Paper for Consideration by TSMAD/DIPWG

Management of dataset attributes in S-100

Submitted by:	TSMAD Vice Chair
Executive Summary:	This paper outlines the issue of dataset attributes and proposes a potential way forward on how to handle dataset attributes in S-100 and in S-10X product specifications
Related Documents:	S-100
Related Projects:	S-101, S-100 Based Product Specifications

Introduction / Background

Dataset attributes are used to provide information for the entire dataset rather than individual features that are contained within the dataset itself. Over the years, there has not been a consistent methodology established on how to model and use dataset attributes. The intent of this paper is to provide several alternatives for modelling dataset attributes within S-100 and S-100 based product specifications for TSMAD/DIPWG to decide on. The resultant decision will then be worked into a formal proposal for the S-100 working group to incorporate into S-100 and be made available for other product specification developers to utilize.

Analysis/Discussion

Throughout the years of developing S-101 one of the main issues that kept coming up was how dataset attributes should be handled, as there wasn't any clear guidance in S-100. In S-101, the dataset attributes are handled via the Catalogue File metadata, these attributes include producing agency, minimumDisplayScale, maximumDisplayScale, and verticalDatum.

The intent of storing the dataset attributes within the XML metadata catalogue was so that they could be exposed as discoverable metadata without having to open up the dataset itself. However, it was noted that some of these dataset attributes need to be carried on the dataset itself in order to make it easier for the machine to process the information. This issue applies specifically to data that is intended for use on an ECDIS. At one point these dataset attributes were defined within the feature catalogue, but it was determined at the Test Strategy Meeting in September 2014 that the feature catalogue was not the appropriate place for these attributes, especially since the S-100 Feature Catalogue model did not support this modelling. A couple of ideas were presented at the meeting that require further discussion by TSMAD/DIPWG.

They are as follows:

1. Model dataset attributes as an information type and then create a general metadata catalogue that contains the dataset attributes.

   Question: If this approach were taken, would the dataset attributes have to be registered as information types within the S-100 registry? What if they were already registered as a Simple Attribute also?

   Question: Does this mean that we would be publishing an S-100 Metadata catalogue that contains all possible dataset attributes for a different variety of products?

   Question: How is this any different than the S-100 XML metadata catalogue that is defined in S-100? The only key difference is that not all of these attributes have been "registered" as they are not associated with a real world feature. Some attributes have been registered as a simple attribute – such as soundingDatum.

2. For those products that use S-100 8211 encoding they can be stored In the Dataset General Information Record. However, if the dataset attributes are modelled as an information type then an INAS field would have to be added to the S-100 record.

   Question: Do the dataset attributes need to be modelled as information types? If not then they can be stored in the existing ATTR field.

Question: If they are stored in the existing ATTR field how are they documented within S-100 and within the product specification?

Question: How are dataset attributes handled in the GML encoding?

Question: How are dataset attributes handled in the HDF-5 encoding?

Further to the discussion above coming from the test strategy meeting there are some additional considerations. In order to support the concept of machine readable data the Feature Catalogue is used to define the content that would be expected within a given Product dataset. In that sense why not reconsider whether the concept of dataset attributes and related information would also be defined in the same catalogue. For example, a separate class for S100_Dataset with a one to one relation to the S100_FC_FeatureCatalogue and optional S100_FC_AttributeBinding and S100_FC_InformationBinding. In that way the attributes and information types applicable to the dataset for the product could be defined within the one catalogue and also used by other types in the same catalogue.

In addition, any decision on modelling dataset attributes needs to be applicable to the different types of encodings that are utilized in S-100. Or if different approaches are used for different encodings, each approach needs to be documented within S-100.

Conclusions
As noted above, there has not been a consistent approach on how to handle dataset attributes within S-100 based product specifications. It should be noted that as more and more product specifications are developed against S-100, the new S-100 working group will have to provide a methodology that can be used for all types of products. It is hoped that the discussion from the questions above will result in follow on actions, to provide a way forward in the use of dataset attributes.

Action Required of TSMAD/DIPWG
The TSMAD/DIPWG is invited to:
   a.   discuss the questions regarding dataset attributes proposed in this paper
   b.   record any follow on actions for the S-100 working group.