- TMSAD18-16.3G –

Joint TSMAD18 & DIPWG1 MEETING

4th to 8th May 2009 (Ottawa, Canada)

Service Delivery
S-63, S-57 and S-101 - Perspectives

I called this paper "Perspectives" because I wanted to share some of the perspectives that the DPSWG and the various data servers (people who distribute ENCs) and ECDIS manufacturers have shared with UKHO and DPSWG over the last few years of developing ECDIS/ENC capability in the hope we can learn from our experiences in the development of new standards for ENC.

We are at a particularly crucial period of development of S-100 with the writing of the S-101 product specification and this paper discusses some of the different aspects of onward ENC distribution and how those aspects can be thought about as we develop S-101 further. The hope in putting these views forward now is that we can come up with a product specification that is more flexible, more efficient and more able to stand up to the developments of the world of electronic navigation around it.

I chair the DPSWG, the Data Protection and Security Working Group for the IHO and we are charged with definition and maintenance of S-63. DPSWG arose out of the shift to the IHO of the old “Primar Security Scheme”, PSS, an early prototype for commercial ENC distribution. PSS became S-63 when it moved to IHO and has since undergone two issues with the latest one being incorporated into the iec61174 edition 3 release as well and used in type approval for new ECDIS.

S-63 provides a means of taking an unencrypted data product, an ENC, and outputting a service which a mariner can use to gain carriage compliance with their ECDIS. It is a "layered" service where the “S-63-ness” is provided on top of the underlying S-57 data. S-63 provides compression, encryption and digital signatures facilities throughout the supply chain for ENCs and is used to provide predominantly commercial services for ENCs. In this sense both S-57 and S-63 define an interconnected encoding of ENC data which results in a hybrid product being delivered to the end user (is this optimal? More about this later).

DPSWG held its 7th meeting in Monaco on 31st March/1st April and one of the days was aimed at areas of ENC distribution and ECDIS implementation where potential improvements could be made in the context of a fresh approach, S-101. Some of these discussions I have reproduced here and some may be talked about elsewhere (e.g. Richard's paper on auxiliary file management). The topics I have selected are there to provoke discussion and give a flavour of the kind of issues we deal with in DPSWG.

S-63. What works!

I’m going to start by reviewing the way that the core elements of S-63, the copy protection and digital signatures are defined and then look in a bit more detail about how compression and data encoding are related. I’ve concentrated on the elements which are familiar to most and which work well in the distribution chain.

Permits

Permits are the units of copy protection defined by S-63. Permits have worked very well in the whole S-63 scheme by providing one of the building blocks of a commercial ENC service to service providers. The basic idea is that an ECDIS can be "tagged" with a unique ID number (normally implemented as a hardware dongle on the ECDIS). Unlock keys for the data are then constructed, wrapped up encryption keys, which will unlock data only on that particular ECDIS. What this provides is a copy protection mechanism where:

a. A user can only install and use data on a system as long as they have the correct hardware id plugged in, i.e. only a single copy of the data can be used against that license.
b. It is impossible to copy data from one machine to another without a new set of keys being issued for that license.

This mechanism is actually very simple but provides the basic foundation that all commercial ENC services are built on and is used in all non-SENC ENC distribution services.

No system of copy protection is 100% foolproof but the basic mechanism works well and has been implemented consistently across all ECDIS manufacturers (there are currently over 100 signed up to the IHO-administered data protection scheme). Permits are an example where the some elements of the ENC product specification are used to define the S-63 elements essential to a commercial protection scheme.

The copy protection aspect of S-63 is dependent on:

1. ENC "cells" - the foundation for encryption is file based, so S-63 needs files of data to encrypt with a key for unlocking.
2. Unique cell names - we issue a key per cell so a unique cell name is a dependency
3. Edition numbers (although we rarely use these now!) were a dependency as changes to encryption keys were made dependent on new editions of cells. Although this tie is not generally used any more the important point is that there is/was a link from an element of the S-57 product specification through to S-63.

Could this scheme be improved – well, there’s probably some modifications we can make but I suspect that although the details may change, the basic idea of unlocking data against a specific ECDIS will continue.

Digital signatures and CRC values.

