10 Data Product Format (Encoding)

10.1 Introduction

The S-102 data set must be encoded using the Hierarchical Data Format standard, Version 5 (HDF5).

Format Name: HDF5

Version: 1.8

Character Set: UTF-8

Specification: https://www.hdfgroup.org/

The key idea behind the S-102 product structure is that each coverage is a feature. Each of these features is co-located with the others. Therefore, they share the same spatial metadata and each is required to correctly interpret the others.
For the use of HDF5, the following key concepts (S-100 Part 10c, clause 10c-5.1) are important:

⟩ *File* - a contiguous string of bytes in a computer store (memory, disk, etc.), and the bytes represent zero or more objects of the model;

⟩ *Group* - a collection of objects (including groups);

⟩ *Dataset* - a multidimensional array of data elements with attributes and other metadata;

⟩ *Dataspace* - a description of the dimensions of a multidimensional array;

⟩ *Datatype* - a description of a specific class of data element including its storage layout as a pattern of bits;

⟩ *Attribute* - a named data value associated with a group, dataset, or named datatype;

⟩ *Property List* - a collection of parameters (some permanent and some transient).

In addition, datasets may be a compound (a single record consisting of an array of simple value types) and have multiple dimensions.

10.2 Product Structure

The structure of the data product follows the form given in S-100 Part 10C – HDF5 Data Model and File Format. The general structure, which was designed for several S-100 products is given in Figure 10-1.
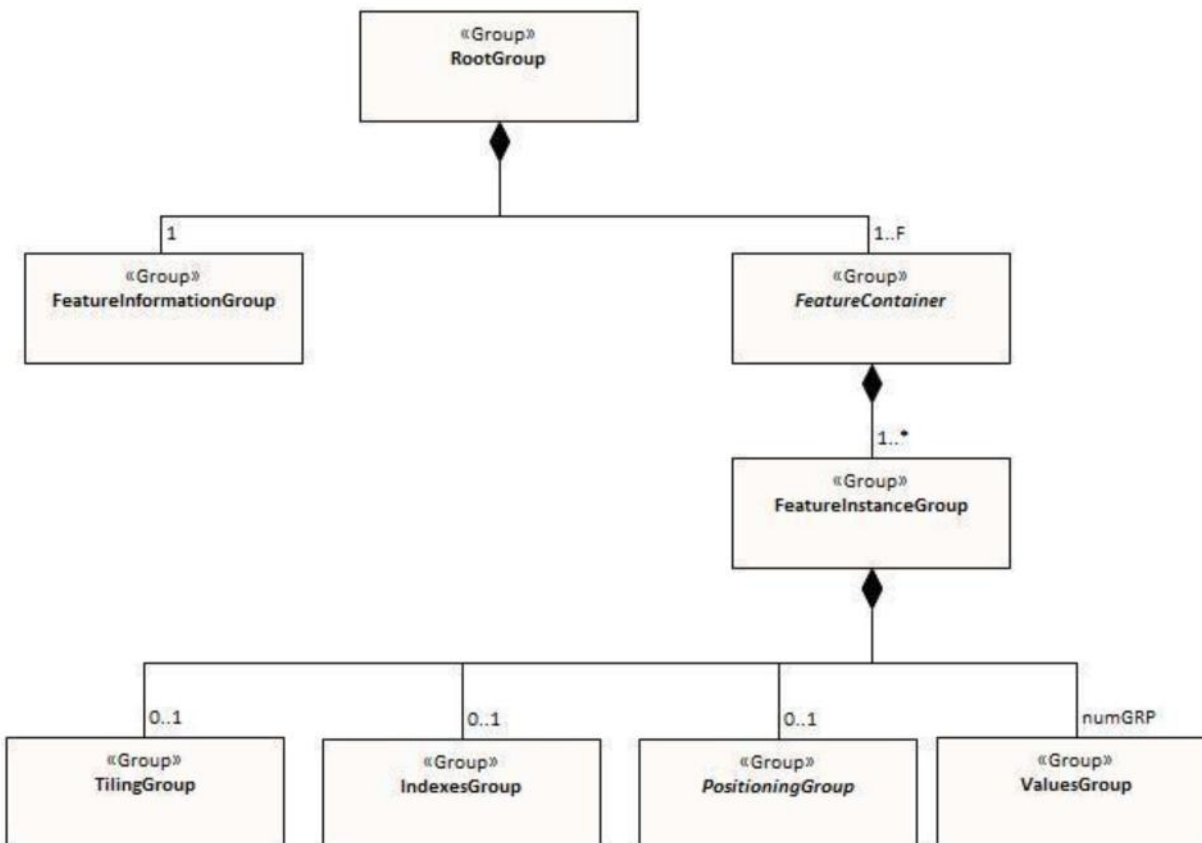

Figure 10-1 - Outline of the generic data file structure

Figure 10-1 shows the four levels defined within the HDF encoding. Below is a further definition of these levels.

