Paper for consideration by TSMAD

Schema Definition and Validation Files

Submitted by:	Jeppesen
Executive Summary:	This paper outlines considerations for the management of format definition and validation files in S-100 data products and proposes bundling this issue with similar issues for feature and portrayal catalogues.
Related Documents:	(1) S-100 Ed. 1.0.0
Related Projects:	(1) S-100

1   Introduction / Background

Data encoded in XML or other formats may use external schema definition files to specify the structure and content of dataset files. Several schema definition languages are in use. For example the structure of an XML file can be specified using Document Type Definitions (DTDs), XML Schema Definition Language (XSD - there are 2 current versions), or REgular LAnguage for XML Next Generation (RELAX NG). The ISO 19757 series of standards defines several different Document Schema Definition Languages (DSDLs) for validating XML or SGML documents. Other languages have been specified by the World Wide Web Consortium (W3C) and similar groups.

S-100 product specification writers and application developers will need to consider the location, maintenance, and distribution of such schema definition files to OEMs and/or end users.

2   Terms and Abbreviations

DSDL   Document Schema Definition Language
DTD   Document Type Definition
FDF   Format Definition File
OASIS   Organization for the Advancement of Structured Information Standards
RELAX-NG   REgular LAnguage for XML Next Generation
SGML   Standard Generalized Markup Language
W3C   World Wide Web Consortium
XML   Extensible Markup Language
XSD   XML Schema Definition (language)

3   References

ISO 19131: Geographic Information – Data Product Specifications (2007), as amended by Amdt. 1. (2011).

ISO 19757-3: Information technology Document Schema Definition Languages (DSDL) Part 3: Rule-based validation - Schematron - First Edition.

ISO 19757-11: Information technology — Document Schema Definition Languages (DSDL) — Part 11: Schema association. (2011)

S-100: Universal Hydrographic Data Model, Edition 1.0.0, January 2010.

TSMAD25-4.3.2: Revisions and Extensions to S-100 Edition 1.0.0, 2013. Paper at TSMAD 25, Jan. 2013. URL: http://www.iho.int/mtg_docs/com_wg/TSMAD/TSMAD25/TSMAD25-4.3.2_S-100_GapsandExtensions-Final.pdf.

XML10: XML Schema Part 1: Structures & Part 2: Structures. W3C Recommendation 28 October 2004. URL: http://www.w3.org/TR/xmlschema-1/, http://www.w3.org/TR/xmlschema-2/.

XML11: W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures & Part 2: Datatypes. W3C Recommendation 5 April 2012. URL: http://www.w3.org/TR/xmlschema11-1/, http://www.w3.org/TR/xmlschema11-2/.

XMLCAT: XML Catalogs. OASIS Standard V 1.1, 2005. URL: https://www.oasis-open.org/committees/download.php/14810/xml-catalogs.pdf.

4   Discussion

A schema definition file is any machine-readable file that specifies the data format, or constrains data content or objects' relationships in a dataset, data package, or stream. Validation files contain rules for validating parts of the dataset such as individual objects or groups of objects, but do not define the entire dataset structure. In ISO terminology both are DSDL files. Schema definition and validation files are intended for use by data validation software modules.

DSDL files can supplement other DSDL files – e.g., a Schematron file can be a supplementary validation file that defines rule-based validation for XML data, the data format and dataset structure being specified by an XSD file. Validation rules may also be embedded in other DSDL constructs (e.g., Schematron assertions embedded in XML schemas) but embedding cannot be universally assumed. In other words, there may be more than one DSDL file for the same encoding.

Relevant issues for supply, update, and distribution of these files are outlined below:

Validation: To improve application performance and reduce application complexity, the validity of data received at the user application end of the data stream would ideally be guaranteed by the immediate upstream supplier. Obviously this would make schema definition and validation files unnecessary at the end-user application stage. Ideally this condition would also hold at every transfer stage in the data supply chain. It is unlikely that this condition can be imposed on all S-10x products and transfer modes, and whether acceptance of data from outside the local system involves schema-aware validation can be determined only in the context of the particular application domain.

Location: Schema definition files may be on the local platform or at a fixed Internet address. XML data can identify the location of the applicable XML schema using a *schemaLocation* attribute. Schemas may import other schemas and resources, potentially from Internet locations. An OASIS standard [XMLCAT] provides for "XML catalogues" that can map identifiers for external resources to locally accessible URIs, and this mechanism is widely used for substituting local copies of schema definition files or other resources nominally located at a fixed Internet address. Since Internet access cannot be assumed for shipboard applications, a local copy would be needed even if the nominal location is a fixed Internet address. Also, transactional delivery modes (e.g., web services) would not deliver schemas with each transaction.

Part 11 (Product Specifications) §12 should be more explicit about how product specifications can specify locations (fixed Internet addresses) of schemas and validation files. Guidance for writers of product specifications about packaging of schema definition and data validation files (e.g., exchange set layout to separate schema files from support files) may be appropriate.

Updating: There may be different current versions of the DSDL (e.g., HTML 4 and 5, XML 1.0 and 1.1). There is also the possibility of differences of structure and format due to updates to the application schema. Trying to make a schema definition file general enough for multiple versions is, in general, likely to either be impossible or make it useless for validation purposes. Packaging and delivery guidance should take into account the need to manage multiple versions.

Distribution: Application domains which need schema definition files locally will either have to define a delivery method (e.g., packaged with OEM applications, manual or automatic download from IHO or OEM Internet servers) or deliver them with exchange sets. Conventions to reduce exchange set size may be definable, for example, DSDL files might potentially be delivered only in new issue exchange sets. This too depends on application domain requirements and constraints.

Feature catalogues: Considerations for the management and distribution of feature and portrayal catalogues are very similar, and these catalogues are also part of product specifications. Bundling the question of schema and validation file management with feature and portrayal catalogue management will allow consistent solutions to similar problems.

5   Conclusion

Whether schema and validation files need be included in exchange sets should be left to writers of individual product specifications. For data products which need to make them available to end-user applications, addressing the questions of management and distribution can and should be bundled with the similar questions for feature and portrayal catalogues. It should be taken up at the same time those issues are addressed.

6   Recommendations

1. Clarify Part 11 (Product Specifications) with guidance about specifying the location of schema and validation files and the need for product specifications to address exchange set layout, updating, and distribution issues.

2. Address distribution and updating along with the similar issues for feature and portrayal catalogues.

7   Justification and Impacts

Justification: The proposed guidance answers likely questions arising during development of product specifications or encodings that use schema and validation files. Dealing with management and distribution along with similar issues for feature and portrayal catalogues allows harmonized solutions to similar questions.

Impacts: No impact on data products. The specification template will need to be updated with the proposed guidance. Any planned action on distribution and updating of feature and portrayal catalogs will be expanded to include schema and validation files.

8   Actions Requested

TSMAD is invited to:

- include the questions of management and distribution of schema and validation files in any future discussions of the management and distribution of feature and portrayal catalogues.