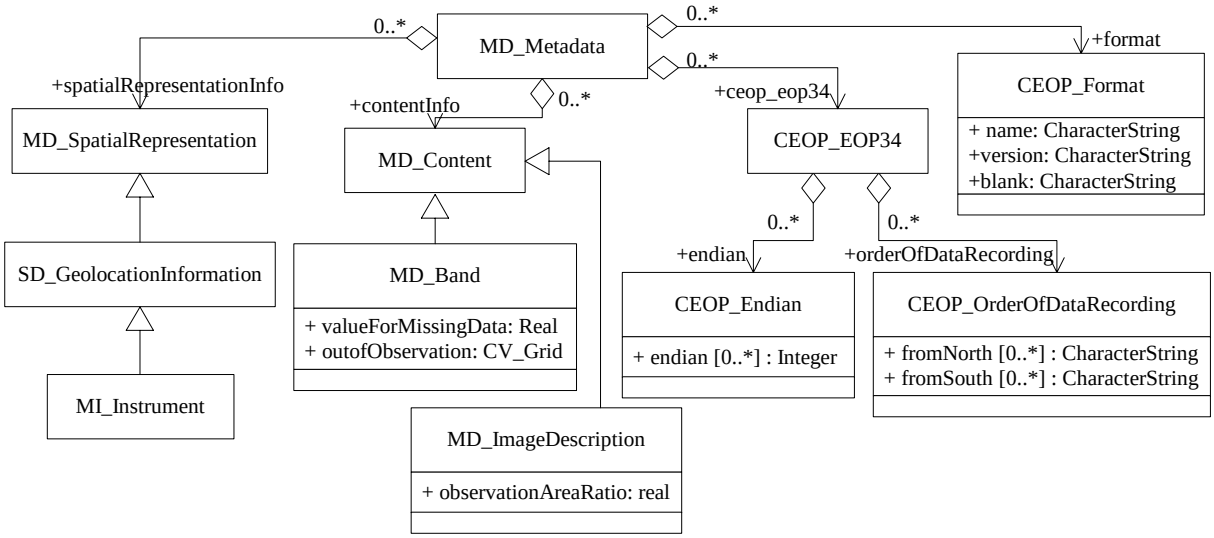

Figure 3. UML model of specific metadata elements (an example)

- *valueForMissingData*::MD_Band – whether data are not captured though a grid-cell is in the observation area
- *outOfObservation*::MD_Band – description of a grid-cell when the grid-cell is out of the observation area
- *observationAreaRatio*::MD_ImageDescription – observation area ratio
- *endian*::CEOP_Endian – a mixture of endian by organizations and systems that generated the satellite geo-coded image product.
- *fromNorth*::CEOP_OrderOfDataRecording and *fromSouth*::CEOP_OrderOfDataRecording – some data products have pixels starting from the north, while some others have pixels starting from the south. The information of "from north" and "from south" is needed to describe the order of data recording special attributes for CEOP application.
- *name*::CEOP_Format – name of the data transfer format
- *version*::CEOP_Format – version of the format (date, number, etc)
- *blank*::CEOP_Format – blank, special attribute for CEOP application

3.3 Extended metadata elements

Figure 4 presents the UML model of extended metadata elements for imagery, which is proposed based on ISO 19115 metadata standard, ISO 19115-2 metadata extension for imagery and ISO 19130 sensor model. It includes six main metadata classes as follows. *MD_Metadata* defines the imagery metadata. *MD_Identification* identifies the imagery data, including a graphic overview of the data, keywords describing the resource, and data identification. *DQ_DataQuality* describes an assessment of the quality of the dataset as well as sources and production processes used in producing a dataset. *MD_SpatialRepresentation* concerns the mechanisms used to represent spatial information in a dataset. *MD_ReferenceSystem* describes the spatial and temporal reference systems used in a dataset. *MD_Content* identifies the feature catalogue used and content of the coverage dataset. *SD_GeolocationInformation* defines the geographic location corresponding to image location.

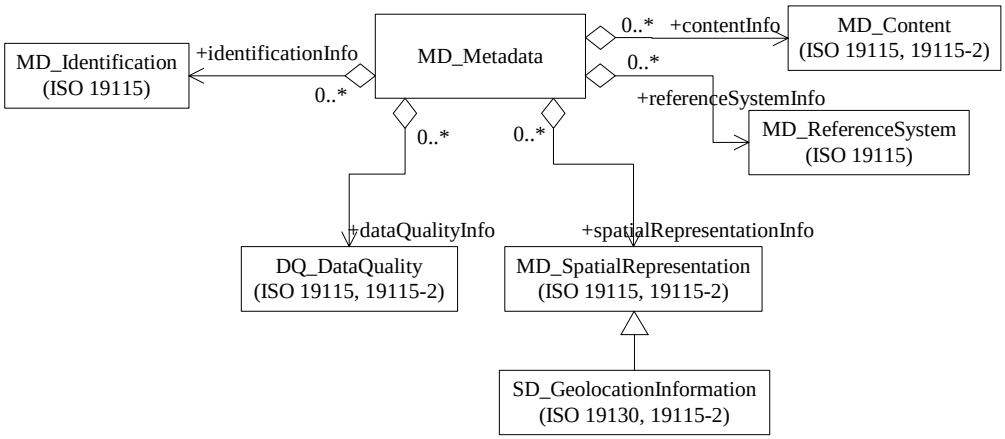

Figure 4. UML model of extended metadata elements

3.3.1 Identification information

This information provides an outline of the content and format of the imagery data. In ISO 19115, it defines general metadata elements of identification information that apply to any geographic data. That information relevant to imagery is: MD_DataIdentification, MD_BrowseGraphic and MD_Keywords. MD_DataIdentification describes information required to identify a dataset. MD_BrowseGraphic describes a graphic that provides an illustration of the dataset. MD_Keywords provides keywords, their type, and reference source. Figure 5 shows the UML model of identification information for imagery. The definition of each metadata element is described as follows.

