S-100WG6 xx-xx

Paper for Consideration by S-100WG6

Add the possibility to store feature oriented discrete coverage in HDF5

Submitted by:	BSH (Germany)
Executive Summary:	Add the possibility to store feature oriented discrete coverage in Part 10c of S-100 Specification.
Related Documents:	S-100 Specification Edition 4.0.0 ISO 19129 S102PT6(2020) Germany/BSH S-102 extension as proposal to implement source metadata
Related Projects:	S-102 Product Specification

Introduction / Background

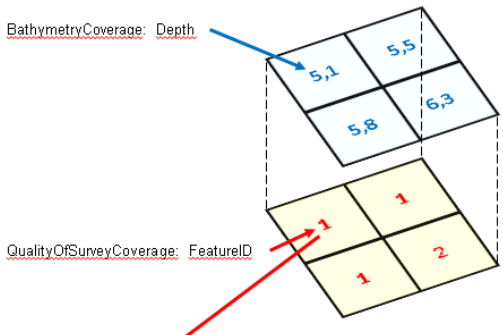
The BSH has undertaken the task to enable the creation of S-102 datasets and has identified an area for enhancement in the S-100 specification to provide more possibilities to transport information.
We consider future S-102 data products as regular tiles that can be inserted into the tiling scheme of SOLAS ENCs. That is, within an S-102 tile, there are typically multiple survey datasets with different metadata properties.
The starting point for our activities is the fact that the S-102 PS currently does not provide for information to describe the quality (e.g. age) or origin (e.g. authority) of the survey data. For example, the survey date, data provider or measurement method cannot currently be specified in a S-102 file.
If we assume that we want to use S-102 data for navigation in the future, and that is the current focus of PS, we need to ensure consistency with the S-101 base file. Moreover, we received a note and needs from our local pilots with needed information for a safe navigation. That means, the features (sounding, depth contour, depth area) and metadata in a S-101 ENC as well as the associated measurement and metadata in a S-102 data product must be consistent with each other.

Analysis/Discussion

Part 3 chapter 7.4 and Part 8 chapter 5.3 of the S-100 describes a possible usage of discrete grid coverages. In such a grid coverage, each grid node stores an ID, which is a reference to a feature in an information table. The information table contains one or many features with one or many attributes.
In practice it means for the S-102, that exists a grid with depth information (BathymetryCoverage). This coverage is a DTM of the seabed. The DTM is created using various surveys with different ages, surveying methods and from different data providers. In S-101, the QualityOfSurvey feature class can be used to code this information.
To obtain detailed information about each survey in S-102, we want to add a discrete grid coverage (named QualityOfSurveyCoverage) with feature IDs. The feature ID from the QualityOfSurveyCoverage grid is linked to a HDF5 CompoundDataset (like a Table) with the attribute values for the feature. As a result, the information is assigned unambiguously to each grid cell and stored non-redundantly in the HDF5 file. The figure below illustrate the previously described technique in a graphical form.

Draft

Note: FOR REASONS OF ECONOMY, DELEGATES ARE KINDLY REQUESTED TO BRING THEIR OWN COPIES OF THE DOCUMENTS TO THE MEETING



Feature Attributes

ID	Data provider	Survey date	Measure method	...
1	BSH	2021-02-16	Multi-beam	
2	WSA HH	2018-07-23	Single-beam	
...				

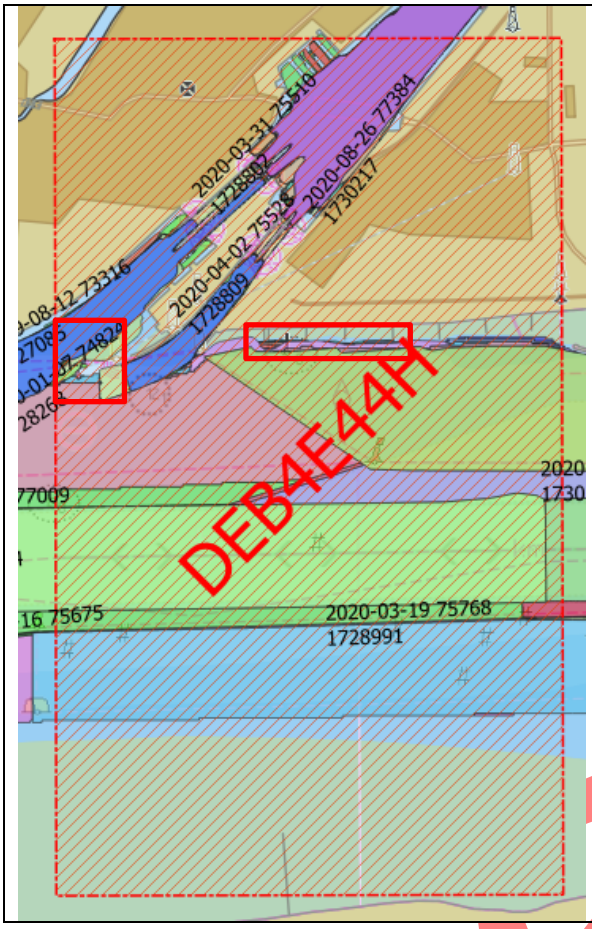
The S-100 also describes a concept for storing object metadata in the form of a vector geometry and GML (cf. 10c-9.2.3). In the next paragraphs, we will compare the differences and show the advantages and disadvantages.

