Project input to IHO standardisation S-102 Bathymetric Surface Product Specification

Written by Svein Skjaeveland - Electronic Chart Centre
Second Version - March 2019

Summary
The IHO standard S-102 Bathymetric Surface Product Specification 1.0.0 was released in 2012. It was then the first official standard built on the S-100 concept. The last 3 years there has been ongoing work by the IHO S-102 project team to release an improved version of the standard.

During our project period several topics discussed has resulted in us giving input to the IHO S-102 Project Team for standardisation. This appendix summarizes and shortly describes the input given. Some of our suggestions have resulted in direct changes to the standard, and some have been postponed to future discussions.

During the project period the S-102 Project Team has conducted 2 physical meetings:
- 16 & 18 March 2017 Genoa Italy
- 09 & 12 April Singapore

Besides those meetings work has been conducted using email communication. A Nordic workshop was also held in Sweden during the project period.

This annex discusses further the following topics:

Input given resulting in direct changes:
1. Coordinate Reference Systems
2. New edition Coverage
3. Naming convention
4. Dataset Size
5. Gridding Methodology
6. Purpose
7. Data integrity and encryption
8. Data Quality

Input given postponed to future discussions:
9. Tiling Schemes
10. Standard Align Grids
11. Validation
12. Single vs Multiple tile data products
13. Multi resolution data products
14. Grid resolution quad-tree approach

Input given resulting in direct changes

1. Coordinate Reference systems

The standard did not limit the use of coordinate reference systems sufficiently, resulting in unnecessary challenges for OEMs and Service Providers. They would have to implement support for too many coordinate reference systems. A limited number of systems using the ESPG code was suggested. Based on our feedback the following limitation has been included in the draft new edition:

Table 5.1 – S-102 Coordinate Reference Systems (EPSG Codes).

EPSG Code	Coordinate Reference System
4326	WGS84
32601 – 32660	WGS 84 / UTM Zone 1N to Zone 60N
32701 - 32760	WGS 84 / UTM Zone 1S to Zone 60S
5041	WGS 84 / UPS North (E,N)
5042	WGS 84 / UPS South (E,N)
The full reference to EPSG can be found at www.epsg-registry.org.	

2. New edition coverage

In order to avoid confusion for end users and distributors when issuing new editions of a dataset, the new edition should cover the same area as the previous edition. Based on our feedback this requirement has now been included in Dataset Management in the draft new edition:

11.2.1 Dataset Management

Two types of dataset files may be produced and contained within an exchange set: □
New dataset: Initial.

- New edition of a dataset: Includes new information. New editions must cover the same area as its predecessor.

3. Naming Convention

The work conducted with S-102 has been done in conjunction with working on other S-100 Product Specifications, especially S-101. During this work a suggestion to a common naming convention for all S-10x products has been accepted. Based on our feedback this requirement has now been included in Dataset file naming in the draft new edition:

11.2.3 Dataset file naming

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

102YYYYØØØØØØØØØØ

- 102 – the first 3 characters identify the dataset as an S-102 dataset (mandatory).
- YYYY - the fourth to seventh characters identify the producer code of the issuing agency (mandatory for S-102). Where the producer code is derived from a 2- or 3-character format, the missing characters of the producer code must be populated with zeros (“00” or “0” respectively) for the sixth and seventh characters of the dataset file name, as required.
- ØØØØØØØØØØ - the eighth to the maximum seventeenth 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).

