

Paper for Consideration by S-100 WG8

ECDIS dataset handling

Submitted by:	CCG (Eivind Mong) and Teledyne (Hugh Astle)
Executive Summary:	Consistent handling of datasets across product specs
Related Documents:	S-101 Annex D Dataset Loading, S-124, S-98
Related Projects:	S-100, machine readable generic products

Introduction / Background

One of the intentions for S-100 is that system implementation guidance should be handled at a common level so that an S-100 supporting system could in theory support any S-100 product based on the supported version of S-100 and the compliancy level aimed for by the system and service.

The topic of dataset loading and rendering have been ongoing, and recently, active in the S-101PT and dedicated sub working group. Having guidance about how an end system should handle datasets should not be part of an individual product spec because this approach will essentially amount to each product specification describing a custom implementation.

S-124 also includes some custom implementation guidance related to handling of datasets. Specifically, regarding the cancellation of datasets and to identify a list of active datasets.

The goal is to have such functionality defined in a common way to avoid product specific implementations.

Analysis/Discussion

Dataset loading by Scale

S-101 has been focused on how datasets will be loaded based on dataset scale properties. Other scaled products may also need the same guidance as to how the system chooses which datasets to load. Dataset loading logic should be described in a common place such as S-98.

Overlapping Coverage

S-101 is defined such that data at the same scale shall not overlap. This restriction may not apply to other products, especially if the portrayal of a dataset does not hide other data. Products such as S-123 may have overlapping content, even beyond national borders.

Scaleless data

Load all active (e.g., activated by a Predefined Combination from S-98) scaleless datasets that overlap the area of interest. Perhaps consider if there should be a loading or rendering order such as to sort by issue date to place the most recent on top.

Historic data

The logic should include how to handle historic datasets. If the system allows a user to set a viewing date/time then presumably only the datasets/updates active at that time would be included.

A dataset would be considered active if:

- The viewing date is after the dataset issue date and time.
- The dataset has not been replaced with a new edition at the viewing date.
- The dataset has not been cancelled at the viewing date.
- The viewing date is within the S100_DatasetDiscoveryMetadata/ temporalExtent.

S-124 Dataset model

In S-124 there is a 1:1 modelling of NAVWARN and a dataset. Each NAVWARN is issued as a new/separate dataset. Some reasons for this are:

- Easier to manage as separate documents, similar pattern to NAVWARNs prior to S-100, which simplifies backwards compatibility.
- An end system can minimize data usage by filtering NAVWARNs by geometry such as intersecting an area of interest or route.
- Take advantage of expected ECDIS behaviour to load/view datasets that are active.
- A fixed set of dataset boundaries could necessitate splitting NAVWARNs between datasets, resulting in partial information which is a significant deviation from current practice, which may impact STCW training.

In-force vs cancelled

A NAVWARN which is in-force is one that is active. It would not be in-force if it is not within a defined time period described in the dataset or when it has been cancelled by another NAVWARN.

The process of cancelling datasets is linked to the NAVWARN messages and legacy system behaviour. This means that datasets are only edition 1 and then eventually cancelled. S-124 has the following types of datasets drawn from this legacy;

- New dataset – new MSI information
 - In S-100 a new dataset for each NAVWARN
- New dataset self-cancelling – new MSI information with limited validity
 - In S-100 use temporalExtent to give self cancellation time.
- New dataset with cancellation – new information updating previously issued information and cancelling the previously referenced dataset(s)
 - In S-100 there is no clear means to refer to another dataset
- New dataset with cancellation self-cancelling – new information updating previously issued information, and with limited validity
 - In S-100 a new dataset can be self-cancelling through temporalExtent but there is no way to refer to other datasets that are cancelled
- In-force bulletin – list of currently active message
 - In S-100 a new exchange set or dataset can be made but there is no way to refer to other datasets that are still in-force

Investigation into use of S-100 Ed 5.x metadata

S-124 datasets contain the attributes to describe the behaviour needed, however the S-100 principle is that the metadata should be used to enable the system to know what behaviour it should perform in response to the data received, e.g. load, update, terminate. S-124 therefore tries to duplicate some of this information to simplify backwards compatibility but expects that over time and with experience some of this may be better streamlined.

The S-124 Project Team, responsible for S-124 development, have attempted to use S-100 metadata to facilitate dataset handling and have identified that a way to refer to other datasets would be useful to support actions such as identifying previously issued datasets that are cancelled or to list a set of datasets which are still active to form an in-force bulletin.

There is a way in S-100/S-101 to cancel a dataset by issuing an update which identifies the dataset as cancelled but there are some issues with this approach:

- How to encode an S-100 GML update dataset? Presumably it would be empty.
- How to define that a new dataset cancels previously issued datasets?
- How to issue a list of active datasets?

Cancel using metadata

It would be optimal to be able to identify datasets to be cancelled using metadata without having to create an update dataset file (undefined in GML) or having to deliver the same dataset again.

Part 17 17-4.4.1 S-100 states that a dataset can be cancelled using metadata but does not seem to describe how this would work.

