Paper for Consideration by S-100 WG

Attribute encoding for the ISO 8211 encoding

Submitted by:	SevenCs GmbH
Executive Summary:	Proposal for amendment of S-100 Part 10a
Related Documents:	S-100 Part 10a
Related Projects:	

Introduction / Background
According to S-100 Part 10a attributes of feature and information types are stored as character strings. So far there are no rules how the different value domains should be encoded. This paper will propose an amendment to close this gap.

Analysis/Discussion
Attributes have been stored in ISO 8211 files as string already in S-57. The main standard did not have any rules though in the ENC product specification a rule on the encoding of numeric attribute values exists (non-significant zeros are prohibited).
In S-100 different encoding exists where the storage of attributes is handled differently. The HDF5 encoding uses built-in types of HDF5 and the XML encoding can also use existing XML types for the encoding. The latter is not explicitly defined in the S-100 standard; an amendment would be required for that but is outside the scope of this proposal.
The ISO 8211 encoding part (10a) does not have any rules for the encoding of attribute values. A section for the Attribute field exists: 4.1- Attribute field with sub-sections on encoding, updating and the definition on the structure. It turns out that the numbering of the clauses is not correct in Edition 4.0.0 and will be corrected by this proposal.
Attributes, strictly speaking thematic attributes are defined by the GFM in Part 3 of the standard. Attributes values are carried only by simple attributes; complex attributes are a composition of other attributes. The following types are defined with the value type in brackets:

- S100_GF_TextAttributeType [CharacterString]
- S100_GF_IntegerAttributeType [Integer]
- S100_GF_RealAttributeType [Real]
- S100_GF_BooleanAttribute [Boolean]
- S100_GF_EnumerationAttributeType [Positive Integer]
- S100_GF_DateAttributeType [Date]
- S100_GF_TimeAttributeType [Time]
- S100_GF_DateTimeAttributeTime [DateTime]
- S100_GF_URIAttributeType [URI]
- S100_GF_URLAttributeType [URL]
- S100_GF_URNAttributeType [URN]
- S100_GF_TruncatedDateAttributeType [S100_TruncatedDate]
- S100_GF_CodeListAttributeType [???]

The S100_GF_CodeListAttribute is not really a simple attribute since it does not carry exactly one value. In the ISO8211 encoding it would not be possible to encode it. The GFM should be corrected. This proposal will not include it in the list of simple attributes. For the remaining types we will define the encoding instructions.

Encoding rules are already given for the Date, Time, DateTime, and S100_TruncatedTime in Part 1.

URI, URL, and URN are already mentioned in part 1 and references are given to the appropriate RFCs. (Request for Comments – technical specifications for the Internet)

As a side note it might be worth to be mentioned that for the XML type anyURI the processors are advised to NOT do any syntax checks for this type.

Conclusions

The clause 4.1 Attribute field should be amended as follows.

1. The section 4.1.1 Encoding rules stays unchanged
2. The section 4.2 Updating of the Attribute field becomes section 4.1.2 but its content stays at it is
3. A new section 4.1.3 Encoding of Attribute Values will be inserted as described below
4. The section 4.3 Attribute filed structure will be renumbered to 4.1.4

Note:

The changes of the numbering will also affect all other subsequent sections of clause 4.

The new section should read as:

10a-4.1.3 Encoding of attribute values

The following rules must be used to encode attributes values in the ATVL sub-field

AttributeType (Type)	Format	Remarks	Examples
S100_GF_TextAttributeType [CharacterString]	Any characters that are supported by the encoding (Here - UTF8)		whiskey водка
S100_GF_IntegerAttribute [Integer]	Signed Integer numbers of arbitrary length. The numbers must be in the decimal system.	Positive numbers should not use the '+' sign Negative numbers must have a leading '-' sign No whitespaces between the sign and the number. No other characters than the digits '0' to '9' in the number No non-significant zeros are prohibited.	Valid 42 -1224566 Not Valid 0012 123,234 - 12
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 no thousands separator must be used. Positive numbers should not use the '+' sign	123.456 -42 1E-5 -2.45E7

		Negative numbers must have a leading '-' sign The exponential form is permitted. No non-significant zeros are prohibited, if there are only zeros right of the decimal point the decimal point should be omitted. The values INF, -INF, and NaN are not allowed.	
S100_BooleanAttributeType [Boolean]	true or false	Alternatively, 1 or 0 are permitted.	true 1 0 false
S100_EnumerationAttributeType [Positive Integer]	Integer numbers > 0 in the decimal system	Non-significant zeros are not allowed. The '+' sign should not be used.	
S100_DateAttributeType [Date]	CCYYMMDD complete representation, basic format according to ISO8601	Representation with reduced resolution or truncated representations are not allowed)	20211223
S100_TimeAttributeType [Time]	complete representation, basic format according to ISO8601	Optional a time zone can be added. Fractions of a second are permitted but should be omitted if all zero	173500 183942+Z 201700-0500
S100_DateTimeAttributeType [DateTime]	Combination of date and time representation as above	The character 'T' is used as the separator between the date and the time part.	20211223T170000
S100_URIAttributeType [URI]	An URI according to the RFC 3986	URI do not contain white spaces, start with a 'scheme' followed by a ':'	tel:+1-816-555-1212
S100_URLAttributeType [URL]	An URL according to the RFC 3986		http://registry.iho.int
S100_URN_AttributeType [URN]	An URN as specified in RFC 2141	'scheme' = urn	urn:mrn:iala:aton:us:1234.5
S100_TruncatedDateAttributeType [S100_TruncatedDate]	A truncated version of a date		---01 --1224

Notes:
Though, the rules prohibit the use of non-significant zeros and the '+' sign, processors should be robust enough to support them.
For URI, URL, and URN it is recommended the processors do not perform any syntax checks.

Recommendations:

Insert the proposed section in part 10a.

Action Required of [S-100 WG]
The S-100-WG is invited to:

     a.   endorse the paper

     b.   discuss the topics

     c.   conclude on the action list

     d.   decide on the next steps forward