

Paper for Consideration by S-101PT

Attribute formats and constraints in S-101

Submitted by:	France, Teledyne
Executive Summary:	Attribute formats and constraints in S-101 documentation could be improved.
Related Documents:	S-100 5.2.0, S-101 2.0.0: FC and DCEG.
Related Projects:	S-101.

Introduction / Background

1. Attribute formats and constraints in S-101 are not always consistent with S-100.
2. The format and resolution of some attributes of type "Real" as shown in the S-101 DCEG 1.2.0 may allow the encoder think that there are some restrictions on the allowable values that make it impossible to encode some real-world situations.
3. This paper reviews the current guidance in S-100 Ed. 5.2.0 and S-101 Ed. 1.2.0 and suggests some improvements.

Analysis/Discussion

4. Attribute types and formats in S-100 and S-101.
 - S-100, 10a-5.1.4 lists 13 attribute Types and provide details on their formats. Thus, the format of an attribute is a direct consequence of the Attribute Type. Ex.:

S100_GF_RealAttribute [Real]	Decimal floating-point numbers in the value domain of double precision numbers (IEEE 64-Bits).	The decimal separator is always a point ('.') and there must be no thousands separator used. Positive numbers should not use the '+' sign. Negative numbers must have a leading '-' sign. The exponential form is permitted. Non-significant zeros are prohibited. If there are only zeros to the right of the decimal point the decimal point should be omitted. The values INF, -INF, and NaN are prohibited.	123.456 -42 1E-5 -2.45E7
--	--	--	-----------------------------------

- S-101 PS 4.6.3 (Simple attributes) states:

4.3.6.1 Simple attributes

S-101 uses seven types of simple attributes; these types are listed in Annex A – *Data Classification and Encoding Guide*, clause 2.4.2. Descriptions of the simple attributes included in S-101 can be found in Annex A, Sections 27, 28 and 30.

- The table below compares Attribute Types in S-100 and S-101 DCEG:

S-100 - 10a-5.1.4		S-101 DCEG - 2.4.2
S-100 Attribute Type	Value Type	Attribute type
S100_GF_TextAttributeType	CharacterString	Text (TE)
S100_GF_IntegerAttribute	Integer	Integer (IN)
S100_GF_RealAttribute	Real	Real (RE)
S100_BooleanAttributeType	Boolean	Boolean (BO)
S100_EnumerationAttributeType	PositiveInteger	Enumeration (EN)
S100_DateAttributeType	Date	Not Used
S100_TimeAttributeType	Time	Time (TI)
S100_DateTimeAttributeType	DateTime	Not Used
S100_URIAttributeType	URI	Not Used
S100_URLAttributeType	URL	Not Used
S100_URN_AttributeType	URN	Not Used
S100_TruncatedDateAttributeType	S100_TruncatedDate	TruncatedDate (TD)
S100_CodelistAttributeType	CharacterString	Not Used

Note: in Edition 1.2.0 of the DCEG, Simple attribute types “Free Text” (TE) has been amended to “Text” in 2.4.2. This is to be aligned with in section 27 to 30.

- S-101 DCEG, sections 27 to 30 provide the following details on the attributes, among which “Attribute Type” and “Format”. Example:

27.191 waterway distance
<p>Waterway distance: IHO Definition: The length of the space between two points along a waterway. (Adapted from Oxford English Dictionary).</p> <p>Attribute Type: Real</p> <p>Unit: Defined by the sub-attribute distance unit of measurement (see clause 27.89).</p> <p>Resolution: 0.1</p> <p>Format: xx.x</p>

“Format” as listed in the DCEG does not conform to any formal description. The information supplied by the attribute type and the Precision are self-sufficient.

One issue with indicating the format here is that it may not be suitable for all cases. In the situation above, the waterway maybe defined in nautical miles, in which case a resolution of 0.1 is not sufficient.

To prevent such issues, it is suggested to remove references to “Format” in sections 27 to 30 of the DCEG, as this is covered by Attribute Type.

