TSMAD 23-4.5.11

Paper for Consideration by TSMAD

SI/SD impact review

Submitted by:	UKHO (Tom Richardson) & Jeppesen (E Mong, R Malyankar)
Executive Summary:	SI/SD introduces many added benefits to S-101 ENCs but is a changed concept from the S-57 ENCs. This paper attempts to highlight the benefits and impacts of SI/SD.
Related Documents:	S-101 Product Specification
Related Projects:	N/A

Introduction

The idea of SI/SD was first specified at TSMAD 20 (TSMAD20-16A[1]) in Rostock, Germany. Since then a few rounds of discussions have taken place due to some worries about the impact of SI/SD in terms of costs associated with adopting the concept. This paper seeks to address those concerns and show that the costs will be manageable and justifiable.

SI/SD is here proposed as an optional and optimal last step in the S-101 adoption process. The nature of the SI/SD concept is fully harmonized and consistent data in both the horizontal and vertical planes of the S-101 data stream. Given the envisioned role of S-100 in eNavigation, and S-101 as a core layer in the IMO concept of INS, this highly increased data quality will become a requirement for the data to function in the way the users expect.

1. Benefits

It is expected that SI/SD will introduce a number of improvements to vertical and horizontal consistency in ENCs. This effect is expected since it will only be possible to have one position for SI features, which will require that SD features and attributes (to ensure both geometry and properties are consistent) follow the same accuracy to ensure accuracy across scales. In S-57 it has been possible to have as many positions for a given feature as there are usage bands the feature is used on. The improved accuracy of the navigational data is seen as the single most important benefit of the SI/SD concept.

With SI features being stored only once, there will be an overall reduction in data volume for end users. The reduction will depend how many features are migrated into SI, with an expected decrease of data volume with an increase of SI features.

An informal review by Jeppesen Russia of Notice to Mariners have shown that 60-80% of all corrections by a hydrographic office were of an SI nature. Assuming this applies to all hydrographic offices, updating efforts within the SI/SD concept should be reduced for both HOs and user, as 60-80% of all corrections will only have to be applied once, as opposed to all applicable scales in the flat dataset[2] regime.

[1] http://88.208.211.37/mtg_docs/com_wg/TSMAD/TSMAD20/TSMAD20_DIPWG2-16.1A_SI-SD.pdf

[2] "Flat datasets" is the same as LayerID refers to as complete dataset, but is considered as a more appropriate term.

