



## **Guía Básica de Características de archivos .XML**



## 1 Índice

1	Índice .....	2
2	Introducción .....	3
2.1	Propósito .....	3
2.2	Alcance .....	3
2.3	Audiencia .....	3
2.4	Glosario .....	3
3	Definición.....	4
3.1	XML .....	4
3.1.1	¿Qué es un XML? .....	4
3.1.2	¿Para qué sirve un XML? .....	4
3.1.3	Marcado en un XML .....	5
3.1.4	Un XML bien formado .....	6
3.2	XSD .....	7
3.2.1	¿Qué es un XSD? .....	7
3.2.2	¿Para qué sirve un XSD? .....	7
3.2.3	Espacio de nombres .....	8
3.2.4	Tipos de datos .....	9
3.2.5	Componentes .....	11



## 2 Introducción

### 2.1 Propósito

---

El propósito de este documento es brindar una introducción al concepto del lenguaje XML y los Esquemas XML, mismos que son la base para la generación de los archivos, así como los estándares que apoyan la implementación de un esquema para poder formar un XML que sea validado para la contabilidad electrónica.

Con este documento se busca apoyar a cumplir los estándares tecnológicos dentro del Servicio de Administración Tributaria, realizando una traducción y explicación de los estándares emitidos por la W3C.

### 2.2 Alcance

---

El documento comprende la definición de patrones de diseño de SOA, Cloud y persistencia de datos. Describiendo principios, buenas prácticas y estándares de diseño.

### 2.3 Audiencia

---

La audiencia principal de éste documento se conforma por:

Contribuyentes que deseen formar un documento XML tomando como base el esquema XSD publicado por el SAT para la Contabilidad Electrónica.

### 2.4 Glosario

---

<b>Término</b>	<b>Significado</b>
<b>XML</b>	Lenguaje de Mercado Extensible
<b>XSD</b>	Esquema de Definición XML
<b>Namespaces</b>	Espacio de Nombres
<b>W3C</b>	



## 3 Definición

### 3.1 XML

---

#### 3.1.1 ¿Qué es un XML?

El lenguaje de marcado extensible (XML por sus siglas en inglés) es un formato simple basado en texto para representar la información de manera estructurada: documentos, datos, configuraciones, libros, transacciones, facturas, y mucho más. Derivado de un antiguo formato estándar llamado SGML (ISO 8879), con el fin de ser más adecuado para su uso web.

El lenguaje extensible de marcas, abreviado XML, describe una clase de objetos de datos llamados documentos XML y parcialmente describe el comportamiento de programas informáticos que pueden procesarlos. XML es un perfil de aplicación o forma restringida de SGML (Lenguaje Estándar de Marcado Generalizado por sus siglas en inglés).

XML fue desarrollado por un Grupo de Trabajo de XML (originalmente conocido como el comité de revisión editorial de SGML) formado bajo el auspicio del World Wide Web Consortium (W3C) en 1996.

#### 3.1.2 ¿Para qué sirve un XML?

XML es uno de los formatos más utilizados para el intercambio de información estructurada al día de hoy: intercambio entre programas, entre personas, entre computadoras y personas, tanto a nivel local como a través de las redes informáticas.

Los objetivos de diseño para XML son:

- ✓ XML debe ser utilizable directamente sobre internet.
- ✓ XML debe soportar una amplia variedad de aplicaciones.
- ✓ XML debe ser compatible con SGML.
- ✓ Debe ser fácil escribir programas que procesen documentos XML.
- ✓ El número de características opcionales en XML debe ser mantenido en un mínimo, idealmente cero.
- ✓ Los documentos XML deben ser legibles por un humano y razonablemente claros.



- ✓ El diseño de XML debe ser preparado rápidamente □ El diseño de XML debe ser formal y conciso.
- ✓ Los documentos XML deben ser fáciles de crear.
- ✓ La brevedad en la marcación es de mínima importancia
- ✓ La especificación, junto con los estándares asociados (Unicode and ISO/IEC 10646 para caracteres, Internet RFC 1766 para las marcas de identificación de lenguaje, ISO 639 para los códigos de nombre de lenguaje, ISO 3166 para los códigos de nombre de país), proveen toda la información necesaria para entender XML Versión 1.0 y construir programas de computador que lo procesen.

### 3.1.3 Marcado en un XML

El comienzo de todo elemento no vacío de XML está marcado por una etiqueta de comienzo.

#### Etiqueta de comienzo:

```
<nombre atributo="valor">
```

El nombre en la etiqueta de inicio y de fin define el tipo de elemento. El atributo y sus valores son conocidos como especificaciones de atributo de un elemento, en donde el "atributo" en cada par es conocido como el nombre de atributo y el contenido de "valor" debe ir entre delimitadores ' o ".

- ✓ **Restricción de buena formación:** Especificación de atributo única. Ningún nombre de atributo puede aparecer más de una vez en la mismo etiqueta de comienzo o etiqueta de elemento vacío).
- ✓ **Restricción de validez:** Tipo del valor de atributo  
El "atributo" debe haber sido declarado; el valor puede ser del tipo declarado para él.
- ✓ **Restricción de buena formación:** Ninguna referencia a entidad externa



Los valores de "atributo" no pueden contener referencias directas ni indirectas a entidades externas.

- ✓ **Restricción de buena formación:** Ningún < en el valor de atributo El texto de reemplazo de cualquier entidad a la cual se haga referencia directa o indirectamente en un valor de atributo (diferente a "&lt;") no debe contener <.

El final de todo elemento no vacío de XML está marcado por una etiqueta de fin.

#### **Etiqueta de fin:**

```
</nombre>
```

El texto entre la etiqueta de comienzo y la etiqueta de fin es llamado el contenido del elemento.

### **3.1.4 Un XML bien formado**

Un objeto de texto es un documento XML bien formado si:

- ✓ Tomado como un todo, se empareja con la producción marcada como documento.
- ✓ Cumple con todas las restricciones acerca de buena formación descritas en la especificación
- ✓ Cada una de las entidades procesadas referenciadas directa o indirectamente en el documento está bien formada.

**Nota:** Emparejar con la producción documento implica que contiene uno o más elementos, los cuales están conforme a la sintaxis explicada en el punto anterior.

## 3.2 XSD

---

### 3.2.1 ¿Qué es un XSD?

Un esquema XML es un lenguaje para expresar las restricciones sobre los documentos XML.

Existen diferentes tipos de lenguajes de esquema de uso generalizado, pero las principales son Documento de Tipo de Definiciones (DTD), Relax -NG, y XSD (Definición de Esquemas XML). Desde esta página se puede encontrar más información acerca de las DTD y XSD W3C, ya que esos son los lenguajes de esquemas primarios definidos en el W3C.

### 3.2.2 ¿Para qué sirve un XSD?

Un esquema puede ser utilizado:

- ✓ Para proporcionar una lista de elementos y atributos en un vocabulario.  
Asociar tipos, tales como número entero, cadena, etcétera, con los valores que se encuentran en los documentos.
- ✓ Para restringir el lugar dónde pueden aparecer los elementos y atributos, y lo que puede aparecer dentro de esos elementos.
- ✓ Lograr que la documentación sea legible y procesable por máquina.
- ✓ Para dar una descripción formal de uno o más documentos.

La información contenida en documentos de esquema a menudo es utilizada por los sistemas de edición compatibles con XML para que puedan ofrecer a los usuarios los elementos con más probabilidades de ocurrir en cualquier lugar dado en un documento.

Comprobación de un documento contra un esquema es lo que se conoce como la validación contra ese esquema. Validar contra un esquema es un elemento importante dentro del aseguramiento de la calidad.



El Lenguaje de Modelado de Servicio (SML) proporciona un marco para relacionar múltiples documentos XSD a uno o más documentos en un solo episodio de validación.

Dado que el XSD soporta la asociación de tipos de datos con el contenido de atributos y elementos, también se utiliza para el enlace de datos, es decir, para los componentes de software que leen y escriben representaciones XML de los objetos del lenguaje de programación.

### **3.2.3 Espacio de nombres**

Un espacio de nombre XML es identificado mediante una referencia IRI; los nombres de elemento y atributo pueden ser colocados en un espacio de nombre XML mediante los mecanismos descritos más adelante.

Un nombre expandido es un par constituido por un nombre de espacio de nombre y un nombre local. Para un nombre N en un espacio de nombre identificado mediante un IRI I, el nombre del espacio de nombre es I. Para un nombre N que no está en un espacio de nombre, el nombre del espacio de nombre no tiene asignación de valor. Definición: En ambos casos el nombre local es N. Es esta combinación del espacio de nombre IRI manejado universalmente y los nombres locales del vocabulario la que resulta efectiva para evitar las colisiones de nombres.

Las referencias IRI pueden contener caracteres que no están permitidos en los nombres, y frecuentemente son inconvenientemente largas, de manera que los nombres expandidos no son usados directamente para nombrar elementos y atributos en los documentos XML. En cambio, se utilizan “nombres cualificados”. Un nombre cualificado es un nombre sujeto a la interpretación del espacio de nombre. En los documentos que se adecuan a esta especificación, los nombres de elementos y atributos aparecen como nombres cualificados. Sintácticamente, son nombres con prefijo o nombres sin prefijo. Se provee una sintaxis de declaración basada en atributos para vincular los prefijos a los nombres de espacio de nombre y para vincular un espacio de nombre por defecto que se aplica a los nombres de elementos sin prefijo; éstas declaraciones están limitadas a los elementos en los cuales aparecen, de modo que pueden darse distintas vinculaciones en distintas partes de un documento. Los procesadores que se adecuan a esta especificación deben reconocer y actúan sobre estas declaraciones y prefijos.





Definición: Un espacio de nombre (o más precisamente, un vínculo de espacio de nombre) se declara utilizando un conjunto de atributos reservados. El nombre de este atributo debe ser **xmlns** o comenzar con **xmlns:** Estos atributos, como cualquier otro atributo XML, se pueden proporcionar directamente o pueden tener un valor por defecto.

El valor normalizado del atributo debe ser una referencia IRI - el nombre de espacio de nombre que identifica al espacio de nombre - o una cadena vacía. El nombre del espacio de nombre, para servir a la finalidad deseada, debería tener las características de unicidad y persistencia. El objetivo es que no sea directamente usable para la obtención de un esquema (si es que existe alguno). Sin embargo, debería tenerse en cuenta que los URL ordinarios pueden utilizarse de modo tal que se alcancen estos mismos objetivos.

Si el nombre de atributo concuerda con *PrefixedAttName*, entonces el *NCName* da el prefijo del espacio de nombre, usado para asociar nombres de elementos y atributos con el nombre del espacio de nombre dado por el valor del atributo en el ámbito del elemento al cual está ligada la declaración.

Si el nombre del atributo concuerda con *DefaultAttName*, entonces el nombre de espacio de nombre dado por el valor del atributo es aquel del espacio de nombres por defecto en el ámbito del elemento al cual está ligada la declaración.

### 3.2.4 Tipos de datos

- ✓ String

El tipo de datos de *cadena* representa cadenas de caracteres en XML. El espacio de valores del tipo *cadena* es el conjunto de secuencias de longitud finita de cero o más caracteres. Un carácter es una unidad atómica de comunicación en XML.

La igualdad de las cadenas *string* es la identidad, sin importar el orden de las mismas.

- ✓ Boolean



Representa los valores de la lógica de dos valores, tiene el espacio de valores de dos valores lógicos: verdadero y falso.

- ