TSMAD18-16.3F

18th TSMAD MEETING
4th to 8th May 2009 (Ottawa, Canada)

S-101 Coverage Referencing System – Discussion Paper.

(Tony Pharaoh – IHB)

Executives Summary: This discussion paper proposes a universal, global spatial indexing system that could be used for S-101 data.

Introduction

Since the publication of S-57 Edition 3.0, ENC production and coverage has grown significantly, and new concepts and structures have emerged. Partnerships between Hydrographic Offices, RENCs, data producers and data distributors have resulted in the provision of a range of services that include ENC quality checks, validation, distribution, updating and others. ENCs are still however distributed as discrete cells which are based on the limitations imposed by technologies of the middle 1990's. (e.g. - the 5 MB limit for cell sizes).

During the S-101 workshop that took place at the IHB (March 2008), a breakout group was tasked with answering the following question; "What improvements can be made to avoid low-level data warning messages appearing on loading ENCs?" One of the conclusions of this group was; *"End users want to see all charts as a managed database instead of individual cells. Especially they need a date up to which this database is updated"*.

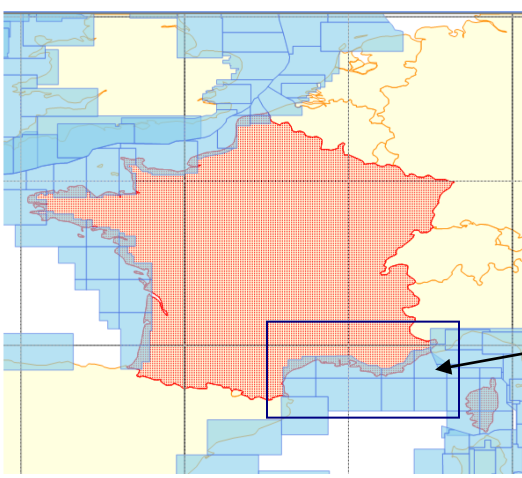
The S-57 ENC product specification provides little guidance on how cell should be should be schemed, indexed and numbered. Almost all HO's have developed their own schemes, based on their national chart requirements. They usually have no continuity with other neighbouring systems and have no relevance or meaning within a global context. The only way to determine where cells are located or which cells overlap is to view their limits in a catalogue. The cell limits (provided in an ENC catalogue file) seldom represent the actual ENC dataset coverage, and it is necessary to load the bounding polygon (contained in dataset) in order to understand the extent of a cells data coverage. It is proposed that national schemes are largely based on paper chart limitations and are no longer relevant for S-101 ENC production, distribution and updating

In an ideal world an ECDIS should be able to receive one large dataset that provides contiguous, harmonized coverage for its intended voyage. There are two main reasons for ENC's to be divided into cells. The first is to ensure that datasets do not exceed the internal resources (memory) of the ECDIS system. (One very large cell would overwhelm these resources, however smaller cells allow the ECDIS to load/unload cells as the vessel progresses along its intended rout). The second reason is for updating purposes (i.e. to address an update to a certain feature within a certain cell).

As many HOs move towards implementing seamless databases covering their entire areas of responsibility, the requirement to provide data as small ENC cells is questioned. This paper proposes that S-10X products should not be constrained to paper sheet bound limits, or data loading limits. Data producers (HOs) should be able to provide larger datasets which may, for example cover portfolio areas covering important navigational routs. Data segmentation (i.e. breaking down into smaller cells) could then be carried out by data producers, Value Added Resellers (VAR/Service providers), data distributors or within the ECDIS (during SENC conversion). In order to achieve this in a uniform/consistent manner a suitable grid reference system must be adopted (This will allow cells to be easily identified and addressed for updating purposes, and will be the basis for indexing additional overlays such as bathymetric grids). It is proposed that the C-square spatial indexing system (described below), should be considered for S-101 use. This index system is well suited to marine datasets (which frequently have an irregular footprint). The C-squares indexing system is described in annexure A below. An example of how the system could be used to index cell located within a dataset folio covering the south coast of France is provided below.

Example – coastal ENC portfolio coverage.

