S101PT 4-4.4

Paper for Consideration by S101PT

S-57 to S-101 ENC conversion study outputs

Submitted by:	IIC Technologies
Executive Summary:	A study into the conversion of ENCs from S-57 to S-101 and the possibilities for optimising the conversion process has been in progress since Summer 2018. This paper presents the conclusions of the study and suggests opportunities for further development.
Related Documents:	IHO S-101 Product Specification, DCEG IHO S-57 UOC, KHOA Paper S-100WG Singapore.
Related Projects:	S-58, S-64

Background

Following the S-100WG meeting in Busan in 2018 a piece of work co-sponsored by NOAA and KHOA was commissioned to look at the process of conversion of S-57 ENC data into S-101 form. Against a background of the ongoing development of the S-57 to S-101 converter and associated viewer the following objectives were identified:

1. An examination of the current converter application, its operation and results gained.
2. To systematically look at how S-57 could be "optimised" to prepare data for conversion to S-101 and what the scope for performing this optimisation is in advance of S-101 migration
3. Report, summarise and suggest next steps for the conversion process.

An initial report was presented at the S-100WG TSM meeting in Busan and reviews and refinements made to the report following that meeting. A number of subsequent objectives were identified from the initial report and progress towards those have been made. A broader review of the content of the report is now in progress and presentation of this paper concludes the development of the study. This paper presents the essential elements of the research and conclusions generated.

Optimising S-57 to S-101 conversion

A systematic methodology was identified and defined during a process of initial investigation using up to date toolsets and data. These were:

1. The ESRI converter tool at its major release
2. The SPAWAR S-101 viewer which also contains numerous analysis tools for data inspection.
3. Data to support the study was made available by several member states. Support from S-100WG members was also gratefully received during the work and the writing of the subsequent report

The core of the study's activities was development of the methodology and a side-by-side comparison of the S-57 Use of the Object Catalogue (UOC) with the current version of the S-101 DCEG. These documents contain extensive and detailed descriptions of the encoding for "equivalent" real world features relevant to the purpose of ENC data arranged in semantic groupings. The study of the two encoding guides was further supplemented by examples drawn from interpretations of the UOC by participating member states – this common practice details how individual member states adapt, enhance (and occasionally) restrict the encoding of ENCs depending on their own internal policies, regulation, cartographic practices and other constraints.

Conceptual Basis.

The last time such a study was performed was when a systematic study of DNC to ENC conversion was carried out by the HIHWG in 2006. There are some similarities but also several notable differences:

1. The domains of reference between DNC and ENC were fundamentally different. Although based around navigational features there were numerous conceptual differences where conversion required manual intervention and reference to an overarching model of real world features.
2. In this study S-101 is largely derived from the model of real-world features defined in the IHO S-57 Object catalogue whose encoding is documented by the S-57 UOC. Chart content for ENC is drawn from a single conceptual entity of information necessary and sufficient for safe and efficient navigation under the SOLAS convention.
3. The geometry model of S-101 is derived directly from the underlying ENC model and, barring levels of abstraction is directly convertible with minimal topological change. This allows conversion to be focused almost entirely on the semantic meanings and feature/attribute conversions, tying the two products together much closer.
4. The ENC ecosystem is a well developed and thoroughly tested one having been accepted and adopted worldwide by the commercial shipping industry under the SOLAS convention.

The concept of a "domain of reference" in the S-100 context is a significant one and during the course of the study it has made more sense to consider geospatial features as belonging to a particular domain of reference as specified by its individual product specification feature catalogues. In the S-100 ecosystem a product specification defines a feature catalogue which defines completely all possible features representable by that product specification. The question as to whether IHO S-57 defines (similarly) a domain of reference is still an open one (thought to be solvable by defining an S-100 feature catalogue which provides a one-to-one correspondence with S-57 features). Assuming this is possible (big assumption) then data conversion is a matter of transformation from one domain of reference to another (S-57 to S-101).