Current approach

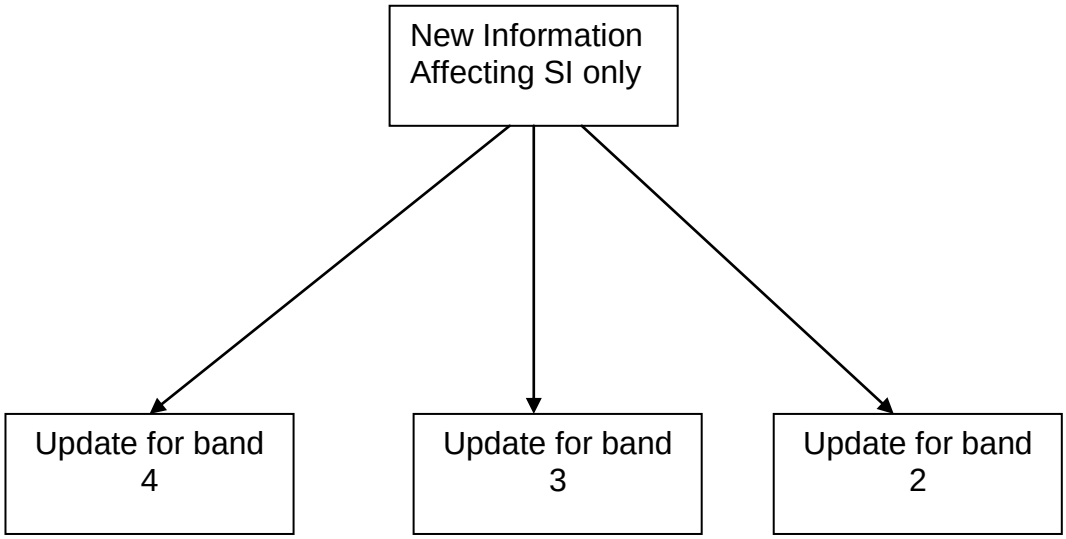

Figure 1

SI/SD approach

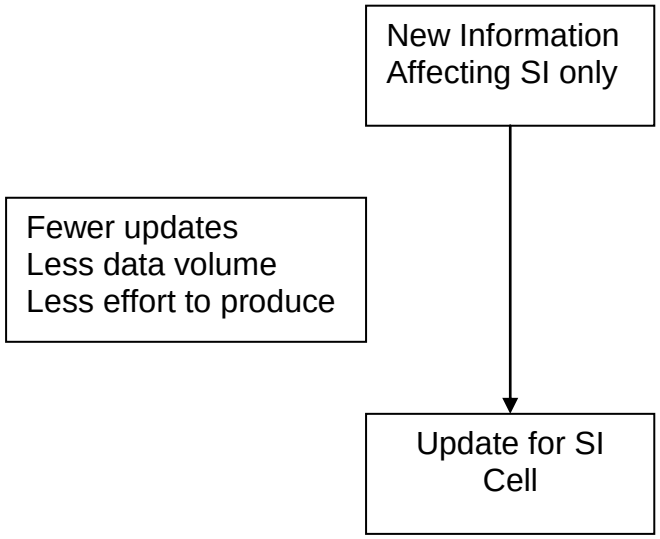

Figure 2

SI/SD increases readiness for eNavigation as concepts of data streams will become more apparent with eNAV, structures and principles in SI/SD can be reused and will allow for easier interaction with non-IHO data as harmonization issues will be reduced due to the store once use multiple times principle that SI/SD introduces.

SI/SD is the next evolution in hydrographic data and a potentially good showcase for the possibilities that S-100 has introduced over S-57.

1.1 Data Volume Comparison

The Fujairah test case presented at TSMAD 22 was limited in its application of SI features to the SI dataset. Only a few significant features in the areas covered by the Usage bands 4 and 5 datasets, were moved into the SI dataset and the test case therefore does not show the complete volume reduction that can be expected. However, volume reduction was still achieved.

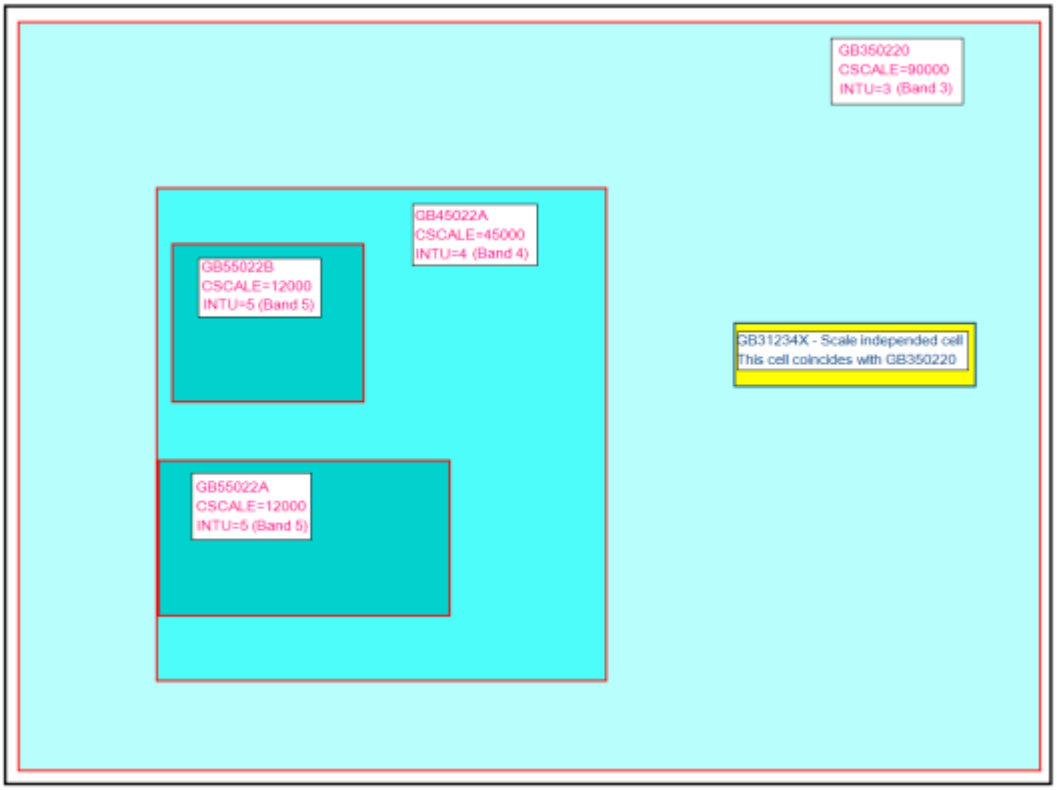

Figure 3 – Fujairah test case coverages

A comparison of the Fujairah test case is shown in Table 1. The test case has two ports where the SI features were moved into an SI dataset along with a number of SI area features. This reduces the data volume of the SD cells by around 10% but the SI cell results in overall data volume which is the same. The real difference is for updates, when only SI data is affected data volume is substantially less using SI/SD. Given that the majority of updates are for SI features by extrapolating this over all ENCs a significant reduction in update data volume can be achieved.

usage band	Original with ER applied	SD	SI
			82
3	342	320	
4	316	288	
5A	219	200	
5B	144	132	Difference
	1021	940	+1
Update 1	12	0	4
		Difference	8

Table 1

Data volume comparison is complicated at best due to a multitude of factors and scenarios. For example, files made in different production may have different file size; another factor is what files the ECDIS has on board. In the extreme case a user may need only one SD dataset and also therefore need the whole SI dataset. In such a case there could be a data volume increase due to the SI/SD concept. However, it is assumed most users will need coverage based on a waterway or region and overall data volume savings are expected for the majority of users, particularly for updates.