Assuming that the coastal ENC coverage for the south coast of France was provided as a portfolio dataset, it would be possible for, either SHOM, PRIMAR /IC-ENC, or a data distributor to segment the dataset onto cells according to the users requirements. (A further possibility is that data segmentation could be undertaken by the ECDIS – possibly during SENC conversion). Irrespective of where the segmentation takes place, the cell footprint can be stored as a unique, easily identifiable C-square (text) code. The complexity of the code depends on the complexity of the cell footprint. In figure 3 below, the 1 x 1 degree cell would be identified by the following text code 1400:226. Presently the only way to identify the location of cell FR303110 would be to consult an ENC catalogue, however any person (or system) understanding the C-square notation could very easily determine its location from the text code. The data coverage extent of cell FR303120 (shown in figure 3) could be indexed more precisely using the 0.5 deg C-square notation - 1400:235:1|1400:235:2. (The existing extent of this cell is shown in green and the c-square cells are shown in red). These codes could be stored in an associated text file and would enable systems to easily identify overlapping cells (between usage bands) without the need for complex spatial analysis (i.e. no special, vector-based searching overhead). The "nested" nomenclature means that searching can be carried out at any level of the hierarchy equal to, or greater than, the encoded resolution.



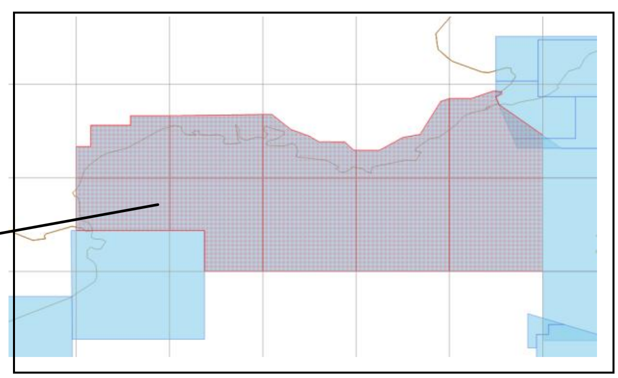


Figure 1

Figure 2

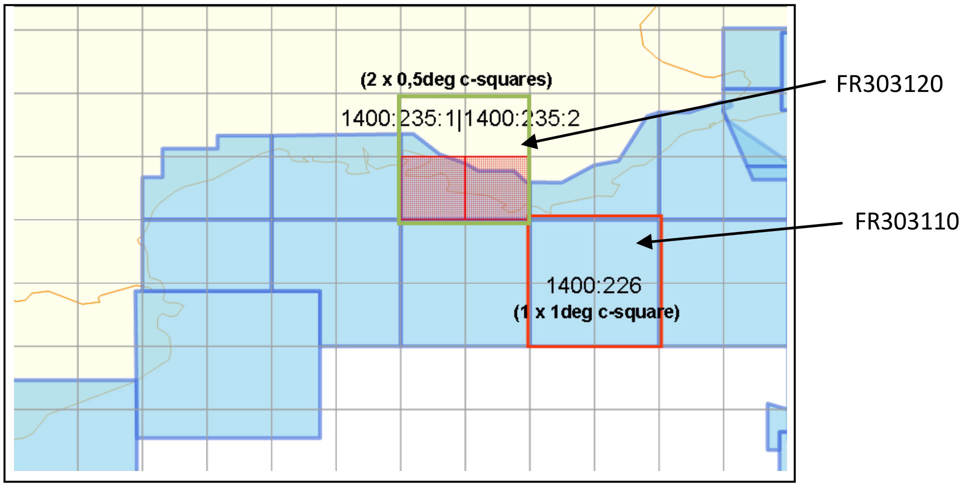


Figure 3

Some Advantages for S-101

Some of the advantages of implementing a uniform grid reference system are listed below;

- The principal benefits include (a) the flexibility to represent a wide variety of shapes and sizes of a datasets footprint, (b) the relative simplicity of the storage and query mechanisms for the c-square strings, and (c) the ability to decouple this information from the base data for the purpose of spatial searching and display. Being standard ASCII text, C-squares strings can also easily be included in metadata exchange sets, and are also well suited for expressing in an XML format.

