Proposed Specification for Auxiliary Information Layer Integration for use with ENC S.10x

TSMAD20/DIPWG2-18B

prepared by IDON Technologies Inc.

for Lynn Patterson, Marc Journauld
Canadian Hydrographic Service
Fisheries and Oceans Canada

31 March 2010

Proposed Specification for Auxiliary Information Layer Integration for use with ENC S.10x

Introduction

This document describes the integration of auxiliary data layers to be used in conjunction with S.101 ENC data in navigational products. It is based[1] on the emerging standards in the International Organization for Standardisation (ISO) and the International Hydrographic Organization (IHO) S.100. It includes the use of Imagery and Gridded Data (Coverage) and Metadata components together with Vector data components.

ISO has developed a suite of geographic information standards that addresses all areas of geographic information. The ISO and IHO standards are based on a "General Feature Model" as defined in the ISO standard 19109 Rules for Application Schema. In accordance with the General Feature Model, all geographic information is feature oriented. Features have properties, associations and constraints. Among the properties of a feature are its location, spatial geometry, and associated attributes. This concept applies equally well to boundary defined features such as those used in "vector" based charts, as to gridded data sets representing bathymetry or other coverage data sets such as currents. In the case of bathymetry, the coverage feature is the area of sea bottom for which a depth surface is defined. The location is defined by a spatial referencing system and the geometry by a grid. The attributes consist of a set of depth measurements and associated metadata. The distribution of the attributes across the area covered by the grid is described by a coverage function.

The IHO S-100 Imagery and Gridded Data component describes the concept of coverages and the basic structure for a coverage data set. The concept of coverage geometry is defined in ISO standard 19123 and metadata in ISO standard 19115. The framework for Imagery, Gridded and Coverage data is described in ISO standard 19129. These standards are in place and can be used to produce a product specification. The ISO IGD component of S-100 is closely aligned to the ISO 19129 standard since the original ISO 19129 draft document was based on the IHO S-100 Imagery and Gridded data component of S-100. Many of the Auxiliary Information Layers for S.100 are coverage type data.

The approach of treating all data as feature data allows the easy mixing of different data types. This means that multiple "layers" or "groups" can be defined and co-presented as integrated data. This layering concept is very important for providing additional aids to navigation. The S.101 product specification for the Electronic Nautical Chart describes the essential data required to drive an Electronic Chart Display Information System. An official ECDIS system is considered as a replacement for the official paper chart and so can only display the ENC data in a prescribed way. However, the user requirements show that the mariner at times wants access to other data that can be displayed together with the ENC data. In particular the user requirement has identified a need

[1] This work is based the requirements identified by the user community through the St Lawrence Project of the Canadian Hydrographic Service and from input from the US Naval Oceanographic Office with respect to the structuring and encoding of bathymetry data. This document also makes use of work on an Elevation Surface Model that has been prepared by the US National Geospatial Intelligence Agency for application in land based military systems, which was presented to the IHO TSMAD sub-committee on S-100 in November 2006 and which is now under study by the international Defence Geospatial Information Working Group (DGIWG). IHO has a cooperative agreement with DGIWG and a liaison status with ISO. DGIWG is also working with the NATO GMWG committee to address a set of Additional Military Layers (AMLs) that are military auxiliary layers that can used together with S.101 (or the older S.57) ENC information. This document also supports AML layers as auxiliary layers.

for high resolution bathymetry, real time water levels, currents and ice information. This information may be addressed as optional overlay information on an Electronic Chart System ECS.

In order to be able to mix different types of data together it is necessary to orient this data using a common tiling scheme, spatial referencing system, and other common system wide attributes. Each additional type of data needs to be defined completely in its own product specification. This document introduces the concept of having a standardized set of auxiliary data layers or groups that can be used together, and describes the metadata and tiling schemes needed to achieve this. All of the components in the layers will derive from the same S.100 schema so they will closely interwork. The case of high resolution bathymetry is used as an example, but the details of this data type are described in a separate document S.102 Bathymetric Attributed Grid.

Proposed Specification for
Auxiliary Information Layer Integration
for use with ENC -- S.10x

1 Scope

This document specifies a structure for the integration of Auxiliary Information Layer data together with S.101 ENC base data or other S.100 compatible base data. The structure addresses the organization of metadata and the coordinate and spatial referencing issues.

2 Conformance

Conformance to this standard addresses only conformance to the common structure needed for the integration of auxiliary information layers. The data content and encoding within a layer is addressed within the conformance clause for that layer which is in addition to the conformance addressed in S.100. Data complying with this specification shall also comply with the conformance clauses inherited from the referenced ISO and IHO standards.

3 Normative references

The following normative documents contain provisions that, through reference in this text, constitute provisions of this document.

IHO S.100 IHO Universal Hydrographic Data Model
ISO 19101-2:2008 Geographic Information - Rules for Application Schema
ISO/TS 19103:2005 Geographic information - Conceptual schema language
ISO 19106:2004 Geographic Information - Profiles
ISO 19109:2005 Geographic Information - Rules for Application Schema
ISO 19111:2003 Geographic information - Spatial referencing by coordinates
ISO 19115:2003 Geographic information - Metadata
ISO 19115-2:2009 Geographic information - Metadata: Extensions for imagery and gridded data
ISO 19123:2005 Geographic information - Schema for coverage geometry and functions
ISO 19129:2009 Geographic information - Imagery gridded and coverage data framework
ISO 19131:2007 Geographic information - Data product specifications
ISO/IEC 19501:2005, Information technology — Open Distributed Processing - Unified Modelling Language Version 1.4.2
Note: a summary of UML is given in S.100 Part 1

4 Terms, and definitions

4.1 Terms and definitions

Terms and definitions have been taken from the normative references cited in clause 3. Only those, which are specific to this document, have been included and modified where necessary.

4.1.1 continuous coverage

coverage that returns different values for the same feature attribute at different **direct position**s within a single **geometric object** in its **spatiotemporal domain**
[ISO 19123]

4.1.2 coordinate

one of a sequence of numbers designating the position of a point in N-dimensional space
[ISO 19111]

4.1.3 coordinate reference system

coordinate system which is related to the real world by a datum
[ISO 19111]

4.1.4 coverage

feature that acts as a function to return values from its range for any direct position within its spatial, temporal, or spatiotemporal domain
[ISO 19123]

EXAMPLE Examples include a digital image, polygon overlay, or digital elevation matrix.

NOTE In other words, a coverage is a feature that has multiple values for each attribute type, where each direct position within the geometric representation of the feature has a single value for each attribute type.

4.1.5 coverage geometry

configuration of the domain of a coverage described in terms of coordinates
[ISO 19123]

4.1.6 depth

distance of a point from a chosen reference surface measured downward along a line perpendicular to that surface
[ISO 19111:2006]

NOTE Height is distinguished from elevation in that it is a directional measurement A depth above the reference surface will have a negative value.

4.1.7 direct position

position described by a single set of coordinates within a coordinate reference system
[ISO 19107]

4.1.8 domain

well-defined set
[ISO 19103]

NOTE Domains are used to define the domain set and range set of operators and functions.

4.1.9 evaluation

< coverage>
determination of the values of a **coverage** at a **direct position** within the **spatiotemporal domain** of the **coverage**
[ISO 19123]

4.1.10 feature

abstraction of real world phenomena
[ISO 19101]

NOTE A feature may occur as a type or an instance. Feature type or feature instance should be used when only one is meant.

4.1.11 feature attribute

characteristic of a feature

[ISO 19109]

NOTE A feature attribute type has a name, a data type and a domain associated to it. A feature attribute instance has an attribute value taken from the value domain of the feature attribute type.

4.1.12 function

rule that associates each element from a **domain** (source, or domain of the function) to a unique element in another domain (target, co-domain, or **range**)
[ISO 19107]

NOTE The range is defined by another domain.

4.1.13 geometric object

spatial object representing a set of **direct positions**
[ISO 19107]

NOTE A geometric object consists of a **geometric primitive**, a collection of geometric primitives, or a geometric complex treated as a single entity. A geometric object may be the spatial characteristics of an object such as a **feature** or a significant part of a feature

4.1.14 grid

network composed of two or more sets of **curves** in which the members of each set intersect the members of the other sets in a systematic way
[ISO 19123]

NOTE The curves partition a space into grid cells.

4.1.15 grid point

point located at the intersection of two or more curves in a **grid**
[ISO 19123]

4.1.16 height

distance of a point from a chosen reference surface measured upward along a line perpendicular to that surface
[ISO 19111:2006]

NOTE Height is distinguished from elevation in that it is a directional measurement.

4.1.17 range

<coverage>
set of values associated by a **function** with the elements of the **spatiotemporal domain** of a **coverage**
[ISO 19123]

4.1.18 record

finite, named collection of related items (objects or values)
[ISO 19107]

NOTE Logically, a record is a set of pairs <name, item >.

4.1.19 rectified grid

grid for which there is a linear relationship between the grid coordinates and the coordinates of an external coordinate reference system
[ISO 19123]

NOTE If the coordinate reference system is related to the earth by a datum, the grid is a georectified grid.

4.1.20 referenceable grid

grid associated with a transformation that can be used to convert grid coordinate values to values of **coordinates** referenced to an **external coordinate reference system**
[ISO 19123]

4.1.21 spatiotemporal domain

<coverage>
domain composed of **geometric objects** described in terms of spatial and/or temporal **coordinates**
[ISO 19123]

NOTE The spatiotemporal domain of a **continuous coverage** consists of a set of **direct positions** defined in relation to a collection of geometric objects.

4.1.22 surface

connected 2-dimensional **geometric primitive**, representing the continuous image of a region of a plane
[ISO 19107]

NOTE The boundary of a surface is the set of oriented, closed curves that delineate the limits of the surface.

4.1.23 tiling scheme

a discrete grid coverage that is used to partition data into discrete edge matched sets called tiles

4.1.24 vector

quantity having direction as well as magnitude
[ISO 19123]

NOTE A directed line segment represents a vector if the length and direction of the line segment are equal to the magnitude and direction of the vector. The term vector data refers to data that represents the spatial configuration of features as a set of directed line segments.

5 Symbols and abbreviated terms

5.1 Abbreviations

This product specification adopts the following convention for presentation purposes:

DGIWG	Defence Geospatial Information Working Group
ECDIS	Electronic Chart Display Information System
ECS	Electronic Chart System
ENC	Electronic Nautical Chart
IHO	International Hydrographic Organization
IMO	International Maritime Organization
SOLAS	Safety Of Life At Sea
TIN	Triangulated Irregular Network
UML	Universal Modelling Language

5.2 Notation

In this document conceptual schemas are presented in the Unified Modelling Language (UML). Several model elements used in this schema are defined in ISO standards developed by ISO TC 211, or in IHO S-100. In order to ensure that class names in the model are unique ISO TC/211 has adopted a convention or establishing a prefix to the names of classes that define the TC/211 defined UML package in which the UML class is defined. Since the IHO standards and this product specification make use of classes derived directly from the ISO standards this convention is also followed here. In the IHO standards the class names are identified by the name of the standard, such as "S100" as the prefix optionally followed by the bialpha prefix derived from ISO. For the classes defined in this document the prefix is "S10x". In order to avoid having multiple classes instantiating the same root classes, the ISO classes and S.100 classes have been used where possible; however, a new instantiated class is required if there is a need to alter a class or relationship to prevent a reverse coupling between the model elements introduced in this document and those defined in S.100 or the ISO model.

Table 1 - Sources of externally defined UML classes

Prefix	Standard	Package
CI	ISO 19115	Citation and Responsible Party
CV	ISO 19123	Coverage Core & Discrete Coverages
DQ	ISO 19115	Data Quality Information
DS	ISO 19115	Metadata Application Information
EX	ISO 19115	Metadata extent information
GF	ISO 19109	General Feature Model
GM	ISO 19107	Geometry Root
IF	ISO 19129	Imagery Gridded and Coverage Data Framework
LI	ISO 19115	Data Quality Information
MD	ISO 19115	Metadata entity set information
MI	ISO 19115-2	Metadata entity set imagery
S100	IHO S.100	IHO Standard for Hydrographic Data
SC	ISO 19111	Spatial Referencing by Coordinates
SD	ISO 19130	Sensor Model

6 Auxiliary Data to ENC and S.100

The S.100 standard permits the definition of many types of data related to navigation at sea or on inland waters. One type of vector data, corresponding to a paper navigational chart, is a special case in that it corresponds to the official paper chart and is required for safe navigation under the IMO SOLAS conventions. This type of data is defined by the product specification S.101 - ENC.

Many different types of auxiliary data may be defined as an aid to navigation. This additional information may be displayed together with all or some of the vector chart data from an ENC as specified in S.101. An Electronic Chart System (ECS) does not provide legal paper chart equivalency per the SOLAS conventions, but it is very useful to a mariner for planning and obtaining a broader sense of the environment. Whenever any auxiliary data is used with ENC data the system is no longer a paper chart equivalency. As such the system then operates as an ECS.

