TECHNICAL REPORT

Validation of S-100 datasets.

Release: v1.0.0 (Internal Draft 7.2.0)
Date: 18th June 2022
Submitted to: JPA Project

IIC Technologies Ltd
Part of the IIC Technologies Group
The Catalyst
York Science Park
Baird Lane
York
YO10 5GA

E-mail jonathan.pritchard@iictechnologies.com

Website www.iictechnologies.com

1 Introduction

This report covers the initial formulation of validation tests in relation to a number of S-100 product specifications. As S-100 is rolled out to stakeholders, and S-100 ECDIS is developed, its key feature, multiple product specifications, will have a profound effect on the end user's experience of the ECDIS. This report looks at a number of areas in relation to validation which have been researched using live data and existing S-100 catalogues and product specifications.

The aim is to propose validation tests including those between multiple product specifications, aimed at increasing the harmonisation of datasets and user experience of them for live navigation on the S-100 ECDIS.

2 Background

Validation tests are an integral component of IHO standards with a long history of active participation by member states and development by standards bodies and stakeholders. Validation tests currently cover S-57 ENCs well and have been through many sets of revisions. These tests cover semantic, formatting and structural aspects of compiled S-57 ENC data. The vast majority of these tests can be carried out within the cell itself, without reference to other datasets or external data.

Tests for horizontal and vertical consistency also exist, although few are formally part of S-58. These tests cover edge matching between cells and the consistency of feature representation and geometry between different scales. They are, therefore "inter-cell" validation tests, carried out on multiple cells of the same type – this idea of validation between cells becomes more key with S-100 as data used for portrayal to the end user is frequently split across multiple datasets.

The scope dealt with in this report is S-100 ECDIS functionality. S-100 ECDIS has progressed in recent history substantially, with drafts of key IMO documents which will enable its development. The ECDIS Governance Document is referenced as well. This draws together many of the threads of the discussions around S-100 ECDIS and the topic of validation in the context of ECDIS is one of key importance for the development of data production flowlines by producing agencies.

Many of the tests here are generic and could be applied to many datasets but many are also specific to the implementation of S-100 Water Level Adjustment and User defined safety contour in S-98 Annex C. As such, these tests may change as S-98 is refreshed and updated (from testing under v.1.0.0 to v2.0.0). For example, the current edition of Annex C only includes rectangular gridded data so part of validation is ensuring that only these datasets are distributed to S-100 ECDIS. Currently no exhaustive list of the elements of S-100 required for ECDIS implementation exists (as noted in the ECDIS Governance Document), so such subjects are not completely clear for manufacturers (test data will help by showing those elements which definitely are required but this is not a long term, definitive solution to the problem).

S-98 Annex C is a core component of S-100 ECDIS functionality and is a bespoke process implemented by ECDIS manufacturers which gives the user a user selectable safety contour and adjusted depth values in ENC according to the available coverage of S-101, S-104 and S-102 (shown in the diagram below). The chapter dealing with this in S-98 Annex C is a new one and will require extensive testing to ensure it meets user requirements and is safe.

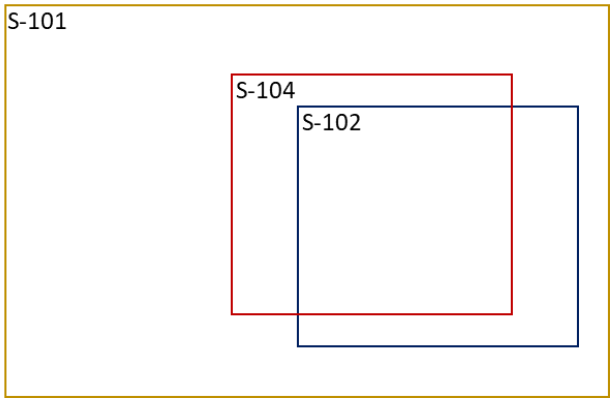


Figure 1: Selective coverage of S-100 products

The tests described in this report follow a structured methodology which scales across datasets, and their interaction together within the S-100 ECDIS. The report is, therefore, a combination of a method for extending validation testing to multiple cells of the same type, and also to multiple datasets conforming to different IHO S-100 product specifications. More details on the methodology is contained in the next section.

3 Methodology

3.1 Introduction.

This section presents the overall methodology used. Subsequent section implement the methodology in a series of defined tests against defined S-100 product specifications.

As discussed in the background section validation of other S-100 product specifications requires a top-down refresh of what we exactly mean by validation. Validation of ENCs is currently predicated on a single-product universe where ENC at different scales is the only product available to the mariner, e.g.

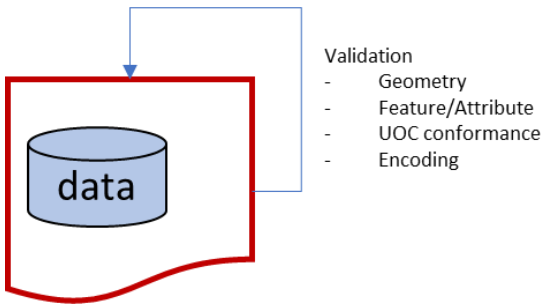


Figure 2: Single (S-57) Product Validation

Much of the existing S-58 validation is currently focused on cells and the encoding, relationships and geometry within the cell itself, as illustrated in the example above.

Additional validation tests exist in order to validate cells against their neighbours (horizontally) or vertically, against different scales covering the same area. These validation processes are illustrated in the following diagram. These validation tests include horizontal edge matching and cartographic generalisation between scales.

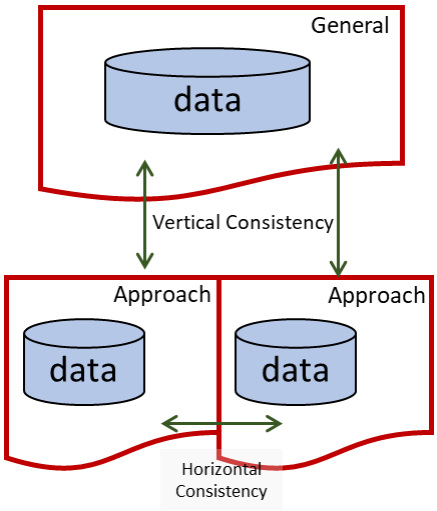
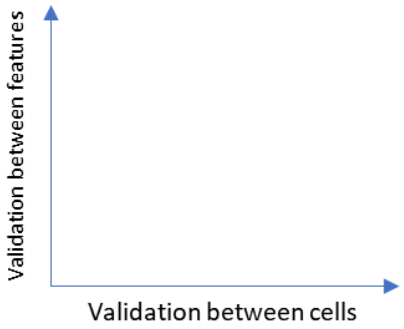


Figure 3: Validation between different cells of the same product type

ENC validation therefore consists of the following two discrete activities across two axes.

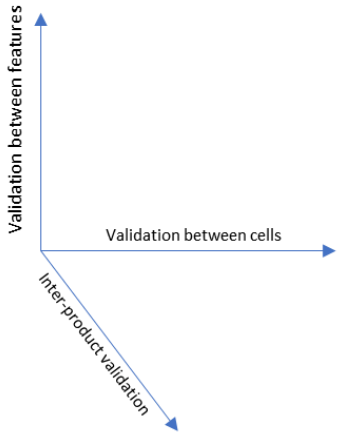
1. Intra-Cell validation. Validation tests to ensure a cell validates against itself.
2. Inter-Cell validation. Tests to ensure the cell validates against those it spatially intersects, touches or joins with.

This is illustrated by the coordinate axes below:



These tests, of course are based only on a single data product. S-100, of course, is predicated on the existence of multiple datasets, product specifications representing a variety of marine phenomena. Therefore the set of possible validation tests between product specifications gives another axis on which the methodology is based.

So, in order to construct a set of validation tests which are meaningful, datasets themselves must be consistent, correct and valid. Then validation between cells of the same "type" and finally validation between cells of different types.



In practice the validation between cells of the same type is reasonably simple (replacing the existing horizontal/vertical consistency checks) with most of the complexity being added into the validation tests focused on multiple product specifications. This complexity can be managed, however (helped by the simplicity of the underlying algorithms in the S-100 ECDIS).

Under this model the scale dependent "ENC stack" is validated not only against itself, but against the other products that might, potentially, be replicating/enhancing or representing similar physical phenomena. The S-100 ECDIS Governance Document defines three categories of data for use in the ECDIS:

1. Chart data (S-101 / S-401)
2. Chart-related data
3. Nautical Publications

The main validation between product specifications required is between (2) and (1) and between (3) and (1). This document focuses on validation of products in (2) and between (2) and (1).

3.2 Validation Description.

The reason for visualising validation like this is to form a "framework" view of what validation actually is. This is important because validation is designed to increase data quality for the end user and reduce the likelihood of ECDIS anomalies – in the S-100 ECDIS context the presence of multiple layers on the ECDIS increases the complexity of this task.

In this context a simple definition of a validation test is:

A validation test is a set of rigorously defined conditions across one or more data inputs which yields a single boolean result together with any textual observations.

In programming terms a validation test can be seen as a Boolean function returning a single true/false value along with any noted observations.

[Boolean,String] : isValid($dataset_1$, $dataset_2$ …. $dataset_N$)

the operation of the isValid function, which map the dataset(s) inputs to Boolean True/False is the validation test. The String returned is any comment/observation is returns.

Validation tests, traditionally are described using the following description form:

Identifier	Description	Message	Solution	Conformance class	Classification

This is based on the description within S-58 (example shown).

The validation checks are laid out as follows;

3.4 Checks relating to Use of the Object Catalogue for ENC					
No.	**Check description**	**Check message**	**Check solution**	**Conformity to:**	**Cat**
1500a	For each CBLARE feature object which is WITHIN OR OVERLAPS a LNDARE feature object of geometric primitive area.	CBLARE object overlaps a LNDARE object.	Amend objects to remove overlap.	Logical consistency	W
1500b	For each SBDARE feature object which is WITHIN OR CROSSES a LNDARE feature object of geometric primitive area.	SBDARE object is within or crosses a LNDARE object.	Amend objects to remove overlap.	Logical consistency	W
~~1501~~	*Check removed.*				

Our input values are the datasets of data within a defined spatial region and may be of any S-100 conformant product specification. Consideration of other S-100 products, increases the input datasets in respect of the three axes described in the diagram.

The headings in the candidate tests within this report are slightly different – taking into account their category (whether they are single product tests or inter-product tests) and the product(s) they relate to. The conformity heading is broadly the same as the category of test identified and unique identifiers have been left for later definition. The Check Description/Message and solution will require further definition once the initial list is developed but the description in the table is designed to fully define the conditions the test evaluates, in most cases the remedy is fairly simple.

3.3 Categorisation of tests

We should probably further recognise that there are two broad kinds of tests within the S-100 ECDIS scope.

1. Tests of "syntax", correctness according to format/encoding and data structure – this covers:
    a. Conformance with Feature Catalogue content. All datasets are associated with a single numbered
    b. Conformance with the encoding format standard (ISO8211, GML HDF5) and any standardised fields or data structures required by the S-100 at a framework level.
    c. Conformance of metadata / correctness of metadata
2. Further tests of which look at the data content itself either on its own (in relation to other features in the dataset), in relation to its defining standard, or in relation to other dataset(s). For example:
    a. Requirement for a coherent grid structure
    b. A complete grid structure (existence of all points)
    c. Validity of data points
    d. Data points values within expected ranges

3.4 Scope and Limitations of the initial tests specified.

There is an unanswered question at the outset of the project, how do we define a "valid" dataset at all, given that S-100 is intended for use in multiple situations and by multiple end users. Like S-58 the following criteria are suggested for how "ideal" tests should be structured.

Our scope considered here is ECDIS (as stated in the introduction). Some validation tests may well be non-ECDIS specific and relate to the encoding/format (and actually could be extended to other

product specs which also use HDF5) but most are specific to the operation of the S-100 ECDIS in relation to the data loaded and in relation to the ECDIS operation under IMO conditions of operations.

One important element not considered is whether validation should take into account contents of the interoperability catalogue. Because the catalogue potentially suppresses features, it makes the ECDIS functionality partly dependent on the DCEG encoding of the individual datasets. Should there be validation tests to make sure datasets are "interoperability-proof"? this should be the focus of future work in this area as potentially data producers' data compilation in respect of interoperability will require validation – this report assumes suppression of S-101 features by the corresponding S-10X data as per the current interoperability catalogue descriptions.

3.5 Document Sources

The following documents have been used in the preparation of this report.