- The C-square notation can be extended indefinitely to give any desired resolution. This will allow the extent of a cells dataset coverage to be specified far more accurately in an associated ENC catalogue file;

- It is not tied to a national schemes and overcomes some of the limitations of bounding box spatial referencing;
- It could provide a common spatial indexing system for both Nautical Publications, ENC data (and other overlays), and could facilitate the implementation of harmonised distribution and updating services. (e.g. If text and picture files were associated with a c-square grid reference it would be very easy to determine which cell were affected by an update to these files);
- It compliments existing S-100 metadata specification;
- It provides a standardised structure that could be implemented within production databases, web portals, distribution systems and OEM (e.g. ECDIS) systems;
- It is compatible with other grid reference systems used within the marine environment (e.g. WMO reference system).
- It could be used for implementing other applications (e.g. search and rescue overlays within an ECDIS system).
- Data producers will not have to be constrained by fixed, uncoordinated ENC cell schemes. Data could be provided as larger datasets covering portfolio areas, and segmentation could be undertaken by distributors.

- Data distributors would have greater flexibility in providing better services to the mariner.

Disadvantages for S-101

There may be incompatibility issues with S-57 ENCs.

- Data producers may wish to keep their proprietary chart schemes.
- Some data producers may not be in a position to produce seamless production databases.

Annex A

C-Squares Spatial Indexing System

C-Squares is a reference system for the storage, querying, display, and exchange of "spatial data" locations and extents in a simple, text-based, human and machine readable format. It uses numbered (coded) squares on the earth's surface measured in degrees (or fractions of degrees) of latitude and longitude as fundamental units of spatial information. These can be quoted as single squares (similar to a "global postcode") in which one or more data points are located. A wide variety of shapes and sizes of spatial data "footprints" can be represented by simple strings of codes. Its notation can be extended indefinitely to give any desired grid resolution.

As its primary level of subdivision, c-squares uses the World Meteorological Organization's (WMO) ten degree global grid square notation where the initial digit **1**, **3**, **5** or **7** indicates the global quadrant NE, SE, SW and NW, respectively. This is shown in Figures 4 and 5 below.

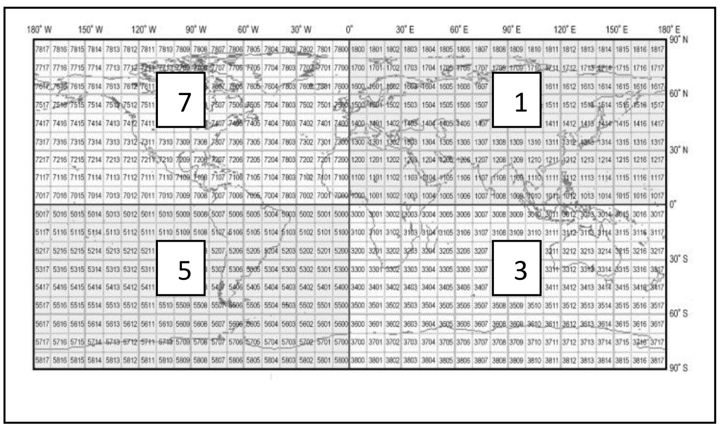


Figure 4

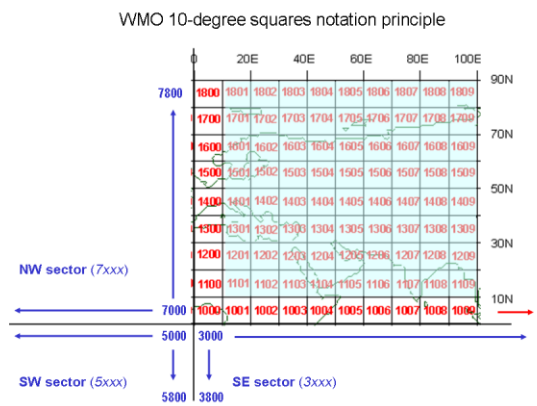


Figure 5

The notation for further subdivisions into 5 degree and 1 degree grid cells is shown in Figures 6 and 7 below.