Auxiliary data is defined in layers each with its own product specification. Interaction with the user community[2] has identified high definition bathymetry as the most important auxiliary layer. Other

[2] The review of user requirements was done by the Canadian Hydrographic Service through the Mariner's Workshops in Canada.

important additional layers are real time water levels, currents and ice information, however there are many other potential data types. Each auxiliary layer type, of international interest, should be defined in its own product specification standardized through IHO. Additional types may also be defined nationally or in other international arenas. The WMO and the Ice services of those nations that have ice infested waters have already defined the required feature types for an Ice chart layer. The military Additional Military Layers, by the NATO GMWG, are special layers of data of interest in military applications.

The following table identifies several types of data that may be organized into auxiliary layers. These are given here only as an example of auxiliary layers to show the different types of data geometries involved. The data is organized spatially in terms of several different geometries. Bathymetry data which is described in S.102 is a continuous coverage quadrilateral grid. Many of the other data types identified in the table may also be handled as a quadrilateral grid so a general model for representing data in a quadrilateral grid coverage is given in this document.

Table 2 - Auxiliary Layers of Data and their Geometries

Data Type	Data Organization	Geometry Elements
Electronic Nautical Chart (per IHO **S-57, S-101**)	Vector (per ISO 19107)	Point, Curve, Surface
Bathymetry	Continuous Coverage (Per ISO 19123)	Quadrilateral Grid
Water Level	Continuous Coverage	Quadrilateral Grid (X,Y,T and L)
Tide Gauge & Tide Prediction	Continuous Coverage	Quadrilateral Grid
Real Time Surface Current	Point coverage	Vector Field (vectors of direction and intensity at point locations)
Current	Continuous Coverage	Quadrilateral Grid
Ice Chart	Discrete Coverage (scanned raster image of a paper product) Vector (using ICE Feature Catalogue)	Quadrilateral Grid Point, Curve, Surface
Ice Image	Continuous Coverage (RADARSAT Satellite Image)	Quadrilateral Grid
Ice Route	Vector Data (complementary to ENC)	Point, Curve
Ice Forecast	Continuous Coverage	Quadrilateral Grid
Ship own position	Vector	Point
Other ships positions	Vector	Points
Bottom Normative Bathymetry	Continuous Coverage	Quadrilateral Grid
Bottom material classification	Discrete Coverage (material type)	Quadrilateral Grid
Wind	Continuous Coverage	Quadrilateral Grid
Notice to shipping	Vector	Point data with Text attributes
wave model	Continuous Coverage	Quadrilateral Grid

The reference system must be the same for the data types identified above for the data to be integrated into a compound data set.

The various data types can be organized either as 2 dimensional vector data sets, or as continuous coverages or discrete coverages. Elevation (or depth) is expressed as an attribute (called 2.5 D because there can only be one elevation value per X, Y location in each layer.)

Vector data is represented by the geometry element **GM_Object** from the ISO 19107 and can be of type Point, Curve (Line) or Surface (Area). The feature data dictionary used to describe feature types is the IHO feature data dictionary and the complementary ICE feature object data dictionary. Any other features required for other feature types would have to be added to the IHO feature object dictionary or to a special feature dictionary and feature catalogue for that data type.

Coverage data is represented by the geometry element **CV_Coverage** from the ISO 19123 standard. Coverage data may be either continuous or discrete. Continuous coverages exist for situations where grid values are numerical entities and where interpolation can occur. For example, it is possible to interpolate intermediate water levels between tide gauges according to an interpolation formula. Discrete coverages exist for situations where grid cell values represent discrete quantities such as sand, rock, mud. A scanned paper ice chart is a discrete coverage because the grid cell values correspond to elements from a synthetic image that is discrete. One can not interpolate. Cells correspond to different types of ice as portrayed on the chart and to the shoreline and other lines from the original chart. In integrating data certain discrete coverage cell types can be retained and others discarded.

Several types of vector data elements (GM_Objects) are used. These are Point, Curve and Surface. Ships own position is the simplest geometry consisting of only a single point with a bearing and size thematic attribute. The S-101 ENC chart data uses Point, Curve and Surface. Topological objects are not used because they would need to be recalculated as part of the data integration if used. They can be calculated when required when the data set is processed in an extended Electronic Chart System (ECS)

Several different types of coverage geometry may be used for different classes of auxiliary data. This specification primarily addresses the Spatially Referenced Quadrilateral Grid (which is used for bathymetric data in S.102). The Riemann Hyperspatial Grid (Quad Tree) coverage, point set coverage and TIN coverage may be used for other organizations or types of data. The most complex continuous coverage type identified is a vector field which consists of a set of points at non uniform positions with a three dimensional vector as the attribute for each point location, representing the direction and value of the quantity, such as water flow (current).

Both the continuous and discrete referenced quadrilateral grid coverages may be used for different types of data. Since the difference between the continuous and discrete referenced quadrilateral grid is only the type of data these two data models are almost identical. The continuous coverage is used to express bathymetric data in S.102 and the discrete coverage is used to express a tiling system for all auxiliary layers.

The **2D vector data model** is identical to the S-100 vector model and is **not reproduced here**. The coverage types of uniform continuous quadrilateral grid, Quad Tree continuous quadrilateral grid Tin and Point Coverage shown in the following sections.

7 Data Structure for Coverage Data

7.1 Coverage function

A coverage is a subtype of feature in that it represents real world phenomena in terms of a set of attributes. The attribute values may be organized in a grid. These values drive the coverage function. A coverage function allows one to interpolate attribute values across a spatiotemporal domain. A continuous coverage function returns a distinct attribute value for each and every position in the domain. Bathymetry, as described in S.102, is inherently a continuous coverage since there are real values on the sea floor or river bottom between the data points that make up the grid. The Bathymetric Quad Grid Coverage represents a set of depth values assigned to the points in a 2D grid. The interpolation method to be used is described as part of the metadata associated with a coverage function. Many other auxiliary data types are also continuous coverages.

The ISO standards permit several organizations of grids with different grid traversal orders, and variable or fixed grid cell sizes. Two grids are of primary interest for the support of auxiliary layers to S.100. These are the simple quadrilateral grid with equal cell sizes traversed by a linear sequence rule, and the variable cell size quadrilateral grid traversed by a Morton Order sequence known as the Quad Tree for a two dimensional grid. The specification S.102 for Bathymetry only makes use of the simple quadrilateral grid with equal cell sizes. Two other coverage types, the Triangular Irregular Network (TIN) and the point set coverage are also of interest for the support of auxiliary layers to S.100..

7.2 Quadrilateral Grid Coverage

A quadrilateral grid coverage data set consists of a sequence of data values organized as a continuous quadrilateral grid coverage together with metadata. The data model describes the relationship between the various attributes and classes that describe the coverage. Except for the set of data values all of these attributes are just another form of metadata. They are attributes that must be provided along with a data set.

Figure 1 gives a simplified representation of the structure of a data set of gridded coverage data such as bathymetry data containing a set of gridded data elements and a set of metadata. The metadata set includes the metadata elements taken from the S-100 metadata standard as well as the attributes that define the grid, which are effectively the metadata of the grid. In some encoding schemes the attributes that define a grid are included with the grid, whereas in others, they are carried with the remaining metadata items.

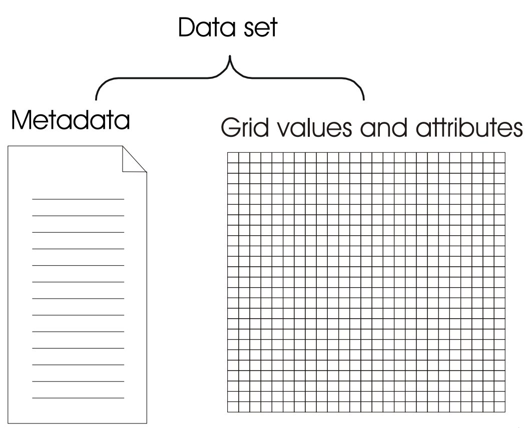

Figure 1 - Metadata and Data

The model in Figure 2 is a more detailed description of the same structure expressed using the UML modelling language and using the modelling elements defined in the ISO geographic information suite of standards and the IHO S-100 standard. It illustrates that a data set of S-100 compliant bathymetric data contains both metadata and a grid values matrix of gridded data. This model is similar to the model given in S.100 Part 3 clause 7.2 except that here the classes are established in the S.100 name space.

The model in Figure 2 states that:

> An S-100 data set (**S100_DataSet**) of auxiliary data references an S-100 Image and Gridded Data Collection (**S100_IGCollection**). The relationship allows a 1 to many (**1..***)

multiplicity on both ends of the relationship from the data set to the collection. Multiple data sets may reference the same collection and one data set may reference multiple collections. That is to say, that any amount of data may be included in a data set.

An S-100 Image and Gridded Data Collection (**S100_IGCollection**) is a group (collection/layer) of gridded data sets with a common theme as expressed in the metadata. Instances of the collection are described by a set of S100 Collection Metadata (**S100_CollectionMetadata**). This relationship also has multiplicity on both ends of the relationship. That is, multiple sets of metadata may describe the instances of the collection. The relationship shows navigability on both ends indicating that the collection refers to the metadata and the metadata refers back to the instance of the collection which it describes.

The metadata also refers to the coordinate reference system (**S100_CRS**). This information is carried in the metadata but it describes the location of the Quadrilateral Grid Coverage. This is shown by a second relationship from the Grid Coverage element to the coordinate reference system. The details of how a coordinate reference system is expressed is described in the standard ISO 19111.

The Quadrilateral Grid Coverage (**S100_GridCoverage**) is defined as a subtype of the S.100 Imagery and Gridded data types type (**S100_CoverageData**) as defined in the standard S.100. The Grid Coverage has two components the Grid Value Matrix and the Grid Value Cell. Both of these objects are also taken from ISO 19123. The Quadrilateral Grid Coverage is a component of the **S100_IGCollection**. That is the metadata and quadrilateral grid coverage make up the collection. The component relationship between **S100_IGCollection** and **S100_GridCoverage** derives from an equivalent relationship in ISO 19101-2 Reference Model - Imagery**.**

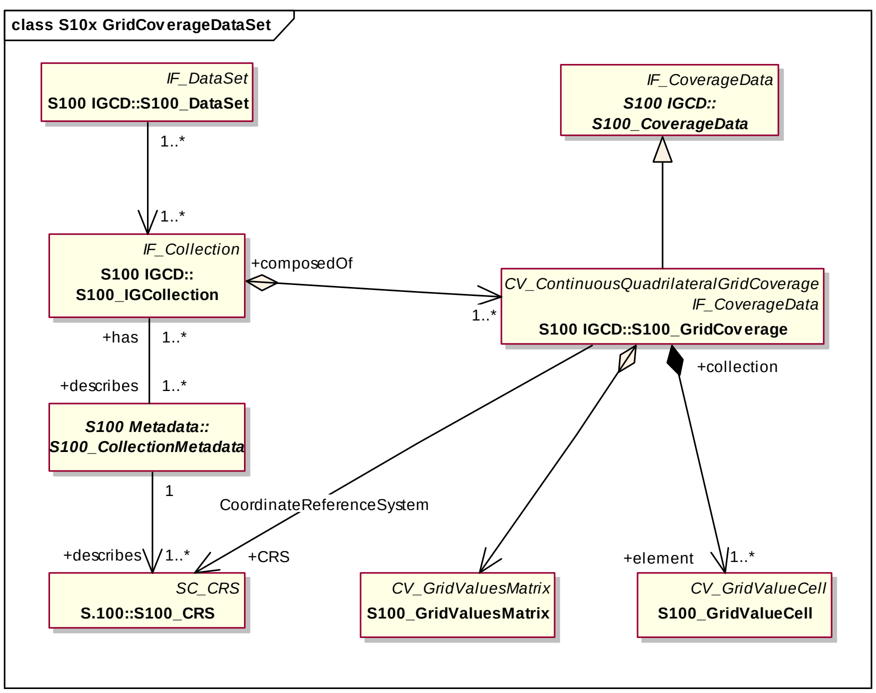

Figure 2- Grid Coverage Data Set with Associated Metadata

A more detailed model is shown in Figure 3 with the attributes that are inherited from the root classes defined in the ISO standards.

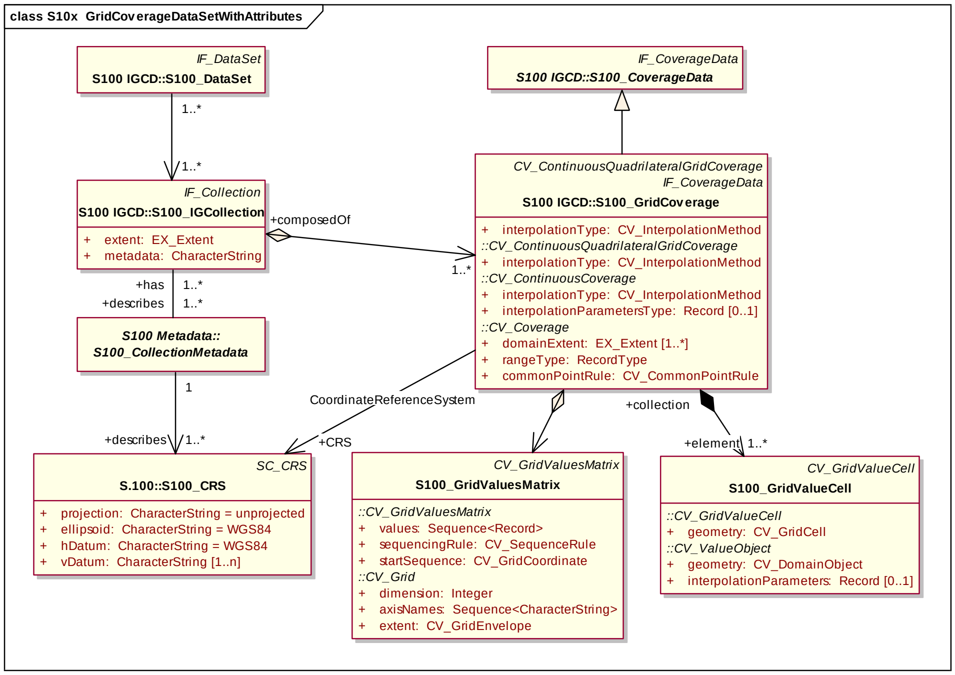

Figure 3- Grid Coverage Data Set with Attributes

This model shows that the data set is composed of instances of a subtype of **CV_ContinuousQuadrilateralGridCoverage from ISO 19123 and IF_CoverageData from ISO 19129**. The grid inherits the attributes: domainExtent, rangeType, commonPointRule, interpolationType, and interpolationParameter from its supertypes. The CV_GridValuesMatrix and the CV_GridValuesCell are components of the CV_ContinuousQuadrilateralGridCoverage. That is, the coverage is described by a matrix of grid cells. CV_GridCell describes the geometry (size) of a grid cell and the matrix contains a sequence of values.

The attribute ***domainExtent*** describes the spatial extent of the domain coverage. The data type EX_Extent is defined in ISO 19115 as part of the metadata.

The attribute ***rangeType*** describes the range of the coverage. It uses the data type RecordType specified in ISO/TS 19103. An instance of RecordType is a list of name data type pairs each of which describes an attribute type included in the range of the coverage. The name field is used to identify the type of the surface that each value describes. For bathymetry the range type might be "bathymetry:Real" for depth measurements, however, other range types might be used for other types of gridded data.

The attribute ***commonPointRule*** identifies a method for resolving potential conflicts between attribute values resulting from evaluation of a coverage at a direct position when that position falls on the boundary between two value objects, such the edge of two grid cells. A code list **CV_CommonPointRule** is specified in ISO 19123. For elevation coverages appropriate values of the CV_CommonPoint Rule include 'average', 'high', and 'low'. For bathymetry the appropriate value would be "low", meaning the lesser depth value, to ensure that depths are given at their shallowest value to promote safety of navigation. In the case of an elevation model for air charts the value might be "high" to ensure that vertical air obstructions are emphasised. For a coverage such as currents the value might be "average" to indicate the average of the overlapping values.

The class S100_GridCoverage also inherits a relationship to the coordinate reference system **S100_CRS which inherits from SC_CRS**. The coordinate reference system is defined in ISO 19111.

The attribute ***interpolationType*** describes the interpolation method recommended for evaluation of the Grid Coverage. For bathymetry and similar data the value is either 'bilinear' or 'bicubic'.

> **Bilinear interpolation** is used to interpolate feature attribute values at direct positions within a quadrilateral grid using the function:
>
> $v = a_0 + a_1x + a_2y + a_3xy$
>
> **Bicubic interpolation** is also used to compute feature attribute values at direct positions within a quadrilateral grid. Bicubic interpolation uses the function:
>
> $v = a_0 + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2 + a_6x^2y + a_7xy^2 + a_8x^2y^2 + a_9x^3 + a_{10}y^3 + a_{11}x^3y + a_{12}xy^3 + a_{13}x^3y^2 + a_{14}x^2y^3 + a_{15}x^3y^3$
>
> ISO 19123 references sources for algorithms for implementing bilinear and bicubic interpolation.

An S100 quadrilateral grid coverage is a type of **CV_ContinuousQuadrilateralGridCoverage** as defined in ISO 19123. The only thing that distinguishes one grid coverage from another is the meaning of the attributes. For example for a bathymetry grid coverage the attributes in the sequence of records in the Grid Values Matrix represent bathymetry depth values. Depth is a measurement from the water surface toward the centre of the earth. Depth values are positive numbers representing this concept. Depths are the opposite of elevation values that are measured in the direction away from the centre of the earth.

The object **CV_GridValueCell** has the attribute ***geometry*** which defines the geometry of the grid cell. Four grid points maybe used to control the shape of a cell.

The object **CV_GridValuesMatrix** contains the actual matrix of grid values. The attribute *values* is a sequence of value records. Each **Record** may contain one or more values at a particular grid cell. For a single valued coverage, the records consist of a single value per cell. The attribute ***sequenceRule*** indicates the order of the attribute values in the sequence. The simplest rule is linear (Row, Column), but more complex rules may be used. For example a Quad Tree Variable cell size grid is traversed in Morton order. The attribute ***startSequence*** gives the position of the first cell in the sequence.

Attributes are also inherited from the supertype class **CV_Grid**. The attribute dimension describes the dimension of the grid, which for bathymetry is 2 dimensions. The attribute ***axisNames*** defines the names of the axis, which are here Latitude and Longitude. The attribute ***extent*** describes an envelope encompassing the matrix of grid values.

7.3 Quadrilateral Uniform Grid Model

In order to implement the model described in Figure 3 a specific implementation class has been defined in S.100 called S100_Grid. This class is used for a bathymetry grid in the S.102 data layer specification. That is, the S.102 bathymetry data may be used as an auxiliary data layer together with S.101 or other S.100 compatible data. (It may also be used as a stand alone set of information).

The class S100_Grid shown in Figure 4 is a realisation of the class **CV_GridMatrix**. Two types of grids are possible, rectified grids and referencable grids. Rectified grids have a direct linear relationship to a coordinate reference system, whereas a referencable grid requires the use of a transform to convert grid coordinate values to values that can be referenced to a coordinate reference system. Since a bathymetry grid is also a rectified grid it also realizes the class **CV_RectifiedGrid** and inherits the two attributes ***origin*** and ***offsetVectors***.

Figure 4 illustrates the S100_Grid class showing its attributes, as derived from CV_RectifiedGrid and CV_GridValuesMatrix. The actual sequence of values has been extracted into a separate component S100_GridValues. The values are each represented by a Record. A Record is a data type, defined in ISO 19103, that consists of a set of individual data values for the grid cell. Only one value, the depth is mandatory. Other values may be included in the record such as a reliability or quality indicator number.

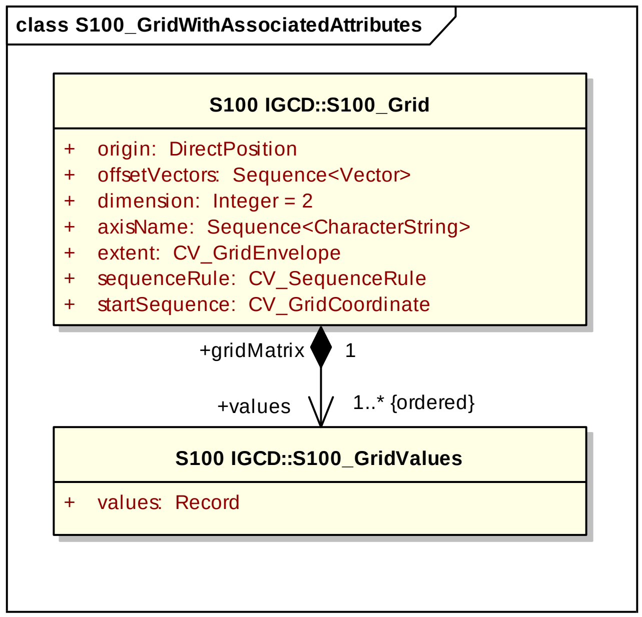

Figure 4 - S100 Grid with Associated Attributes

The attribute ***dimension*** specifies the dimension of the elevation grid. Bathymetry grids are restricted to two dimensions so the value of this attribute shall be "2".

The attribute ***axisNames*** specifies the names of the grid axes.

The attribute ***origin*** specifies the coordinates of the grid origin with respect to an external coordinate system. The data type DirectPosition, specified in ISO 19107, has an association through the role name *coordinateReferenceSyetem* to the class SC_CRS specified in ISO 19111 which specifies the external coordinate reference system. This information is described in the metadata. S100_CRS is equivalent to SC_CRS.

The attribute ***offsetVectors*** specifies the spacing between grid points and the orientation of the grid axis with respect to the external coordinate reference system identified through the attribute *origin*. It uses the data type Vector specified in ISO/TS 19103. For simple grids with equal cell sizes the offset vector establishes the cell size. For variable cell size grids (Quad Tree grids) the offset vector establishes the minimum cell size. The actual cell size is included as an attribute in the data record that describes the level of aggregation of the quad structure. The attribute ***offsetVectors*** implements the geometry of a **CV_GridValueCell** for a simple quadrilateral grid with equal cell sizes.

The attribute ***extent*** specifies the area of the grid for which elevation data are provided. It uses the type **CV_GridEnvelope** specified in 19123 to provide both the **CV_GridCoordinates** of the corner of the area having the lowest grid coordinate values and the **CV_GridCopordinates** of the corner of the area having the highest grid coordinate values. **CV_GridCoordinate** is specified in 19123.

The attribute ***extent*** effectively defines a bounding rectangle describing where data is provided. For simple grids with equal cell sizes, if data is not available for the whole area within this rectangle, then padding with null values shall be used to represent areas where no data is available.

The attribute ***sequencingRule*** specifies the method to be used to assign values from the sequence of elevation values to the grid coordinates. It uses the data type **CV_SequenceRule** specified in ISO 19123. Only the value "linear" (for a **uniform** regular cell size grid) shall be used. The sequence rule for a regular cell size grid is simple. Since the cells are all of the same size the cell index can be derived from the position of the Record within the sequence of Records.

The attribute ***startSequence*** identifies a value of **CV_GridCoordinate** to specify the grid coordinates of the grid point to which the first in the sequence of elevation values is to be assigned. The choice of a valid point for the start sequence is determined by the sequencing rule.

The class **S100_GridValues** is a separate component describing the set of data values that apply to the grid matrix. The attribute ***values*** shall be a sequence of **Records** each containing one or more depth values to be assigned to a single grid point. The Record shall conform to the **RecordType** specified by the ***rangeType*** attribute of the **S100_GridCoverage** with which the **S100_Grid** is associated.

For simple grids with equal cell sizes the **sequenceRule** attribute of an **S100_Grid** equals "linear" and the offset vector establishes the cell size. The attribute ***extent*** specifies the area of the grid for which elevation data is provided.

7.4 Quad Tree Grid Coverage

A Quad Tree is a variation on the quadrilateral grid with variable cell sizes traversed by a Morton Order sequence for a two dimensional grid. This has particular use for some types of data since areas that have similar attribute values, such as bottom cover in a bathymetry data set, may be represented by large cells and areas of rapid change may be represented with smaller cells. This can result in a large compaction[3] of data. It also gives an order to the space. The variable cell size nature of a Quad Tree quadrilateral grid is illustrated in Figure 5.

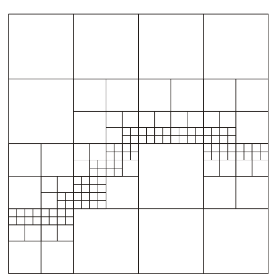

Figure 5 - Example of a Quad Tree Quadrilateral Grid

[3] Note: A compaction of data is different than a compression of data since the structure of the data needed to represent the information is less. In compaction there is less data. In compression there are fewer bits representing the same amount of data. This is addressed in ISO 19129.

The Quad Tree grid may be extended to a multidimensional Hyper Tree where bounded attribute values such as depth are represented as dimensions in hyperspace. There are advantage to sorting and processing such hyperspatial data.

The model given in Figure 4 also applies to a Quad Tree type of quadrilateral grid. The only difference is that the attribute value for **CV_SequenceRule** becomes "Morton" and the attribute ***offsetVectors*** specifies the minimum cell size.

7.5 TIN Coverage

A Triangular Irregular Network (TIN) provides a variable density coverage that can be particularly useful for data analysis. It consists of a set of triangles covering an area. The variable size triangle nature of a TIN coverage is illustrated in Figure 6.

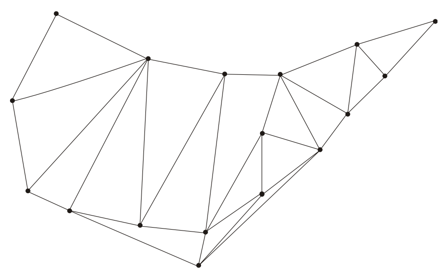

Figure 6 - Example of a TIN Coverage

A template model for a TIN coverage is given in S.100 Part 8. Figure 7 illustrates the S100_TIN_Coverage class showing its attributes and inherited attributes, together with the S100_Triangle class and the S100_VertexPoint class, The S100_TIN_Coverage class inherits from CV_TINCoverage from CV_ContinuousCoverage in ISO 19123, and S100_Triangle inherits from CV_ValueTriangle from CV_ValueObject from ISO 19123.