1. Product Specifications
    - S-102
    - S-104
    - S-111
    - S-101, S-101 Feature Catalogue and S-101 DCEG
2. S-98 Annex C v.1.0.0
3. WLA and USSC Sections of S-98 as well as contributing documents to their development

3.6 S-98 Annex C functionality.

A key driver in the operation of the S-100 ECDIS is the execution of the User Selected Safety Contour and Water Level Adjustment algorithm. The inter-product validation tests in the initial register (next section) are partly designed to ensure this functionality (in S-98 Annex C) is able to be executed without presenting the end user with an unsafe or incorrect picture.

The relevant S-98 Annex C sections contain specifications for OEMs detailing how the functions are to be developed. There is a consequent need to develop validation tests to ensure data is able to be used. The functions are based on areas of coverage where S-102 and S-104 overlap with S-101.

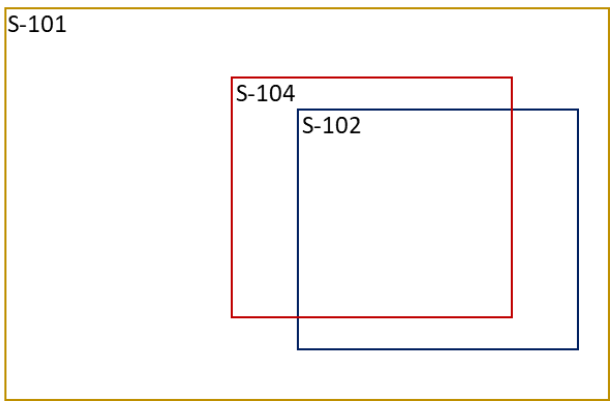


Figure 4: S-102, S-104 and S-101 coverage

The algorithms proposed in s-98 Annex C provide the required interoperability between S-102/S-104 and S-101 and can be executed across the whole visual display at a single instant in time, or projected across a route using S-104 values to adjust water level as appropriate. This process is fairly complex for the OEM to implement, although based on simple operations and avoids complex interpolation. It is also always biased to "shoalest" values when a number of values are in close proximity. The complexity is illustrated by the following diagram, taken from the S-98 Annex C draft.

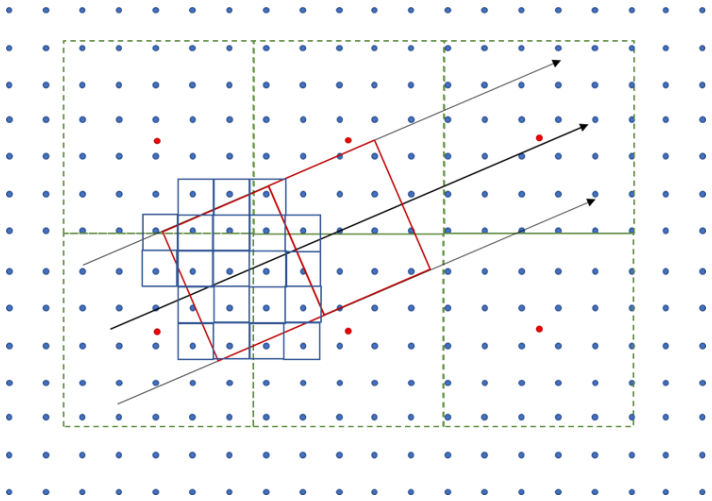


Figure 9: Adjustment of S-102 values by S-104

For these reasons of complexity, the validation tests of S-104/S-104 against S-102 are generally labelled as "Critical". Warnings are generally validation tests which might suppress the S-100 ECDIS "Added Value" functionality, only giving rise to confusion rather than error on the S-100 ECDIS.

4 Dual Fuel and Governance

There are several aspects of the validation tests created which reflect conclusions in the recent Dual Fuel and S-100 ECDIS Governance document. Above all, it highlights the need for an integrated approach to the development of S-100 ECDIS and its "operating model". The operation of the S-100 ECDIS is considerably more complex than the S-57 system defined in the current IMO Performance Standard. In order to be certain that the ECDIS will display data safely to users the interoperability algorithms are designed to be simple to implement.

The complexities of the inter-product validation tests highlights how important it is for these aspects of ECDIS operation to be easily understandable for both implementers and data producers, and for data producers to fully understand the impacts of release of auxiliary S-10X data to their end users.

The Governance document highlighted gaps in support to the implementing community and OEMs and noted the substantial requirements for test data to support S-164 (for formal type approval testing) and OEM implementation. Added to that should be a requirement for representative test data for validation test development and implementation.

The portrayal of S-102, S-104 and S-111 under the combination of the interoperability framework, S-98 Annex C and dual fuel mode has yet to be tested fully. Some of the tests here make reference to the S-10X products and the underlying "largest scale ENC". Similarly dual fuel portrayal requires an equivalence of CSCL/M_CSCL and maximumDisplayScale attributes – these are still under

development and it is possible these validation tests will need review and refinement as the S-100 ECDIS loading strategy is finalised and documented.

The Governance document also noted the question of the impact on data producers of the issuing of multiple data products and the importance of ensuring they are (a) up to date at all times and (b) self-consistent with each other. The validation tests specified make these requirements explicit and should be seen as a baseline for designing data production/distribution systems for multiple products. It will also make it clear the difference between the use of S-10X products for data exchange between institutions/organisations and the distribution of such datasets for use in primary navigation.

5 Initial register of tests.

This is an initial register of tests compiled using individual product specifications, and mapping between features in different product specifications. The column headings are as follows:

1. Category. The broad area of functionality the test refers to
2. Description. Description of the conditions the test is attempting to capture.
3. Inputs. Whether tests are designed to be run on single datasets or between datasets of different type/product.
4. Products. Which products the test is focused on
5. Classification “C” – critical, “W” – warning. A draft classification based on possible user/ECDIS impacts.