2. Impacts and Costs

Implementing SI/SD will require hydrographic offices to re-scheme and improve data, see 2.1. The level of re-scheming needed is estimated to be fairly individual for each hydrographic office depending on their own desire to adhere to the proposed rules for re-scheming. The majority of the impact will come with the requirement to harmonize data in the vertical plane. However, this effort is seen as an effort that any hydrographic office desire to do regardless.

Production systems are moving towards SI/SD, or a variation thereof, as this is required to assist hydrographic offices in the work load management. Implementation of such production systems will over time usually results in a single harmonized model of the area of responsibility, which is then use to generate data streams (paper charts, ENC, coastal zone management, etc). For SI data this will result in a concept of “maintain once and use multiple times”, which will reduce maintenance efforts and risk of quality issues due to data inconsistency.

The impact on users from SI/SD, in isolation, is considered fully positive as SI/SD data will be better harmonized and consistent. Moreover, a reduction in data transmission is expected as the base data and update volumes will be reduced. Apart from the improvements in data consistency, SI/SD should be “invisible” to the end user. Implementation of S-100 support is considered a necessity, regardless of SI/SD, and therefore SI/SD is not a direct cost impact on the user.

2.1 Re-scheming example

Below is a fictional example of how re-scheming can be done in the British part of the English Channel. The large SI datasets have coverage corresponding approximately to major sea routes or waterways and can be populated gradually as underlying datasets are adapted to the SI/SD concept.