Level 1: At the top level lies the Root Group, and it contains the Root Metadata and two subsidiary groups. The Root Metadata applies to all S-100 type products.

Level 2: The next Level contains the Feature Information Group and the Feature Container Group. The Feature Information Group contains the feature names (**BathymetryCoverage**, **TrackingListCoverage**) and the feature attribute codes. The Feature Container Group contains the Feature Metadata and one or more Feature Instance Groups.

Level 3: This contains one or more Feature Instances. A feature instance is a bathymetric gridded data for a single region, or a tracking list of nodal overrides.

Level 4: This contains the actual data for each feature.

In Table 10-1 below, levels refer to HDF5 structuring (see S-100 Part 10c, Figure 10c-9). Naming in each box below the header line is as follows: Generic name; S-100 or S-102 name, or [] if none; and (*HDF5 type*) group, attribute or attribute list, or dataset.

Table 10-1 - Overview of S-102 Data Product

LEVEL 1 CONTENT	LEVEL 2 CONTENT	LEVEL 3 CONTENT	LEVEL 4 CONTENT
General Metadata (metadata) *(h5_attribute)*			
Feature Codes Group_F *(h5_group)*	Feature Name BathymetryCoverage *(h5_dataset)*		
	Feature Name TrackingListCoverage *(h5_dataset)*		
	Feature Codes featureCode *(h5_dataset)*		
Feature Type BathymetryCoverage *(h5_group)*	Type Metadata (metadata) *(h5_attribute)*		
	Feature Instance BathymetryCoverage.01 *(h5_group)*	Instance Metadata (metadata) *(h5_attribute)*	
		First data group Group.001 *(h5_group)*	Group Metadata (metadata) *(h5_attribute)*
	X and Y Axis Names axisNames *(h5_dataset)*		Bathymetric Data Array values *(h5_dataset)*
...			
Feature Type TrackingListCoverage	Type Metadata (metadata) *(h5_attribute)*		
	Feature Instance TrackingListCoverage.01 *(h5_group)*	Instance Metadata (metadata) *(h5_attribute)*	
		First data group Group.001 *(h5_group)*	Tracking_List values *(h5_dataset)*