- *spatialRepresentationType*::MD_DataIdentification – method used to spatially represent geographic information
- *spatialResolution*::MD_DataIdentification – factor that provides the density of spatial data in the dataset
- *extent*::MD_DataIdentification – information on the bounding box, bounding polygon, vertical, and temporal extent of the dataset
- *fileName*::MD_BrowseGraphic – name of the file that contains a graphic that provides an illustration of the dataset
- *fileDescription*::MD_BrowseGraphic – text description of the illustration
- *fileType*::MD_BrowseGraphic – format in which the illustration is encoded, such as GIF, JPEG etc
- *keyword*::MD_Keywords – commonly used word(s) or formalised word(s) or phrase(s) used to describe the subject
- *type*::MD_Keywords – subject matter used to group similar keywords
- *westBoundLongitude*::EX_GeogaphicBoundingBox – western-most coordinate of the limit of the dataset extent)
- *eastBoundLongitude*::EX_GeogaphicBoundingBox – eastern-most coordinate of the limit of the dataset extent)
- *southBoundLatitude*::EX_GeogaphicBoundingBox – southern-most coordinate of the limit of the dataset extent)
- *northBoundLatitude*::EX_GeogaphicBoundingBox – northern-most, coordinate of the limit of the dataset extent)
- *geographicIdentifier*::EX_GeogaphicDescription – identifier used to represent a geographic area
- *extent*::EX_TemporalExtent – date and time for the content of the dataset

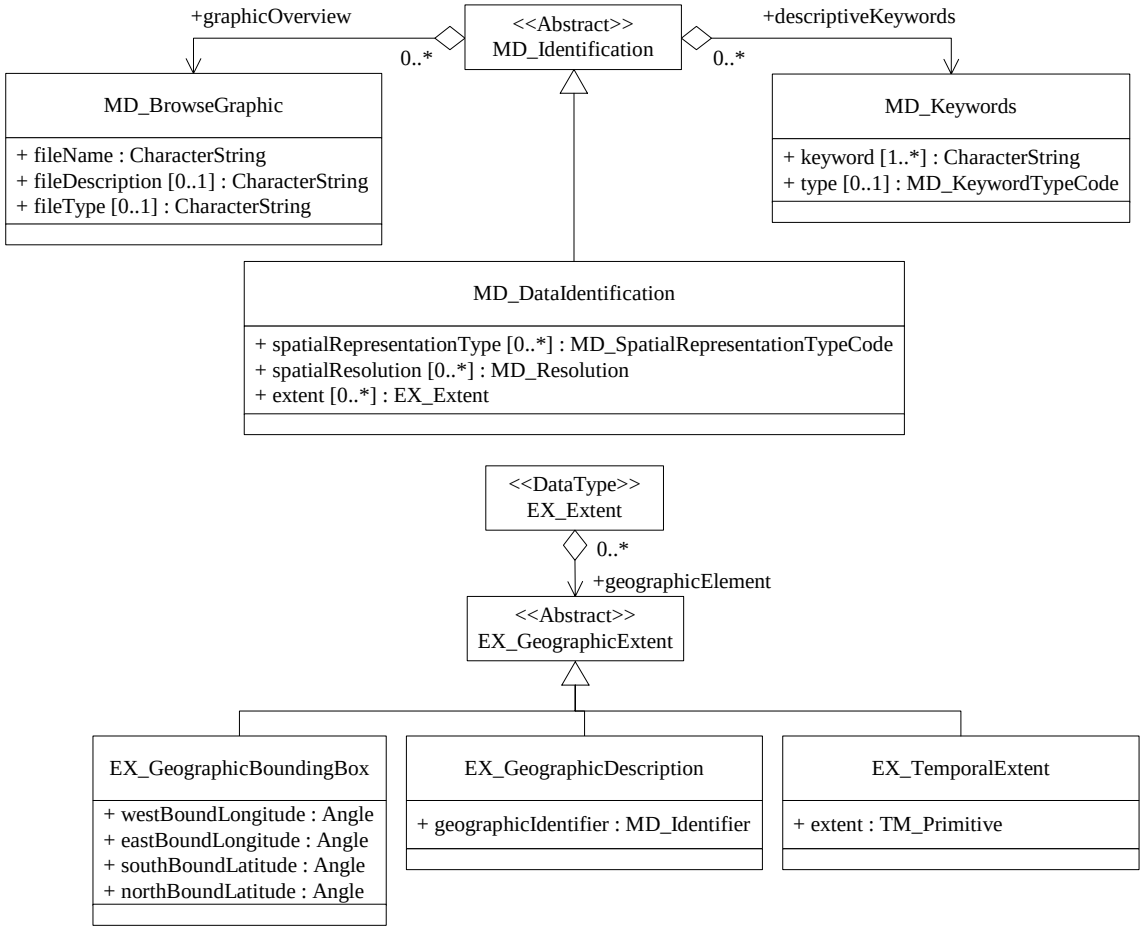

Figure 5. UML model of identification information for imagery

3.3.2 Data quality information

This information provides metadata classes and elements that describe the accuracy of the data and how the data presented in the current dataset are derived from the original measurement. Imagery data are not raw measurements but are obtained after processing. Metadata on the processing procedure are often of particular importance to imagery. LI_Lineage, LI_Source and LI_ProcessStep are the important information that can describe the algorithm used to derive the imagery and the processing used to create it. This information is defined in ISO 19115. However, data quality information defined in ISO 19115 is currently very general, as it must apply to any geographic data. This information can be extended to include more detailed elements about processing imagery data, such as LE_Processing, LE_ProcessStepReport, and LE_Algorithm, which can be found in ISO 19115-2. In addition, DQ_Element in ISO 19115 defines all values obtained from applying a data quality measure or the outcome of evaluating the obtained value. The class of DQ_PositionAccuracy, however, is only essential to describe positional accuracy information of an image. Figure 6 shows the UML model of data quality information for imagery. The definition of each metadata element is described as follows.

