Paper for Consideration by S-100 WG TSM7

Potential Changes to S-100 HDF5 for S-104

Submitted by:	Greg Seroka & Kurt Hess on behalf of TWCWG S-104 Project Team
Executive Summary:	Describes potential changes to the S-100 HDF5 structure to support S-104 and other time series data.
Related Documents:	1/ S-104 Ed. 0.0.7 2/ S-111 Ed. 1.0.1 3/ S-100 Ed. 4.0.0

Introduction / Background

The 'Water Level Information for Surface Navigation' Product Specification (S-104, Ed. 0.0.7) is intended, in part, to describe how an HDF5-formatted file can contain water level data of three types: time series data at a number of stations, data such as forecasts on a regular grid, and data on an ungeorectified grid. In development of proposed configurations of HDF5 files for S-104, the TWCWG S-104 Project Team focused on a proposed HDF5 file format to hold the time series data, since gridded data can be structured in ways similar to those described in the S-111 Product Specification. The major alterations in the HDF5 structure resulted from accommodating the two use cases, and incorporating time series data with non-uniform time intervals.

The S-104 draft specification describes two use cases for display of time series data. After selection of a station, the possibilities are:

- display of a graphic showing station name, water level height, and trend at a specific time ("pick report", Figure 1 in Appendix A), and
- graphic plots of multiple types of time series data over a specific time period (Figure 1).

In the graphic plot use case, multiple time series of different water level types (e.g. observed, predicted, forecast) are shown from a single station. The start and end times, number of records, and time intervals may differ for each series. The series may also have variable time intervals. S-100 Ed. 4.0.0 Part 10c-6 mentions as a goal that the S-100 HDF5 profile must apply to "either static data or time series data (for any of the other kinds), with fixed or variable intervals." The S-104 Project Team is therefore supporting time series with variable, or "non-uniform", time intervals. However, S-100 does not describe how to do this.

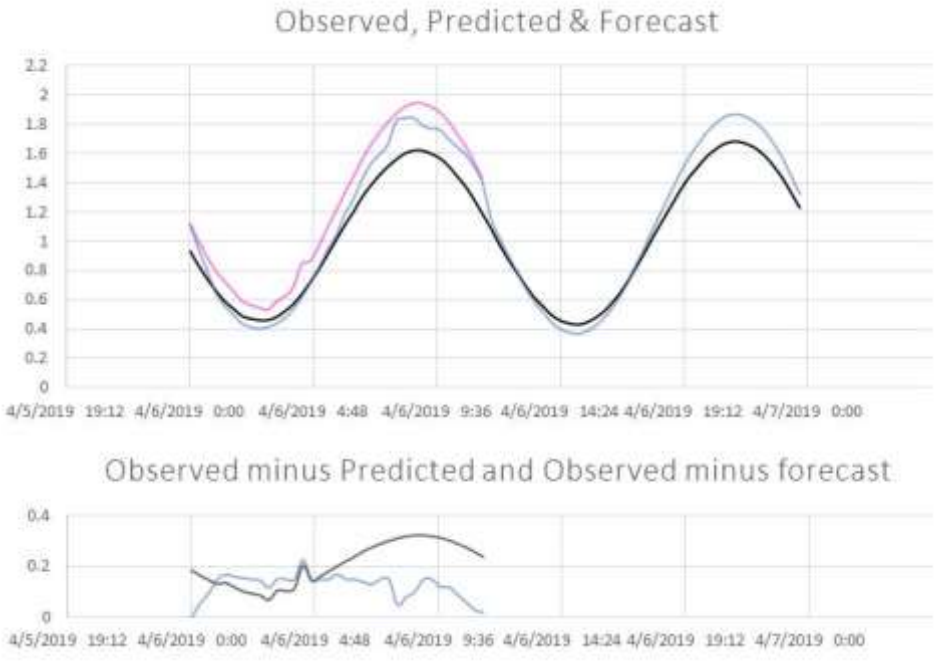


Figure 1. Sample graphic showing five separate time series of water level data. The type of data (observation, prediction, etc.) is represented by the line colour. (see S-104 presentation at TWCWG4, slide 8)

Summary of Changes

The S-104 Project Team has developed through consensus the proposed changes to the S-100 HDF5 structure described in this paper, and the result includes all the necessary water level and supplementary data in a single HDF5 file and in an efficient and easily readable format. The team is testing and finalizing the HDF5 structures, so this paper is meant to start the discussion and get feedback at the S-100 level before formally submitting an S-100 Change Proposal. The changes to S-100 that may need S-100WG approval are:

1. **Non-uniform time intervals**: Currently, S-100 Ed. 4.0.0 details support for only uniform time interval time series data. To support non-uniform time interval data, we introduce a metadata variable *timeIntervalIndex* to indicate that the time interval is non-uniform; its value is either (a) a positive number representing a uniform time interval in seconds, or (b) a negative number denoting non-uniform intervals. The variable *timeRecordInterval* (see S-100 Ed. 4.0.0 Table 10c-12) is then used only for gridded data. The time point of each value, required for non-uniform time interval data, is described below in Item 4 (Triply-compound arrays).
2. **Fixed-station Values Group structure change**: Currently in S-100 Ed. 4.0.0, for all coverage types except moving platforms, each Values Group represents one time record. For fixed stations, we propose to have each Values Group represent one station, and the length of the Values Group be the number of times, not number of stations. This would be consistent with S-111 Ed. 1.0.1.
3. **Expansion of the Values Group metadata**: S-100 Ed 4.0.0 only includes *timePoint* in the metadata at the Values Group level (Table 10c-18). We propose to expand the Values Group metadata to include:
    a. *stationName* and *stationNumber* for station identification
    b. *startDateTime*, *endDateTime*, and *numberOfTimes* as these may differ between time series
    c. *timeIntervalIndex* to accommodate non-uniform time intervals.
4. **Triply-compound arrays**: S-100 Ed. 4.0.0 allows for doubly-compound arrays (e.g. S-111 *surfaceCurrentSpeed* and *surfaceCurrentDirection*). It is not clear if S-100 Ed 4.0.0 allows for triply-compound values datasets. These would be needed to support non-uniform time interval data, as each array element must have three values, e.g. *waterLevelHeight*, *waterLevelTrend*, and *waterLevelTime*.
5. **Other changes we think are allowed in S-100**:
    a. Adding product-specific metadata *pickPriorityType* at the Feature level for fixed station data, to indicate the default priority of time series to display in the pick report when more than one type of series is available (e.g. prediction, observation, forecast).
    b. Migrating optional S-100 variable *numberOfTimes* from the Instance level to the Values Group level.

These changes are described in more detail below. The full document describing the proposed configurations of HDF5 files for S-104 is reproduced in Appendix A of this paper.

References

S-104 – Water Level Information for Surface Navigation Product Specification Ed. 0.0.7 (TWCWG4, April 2019). TWCWG4-04.2. At TWCWG4 documents page https://www.iho.int/mtg_docs/com_wg/IHOTC/TWCWG4/TWCWG4.htm.
S-104 – TWCWG4 Presentation (April 2019). TWCWG4-Presentations. At TWCWG4 documents page https://www.iho.int/mtg_docs/com_wg/IHOTC/TWCWG4/TWCWG4.htm.
S-111 – Surface Current Product Specification Edition 1.0.1 (June 2019).
S-100 – Universal Hydrographic Data Model Edition 4.0.0 (December 2018).
Appendix A – Proposed Configurations of HDF5 Files for S-104

S-100 HDF5 Changes

1. **Non-uniform time intervals**

S-100 Ed. 4.0.0 Part 10c-6 mentions as a goal that the S-100 HDF5 profile must apply to "either static data or time series data (for any of the other kinds), with fixed or variable intervals". As suggested by an S-104 Project Team member, observation time series may have variable time intervals. Therefore, S-104 is including support for time series with variable, or "non-uniform", time intervals.

In a series with uniform time intervals, the time point of each array element is determined by the series' starting date and time and the time interval. With a non-uniform interval, each element must have its own date and time.

There is also a need for a variable to indicate that the interval is non-uniform. We introduce the metadata variable *timeIntervalIndex* to fill this need. Its value can be either (a) a positive number representing a uniform time interval

in seconds, or (b) a negative number to denote non-uniform intervals. The optional variable *timeRecordInterval* (see Table 10c-12, S-100 Ed. 4.0.0) is supplied in the metadata only for gridded data. For case (a), *timeRecordInterval* is numerically equivalent to *timeIntervalIndex.*

2. Fixed-station Values Group structure change

In S-100 Ed. 4.0.0, in Part 10c-9.11 (third paragraph below Table 10c-18), it states "For all coverage types except moving platforms, the number of Groups is the number of time records." But to accommodate varying start times, end times, time intervals, and number of records across different fixed stations, this should be changed to include fixed station data (*dataCodingFormat*=1), so that each Values Group represents one station, and the length of the Values Group is the number of time records. This would be consistent with the examples shown in S-100 Ed. 4.0.0 Figure 10c-9 and S-111 Ed. 1.0.1.

The first paragraph below Table 10c-18 should now read "Time series data for all except the fixed station and moving platforms format (dataEncodingFormat = 1, 4) are encoded in successive groups contained within the instance group." The third paragraph below Table 10c-18 should now read "For all coverage types except fixed stations and moving platforms, the number of Groups is the number of time records." And add "For fixed stations, the number of Groups is the number of stations". Also, Table 10c-17, second row, should be revised from 'values: 1-dimensional Array, length = numberOfStations' to 'values: 1-dimensional Array, length = numberOfTimes'. Table 10c-18, second row, should also be revised to reflect this change. Please see Item 3 (Expansion of the Values Group metadata) below for specific revisions to Table 10c-18.

Below shows the Fixed-station Values Group structure before and after this change:

Before:

Group_001=Data at Time No. 1
values across all stations
Group_002=Data at Time No. 2
values across all stations
Group_003=Data at Time No. 3
values across all stations

After:

Group_001=Data at Station No. 1
values across all times

3. Expansion of the Values Group metadata

Along with the fixed station Values Group structure change, and because each time series may have different characteristics (e.g. start and end time, time intervals, number of records), each Values Group will need additional attributes. Thus, an expanded new metadata block is required for the Values Groups. The current S-100 Ed. 4.0.0 structure of the data Values Groups is depicted in Table 1 below (see S-100 Ed. 4.0.0 Table 10c-18). The only metadata listed currently at this level is *timePoint*.

Table 1 – Existing S-100 structure of value group (S-100, Table 10c-18).

Group	HDF5 Category	Name	Data Type	Data Space
/Group_NNN	Attribute	timePoint (optional)	String (date-time format)	Time point for time series data For other types of data, it can be used to indicate the time for the whole grid
	Dataset	values	Compound	Array of Compound type, with array rank depending on dataCodingFormat and spatial dimension, as described in S-100 Table 10c-17
	Dataset	gridIndex	Bitfield	Required for dataEncodingFormat = 5 or 6 Described in Table 10c-17
	Dataset	cellScale	Compound	Required for dataEncodingFormat = 6 Described in Table 10c-17

In the expanded Values Group metadata (Table 2), the variables *stationName* and *stationNumber* have been added for both identification and possibly for inclusion in the text of the graphic plot. The variables *startDateTime, endDateTime*, and *numberOfTimes* are included here as these may differ between each time series. To accommodate non-uniform time intervals, *timeIntervalIndex* is also added here. Note that additional variables such as Marine Resource Names (MRN) and station category (e.g., high or long-term, medium, or low) could be added here.

Table 2 – New expanded Values Group metadata for S-104.

N	Name	Camel Case	Mult.	Data Type	Remarks and/or Units
dataCodingFormat = 1					
1	Name of the station	stationName	0..1	Character	
2	Station identification number	stationNumber	0..1	Integer	
3	Number of time records	numberOfTimes	1	Integer	
4	Index for time interval	timeIntervalIndex	1	Integer	Positive value denotes timeRecordInterval (sec). Negative value denotes non-uniform time interval.
5	Valid time of earliest value	startDateTime	1	Character	DateTime format
6	Valid time of latest value	endDateTime	1	Character	DateTime format
dataCodingFormat = 2 or 3					
1	Time stamp	timePoint	1	Character	DateTime, gridded data

Just like for Item 2 (Fixed-station Values Group structure change), Table 10c-18, second row, should also be revised to indicate the expanded Values Group metadata. Table 3 shows this change.

Table 3 – New S-100 structure of value group [to replace S-100 Table 10c-18 (Table 1 in this paper)].

Group	HDF5 Category	Name	Data Type	Data Space
/Group_NNN	Attributes	(Values Group metadata attributes)	(see Table 2)	Single-valued attributes as described in Table 10c-XX
	Dataset	values	Compound	Array of Compound type, with array rank depending on dataCodingFormat and spatial dimension, as described in S-100 Table 10c-17.
	Dataset	gridIndex	Bitfield	Required for dataEncodingFormat = 5 or 6 Described in Table 10c-17
	Dataset	cellScale	Compound	Required for dataEncodingFormat = 6 Described in Table 10c-17

4. Triply-compound arrays

To support non-uniform time interval data, an optional feature attribute *waterLevelTime* has been added to the S-104 Data Classification and Encoding Guide. This feature attribute would be required for non-uniform time interval data, to represent each element's date-time. With uniform time interval data, the values dataset would have two elements (doubly-compound): height and trend. With non-uniform time interval data, the values dataset would need to have three elements (triply-compound): height, trend, and date-time.

It is unclear to us whether triply (or higher order) compound values datasets are allowed in S-100 Ed. 4.0.0. Table 10c-17 states for fixed stations that the HDF5 dataset is comprised of a compound array, "one component for each attribute specified in the corresponding feature information dataset in the Feature Information group (Table 10c-8)".

5. Other changes we think are allowed in S-100

We added product-specific metadata and migrated an optional variable downward from one metadata level to another, but we think these changes are allowed in S-100.

The first is that we added *pickPriorityType* to the Feature metadata level under *dataCodingFormat*=1. If there are more than one time series available at a station (e.g. observation, forecast, and prediction), then there needs to be an attribute that defines which series is used by default in a pick report. The variable *pickPriorityType* indicates the priority of time series (highest to lowest priority) for the ECDIS to display. For example, if *pickPriorityType* = "2,1,5" (a character string), the system will select prediction series (data type=2), then observation series (data type=1) if the prediction is unavailable, then forecast series (data type=5) if the previous two are not available.

Finally, we migrated optional S-100 variable *numberOfTimes* from the Instance level to the Values Group level, as shown in Item 3 above.

Recommendations

We recommend the changes to S-100 HDF5 described above be considered. The changes will allow for HDF5 files to include in an efficient and easily readable format all the required information for water level time series data. These changes should also benefit other S-XXX HDF5 time series data, such as S-111.

Action Required of the TSM

The TSM is invited to:

a. Discuss the proposed changes to the S-100 HDF5 format.

b. Make a recommendation as to whether the changes are sufficiently justified and should be included in an S-100 Change Proposal, or if items should be added/deleted/modified.

c. Endorse the recommendation set forth in the Recommendations section, as modified during the meeting.

d. Take further action as appropriate.

Appendix A. PROPOSED CONFIGURATIONS OF HDF5 FILES FOR S-104

K. Hess
September 13, 2019

1. INTRODUCTION

The 'Water Level for Surface Navigation' Product Specification (S-104, Ed. 0.0.7) is intended, in part, to describe how an HDF5-formatted file can contain water level data of three types: time series data at a number of stations, data such as forecasts on a regular grid, and data on an ungeorectified ('irregular') grid. Here we focus on a proposed HDF5 file format to hold the time series data, since gridded data can be structured in ways similar to those described in the S-111 Product Specification. The major adaptations in the HDF5 structure resulted from accommodating the two use cases, and incorporating time series data with non-uniform time intervals.

Use Cases

The S-104 draft specification describes two use cases for display of time series data. After selection of a station, the possibilities are:

- display of a graphic showing station name, water level height, and trend at a specific time (Figure 1), and

- graphic plots of multiple types of time series data over a specific time period (Figure 2).

All water level and supplementary data to support both these use cases is to be supplied in a single HDF5-formatted file.

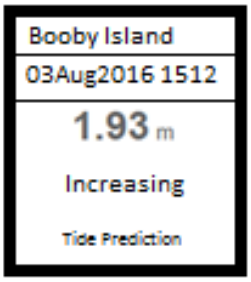


Figure 1. A sample pick report for a water level station (see S-104, Ed. 0.0.7, Sec.9.2.2). The information displayed includes the station name, the valid date and time (local or UTC), the height, the trend, and the type of water level data.

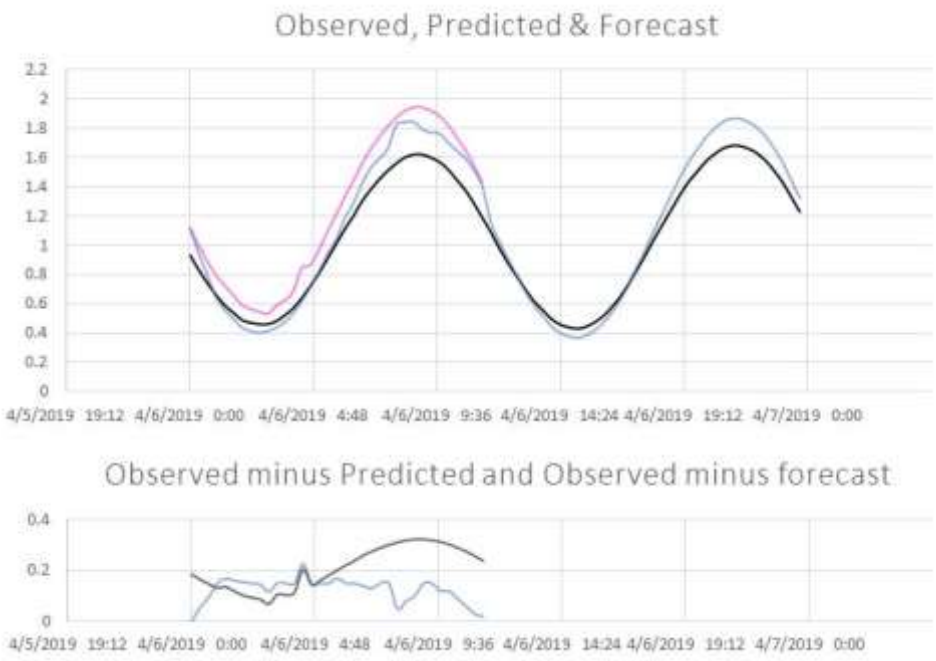


Figure 2. Sample graphic showing five separate time series of water level data. The type of data (observation, prediction, etc.) is represented by the line colour. (see S-104 presentation at TWCWG4, slide 8)

The approach used here is to expand the S-100 HDF5 format to include additional HDF5 attributes. These are (a) additional water level station information for the pick report, (b) establishing a default priority of the types of time series when more than one are available, and (c) migrating optional attributes downward from the Feature to Instance level, or Instance level to the Values Group level.

Non-uniform Time Intervals

Another important change to HDF5 file format is the inclusion of time series with non-uniform time intervals. In a series with uniform time intervals, each pair of height and trend variables is contained in a doubly-compound dataset (i.e., each element in the array has two values). The time point of each pair of variables is determined by the series' starting date and time and the time interval. With a non-uniform interval, each pair of height and trend must have its own date and time. The triply-compound array thus has three elements: height, trend, and date-time.

Adding a date-time variable will require a new S-100 Feature Attribute, to be called *waterLevelTime,* to be added to the existing attributes *waterLevelHeight* and *waterLevelTrend.* There is also a need for a variable to indicate that the interval is non-uniform. We introduce the metadata variable *timeIntervalIndex* to fill this need. Its value can be either (a) a positive number representing a uniform time interval in seconds, or (b) a negative number to denote non-uniform intervals. The variable *timeRecordInterval* (see Table 10c-12, S-100 Ed. 4.0.0) is used only for gridded data.

2. EXAMPLE: DATA FOR TWO STATIONS AND USE CASES

As mentioned above, all time series contain at least the height and trend, and some series may contain a date-time variable. Among the series, there may be differences in start times, end times, time intervals, and number of records. In reference to the discussions below, the following table is provided to show the list of proposed S-104 water level data types (Table 1).

Table 1. Types of water level data, based on their source. The Residual series are to be supplied by the HO, not calculated by the EDCIS.

N	Name	Camel Case	Mult	Data Type	Remarks and/or Units
	Type of water level data	typeOfWaterLevelData	1	Enumeration	1: Observation 2: Astronomical prediction 3: Analysis or hybrid method 4: Hydrodynamic model hindcast 5: Hydrodynamic model forecast 6: Residual: predicted minus observed 7: Residual: analysis minus observed 8: Residual: hindcast minus observed 9: Residual: forecast minus observed

Suppose there are two water level stations, each with a different set of time series data:

Station No. 1 has only one time series: an astronomical prediction

Station No. 2 has three series: an astronomical prediction, a model-based forecast, and real-time observations

For each of the stations in the data file, we assume that the ECDIS user may select (by means of placing the cursor over the station symbol on the display) a pick report showing height and trend, a graphic plot of one or more of the available time series, or both. In addition, either case may display the station name, station number, vertical datum, dates and times, and/or other available information.

Use Case: Pick Report

Suppose an ECDIS user selects Station No. 1. For the pick report, the height and trend will come from the only available series, the predictions.

Now suppose an ECDIS user selects Station No. 2. For the pick report, the height and trend will be available from one of three possible series. Thus there needs to be an attribute that defines which series is used, for example by a priority list, or the capability for the user to select the series, or both.

Here we introduce a variable, *pickPriorityType,* to indicate the default priority of time series (highest to lowest priority) for the ECDIS to display when more than one type of series is available. For example, using the data type information in Table 1, if *pickPriorityType* = “2,1,5” (a character string), the system will select prediction series (data type=2), then observation series (data type=1) if the prediction is unavailable, then forecast series (data type=5) if the previous two are not available. There must be as many numbers in the string as there are data types in the file. This parameter would be an HDF5 Feature-level attribute.

Use Case: Graphic Plot

For Station No. 1, if the user selects a plot of the time series, the time series of predicted height values will be plotted. The trend is not plotted.

For Station No. 2, if the user selects a plot of the time series heights, all available (three) time series heights (but not trends) will be plotted, unless an option exists for deselection of some series. The colour and line type (solid, dashed, dotted, etc.) for each series is to be defined in the S-104 Product Specification.

3. HDF5 FILE STRUCTURE

The data for these two stations, using the HDF5 format outlined in Option A (cf. email of May 21: 'Potential Configurations of HDF5 Files for S-104'), can be structured as shown below (Figure 3). A sample of the pertinent attributes are listed at each level; all attributes are shown in Tables 2-5 below. As explained above, each time series may have different characteristics (e.g., start and end time, time interval index, number of records), so that each Values Groups will need additional attributes.

HDF5 File
Attributes: Product Specification, dateTime of issue, bounding box, datums, geographic identifier

Feature: WaterLevel
Attributes: number of instances, time uncertainty, common point rule

Instance: WaterLevel.01=Predictions Only (data type 2)
Attributes: data type, number of Values Groups

Values Group: Group_001=Data for Station No. 1
Attributes: station information, series start and end time, time interval index, number of records

Values Group: Group_002=Data for Station No. 2
Attributes: station information, series start and end time, time interval index, number of records

Instance: WaterLevel.02=Real-time Observations Only (data type 1)
Attributes: data type, number of Values Groups

Values Group: Group_001=Data for Station No. 2
Attributes: station information, series start and end time, time interval index, number of records

I**nstance**: WaterLevel.03=Forecasts Only (data type 5)
Attributes: data type, number of Values Groups

Values Group: Group_001=Data for Station No. 2
Attributes: station information, series start and end time, time interval index, number of records

Figure 3. Plain-language outline of the proposed HDF5 file structure for the S-104 product as applied to data for two stations. Note that Group_F, the uncertainty dataset, and the axisNames dataset are not shown.

Note that 'station information' is a set of parameters, including identification (a unique code, such as a name, number, or character string, which identifies the station), MRN, vertical datum, etc.

The ECDIS will read the HDF5 file and, for each station, be required to keep track of which time series are available, and for each series, keep track of the attributes such as type, start time, time interval, etc.

4. METADATA AND ATTRIBUTES

Following the S-100 Ed. 4.0.0 samples, the draft S-104 metadata is divided into four sections, corresponding to the General Metadata (Table 2), the Feature Metadata (Table 3), the Instance Metadata (Table 4) and the Values Group metadata (Table 5). These are shown in the following tables.

General Metadata

Here the bounding box applies to all stations (and to grids for other d*ataCodingFormat* values). Also, the vertical datum is assumed to apply to all time series.

Table 2 – General Metadata, related to the entire HDF5 file (see S-100, Table 10c-6; S-111 Table 12.1).

N	Name	Camel Case	Mult.	Data Type	Remarks and/or Units
1	Product specification number and version	productSpecification	1	Character	This must be encoded as 'INT.IHO.S-104.X.X', with Xs representing the version number
2	Time of data product issue	issueTime	1	Character	Time (UTC). E.g., 123000Z
3	Date of data product issue	issueDate	1	Character	Date must be consistent with issueDate in discovery metadata.
4	Horizontal datum	horizontalDatumReference	1	Character	EPSG
5	Horizontal datum number	horizontalDatumValue	1	Integer	4326 (for WGS84)
6	Epoch of realization	epoch	0..1	Character	Code denoting the epoch of the geodetic datum used by the CRS. For example G1762 for the 2013-10-16 realization of the geodetic datum for WGS84
7	Bounding box	westBoundLongitude	1	Float	Area encompassing all feature instances
8		eastBoundLongitude	1	Float	Units are Degrees.
9		southBoundLatitude	1	Float	
10		northBoundLatitude	1	Float	
11	Geographic locator	geographicIdentifier	0..1	Character	Description
12	Metadata file name	metadata	1	Character	Name of XML metadata file for the HDF5 file. Form: MD_<hdf file name>.XML.
13	Vertical datum reference	verticalDatum	0..1	Enumeration	See S100_VerticalAndSoundingDatum

NOTE: For the *Enumeration* data type, the entry is a single integer.

Feature Metadata

Here the number of instances (*numInstances*) denotes the number of types (see Table 1) of time series. The new variable *pickPriorityType* appears in the section for *dataCodingFormat* = 1.

Table 3 – Feature Metadata, pertaining to the Water Level feature (see S-100, Table 10c-10; S-111, Table 12.2).

N	Name	Camel Case	Mult	Data Type	Remarks and/or Units
1	Data organization index, used to read the data (Table 10.1)	dataCodingFormat	1	Enumeration	1: Time series at fixed stations 2: Regularly-gridded arrays 3: Ungeorectified gridded arrays
2	Dimension	dimension	1	Integer	The (spatial) dimension of the feature instances. For water levels, use 2.
3	Common Point Rule	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. 1: average 2: low 3: high 4: all (recommended)
4	Horizontal position uncertainty	horizontalPositionUncertainty	1	Float	-1.0 (unknown) or positive value (m)
5	Vertical position uncertainty	verticalUncertainty	1	Float	-1.0 (unknown) or positive value (m)
6	Time uncertainty	timeUncertainty	1	Float	-1.0 (unknown) or positive value (s)
7	Number of feature instances	numInstances	1	Integer	Num. of series types, gridded forecasts, etc
	ADDITIONAL METADATA FOR S-104				
8	Methodology	methodWaterLevelProduct	0..1	Character	Brief description of tide gauge type, forecast method or model, etc.
9	Min. water level height in dataset	minDatasetHeight	1	Float	
10	Max. water level height in dataset	maxDatasetHeight	1	Float	
dataCodingFormat = 1					
11	Order of series in pick report	pickPriorityType	1	String	E.g., “2,1,4,5,3” (without quotes) See Table 1. Total numbers (here 5) must be equal to numInstances
dataCodingFormat = 2					
11	Sequencing Rule	sequencingRule.type	1	Enumeration	Method to be used to assign values from the sequence of values to the grid coordinates. Components: type: Enumeration CV_SequenceType For example 1 (for ‘linear’)
12		sequencingRule.scanDirection	1	String	scanDirection: String <axisNames entry> (comma-separated). For example “latitude,longitude”
13	Interpolation Type	interpolationType	1	Enumeration	Interpolation method recommended for evaluation of the S100_GridCoverage Values: CV_InterpolationMethod (ISO 19123).
dataCodingFormat = 3					
11	Interpolation Type	interpolationType	1	Enumeration	Interpolation method recommended for evaluation of the S100_GridCoverage Values: CV_InterpolationMethod (ISO 19123).

Instance Metadata

Here, each Instance corresponds to a single type of time series (see *typeOfWaterLevelData*). The bounding box is optional here, but may be included if the stations for this *typeOfWaterLevelData* is a small subset of the total. Also, the variables *dateTimeOfFirstRecord* and *dateTimeOfLastRecord* refer to all time series in this Instance only. The variable *numberOfStations* appears here since it is used to determine array sizes.

Table 4 – Instance Metadata, pertaining to the feature instance (e.g., WaterLevel.01) (see S-100, Table 10c-12; S-111, Table 12.3).

N	Name	Camel Case	Mult.	Data Type	Remarks and/or Units
1	Bounding box	westBoundLongitude	0..1	Float	Area of grid, set of stations, etc.
2		eastBoundLongitude	0..1	Float	Units are decimal degrees.
3		southBoundLatitude	0..1	Float	
4		northBoundLatitude	0..1	Float	
5	Time interval	timeRecordInterval	0..1	Integer	The interval between time records. Units: Seconds. Use only for dataCodingFormat = 2 or 3.
6	Valid time of earliest value	dateTimeOfFirstRecord	0..1	Character	DateTime format, first record in the Instance
7	Valid time of latest value	dateTimeOfLastRecord	0..1	Character	DateTime format
8	Number of values groups	numGRP	1	Integer	Number of Values Groups. For time series, equals the number of stations. For grids, equals the number of time points.
9	Instance chunking value	instanceChunking	0..1	Character	For example “1,256” (without quotes). If present, overrides attribute value in Group_F (See Sec. 10.2.2)
Additional attributes for S-104					
10	Type of water level data	typeOfWaterLevelData	1	Enumeration	(Table 1)
dataCodingFormat = 1					
11	Number of fixed stations	numberOfStations	1	Integer	Number of individual fixed stations in this instance. Must equal *numGRP*.
dataCodingFormat = 2					
11	Longitude of grid origin	gridOriginLongitude	1	Float-Double	Degrees
12	Latitude of grid origin	gridOriginLatitude	1	Float-Double	Degrees
13	Grid spacing, long.	gridSpacingLongitudinal	1	Float-Double	Degrees
14	Grid spacing, lat.	gridSpacingLatitudinal	1	Float-Double	Degrees
15	Number of points, long.	numPointsLongitudinal	1	Integer	numCOL
16	Number of points, lat.	numPointsLatitudinal	1	Integer	numROW
17	Start sequence	startSequence	1	Character	E.G., “0,0” (without quotes)
dataCodingFormat = 3					
11	Number of nodes	numberOfNodes	1	Integer	

Values Group Metadata

An expanded new metadata block is required for the Values Groups. The variables *stationName* and *stationNumber* have been added for both identification and possibly for inclusion in the text of the graph. Note that additional variables such as Marine Resource Names (MRN) and station category (e.g., high or long-term, medium, or low) can be added here. The series start and end times, number of records, and time interval index are included since they may differ for each series.