4. Dataset Size

During the project period the average size of the test datasets created has been in the range 150 – 200 MB. During lengthy discussions we proposed a standard acceptance of dataset size to a minimum of 200 MB. In the end an acceptance dependent on digital or physical transfer media was accepted. However, a note added specifically identify these sizes as informative information only, letting the final selection of an appropriate file size be decided by the data producer. Based on our feedback the following information has been included in Dataset size in the draft new edition:

11.2.2 Dataset Size

S-102 delivery will take place in one of two forms: 1) network transfer to platform (that is, internet download), or 2) physical media transfer (that is, removable media). An example scenario for each method has been provided below:

> ***Note:*** *The use of 10 MB and 256 MB in this and other sections should be treated as informative information only. Additionally, any computed values associated with either file size limit should be treated as approximate answers. Final selection of an appropriate file size limit or grid resolution is left to the discretion of the data producer.*

Network Transfer: To minimize overall file size, the HO produces a 10 MB file for wireless transmission to marine vessels. In uncompressed form, this file would contain roughly 600 nodes by 600 nodes.

Physical Transfer: The HO produces a 256 MB file which can be personally delivered to a ship in port. In uncompressed form, this file would contain roughly 5700 nodes by 5700 nodes.

> *Table 11.1: Informative Grid Resolutions and Tile Size at Chart Scale, provides general information to aid in the compilation of S-102 data for specific charting scales.*

> *Annex F: S-102 Dataset Size and Production, discusses in greater detail the physical size components of an S-102 file.*

5. Gridding Methodology

A bathymetric grid can be calculated differently using different calculation algorithms. It is important to convey this information for the end user to know which methodology was chosen for grid calculation. Based on our feedback the following metadata has been included in the dataset discovery metadata in the draft new edition:

griddingMethod	Algorithm used to calculate grid values	0..1	Class	1. basicWeightedMean 2. shoalestDepth 3. tpuWeightedMean 4. cube 5. nearestNeighbour 6. naturalNeighbour 7. polynomialTendency 8. spline 9. kriging

A more descriptive section of the various gridding methods has been added – some of them using uncertainty in different ways. Based on our feedback Annex G has been added to the draft new edition:

Annex G. S-102 Gridding Methods

- The **Basic Weighted Mean** algorithm computes an average depth for each grid node. Contributing depth estimates within a given area of influence are weighted and averaged to compute the final nodal value.
- The **Shoalest Depth** algorithm examines depth estimates within a specific area of influence and assigns the shoalest value to the nodal position. The resulting surface represents the shallowest depths across a given area.
- The **Total Propagated Uncertainty (TPU) Weighted Mean** algorithm makes use of the elevation and associated total propagated uncertainty for each contributing depth estimate to compute a weighted average depth for each nodal position.

  Note: TPU is a measure of the expected accuracy of the depth estimate when all relevant error/uncertainty sources have been considered.
- The **Combined Uncertainty and Bathymetric Estimator, or CUBE** makes use of the elevation and associated total propagated uncertainty for each contributing depth estimate to compute one or many hypotheses for an area of interest. The resulting hypotheses are used to estimate statistical representative depths at each nodal position.
- The **Nearest Neighbour** algorithm identifies the nearest depth value within an area of interest and assigns that value to the nodal position. This method does not consider values from neighbouring points.
- **Natural Neighbour** interpolation identifies and weights a subset of input samples within the area of interest to interpolate the final nodal value.
- The **Polynomial Tendency** gridding method attempts to fit a polynomial trend, or best fit surface to a set of input data points. This method can project trends into areas with little to no data, but does not work well when there is no discernible trend within the data set.
- The **Spline** algorithm estimates nodal depths using a mathematical function to minimize overall surface curvature. The final "smoothed" surface passes exactly through the contributing input depth estimates.
- **Kriging** is a geostatistical interpolation method that generates an estimated surface from a scattered set of points with a known depth.