Note: This would confirm the statement in DCEG 1.3.1: “NOTE A feature attribute type has a name, a data type and a domain associated to it.”

5. Attribute constraints in S-100 and S-101.

- S-100 2a-4.2.10 define the following attribute constraints:

2a-4.2.10 S100_CD_AttributeConstraints

Role Name	Name	Description	Mult	Type	Remarks
Class	S100_CD_AttributeConstraints	Constraints of a simple attribute	-	-	
Attribute	stringLength	Shall be represented as a positive integer (that is, greater than zero) that specifies the maximum number of characters that may be assigned to the text attribute type. If not specified, then the text length shall be unconstrained	0..1	PositiveInteger	
Attribute	textPattern	A character string that specifies a scheme of one or more constraints on the structure of the text values that may be assigned to the attribute. This shall be achieved by using a regular expression. W3C XML Schema Part 2: Datasets Second Edition, Appendix F (Regular Expressions) shall be used to define text patterns in this standard	0..1	CharacterString	
Attribute	range	Specifies the range of allowed numeric values	0..1	S100_NumericRange	
Attribute	precision	Specifies the precision of a real number	0..1	NonNegativeInteger	

- The constraints are the detailed in S-100, 13-8.1.1.1:

13-8.1.1.1 AttributeConstraints CreateAttributeConstraints(integer *stringLength*, string *textPattern*, string *rangeLower*, string *rangeUpper*, string *rangeClosure*, integer *precision*)

Return Value

AttributeConstraints

A Lua table containing an attribute constraints object.

Parameters

stringLength: integer or nil

The maximum number of characters that may be assigned to the text attribute type. If this value is nil, the length is unconstrained.

textPattern: string or nil

A regular expression defining the structure of text values that may be assigned to the attribute. If this value is nil, the structure is unconstrained.

W3C XML Standard Part 2 Appendix F (Regular Expressions) shall be used to define the text pattern.

rangeLower: string or nil

Specifies the lower range of allowed values for the attribute. If this value is nil, there is no lower value constraint.

rangeUpper: string or nil

Specifies the upper range of allowed values for the attribute. If this value is nil, there is no upper value constraint.

rangeClosure: string or nil

Defines the closure operations for the lower and upper ranges. This is one of enumerated values as defined in Table 1-3. This must be specified if either or both the lower or upper ranges are specified.

precision: integer or nil

If specified, indicates the precision of a real number.

- Attribute constraints in the S-101 FC: although S-100, 2a-3.5 states that attribute constraints can be specified in the FC, very few instances are present in the 1.2.0 S-101 FC. Ex:

```
<S100FC:S100_FC_SimpleAttribute>
  <S100FC:name>Text Offset Bearing</S100FC:name>
  <S100FC:definition>The angular distance measured from true no
  <S100FC:code>textOffsetBearing</S100FC:code>
  <S100FC:valueType>integer</S100FC:valueType>
  <S100FC:uom>
    <S100Base:name>Degree</S100Base:name>
    <S100Base:symbol>°</S100Base:symbol>
  </S100FC:uom>
  <S100FC:constraints>
    <S100CD:textPattern>xxx</S100CD:textPattern>
    <S100CD:range>
      <S100Base:lowerBound>0</S100Base:lowerBound>
      <S100Base:upperBound>360</S100Base:upperBound>
      <S100Base:closure>closedInterval</S100Base:closure>
    </S100CD:range>
  </S100FC:constraints>
</S100FC:S100_FC_SimpleAttribute>
```

For practical reasons, the S-101 FC is currently built from DCEG sections 27 to 30. Ideally, it should be the contrary and the FC should be the base reference as it facilitates machine processable checking.

It is proposed to systematically populate attribute constraints, including regex for Text attribute in the S-101 FC (refer to annex A of the paper).