S-63 has a mandatory data integrity check in the form of digital signatures. These are encrypted CRC values and the ECDIS is able to check that a cell has not been corrupted in transit to the ECDIS. It also gives the user an assurance the digital signature was created by a "trusted party", a reputable digital chart provider, a hydrographic office or other service provider. Implementation has been difficult but the assurance is undoubtedly there and is a mechanism we would like to maintain in future versions of the standard. The digital signature scheme is the same in outline as the SSL (or “https:…” you see in web browsers) hierarchy which is used by internet applications worldwide for global e-commerce

S-63s digital signatures depend much more heavily on elements of the ENC product specification such as file/folder layout and filenames:

a. ENC cell names and directory structure. We place digital signatures next to the cells which they represent in the folder structure.
b. ENC naming conventions. We rely on the usage band to map to a unique digital signature name.

c. CATALOG.031 (we store signature details within the catalog.031 file).

In terms of improvements better algorithms could be used and the concept of data chain certification has been raised where data can be signed at each stage in the chain, even potentially when it is reformatted or grouped differently. Digital Signatures need to be articulated clearly to the customer though what it means and I think that communication has not really been 100% successful in ECDIS rollout. Interestingly enough digital signatures replicate CRC32 values in the catalog.031 file and our blind acceptance of these two mechanisms may not be worthwhile in the future.

It's worth noting that in the transition to digital signatures their operation could become seamless with no changing of keys by users and manually intensive, confusing processes. We need to emphasise that the authentication elements of S-63 are very powerful and can provide a "big green tick" or padlock depending on your choice of GUI to a user.

Service management.

We are able to install, check, translate to SENC, provide customer support, check holdings, compile management information and provide a fairly wide range of products and services for people with a need for ENC data. This all works and generally works pretty well. It's not an online solution though yet (even though many online distribution schemes are starting up) and it's still pretty weighty and manually intensive in some areas of the chain.

Service management has evolved and various parts of the S-57 product specification are used such as cell names (again!), file/folder layouts, update methodologies, media layouts.

Better Service elements

What are the new or improved service elements we need?

1. Better control of auxiliary file management. One of the big holes in the ENC product specification is the management of auxiliary files. The association of text/tiff files with cells and the formats/conventions used causes numerous service problems and needs a fresh look.
2. Better visibility and separation of metadata from underlying chart data. We provide numerous different files with metadata and service-specific data within them (catalog.031, PRODUCTs.TXT, cell names, SERIAL.ENC as well as new media files and digital signatures). This is largely because the decision was taken to encrypt entire cells and not leave headers or metadata in plaintext. A fresh look at what needs to be encrypted to provide copy protection needs to be done and reduction in the duplication of this data (do we really need to store update numbers in cells, in their filenames,products.txt and the catalog.031?)
3. Dealing with multiple data providers and services sensibly. It needs to be clear to the ECDIS which "service provider" data comes from and the standard(s) should define a flexible set of protocols and structures for letting this be done. For the first time we have got S-63 included in ECDIS type approval thus ensuring new systems will have to import and manage data correctly as well as display it for primary navigation.
4. Data structure itself. There is more of this later but it is worth saying that loading and import of ENC data can be very torturous for ECDIS and structures to allow that to work better MUST be considered in the writing of the standards as well as facilities to expand management by service providers to their customers. Our discussion with manufacturers leads me to think there must be better ways of structuring the data so it can be more easily imported into ECDIS. It should be possible for a service provider to deal

exclusively with the service elements (data install, removal, management, revocation etc…) with the ECDIS' support rather than fighting against multiple different methodologies for data import on different manufacturer ECDIS.

What's the point?

The point of illustrating these elements of S-63 is to communicate how they are built on "top" of elements of the S-57 product specification and how they are dependent, therefore, in a lot of ways on the supply chain producing data which conforms to the product specification.

It is interesting to note that most of the areas S-63 is dependent on are the areas of the S-57 product specification which define the "encoding" of the data, i.e. its physical rendering either on hard or soft media. Now if you read the standard it is also those elements which look the most dated e.g. floppy disk layouts, 5Mb cell limits, DOS (8.3) filenames etc whereas technology has moved on significantly in that time (even ECDIS technology which admittedly moves somewhat slower than our mainstream desktop PCs).