6. <u>Purpose</u>

The standard support products being produced for 2 different purposes, navigation and other purpose. It was important that the purpose information could be encoded to inform the user whether the product was intended for safe navigation or not. Based on our feedback the following metadata has been included in the dataset discovery metadata in the draft new edition:

purpose	The purpose for which the dataset has been issued	1	Class MD_Identification>purpose	1. navigation 2. other

7. <u>Data Integrity and encryption</u>

Having sufficient data integrity and encryption is an absolute necessity for being able to trust the data and the data delivery. A whole new chapter has been written in S-100 on this topic, Part 15 – Encryption and Data Protection. As part of this work, data integrity and encryption has also been incorporated in this product specification. This also includes the incorporation of the necessary metadata fields for information. Based on our feedback the following information has been included in the Data Integrity and Encryption in the draft new edition:

11.4 Data integrity and encryption

S-100 Part 15 defines the algorithms for compressing, encrypting and digitally signing datasets based on the S-100 Data Model. The individual product specifications provide details about which of the elements are being used and on which files in the dataset.

11.4.1 Use of Compression

The data producer decides if compression will be used on the S-102 product files (HDF5). It is expected that a hydrographic office will make a policy decision and that all the S-102 datasets from the producer will be either compressed or uncompressed.

It is recommended to compress all the dataset files, e.g. HDF5 files. The ZIP compression method defined in S-100 Part 15 shall be applied to the product files.

The use of compression will be encoded:

Since information about compression is encoded in the S-102_ExchangeCatalogue, it is implicitly applied to all the dataset files in the exchange set. It will not be possible to create an exchange set where some HDF5 files are compressed while others are not. In cases where a data distributor produces an integrated S-102 product, all sources are required to be either compressed or uncompressed at time of integration. In this situation the digital signature encoded into source data (i.e. original data producer) will be replaced with that of the distributor (Data Server).

11.4.2 Use of Data Protection

It is recommended to encrypt all the dataset files, e.g. HDF5. The encryption method defined in S-100 part 15 shall be applied.

8. Data Quality

Quality discussions have been part of every project meeting. Being able to see which quality a product have is important for the end user. On our initiative the IHO Data Quality Working Group (DQWG) was approached to consider how quality should be maintained within the S-102 standard. As a result of DQWGs recommendations a data quality section has been included in the standard. A new section is now included describing the following data quality components:

- Completeness
- Logical Consistency
- Positional Accuracy
- Temporal Accuracy
- Thematic accuracy

As for sufficient methodology to portray data quality, this is still ongoing work and not part of the proposed draft new edition. There are still unresolved issues related to the display of uncertainty, and this is also relevant to other S10x products where uncertainty is a factor. The DQWG is tasked to find a solution for quality portrayal, and the result of their work may affect how uncertainty will be portrayed for S-102 products.

The issue of data quality is also strongly related to the gridding methodology as mentioned above in this annex.

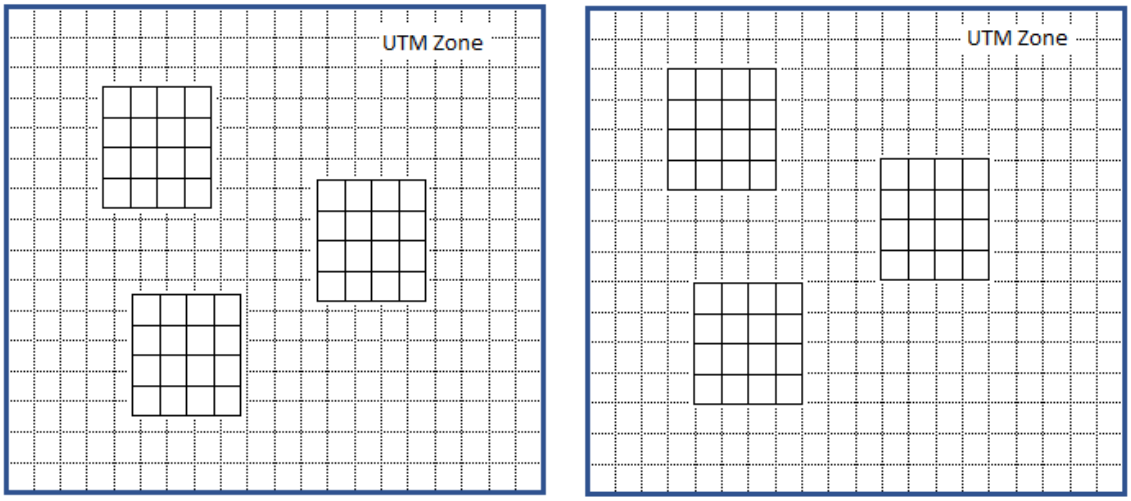
Input given postponed to future discussions