Category	Description	Inputs	Products	Classification
Dataset Coverage and Datums				
	Spatial limits must match CATALOG.XML	**Dataset / CATALOG.XML**	**All**	**C**
	Temporal limits must match CATALOG.xml	**Dataset / CATALOG.XML**	**All**	**C**
	Temporal limits start must be before temporal limits end	**CATALOG.XML**	**All**	**C**
	Spatial limits must be well defined polygon(s)	**CATALOG.XML**	**All**	**C**
	Horizontal Datum must be WGS84. horizontalDatumValue = 4326 and horizontalDatumReference = “EPSG”	**Dataset**	**All**	**C**
Data Values	**Depth must be negative up**	**Dataset**	**S-102**	**C**
	All depth values must be Real numbers	**Datset**	**S-102**	**C**
	All Water Level values must be real numbers	**Dataset**	**S-104**	**C**
	All Current speed numbers for Current Speed shall be real numbers. All trend indicators shall be integers, 1, 2 or 3	**Dataset**	**S-104**	**C**
	-90.0°<gridOriginLatitude<90.0° and -180.0°<gridOriginLongitude<180.0°	**Dataset**	**All**	**C**

Category	Description	Inputs	Products	Classification
	Trend should be in the correct direction (i.e. values following trend up should be >= to values and trend down should be <= to values). Trend values must validate against the following formulas. 1. Decreasing: trend <= -*waterLevelTrendThreshold* 2. Increasing: trend >= +*waterLevelTrendThreshold* 3. Steady: -*waterLevelTrendThreshold* < trend < +*waterLevelTrendThreshold*	**Dataset**	**S-104**	**W**
	typeOfWaterLevelData=2,3 or 5. Only predictions (astronomical, hybrid or hydrodynamic forecasts are suitable/useful for ECDIS). Others will be ignored			**W**
	Current direction must be >=0.0 and <360.0	**Dataset**	**S-104**	**C**
	Reserved names for groups. Must be Group_NNN	**Dataset**	**All**	**C**
	All data points must be valid values within ranges in dataset metadata. **maxDatasetCurrentSpeed shall not be exceeded.** **No data values < minDatasetCurrentSpeed**	**Dataset**	**S-111**	**W**
	Number of null speed values must be equal to number of null values in the direction grid.	**Dataset**	**S-111**	**C**
Coverage	**Dataset coverage must not overlap (temporal and spatial)**	**>1 Dataset , CATALOG.XML, S-128 (service)**	**All**	**C**
Grid Structure				
	Well defined grid	**Dataset**	**All**	**C**
	All grid dimensions > 0	**Dataset**	**All**	**C**

Category	Description	Inputs	Products	Classification
	Grid Completeness. All grid points must have a valid value or noData value	**Dataset**	**All**	**C**
	Grid Consistency. Correct number of values per row/column. Number of values in data must equal numPointsLongitudinal and numPointsLatitudinal from dataset metadata.			**C**
	Calculated grid from gridSpacing, numRows and numCols must not be outside defined coverage in metadata. *Longitude Limit = GridOriginLongitude + (NumCol-1) * (gridSpacingLongitudinal).* [from S-111 Eqn. 4.1] *Latitude Limit = GridOriginLatitude + (NumRow-1)(gridSpacingLatitudinal).* [from S-111 Eqn 4.2]	**Dataset**	**All**	**C**
	Calculated grid coverage shall not cross +/-180.0° longitude nor exceed +/- 90.0° latitude	**Dataset**	**S-104 / S-111**	**C**
	Regular Grid only is allowed for S-100 ECDIS (currently). S100_GridCoverage type (b) (Table 4.1). This is dataCodingFormat=2 in the dataset metadata.	**Dataset**	**S-102, S-104**	**C**
	interpolationType shall always be nearestNeighbour (see notes)	**Dataset**	**S-102/S-104**	**W**
	timepoint shall always be after issueDate	**Dataset**	**S-104,**	
Resolution				
	Maximum Resolution (See notes)	**Dataset vs Dataset**	**(S-102/S-104) and S-101**	**C**
	Minimum Resolution (See notes)	**Dataset vs Dataset**	**(S-102/S-104) and S-101**	**C**
Dataset Metadata				
	(Unmatched fields between CATALOG.XML and Dataset)			

Category	Description	Inputs	Products	Classification
	Issue Date and issueTime in future in CATALOG.XML. This should match issueDate and issueTime in dataset header metadata.	Dataset	All	C
	Invalid feature code (10c-8)	Dataset	All	C
	uom.name or name doesn't match feature catalogue.	Dataset	All	C
	Invalid HDF5 datatype	Dataset	All	C
	Non strings in the feature description	Dataset	All	C
	Not all feature codes described.	Dataset	All	C
	Missing feature descriptions	Dataset	All	C
	Negative time interval	Dataset	All	C
	Times outside time of earliest values and time of latest value	Dataset	All	C
	numGrp shall be equal to the number of Groups in the HDF file.	Dataset	All	C
Cross Validation				
	Vertical Datum/Sounding Datum must be the same in areas of overlapping coverage or WLA/USSC can not be computed.	Dataset vs Dataset	S-101, S-102, S-104	C
	S-102 depth values (excluding noData, defined by bounding rectangle) shall only spatially intersect the following Group 1 features (S-101): 1. Depth Area 2. Dredged Area	Dataset vs Dataset	S-102/S-101	C
	S-102 values within defining bounding box should not be shoaler than 1. DepthArea.DepthRangeMinimumValue (any feature as specified by feature catalogue) 2. Sounding depth value (ZCOO/COMFZ)	Dataset vs Dataset	S-102/S-101	C

Category	Description	Inputs	Products	Classification
	3. ValueOfSounding (any feature with valueOfSounding in S-101 feature catalogue)			
	Where polygons with depth attribution (S-101) are not enclosed by S-104 coverage WLA can not be calculated. See S-98 Annex C	Dataset vs Dataset	S-101 / S-104	W
	Where polygons with depth attribution (S-101) are not enclosed by S-102 coverage, user selected safety contour can not be calculated.	Dataset vs Dataset	S-101 / S-102	W
	Values in S-111 for surfaceCurrentSpeed shall not exceed the values in the spatially intersecting S-101 data for Tidal Stream flood/ebb – speedMaximum (where this value is set/non-null)	Dataset vs Dataset	S-101/S-111	W
	Depth of Current (surfaceCurrentDepth in dataset Metadata, relative to depthTypeIndex=seaSurface) must be valid. Must be within S-102/S-104, S-101 or S-101 WLA value.	Dataset vs Dataset	S-101/S-111	W
	When Water Level (S-104) is zero, for all inter-tidal locations (S-101), the current speed and direction must be zero valued (or noData) (S-111)	Dataset vs Dataset	S-101/S-104/S-111	W

6 Validation Test Notes:

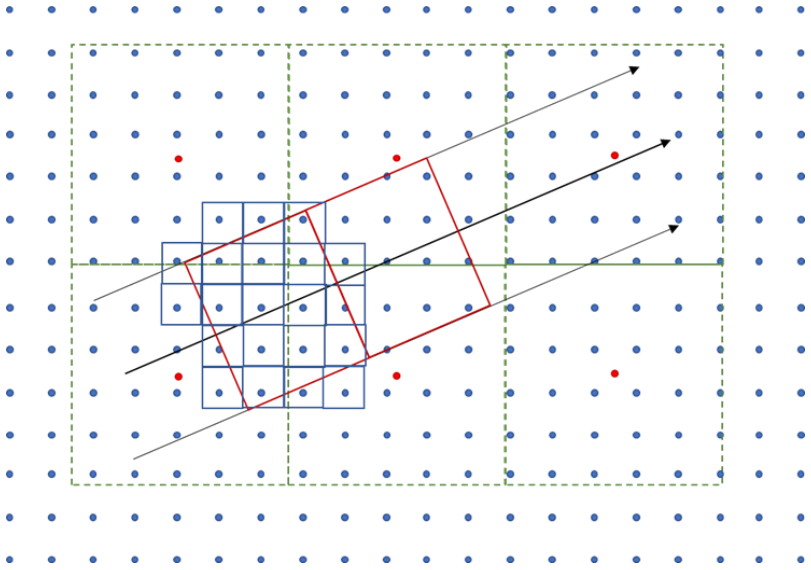
6.1 S-102 / S-104 Maximum/Minimum Resolution

The issue of whether a minimum or maximum resolution should be imposed via validation should be discussed with the S-100WG as a whole as it has a bearing on S-100 ECDIS implementers, as well as data producers. In terms of minimum resolution it is likely that unless data is of a certain minimum threshold resolution (in relation to the largest scale ENC coverage corresponding to the dataset, then both User Selected Safety Contour and Water Level Adjustment will be unusable (and, in all cases, worse than the corresponding ENC). Although the user is free to not use S-102/S-104 in these cases it would seem appropriate to put a minimum resolution in place to ensure data is fit for purpose, given the extensive use likely to be made of this feature.

Similarly, a maximum resolution is probably appropriate, in order to provide ECDIS implementers with a realistic estimate of maximum data volumes/density expected. A measure of maximum density may be found either in the average pixel size onscreen or via the old 0.3mm rule (which determines the minimum distance between vertices under S-57 ENC). Assuming the same 0.3mm, as measured at MaximumDisplayScale a maximum resolution could be recommended. The logic of this is that at higher resolution this requires the OEM to resolve the display of multiple depth values with the same (displayed) physical location to the user, as well as their integration into the user selected safety contour and water level adjustment sections of S-98 Annex C. A consultation with data producers and OEMs is probably the best way to define any validation tests over minimum/maximum resolution for S-102 (and possibly S-104) within S-100 ECDIS and the results used to refine the validation tests noted in this report.

6.2 Regular grid only

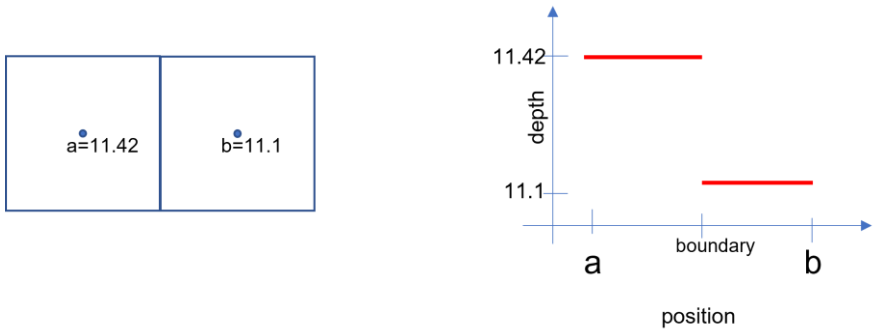
The S-102, S-111 and S-104 product specifications have a broad array of grid formats, as set out in S-100 and are flexible on interpolation between different grid points. However, for simplicity of implementation by OEMs, when S-98 Annex C was initially drafted, only regular grids were considered and these have been assumed within that part of S-98 Annex C. Therefore, as the main use case for S-102 and S-111 is these sections of S-98 (fulfilling the requirement for interoperability on the S-100 ECDIS) a validation test has been included which restricts grid arrangements to regular grids and S-104 values to predicted forecasts only. The regular grid is used to define areas of extent for individual values in the grids concerned, both for delimiting safe/unsafe water (for user selected safety contour) and proximity to route / depth values for S-104 as part of the Water Level Adjustment (WLA) sections, shown below.



WLA along a route segment (from S-98 Annex C)

As S-98 is tested and refined it may be possible to expand the type of grids that S-100 ECDIS is capable of processing.

The method of interpolation used by S-100 ECDIS is also fixed as nearest neighbour. Although datasets may well use other interpolation types (as defined by S100_CV_InterpolationMethod) the S-100 ECDIS will always use a nearest neighbour interpolation, as illustrated below.



a.

Figure 3: Extents of each S-102 points showing nearest neighbour interpolation (S-100 Part 8-7.1.4)

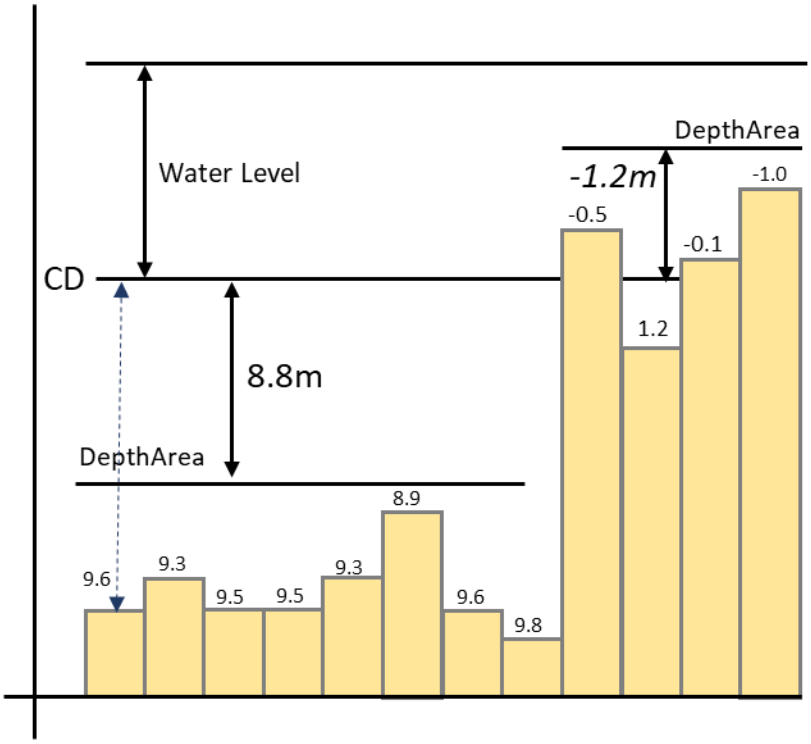
Although this is not necessarily an issue for S-100 ECDIS calculations under S-98 Annex C, it is something data producers should be aware of and hence a validation Warning is probably appropriate. The differences in values for different interpolation methods may be minimal (within reasonable bounds of dataset resolution) but the data producer should be aware of such differences before issuing data for use by end users to avoid any mis-apprehensions.