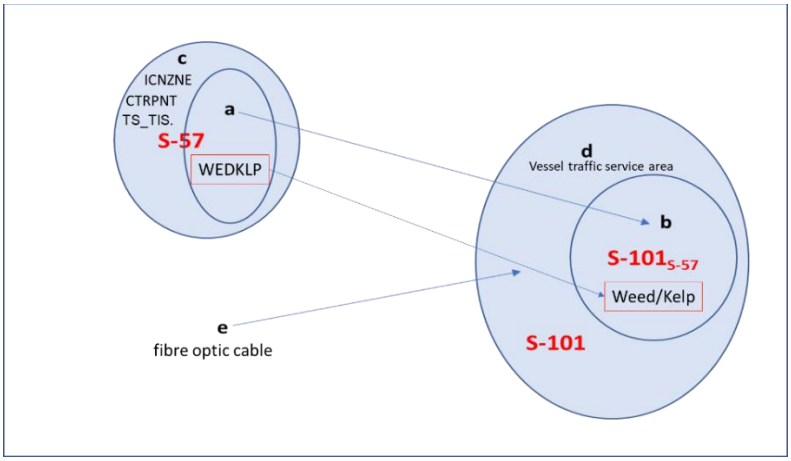


Figure 1: S-57 and S-101 domains of reference.

In the diagram shown the domain of features defined by IHO S-57 UOC is represented in the left hand set. The right hand set shows the domain of features covered by IHO S-101 (these can be any of the components of the product specification – features, attributes, associations, feature bindings, enumeration values etc). The elements of the conversion between S-57 and S-101 are categorised as follows:

a. Features which translate directly from S-57 to S-101

b. S-101 equivalents of S-57 features

c. Features which can't be translated into S-101, i.e. that have no equivalent. The examples given are Control Points and Incineration Zones

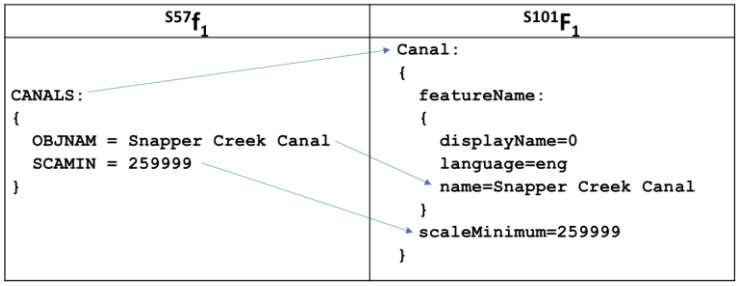
d. Features for which related features may have existed in S-57 but which are modelled in a fundamentally different way or enhanced within S-101 – the example is Vessel Traffic Service areas. There are a number of these aggregated groupings within S-101

e. Real world features which were never modelled in S-57 but which have S-101 representations – these are items which were "missed" for a number of reasons from the S-57 view of the world. The example given is fibre optic cable.

This provides a conceptual model for the study. By examining the UOC vs the S-101DCEG a good coverage of possible feature conversions is possible

Encodings, GML, XML, JSON etc…

A notation for features in both S-57 and S-101 was developed during the writing of the report which facilitates side by side comparison of data before and after conversion. This was used to examine and notate the many converted instances during the process of the UOC/DCEG comparison:



This systematic representation of the individual domains of S-57 and S-101 universe is used to show a number of interesting observations on conversion which are contained in the output of the report. The encoding purposely used a JSON-like syntax to emphasise its neutral nature, as opposed to both iso8211 and GML which contain a substantial amount of standard specific encoding content. A neutral XML or JSON-like encoding has proved very useful to show conversions between different domains and also to represent transformations between them and a more concrete version of this will be included in the report's outputs.

UOC comparison

The methodology examined all instances where a "real world feature" was equivalently encoded by the two documents (S-57 UOC and S-101 DCEG) and what, if any, transformation process was required in order to produce S-101 versions of the existing S-57 data. In many cases the S-57 data is directly transformed into an equivalent S-101 feature (and attribute). There are, however a number of gaps, best exemplified in the following diagram:

Conclusions

An initial summary of conclusions was presented at an early stage of the report's publication which were then supplemented with the output from further work after the initial study. Currently the report is awaiting a broader

review from stakeholders and will be finalised in June 2019 following receipt of any further review comments and completion of a number of follow on tasks

Toolkit Assessment: The current toolset works! It is possible today to convert S-57 data to S-101 and view and inspect it within a viewer. This in itself is a major achievement and enables the more detailed consideration of the underlying data to take place. The current converter has been enhanced to the point where detailed reporting of its results are output and optimisation of data for the conversion is possible.