- *description*::LI_Source – detailed description of the level of the source data
- *scaleDenominator*::LI_Source – denominator of the representative fraction on a source map
- *sourceReferenceSystem*::LI_Source – spatial reference system used by the source data
- *sourceCitation*::LI_Source – recommended reference to be used for the source data
- sourceExtent::LI_Source – information about spatial, vertical, and temporal extent of the source data
- *description*::LI_ProcessStep –description of the event, including related parameters or tolerances
- *rationale*::LI_ProcessStep – requirement or purpose for the process step
- *dateTime*::LI_ProcessStep – date and time or range of date and time on or over which the process step occurred
- *processor*::LI_ProcessStep – identification of, means of communication with persons and organizations associated with the process step
- *identifier*::LE_Processing – information to identify the processing package that produced the data
- *softwareReference*::LE_Processing – reference to document describing the processing software

- *procedureDescription*::LE_Processing – additional details about the processing procedures
- *documentation*::LE_Processing – reference to documentation describing the processing
- *runTimeParameters*::LE_Processing – parameters to control the processing operations
- citation::LE_Algorithm – information identifying the algorithm and version or date
- description::LE_Algorithm – information describing the algorithm used to generate the data
- *fileName*:: LE_ProcessStepReport – name of the processing report
- *fileDescription*:: LE_ProcessStepReport – textual description of what occurred during the process step
- *fileType*::LE_ProcessStepReport – type of file that contains the processing report

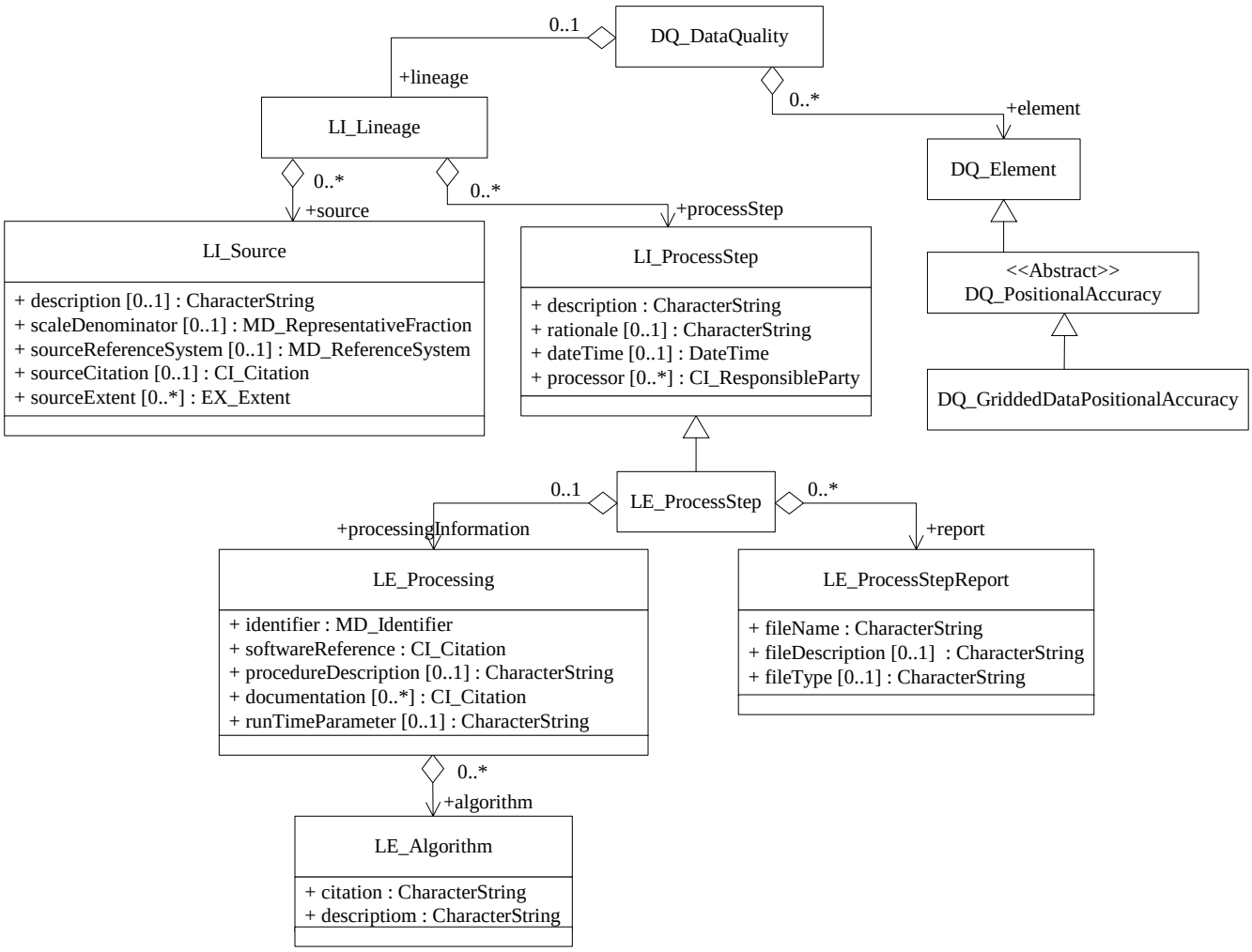

Figure 6. UML model of data quality information for imagery

3.3.3 Spatial representation information

This information provides the spatial representation for imagery. MD_GridSpatialRepresentation, defined in ISO 19115, is essential to the description of any geolocated imagery data. MD_Georectified applies to a grid whose cells are regularly spaced in geographic coordinate system and whose cells in the grid can be geolocated according to its grid coordinate, grid origin, cell spacing, and orientation. MD_Georeferenceable applies to a grid whose cells are irregularly spaced in the geographic coordinate system and whose cells in the grid can be geolocated using geolocation information supplied with the data but not from the grid properties. This information can be extended to provide further details needed to reference imagery data to a geographic coordinate system, which can be found in ISO 19115-2, such as MI_Georectified and MI_GeorefencingDescription. MI_Georectified is the subclass of MD_Georectified that contains additional information used to further specify georectification details of the imagery data. MI_GeorefencingDescription is the subclass of MD_Georeferenceable that contains additional information used to support georectification of the

imagery data. Figure 7 shows the UML model of spatial representation information for imagery. The definition of each metadata element is described as follows.