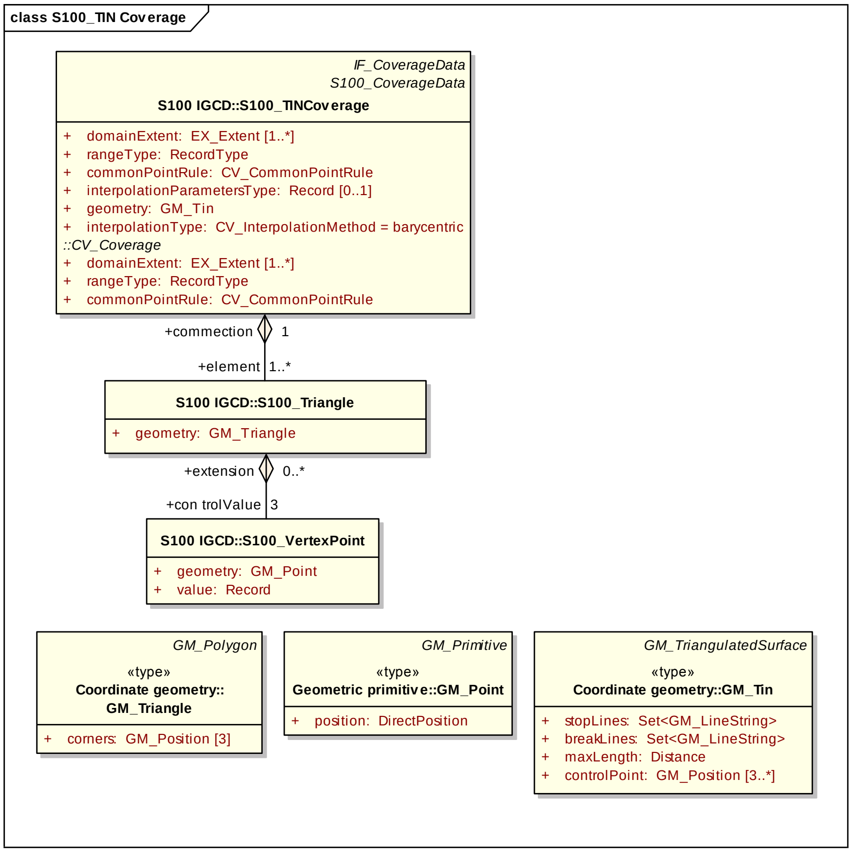

Figure 7 - TIN Coverage

The attribute **domain*extent*** specifies the area of the TIN coverage for which data are provided. It uses the type **EX_Extent** specified in 19115 whose attribute **description** is a character string used to describe the extent.

The attribute ***commonPointRule*** specifies the method to be used to assign a value when there are more than one vertex point at a position, such as might occur when two TIN coverages are edge matched. For bathymetry the "low" value is used to ensure safety of navigation. Otherwise the "average" value is used.

The attribute ***interpretationMethod*** takes a value from the code list provided in **CV_ *InterpretationMethod*** from ISO 19123 and is "barycentric" for a TIN coverage.

The attribute ***geometry*** takes a value from the class **GM_TIN** from ISO 19107. It allows the description of additional parameters to describe a TIN.

The class **S100_Triangle** is a separate class describing each triangle in the set of triangles. The attribute **geometry** has the value type **GM_Triangle** which has 3 corners represented as **GM_Positions** using the class **S100_VertexPoint**.

The class **S100_VertexPoint** is a separate class describing triangles vertex. The attribute **geometry** has the value type **GM_Point** which is a coordinate position represented by DirectPosition from ISO 19107.

The attribute ***values*** shall be a sequence of **Records** each containing one or more depth or other values to be assigned to a single triangle vertex point.

7.6 Point Coverage

A point set coverage provides a variable density coverage based on a set of point measurements. This is the classical method of describing bottom cover by hydrographic soundings. A set of soundings is a discrete coverage since it consists of an aggregation of points, but if it is intended to represent the bottom surface it is permissible to interpolate between the points and generate a continuous surface.

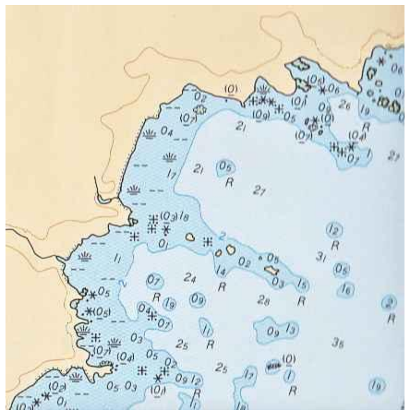

Figure 8 - Example of a Point Coverage as Soundings

Note chart example not for navigation.

A template model for a Point coverage is given in S.100. Figure 9 illustrates the S100_Point_Coverage class showing its attributes, together with the S100_VertexPoint class, The S100_Point_Coverage class inherits from CV_DiscreteCoverage in ISO 19123, and S100_VertexPoint inherits from CV_GeometryValuePair from ISO 19123.

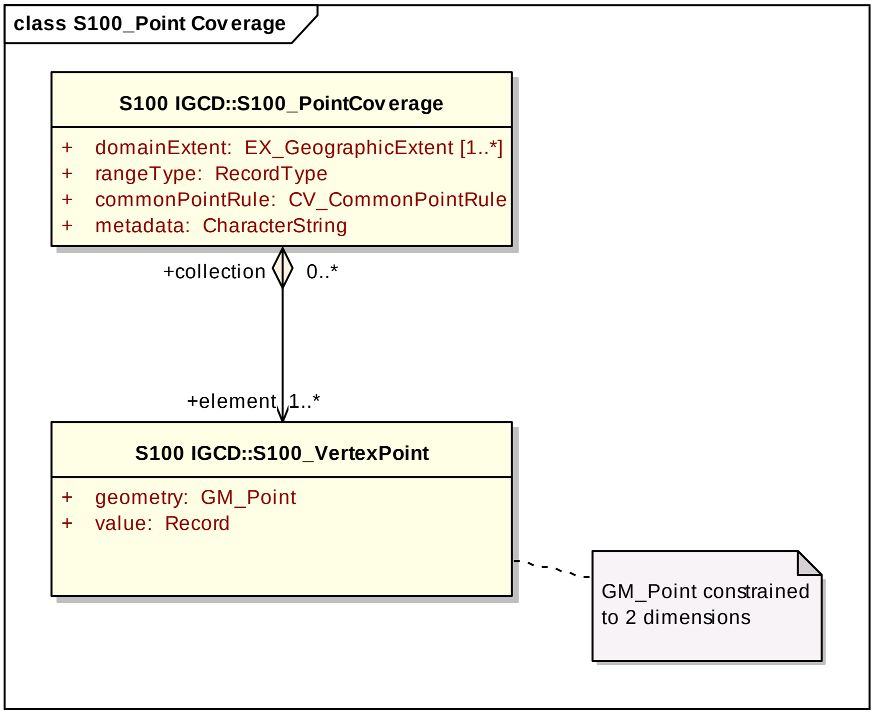

Figure 9 - Point Coverage

The attribute **domain*extent*** specifies the area of the Point coverage for which data are provided. It uses the type **EX_GeographicExtent** specified in 19115.

The attribute ***commonPointRule*** specifies the method to be used to assign a value when there are more than one vertex point at a position, such as might occur when two Point coverages are edge matched. For bathymetry the "low" value (low depth = not as deep) is used to ensure safety of navigation. Otherwise the "average" value is used.

The attribute ***metadata*** is a character string identifying any additional metadata about the point coverage.

The class **S100_VertexPoint** is a separate class describing triangles vertex. The attribute **geometry** has the value type **GM_Point** which is a coordinate position represented by DirectPosition from ISO 19107.

The attribute ***values*** shall be a sequence of **Record** each containing one or more values to be assigned to a single vertex point.

7.7 Vector Field Point Coverage

A vertex point in a point coverage can contain more than a simple value such as a depth. For example it could contain a vector quantity such as the strength and direction of flow to represent a current. In this case the coverage becomes a vector field. Figure 10 illustrate a vector field of currents as might be used in an auxiliary layer of data.

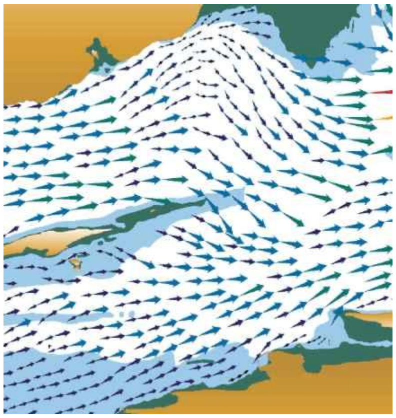

Figure 10 - Example of a Vector Field Point Coverage for Currents

The model for a vector field point coverage is the same as for any other point coverage and is given in Figure 9. However for a vector field the attribute ***values*** contains a **Record** that describes the vectorial quantity.

8 Tiling

A tile is a set of data that is edge-matched with other data within a tiling scheme. This allows for the handling of data of different densities or data over an extended or irregularly shaped area. A common tiling scheme is a prerequisite for establishing a set of common auxiliary layers that may be applied to hydrographic data. The tiling scheme must match the tiling scheme that the ENC data is delivered in, or the ENC data must be cut to fit a new tiling scheme as the base data for the set of auxiliary layers. This clause on tiling schemes is general to all of the product specifications that build auxiliary layers.

A tile naming convention is required to uniquely identify the tiles. This is a common issue to the cell (tile) naming issues in ENC.

Tiles must fit together. Some producers of ENC data provide additional data beyond a tile to ensure an overlap and identify the actual valid area using the MCOVR attribute. This approach is acceptable for ENC data as long as the data identified by MCOVR fits together in the tiling coverage.

A tiling scheme is effectively a second higher level of grid. That is, a tiling scheme behaves as a discrete grid coverage where the grid elements are themselves grid coverages.

The simplest approach to handling higher density data is to build a tiling mechanism that supports both higher and low-density tiles. This means that higher density data could be mixed with low-density conventional data. That is, certain areas may be surveyed in high density and this data may be combined with the low-density data in a single coverage. This is simple since cells at the boundary of tiles may be coincident with cells of different resolution. One must consider the cells as data samples at a varying-sampling rate. There will be a need to conflate the surfaces using rules based on accuracy of the surfaces. Variable density tiling is illustrated in Figure 11.

Note that a Quad Tree could be used to produce a variable size tiling system, or multiple discrete tile grids, one for each tile density, could be overlaid. The later approach is often the simpler.

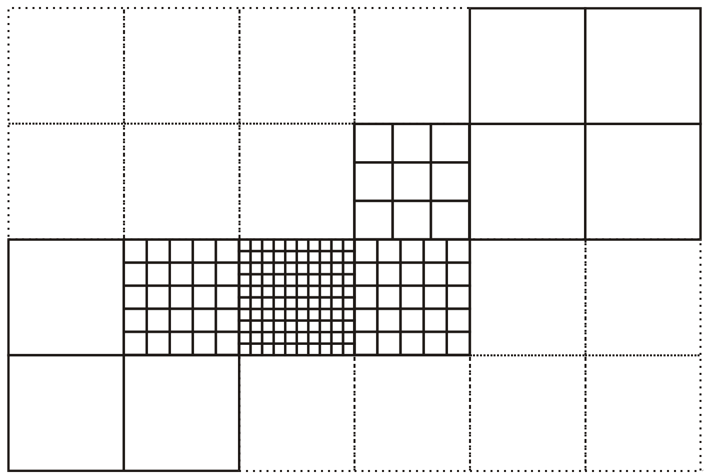

Figure 11 - Variable Tiling Density using a Quad Tree approach

8.1 A Quadrilateral Grid tiling scheme

An example of a discrete quadrilateral grid used as a tiling scheme is shown in Figure 12. This example is the general tiling scheme for the St Lawrence river area as used in the CHS St Lawrence Project. The figure is just an approximation to illustrate how tiling will work, and does not illustrate the correct placement or size of tiles.

The type of tiling scheme shown is a simple quadrilateral grid with equal cell sizes traversed by a linear sequence grid rule. A particular data set may include a few (2 to 4) tiles. The tiling scheme illustrated is a three level system including Harbour, Coastal and Overview tiles. This is accomplished by having three separate discrete quadrilateral grid coverages that are aligned and have discrete cell sizes corresponding to each of the tile sizes.

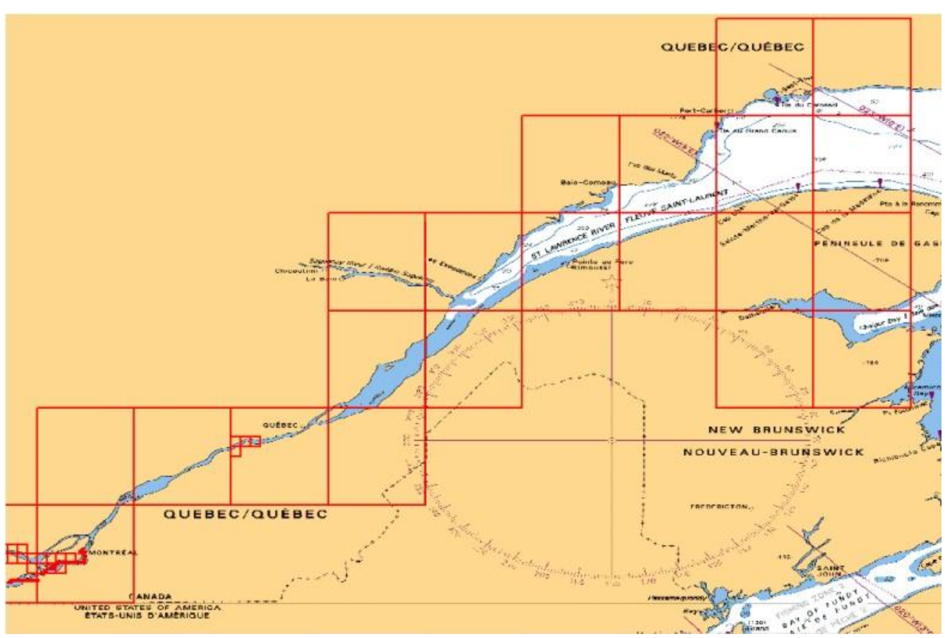

Figure 12 - St Lawrence Tiling Example

8.2 Tiling Model

There are two elements to the tiling model, the description of the tiling scheme and the identification of each individual tile. Figure 13 shows that a tiling scheme is associated with a set of data forming a **S100_IGCollection**. The identification of an individual tile is associated with the metadata for a particular instance of the collection.

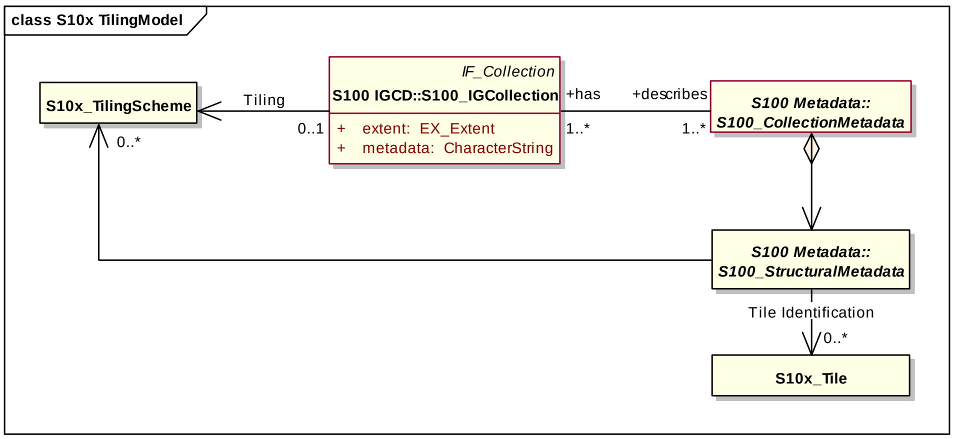

Figure 13 - Tiling Model

An individual tile may be identified either by a tile ID or by the tile boundary. The tile boundary is described by the geometry type **GM_CurveBoundary**. This is illustrated in Figure 14.

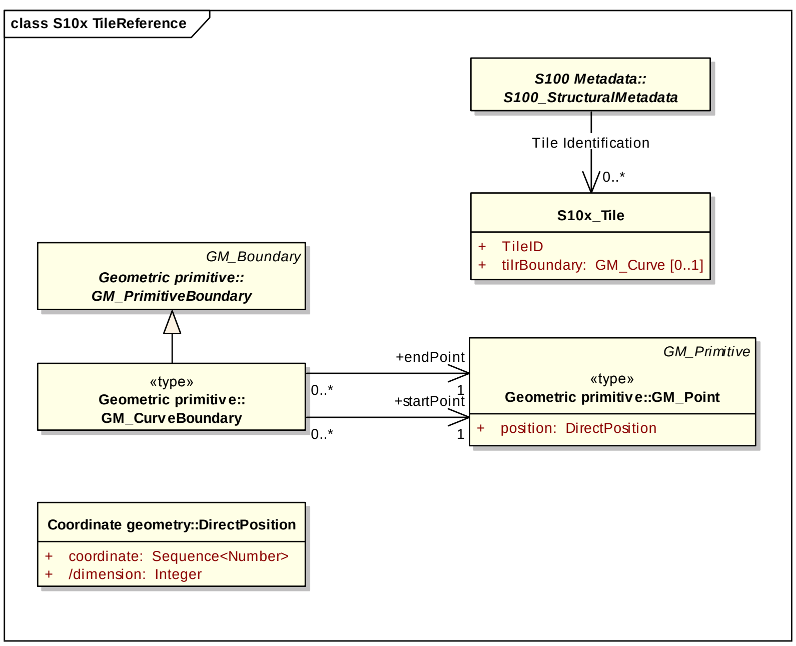

Figure 14 - Tile Reference

The full description of the tiling scheme requires the description of a complete discrete coverage that has as its elements the tiles. The S10x_TilingScheme is a realization of the CV_GridValuesMatrix and the CV_DiscreteSurfaceCoverage. This is illustrated in Figure 15. The description of the tiling scheme includes similar attributes to those used to describe the gridded data within each tile. The difference is that the attributes apply to the larger tile grid and not to the grid data points in the data set(s).

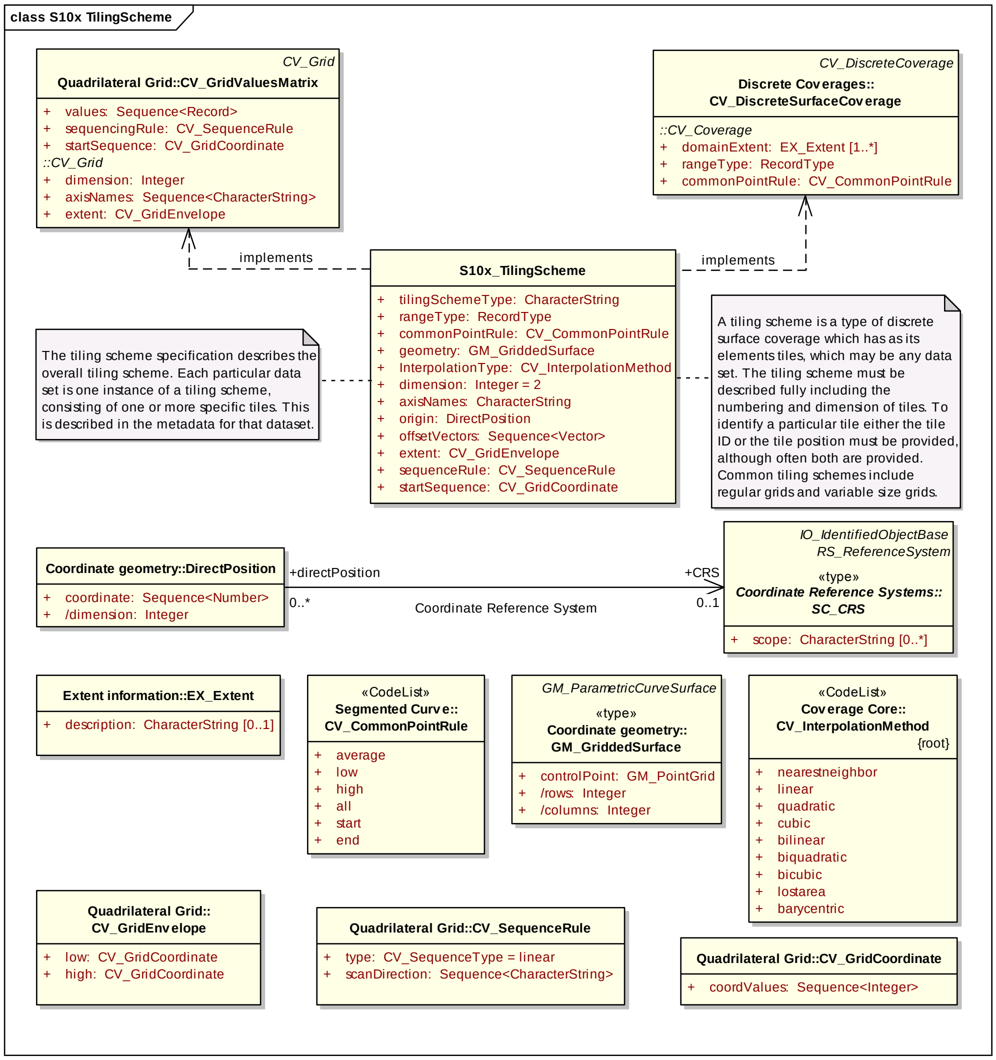

Figure 15 - Tiling Scheme

9 Coordinate Reference System

9.1 Types of referencing

Position relative to the earth is described in terms of a coordinate reference system. Three-dimensional geospatial data positions are usually referenced to a compound coordinate reference system (ISO 19111) consisting of a 2-dimensional horizontal reference system and a one-dimensional vertical reference system.

9.2 Horizontal reference systems

There are many possible horizontal reference systems. A register of such systems is being established by ISO TC211, and industrial registers already exist such as through the European Petroleum Survey Group (EPSG). IHO describes its specification of Coordinate Reference Systems in S.100 Part 6. Where necessary different coordinate reference systems may be used. However, for the aggregation of auxiliary layers to work it is necessary for there to be one common horizontal reference system for all of the layers.

The default horizontal reference system established for S.10x compliant sets of auxiliary layers shall be the World Geodetic System 1984 (WGS84). This shall be identified in the metadata.

9.3 Vertical reference systems

A vertical reference system consists of a surface identified as a datum from which heights are measured and an axis normal to the surface through the point for which the height is stated. The WGS 84 ellipsoid, which is the fundamental datum used by the Global Positioning System, is commonly used as the vertical reference surface for elevation measurements on land. Reference to a constant gravity surface is useful in terrestrial applications. Since water depths vary with the tides, bathymetric data is normally referenced to a datum defined in terms of tide state. Bathymetric data in a river or lake is often referred to a special river or lake datum, such as the Great Lakes Datum.

In order for the auxiliary data described in this product specification to be able to be used with the base vector data (derived from or from) ENC, the vertical datum must be the same as described for the ENC data. The vertical datum shall be referenced in the metadata.

The use of a sounding datum based on sea level as opposed to an ellipsoid (or geoid) height is the major difference between hydrographic data and land elevation data. The measurement of depth from sea level is important for navigation, and since the level of the sea varies with the tides and other conditions the lowest level is normally used to provide a safety margin. Because the land elevation is measured to a different reference the coastline as determined from the land and from the sea is seldom coincident. Data over the land and over the sea can be used together as long as the user is aware that they are normally measured against different references. That is, if auxiliary data layers combine information on land and over the water then it is necessary to make the user aware of the possibly different coastlines that may result from the two measurement surfaces. This is of primary importance for sets of auxiliary data such as the NATO AML that may include a transition from the sea to the land. If they are different then the vertical reference systems of the various layers of auxiliary data must be related as part of the auxiliary data. Additional features may be required to show areas, such as the difference between the land referenced shoreline and the sea referenced shoreline.

10 Data Set Structure

10.1 Data Set Components

All of the auxiliary data sets have a similar structure, both vector and coverage data. A Data Set may contain one or more collections. The collection is related to the tiling scheme as described in clause 8. The collection shown is the IGCollection but the collection applies equally to vector based data. A component is the set of features comprising the layer of vector or coverage data.

Metadata is associated at several levels. Metadata maybe associated with the data set as a whole, or with the coverage or vector data. Metadata may also be associated with particular data elements where needed as part of the data value record in a coverage or on particular GM_Object vector elements. More detailed metadata at a lower level overrides general metadata for an entire collection. Metadata may also be associated with particular tiles.

Metadata is organized into modules. The organization into modules is important for managing auxiliary layers of data because some metadata crosses layer boundaries and some is related only to particular layers. The modules separate out the metadata elements that apply only to a single layer from the metadata that must be integrated across layers. The Discovery Metadata Module relates to the data set as a whole whereas other metadata applies to the data collection (**S100_IGCollection**). The Collection Metadata Module refers to the Discovery Metadata Module, the Structure Metadata Module, the Acquisition Metadata Module and the Quality Metadata module as sub-components. That is, some elements from the Discovery Metadata Module may also be repeated or overwritten at the collection level.

The overall structure of a data set is illustrated in Figure 16 below.

In a data exchange format there are three classes that need to be implemented for each tile in a collection. These are:

S10x_TilingScheme
S100_VectorData or S100_CoverageData component, and
S100_CollectionMetadata

The S100_Collection Metadata, of course includes the elements used from the four metadata modules, discovery, structural (and tile ID), acquisition and quality metadata. For the entire data set the S100_DiscoveryMetadata class must be implemented.

In any particular transmittal (data exchange) a whole or part of a data set, or several data sets may be communicated. This is identified in the S100_Transmittal. The transmittal is what is actually communicated and it is the transmittal that needs protection from inappropriate copying or use.