“In addition to fileless dataset cancellation using fields in the Catalogue metadata file a dataset may be cancelled by the Data Producer by the issuing of a cancellation update”

The S100_DatasetDiscoveryMetadata has mandatory fields such as filename, signature and compression which would not be needed if an actual file is not included in the exchange set such as when a dataset is cancelled. See discovery vs delivery metadata below.

Association between new dataset and datasets which are cancelled by the new dataset

In S-100 part 17 S100_DatasetDiscoveryMetadata has fields to describe that a cancelled dataset is replaced by other datasets using names of the replacement. The attribute 'replacedData' Boolean indicates there is a replacement and one or more 'dataReplacement' strings can hold the names of the replacement datasets. This might be made a bit more structured by using an MRN as the name of the replacement dataset. It seems a little awkward or inverted compared to how a NAVWARN can cancel other datasets. The exchange set could include metadata of several cancelled datasets and new datasets. In some cases, a new NAVWARN might replace a previous one with new information or it might only exist to cancel previously issued NAVWARNs without replacing them.

It would simplify the NAVWARN use case if a new dataset could refer to other datasets such as a list of datasets to be cancelled by this new dataset.

In-force Bulletins

S-124 has an additional guidance relating to how an ECDIS system should automatically reconcile S-124 datasets. In addition to handling a dataset cancel event there is a concept in S-124 of issuing a bulletin of in-force notices which can be used as a 'fail safe' to cancel notices not in the list if a cancellation has been missed by some glitch in data delivery. The wording in S-124 is

Validity is also indicated by the NAVWARN being present in the latest in-force bulletin. Any dataset not found in the latest in-force bulletin must be considered not valid.

it also states that

If the in-force bulletin contains one or more NAVWARNs that are not present in the system, an indication should be given.

There is also a note that the normal behaviour is that datasets be cancelled explicitly, and the use of the bulletin is a sort of validation check.

NOTE: The in-force list should not actively be used as a means to cancel S-124 datasets, its role in dataset cancellation should only be as a failsafe in the event that a service interruption has caused the user system to miss one or more datasets that cancels earlier information

There would be bulletins from each S-124 services, which are normally issued as series (e.g. NAVAREA XVIII, Canada Great Lakes) and the bulletins would only list the NAVWARNS issued from that series.

The functionality to reconcile datasets using dataset identification embedded as attributes on features/information within the dataset is not a standard defined behaviour in S-100 and no obvious means of using S-100 metadata exists in S-100 Ed 5.x.

There may be a perceived overlap with the ongoing efforts to use S-128 catalogue datasets to validate dataset holdings in user systems, however the expectation is that S-128 datasets would only include an description of an S-124 service, its geographic coverage, and not a list of all the individual datasets active at a given time. A given NAVWARN service may issue several thousand NAVWARNS per year which might be a lot of overhead for an S-128 service and recipients of S-128 which are not interested in S-124.

Perhaps the bulletin could be used as part of a data validation check but then it would only report datasets that should have been cancelled or that are missing, and it would leave the user to resolve the data since there is no mechanism to change or cancel a dataset using a validation check. Moreover, in several S-100 discussions it has been stated that ECDIS or other end systems would not be expected to be executing validation checks as the expectation is that invalid data would never be distributed.

Perhaps there should be a generic S-100 implementation described where a system can produce a report identifying discrepancies between existing holdings and a list of available datasets. With some basic actions such as to archive data which is no longer available or to order data which is not being held but is desired.

Discovery vs Delivery metadata

Currently the discovery metadata in the exchange set is somewhat mixed with 'delivery' metadata. The results of a query to search for available data in a region should not need to include actual filenames and signatures. These would only be needed to validate content in an exchange set where the data is delivered.