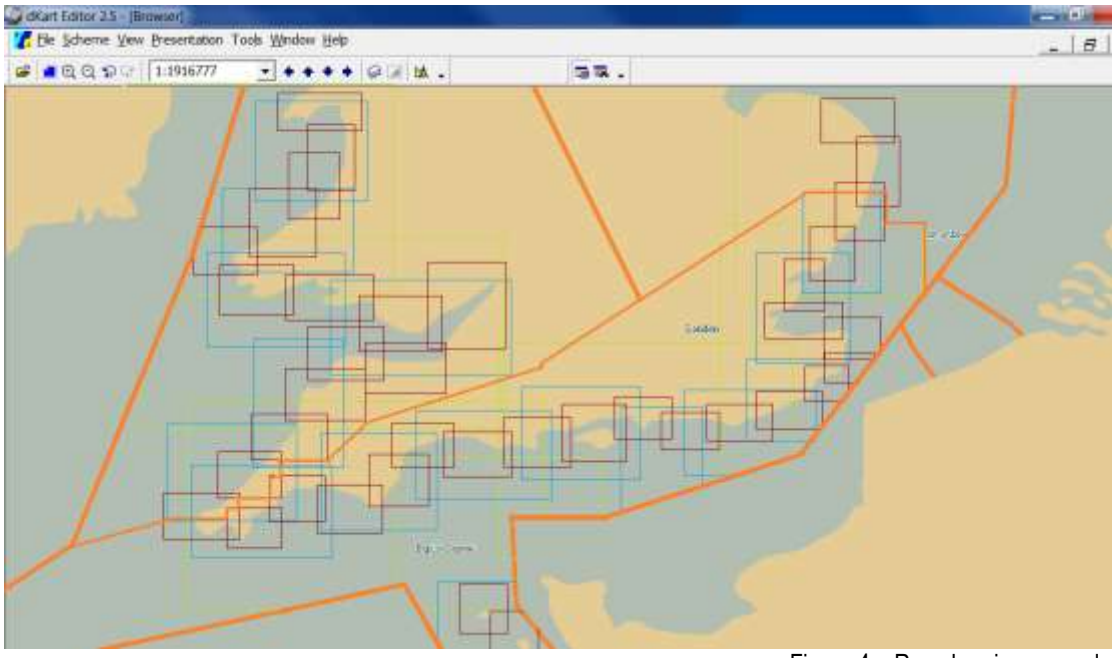

Figure 4 – Re-scheming example

2.1.1 Guidance for scheming

- It is recommended to plan SI coverage based on sea routes or water ways. This will reduce the applicable updates to only the area of interest. E.g. the British coast can be divided into 4 major water ways; The English Channel, The Irish Sea, The North Sea, and Scotland and Northern Ireland. In this example updates affecting Scotland will not be applied to the SI cell for the English Channel and vice versa. As such a mariner transiting the English Channel will only receive data traffic that is applicable to his area of interest.
- Countries with a short coastline may wish to take a different approach and utilize only one SI cell for their entire area of responsibility (e.g. Belgium).
- SD cells must be covered by one or more SI cells. This allows producers to retain a paper chart based SD cell layout.
- Implementation of SI/SD scheming in a waterway or region can be done en masse or gradually. For example, a hydrographic office may replace all flat datasets in an area with complete SI/SD coverage, or may choose to do this gradually by extending the SI/SD coverage as the vertical data harmonization is being done by terminating flat datasets and issuing new edition SI dataset to extend the coverage of a SI dataset and new SD datasets to replace the flat datasets.
- SI/SD coverage must form a vertically complete coverage[3] as well as horizontally complete coverage[4] before it can replace the coverage of flat datasets. This means that flat datasets in all scales in an area, set to be transitioned into SI/SD, must be migrated together, as opposed to retaining flat datasets for some scales and using SI/SD datasets for other scales.
- Annex B contains an example of how gradual transition from flat datasets to SI/SD datasets can be done in accordance with the above guidance.
- In some parts of the world there are few local resources to produce navigational charts and one or more foreign authority has assumed this responsibility. Where multiple foreign authorities are involved in producing coverage for an area, it is recommended that a dialogue between all involved parties is started to resolve responsibilities.
- If it is not possible to resolve responsibilities, and coverage of an area is produced by more than one authority, ECDIS should have a possibility for the user to select which data has priority.

2.2 Data flow diagram for production and distribution

The dataflow is not expected to be dramatically different from today. There will be challenges in the transition period as the coverage is developed. Mainly this will be on workload and will be similar to when S-57 was introduced in that the coverage of S-101 is built gradually. S-57 ENCs will have to be maintained simultaneously with S-101 flat files and SI/SD.

[3] "Vertically complete coverage" means coverage in all scales in a region.
[4] "Horizontally complete coverage" means complete data coverage in a region regardless of scale.

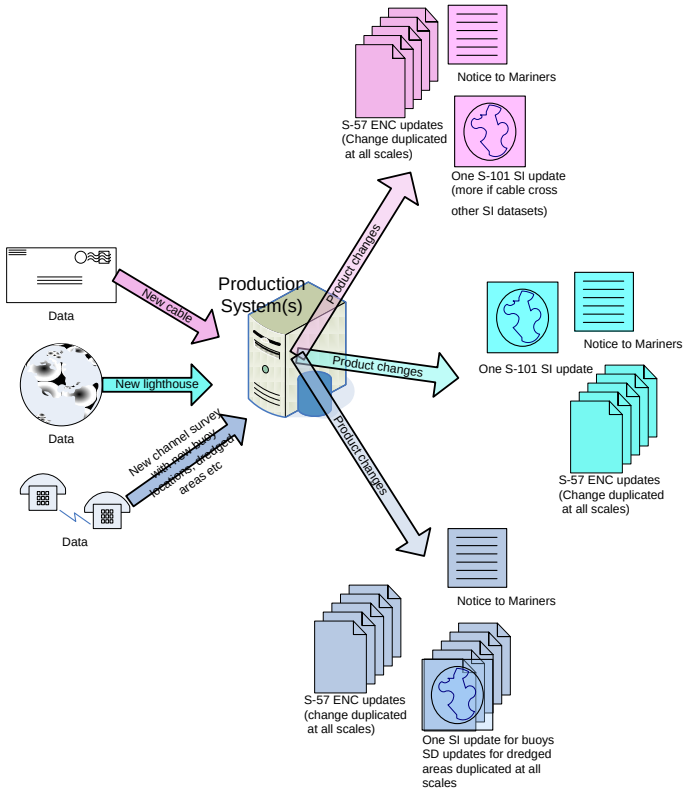

Figure 5 – Production flow

For most hydrographic office, the flow is expected to be as today, with some hydrographic offices distributing only through RENCs, others via service providers and others still may choose to do both or even direct to end user.

Packaging data for the end user is something the service providers already do and know well. The impact on service providers is a new data stream that needs to be accounted for in the current processes. eNAV is expected to introduce more data streams, therefore a scalable distribution process is seen something all service providers need to have in place.