- *numberOfDimensions*::MD_GridSpatialRepresentation – number of independent spatial-temporal axes
- *axisDimensionsProperties*::MD_GridSpatialRepresentation – information about spatial-temporal axis properties
- *cellGeometry*::MD_GridSpatialRepresentation – identification of grid data as point or cell
- *transformationParameterAvailability*::MD_GridSpatialRepresentation – indication of availability of parameters for transformation between image coordinates and geographic coordinates
- *checkPointAvailability*::MD_Georectified – indication of availability of geographic position points to test the accuracy of the georeferenced grid data
- *checkPointDescription*::MD_Georectified – description of geographic position points used to test the accuracy of the georeferenced grid data
- *cornerPoints*::MD_Georectified – earth location in the coordinate system defined by the Spatial Reference System and the grid coordinate of the cells at opposite ends of grid coverage along two diagonals in the grid spatial dimensions
- *centerPoint*::MD_Georectified – earth location in the coordinate system defined by the Spatial Reference System and the grid coordinate of the cell halfway between opposite ends of the grid in the spatial dimensions
- *pointInPixel*::MD_Georectified – point in a pixel corresponding to the earth location of the pixel
- *transformationDimensionDescription*::MD_Georectified – general description of the transformation
- *transformationDimensionMapping*::MD_Georectified – information about which grid axes are the spatial axes
- *controlPointAvailability*::MD_Georeferenceable – indication of existence of control point(s)
- *orientationParameterAvailability*::MD_Georeferenceable – indication of availability of orientation parameters
- *orientationParameterDescription*::MD_Georeferenceable – description of parameters used to describe sensor orientation
- *georeferencedParameters*::MD_Georeferenceable – terms that support grid data georeferencing
- *parameterCitation*::MD_Georeferenceable – reference providing description of the parameters

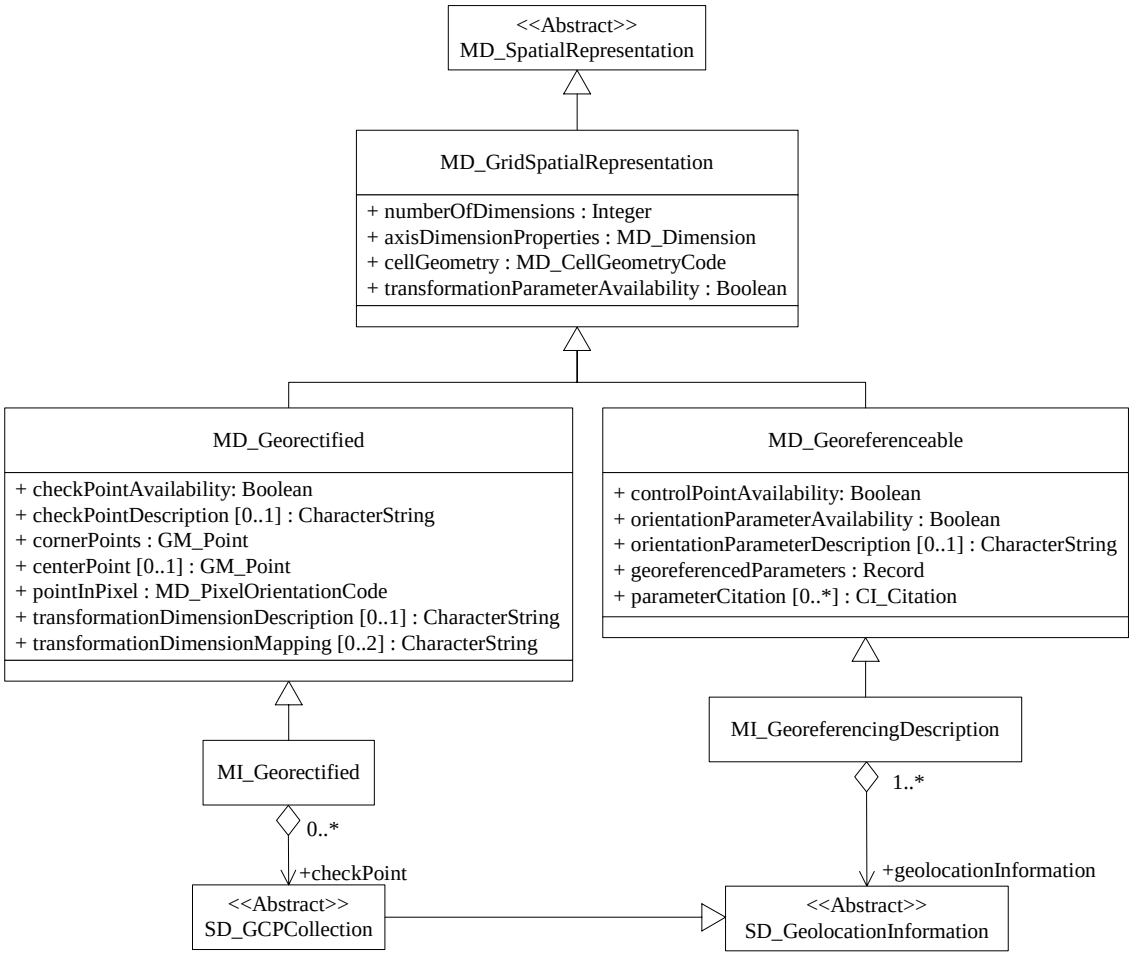

Figure 7. UML model of spatial representation information for imagery

3.3.4 Geolocation information

This information determines the geographic location corresponding to an image. ISO 19130 defines the sensor model and external information required for geolocation of an image. It is defined that the geolocation information should be provided in SD_GCPCollection or SD_SensorModel. SD_GCPCollection defines collection of ground control points. SD_SensorModel describes the sensor model information, consisting of SD_SensorParameters and SD_PlatformParameters. SD_SensorParameters defines basic sensor parameters including the position, orientation, and operational mode at a given time. SD_PlatformParameter defines basic identification information for the platform. Figure 8 shows the UML model of geolocation information for imagery. The definition of each metadata element is described as follows.

- *collectionIdentification*::SD_GCPCollection – identifier of the GCP collection
- *collectionName*::SD_GCPCollection – name of the GCP collection
- *coordinateReferenceSystem*::SD_GCPCollection – coordinates in which ground control points are defined
- *geographicCoordinates*::SD_GCPLocation – geographic or map position of the control point in either two or three dimensions

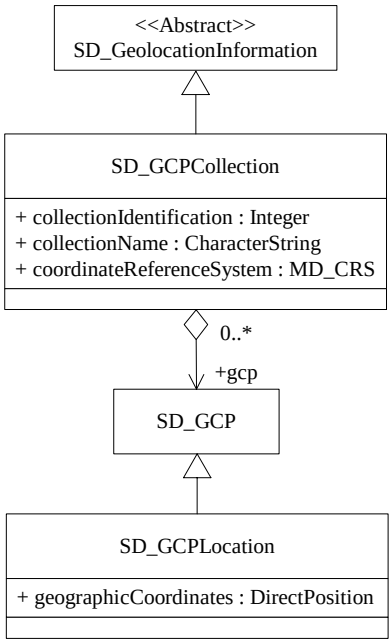

Figure 8. UML model of geolocation information for imagery

3.3.5 Reference system information

This information describes the spatial and temporal reference system, important information required for any image that is geographically located. In ISO 19115, MD_CRS describes metadata about a coordinate system consisting of MD_ProjectionParameters and MD_EllipsoidParameters. MD_ProjectionParameters includes a set of parameters that describe the projection, and MD_EllipsoidParameters includes a set of parameters that describe the ellipsoid. Figure 9 shows the UML model of reference system information for imagery. The definition of each metadata element is described as follows.

- *referenceSystemIdentifier*::MD_ReferenceSystem – name of reference system
- *projection*::MD_CRS – identity of the projection used
- *ellipsoid*::MD_CRS – identity of the ellipsoid used
- *datum*::MD_CRS – identity of the datum used

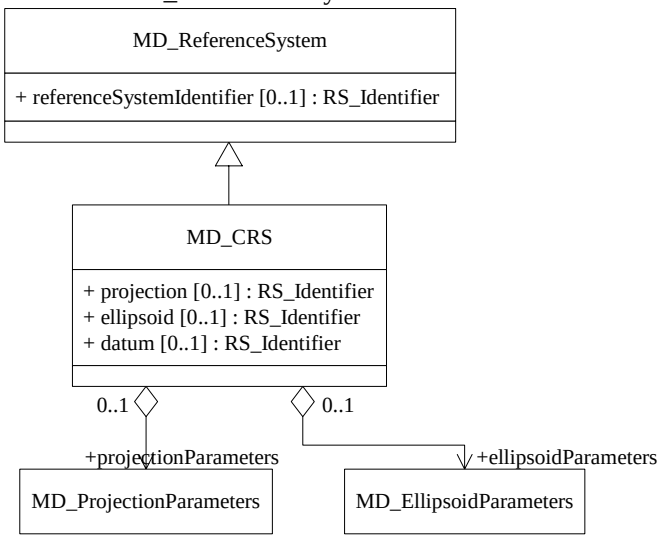

Figure 9. UML model of reference system information for imagery

3.3.6 Content information

This information contains the metadata classes and metadata elements required to describe attributes of the imagery content and how they are represented. ISO 19115 defines metadata elements of content information for any geographic data. MD_CoverageDescription is required, which contains information about the content of a grid data cell. MD_RangeDimension class identifies the range of each dimension of a cell measurement value,

and MD_Band class describes the range of wavelengths in the electromagnetic spectrum used. MD_ImageDescription class provides metadata about the image's suitability for use. ISO 19115 very generally describes content information for geographical data. This information can be extended to include more detailed information for imagery data, such as MI_Band, which is described in ISO 19115-2 by defining additional attributes for specifying properties of individual wavelength bands in the dataset. Figure 10 shows the UML model of content information for imagery. The definition of each metadata element is described as follows.

- *attributeDescription*::MD_CoverageDescription – description of the attribute described by the measurement value
- *contentType*::MD_CoverageDescription – type of information represented by the cell value
- *sequenceIdentifier*::MD_RangeDimension – number that uniquely identifies instances of bands of wavelengths on which a sensor operates
- *descriptor*::MD_RangeDimension – description of the range of a cell measurement value
- *maxValue*::MD_Band – longest wavelength that sensor is capable of collecting within a designated band
- *minValue*::MD_Band – shortest wavelength that sensor is capable of collecting within a designated band
- *units*::MD_Band – units in which sensor wavelengths are expressed
- *peakResponse*::MD_Band – wavelength at which the response is the highest
- *bitsPerValue*::MD_Band – maximum number of significant bits in the uncompressed representation for the value in each band of each pixel
- *toneGradation*::MD_Band – number of discrete numerical values in the grid data
- *scaleFactor*::MD_Band – scale factor which has been applied to the cell value
- *offset*::MD_Band – the physical value corresponding to a cell value of zero
- *bandBoundaryDefinition*::MI_Band – designation of criterion for defining maximum and minimum wavelengths for a spectral band
- *nominalSpatialResolution*::MI_Band – smallest distance between which separate points can be distinguished
- *polarisation*::MI_Band – polarisation of the transmitter or detector
- *illuminationElevationAngle*::MD_ImageDescription – illumination elevation measured in degrees
- *illuminationAzimuthAngle*::MD_ImageDescription – illumination azimuth measured in degrees
- *imagingCondition*::MD_ImageDescription – conditions affecting the image
- *imageQualityCode*::MD_ImageDescription – specification of the image quality
- *cloudCoverPercentage*::MD_ImageDescription – area of the dataset obscured by clouds
- *processingLevelCode*::MD_ImageDescription – image distributor's code that identifies the level of radiometric and geometric processing
- *compressionGenerationQuantity*::MD_ImageDescription – count of the number of lossy compression cycles performed on the image
- *triangulationIndicator*::MD_ImageDescription – indication of whether or not triangulation has been performed upon the image
- *radiometricCalibrationDataAvailability*::MD_ImageDescription – indication of availability of the radiometric calibration information for generating the radiometrically calibrated standard data product
- *cameraCalibrationInformationAvailability*::MD_ImageDescription – indication of availability of constants that allow for camera calibration corrections
- *filmDistortionInformationAvailability*::MD_ImageDescription – indication of availability of Calibration Reseau information
- *lensDistortionInformationAvailability*::MD_ImageDescription – indication of availability of lens aberration correction information

