DITA 1.3 proposed feature 13112
RelaxNG for DITA Vocabulary

An alternative approach based on XML Schema

Written by: Guillemette Borel, Jean-Luc Borie, Michel Doméon, Nicolas Dupuy, Alberto Franzetti, Jean-Jacques Thomasson, Philippe Zingoni.

Date: July, 15th 2013

In his paper dated June 19th, Eliot Kimber presented the pros in favor of Relax NG use as the basis in modeling technique for DITA 1.3. We would like hereby to comment that paper.

The new recommendation 1.1 of XML Schema (dated April 5th , 2012) incorporates notable enhancements such as the possibility to:

- define business rules with <assert> element,
- change the content model of an element depending of the value of an attribute in the document instance (<alternative> mechanism),
- interleave any elements within a content model (<openContent>),
- define error messages (<error>)
- have revision tracking within schemas.

Nevertheless, the most interesting enhancement for DITA is certainly the new <override> mechanism which directly solves the initial difficulties adressed by Eliot in the use of the redefine mechanism.

Although we could avoid the <override> mechanism to get valid schemas for DITA (refer to our presentation to the DITA TC on July 31st, 2012), its use allows a major simplification of the DITA XML Schema design specifically in regards of a specialization perspective.

Thanks to the work of the WG on XML Schema recommendation, the version 1.1 was indeed an interesting update of version 1.0.

With the previous version 1.0, our group had already used a modeling technique, based on the concept of “libraries of component”, which bring a set of positive enhancements for the management of the model itself:

- The model is easier to manage as each component can be isolated, analyzed and tested by its own.
- Domaining and specializing are easy to implement
- Model may be logically organized in different folders
- Extensibility has been proven as we easily succeeded to integrate the S1000D vocabulary.

The interconnection between DITA and S1000D was so simple to achieve that this shows the benefit of such approach, allowing DITA to easily embrace other industry specific vocabularies.

The experience with S1000D express that the design of schemas using libraries of logical components facilitates the development of applications (transformation tools, publication engines, design tools).

As mentioned above, XML Schema fits well with a design based on libraries of logical components which offers the possibility to have dedicated “sub-schema” depending on the modeling strategy: you can have one schema dedicated to definition of all, or part of, the elements names (and only the names), another for the attributes names, and others for the definition of types. This modeling technique is possible thanks to the two following features:

- XML Schema does not enforce each schema to have a root element
- XML Schema allows recursive inclusions of schemas.

Another point to be taken into account for the industry, is also the numerous operational business standards which are already based on XML Schema (http://en.wikipedia.org/wiki/List_of_XML_Schemas)

When designed as we propose, with a concept of components libraries, XML Schema is not verbose: When the Relax NG schemas for DITA 1.3 are about 14 694 lines long (SGML comments and documentation being excluded) our modular XML schemas are only 10 285 lines (same scope).

The conversion from RNG DITA 1.3 to XML Schema that we have tested with the proposed converter from SyncroSoft/Oxygen has produced a 24 450 lines long monolithic XML Schema. In comparison, with our design technique, that last schema would only be 11 761 lines long. Much ‘slimmer’ than through a traditional XSD design.

Relax NG has not been updated since December 3rd, 2001. There can be a risk of ageing. It is not a popular language in the industry, and is currently still not supported by some major XML tools and only few people, in that market, have skills on Relax NG.

In its tutorial on DITA dated 2011/06/16 Eliot Kimber was giving a series of good reasons explaining why he preferred XSDs to DTDs.

Relax NG is also limited to a hierarchical cutting of a model, with mandatory root elements per grammar and does not allow recursive schemas.

At least, Eliot mentioned the possibility of using exceptions in Relax NG, but as soon as somebody will use that mechanism, then, the conversion to XML Schema will become impossible or extremely verbose. It makes impossible the use of this mechanism in base DITA Relax NG.

Regarding all these comments, our recommendation would be that the DITA TC analyzed carefully the future XML model used to support the DITA vocabulary.
Our vision is that the use of component libraries based on the last XSD schema specification would better fit to the DITA community. It would enhance the management of DITA structures, allows future possibilities of evolution and would still be based on an open, evolving and well supported W3C standard.

Enclosed models: Proposed XSD for DITA 1.3 with S1000D vocabulary being interconnected.