NOTE: This metadata is beyond what is specified in S-100 Ed. 4.0.0, so may require approval.

Table 5 – Values Group metadata (see S-100, Table 10c-18).

N	Name	Camel Case	Mult.	Data Type	Remarks and/or Units
dataCodingFormat = 1					
1	Name of the station	stationName	0..1	Character	
2	Station identification number	stationNumber	0..1	Integer	
3	Number of time records	numberOfTimes	1	Integer	
4	Index for time interval	timeIntervalIndex	1	Integer	Positive value denotes timeRecordInterval (sec). Negative value denotes non-uniform time interval.
5	Valid time of earliest value	startDateTime	1	Character	DateTime format
6	Valid time of latest value	endDateTime	1	Character	DateTime format
dataCodingFormat = 2 or 3					
1	Time stamp	timePoint	1	Character	DateTime, gridded data

5. SAMPLE HDF5 FILES

The following figures show results of a test HDF5 file created for the two stations described above in Secs. 3, 4, and 5. A number of assumptions have been made for convenience, but the following points are open for further discussion by the Project Team and TWCWG:

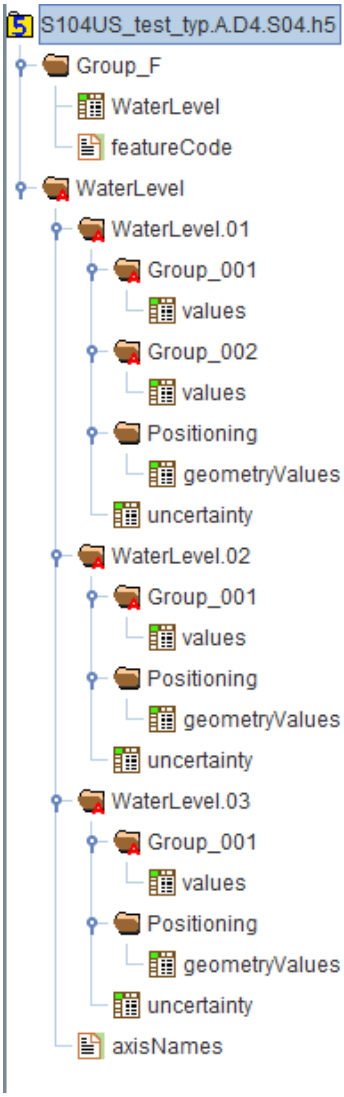
Datasets for Positioning and Geometry (Figure 4a) must be repeated in each Instance, according to S-100 Ed 4.0.0. This arrangement seems to be redundant and may lead to increased file size.

The Bounding Box for the entire set of stations appears in the General metadata (Figure 4b). Similar data, for a subset of stations, has not been included here in the Instance metadata (Table 4).

The Vertical Datum enumeration (Figure 4b) applies to all series. If some series have a different datum, this value would have to be migrated downward.

In the Group_F Feature dataset (Figure 4c), the fill values (denoting a missing value) used here are somewhat arbitrary, and absent as applied to trend, since ‘4’ (unknown) can be used. The upper and lower limits to heights and date-time are somewhat arbitrary.

The Uncertainty default value (-1.) is arbitrary.



(a)

```
/ (96)
  Group size = 2
  Number of attributes = 13
    eastBoundLongitude = -76.29
    epoch = G1762
    geographicIdentifier = Chesapeake Bay
    horizontalDatumReference = EPSG
    horizontalDatumValue = 4326
    issueDate = 20190717
    issueTime = 231100Z
    metadata = MD_S104CCnnnnnnnn.XML
    northBoundLatitude = 39.250123
    productSpecification = INT.IHO.S-104.0.0
    southBoundLatitude = 38.74
    verticalDatum = 12
    westBoundLongitude = -76.41
```

(b)

	code	name	uom.name	fillValue	dataType	lower	upper	closure
0	waterLevelHeight	Water level height	meters	-999.	H5T_FLOAT	-99.99	99.99	closedInterval
1	waterLevelTrend	Water level trend			H5T_NATIVE_INT16	1	4	closedInterval
2	waterLevelTime	Water level time	DateTime		H5T_C_S1	19000101T000000Z	21500101T000000Z	closedInterval

(c)

Figure 4 – (a) HDFView of proposed file structure, (b) General HDF5 attributes (see Table 2), and (c) the WaterLevel dataset in Group_F.

```
WaterLevel (6048, 4)
  Group size = 4
  Number of attributes = 11
    commonPointRule = 4
    dataCodingFormat = 1
    dimension = 2
    horizontalPositionUncertainty = -1.0
    maxDatasetHeight = 2.911
    methodWaterLevelProduct = prediction, obs, fcst
    minDatasetHeight = 0.665
    numInstances = 3
    pickPriorityType = 2,1,5
    timeUncertainty = -1.0
    verticalUncertainty = -1.0
```

(a)

```
WaterLevel.01 (8360, 4)
  Group size = 4
  Number of attributes = 3
    numGRP = 2
    numberOfStations = 2
    typeOfWaterLevelData = 2
```

(b)

Figure 5 – (a) Feature (WaterLevel) attributes (see Table 3) and (b) Instance (WaterLevel.01) attributes (see Table 4).