6.3 Cross-Validation with S-101

A number of tests have been proposed which cross-reference content between S-102/S-104 and S-111 with the underlying ENC (S-101). These are new proposals which are likely to require some refinement and should be seen as extensions to existing horizontal/vertical consistency tests done on adjoining/overlapping ENCs.

The S-98 Annex C algorithms will require testing by implementers, particularly water level adjustment of features which have depth "attribution" (e.g. depthRangeMinimumValue and valueOfSounding). The test proposed here are designed to validate data which is focused on the S-98 Annex C algorithms and so, as they mature and are updated, the validation tests may require adjustment.

The following diagram shows the reasoning behind the validation test of ENC compiled depths (e.g. DepthAreas) and S-102 point depths. This is a profile view of ENC depths, as Depth Areas, and drying heights.



The depths of the S-102 are shown as individual measurements at particular points (in practice the S-102 depths' sign may be negative or positive, determined by dataset metadata – in the diagram the S-102 values are represented in the same way as the ENC data (positive downwards)). When the ENC is compiled it is shoal-biased and depth values are rounded to the shoalest value (and number of decimal places) of the S-102 depths within the region – in the diagram this is the numerically smallest value (often with a tolerance factor).

Similarly drying heights are similarly rounded to the highest values above CD (so that they are shoal-biased when any water level is applied to them).

So, from a validation perspective the S-102 values should be “deeper” than the ENC depth areas (and other depth values contained in the cell). Conversely, when considering drying heights the S-102 data should be less (numerically) than the values in the intersecting ENC data. Rounding of S-102 data for portrayal has not been addressed currently (and the S-98 Annex C algorithms assume no rounding, comparing S-102 values at the their full precision with the S-101 values.

The situation with S-111 and S-104 in respect of S-101 is more complex. With S-102 cross-validation there is no temporal aspect to be taken into consideration but S-104 and S-111 both contain values which vary with time. However, a similar approach can be taken with S-104 and S-111 as there are still tide/current features contained within S-101 ENCs and producers who do not produce S-104/S-11 will rely on these values to publish summary data. So, in a similar way to S-102/S-101 validation it is a reasonable assumption that the values in the ENC should be those which are “safest”. Maximum values for tidal stream for instance should be a true “maximum” in the S-101 ENC with the values in S-104 not exceeding them (this is the proposed validation test). The more complex Tidal Stream Panel Data S-101 feature could, in theory be reconciled with S-111 data but this would require the alignment of S-111 against actual tidal values for true validation to take place. This would require the validator to have access to tidal prediction (HW/LW times) to match up the two datasets – this is, in theory possible, but beyond the scope of these initial validation tests. A basic test against maximum values is a start in cross-product validation and could be further examined by the relevant IHO working groups.

However, it should be borne in mind that S-111 is not used within S-98 Annex C and is an extra layer on the ENC. As it may (subject to S-98 interoperability) suppress tidal data it should be validated to ensure the user picture is as intended by the data producer.

The correction of validation errors is likely to be down to the data producer, either to adjust the ENC in respect of the S-10X product or vice-versa. As was noted in the ECDIS Governance Document this will focus data producers on the currency of data being issued and will place a requirement to ensure data which is being made available for primary navigation is up to date and self-consistent.

6.4 Parsing of HDF5 data

One element missing from the set of validation tests is a basic validator of data “format” – Part 10c reference HDF5. In the context of IHO standards and S-100 a basic validator of a file format would be a step forward. In S-57 the ISO8211 encoding is partly validated by S-58 but no defined level of validation against the ISO8211 standard is defined.

In S-100 the situation is more complex, with encodings for ISO8211 (using concatenated data structures – absent from S-57), XML formats (CATALOG.XML and GML under Part 10b) and HDF5 for Part 10c. XML/GML are reasonably easy to parse and sanity test form a basic format point of view, using simple XML validation. ISO8211 is also fairly simple to construct a format validator for as a number of public domain and sector specific tools built on S-57 parsers could be used. HDF5, similarly, is an open format

standard and tools such as the Unidata netCDF HDF5 / CDM viewer[1] could easily be leveraged to provide a basic level of HDF5 format checking prior to more formalised validation testing. As an SDK exists for HDF5 an automated checker could also easily be put in the public domain to ensure basic sanity tests can be run.

[1] Available at: https://www.unidata.ucar.edu/software/netcdf-java/

7 Testing Validation of S-102 against S-101

A testing process has been carried out of S-102 against S-101 according to the tests described in the tables in the previous section. This used test data issued by UKHO in an area near the Isle Of Wight in southern UK and comprises S-101 and S-102 data for the area shown in the image below (All the data is available through the Admiralty digital portal – all is available as PostGIS data as well).