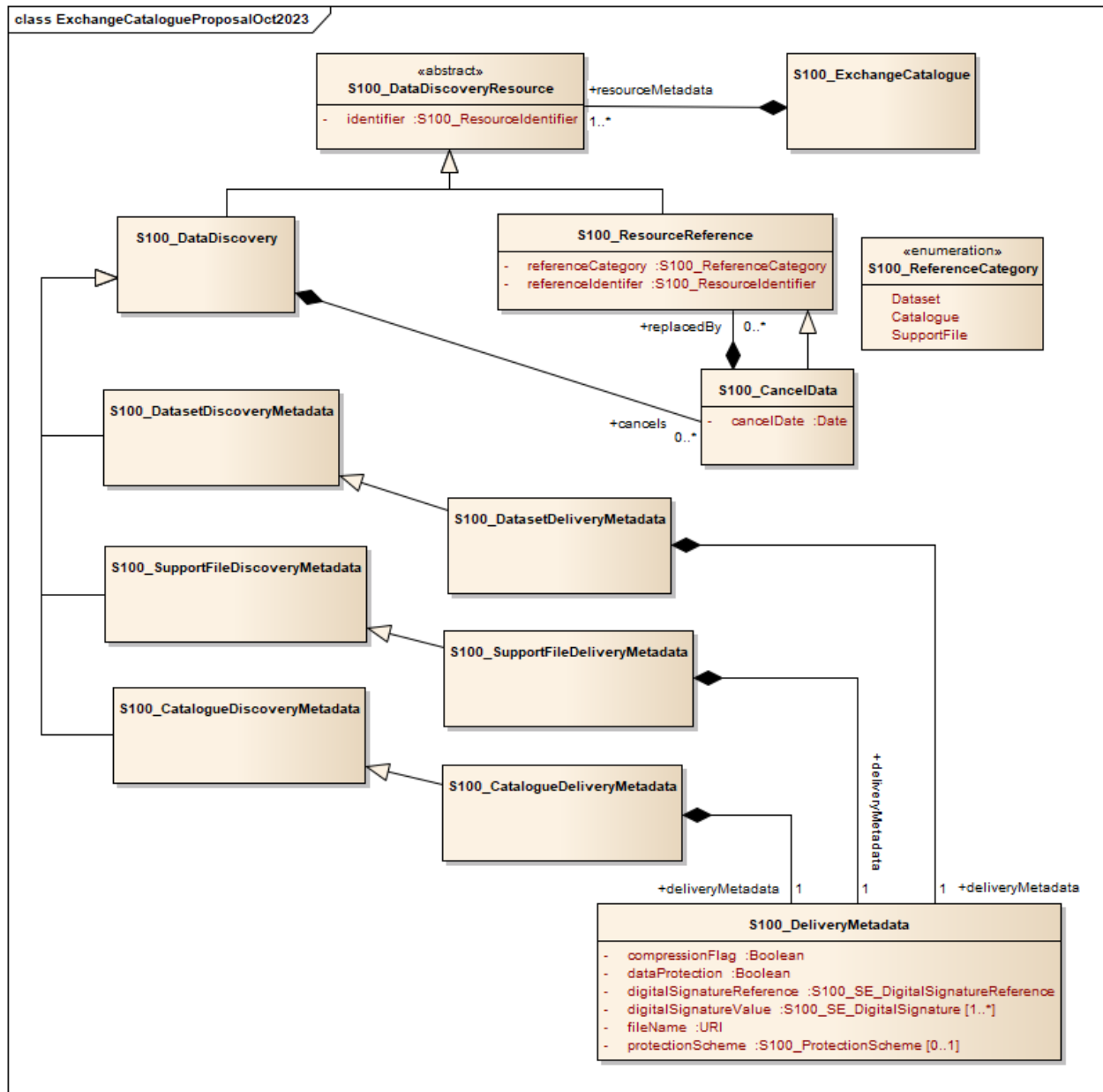
How would we distinguish between discovery data – without filenames and signatures and delivery metadata where filenames and signatures are mandatory?

It would be useful to be able to provide an exchange set that is meant just for discovery and only provides the discovery metadata of a set of datasets rather than delivering the actual datasets.

The exchange structure model could be enhanced to make the dataset discovery class be used for describing content for discovery and have an additional derived class for delivery metadata where the filename, signature etc is mandatory.

Potential restructuring of Exchange metadata hierarchy

- Add a base class for data discovery with a commonly defined unique resource identifier.
- Separate discovery from delivery content.
- Add a simple reference record to refer to a resource such as a dataset, catalogue or support file
- Add a simplified record to cancel a dataset.
- Add a reference from a resource (e.g., new dataset) to a cancelled resource (e.g., previous dataset).



For discovery or a list of available datasets an exchange set would include the Discovery metadata objects and not the delivery metadata. This could be used to discover available datasets such as based on a set of criteria. A discovery service would provide such a package. An in-force bulletin would be such an exchange set.

To cancel a dataset all that is needed is a unique resource identifier and the date it is cancelled. If a new dataset is used to cancel another dataset, then it can refer to the cancelled datasets. A cancelled dataset can optionally have a list of references to datasets which replace it.

Conclusions

Guidance for handling datasets such a loading and rendering should be defined at a common level and not require custom implementation for a specific product.

There is a gap in requirements for handling products such as S-124 NAVWARNS in terms of behaviour defined in S-100 or S-98. S-412 may have similar requirements.

Some improvements to exchange set metadata could facilitate handling of situations such as cancelling a dataset or providing a list of available datasets. System behaviour related to these activities could be defined generically in S-100 or S-98.

Recommendations

Recommendation that the dataset handling guidance be carried in a common place such as S-98 and expanded to describe dataset loading for active scaled and scaleless datasets over a mix of product specs.

Consideration be given to the requirement to reconcile active datasets against a list of active datasets which is itself a dataset.

Consider adjusting exchange set metadata to model discovery metadata which could be used on its own without actual datasets or data files.

Consider adjusting exchange set metadata to simplify the operation of cancelling datasets.

Justification and Impacts

Ideally S-100 and S-98 should define behaviour required by end user systems to be able to ingest and handle any S-100 product based on the same version of S-100.

Action Required of S-100 WG

The S-100 WG is invited to:

- a. Note this paper.
- b. Acknowledge the issues related to handling of datasets across products.
- c. Identify solutions such as the recommendation to define dataset loading at a common level such as S-98.
- d. Improve the flexibility of exchange sets for data discovery and cancellation.
- e. Note and identify actions to investigate solutions for 'failsafes' on handling of datasets.