S-63 largely cares little for the content of the data except insofar as it uses the binary encoded content to define digital signatures. Subjects such as s-57 3.1.1, ASLs etc and dynamic catalogues largely are bypassed by S-63, the object->cell encoding is just a vehicle for translating data into a form for distribution. As long as the signatures (and associated CRC32 checks) are complete S-63 doesn't look "inside" the cells at all really. There are a couple of exceptions but the vast majority of S-63 processing could be done on top of ANY file type, not just S-57s contents.

So, when we started discussing S-101 it was the encoding of the data which particularly interested me in the context of S-63. It is this area that I think we have the best opportunity for TSMAD and DPSWG to work together to define a data standard which will be able to give the end user a useable product. I have suggested through the product specification draft that "encoding" is defined within the standard as an example only and that the mechanisms for commercial distribution of data are defined elsewhere. I think it's highly likely a future S-63 (whatever it is called) will use many of the elements within a default encoding specified in S-101 but we need to ensure that the standards allow for redefinition to allow a smoother and more efficient supply chain and to give service providers maximum opportunity to add value.

Additionally, trying to "future-proof" the standard as much as possible is very important. If future encodings are required (e.g. protocols for online distribution) then, if they don't change the underlying data, there's no need to completely revise the standard. An approach using annexes for different encodings would smooth the process considerably. The best product specification for S-101 would allow a future S-63 to build on an example encoding all the mechanisms and structures needed for data distribution without having to layer everything on top. That way data can be optimized for ECDIS import and redundancy can be reduced.

What's also important to note as I boldly state that encodings should be refined and altered is that the key requirement of data "originators", mostly hydrographic offices', is for data integrity and this is the biggest hurdle to get over if we are to allow any tinkering with encodings. It is a clear requirement from the hydrographic office community that data should not be "changed" as it passes through the distribution chain but rarely have we defined either what the "data" is or what we mean by "changed". Bearing in mind that ALL ENC data is ultimately re-formatted into SENC I think a goal for S-101 should be to identify what ENC "data" is, to my mind it's the set of objects, attributes, geometric primitives and metadata that are grouped together into a cell. It's absolutely NOT the content of an iso8211 file. From a data integrity perspective I believe if an encoding is able to reproduce with 100% precision the data content of a cell then the data has not been "changed" - it may have been "transformed" but the underlying data content is the same.

If we can get this concept into our standards and our data distribution methodologies then we open the door to much more flexible data encodings and distribution models whilst retaining our core requirement of un-"changed" data passing through the chain.

Encodings and data size.

There may be many other benefits in taking a wider view on encoding and identifying ENC data with objects rather than files. Whilst I was writing this paper the first of the Chinese ENCs has been released for commercial sale. These ~250 cells represent probably the last of the BIG countries to come online and the growth of numbers of cells in commercial ENC services will probably enter a plateau at around the <9,000 mark. What's significant is that now we're in a position to "profile" all ENCs of the world, in short to consider the "world dataset" and try to define future standards based on the profile of those ENCs rather than the other way around.

The other concern for me as a data distributor is sheer volume of redundancy in the supply chain. I send out multiple CDs of all the data I have and then every week I send out new editions (where there may only be half a dozen changes in some cells), text files which I've already sent out and updates which are marvelously small and compact (sometimes!). But I would encourage us to think about what updates actually are – they're updated objects and we should be asking if all the data we send is actually required by the ECDIS itself. From a service perspective it does make sense to blindly forward on the data as received as it's simple but the volume of repeated data is staggering and does make you question whether there is a better way to reduce size and increase efficiency at the expense of a more complex life for the data distributor.

Loading time is still an issue for ECDIS users using S-63 and whilst a neutral transfer standard will probably never approach the speed of raw SENC loading we need to be mindful of the effect that our ENC encoding practices and distribution policies have on the end user. It was discussed at length at DPSWG by one of the ECDIS manufacturers that certain object attribution could dramatically reduce the amount of work every ECDIS needs to do when converting objects to SENC and I believe this is an area we need to look into more closely, matching up generic ECDIS requirements with ENC compilers' abilities to structure cells at source. Some of this ECDIS "pre-processing" could be done at the S-63 level as well prior to its onward distribution to end users.