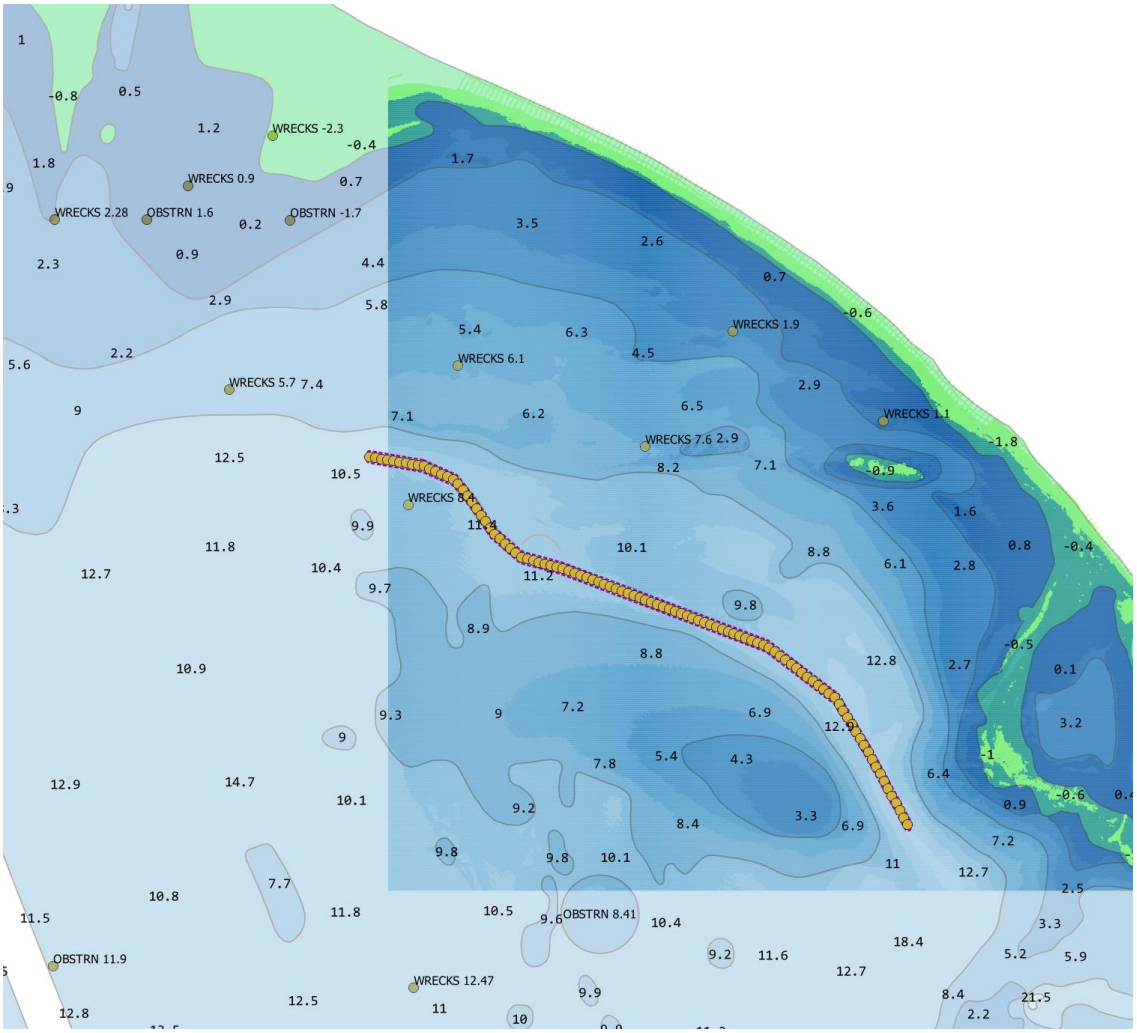


Figure 5: UKHO S-102/S-101 data (PostGIS/QGIS)

A route has been plotted through the region and the S-102 data overlaid on ENC polygons and features.

As part of this testing, a number of tests were run, including sanity tests of values, formats, HDF5 formatting and metadata examinations of coverage and grids.

In order to experiment with cross-product validation the depths in the ENC (as defined by the S-101 depth areas and valueOfSounding attributes were cross referenced against the 500k depth value points

in the S-102 dataset. This involved using the rectangular extents of each S-102 point (as described in the previous section using linear interpolation) against DepthArea polygons and features with value of sounding values.

This process, although compute intensive (the S-102 dataset contains 530k points, each of which has a rectangular extent which requires a spatial join with every DepthArea feature and feature with ValueOfSounding or depthMinimumValue) is achievable when all the data has been ingested into a spatial database in a common form.

It is interesting to note that, because the S-102, S-104 and S-111 datasets consist of a very small (typically 1 or 2) “features” (in the S-100 sense of the word) the number of validation tests is actually fairly small. The challenge is actually the scale on which they need to be carried out. This is in contrast to ENC validation where the numbers of feature classes is much larger (typically around 60-100 in an average ENC) but the number of features in each class is much smaller. One of the conclusions of this report is that it is likely the number of inter-product tests will remain small in number but fine attention will need to be paid to how they are defined and the practicalities of executing them in large quantities. It is also noted that, because the S-10X products considered are, essentially, simpler in structure than ENC, the action to “fix” validation errors is normally straightforward – grid structure, data values, metadata, and basic conformance with product specifications require only correction to the product specification. Inter-product validation errors will require more complex fixes to either product which can be resolved (in most cases) only by the data producer.

The tests performed were the following:

1. Basic parsing, textual conversion and database upload from HDF5 to the PostGIS/QGIS validation subsystem.
2. Reconciliation of each S-102 depth value (as its bounding rectangle, bounded by linear interpolation) with the intersecting DepthArea of the S-101 ENC. This was done across depth areas of positive depth and intertidal areas)
3. Reconciliation of Soundings with S-102 points
4. Features with valueOfSounding (these were Wreck, UnderwaterAwashRock, FoulGround, Obstruction)
5. Features with depthMinimumValue aside from Depth Area (these were DredgedArea for the selected dataets).

The only validation “error“ found during the process is illustrated in the following screenshot:

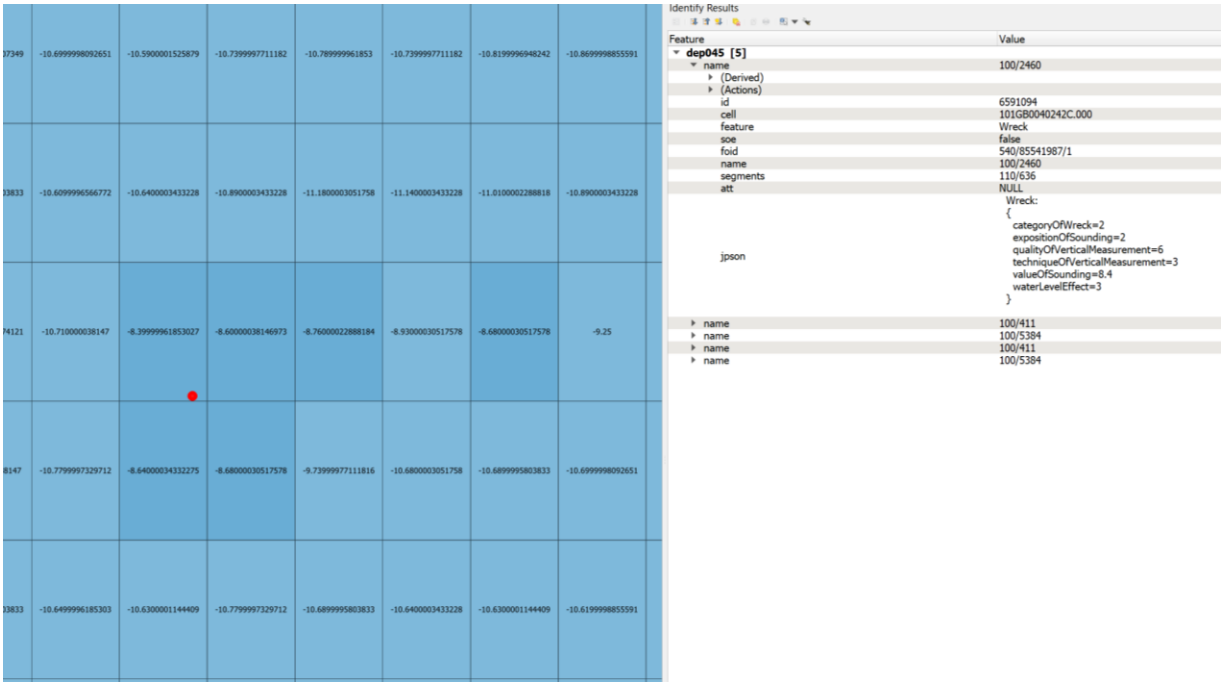


Figure 6: Validation cross-reference test results.

This concerns a Wreck feature (with a coincident Obstruction) with a valueOfSounding of 8.4m. As can be seen in the above image, the S-102 value corresponding to the point feature (which is within the bounding rectangle of the S-102 point) is 8.39999961853027 – obviously this is only marginally (fractions of a millimetre) shoaler than the charted value but it does show that the validation process works and produces results which are intuitive (and perhaps raises the question of whether validation tolerances should be defined for test execution).

No good examples have been found for reconciling S-111 data with tidal data in S-101 although work is continuing with US datasets. This will be fed back as this work concludes.

8 Conclusions.

Some conclusions based on the development accomplished for this task are listed below:

1. The current model of validation, contained within S-58 is a comprehensive framework for single-dataset validation. Inter-cell validation (of the same product) for the purposes of horizontal and vertical consistency is also performed by validators. However, there needs to be a more formal statement of what validation "is" and its operations so that both inter-product and cross-product validation can be defined consistently
2. The emergence of S-98 Annex C and its algorithms for user selected safety contour and water level adjustment are a significant step for S-100 ECDIS. Data Producers need to be aware of the behaviour of the ECDIS when loaded with multiple S-100 products and consequently there is a need to extend validation across multiple products at different scales
3. There are impacts on data producers in terms of release of S-102, S-104 and S-111 data. These need to be explicitly described so that data producers can balance their risk with the advantage to their end users of access to these ENC-aligned datasets. The value of automated validation tests will help considerably to mitigate any risks and show up any mis-alignment of ENCs with auxiliary datasets.
4. Reconciliation of S-102 with S-101 (and S-104) is reasonably straightforward to describe and the tests in this document should form a basis for a more formal description within a (to be defined) S-58-like validation standard which deals with cross-product validation focused on S-100 ECDIS. Validation of S-111 is more complex because a full validation against S-101 tidal features requires alignment against HW/LW times which validators may not have access to. Some validation against maximum values is certainly possible but a more complete validation suite may require input from IHO working groups to complete.
5. There is a lack of representative data against which to develop tests. UK, KHOA and US all have datasets which have been invaluable but they do not test all possibilities for S-101 features and often do not overlap as required for complete testing. S-98 Annex C testing will also require a large quantity of representative/authoritative test datasets and this joint requirement should be noted by the S-100 working groups responsible.
6. Further cross-validation of S-101 and S-111 remains to be done, once suitable test data has been found. This will possibly refine the S-111/S-101 tests. Additionally, a more general framework for reading the HDF5 files and populating a spatial database with coverage polygons for each data point is being developed which will enable the validation tests developed to be scaled up across regions of interest. This will be demonstrated at a later date.

9 Original Task Statement of Work

9.1 T2. Key Goals

This task aims to define a number of validation tests on selected IHO S-100 product specifications, specifically S-104, S-111 and S-102.

Much of the current test development effort in S-100 is concerned with re-definitions of existing tests within S-58, primarily for S-101 ENCs, and conformance with the broader S-100 framework requirements. An initial view of such testing has been produced and is in the process of review.

Little work has been done in two areas of validation testing:

1. Validation of other product specifications against independent criteria, forming a set of tests with equivalent rigour to S-58. Such tests would ensure non-ENC data is fit to be imported into an S-100 ECDIS and used for primary navigation by the end user.
2. Validation between different product specifications, in two important areas
    a. Validation between a non-ENC product specification and the S-100 product which overlays it, for example S-101/S-102 or S-101/S-104
    b. Validation between S-101 and the legacy S-57 ENC it replaces

This task should develop an initial framework for development of tests for (1) above in which tests can be placed. This should categorise tests, set out a language for their definition and a set of categories in which results can be placed (compatible with existing S-58 and its S-100 equivalent). (2) should also be set out in some detail, with particular reference to the S-57 to S-101 conversion document currently in draft by the ENC conversion subgroup of ENCWG and S-101PT.

This task should also develop an initial view on how validation testing can be carried out for data producers who are releasing multiple S-100 products in a single area and advise on how data production can be modified to take into account both dataset validation and cross-product validation. The impact of the interoperability specification should also be accounted for in the outputs of the task.

9.2 T2. Deliverables

IIC will deliver the following in respect of this task.

- A initial set of tests for each of the nominated product specifications
- Draft cross validation tests between each of the product specifications and overlapping IHO S-101 content.
- A set of test datasets illustrating some of the test cases and showing how validation test tools should interpret the results.
- A report detailing the framework for test development, documenting the tests defined and the datasets compiled. How the testing framework relates to S-100 ECDIS, Dual Fuel Mode and the emerging S-100 ECDIS Governance Document should also be covered.