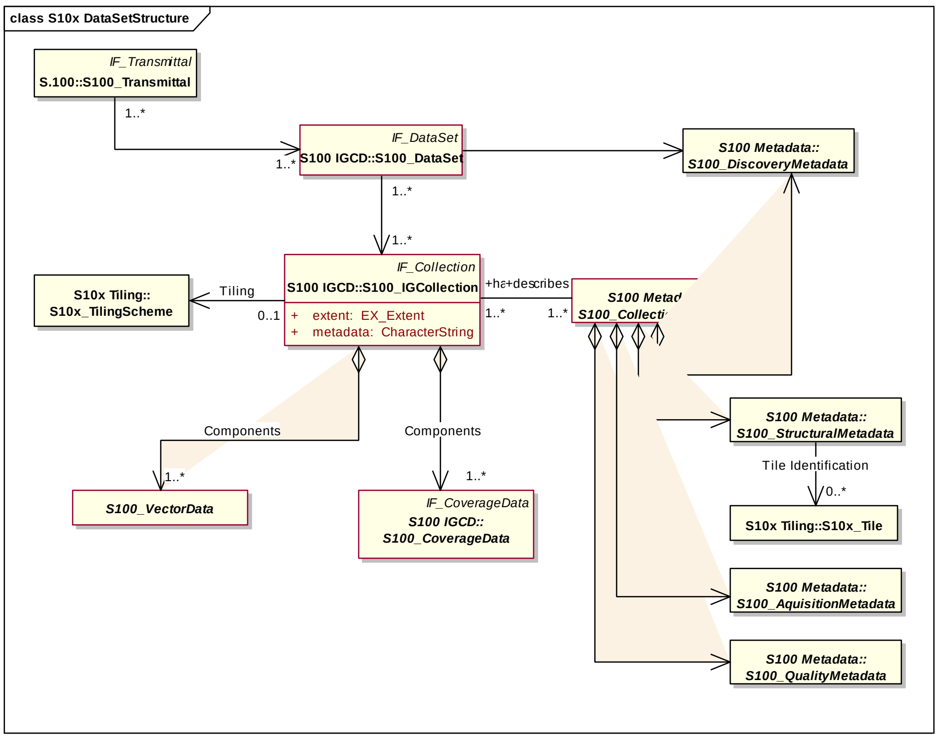

Figure 16- Data Set Structure

10.1.1 Data Set Class

A data set is an identifiable collection of data that can be represented in an exchange format or stored on a storage media. A data set can represent all or a part of a logical data collection and may include one or many tiles of data. The content of a data set is defined by the Product Specification (per ISO 19131) for that particular type of data and is normally suited to the use of that data. A product specification for a particular data type needs to indicate the organization of that data product. For example, a simple gridded bathymetry product may have only one elevation grid coverage, and a tiling scheme that indicates that every data set contains one tile. More complex products may include several collocated coverages and more complex tiling schemes such as a quad tree based variable size tiling scheme, where one data set may contain more than one tile. The data set is the logical entity that can be identified by the associated discovery metadata, not the physical entity of exchange.

10.1.2 S100_Discovery Metadata Module

Associated with a data set is a set of discovery metadata that describes the data set so that it can be accessed. It consists of the "core" metadata defined in ISO 19115.

10.1.3 S100_Transmittal

A transmittal is the encoded exchange format used to carry all, part of, or several data sets. It represents the physical entity of exchange. The transmittal is dependent upon the encoding format and the exchange media. A transmittal on a physical media such as a DVD may carry a number of data sets, whereas a transmittal over a low bandwidth telecommunications line may carry only a small part of a data set. Any metadata carried with a transmittal is integral to the transmittal and may be changed by the exchange mechanism to other exchange metadata as required for the routing and delivery of the transmittal. A common exchange mechanism would be to carry a whole data set on one physical media such as a CD-ROM. Transmittal metadata is not shown because any transmittal metadata, exclusive of the information in the Discovery Metadata Module, is dependent upon the mechanism used for exchange, and may differ from one exchange media or encoding format to another. An example of transmittal Metadata would be counts of the number of data bytes in a unit of exchange.

10.1.4 S100_IGCollection

An S100_IGCollection represents a collection of bathymetry data. A collection may include multiple different data types over a particular area, or multiple coverages of data of the same coverage type, but representing different surfaces. For example a collection may consist of a grid coverage and a point set over the same area, where the grid coverage represents an elevation surface and the point set a number of soundings.

10.1.5 S100_Collection Metadata Module

Associated with an S100_IGCollection is a set of collection metadata that describes the data product as represented in the collection. It consists of a number of sub-components that include the Discovery Metadata Module as well as the Structure Metadata Module, the Acquisition Metadata Module and the Quality Metadata Module. Metadata from the Discovery Metadata Module may be applied to an elevation collection so that the entire collection may be discovered. The other metadata modules are descriptive metadata defined in ISO 19115.

11 Data Certification and Rights Management

The concept of separating data content from its carrier derives from the ISO Open System Interconnection (OSI) architecture from ISO 7498 and ISO 7498-2 which relates to secure communications. The OSI model defines seven layers that apply to a communications system. These are the Application, Presentation, Session, Transport, Network, Data Link and Physical layers. Peer processes at each layer communicate with their equivalent peer processes at the corresponding layer in the other system in a communication exchange. The layers are plug compatible. That is, one could communicate over an internet link and use TCP/IP and Ethernet to provide the lower communications layers, communicate over a modem on a telephone line using SLIP or load the same data on a media such as a CD-ROM. Modern communications systems depend upon the interchangability of layered protocols to provide flexibility. All layers are not present (or are null) in all systems.

Table 3 - OSI Layers

OSI Model		
Information type	Layer	Purpose
Data	7. Application	Content as used by an application
	6. Presentation	Data encoding or representation including encryption
	5. Session	System to system interaction
Segment	4. Transport	End to end connection and data flow control
Packet	3. Network	Path through network
Frame	2. Data Link	Physical network addressing
Bit	1. Physical	Media, and signals

Layers provide services to other layers. That is, the Presentation layer provides an encoding service to the Application layer. The session layer provides control over the communications dialogue. For example, dialling a telephone is a session layer function, whereas the analogue or digital voice data is a presentation layer encoding of the voice message at the application layer.

Digital Rights Management is Session layer functions that must be supported by Presentation layer encoding and an Application layer structure to support the Digital Rights Management functions. These functions need encodings and application layer structures to operate, but they are separate entities from the information content because they support a separate session level function. Telephone dialling also needs an encoding of touch tones, but that is separate from the voice message. The two would interfere if they went on at the same time.

Digital Rights Management provides a method of ensuring that the owner of the data retains the rights to the data. With respect to ENC data and Auxiliary data layers this means either data encryption or verification. Encryption protects a whole set of data so that only an authorized user may have access to it. A Digital Signature is one type of verification where only a verification function is encrypted so that one can verify that the data has not been altered (intentionally or accidentally). Because they support a separate function Encryption and Verification do not fit simply into the Geographic Information content model upon which S.100 is based.

Encryption and Verification can be applied in different ways dependent upon what needs to be protected and why. Two use cases are identified here. The data content layers identified in product specifications making use of S.10x can be used in different systems and different Digital Rights Management approaches may be applied.

Use Case 1 - Rights Protection for the Sale of Data

The business model for the sale of ENC data is different in various countries. In some the information is sold, and there is a need to provide an encryption around all or most of the data. Certain metadata may be excluded so the data can be identified. The encryption envelope is related to what is communicated, so it could relate to the **S100_Transmittal** which may be all, a part of or several set of data in one communications. Alternatively, encryption could relate to particular instances of **S100_DataSet**. This would preclude communicating only a part of a data set and require multiple encryption envelopes for multiple data sets. The use of encryption is so related to the business model that it should not be allowed by S100 and S01x, but should be external to the standard.

Use Case 2 - Verification

Navigational information is critical and it is important to users that they know that they have uncorrupted verified data. A digital signature is an extra data element that contains a function derived from the data that is encrypted. It can be encrypted using public key encryption techniques so it may be decoded by anyone, but only generated by a verified source. A Digital Signature is closely linked to the data. A verification key links the data through the metadata's lineage section. This allows a description of the purpose and other information about the Digital Signature to be recorded.

12 Metadata

12.1 Metadata Context

The metadata associated with a layer of auxiliary data is a profile of the S-100 metadata specification, which itself is taken from the ISO 19115 Metadata specification and the ISO 19115-2 Imagery Metadata specification. The quality metadata elements make use of the metadata classes defined in ISO 19115 with the attributes described in ISO 19138 Data Quality Measures. The acquisition metadata makes use of metadata classes defined in ISO 19115 Metadata together with attributes defined in ISO 19130 Imagery sensor models for geopositioning. Since ISO is developing 19130-2 to cover sonar and other methods of acquisition of direct interest to hydrography, there may be some future enhancements to this area of metadata.

The ISO standard on profiling 19106 defines two classes of profile. A class 1 profile is a pure subset of a base standard, whereas a class 2 profile allows for the extension of the metadata classes within the context defined in the base standard. It is permitted in a class 1 profile to make character strings into code lists, and to introduce other constraints. A class 2 profile can add new metadata and other classes of information. The intent is to promote interoperability. A system that implements a base standard should be able to recognize and process all of the data fields in a profile. If a profile is of class 1 and consists only of constraints, or codification of lists, then a system built to the base standard can process all of the metadata fields at least to the level specified in the base standard. If a data set makes use of a profile of class 2 then a system built to a base standard can recognize and process all of the data fields defined in the base standard and gracefully ignore any extensions. That is, any extensions must provide supplementary information in a form that can be recognized by a system implementing the base standard and ignored without invalidating the context of the other information.

The IHO S-100 metadata is a class 1 profile of ISO 19115 Metadata. It also adds some metadata elements from ISO 19115-2. As such any system compliant with the ISO 19115 standard can recognize and process any data defined using the IHO S-100 standard. For example the ISO standard may contain a character string data field to express the source of a data set, where IHO includes a code list of the IHO member states as sources of hydrographic data.

A product specification for bathymetric data (S.102) may be a class 2 profile of the ISO standards because it would incorporate data fields for other facets of the information not necessarily addressed in the collection of ISO standards profiled in the product specification. An application such as data discovery that only makes use of a subset of the ISO metadata would operate on the product because it would be a compliant class 1 profile within that context, however other applications that made use of the entire set of information within the product may be specific applications such as navigation.

Metadata is broken down into several modules.. Metadata may be applied at several levels within the structure of a data set. It is usually information from one or several modules that apply at each level. For example, only discovery metadata applies at the data set level, whereas all types of metadata apply at the collection level. Only specific metadata elements are extracted at the individual geometry component level such as the grid cell level and at that level are included as attributes in the data record. Metadata is hierarchical, so that any metadata element at a higher level applies unless superseded by more detailed metadata at a lower level. That is, metadata elements only need to be described at the highest level since metadata at a lower level overrides metadata inherited from a higher level. Typically it is the acquisition and quality metadata modules that are applied at the lower levels.

The metadata modules are illustrated in Figure 17.

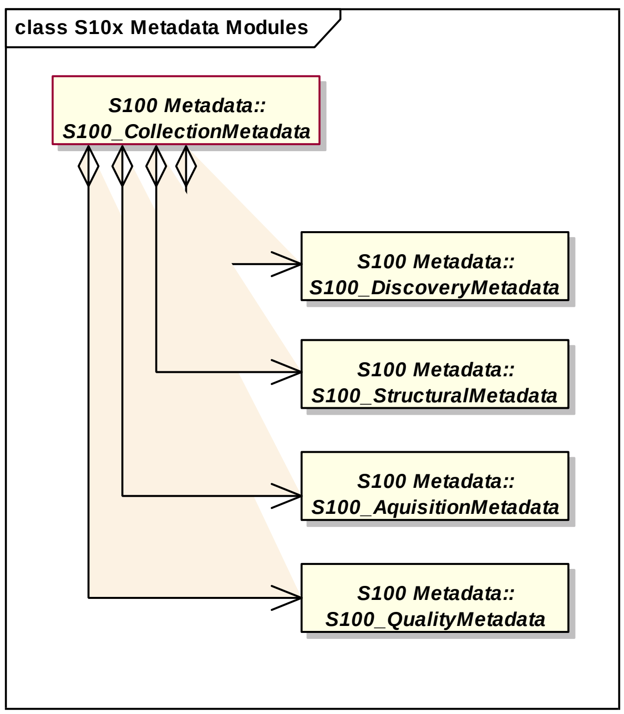

Figure 17- Metadata Modules

The S100_DiscoveryMetadata applies to the whole data set. This is shown by the reference from the S100_DataSet to the S100_DiscoveryMetadata module as shown in Figure 11. ISO 19115 indicates that certain of the metadata fields in the discovery metadata module are mandatory at the data set level. Only a very few metadata fields are actually mandatory, but many more are conditional. If a product specification indicates that some class of metadata is required in that product specification, then other metadata associated with and conditional on that metadata element becomes mandatory.

The metadata modules are linked to the data set or data collection by reference. The attribute "+metadata" contains a character string that may contain a file name for an associated file of metadata or any URL at which the metadata may be located. This approach decouples the metadata from the data allowing different encoding techniques or storage techniques to be used to carry the metadata and the coverage data values. For example, the metadata may be carried in an XML file compliant with ISO 19139 or in a NetCDF4/HDF5 format.

S.10x only describes the S100_Discovery and S100_Structural metadata modules. The S100_Aquisition and the S100_QualityMetadata modules relate only to a specific layer and not integrated across layer boundaries and therefore do not need to be described in detail in this document. Their details are described in S.100.