Is it possible to achieve more efficient ENC encoding? Well, all these updates are sent out as "files" still, rather than looking at them as objects and we send out an enormous number of duplicates using this methodology. Additionally they're all iso8211 encoded at the bottom level and whilst iso8211 is great at describing data within a single file its self-describing nature is repeated in every single cell. iso8211 also uses a "catalogue" like structure at the beginning of every record with the names of the fields and their length (even though the standard tells you the sequence of fields to expect in each record). What this leads to is a large amount of data within an iso8211 file which isn't "ENC data" – it's header, catalogue, fields which describe length and pointers to other parts of the file. I did some analysis of some GB cells and got the following results:

Cell Name	Records	Total size	Data Size
GB203552.000	269	533,152	378,010
GB203593.000	1,987	327,800	214,897
GB203596.000	4022	666,528	437,442

As you can see from the table above that 66-71% of these cells is data, the rest is record overhead!

It would be interesting to see what results could be gleaned from the current world ENC portfolio. I actually had a stab at this and got similar results. Of the 24,424 base and update cells which constituted the last UKHO AVCS base reissue about 8,000 of them were base cells. These on average contain 58% data. The remaining update cells contain about 23% data only. At the very bottom of the pile are updates with only one record. GB5A0052.002, a cancellation cell, contains 3% data only in a file containing 1,887 bytes. There are, however cells with the opposite effect GB200760.001 is 97% data. This is an update which added a small number of features with a high geographical content[1].

Our service is currently 9 CDs worth and comprises about 40,000 folders to hold all the data. This will never make the transition to full scale online usage though as the underlying data is simply too large in its encoding. Once data is encrypted it effectively turns into a random byte sequence which can not be compressed any further so I believe in order to make the online transition we need to open up S-63 to the underlying encoding of the data and allow development of better compression algorithms or we will be stuck with online services unable to process the enormous quantities of "new edition" data which we send out weekly.

Service providers, particularly SENC distributors will tell you that compression is ALL about data formats = data encodings and they would be right! The trick with data compression is squeezing the same amount of information into a smaller space by using predictive modeling and coding methods. The amount of information (a mathematically definable quantity) is identical (at least, it should be!) in all possible ENC compression methods - there is a (similarly mathematically definable) minimum size for encoding ENC data but I don't think we're anywhere near that at the moment for two reasons:

- Our current compression method compresses the entire cell and the "data content" of a cell is not precisely the same as its "file content".

- None of our compression methodology uses the ENC-ness of the data to achieve any data savings. We basically used an off the shelf compression algorithm and applied it blindly to the entire cell content. We now know that ~70% of a cell's content is concerned with the representation of geographical coordinates but our compression methodology

[1] Worth mentioning in the S-63 context though is that the overhead is very compressible. The true size once compressed over and above the information-theoretic minimum is not as bad as the percentages suggest.

This is an issue I believe from two standpoints: (1) We are sending out enormous quantities of data which has a large amount of redundancy within it, i.e. data which isn't actually ENC data and (2) Our data integrity mechanisms and distribution policies mandate that this non-ENC data is maintained throughout the chain all the way to the ECDIS. This also illustrates that we can gain a lot of perspective from examining the world ENC dataset as a whole and that standards written defining one structure may not scale – we need other standards to allow us to scale up data distribution to the size of the world dataset and into the bandwidth we know is available to the global end user.

Summary

So, in summary: how can S-101 help DPSWG (and how can DPSWG help S-101) produce a better specification to feed the ENC supply chain? These are the recommendations I would make to both of our groups making standards for ENC distribution.

1. Don't tie the encoding down to a particular standard. It is better to allow a standard like S-63 to take elements of an encoding (catalog structure, file types, cell names) and build other mechanisms on top of it so that it's less of a layer model and more of a harmonised standard. S-101 started by saying encoding and data are independent and it is time to start reaping the benefits of that as we develop S-101 further.
2. Be clear as to what the "data" is. Distribution of ENC is completely 100% based on "cells" as "files". In the sense of the standard though a cell is just a collection of data objects. We need to be clear what the ENC data is, precisely and get away from the suspicion and fear that surrounds "changing" of files.
3. There needs to be a fresh look at S-63 in conjunction with TSMAD definition of S-101 in order to come up with a harmonized pair of standards defining ENC structure, encoding and copy protection. Many of the S-63 elements have proved their worth, others need refining but we're in a position to learn a huge amount from our experiences as data distributors.
4. Ensure the structure of datasets is sensible and scalable and able to represent the world dataset with as little repetition and redundancy as possible.
5. Be media independent. We shouldn't make any predictions as to what media data will be embedded within, whether that's hard, soft or programmatic. We need to concentrate on