Paper for Consideration by S100TSM9

ISO8211 Description of complex attributes

Submitted by:	PRIMAR, SevenCs
Executive Summary:	A need for a more extensive description of the ISO 8211 encoding of complex attributes in S-100 Part 10a is needed.
Related Documents:	S-100 5.0.0.
Related Projects:	

Introduction / Background

S-100 Part 10a (ISO/IEC8211 Encoding) gives examples of how the Attribute Index (ATIX), Parent Index (PAIX) and Attribute Instruction (ATIN) fields are encoded and what happens when attribute fields are updated. During implementation of the 8211 profile in PRIMAR it turned out to be problematic to implement and understand the logic related to complex attributes.

As a consequence, it is proposed to add clarifying information in S-100 10a-4.

Analysis/Discussion

Currently in S-100 5.0.0 the following figures and tables are available in S-100 10a-4.1.1 and 10a-4.1.2 :

10a-4.1 Attribute field

10a-4.1.1 Encoding rules

In S-100 attributes can be either simple or complex. Simple attributes have values whereas complex attributes are an aggregation of other attributes, either simple or complex. The following diagram shows an example of a feature type with both simple and complex attributes.

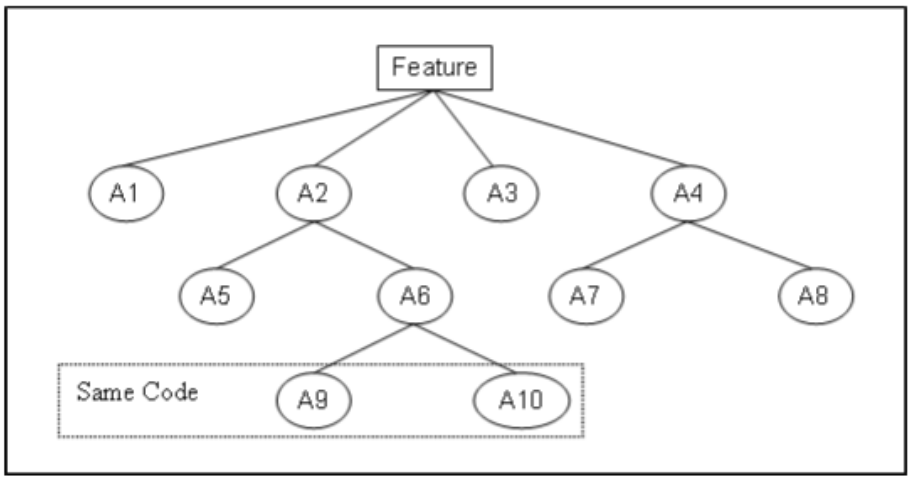

Figure 10a-2

The feature has four attributes: A1, A2, A3, and A4. A1 and A3 are simple attributes; A2 and A4 are complex attributes. A2 comprises two attributes (A5 and A6) where A5 is a simple one and A6 is another complex attribute. A4 and A6 are two complex attributes; both consist of two simple attributes.

Another characteristic of attributes is the cardinality. This indicates how many attributes of the same kind (the same code in a feature catalogue) are used at the same parent. The same parent means that they are all top level attributes or belonging to the same instance of a complex attribute. In the example above A9 and A10 are assumed to have the same code.

With the concept of cardinalities larger than one, an attribute can be seen as an array of attributes. To access an attribute in such an array one needs not only the code of that attribute but also the index of that attribute. Note that the order in such an array may be meaningful and must be maintained by the encoding.

Taking all of the above into account an attribute can be uniquely addressed by three values:

1. The attribute code;
2. The index of the attribute (starting with 1);
3. The parent of the attribute.

To complete the example above, the following table defines codes and values of the attributes:

Attribute	Code	Attribute Index	Value	Remarks
A1	21	1	Vachon	
A2	22	1		complex
A3	23	1	12	
A4	24	1		complex
A5	25	1	42.0	
A6	26	1		complex

A7	27	1	123	
A8	28	1	Canada	
A9	29	1	17	same code as A10
A10	29	2	43	same code as A9

To encode an attribute a set of five items is necessary: the three mentioned above plus an update instruction and the value of the attribute. To specify the parent of the attribute an index is used. This index points to the n^{th} tuple in the ATTR field starting with 1. The following table shows the encoding of the example:

Index	NATC	ATIX	PAIX	ATIN	ATVL	Remark
1	21	1	0	Insert	Vachon	A1
2	22	1	0	Insert		A2 - composite
3	25	1	2	Insert	42.0	A5
4	26	1	2	Insert		A6 - composite
5	29	1	4	Insert	17	A9
6	29	2	4	Insert	43	A10
7	23	1	0	Insert	12	A3
8	24	1	0	Insert		A4 - composite
9	27	1	8	Insert	123	A7
10	28	1	8	Insert	Canada	A8

Note that here the pre-order traversing is used to define the order of tuples in the field. This keeps all part of a complex attribute together and guarantees that the parent is always stored before the child. The pre-order traversing is defined as follows:

1) Encode the root;
2) Then encode the sub-trees from left to right.

This traversing order is mandatory within this standard.

Note also that the ATIN subfield (Attribute update Instruction) will always be 'Insert' for encoding base data attributes. The other ATIN values (Modify, Delete) are only needed for updating the ATTR field.

All values of attribute are stored as character strings even if the value domain is a numeric type. UTF-8 will be the only encoding allowed in S-100 for such character strings. This allows the encoding of all characters of the first multilingual plane of ISO 10646. There is no other encoding for national character sets necessary.

10a-4.1.2 Updating of the Attribute field

To update an attribute the attribute must be uniquely identifiable and once identified instructions are needed to affect that attribute. The Attribute Update Instruction indicates whether an attribute is to be deleted from the field; modified, or inserted. Deletion and modification implies that the attribute exists. Deletion and insertion may change the indices of other attributes in an array of attributes and therefore must be taken into account when the attribute field is updated. Instructions must be applied in sequence in order that the indices used are identifying the correct attributes components on subsequent updates.

To demonstrate the updating of attributes the example above should be modified as shown in the following figure.

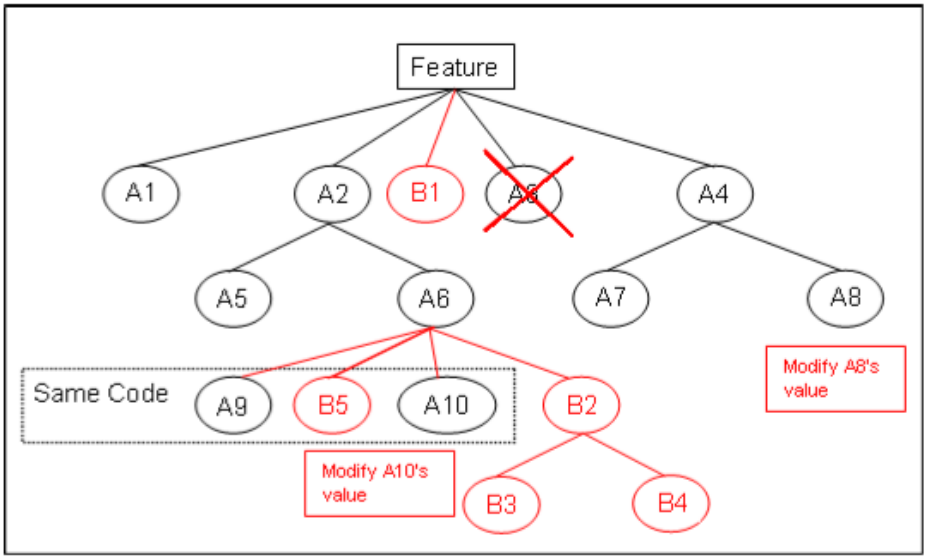

Figure 10a-3

The details are:

Attribute	Code	Attribute Index	Value	Update Instruction	Remarks
B5	29	2	32	Insert	Will change A10's index to 3
A10	29	3	7	Modify	
B2	35	1		Insert	complex
B3	36	1	32	Insert	
B4	37	1	123	Insert	
B1	32	1	abc	Insert	
A3	23	1	1,2	Delete	
A8	28	1	Germany	Modify	

In order to identify B5, A10 and B2 the entries for A2 and A6 must be inserted. The same is true for A4 (to identify A8). The complete field will look like:

Index	NATC	ATIX	PAIX	ATIN	ATVL	Remark
1	22	1	0	Modify		A2 - complex
2	26	1	1	Modify		A6 - complex
3	29	2	2	Insert	32	B5 - Will increase the ATIX of A10
4	29	3	2	Modify	7	A10 - now with ATIX 2
5	35	1	2	Insert		B2 - complex
6	36	1	5	Insert	22	B3
7	37	1	5	Insert	123	B4
8	32	1	0	Insert	abc	B1
9	23	1	0	Delete		A3
10	24	1	0	Modify		A4 - complex
11	28	1	10	Modify	Germany	A8

Note that in order to delete a complex attribute it will be adequate to delete the root entry of that attribute. For example, to delete A2 only one entry (22, 1, 0, Delete) has to be encoded.

Due to other changes being implemented in Part 10c the numbering of the chapters have changed. The following suggested changes are numbered in accordance with the new chapter numbering.

It is proposed to add the following to S-100 10a-5.1.1 (Previously S-100 10a-4.1.1 in S-100 edition 5.0.0) in the 5th paragraph:

Taking all of the above into account an attribute can be uniquely addressed by three values:

1. The attribute code – encoded with the sub-field NATC;
2. The index of the attribute (starting with 1) – encoded with the sub-field ATIX;
3. The parent of the attribute – encoded with the sub-field PAIX.

The Parent Index (PAIX) is solely used for defining the tree inside the field only, and there is no need to preserve it in an internal system (e.g. SENC) structure.

To complete the example above, the following table defines codes and values of the attributes:

It is proposed to add the following to S-100 10a-5.1.2 (Previously S-100 10a-4.1.2 in S-100 edition 5.0.0):

10a-5.1.2 Updating of the Attribute field

To update an attribute the attribute must be uniquely identifiable and once identified instructions are needed to affect that attribute.

The identifiaction of an attribute is defined by the Numeric Attribute Code (NATC) and the Attribute Index (ATIX).
The Attribute Update Instruction indicates whether an attribute is to be deleted from the field; modified, or inserted. Deletion and modification imply that the attribute exists. Deletion and insertion may change the indices of other attributes in an array of attributes and therefore must be taken into account when the attribute field is updated, thus the ATIX sub-field must use then the corrected value. Instructions must be applied in sequence in order that the indices used are identifying the correct attributes components on subsequent updates.

To demonstrate the updating of attributes the example above should be modified as shown in the following figure. Note that to identify a node in the attribute tree it is necessary to specify all nodes from the root of the tree. In order to insert the attribute B5 both the nodes A6 and A2 must be encoded before and marked as modified.

Conclusions

- Information has been added to better describe the NATC, ATIX, PAIX and their use.

Action Required of S100TSM9

The S100TSM is invited to:
Note the report and discuss proposed changes.