1. Size of data
   A GML is a text orientated file format. It consists of many redundant and describing format elements. This results in a bloated text file. The solution of a discrete grid coverage is more direct. Besides, it can be compressed in HDF5. The following table shows the additional memory required for a specific example and compares the two options "Discrete Grid Coverage" within a HDF5 file versus "separate GML file" that is included in the same exchange set.

Data element	**S-102 v2.0.0**	**S-102 v2.x.x + Discrete Grid Coverage**	**S-102 v2.x.x + GML**
BathymetryCoverage	3 060 kByte	3 060 kByte	3 060 kByte
QualityOfSurveyCoverage (Discrete Grid & CompoundDataset)		65 kByte	
QualityOfSurvey (GML)			547 kByte
QualityOfSurvey (GML Zip compressed)			156 kByte
Size	**100 %**	**102,1 %**	**117,9 % / 105,1 %**

2. Consistency of data
   A GML metadata file with vector geometry would be stored outside the actual HDF5 dataset. An assignment of the files is only possible via a naming convention. Replacing the HDF5 dataset in the case of a new edition should then also always result in a change to the metadata file. We see this as a possible error scenario, because it must be technically ensured that both files are updated because there is no direct link between them. Thus, it could be the case that the metadata no longer matches the actual dataset. Renaming and deleting datasets could also lead to incorrect consistency between the data. On the other hand, the data of the discrete grid coverage are completely contained in the HFD5 file. Therefore, all data are always consistent with each other.
3. Geometry issues
   A HDF5 datasets can consist of several individual surveys that are combined into a digital terrain model (DTM). An HDF5 dataset is oriented in its extent to the SOLAS compliant ENC cells.
   The individual surveys of the DTM overlap in extent and are not measured at the same time. Therefore, new measurements superpose older ones.
   Because of the superposition's, very complex geometries of the respective surveys are created. As a survey ages, more and more areas are superposed by newer surveys. Over time, this results in very complex geometries with many enclaves and exclaves. The description of these geometries is extensive in GML, thus creating a large demand for storage. With a discrete grid coverage, the memory requirement is not as dependent on the complexity of the geometries and is much more scalable. Furthermore, the processing of these complex vector geometries is technically more demanding than with grids. An example of the complex geometries is shown in the following figures.