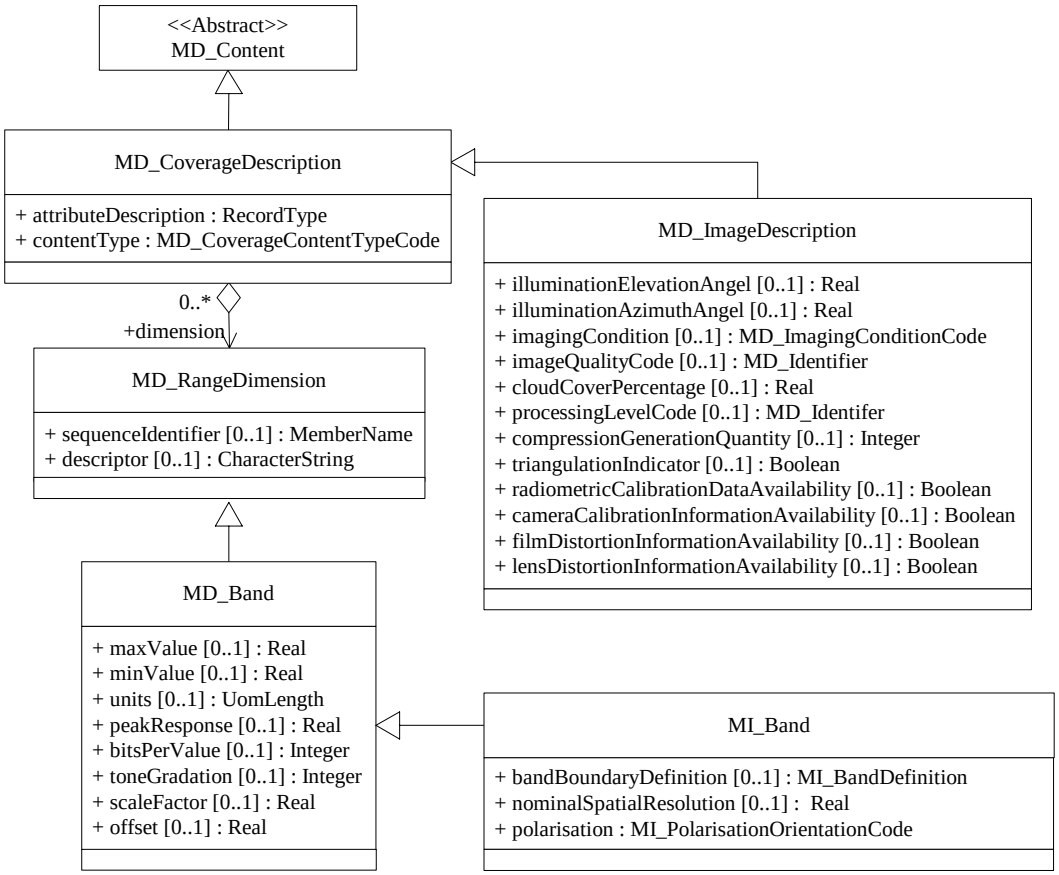

Figure 10. UML model of content information for imagery

3.4 Metadata elements design

Figure 11 shows how to determine metadata elements in applications. When designing metadata elements, it is first required to answer the questions for the user requirements as follows: (1) What: Does a dataset on a specific topic exist? (2) Where: For which specific place? (3) When: For which specific date or period? (4) Whom: Who is the point of contact to learn more about or to order the dataset? The answer to these user requirements is the basic step for the collection of metadata items in applications.

For a specific application, metadata elements, including core elements, extended elements, and specific elements, should be identified before the establishment of metadata structure. All core metadata elements defined in the imagery metadata model should be included in the design. To identify the extended elements, the application specific elements must be compared to the imagery metadata model. Some elements are reflected in the model, and some are not. The items reflected in the model are determined as extended elements. The items non-reflected in the model can be added as specific elements for the requirements of applications. These additional items will be helpful for users in deciding the availability and interoperability of datasets.

After all application specific metadata elements have been identified, they can be then organized as metadata classes according to the structure of imagery metadata as shown in Figure 1(a).

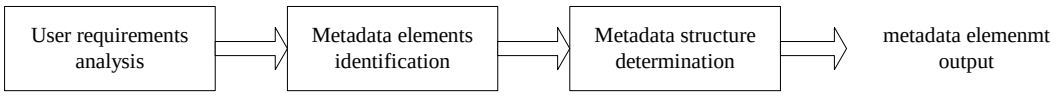

Figure 11. Metadata elements design

4 CASE STUDY: CEOP SATELLITE METADATA

CEOP (Toshio, 2005) is currently establishing an integrated global observing system of water and energy cycles for both scientific and social needs, which necessitates integrating various CEOP satellite imagery data. To facilitate the accessibility of the data collected from information sources and maximize their retrieval and

exchange, information sharing and dissemination is being established in CEOP. The development of satellite imagery has become a very important basic research tool.

Satellite products are supplied by JAXA, NASA, ESA and EUMETSAT, which include full scenes of data over the CEOP focused research areas, level 3 global gridded data and subset scene data (JAXA, 2005). Satellite products contain two major parts: the header and the data. The data part records the observed data, position data, etc. The header part is a metadata standard for satellite imagery data, which describes the outline of the satellite product and the main characteristics of the product. The header part is from Header Information of Each Data File (CEOP, 2003), which includes 21 necessary attributes such as filename, sensor, product, observation date and time, image size, data type, data unit, scale factor, observation channel, reference site, latitude/longitude in center of lower left pixel, grid size, missing value, observation area ratio, subset software version, processing date, processing center, input original filename, original file processing center, HDF library version, and blank. This information becomes the most important source to decide metadata structure and content. The CEOP satellite metadata can be designed from this satellite data file header.

4.1 CEOP satellite metadata element

We can define the mapping between CEOP items in the header and the proposed imagery metadata model. By mapping results, CEOP metadata elements are designed as shown in Table 1. In addition, some additional metadata elements are also required to describe CEOP specific information of metadata set, identification, citation and responsible party. Table 2 gives the results of additional items required.

Table 1. CEOP metadata elements for satellite imagery data

Description	Metadata Class	Metadata Element	Explanation
Filename	MD_Identification	citation	core metadata
Sensor	MI_Instrument	citation	extended metadata
		type	extended metadata
		description	extended metadata
Product	MD_DataIdentification	topicCategory	core metadata
	or MD_Keywords	keyword	extended metadata
Observation date and time	EX_TemporalExtent	extent	extended metadata
Image size	MD_GridSpatialRepresentation	numberOfDimensions	extended metadata
		axisDimensionProperties	extended metadata
		cellGeometry	extended metadata
	MD_RangeDimension	sequenceIdentifier	extended metadata
Data type	MD_Band	bitsPerValue	extended metadata
Data unit	MD_CoverageDescription	attributeDescription	extended metadata
Scale factor	MD_Band	scaleFactor	extended metadata
Observation channel	MD_RangeDimension	descriptor	extended metadata
FOV	MI_Band	nominalSpatialResolution	extended metadata
Reference site	CI_Citation	title	core metadata
	CI_Citation	date	core metadata
	EX_GeographicDescription	geographicIdentifier	extended metadata
LAT./LON. in center of lower left pixel	EX_GeographicBoundingBox	westBoundLongitude	extended metadata
		eastBoundLongitude	extended metadata
		southBoundLatitude	extended metadata
		northBoundLatitude	extended metadata
LAT./LON. and height of satellite in scanning scene center	MI_Instrument	origin	specific metadata
Grid size	MD_GridSpatialRepresentation	numberOfDimensions	extended metadata
		axisDimensionsProperties	extended metadata
		cellGeometry	extended metadata
		transformationParameterAvailability	extended metadata
Missing value	MD_Band	valueForMissingData	specific metadata
		outOfObservation	specific metadata
Observation area ratio	MD_ImageDescription	observationAreaRatio	specific metadata
Subset software version	LE_Processing	softwareReference	extended metadata