9. Tiling Schemes

Discussions related to national tiling schemes recommendations have been decided postponed to a future edition of the standard.

10. Standard aligned grids

We have highlighted the benefits of producing standard aligned grids within a defined geographical area. Our intention was to implement recommendations in the standard for data product creation outside of national predefined tiling schemes. The most important benefit is to avoid additional hardware calculations to display non-aligned data together. The following figure explains visually the difference between aligned and not aligned datasets in a UTM zone.



Grid not aligned - no common origo **Grid aligned - common origo**

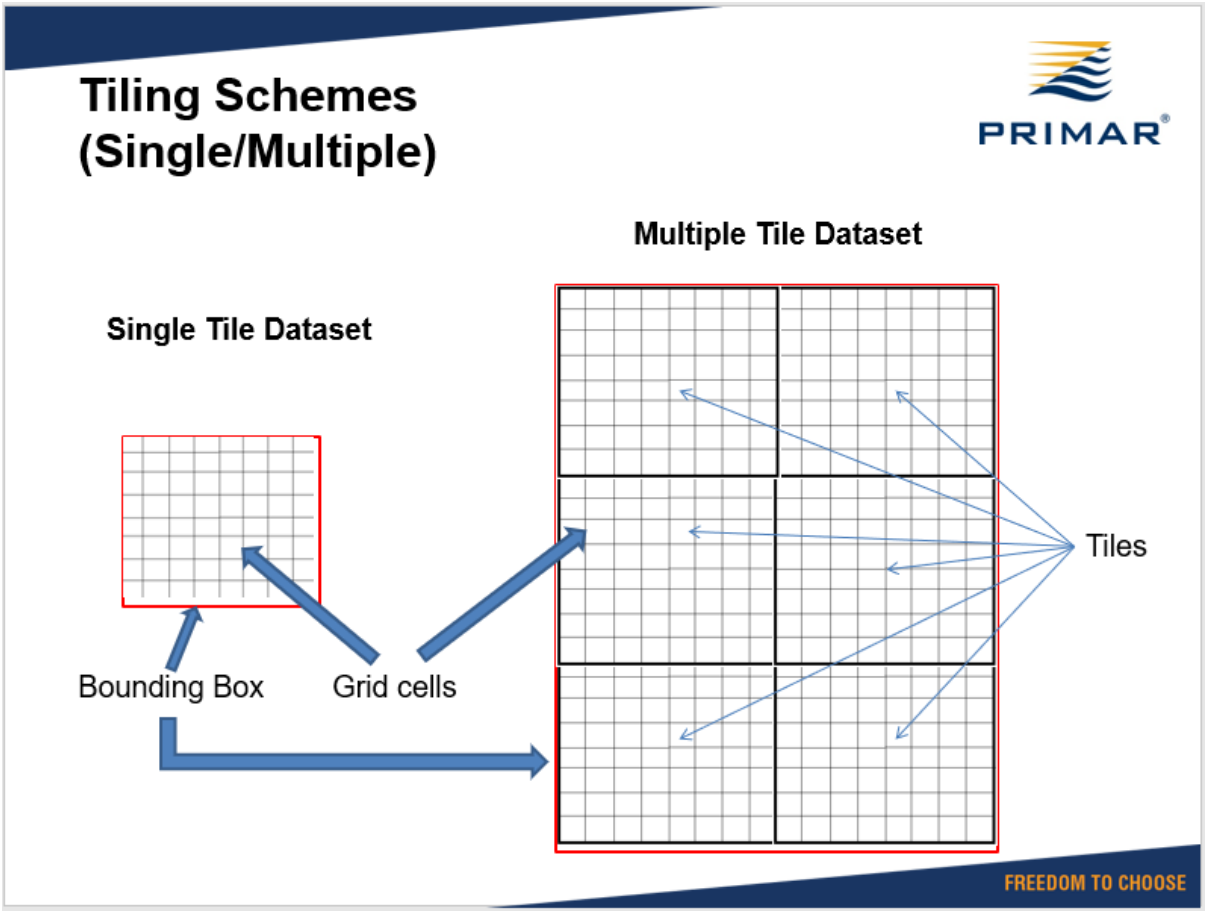
Such recommendations and discussions related to how they should be described has been decided postponed to a future edition of the standard.