Table 10-2 - Hierarchy of S-102 Data Product
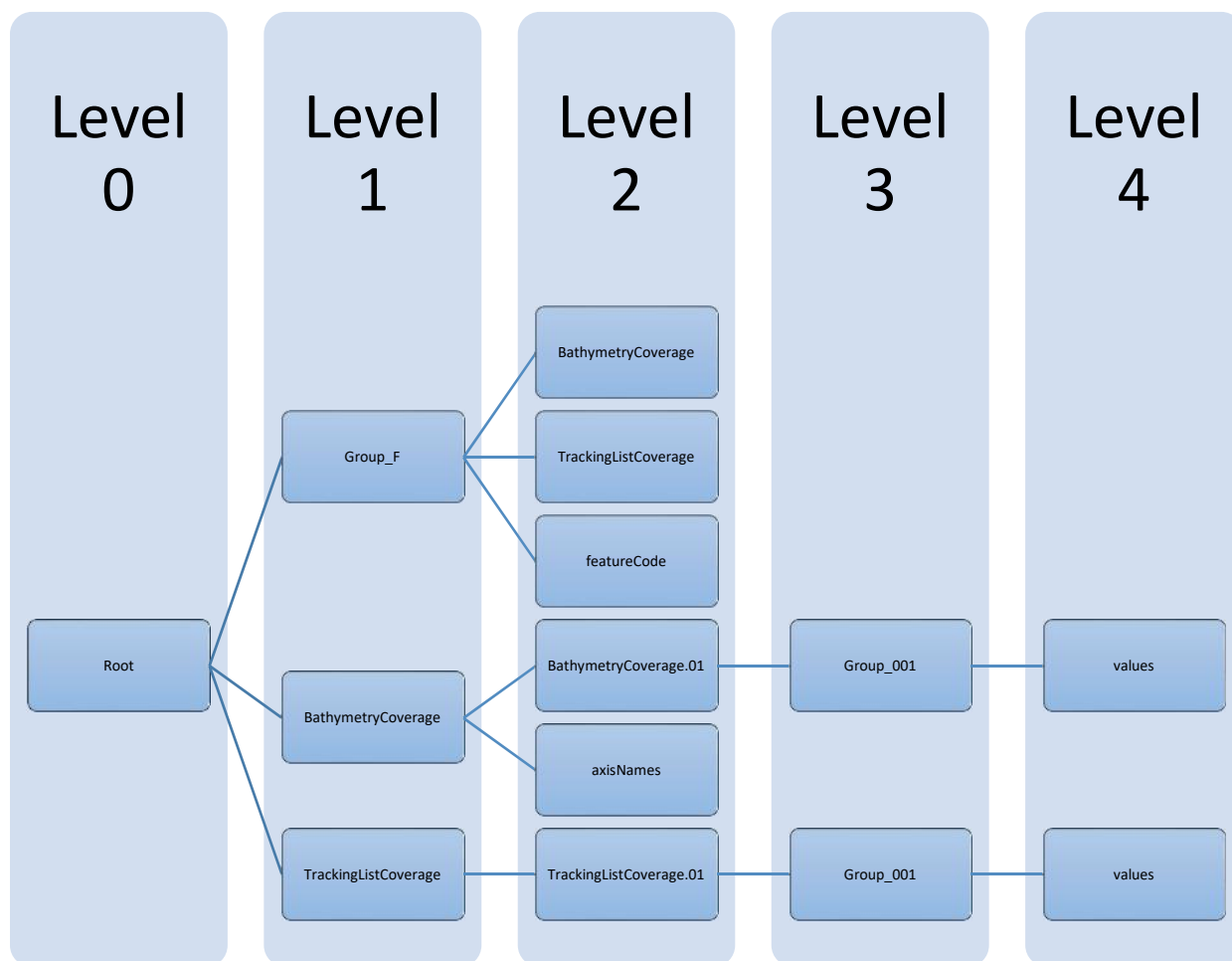
10.2.1 Root Group Attributes

Table 10-3 – Root group attributes

No	Name	Camel Case	Mult	Data Type	Remarks
1	Product specification number and version	productSpecification	1	String	Value: INT.IHO.S-102
2	Time of data product issue	issueTime	0..1	String (Time Format)	
3	Issue date	issueDate	1	String (Time Format)	
4	Horizontal datum	horizontalDatumReference	1	String	Value: EPSG
5	Horizontal datum number	horizontalDatumValue	1	Integer	Example: 4326 (for WGS84)
6	Epoch of realization	epoch	0..1	String	
7a	Bounding box	westBoundLongitude	1	Float	Bounding box intersects the pixel center of each corner cell in a pixel-is-point manner. Half a cell all around the grid is not included in the extents.
7b		eastBoundLongitude	1	Float	
7c		southBoundLatitude	1	Float	
7d		northBoundLatitude	1	Float	
9	Metadata	Metadata	1	String	Name of metadata file
10	Vertical datum reference	verticalDatum	0..1	Enumeration	

The bounding box should use the horizontal units specified by the coordinate system defined by the horizontal datum. The boundbox should intersect the center of the bottom left and upper right cells.

10.2.2 Group_F

No attributes.

This group specifies the S-100 features to which the data applies, and consists of three components:

featureCode – a dataset with the name(s) of the S-100 feature(s) contained in the data product. For S-102, the dataset has two elements. These elements are **BathymetryCoverage** and **TrackingListCoverage**. TrackingListCoverage may be omitted if there are no tracking list points. In such a case, the TrackingListCoverage table in Group_F must also be omitted as well as the entire root TrackingListCoverage Group and child components.

BathymetryCoverage – One of the features described in the featureCode table. This feature contains the standard definition of the feature class.

TrackingListCoverage – One of the features described in the featureCode table. This feature contains the standard definition of the feature class. If no TrackingList points exist, this table may be omitted.

10.2.3 BathymetryCoverage and TrackingListCoverage Tables (in Group_F)

These tables specify feature information. One for each feature in the featureCode array.

There is a single column in the BathymetryCoverage and TrackingListCoverage. This column is an HDF5 compound type with multiple values. Since it is a single column, it is one dimensional despite its two dimensional appearance.

The elements in the compound type are as follows:

code: camel case code of attribute as in feature catalogue
name: long name as in feature catalogue
uom.name: units (uom.name from S-100 feature catalogue)
fillValue: fill value (integer or float value, string representation)
datatype: HDF5 data type, as returned by H5Tget_class() function
lower: lower bound on value of attribute
upper: upper bound on attribute value
closure: type of closure

Table 10-4 - Sample contents of the one-dimensional BathymetryCoverage array

Name	Explanation	S-100 Attribute 1	S-100 Attribute 2
code	Camel Case Name	elevation	uncertainty
name	plain text	elevation	uncertainty
uom.name	Units of Measurement)	metres	metres
fillValue	Denotes missing data	1000000	1000000
datatype	HDF5 datatype	H5T_NATIVE_FLOAT	H5T_NATIVE_FLOAT
lower	Lower bound on attribute	-12000	-12000
upper	Upper bound on attribute	12000	12000
closure	Open or Closed data interval. See S100_IntervalType in Part 1.	closedInterval	closedInterval

As per section S-100 10c-9.5, "All the numeric values in the feature description dataset are string representations of numeric values; for example, "-9999.0" not the float value -9999.0."

While the sample contents are shown in the two attributes columns, these are actually rows in the BathymetryCoverage table. They are also each a single HDF5 compound type and represent a single HDF5 element in the table.

All cells shall be HDF5 variable length strings. The minimum and maximum values are stored in lower and upper columns. Variable length strings allow future proofing the format in the event editing is allowed or correcting these values is required.

TrackingListCoverage table has 5 attribues X, Y, originalValue, trackCode and listSeries.

Table 10-5 - Sample contents of the one-dimensional TrackingListCoverage array

Name	S-100 Attribute 1	S-100 Attribute 2	S-100 Attribute 3	S-100 Attribute 4	S-100 Attribute 5
code	X	Y	originalValue	trackCode	listSeries
name	X	Y	Original Value	Track Code	List Series
uom.name	N/A	N/A	metres	N/A	N/A
fillValue	0	0	1000000	0	0
datatype	H5T_NATIVE_INT	H5T_NATIVE_INT	H5T_NATIVE_FLOAT	H5T_NATIVE_INT	H5T_NATIVE_INT
lower	0.0	0.0	-12000	0.0	0.0
upper	1000000.0	1000000.0	12000	1000000.0	1000000.0
closure	closedInterval	closedInterval	closedInterval	closedInterval	closedInterval

As with the BathymetryCoverage table, the attribute columns are actually rows in the TrackingListCoverage table. The TrackingListCoverage table has five rows and one HDF5 compound type column.

10.2.3 Root BathymetryCoverage and TrackingListCoverage Groups

Table 10-6 – Attributes of BathymetryCoverage feature container group

No	Name	Camel Case	Mult	Data Type	Remarks
1	Data organization index	dataCodingFormat	1	Enumeration	Value: 2
2	Dimension	dimension	1	Integer	Value: 2
3	Common point rule	commonPointRule	1	Enumeration	Value: 1 (average) or other values from S100 Table 10c-19.
4	Horizontal position uncertainty	horizontalPositionUncertainty	1	Float	Value: 0.0 (if not available)
5	Vertical position uncertainty	verticalUncertainty	1	Float	Value: 0.0 (if not available)
6	Number of feature instances	numInstances	1	Integer	Value: 1
7a	Sequencing rule	sequencingRule.type	1	Enumeration	Value: 1
7b		sequencingRule.scanDirection	1	String	Value: "Longitude, Latitude" (without quotes)
8	Interpolation type	interpolationType	1	Enumeration	Value: 1 or other value from S100 Table 10c-21

Table 10-7 – Attributes of TrackingListCoverage feature container group

No	Name	Camel Case	Mult	Data Type	Remarks
1	Data organization index	dataCodingFormat	1	Enumeration	Value: 1
2	Dimension	dimension	1	Integer	Value: 2
3	Common point rule	commonPointRule	1	Enumeration	Value: 1 (average) or other values from S100 Table 10c-19.
4	Horizontal position uncertainty	horizontalPositionUncertainty	1	Float	Value: 0.0 (if not available)
5	Vertical position uncertainty	verticalUncertainty	1	Float	Value: 0.0 (if not available)
6	Number of feature instances	numInstances	1	Integer	Value: 1