Conversion and other tools: A converter forms part of an interlocking set of tools and standards in the ENC world which generate, validate and format ENC data. A converter can form part of an interlocking set of tools for supporting the transition from S-57 to S-101 and, properly configured, could form the basis for production of carriage compliant ENC data in both formats by a single production system

INFORM: The main source of ambiguity in current S-57 encodings is the use of the INFORM attribute. INFORM has been used for many years to encode elements of real world features (often as a part of member state specific encoding guidelines) which don't "fit" exactly with UOC guidelines or, sometimes, within the UOC where S-57's feature catalogue don't have the capability of representing data adequately. A sample of approximately 900 cells showed approx. 66,000 INFORM encodings

Conversely, INFORM can be used to populate many new S-101 features by adopting a structured encoding format and ensuring a suitably configured converter is available to perform the appropriate transformation. The current converter is not designed with such mechanisms in mind but does convert data to a minimum standard of compliance with the S-101 product specification and, to a large degree, the S-101 feature catalogue. There are a number of extraneous features encoded in current S-57 cells which are there to ensure display of data in ECDIS Base mode. As such these are not required in S-101 and could be removed.

Using a combination of structured INFORM encodings a suitably configured converter would be able to convert S-57 data into an "equivalently safe" version for use within an S-101 ECDIS. These terms would need to be made more precise (and suggestions for doing that are in the recommendation section of the report generated).

Converters: The general observation to make is that the more sophisticated a particular converter technology and the more aligned it is to specific encoding guidelines (such as structured INFORM encodings) the more complete the transformation From S-57 to S-101 is. This suggests that converters between these different domains can have a variable level of conversion depending on their complexity and the conformance of the data to tailored S-101-ready encodings within S-57. Although feasible, this is as yet untested although the methodology suggested by the report sets out what would need to be done to achieve such a data conversion.

Existing S-57: A number of observations relating to existing S-57 data became clear during the period of the report and its research. This highlighted some gaps in the existing S-57/S-101 ecosystem which could be filled over time and which would facilitate much easier automation and more sophisticated tools for ENC conversion as S-101 matures:

1. The existence of a machine readable specification for conversion between two domains specified individually by feature catalogues. This would need to take in all nature of conversions from straight dictionary conversion ('LNDARE' -> 'Land Area') to more sophisticated ones such as the grouping of converted LNDARE features under an "Island Group" feature and the insertion of its appropriate associations.
2. The existence of a "feature catalogue" for IHO S-57 data to assist systematic conversion between the two domains. Currently encoding is only mediated by the S-57 UOC which only exists in document form.

No abstract definition exists of the actual operation of the current converter. This restricts the ability of any study of the converter to be systematic. This would not be difficult to do as the conversion is largely dictionary based. This is still in progress. Given the complexities of the DCEG encoding any conversion definition needs to be able to properly represent all conversions explicit in the DCEG including the conversion of features to attributes, C_AGGR to associations and the definition of new features from structured attributes. Additionally, elements like

Skin of the Earth feature substitution may only be possible to describe in the abstract due to the complex geometric processes involved in their conversion.

Data Validation: No equivalent level of e.g. S-58 compliance exists for S-101 data yet. This restricts the ability of a data producer to evaluate whether a converted cell is "safe" (in the navigational sense) or not. This will follow in time as the S-101 ecosystem matures but should be highlighted at this stage as it limits the ability to objectively measure how complete or "good" a conversion is.

Feature/Attribute bindings: During the writing of the report the DCEG was baselined with a comprehensively documented set of feature/attribute bindings which, although commonplace in the existing global ENC dataset, are now undefined. The quantity of these bindings should not be underestimated and, although arrived at through a very systematic and thorough process within the S-101 project team the impact on migration of ENC data to S-101 remains to be seen. The absence of a feature catalogue for S-57 data makes it difficult to automate comparisons between the two global datasets and to quantify the impact on individual cells.

Migration from S-57 to S-101: A graduated level of sophistication of operation of the converter could capture a number of scenarios for S-57 to S-101 conversion in a progressive way as proposed in the INFORM section earlier in this document. The process of conversion should ideally be considered within the context of various different models of how member states intend to migrate from S-57 to S-101 and the assurances required that data remains safe for primary navigation, equivalent etc. These levels of compliance resemble, and could be a part of, "Technical Readiness Levels" for S-101 proposed elsewhere in the S-100 ecosystem.

Test datasets: Preparation of a comprehensive test dataset would be a necessary first step to test that different converters behave equivalently and that the specification for conversion defined is followed correctly. An example test dataset has been prepared as a follow up action to this report which can be demonstrated to illustrate the concept.

Action Required of S-101PT

1. Note the report generated by the converter work and the steps forward made in this area.
2. Note the usefulness of machine readable catalogues for specifying S-101 data and the flexibility it allows when looking to convert between different product specifications. This has applications for broader use of marine geospatial data in an MSDI context.
3. Review the existing paper and present relevant comments.