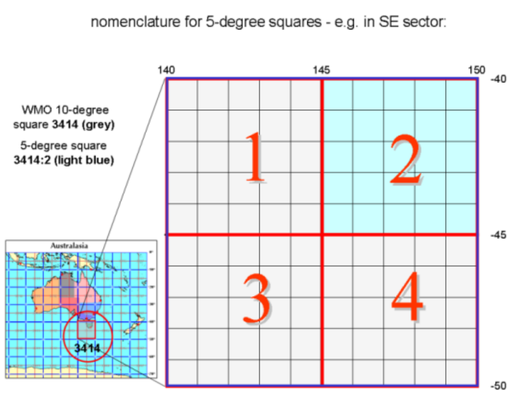


Figure 6

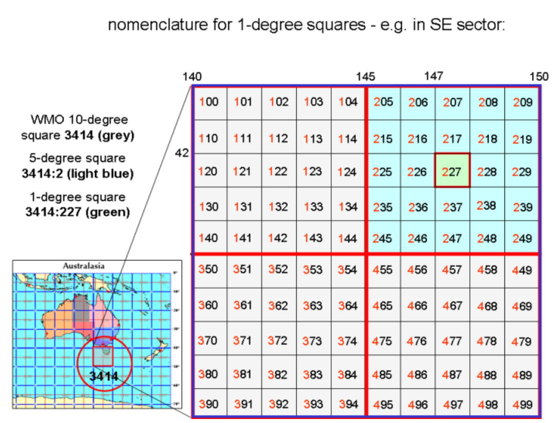
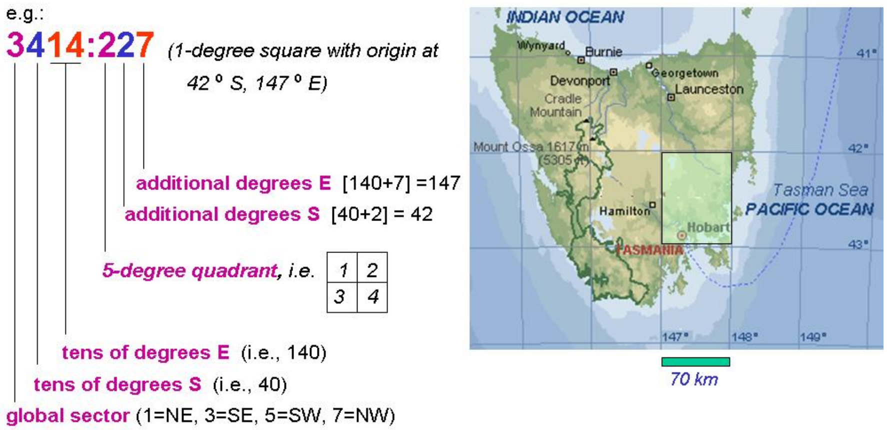


Figure 7

Example code for Hobart - Tasmania



The ability to do simple spatial queries of documents or indexed objects usually relies on either using a controlled list of geographic search words (which are applied consistently to the objects such as counties, oceans, seas, etc ...), or by using some type of numeric units such as latitude and longitude. Lat / long designators are typically either one set of values (for a point), or some type of range (e.g. maximum and minimum values or a centre point plus a radius). These provide a bounding box or a circle which forms an approximation of the spatial extent of the data, and can be searched using numeric functions. True "polygon" boundaries of dataset footprints can also be expressed in simple notation (e.g. by giving lat/long coordinates of a series of points which are then connected together by straight lines to produce the boundary shape). To search these however requires the use of complex spatial queries.

C-squares makes provision for a simple, text-based method of representing either points, or more complex shapes on the earth's surface ("spatial objects") which are easy to index and easy to query (e.g. to answer a question "which of these objects occur within my area of interest"). This can be done very easily using standard text queries within a word processor, text-based database, or internet search engine and does not require sophisticated GIS software.