12.2 Metadata Overview

The metadata packages defined in the metadata component of the S-100 standard are illustrated in Figure 18. This is a "shopping list" of metadata elements and attributes. A particular product specification selects from this list of elements. What the ISO and the S-100 standard does, by providing a list of elements from which to choose, is to ensure that there is a common understanding associated with a particular metadata element when it is provided. Since few metadata elements are mandatory, this is all that is done. It provides a level of interoperability, but it does not provide commonality. To have a common data specification for a particular type of data it is necessary to define a product specification for that data. The product specification provides the commonality. A product specification is needed for each auxiliary layer that identifies the metadata used.

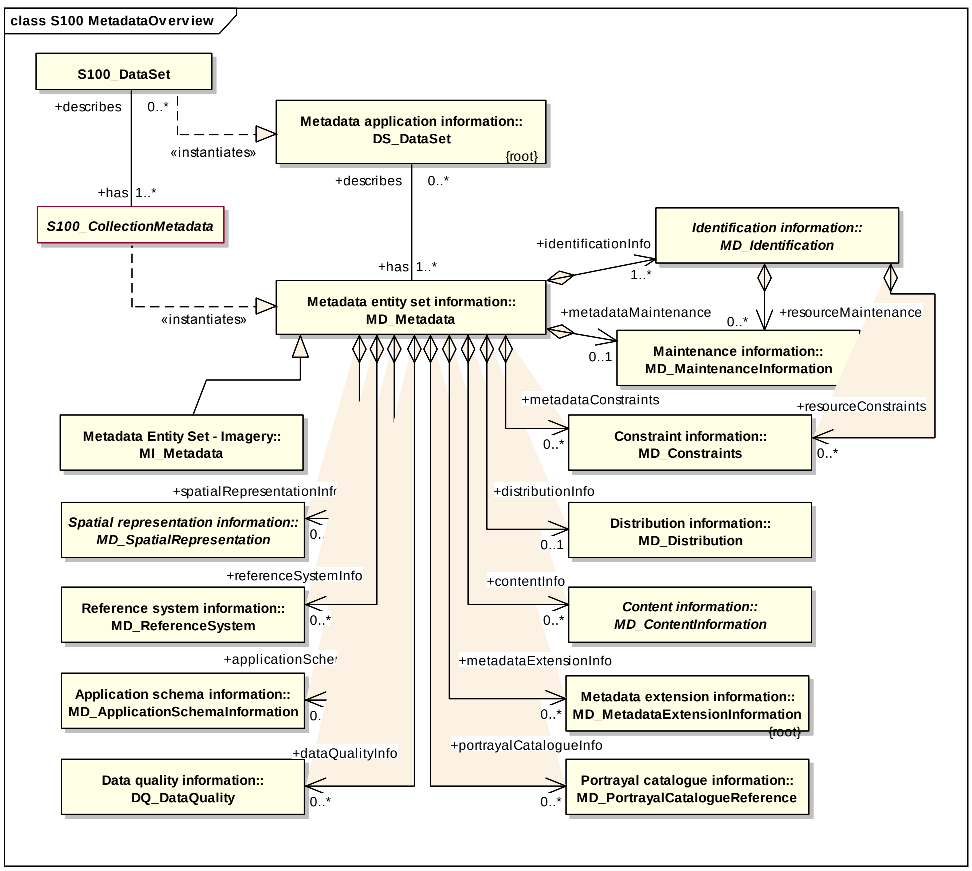

Figure 18- Metadata Overview

The S-100 metadata derives from both the ISO 19115 Metadata standard and the 19115-2 Imagery metadata standard. The MD classes are defined in 19115 and the MI classes in 19115-2. The DQ classes in 19115 are empty classes whose attributes are provided by ISO 19138 Data Quality Measures.

12.3 Discovery Metadata Module

The Discovery Metadata applies to both the data set and a data collection. Elements from the Discovery Metadata Module are derived from the core metadata identified in ISO 19115 plus additional information related to metadata constraints, identification information, descriptive keywords, graphic overview, aggregation information, resource format, resource specific usage and reference system information. This is illustrated in Figure 19.

The description of each metadata field is given in S.100.

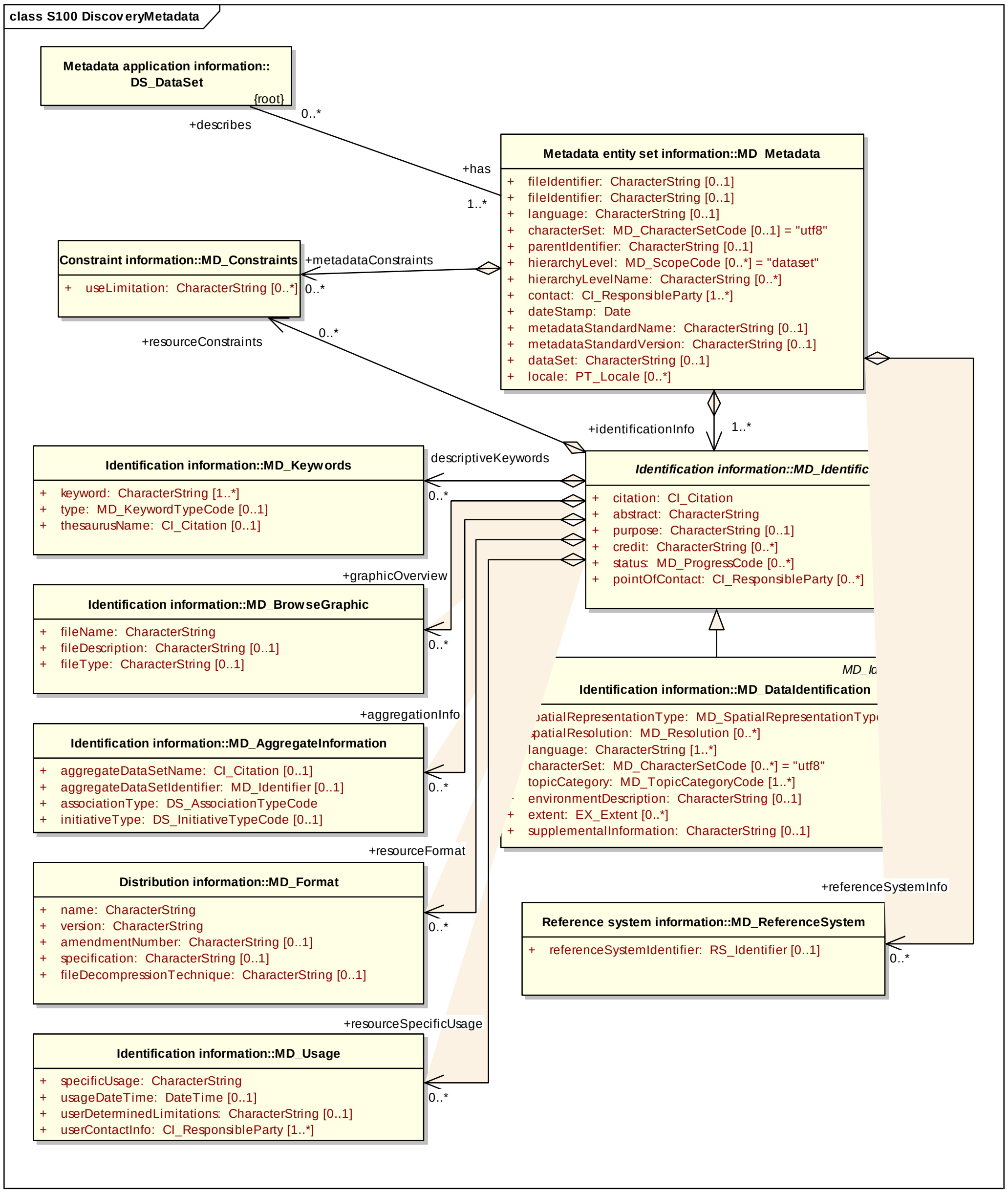

Figure 19- Discovery Metadata Module

The **MD_Metadata** class is the root class for ISO TC211 metadata. It describes the basic characteristics of a data set. Only the ***contact*** and ***dateStamp*** are mandatory attributes. **MD_Identification** is a mandatory component. Within the **MD_Identification** component class only the ***citation*** and ***abstract*** attributes are mandatory. This means that it is possible to have as little as four attributes of mandatory metadata. A product specification for an auxiliary layer establishes which metadata elements are required for that product.

The **MD_Identification** contains information to uniquely identify the data. Identification information includes information about the citation for the resource, an abstract, the purpose, credit, the status and points of contact. The MD_Identification entity is mandatory. It contains mandatory, conditional, and optional elements. The **MD_DataIdentification** contains information.

The **MD_Constraints** information concerning the restrictions placed on data such as legal constraints or security constraints. Legal constraints are important for navigational data.

The **MD_Keywords** contains keywords describing the data set.

The **MD_BrowseGraphic** contains an overview like a thumbnail image of the area over which the data is held. The browse graphic may be held as an external file referenced by name.

The **MD_AggregationInformation** contains information.

The **MD_Format** contains information about the format of the data. This is encapsulation and encoding information carried as metadata, and this metadata element can be changed to represent other formats if the data is transformed from one format to another. Only information at the level of the name and version of the format is carried. Details of the format, such as byte counts, and other low level details are carried as part of the encoding of the data.

The **MD_Usage** contains information related to the aggregation of data within a dataset..

The **MD_ReferenceSystem** contains a description of the spatial and temporal reference system(s) used in a dataset. The reference system may be specified using an identifier through the attribute ***referenceSystemIdentifier*** or through a MD_CRS, which describes projection, ellipsoid and datum parameters. A postal code is an example of an identifier. If coordinate reference system is used it may be taken from a register of geodetic codes and parameters as described in ISO 19127 or a sub-register for S-100 or it may be described explicitly by its geodetic parameters. Reference system information are derived from ISO 19108 Temporal Schema, ISO 19111 Spatial Referencing by Coordinates and ISO 19112 Spatial Referencing by Geographic Identifiers.

12.4 Structure Metadata Module

Structure Metadata provides information describing the content of the data and the distribution information. This is also selected directly from ISO 19115. In addition information about the instance of a tile is provided as attributes in a related class **S10x_Tile**. Providing this tiling information as an attribute of a related class means that there is no requirement to actually extend the ISO metadata standard. The advantage of not extending the ISO standard is that software designed to recognize and process the ISO metadata will work directly on the S-100 metadata.

Figure 20 illustrates the structural metadata including the tile reference.

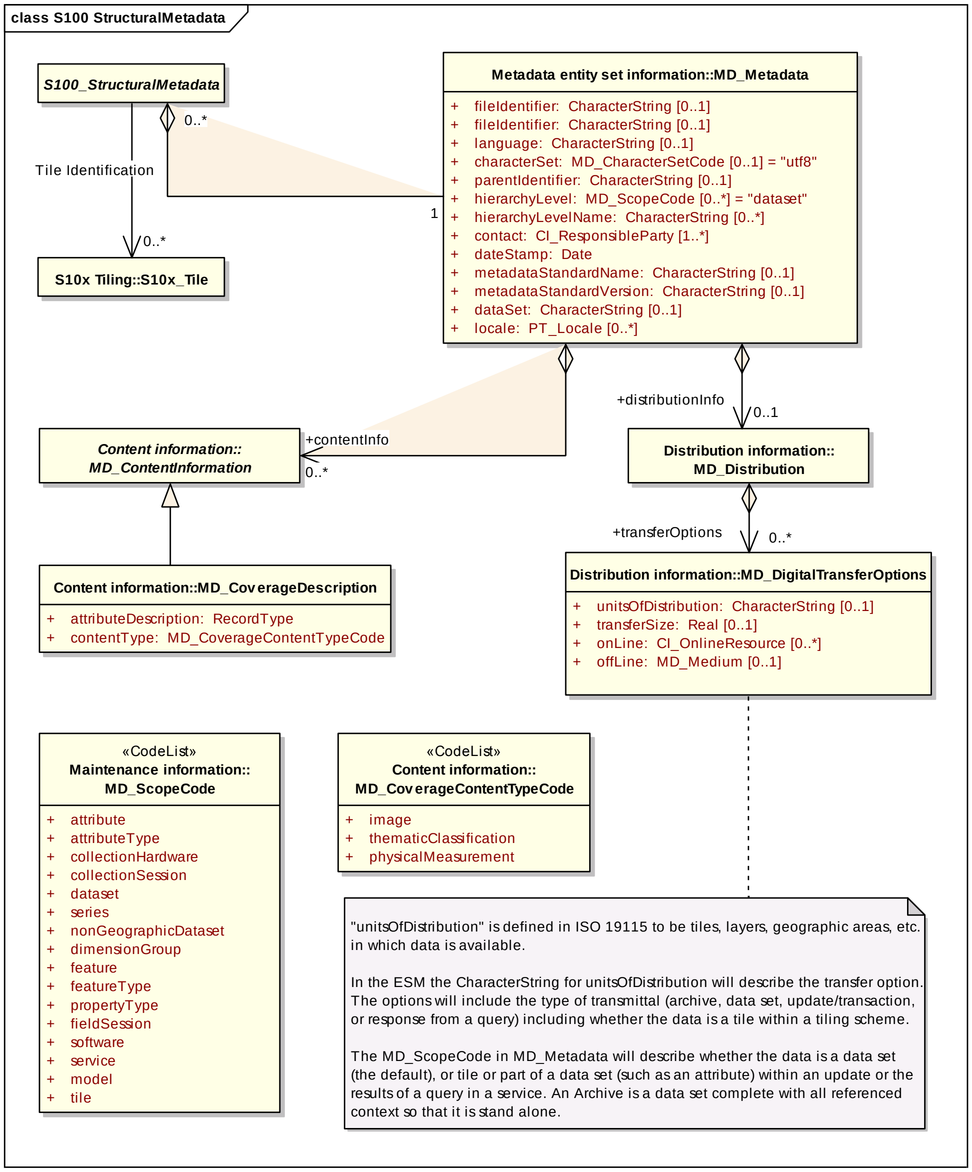

Figure 20 - S-100 Structure Metadata Module

12.4.1 AcquisitionMetadataModule

Acquisition Metadata provides information related to the sensor or other source of the auxiliary layer data. For coverage data this module primarily makes use of the classes defined in ISO 19115-2 Imagery Metadata.