10.2.4 BathymetryCoverage.01 Group

Table 10-8 Attributes of BathymetryCoverage feature instance group

No	Name	Camel Case	Mult	Data Type	Remarks
1a	Bounding box	westBoundLongitude	0..1	Float	Optional. If present, must be identical to root Bounding box attribute.
1b		eastBoundLongitude	0..1	Float	
1c		southBoundLatitude	0..1	Float	
1d		northBoundLatitude	0..1	Float	
2	Number of groups	numGRP	1	Integer	Value: 1
3	Longitude of grid origin	gridOriginLongitude	1	Float	Longitude of grid origin.
4	Latitude of grid origin	gridOriginLatitude	1	Float	Latitude of grid origin.
5	Grid spacing, longitude	gridSpacingLongitudinal	1	Float	Cell size in x dimension.
6	Grid spacing, latitude	gridSpacingLatitudinal	1	Float	Cell size in y dimension.
7	Number of points, longitude	numPointsLongitudinal	1	Integer	Number of points in x dimension.
8	Number of points, latitude	numPointsLatitudinal	1	Integer	Number of points in y dimension.
9	Start sequence	startSequence	1	String	Value: "0,0" (without quotes)

The gridOriginLongitude, gridOriginLatitude, gridSpacingLongitudinal and gridSpacingLatitudinal attributes should be in the same geographic units as the bounding box. Note that this deviates from S100 where it indicates that this should be in Arc Degrees. To avoid loss of precision, resolution (grid) information must remain in the same geographic units as the horizontal datum. This has the effect that gridOriginaLongitude and gridOriginLatitude are identical to westBoundLongitude and southBoundLatitude.

The gridOriginLongitude and gridOriginLatitude are the cell center of the cell physically located at the bottom left of the raster. This cell is the first cell in the values table.

The grid spacing can be calculated as follows:

-) gridSpacingLongitude = (eastBoundLongitude – westBoundLongitude) / (numPointsLongitude – 1);
-) gridSpacingLatitude = (northBoundLatitude – southBoundLongitude) / (numPointsLatitude – 1);

numPointsLongitude and numPointsLatitude must contain the number of cells in the x and y dimensions of the values table.

10.2.5 TrackingListCoverage.01 Group

Table 10-9 Attributes of TrackingListCoverage feature instance group

No	Name	Camel Case	Mult	Data Type	Remarks
1	Number of groups	numGRP	1	Integer	Value: 1
2	Number of fixed stations	numberOfStations	1	Integer	Value: 0

The attribute numberOfStations doesn't apply for tracking list, but is required by S100 when dataCodingFormat is 1 as is the case for tracking list. Therefore a value of 0 should be used.

10.2.6 Group.001 group

This group contains no attributes.

10.2.7 Values Groups (Group.001)

These groups each contain the compound data arrays containing bathymetric gridded data or tracking list data. These groups contain no attributes. These groups are explained below.

For bathymetric gridded data, the dataset includes a two-dimensional array containing both the elevation and uncertainty data. These dimensions are defined by the dataspace of the values table. If uncertainty data is not used, it must be filled with the fillValue specified in the Group_F feature information dataset.

For tracking list data, the dataset includes a single dimension array containing the position (X, Y) of each override, defined as row/col of the bathymetric grid, the original value, the type of override and the index into the metadata that defines the override. The number of overrides in the array is defined by the originator and this dataset could be empty if no overrides were required.

10.2.8 Data Arrays

Within the **BathymetryCoverage**, the elevation and uncertainty values (elevation and uncertainty) are stored in two dimensional arrays named *values,* with a prescribed number of columns (*numCOL*) and rows (*numROW*). This grid is defined as a regular grid (*dataCodingFormat* = 2), therefore; the elevation and uncertainty values will be for each discrete point in the grid. The data array *values* is two-dimensional.

Within the **TrackingListCoverage**, entries are stored in a single dimensional array named *values.* The number of rows in this array is dynamic as entries into this dataset are optional as not all data sources require modifications to the **BathymetryCoverage**. This grid is defined as a point set (*dataCodingFormat* = 1), if it exists.

10.2.9 Summary of Generalized Dimensions
To summarize, there are data Groups contains one of two types of feature datasets. The first contains the elevation and uncertainty data, which are stored in two-dimensional arrays of size *numROW* by *numCOL*. The second is a single dimension array containing information on overrides that were performed on the data in the dataset.

10.2.10 Mandatory Naming Conventions

The following group and attribute names are mandatory in S-100: Group_F, featureCode, and (for S-102) **BathymetryCoverage**, **BathymetryCoverage.01**, **TrackingListCoverage.01**, and Group.nnn.