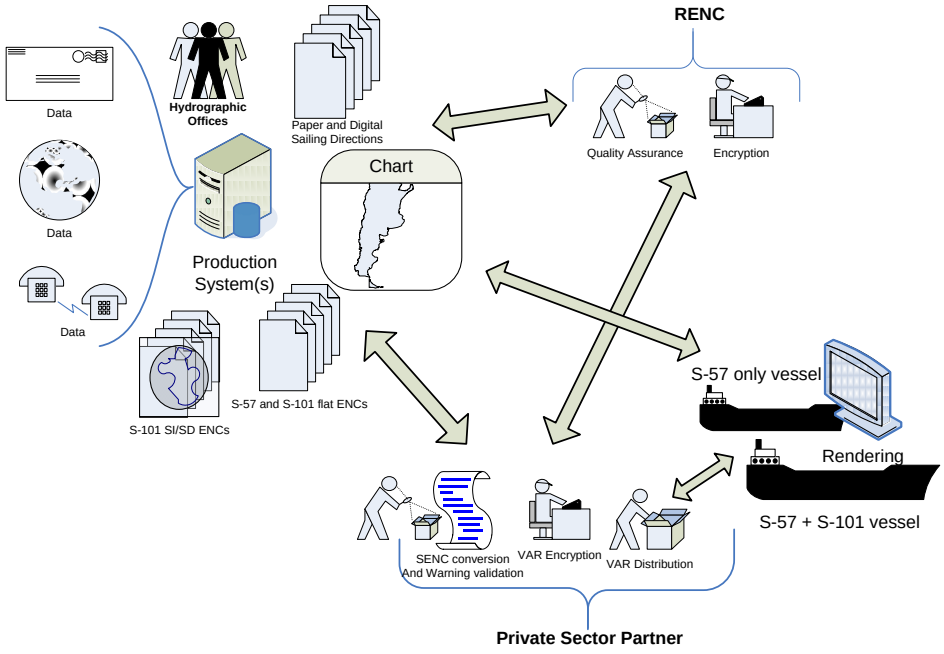

Figure 6 - Distribution flow

3. Impact

Implementing SI/SD will be a challenge to all hydrographic offices. The level of difficulty in that challenge will however vary with each hydrographic office depending on a variety of factors, such as production system, length of coast, staffing, etc. To move to S-101 HOs will have to adapt their production to some extent, so the change to support SI/SD would be covered by these inevitable changes.

SI/SD introduces the store once use multiple times principle to distributed data beyond the production system. This will enable a more direct correlation between distributed data and data in the production system, and possible reducing the labour needed during export.

The SI/SD concept is still some ways off in terms of being ready for specification in S-101. For example there is a need to develop processes for how SI data interact with SD data. This is a requirement that will arise from other data streams too (NPubs, eNAV, etc), and processes established for SI/SD data can be re-used by those data streams.

3.1 SI/SD processes

There is a need to explore options for how SI/SD data interact. This includes how portrayal work when for example wrecks with unknown value of sounding draw their portrayal rule in part on the underlying depth area. If the wreck then is an SI feature and the depth area a SD feature, will it still work and if not what is the mitigation.

More complicated is how area features interact. For example in a TSS, the inshore traffic zone is limited on the coastline. How will this work in a scenario where TSS is an SI feature and the coastline is an SD feature. Here multiple geometries was considered but became too problematic due to the assumption that multiple geometries would require stronger linking in data between specific features in different data streams, meaning if one change, the other must also change, and possibly creating a cascading problem. Particularly if the data stream is outside the control of an HO as will be the case with eNAV. An alternative that seem more feasible is for the linking to be done in the SENC by a “cutting process” where inshore traffic zone is cut by a land area, meaning the binding is a geographical co-location link.

Examples of SI data that directly interact with SD data; TSS (particularly the inshore traffic zone), wrecks without VALSOU draw portrayal VALSOU from DRVAL1 of underlying DEPARE. NPubs, AML, etc.

Examples of SI data that indirectly interact with SD data; buoys marking a channel, light range (leading lights) with navigation line and recommended track, beacons on a clearing line.

4. Rules

A draft set of rules for SI/SD datasets have been provided in Annex A. The rules cover metadata, exchange set packaging, dataset, production and ECDIS application rules for each dataset type. These rules are extendable and adoptable for non-ENC data streams, such as NPUbs. Non-IHO data streams may also make use of these rules as an informative draft to draw upon when establishing rules for the specific data stream.

5. Packaging

Packaging of SI/SD exchange sets will differ from flat datasets only by the addition of SI/SD data sets. Special file naming conventions for distinguishing datasets as SI, SD, or flat datasets may be convenient for production and applications but are not strictly necessary because applications can use the metadata. The dependency rules for S-101 data in Annex A address whether areas where SD data is provided also have the appropriate accompanying SI data and must be enforced when the exchange set is deployed or used, e.g. verified by the ECDIS by comparing coverages when the exchange set is added to a system.

Other products such as nautical publications may have different dependencies. Product specifications must describe these dependencies and provide rules for validation.

5.1 A hypothetical SI+SD exchange set

On each volume within an exchange set there must be a root directory called ENC_ROOT. The catalogue file for the exchange set must be in the ENC_ROOT directory of the first volume of the exchange set. The ENC_ROOT directory of the first volume may also contain a README file, containing ASCII text. Further directories and sub-directories may be defined under the root directory on any volume in the exchange set. The following example shows an example directory structure:

Directory ENC_ROOT
Files:
CATALOG.101 Catalogue file
NL60002100.000 SD cell
NL60002100.001 Update 1 to SD cell
NL60002100.002 Update 2 to SD cell
NLAA001A0X.000 SI cell
NLAA001A0X.001 Update 1 to SI cell
NL1234560X.TIF Support file accompanying SI dataset
NL12345600.TXT Support file accompanying SD dataset
README.TXT README file

6. End user verification of data needed for voyage

Today most users trust in service providers to provide the needed data at the needed time. This situation is expected to continue with SI/SD data.

