



BUNDESAMT FÜR  
SEESCHIFFFAHRT  
UND  
HYDROGRAPHIE

# Clarify and enhance HDF5 encoding instructions for S-102



S-102PT 27 - 28 October 2020

# Data structure in S-102 ed. 2.0.0

- Dataset file naming
  - *Discussion:* Producer code have only 2 character in the Producer Code Register but according to the specification 4 characters should be given.
  - *Recommendation:* Changing the dataset naming convention to 2 character for the producer code.

[S-102 Chapter 11.2.3, Page 34]

## 11.2.3 Dataset file naming

Dataset naming must follow a standard pattern to give implementers greater predictability of incoming datasets. S-102 dataset naming conventions must follow these rules.

### 102PPPPØØØØØØØØØØ.H5

- 102 - the first 3 characters identify the dataset as an S-102 dataset (mandatory).
- PPPP - the fourth to seventh characters is the producer code according to the IHO Geospatial Information Registry, Producer Code Register (mandatory for S-102).
- ØØØØØØØØØØØØØØ - the eighth to the maximum nineteenth characters are optional and may be used in any way by the producer to provide the unique file name. The following characters are allowed in the dataset name: A to Z, 0 to 9 and the special character \_ (underscore).
- H5 - denotes and HDF5 file.

# Data structure in S-102 ed. 2.0.0

- Attribute: commonPointRule
  - *Discussion:* The attribute is an enumeration. S-100 specify the data type enumeration as 1 or 2 byte unsigned integer. In Annex B of S-102 it is an 4 byte signed integer.
  - *Recommendation:* Changing the datatype in Annex B to 1 byte unsigned integer.

```
GROUP "BathymetryCoverage" {
  ATTRIBUTE "commonPointRule" {
    DATATYPE  H5T_STD_I32LE
    DATASPACE SCALAR
  }
}
```

[S-102 Chapter Annex B-1,  
Page 68]

[S-100 Part 10c, Chapter 7, Page 8]

enumeration	Enumeration	Numeric codes must be 1 or 2-byte unsigned integers, range [1, 2 <sup>8</sup> - 1] or [1, 2 <sup>16</sup> - 1]
-------------	-------------	--

[S-100 Part 10c, Chapter 9.6, Page 22]

commonPointRule	1	Enumeration	The procedure used for evaluating the coverage at a position that falls on the boundary or in an area of overlap between geometric objects Values from CV_CommonPointRule (Table 10c-19)
-----------------	---	-------------	---

# Data structure in S-102 ed. 2.0.0

- Attribute: verticalDatum (vertical coordinate reference system)
  - Discussion:* Although there is no feature storable in S-102 which need a vertical datum, the attribute is specified as mandatory. According to S-100 it is a conditional attribute but the condition is not fulfilled.
  - Recommendation:* Removing the attribute in S-102 specification.

multiplicity 

[S-102 Chapter 12.7, Page 55]

verticalDatum	Vertical  Datum of the entire dataset	1	S100_VerticalAndSoundingDatum	This optional S-100 attribute is mandatory in S-102
---------------	---------------------------------------	---	-------------------------------	---

[S-100 Part 10c, Chapter 9.4, Page 16]

verticalDatum	0..1	Enumeration	See S100_VerticalAndSoundingDatum Conditional, if and only if depthTypeIndex=3
---------------	------	-------------	---

- Attribute: soundingDatum in Annex B-1 (HDF5 encoding structure)
  - Discussion:* Missing attribute in Annex B of S-102 product specification.
  - Recommendation:* Adding the attribute in Annex B in the root group.

## Data structure in S-102 ed. 2.0.0

- Attribute: depthTypeIndex

- Discussion:* S-100 suggests an additional attribute for the vertical reference: depthTypeIndex.
  - Recommendation:* Adding the attribute with the value of 4 in the root group of the S-102 specification.