11. Validation

Discussions related to data validation has revealed that this part of the standard has been decided postponed to a future edition of the standard. The issue of validation is closely connected to planned work creating a common validation test section applicable to all S-100 products. The idea is that all products will consist of two validation sections, one covering generic S-100 tests and the other being product specific.

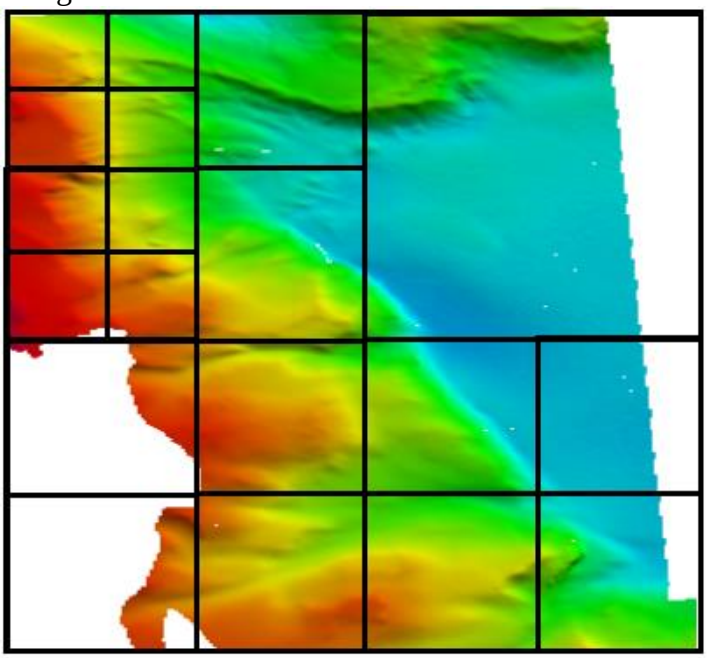
12. Single vs multiple tile data products

Discussions related to single vs multiple tile datasets have been decided postponed to a future edition of the standard. The idea is to have several tiles composing one dataset:



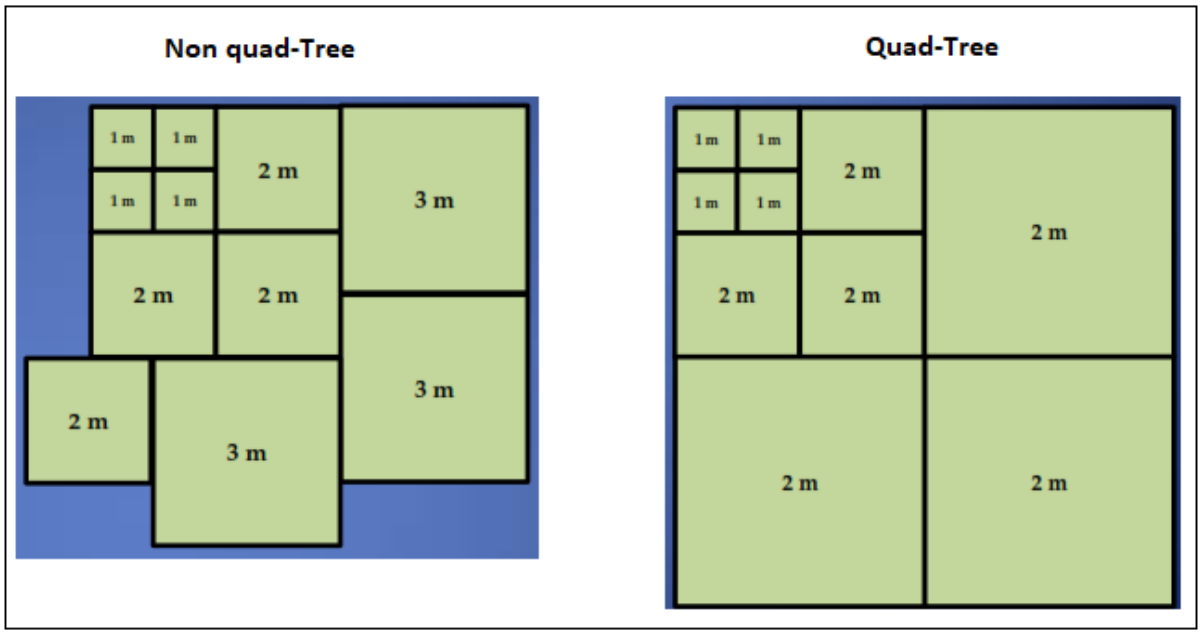
13. Multi resolution data products

Discussions related to multi resolution data products has revealed that this part of the standard has been decided postponed to a future edition of the standard. The idea of multi resolution data products would allow for having areas within a dataset being built of different size grid cells:



14. Grid resolution quad-tree approach

To improve the use of S-102 data when ingesting them in a terrain model a quad-tree approach for resolution was suggested. Such an approach would affect efficient hardware processing time when moving around in a model. An example of quad-tree vs no quad -tree:



A recommendation has been included in the standard where informative grid resolutions related to various display scales are listed:

11.2.2.1 S-102 Grid Resolution and Tiling

Table 11.1 – Informative Grid Resolution and Resulting Tile Size at Chart Scale

Scale	Informative Grid Resolution	Resulting Tile Size @ 10 MB	Resulting Tile Size @ 256 MB
NULL (only allowed on minimum display scale where the maximum display scale = 10,000,000)		Approximate Linear Distance in Nautical Miles (M) for a 600 X 600 node grid	Approximate Linear Distance in Nautical Miles (M) for a 5700 X 5700 node grid
1:10,000,000	900 metres	291 X 291	2770 X 2770
1:3,500,000	900 metres	291 X 291	2770 X 2770
1:1,500,000	450 metres	145 X 145	1385 X 1385
1:700,000	210 metres	68 X 68	646 X 646
1:350,000	105 metres	34 X 34	323 X 323
1:180,000	54 metres	17.5 X 17.5	166 X 166
1:90,000	27 metres	8.7 X 8.7	83 X 83
1:45,000	13 metres	4.2 X 4.2	40 X 40
1:22,000	6 metres	1.9 X 1.9	18.5 X 18.5
1:12,000	3 metres	1.0 X 1.0	9.0 X 9.0
1:8,000	2 metres	0.6 X 0.6	6.0 X 6.0
1:4,000	1 metres	0.3 X 0.3	3.0 X 3.0
1:3,000	1 metres	0.3 X 0.3	3.0 X 3.0
1:2,000	1 metres	0.3 X 0.3	3.0 X 3.0
1:1,000	1 metres	0.3 X 0.3	3.0 X 3.0

This is currently not reflecting a quad-tree approach, and the producer will have to choose resolutions outside of the recommendations to produce accordingly. Further discussions will hopefully result in extended recommendations for the quad-tree approach.