Submitted by:	Data Protection PT and IC-ENC[1]
Executive Summary:	Management of Auxiliary Files in S-101
Related Documents:	
Related Projects	

Ancillary Files in S101

There are two approaches for the management of ancillary files in S-101. This paper articulates both possible approaches together with their pros and cons.

Currently ancillary files are used within S-57 for holding both text and picture (TIF format) information. TIF files,. Although TIF files have been included in ENC data since S-57s inception it is probably a good time to review their suitability within the future S-101 product specification. Observations on current encoding practices within ENCs observed by UK and IC-ENC are:

1. TIF files tend to only add peripheral value to ENCs.
2. TIF images used seem to relate mainly to back of bridge planning activities, not to the navigational use of ECDIS and the information within the images is probably better conveyed through Navigational Publications. In general the information within a TIF image could be encoded within the ENC by other means rather than requiring an explicit image.
3. Implementation of TIF files causes ECDIS manufacturers significant overheads in terms of having to implement an image display function (complete with accompanying palette colours).
4. TIF standards for resolution and sizes of TIF images do not exist and can adversely affect sizes of exchange sets.

The rest of this paper refers to the use and management of TXT files.

The information in ancillary files is (currently) unstructured and referred to from the cell by use of a TXTDSC or other attribute (such as NTXDSC) – the feature attribute holds the name of the file containing the information. No revision control is applied to the ancillary files and consequently ECDIS manufacturers have developed a variety of different methods for their management and storage. The current S-57 situation is almost a "worst of both worlds" eventuality which we would not wish to repeat in the context of S-101.

---

[1] This paper was originally presented in TSMAD28. Although much discussion took place no firm conclusions or actions were taken. It is still an outstanding issue, hence the need to resubmit and come to an agreed set of actions to move it forward. For this reason some of the metrics presented at the end of the paper may be out of date.

The two options suggested for storage and management of ancillary information are described below to illustrate the problem in more detail.

Option 1 – Embedding TXT/ information within cells.

In this option ancillary information is embedded within each cell either as feature attributes or as independent information features linked to from the S-101 features which require the information to be attached.

Advantages:
- No need for external files
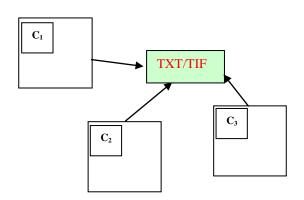- No need for revision control mechanism, uses existing S-101 scheme.

Disadvantages:
- Duplication. Information which is common across multiple cells (either in the same geographic area at different scales or adjoining cells) require copies of the information within them bringing an associated updating burden.

Option 2 – Accessing TXT/TIF information from a central location.

In this option the ancillary information is read from a central location by each cell which requires it. In a file context the TXT/TIF information would be kept in a separate cell composed only of information features which hold the information. The main cells then use a feature to feature reference to link the information to the source feature:
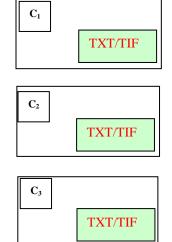
Advantages:
- Allows a "store-once, use many-times" structure around the information. There is no duplication across scales or cell boundaries.
- If a geometry-less cell composed only of the ancillary information is constructed then no new revision control mechanism is required (although management of that cell would be more complex – see below). Each TXT/TIF information feature can be updated using the same protocol as other cells.

Disadvantages:

- Introduces the concept of a dependency between the ENC cells and the cell containing the ancillary information. This would mean the TXT/TIF cell would have to be distributed separately. The mechanism for delivering, installing and maintaining the cell would need to be defined. Whilst the updating mechanism would be the same as other features it should be recognised that each TXT/TIF feature is independent so complete, cumulative sequential updates are not necessarily required.
- If TXT/TIF information was not kept in a cell then a separate updating/management mechanism would be required which would version the information linked to by the cell. This would need to be implemented for any product specifications requiring linkages to external files.

Duplication of TXTDSC entries.

In the current world ENC dataset there are approx 9,700 distinct text files. Of these 8025 are only referenced by a single ENC cell and a further 800 are only referenced by two cells. There are a number of text files which are accessed by many cells (a short list is given below):

AU057_04.TXT	341
IDA10001.TXT	224
GBCELA18.TXT	218
GBUNSDAT.TXT	145
AU030_01.TXT	125
GB011_06.TXT	125
GB007_19.TXT	124
GB060_01.TXT	118
AU168_02.TXT	111
GBCELA19.TXT	109

In all there are 20,327 references are made by distinct cells to the 9,700 text files so if Option 1 was to be taken the amount of space required across all ENCs (given the encoding practices today) would be doubled as opposed to Option 2 where all text files would be held centrally and referred to by reference. Although there are many text files they currently only make up 1% of the total size of the ENC dataset (all text files add up to about 54Mb and only 5Mb when zipped).

Summary.

In summary there is no simple way to distribute and maintain all ancillary data using existing mechanisms because of the nature of cells as a distinct collection of features within a geographic area at varying scales. The tradeoff is between duplication (independence of cells, 1% larger size requirements) and a central store (no duplication, more complex to manage).

Recommendation.

1. Unless a compelling reason for including image files can be found they should not be part of the S-101 product specification. A compelling reason is one where elements of the IMO Performance Standard for ECDIS as interpreted by IHO standards can not be satisfied within the feature object and associated encoding standards.
2. Note the duplication of current TXTDSC information and the small number of (possibly extraneous text files which could be reduced by more efficient encoding?)

3. Decide which of the two options is optimal for future S-101. If Option 2 is decided then a separate management regime may be required to update textual information.