[S-100 Part 10c, Chapter 9.4, Page 17]

depthTypeIndex	1	Enumeration	1: Layer average 2: Sea surface 3: Vertical datum (see verticalDatum) 4: Sea bottom
----------------	---	-------------	--

- Attribute: gridOriginLatitude and gridOriginLongitude

- Discussion:* On using a projected CRS for the dataset, it makes no sense to define the grid origin with geographical coordinates.
  - Recommendation:* Changing the multiplicity of the attribute in S-100 specification and add the compound dataset cellGeometry from S-100 Part 10c table 10c-11 on page 25.

[S-100 Part 10c, Chapter 9.7, Page 26]

gridOriginLongitude	1	Float	The longitude of the grid origin. Unit: Arc Degrees
gridOriginLatitude	1	Float	The longitude of the grid origin. Arc Degrees

# Data structure in S-102 ed. 2.0.0

- Attribute: origin
  - *Discussion:* This is a redundant information to the previous discussed attribute. The UML class of the attribute is `DirectPosition` which should be implemented as `String`. But there is no encoding format defined.
  - *Recommendation:* Removing the attribute in S-102 specification.

[S-102 Chapter 4.2.1.1.1, Page 13]

## 4.2.1.1.1.8 origin

The attribute *origin* has the value class *DirectPosition* which is a position that shall locate the origin of the rectified grid in the coordinate reference system. This attribute is required. There is no default.

```
ATTRIBUTE "origin" {  
    DATATYPE H5T_STRING {  
        STRSIZE 64;  
        STRPAD H5T_STR_NULLTERM;  
        CSET H5T_CSET_ASCII;  
        CTYPE H5T_C_S1;  
    }  
}
```

[S-102 Chapter Annex B-1, Page 70]

## Data structure in S-102 ed. 2.0.0

- Attribute: `scanDirection`
  - *Discussion*: The specification needs a clarification on using a projected coordinate reference system for the dataset.
  - *Recommendation*: Adding information on using a projected CRS for the attribute in S-102 specification. For example:

### 4.2.1.1.7 `scanDirection`

The attribute `scanDirection` has the value class `Sequence<CharacterString>` a list of axis names that indicates the order in which grid points shall be mapped to position within the sequence of records of feature attribute values. The scan direction for layers in S-102 with unprojected coordinate reference system (CRS) use "longitude,latitude" and for projected CRS "X,Y" according to the `axisNames` entries. Reverse scan direction along an axis is indicated by prefixing a '-' sign to the axis name.

[S-102 Chapter 4.2.1.1.7, Page 15]

### 4.2.1.1.7 `scanDirection`

The attribute `scanDirection` has the value class `Sequence<CharacterString>` a list of axis names that indicates the order in which grid points shall be mapped to position within the sequence of records of feature attribute values. The scan direction for all layers in S-102 is "Longitude" and "Latitude" or west to east, then south to north.

# Data structure in S-102 ed. 2.0.0

- Attribute: chunking
  - *Discussion:* The HDF5 format already has a mechanism for chunking. It is not necessary to save the chunk size in an extra attribute.
  - *Recommendation:* Removing the attribute in S-102 specification.

[S-102 Chapter Annex B-1, Page 72]

```

DATASPACE SIMPLE { ( 2, 1 ) / ( 2, 1 ) }
ATTRIBUTE "chunking" {
  DATATYPE H5T_STD_I16LE
  DATASPACE SCALAR
}
  
```

[screenshot of HDFview]

No. of Dimension(s):	2	
Dimension Size(s):	3728 x 2218	
Max Dimension Size(s):	3728 x 2218	
Data Type:	Compound	
Name	Type	Array Size
depth	32-bit floating-point	1
uncertainty	32-bit floating-point	1
Storage Layout: CHUNKED: 256 X 256		





## Data structure in S-102 ed. 2.0.0

- Group: Group.001
  - *Discussion:* According to S-100 the first data group in HDF5 should be named “Group\_001”, but in S-102 the specified name is “Group.001”.
  - *Recommendation:* Changing the group name in S-102 specification.

First data group Group.001 ( <i>h5_group</i> )	[S-102 Chapter 10.2, Page 30]
--	-------------------------------

[S-100 Part 10c Chapter 8, Page 10]

Numeric suffixes preceded by the underscore character (that is, the suffix ‘NNN’) may be added to distinguish groups which would otherwise have the same names (for example, data groups at different time points).
---

Group_nnn	Data for one member of a series; for example, at a time point in a time series, or for different stations. “n” means any digit from 0 to 9. Numbering must use 3 digits, 001-999.
-----------	---

## Annex B: Not found in PS

- *Discussion:* The following attributes are used in Annex B but aren't found otherwise in the specification.
- *Recommendation:* Adding information for the attributes in S-102 specification.
  - / Attribute: geographicIdentifier
  - / Attribute: productSpecification
  - / Attribute: timeOfIssue
  - /BathymetryCoverage Attribute: commonPointRule
  - /BathymetryCoverage Attribute: horizontalPositionUncertainty
  - /BathymetryCoverage Attribute: interpolationType
  - /BathymetryCoverage Attribute: numInstances
  - /BathymetryCoverage Attribute: verticalUncertainty
  - /BathymetryCoverage/BathymetryCoverage.01 Attribute: gridOriginLatitude
  - /BathymetryCoverage/BathymetryCoverage.01 Attribute: gridOriginLongitude

## Annex B: Not found in PS

- /BathymetryCoverage/BathymetryCoverage.01 Attribute: gridSpacingLatitudinal
- /BathymetryCoverage/BathymetryCoverage.01 Attribute: gridSpacingLongitudinal
- /BathymetryCoverage/BathymetryCoverage.01 Attribute: numGRP
- /Group\_F/BathymetryCoverage Attribute: chunking
- /Group\_F/TackingListCoverage Attribute: chunking
- /TackingListCoverage Attribute: commonPointRule
- /TackingListCoverage Attribute: horizontalPositionUncertainty
- /TackingListCoverage Attribute: interpolationType
- /TackingListCoverage Attribute: numInstances
- /TackingListCoverage Attribute: verticalUncertainty
- /TackingListCoverage/TackingListCoverage.01 Attribute: gridOriginLatitude
- /TackingListCoverage/TackingListCoverage.01 Attribute: gridOriginLongitude
- /TackingListCoverage/TackingListCoverage.01 Attribute: gridSpacingLatitudinal
- /TackingListCoverage/TackingListCoverage.01 Attribute: gridSpacingLongitudinal
- /TackingListCoverage/TackingListCoverage.01 Attribute: numGRP

## Data structure in S-102 ed. 2.0.0

*Discussion:* We have a self developed implementation of S-102 version 1.0.0 and version 2.0.0. In the following we give you a file size comparison. Version 1.0.0 use two HDF5 Scalar Datasets, one for the depth and one for the uncertainty. For Version 2.0.0 these was changed to one HDF5 Compound Dataset which contains the depths and uncertainties for each node in the grid. For this comparison we used the same bathymetry coverage and the same maximum compression grade.

We think the uncertainty could be helpful in the production process and for scientific questions but disturbing for displaying in ECDIS and navigational purpose.

- Storage Size S-102 Version **1.0.0**                      5.059 kB
  - Elevation:            5.022 kB
  - Uncertainty:            37 kB
  
- Storage Size S-102 Version **2.0.0**                      5.702 kB    ≈ **12 % larger**
  - BathymetryCoverage: 5.702 kB

*Recommendation:* Changing the specification to store the depth and the uncertainty in two separate coverages of the type HDF5 Scalar Dataset and making uncertainty optional.

Thank you!



BUNDESAMT FÜR  
SEESCHIFFFAHRT  
UND  
HYDROGRAPHIE

