To	:	ISO / TC 295 Project XML / JSON
From	:	Jan Vrijenhoek
Subject	:	Recommendations on XML-Schema and JSON-Schema design proposal
Date	:	2021-02-23

This is a list of recommendations on the first draft version of XML-Schema and JSON-Schema design proposal.

The following persons contributed to this list and the creation of the first WD version of the document "*Exchange formats for Audit Data Collection Standard XML and JSON.pdf*":

Ms. Teresa Tang (SAP)
Ms. Kira Shi (SAP)
Mr. Vincent Yu (SAP)
Mr. Jiwen Jin (Oracle)
Mr. Tianyang Lyu (TC 295 Representative)
Mr. Robert Stamsnijder (NEN)
Mr. Frans van Basten (NEN)
Mr. Jan Vrijenhoek (NEN, author)

(1)
Don't recommend one big XML-Schema (XSD) file, please separate them per table.
From: Jiwen Jin

NL: We changed the proposal to one schema per ADCS table.

(2)
Some data element -> data attribute
From: Jiwen Jin
The data elements relevant to a table can be attributes of root element since they are attributes of data rather than data themselves.

NL: Because we changed the proposal to one schema per table there is maybe no need for extra information anymore, so let's discuss if other information fields are required and where to place it.

(3)
Namespace:
From: Jiwen Jin
Namespace is too general, must be more specific.

NL: In the new proposal we generate a namespace which is unique per message type.

(4)
Table name as Root Tag
From: Jiwen Jin
Root Tag doesn't need to be AuditDataCollection, AuditDataCollection can be the part of the Namespace. My preference is directly take the table name, and then cut the prefix and add 's' e.g. table name: BAS_Employee, then root tag: <Employees>

NL: In the new proposal we take the table name as the Roottag, but we don't change the table names.

(5)
Separate Data Type Schema
From: Jiwen Jin
In XML schema, we usually will create a data type file as a single source to define and maintain all the data type for the data elements in datatype.xsd

NL: We think this is not a good design in case of the ADCS tables.

a. We can imagine that the proposed method of Jiwen is a very good in-house method, and that perhaps other organisations have a completely different in-house design of their messages and data type schemas. But we believe that from the 71 ADCS table schemas any desired in-house format of message schemas and data type schemas can be derived.

b. There are many advantages in having the table structure and associated data type definitions in one schema per table. Changes in other tables have no effect on the schema and ADCS extensions by adding new tables can be entered without any effect on the existing tables.

(6)
Naming of identifier data types.
From: Jiwen Jin
For instance "EmployerID" should have a data type definition named "EmployerIDType".

NL: See issue (5).

(7)
Reference Keys
From: Leo Alewijnse
How do we tackle reference key validation, if we have about 71 XML table files?

Team: This is outside the scope of the exchange formats.

(8)
From: Leo Alewijnse
Business Rules
How do we tackle validation of business rules?

Team: This is outside the scope of the exchange formats.

(9)
We need some header information in each file.
From: Jiwen Jin

NL: BAS_Profile contains already overall information about the ADC set. BAS_Profile can be seen as the "header" for the whole set of files. We prefer to put the context data in that table. It has already Developer Name, Software Name, Software Version, Extraction date etc.

(10)
We want TAX as a repeating group in the XML file, instead of TAX1, TAX2, TAX3, TAX4 as in CSV.
From: Jiwen Jin

NL: This is an CSV issue where you need to number the occurrences of a repeating group. The proposal and samples contain the repeating group TAX.

(11)
How to tackle NULL values?
From: Jan Vrijenhoek
Some fields may have a NULL value, according to ISO 21378:2019
See f.i. "Order Line Transaction Amount" in "SAL Shipments Made Details" (Table 86).

Team: In XML, fields which can have no value must be defined as optional fields, and then can be omitted (also the start tag and end tag must be omitted).

(12)
Empty files
From: Jan Vrijenhoek
According to ISO 21378:2019 (Chapter 5.5,g) which reads: "If an ADCS table cannot be exported, a blank file shall be created with no data contained it", means that all table files need to be sent, even when there are no lines in it.

Team: There are doubts about the rule that all tables must always be delivered, even if they are empty.

(13)
How to tackle Business_Segment_X and Account_Segment_X?
From: Jan Vrijenhoek
Business_Segment_X and Account_Segment_X, are place-holders for organizational levels, where X can be 1,2,3 up to 20 (see ISO 21378:2019 Annex A and Annex B).

An XML-Schema is a fixed definition. It is not possible to add own defined fields.

We are not sure that all auditee will deliver fixed columns named Business_Segment_1, Business_Segment_2, Business_Segment_3, Business_Segment_4 and Business_Segment_5.

This must be checked.

(14)
Required / Optional Status of fields in tables must be reviewed.
From: Jan Vrijenhoek
In ISO 21378:2019 there's a Level indicator 1 or 2 on tables and data elements, but that's not the same as a Required/Optional status in an XML-Schema. So this must be reviewed.

NL: We will make a proposal for all 71 tables and schema's, because this is also discussed in relation to the Customs Proof Of Concept.