The end user will not need to do any extra work if the packaging rules ensure that SD data is accompanied by the appropriate SI data and verification checks by producers and application developers implement these rules. Users of SI/SD datasets should need to perform the same end user verification activities as for “flat” datasets - for example, checking shipboard media and licenses, and confirmation by checking availability, license activation and loading in the ECDIS/ECS. The end user will only need to obtain and check the ENC for a specific zone or area, whether the ENCs are SI/SD or flat datasets.

The rules for applications should ensure that chart catalogues and database lists or licensing screens do not require the user to cross-check SI data against SD data (to reduce user effort). They must also not allow activation of SD data in the absence of accompanying SI data (or at the least, raise a warning if such activation is allowed).

7. Non – VAR distribution

Non-VAR distribution would use the same channels as today, e.g., posting on a Web site. The difference would be in the packaging of exchange sets for new editions or re-issues, which would include both SI and SD datasets in an exchange set package. Tools or business processes for creating exchange sets can implement the packaging rules in Annexe A – Rule 2 in the packaging rules in particular ensures that SD data is not distributed without the corresponding SI data. Tool implementation of this rule might consist of comparing SI and SD coverages in an exchange set. It is also possible to use the coverage metadata for each SI and SD dataset to construct a cross-index which can be used in graphical, web, or text catalogues of datasets.

8. Updating

There are no major changes expected for how updating of data occurs. Figure 5 – Production flow, demonstrates roughly how some types of updates result in only one SI update as opposed to multiple updates for S-57 files. Other scenarios generate updates to both SI and SD datasets, such as the example of an improved channel. Assuming the informal review of Notice to Mariners mentioned in chapter 1 – Benefits, is applicable to all hydrographic offices, there will be an overall reduction of updates in the SI/SD concept, as opposed to S-57 ENCs or the flat dataset concept of S-101.

9. Problems being solved – Use case

It is the authors' opinion that the paper has demonstrated a significant use case for the SI/SD concept. A summary of the major benefits of including SI/SD in S-101 as opposed to S-101 with just flat dataset is as follows;

Improved consistency of data – Store once, use multiple times for SI data will demand strict rules that SD data must adhere to forcing greatly improved data consistency both vertically and horizontally.

Reduced data volume and updating effort – Store once, use multiple times means SI data that in S-57 are replicated across usage bands will be captured once resulting in an overall data volume reduction. Furthermore, updates of SI features require only one update, as opposed to S-57 where such updates are needed in all datasets where the SI feature is present.

Readiness for harmonization with NPubs data – Methods and processes developed for SI/SD data will be transferable for how NPubs data act with ENC data. Furthermore, NPubs data need to be harmonized and integrated with ENC data; else there will be significant data duplication. SI/SD enables this integration.

Readiness for eNAV – Similarly with NPubs, eNAV data will need to be integrated with ENC data and the SI/SD concept help facilitate this by allowing methods and processes developed for SI/SD to be transferred to other data streams.

SI/SD and related benefits as the major selling point of S-101 - It is considered that better data and the ability to update at sea (less update data volume) are of significant benefit to the user.

10. Pricing and distribution

Pricing and distribution of S-57 ENC is a complicated matter with great variation in pricing and methods of distribution between the hydrographic offices. It is not expected that SI/SD will change that.

11. Actions for TSMAD

- Review and discuss this paper
- Endorse the paper and confirm that SI/SD will be included in S-101
- Consider establishing a sub working group to elaborate on the SI/SD concept for S-101 version 1.0.0

Annex A

Candidate Rules for SI/SD datasets

Dataset metadata element *layerId* describes the type of dataset (SI/SD/complete). This element is defined in the S-101 draft.

SCAMIN and SCAMAX attributes may or may not be coded for SI features. If coded, they indicate the limiting scales at which the feature is expected to be displayed. This can be used to avoid situations where small scale charts display minor lights. Alternatively global display rules can be formed to move more of the display priority processing over to the ECDIS.

Metadata rules

1. Dataset metadata elements optimumDisplayScale for SI datasets shall be set to 0.

2. Dataset metadata elements maximumDisplayScale, minimumDisplayScale for SI datasets shall be set to the extreme values of maximum and minimum display scales respectively, specified in S-101 if SCAMAX or SCAMIN respectively are not encoded for any SI feature[5]. If all SI features have SCAMAX encoded metadata element maximumDisplayScale can be set to the highest SCAMAX value. If all features have SCAMIN encoded metadata element minimumDisplayScale can be set to the lowest SCAMIN value.