Reference for	LE_Processing	procedureDescription	extended metadata
processing algorithm	LE_Algorithm	citation	extended metadata
for physical data	LE_Algorithm	description	extended metadata
product			
Processing date	LI_ProcessStep	dateTime	extended metadata
Processing center	LI_ProcessStep	processor	extended metadata
Input original filename	LE_ProcessStepReport	fileName	extended metadata
Original file version	LE_ProcessStepReport	fileDescription	extended metadata
Original file processing	LI_ProcessStep	processor	extended metadata
center			
HDF library version	CEOP_Format	name	specific metadata
	CEOP_Format	version	specific metadata
Blank	CEOP_Format	blank	specific metadata
Endian	CEOP_Endian	endian	specific metadata
Order of data recording	CEOP_OrderOfDataRecording	fromNorth	specific metadata
	CEOP_OrderOfDataRecording	fromSouth	specific metadata

Table 2. Additional metadata elements for CEOP satellite imagery metadata

Description	Metadata Class	Metadata Element	Explanation
Additional metadata	MD_Metadata	contact	core metadata
element for CEOP	MD_Metadata	dateStamp	core metadata
	MD_Identification	abstract	core metadata
	MD_DataIdentification	language	core metadata
	CI_ResponsibleParty	individualName	core metadata
		organisationName	core metadata
		positionName	core metadata
		contactInfo	core metadata
		role	core metadata

4.2 Metadata element list

Based on Table 1 and Table 2, the metadata element list can be further developed. As given below, this list is provided in a tabbed-outline format. It presents the hierarchical structure of CEOP satellite metadata.

```
/* CEOP satellite imagery metadata */
    /* Metadata set information */
    MD_Metadata:
        contact:
            CI_ResponsibleParty:
                individualName: (name of the responsible person- surname, given name, title separated by a delimiter)
                organizationName: (name of responsible organization)
                positionName: (role or position of the responsible person)
                contactInfo:
                    CI_Contact:
                        phone:
                            CI_Telephone:
                                voice: (telephone number by which individuals can speak to the responsible organization orindividual)
                                facsimile: (telephone number of a facsimile machine for the responsible organization or individual)
                        address:
                            CI_Address:
                                city: (city of the location)
                                administrativeArea: (state, province of the location)
                                postalCode: (zip or other postal code)
                                country:  (country of the physical address)
                                electronicMailAddress: (address of the electronic mailbox of the responsible organization or individual)
                    onlineResource: (on-line information that can be used to contact the individual or organization)
                role:
                    CI_RoleCode: (function performed by the responsible party)
        dateStamp: (date that the metadata was created)
    /* Identification information */
    MD_Identification:
        citation: (citation data for the resource(s))
            CI_Citation:
                title: (name by which the cited resource is known)
```

```
            date: (reference date for the cited resource)
    abstract: (brief narrative summary of the content of the resource(s))
    MD_DataIdentification:
        language: (language(s) used within the dataset)
        topicCategory:
            MD_TopicCategoryCode: (high-level geographic data thematic classification to assist in the grouping and search of available
                                    geographic data sets)
        extent:
            EX_Extent:
                geographicElement:
                    EX_GeographicBoundingBox:
                        westBoundLongitude:  (western-most coordinate of the limit of the dataset extent)
                        eastBoundLongitude: (eastern-most coordinate of the limit of the dataset extent)
                        southBoundLatitude: (southern-most coordinate of the limit of the dataset extent)
                        northBoundLatitude: (northern-most, coordinate of the limit of the dataset extent)
                    EX_GeographicDescription:
                        geographicIdentifier:
                            MD_Identifier:
                                authority: (person or party responsible for maintenance)
                                code: (alphanumeric value identifying an instance in the namespace)
                    EX_TemporalExtent:
                        extent:
                            TM_Primitive:
                                beginDateTime: (the temporal position at which the ordinal era began)
                                endDateTime: (the temporal position at which the ordinal era ended)
                                dataFrequency: (frequency with which data are made)
    descriptiveKeywords:
        MD_Keywords:
            keywords: (commonly used word(s) or formalised word(s) or phrase(s) used to describe the subject)
/* Data quality information */
DQ_DataQuality:
    lineage:
        LI_Lineage:
            processStep:
                LI_ProcessStep:
                    dateTime: (date and time or range of date and time on or over which the process step occurred)
                    processor:
                        CI_ResponsibleParty:
                            individualName: (name of the responsible person- surname, given name, title separated by a delimiter)
                processingInformation:
                    LE_Processing:
                        softwareReference: (reference to document describing processing software)
                        procedureDescription: (additional details about the processing procedures)
                        algorithm:
                            LE_Algorithm:
                                citation: (information identifying the algorithm and version or date)
                                description: (information describing the algorithm used to generate the data)
                report:
                    LE_ProcessStepReport:
                        fileName: (name of the processing report)
                        fileDescription: (textual description of what occurred during the process step)
/* Spatial representation information */
MD_GridSpatialRepresentation:
    numberOfDimensions: (number of independent spatial-temporal axes)
    axisDimensionProperties:
        MD_Dimension:
            dimensionName: (name of the axis)
            dimensionSize: (number of elements along the axis)
            resolution: (degree of detail in the grid dataset)
    cellGeometry:
        MD_CellGeometryCode:
            point: (each cell represents a point)
            area: (each cell represents an area)
    transformationParameterAvailablity: (transformation between image coordinates and geographic coordinates)
    /* Geolocation information */
    geolocationInformation:
        SD_GeolocationInformation:
            instrument:
                MI_Instrument:
                    citation: (complete citation of the instrument)
                    type: (code describing the type of instrument)
                    description: (textual description of the instrument)
                    origin: (position of a satellite at any time)
/* Content information */
MD_CoverageDescription:
```

```
    attributeDescription: (description of the attribute described by the measurement value)
    dimension:
        MD_RangeDimension:
            sequenceIdentifier: (number that uniquely identifies instances of bands of wavelengths on which a sensor operates)
            descriptor: (description of the range of a cell measurement value)
            MD_Band:
                bitsPerValue: (maximum number of significant bits in the uncompressed representation for the value in each band of each
                        pixel)
                scaleFactor: (scale factor which has been applied to the cell value)
                outOfObservation: (description of a grid-cell when the grid-cell is out of the observation area)
                valueForMissingData: whether data are not captured though a grid-cell is in the observation area
                MI_Band:
                    nominalSpatialResolution: (smallest distance between which separate points can be distinguished, as specified in
                                instrument design)
    MD_ImageDescription:
        observationAreaRatio: (observation area ratio)
/* CEOP Distribution information */
CEOP_Format:
     name: (name of the data transfer format)
     version: (version of the format (date, number, etc))
     blank: (blank)
/* EOP 3,4 resampling information */
CEOP_EOP34:
    CEOP_Endian:
        endian: (a mixture of endian by organizations and systems that generated the satellite geo-coded image product)
    CEOP_OrderOfDataRecording:
        fromNorth: (pixels starting from the north)
        fromSouth: (pixels starting from the south)
```

