**NIEM Schema Documents: Reference, Extension, Subset, Constraint**

**Reference and extension schemas**

NIEM was founded on the decision to use XML Schema as the formalism for representing data models. All of the data model in NIEM -- NIEM Core, the NIEM domains, the data model in every message specification – all are defined as an XML schema created by assembling a collection of XML schema documents. The NDR is the special sauce that is added to XML Schema to make this possible. Almost all of the rules in the NDR fall into one of these categories:

1. They define the semantics of NIEM XML schemas and NIEM XML data. For instance, subclasses are represented by xs:extension, and subproperties are represented through substitution groups.
2. They support schema reuse and composition. For instance, final="true" makes reuse more difficult, so it is not allowed in a reference schema.
3. They forbid XML schema features that do not work well with NIEM semantic interpretation. For instance, model groups and repeatable sequences do not have a useful interpretation in NIEM, and are therefore not allowed.
4. They forbid XML schema features that are not well supported in COTS tools, or were not well supported in the past. For instance, xs:include can cause problems in schema assembly, and is not interpreted consistently by different validating parsers.

The schemas defining the NIEM data model -- NIEM Core plus the domains – all follow the NDR rules for *reference schemas,* which are designed to maximize potential for composition and reuse (#2, above). Schemas that extend the NIEM model for a particular data exchange follow the NDR rules for *extension schemas*; because these schemas may or may not be reused elsewhere, the rules are less strict about reuse.

**Subset schemas**

The concept of a subset schema is not unique to NIEM. Technically,

A schema S is a *subset schema* of a schema R

if and only if

every document D valid against S is also valid against R

In NIEM, a subset schema is a subset of one of the reference schemas that define the NIEM data model. Subset schemas are convenient in NIEM message specifications because they permit developers to work schema documents that are much smaller than the (large) NIEM reference schemas. The *Subset Schema Generation Tool (SSGT)* allows developers to select and extract their required data components from the NIEM model. The subset schema documents generated by SSGT all conform to the NDR. When combined with extension schema documents, they define the data model for the message specification, and specify the required and optional content for the messages specified.

**Constraint schemas**

Every NIEM message specification has a collection of extension and subset schema documents to define semantics and conformance. However, developers have other purposes for XML Schema, and those extension and subset schema documents are not suitable for some of those uses. In such cases, developers are free to create *additional* schema documents as needed for their purpose. Those purposes include:

1. Constraints that are difficult or impossible to specify in a subset schema. For example, it is difficult to specify different string lengths for nc:PersonSurName and nc:PersonGivenName.
2. Replacing substitution groups with xs:choice. Substitution groups are inconvenient for some XML tools.
3. Optimization for EXI encoding.

Probably because of purpose #1, these additional schemas were known as *constraint schemas* in the documentation for NIEM 2.0.

**Summary**

* The NIEM data model is defined by a collection of *reference schema* documents
* A NIEM message specification selects data components from the NIEM model in a collection of *subset schema* documents
* Data components not found in the NIEM model are added to a NIEM message specification in a collection of *extension schema* documents
* The extension and subset schema documents define the data model for a NIEM message specification, and specify the required and optional content of the messages
* A NIEM message specification may also include any number of *constraint schema* documents that
  + Are not used to define the data model
  + Need not conform to the NDR
  + Satisfy any purpose of the developers