

Paper for Consideration by S-100 WG

Comments on DQWG Review of S-122, S-123, and S-127

Submitted by:	Raphael Malyankar; NIPWG Chair; Eivind Mong
Executive Summary:	Comments on results reported by DQWG for S-122, S-123 and S-127 feature catalogue/XML schema validation. There are some issues with the validator which should be addressed. The project also leads to questions about the overall vision for data modeling and product specification versioning and release management for the S-100 ecosystem as a whole.
Related Documents:	S100WG5-6.11
Related Projects:	S-122; S-123; S-127

Introduction / Background

The feature catalogue and XML data format for S-122 (Marine Protected Areas) and S-123 (Marine Radio Services) were developed in **2017**, based on S-100 **3.0.0** and feature, information types, attribute, and feature/information associations as defined at that time in the IHO GI registry, S-101 drafts, and the N PUB (Nautical Publications) modeling as captured on the NIPWG Wiki. The feature catalogue and XML data format for S-127 (Marine Traffic Management) was developed in **2018** based on S-100 **4.0.0** and feature, information types, attribute, and feature/information associations as defined at that time in the IHO GI registry, S-101 drafts, and the N PUB (Nautical Publications) modeling as captured on the NIPWG Wiki. Moreover, S-100 infrastructure and tools were still in early development at the time and the various feature catalogues and schemas therefore had to be hand-crafted and inspected visually.

There are some problems with the initial version of the validator which should be addressed before it can usefully be applied. The validation effort also leads to questions about the overall visions for data modeling and the versioning and release management of S-100-based product specifications.

Validation results as reported

A check of the first ten reported discrepancies in Step 1 produced the following results:

Table 1. Verification of Step 1 reported discrepancies for S-122

Discrepancy type	Item	Verification result
warning, not in XSD	applicationProfile	defined inline, child of onlineResourceType
warning, not in XSD	ballast	defined inline, child of ApplicabilityType
warning, not in XSD	callName	defined inline, child of ContactDetailsType
warning, not in XSD	callSign	defined inline, child of ContactDetailsType
warning, not in XSD	categoryofShipReport	This is indeed a case mismatch discrepancy between FC and XSD
warning, not in XSD	cityName	defined inline, child of contactAddressType
warning, not in XSD	communicationChannel	defined inline, child of radioCommunicationsType & ContactDetailsType
warning, not in XSD	contactInstructions	defined inline, child of ContactDetailsType
warning, not in XSD	country	defined inline, child of contactAddressType & sourceIndicationType
warning, not in XSD	dayOfWeeksRange	This is indeed a spelling discrepancy between FC and XSD

In short, 8 out of the 10 warning messages are incorrect, apparently due to limitations of the validation tool. This pattern appears to hold true for the remainder of the reported discrepancies as well, i.e., the vast majority of the reported warnings are not actually discrepancies.

In addition:

- There are XML types defined in the XSD which cannot be expected to be found in a feature catalogue, e.g., ISO639-3 (ISO 3-character language codes).
- XML types defined for constraining values also cannot be defined in feature catalogues. For example, the XML type Decimal0.0To360.0 (decimal values for azimuth/bearing), which is used by attributes “flip bearing”, “orientation value”, and “sector bearing”. In the feature catalogue, the range is captured by range and precision constraints applied to each attribute specification.
- S100_TruncatedDate is an S-100 type defined in S-100 (enumeration S100_CD_AttributeValueType). Since this is not an XML builtin type, it has to be defined in the schema (XSD).
- The spatial types, whether they are spatial primitives (e.g., GM_Point, GM_Curve, GM_Surface) or unions of spatial primitive types (e.g., PointOrSurface, CurveOrSurface), will not be defined in feature catalogues at all. (Union types are used because some features can have multiple types of geometry, e.g., either point or curve geometry).
- The treatment of association types (Permission, Inclusion) is described in S-100 10b-8.3.1, and conforms to their treatment in the GML specifications and is not an S-100 innovation.
- DatasetType, MemberType, IMemberType, are required as structural types by the format and will not be defined in feature catalogues.
- Since S-122/S-123 and S-127 were developed a year apart, during which the feature, information type, attribute, and association names, definitions, and types have all evolved, comparisons across different products are unlikely to be useful. Such evolution can be expected to continue in the future as submitting organizations propose amendments to concepts, features, information types, and attributes in the GI registry and the control bodies accept revisions.

In summary, it appears that only a small fraction (about 5%) of the reported discrepancies actually need attention.

Corrected XML feature catalogues and GML schemas for S-122 and S-123 were sent to the NIPWG Chair in January. Release management of these updates is with NIPWG.

The DQWG report also includes a comparison of schemas for different products that yields a number of warnings. For such comparisons, it must be considered that different data models adopt classes, attributes and enumerations differently depending on the needs of different data products and changes between different editions of S-100. Any cross-product specification schema validation should therefore be flexible enough to detect S-100 version differences and other obvious reasons for discrepancies.

Broader Implications

This discussion also leads to some important questions about the overall vision for data modeling in the S-100 “ecosystem” and versioning and release management for S-100-based product specifications (and, as a consequence, for datasets based on those product specifications). Some of the questions are:

- 1) How many versions of **S-100** can reasonably be in use at the same time?
- 2) Is it possible to allow several (more than two) extant versions if the feature catalogue, portrayal catalogue, and metadata components remain stable?
- 3) How many versions of any particular **Product Specification** reasonably be in use at the same time, assuming they are based on the same S-100 version?
- 4) How would the answer to the previous question change if the Product Specifications are based on different S-100 editions or revisions?
- 5) How far can Product Specifications lag behind the most recent edition/revision of S-100? What are the implications for data production, production tools, and applications?

Note that one argument for S-100 has been that each data model component can be based on any evolution approved as “valid” by the appropriate control body (or bodies) for the GI registry, and any such “valid” item can be (theoretically) used for the Product Specification development. As long as they are machine readable and are compliant to an official S-100 edition (which edition being the decision of the group or organization responsible for the Product Specification) all is fine, they can be displayed, assessed, etc.

Conclusions

We appreciate the NL/DQWG initiative and the effort that has been put into developing this validator. Such a tool is long overdue and it will be very useful. However, while automatic validation of format and feature catalogue is in principle useful, the present tool needs to be made more schema-aware before it can usefully be applied. Specifically, it should be upgraded to allow for various considerations: inline definitions, spatial types, the need to define S100_TruncatedDate as an XML type, the use of XML types to implement constraints on attribute values, structural types, and association types in feature catalogues vs. the GML-specified encodings. Should such tools be pursued, they will also need to factor in support for multiple versions of S-100 with potentially different rules.

A discussion of the broader implications mentioned in this paper is also needed, in conjunction with stakeholders in IHO working groups and other organizations. Based on the results, some of the core principles or assumptions of the “S-100 idea” may need to be reaffirmed or revised.

Actions Requested

The S-100 WG is invited to:

- 1) Note this paper.
- 2) Open a discussion with DQWG and the relevant project teams or groups/organisations responsible for current and planned Project Specifications about defining appropriate testing environments and test assumptions.
- 3) Open a discussion of the appropriateness of the current S-100 versioning regime and the necessity or otherwise for a “best practice” or guidance document on versioning and release management for product specifications. This discussion should include stakeholders (project teams, tool and application developers, etc.).
- 4) Open a discussion on adding an improved and mature version of this validator to the S-100 infrastructure or supporting toolset(s).
- 5) Take other action as appropriate.