The Acquisition Metadata Module provides information on acquisition requirements, acquisition plans, acquisition operations and objectives, and information about the instrument and its platform used to acquire the data.

There is a tremendous amount of information that can be provided about particular sensors. A product specification needs to choose carefully the metadata elements relevant to that type of data.

Figure 21 presents an overview of the acquisition metadata classes.

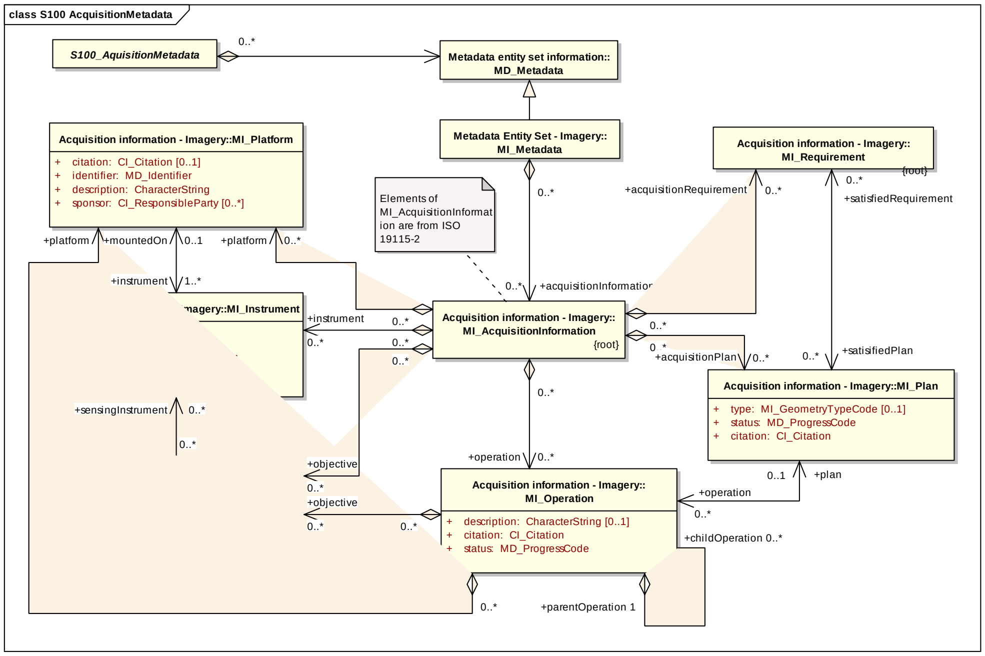

Figure 21 - Acquisition Metadata Module

12.5 QualityMetadataModule

The Quality Metadata Module provides information about the quality of a data set, data collection or specific groups or individual data values. Quality metadata provides information about lineage, source, processing steps and source steps and quality measures. The Quality Metadata Module is a direct subset of the quality metadata provided in ISO 19115. The attributes for quality have been provided by ISO 19138.

Figure 22 provides an overview of the ISO quality metadata. Since there are too many attributes the detailed attributes are not shown here. These are given in the S-100 Quality metadata component.

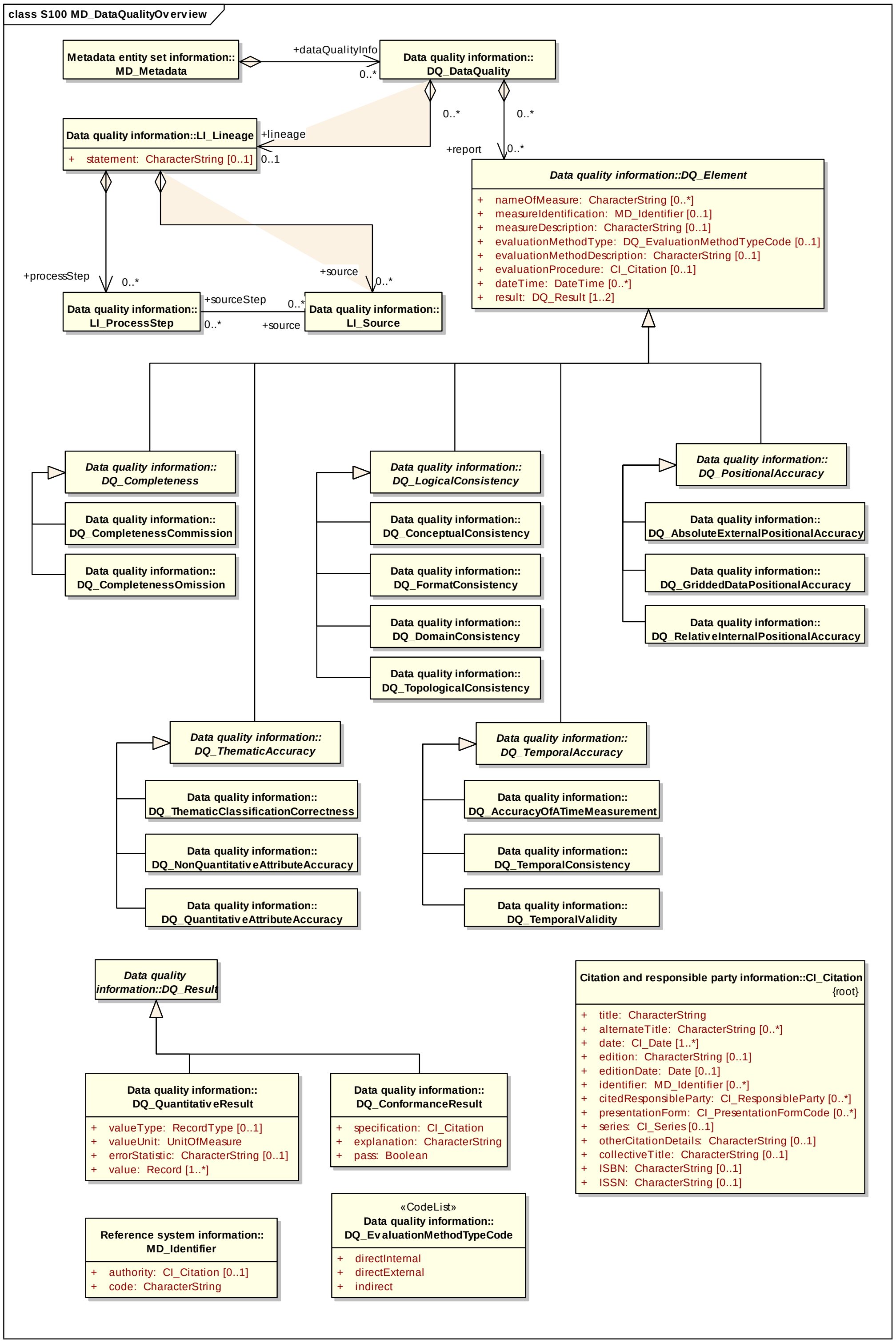

Figure 22 - Quality Metadata Overview

The S100-Quality Measure class structure is derived from ISO 19115 Geographic Information Metadata. The attributes described in the S-100 Quality classes each correspond to independent quality measures. A full description of these measures are contained in ISO 19138 geographic Information Data Quality Measures.

All of the S100_Quality Measures are intentionally optional so that different measures may be used for different types of data. Where multiple attributes describe the same measure in different ways, either only one measure should be used or the measures must be described in a consistent manner.

For bathymetry data the quality measure may be carried as a value associated with each value in the grid value matrix by defining an additional value entity within the value record.

13 Encoding

13.1 Encoding Principals

All data must be encapsulated (delimited) and encoded in order to be exchanged. Encoding is the responsibility of the encoding format. Several encoding formats are available for auxiliary data layers. It is not necessary that the same encoding be used for each layer.

The basic encoding principle used in the ISO suite of standards is the separation of the "carrier" from the "content". The ISO standards define data in terms of an encoding neutral UML model. Elements from this model are then used to create and Application Schema as part of a product specification. The application schema may be encoded using various different encoding formats.

The only difference between the information carried using different encoding formats is information related to the mechanisms used in that particular encoding technique. If some technique requires delimiters and another requires byte-counts then the encoding format is responsible for any encoding level attributes that identify encoding elements.

Different encoding techniques have different levels of efficiency and others introduce limitations so that certain types of data can not be handled, or if it is handled it is handled in an inefficient manner. By describing the content in an abstract manner and separating the carrier from the content, these limitations are isolated from the data itself.

There are many different encoding formats from which to choose. Vector data may be carried as ISO 19136 compliant GML, of ISO 8211 compliant records or even in an industry proprietary format. If an industry proprietary format is used then the limitations that may exist in that format must be taken into account. There is a very wide choice of possible encoding formats for coverage data. A general framework for coverage data is described in ISO 19129. Annex A discusses some of the encoding choices.

Annex A
Coverage Encoding

(Informative)

A.1 Multiple Encoding

The subsections of this annex present several choices for encoding coverage data that may be used in auxiliary layers. This is not an exhaustive list, but is illustrative of the diversity that may be encountered.

A.2 XML plus JPEG Encoding

GML + ISO 19139 + JPEG 2000 (GML JP2)

Coverage data sets such as bathymetric data is primarily a set of values such as depth values together with a associated metadata. The data volume of the metadata is small compared to the data volume of the data values. XML is an easily parsable format for delimiting and identifying structured information. However it introduces significant overhead and is not efficient at handling large arrays of numerical data. Therefore the approach to using XML encoding is to combine it with another encoding for the array of data values.

An XML schema, or Data Type Declaration (DTD), is used in XML to establish the syntactic structure of the format and to identify the semantic meaning of the XML tags. The XML schema for the metadata has already been developed in ISO 19139 and all that is necessary is to select the appropriate portions of that schema to represent the metadata elements used for a set of coverage data in an auxiliary layer such as bathymetry.

JPEG:2000 (ISO 15444) provides a capability to efficiently carry an array of data elements in a lossless manner. The Geography Markup Language GML (ISO 19136) has registered a code in the JPEG header that allows a link to be established to a GML (XML) file.

Note that GML JP2 is an imbedded encoding. That is, the XML data stream is embedded within the JPEG data file. This has the advantage of strongly binding the ISO 19139 metadata to the JPEG data file. The disadvantage is that the file is hard to edit and the normally human readable XML data is not directly available. JPEG can be used the other way around where an application schema is developed in XML which contains a HREF reference to a separate JPEG file. In this configuration the XML can be easily edited and examined, but there is a chance that the two files (the XML/GML and the JPEG) may be separated.

A.3 NetCDF Encoding

NetCDF (Network Common Data Form) is a machine-independent binary data format commonly used in the scientific data community especially for climatology applications such as weather forecasting and in some imagery applications. Because it is a binary format it is an efficient format.

NetCDF includes a data header which describes structure of the file together with data arrays and other information elements. In this way it is similar in structure to ISO 8211. There are some limitations built into the structure to make it simpler.

The new version of NetCDF (version 4) has been implemented on top of the HDF5 data format. The Hierarchical Data Format (HDF) is a multi-object file format for scientific data. It includes several data types, such as, multidimensional arrays, raster images, and tables. New data types can be easily added. The latest version of HDF (version 5) is a significant improvement over previous versions and its object-oriented structure is very flexible.

A.4 NITIF Encoding

National Imagery Transmission Format standard (NITF) is a widely implemented exchange format based on the ISO Basic Imagery Interchange Format BIFF (ISO 12087-5). NITIF is widely used in the defence community. NITF data contains information about the image, the image itself, and optional overlay graphics. An image may have sub-images, as well as other information related to the image such as overlaying text or graphics. Spatial Data Extensions extend NITF functionality in a compatible manner.

Recently NITIF has incorporated the JPEG-2000 (ISO 15444-1 standard. An NITIF encoding is likely to be one of the encodings of the military Elevation Surface Model standard being developed by DGIWG.

A.5 GeoTIFF Encoding

GeoTIFF is an extension of the popular Tagged Image File Format to carry some geospatial metadata together with the image. The elevation values can be carried as an array much like an image. The GeoTIFF specification indicates that it will in the future be extended to carry additional information to support elevation data. The additional metadata identified in this standard can be carried as an optional associated XML file. The advantage of GeoTIFF encoding is that TIFF decoders are very widespread, and therefore coverage data such as bathymetry data expressed in GeoTIFF could easily be viewed, although the other metadata is discarded. This might serve a part of the market for bathymetry data.