Group_001 (15352, 12)
Group size = 1
Number of attributes = 6
endDateTime = 20190710T000000Z
numberOfTimes = 673
startDateTime = 20190703T000000Z
stationName = Station_Alpha
stationNumber = 8000101
timeIntervalIndex = 900

(a)

	waterLevelHeight	waterLevelTrend
0	1.325	4
1	1.491	4
2	1.655	4
3	1.818	4
4	1.975	1
5	2.126	1
6	2.267	1

(b)

Group_001 (24536, 12)
Group size = 1
Number of attributes = 6
endDateTime = 20190704T000100Z
numberOfTimes = 97
startDateTime = 20190703T000100Z
stationName = Station_Beta
stationNumber = 8000102
timeIntervalIndex = -900

(c)

	waterLevelHeight	waterLevelTrend	waterLevelTime
0	1.464	1	20190703T000100Z
1	1.524	1	20190703T001500Z
2	1.578	1	20190703T003200Z
3	1.625	1	20190703T004410Z
4	1.665	1	20190703T010100Z
5	1.697	1	20190703T011540Z
6	1.722	1	20190703T012900Z

(d)

Figure 6 – (a) Attributes for the Values Group (WaterLevel.01,Group_001) (see Table 5) and (b) sample values array for uniform time intervals, and (c) attributes for the Values Group (WaterLevel.02,Group_001) and (d) sample values array for non-uniform time intervals.

	longitude	latitude
0	-76.29	39.250123
1	-76.41	38.74

(a)

	name	value
0	waterLevelHeightUncertainty	-1.0
1	waterLevelTrendUncertainty	-1.0

(b)

Figure 7 – (a) Positioning dataset for WaterLevel.01, and (b) the Uncertainty dataset. Time uncertainty appears in the Feature metadata (Figure 5a).

6. SUMMARY AND CONCLUSIONS

A preliminary structure for an HDF5 file for time series water level data has been discussed. The main topics covered in the above Sections are as follows:

Graphic Plots – The use case for graphic plots has required some specific HDF5 file structures so that types of time series (prediction, observations, etc.) can be easily identified and selected for plotting by the ECDIS.

Pick Report – The use case for a pick report for a station with multiple time series has required a new variable (*pickPriorityType)* to determine a priority in selection among the available series to supply the height and trend.

Non-uniform Time Interval – A new variable (*timeIntervalIndex)* is used to denote whether the time series intervals are uniform or non-uniform. If they are non-uniform, a new S-100 Feature Attribute, *waterLevelTime,* has been introduced to contain the time information.

Changes in the HDF5 structure as described in the S-100 Ed 4.0.0 standard have been proposed. Some of these are not specifically covered in the S-100 section on Overriding Attributes (Part 10c-9.7.1), and may have to be approved by the S-100WG.

An attribute *pickPriorityType* has been added at the Feature level under *dataCodingFormat*=1.

One attribute of the time series (*typeOfWaterLevelData*) has been migrated downward from the Feature level to the Instance level.

One attribute of the time series (*numberOfTimes*) has been migrated downward from the Instance level to the Values Group level.

New attributes have been added at the Values Group level, including tide station and related information, *startDateTime,* and *endDateTime.*

The attribute *timeRecordInterval* at the Instance level is not used for time series but is replaced by the attribute *timeIntervalIndex* at the Values Group level.

A few additional topics arose, but were not directly addressed, during the course of development.

Trend Criteria - There was some discussion of the Trend criteria for differentiating between the categories (Table 6). One proposal is to use a critical value of 0.2 m (change in height over 1 hour) for all HOs. If each HO is allowed to use a different value, then this information must be included in the metadata.

Table 6. Water level trend categories. The trend is based on a 1-hr period, with the magnitude of the change related to a critical value (e.g., 0.2 m).

N	Name	Camel Case	Mult	Data Type	Remarks and/or Units
	Water level trend	waterLevelTrend	1	Enumeration	1: Steady 2: Decreasing 3: Increasing 4: Unknown or not available

User Input - User input for Pick Report and Graphic Plots is a possibility. For example, if there are multiple time series at a station, the user may be able to select the series for the pick report, over-riding the default, or deselecting some of the time series in the graphic plot. These options can be discussed as part of the portrayal specifications.

Redundant Data – When stations have multiple time series, some information is repeated in each Instance. Station information (name, identification number, MRN, etc.) and the geography data is repeated in each Instance. Perhaps these redundancies can be avoided.

Uniform Station Identification – At present there seems to be no standard way of identifying stations. Using a station name may lead to spelling problems or uniqueness conflicts among HOs. Perhaps a uniform approach should be considered.

The next steps are to review the discussion above and create (and read) test files for time series that follow the proposed formats. Another recommended step would be to read the test files and display the data for both use cases.

Following that, creating and reading gridded fields would be advisable, as a test of the metadata.

Once we are confident of the formats, we can then seek approval from the S-100WG for the changes we have proposed.