Packaging rules

1. There can be zero, one, or more than one scale independent cell contained within an exchange set.

2. The coverage of a SD cell in the exchange set must be within the coverage of the scale independent cells in the same exchange set.

3. (File naming rule for SI data was considered but is not technically necessary because the discovery metadata designates files as SI/SD. Convenience during production and maintenance may be improved with a naming convention which distinguishes between SI/SD.)

Data set rules

1. A scale independent cell shall not contain any meta features except for Coverage (which shall be provided) & optionally “Navigational system of marks” (M_NSYS).

2. Rules must be determined about which features within a scale independent cell must have the feature attributes scale minimum and/or scale maximum encoded.

3. Scale dependent cells with the same display scale may overlap. However, data in different SD cells must not overlap. Therefore, in the area of overlap only one dataset may contain data. All other cells must have a meta feature DataCoverage with categoryOfCoverage = 2 covering the overlap area. This rule applies even if several producers are involved. There must be no overlapping data of the same scale, except at the agreed adjoining national data limits, where, if it is difficult to achieve a perfect join, a 5 metre overlapping buffer zone may be used.

[5] According to the S-101 draft the extreme values are 1 and 13. This rule might improve performance, by acting like an “envelope” or bounding box for a dataset, along the ‘scale dimension’.

4. Scale independent cells may overlap but data in different SI cells must not overlap. Therefore, in the area of overlap only one dataset may contain data. All other cells must have a meta feature DataCoverage with categoryOfCoverage = 2 covering the overlap area. This rule applies even if several producers are involved. There must be no overlapping data of the same scale, except at the agreed adjoining national data limits, where, if it is difficult to achieve a perfect join, a 5 metre overlapping buffer zone may be used.

5. Scale dependent datasets shall not contain any of the feature classes designated as scale independent features and vice versa.

Production rules

1. A producer responsible for the population of a scale independent data set will be obligated to produce scale dependent data sets of the same area.

2. Producers shall harmonize updates to scale-dependent spatial objects at one scale with the corresponding scale-dependent spatial objects at other scales.

3. Producers shall validate updates to scale independent (dependent) features against updates to scale dependent (independent) features, and issue all related updates at the same time. Consider the possibility of a cascade effect:

    a. An update to SD dataset A at scale X moves the coastline.

    b. This puts a scale-independent water feature on land.

    c. The coordinates of the water feature are corrected in the SI dataset.

    d. Theoretically this could put it on land for a SD dataset at a different scale.

    Rule 2 above mitigates this problem.
4. Producers shall validate updates to SI features in the context of all SD features directly or indirectly related to the updated feature[6].

Application rules for ECDIS

(Severity of errors and ECDIS behaviour in case of errors TBD.)

1. SI data outside the coverage of available SD datasets shall be clipped (not displayed).

2. Applications shall verify that all coverages of loaded SD data are also inside the (total) coverages of loaded SI data.

3. It shall be an error for the coverage of an SI dataset to overlap the coverage of a dataset coded as “complete” in the dataset metadata.

4. ECDIS may discard datasets in an exchange sets that already exist in SENC if they are the same edition and have the same update number.

5. If datasets from different producers have overlapping coverage of an area, ECDIS shall make it possible for the user to specify which dataset is displayed in the area of overlap. This includes the situation where one producer supplies SI/SD data and the other does not.

[6] It will be necessary to define which SI and SD features are likely to be related, not just those at the same location. The DCEG might do some of this but there is still the possibility of commonsense rules which the DCEG does not explicitly state, like “no buoys on land”.

6. An indication of the presence of overlapping data from a different producer shall be displayed on the screen at user option.

7. Chart catalogues and database/license screens shall not require the user to cross-check SI datasets against SD datasets in order to verify the availability in the system of all of all data needed for a voyage.

Annex B – Example of gradual transition to SI/SD coverage

This is an example of how a gradual transition from flat datasets to SI/SD sets can be done in an area. In this example the flat datasets have been converted from S-57 and therefore have the no coverage areas.

1. Initial status

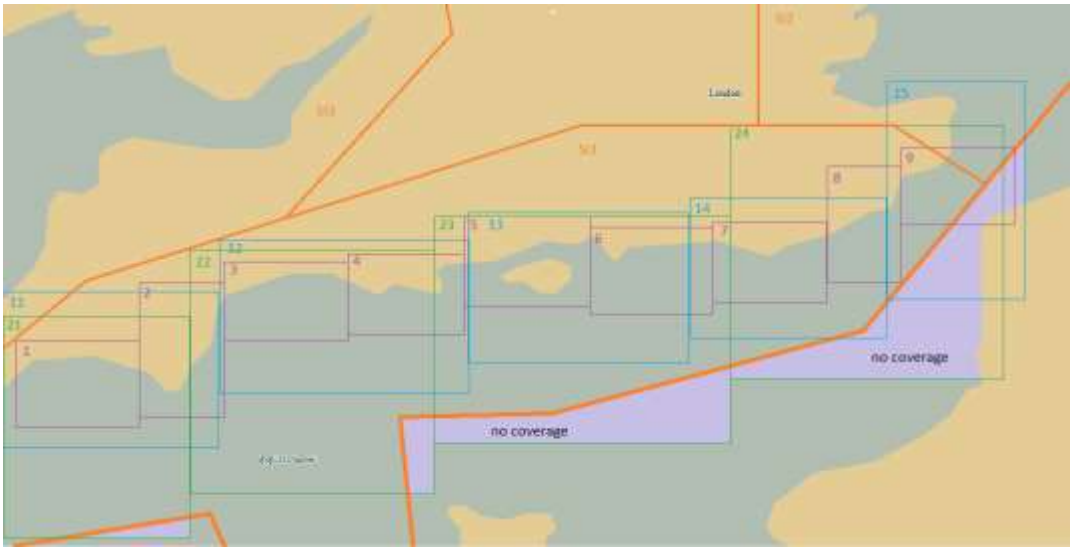

Figure B1

Initially there are 18 flat datasets in different scales and locations within a region planned to be the coverage of an SI dataset named **SI1**. These flat datasets are **1,2,3,4,5,6,7,8,9,11,12,13,14,15,21,22,23** and **24**.

2. First transition

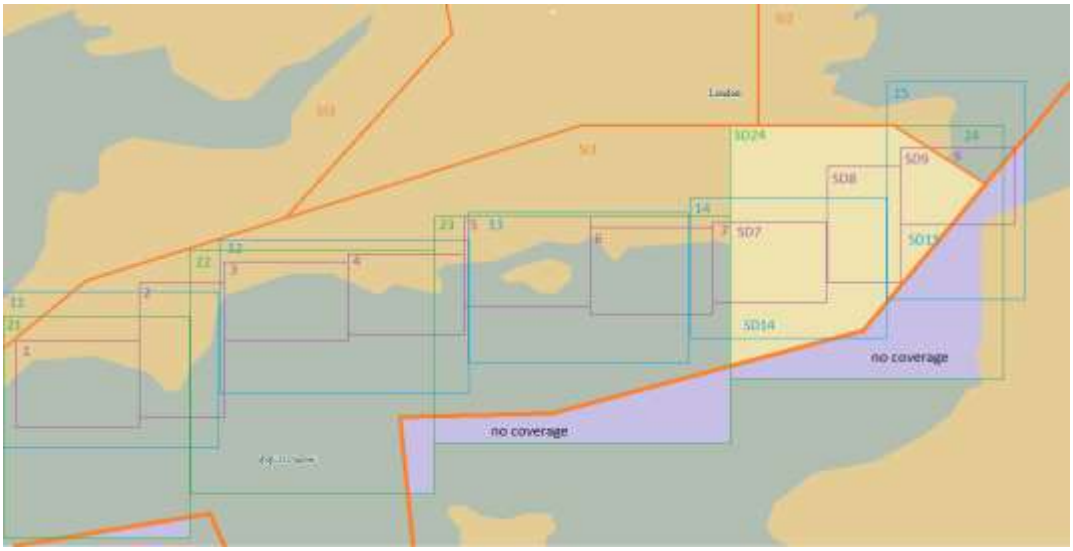

Figure B2

In the first transition, flat datasets **7, 8, 9, 14, 15** and **24** are terminated. And flat datasets **7, 9, 14, 15** and **24** are issued as new editions with coverage excluding the coverage of **SI1** transition 1. Then new SD datasets **SD7, SD8, SD9, SD14, SD15** and **SD24** are issued along with new SI dataset **SI1** with coverage of transition 1. Note that flat dataset **8** remains terminated, as its coverage is completely replaced by **SD8**.

3. Second transition

Figure B3

In the second transition, flat datasets **4, 5, 6, 7, 12, 13, 14** and **23** are terminated. And flat datasets **4** and **12** are issued as new editions excluding the coverage of **SI1** transition 1+2. Then new SD datasets **SD4, SD5, SD6, SD12, SD13** and **SD23** are issued. Datasets **SD7** and **SD14** are terminated and new editions issued that include the coverage previously covered by flat datasets **7** and **14**. **SI1** is also terminated and issued as a new edition with the extended coverage of transition area 1 + transition area 2. Note that flat datasets **5, 6, 7, 13, 14** and **23** remain terminated, as their coverages are now completely replaced by SI/SD coverage.

4. Final transitions

The SI/SD coverage is then completed through two additional transitions (transition 3 and transition 4) to finally conclude with a full coverage of the planned **SI1** area. The flat datasets that previously provided partial coverage of **SI1** coverage area remain as flat files till they are absorbed by additional SI coverage (**SI2** and **SI3**).