5 FURTHER DISCUSSION AND CONCLUSIONS

Imagery metadata provide a short but critical summary of the content, data quality, acquisition conditions, and spatial characteristics of imagery data. The definition of the imagery metadata structure and content is the first essential stage of imagery metadata development. Metadata development includes: (1) imagery metadata implementation, (2) web-based metadata applications, and (3) imagery data discovery.

A primary goal for imagery metadata development is to facilitate access to the metadata and its related geographic data. Different applications may require metadata in different formats, which in turn requires software implementation using common encoding methods for operational ease of use. Within ISO/TC 211, an XML implementation of ISO 19115 is being developed in the ISO 19139 geographic information – metadata – XML schema implementation (ISO 2004). ISO 19139 expands ISO 19115 to define new constraint types that further refine the metadata elements for implementation. It also defines the rules used for deriving XML schema from the ISO abstract UML models. ISO 19139 is currently in draft phase and suitable for implementing metadata for general geographic data. Further work is required to develop XML implementation for imagery metadata, which includes exchange of metadata in data management systems, presentation of imagery metadata elements in a variety of forms and languages, and means to access conformable and available imagery metadata.

Imagery metadata can be applied to the imagery data service of a local or wide-area network. Basically, it requires: (1) Access to metadata concerning description of services, locations on the Internet, and means of accessing these services; (2) Discovering what kinds of services can be used to access the specific data holdings; and (3) Development of a common model for publishing and binding the web services.

Based on the existing ISO metadata standards (such as ISO 19115, ISO 19115 part 2) and ISO imagery standards (such as ISO 19130), we have proposed an integrated imagery metadata model. This model defines a common set of imagery metadata elements, given their definitions and inherent dependencies, and establishes metadata extensions for professional applications and organizational requirements, which can provide an abstract structure and content for describing imagery data. On the basis of our current research, we are now developing data services to use these results, including: (1) Development of a tool of service interface to use imagery metadata; (2) Development of on line access to imagery metadata from an archive system; and (3) Development of a prototype imagery data discovery system. Thus, these metadata can be applied to various applications from data description and data integration to data service and data discovery.

6 ACKNOWLEDGEMENTS

This research is part of the project "The establishment of water cycle informatics," which is funded by Japan Science and Technology under the grant No. 040300000455. The work is currently done under several international research collaborations. The authors would like to thank Prof. Toshio Koike (The University of Tokyo), Mr. Osamu Ochiai (JAXA), Ms. Robin Pfister, etc (NASA), Mr. Ben Burford (RESTEC), Mr. Steve Williams (UCAR), Mr. Frank Toussaint and Ms. Heinke Hoeck (MPI) for their many helpful suggestions on metadata development.

7 REFERENCES

CEOP (2003) Header information of each data file (version 02). Retrieved from the World Wide Web: http://monsoon.t.u-tokyo.ac.jp/ceop/data/eop-1/satellite/doc/header_table_r02.txt.

ISO/TC 211 (2001) ISO/TC 211 Geographic information/Geomatics Final text of CD 19119 Geographic information - Services, 2001-08-29, 77pp.

ISO/TC 211 (2002) ISO/TC 211 Geographic information/Geomatics Text of 19115 Geographic information - Metadata, 2002-10-28, 143pp.

ISO/TC 211 (2004) ISO/TC 211 Geographic information/Geomatics Draft technical specification 19139 Geographic information – Metadata – XML schema implementation, 2004-06-30, 145pp.

ISO/TC 211 (2005) ISO/TC 211 Geographic information/Geomatics CD 19115-2 Metadata - Part 2: Metadata for imagery and gridded data, 2005-11-14, 47pp.

ISO/TC 211 (2005) ISO/TC 211 Geographic information/Geomatics CD 19130.2 Geographic information - Sensor data model for imagery and gridded data, 2005-02-15, 160pp.
JAXA (2005) WGISS test facility For CEOP project plan (version 1.1). Retrieved from the World Wide Web: http://jaxa.ceos.org/wtf_ceop/materials/WTF_CEOP_PP1_1a.pdf.

JAXA (2005) WGISS test facility For CEOP implementation plan (version 1.1), Retrieved from the World Wide Web: http://jaxa.ceos.org/wtf_ceop/materials/WTF_CEOP_IP1_1.pdf.

Koike, T. (2005) Coordinated Enhanced Observing Period (CEOP) Phase 1 Achievements, Committee on Earth Observation Satellites 9th Plenary Meeting, London, UK.

Shibasaki Group (2004) Report on metadata design for integrating CEOP satellite imagery, reference site data and simulation result data for "global water informatics" project (Version 1.0), *CEOP International Metadata Meeting*, Tokyo, Japan.