Draft

	Overview HDF5 dataset with Detail 1 and Detail 2

Draft

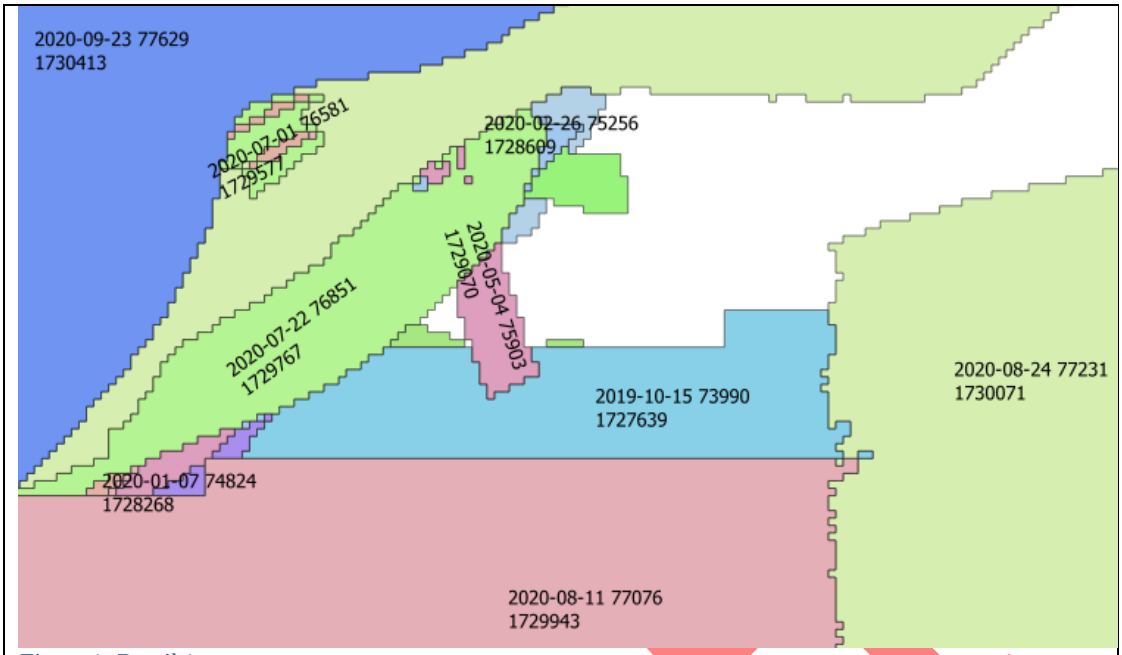


Figure 1: Detail 1

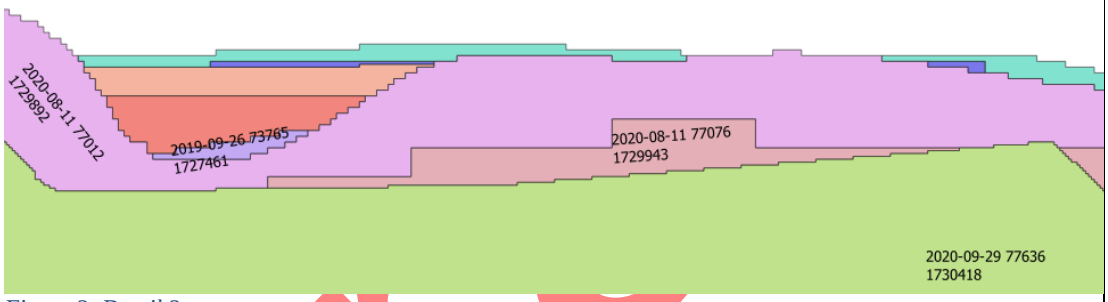
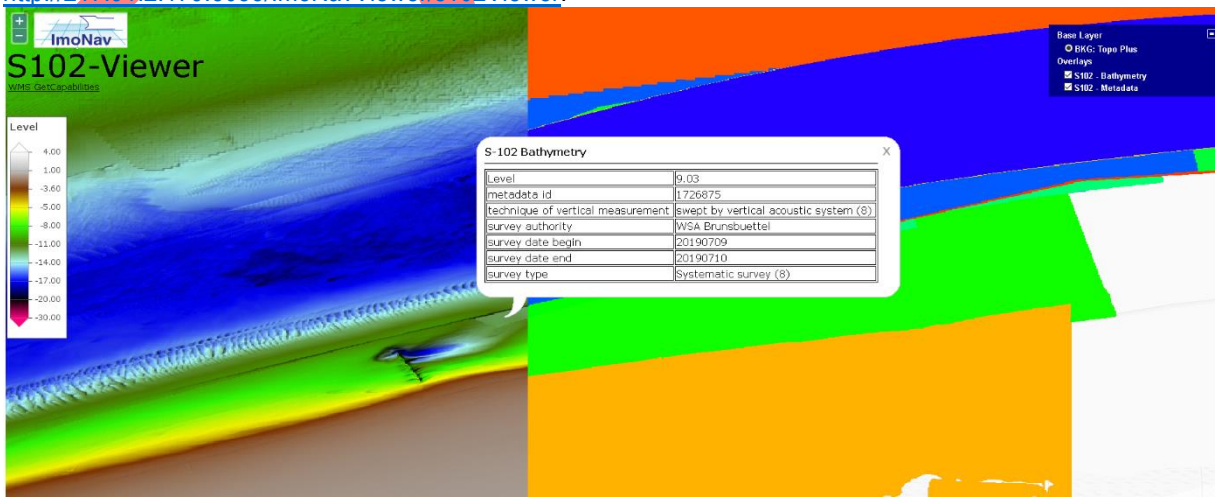


Figure 2: Detail 2

An example of a grid with bathymetric data and a grid with metadata is shown in the figure below. The viewer enables the user to get a report for each grid cell. It is possible to access the demo server from this website http://217.91.2.179:8085/ImoNavViewer/s102Viewer.



Conclusions

1. Differentiated encoding of metadata is necessary for the description of different surveying datasets within a HDF5 file.
2. The solution of a discrete grid coverage allows us to unambiguously assign metadata to each grid cell with a minimum of additional memory requirement.

Note: FOR REASONS OF ECONOMY, DELEGATES ARE KINDLY REQUESTED TO BRING THEIR OWN COPIES OF THE DOCUMENTS TO THE MEETING

3. A HDF5 dataset and an external file with additional information are not necessarily always consistent with each other. Storing the data in a single file is therefore safer and preferable as a solution.
4. Handling complex geometries of the additional data is more efficient, scalable and easier in grids.
5. To fulfil the customer requirements it is necessary to transport additional information in a proper way. It seems that the vector geometry GML is not the optimal way to transport this kind of information.

Recommendations

We recommend adjusting the S-100 specification in order to assign differentiated metadata in a HDF5 file. We would like to encourage further discussion about the necessary metadata in a S-102 data product in order to achieve alignment especially regarding the data quality between a S-102 file and a S-101 base cell.

Justification and Impacts

With this enhancement of the S-100 we are able to create consistency between S-102 and S-101.Furthermore the enhancement enable the HO's to fulfill customer requirements to additional data.

Action Required of [HSSC][Relevant HSSC WG]

The [HSSC] [Relevant HSSC WG] is invited to:

a. endorse …….
b. agree ……
c. note ……...

Draft