

Paper for Consideration by S-100WG

S-101 Producer codes and Naming convention

Submitted by:	PRIMAR/IC-ENC
Executive Summary:	Limitations in producer codes, file naming duplication and ENC update limitations are discussed. A proposal to S-100 products naming convention is presented.
Related Documents:	S-100WG2-8.5A, S-100WG2-8.5B, S100WG2-11.6
Related Projects:	S-100, S -101, S-1xx

Introduction / Background

S100WG2 meeting discussed limitations in producer codes (S-100WG2-8.5A), S-100 naming convention (S-100WG2-8.5A/S100WG2-8.5B) and ENC Update limitations (S100WG2-11.6)

As result of discussions PRIMAR and IC-ENC were given the action 08 (reference 8.5): RENCs to further investigate the use of Producer Codes and file naming conventions for S-101, including file name duplication.

As a result of discussions PRIMAR and KHOA were given action 45 (reference 11.6): Draft a recommendation paper for the next meeting on file naming convention, taking into consideration problems associated with ENC Update Limitations.

Analysis/Discussion

Producer codes, file naming conventions and possible Update limitations are important elements in the discussion of a dataset file name. From the RENCs opinion, there are huge advantages in having human readable file names instantly giving information about producing agency, type of product and (if available) intended (specific) usage. The effective S-57 file management operations provided by RENCs today is based on informative file names used as product identifiers in the data distribution services. Apart from the immediate recognition advantage, duplicate filenames and duplicate product identification are avoided. It is also of value for OEMs if they do some high level ENC file management without the need to inspect the content of the dataset discovery metadata.

Also, if enough of the file name characters are optional, this adds flexibility for the data producers when defining product naming schemes. More elaborate information is available in S100WG2-8.5B.

Producer codes

As paper S-100WG2-8.5A outlines, one challenge is when producer codes are extended due to several national producers. In S-57 this is now solved by replacing the second producer code character with a number.

E.g. United States has producer code US. Additional producers have code U1, U2, U3. As other nations also have producer codes starting with U (Uruguay UY, Ukraine UA), they will be given U4, U5, U6 and so on for their additional producer codes. The limitation for producer codes are then 9 variants under a country code first character (assuming restriction applies to only use numeric digits 1-9).

Extending the producer codes to 3 characters could be a solution. No producer codes contain the same two characters. Following the above example, NOAA would then have the producer code US1 and additional US national producers US2, US3, US4. However, the limitations for producer codes within one country would be 9 variants (assuming restriction applies to only use numeric digits 1-9).

Extending the producer codes to 4 characters could be another solution. Following the above examples, NOAA would then have the producer code US01 and additional US producers US02, US03, US04. The limitations for

producer codes within one country would then be 99 variants (assuming restriction applies to only use numeric digits 1-99).

For S-57 ENC production, having more than 9 national producers do not seem likely. For S-100 related products however, anticipating more than 9 national producers for a whole range of products may seem more likely.

Potentially, given the likely use of S-100 within e-Navigation there may be products with large numbers of “producers”. The examples of an aid to navigation or a vessel which outputs a data product such as a status report or route can be considered possible scenarios. In such cases it may be more appropriate to include the S-100 product identifier and a number of optional characters only

Update limitations

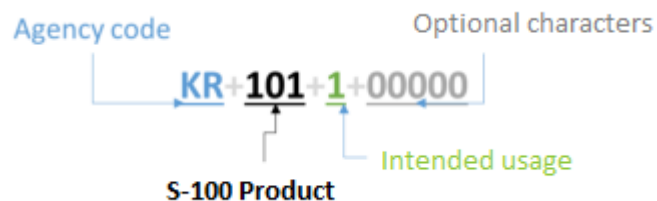
The issue with update limitations are related to file name extensions. If extensions are used to separate between products, problem occurs when an update extension number equals another products file extension.

E.g. An S-101 update file with extension .102 would be equal to an S-102 product file with .102 extension. Further elaboration is available in paper S100WG2-11.6.

If, however, the products are separated based on file name information, the extension issue is no longer a problem. Therefore, if the S-100 product name were to be included in the file name, the update limitation issue would be solved.

Naming convention

Several recommendations were discussed in S100WG2-8.5A. The alternative solution for recommendation 3c described file naming if Agency code information should be included in the datasets. A proposed solution included information on producing agency, specified S-100 product, intended (specific) usage and optional characters as outlined in the following figure:



If intended (specific) usage is not defined, the character 0 must be used.

This topic was further discussed in paper S100WG2-8.5B, and an alternative solution was proposed: KR(X)+101+1+000000000. Both solutions offer unique filenames for all S-100 product related datasets.

Using 2, 3 or 4-character producer codes and 5 or 10 optional characters give the following options:

- Option 1: KR+101+1+00000
- Option 2: KR1+101+1+00000
- Option 3: KR1+101+1+000000000
- Option 4: KR01+101+1+00000
- Option 5: KR01+101+1+000000000

Option 3 and 5 supports added flexibility for data producers with more optional characters available.

Option 2 and 3 supports 9 variants for additional producers.

Option 4 and 5 supports 99 variants for additional producers.

Our recommendation, to incorporate flexibility for the data producers and to prepare for up to 99 national producing agencies, would be option 5 as outlined in the following figure:



If intended (specific) usage is not defined, the character 0 (or another predefined character) must be used. Agency codes should start with 01, meaning the first producing agency in a producing nation have agency code XX01 (where XX is agency code according to S-62).

10 characters have been assigned for individual product definition. Characters A to Z, 0 to 9 and the special character _ (underscore) can be used for product definition. All 10 characters must be assigned a value.

Some examples of file names:

a)

Producing Agency: Korea (KHOA – KR01)
 S-100 Product: S-101
 Intended usage: 1
 Optional Characters: ID12345678
 File name: **KR011011ID12345678**

b)

Producing Agency: (United States additional producer)
 S-100 Product: S-129
 Intended usage: 3
 Optional Characters: QWERTYUIOP
 File name: **US021293QWERTYUIOP**

c)

Producing Agency: Finland (Finnish Transport Agency – FI01)
 S-100 Product: S-102
 Intended usage: N/A
 Optional Characters: YKSIKAKSIA
 File name: **FI011020YKSIKAKSIA**

Conclusions

It is important to have information on producing agency, specified S-100 product, intended (specific) usage and optional characters in S-100 file names to support future file management operations. Providing unique names to avoid file name duplication and product ID duplication is important. Carrying the S-100 related product information in the filename resolves the issue with update limitations. Extending Agency Code to 4 characters supports the possibility for having up to 99 different national producers and resolves the issue with producer codes.

Action Required of S100WG

The group is invited to discuss this paper and decide if proposed solution is acceptable for an S-100 naming convention.

The group is invited to consider this proposed solution for S-101 and seek further review and development with a view to including guidance in S-100. This should take into account the observation that this convention may not be suitable for all S-100 product types.