- Attribute constraints in the S-101 DCEG: the constraints are mentioned in sections 27 to 30. Ex.:

27.98 flip bearing

<p>Flip bearing: <u>IHO Definition:</u> The bearing at which text is re-located to the opposite side of a feature when screen display is oriented away from true north.</p> <p><u>Attribute Type:</u> Real</p> <p><u>Unit:</u> Degree (°)</p> <p><u>Resolution:</u> 1°</p> <p><u>Format:</u> xxx</p> <p><u>Minimum value:</u> 0</p> <p><u>Maximum value:</u> 360</p> <p><u>Example:</u> 180 for a flip bearing of 180 degrees</p> <p><u>Remarks:</u></p> <ul style="list-style-type: none"> • No remarks.
--

Constraints indicated in the DCEG should reflect the FC.

textPattern constraint will be implemented in the FC with the use of Regular Expressions (regex). An example could be “/((\.|+){0,3}\|){1,2}/gm” for attribute signal sequence (reference: <https://regex101.com/r/7EYzFw/1>). As these expressions are not easily understandable for human being, it is proposed to use “Indication” to provide equivalent guidance in the DCEG for the end user.

To align the DCEG with the FC, it is also proposed to replace “Resolution” by “Precision” and add guidance in 2.4.2 explaining that “Precision” indicates the number of decimal digits.

The table below summarize the proposed changes (in red) for attribute constraints in S-101 2.0.0:

S-100 - 5.1.0		S-101 - 1.2.0		S-101 - 2.0.0	
S-100 - 2a-4.2.10	S-100 - 13-8.1.1.1	FC	DCEG	FC	DCEG
stringLength	stringLength	stringLength		stringLength	Remark
textPattern	textPattern	textPattern		textPattern	Remark
range	rangeLower	lowerBound	Minimum value	lowerBound	Minimum value
	rangeUpper	upperBound	Maximum value	upperBound	Maximum value
	rangeClosure	closure		closure	
precision	precision	precision	Resolution	precision	Precision
		/	Format	/	/

A few inconsistencies have been identified between the S-101 FC and DCEG (non-exhaustive list):

- For attribute **source**, stringLength constraint is “150” in the FC, whereas there is no guidance in the DCEG; textPattern constraint is “c...” where there should be no constraint;
- For attribute **sector bearing**, textPattern constraint is “xx.xx” in the FC. This should be replaced by precision = 2;
- For attribute **depth range minimum value**, textPattern constraint is “sxxxxx.xx; s = sign, negative values only” in the FC. This is useless as already specified by the attribute type (Real).
- For attribute **communication channel**, indication “Each Channel should be indicated in square brackets by 4 digits and up to 4 characters (A-Z).” should be changed to “Each Channel should be indicated by *up to 4 characters (A-Z) followed by 4 digits.*” (There seems to be no reason for having the text between brackets).

Other inconsistencies have been identified between the S-101 FC and the IHO Registry:

- Attribute "orientation value" has a range closure: "0.00 = 360.00" which does not conform with S-100 closure constraint (should be “getInterval” or “[0,360)”);

- Attribute precision is listed as (for example) "0.01" where it would be "2" (number of decimal digits) in the FC. Note: S-100, defines Precision as a non-negative Integer;
- GI Registry provide indication for "Format" (which seems useless), but no "Unit" (which is essential) for simple attributes;

Conclusions

It is recommended to review the way attribute formats and constraints are documented in S-101.

Recommendations

It is recommended to:

- A. Systematically encode the attribute constraints (including regex) in the S-101 Feature Catalogue;
- B. Review attribute guidance in the DCEG, (sections 27 to 30):
 - B1: remove guidance on "Format";
 - B2: replace "Resolution" by "Precision" and add guidance in 2.4.2;
 - B3: amend "Free Text" to "Text" in sections 27 to 30;
 - B4: ensure the guidance on the constraints will be consistent with those in the FC.
- C. Review (out of scope of the S-101PT) the presentation of IHO GI Registry to have it consistent with how attributes and constraints are defined in S-100.

Action Required of S-101PT

The S-101PT is invited to:

- a) Discuss this paper;
- b) Agree with the recommendations.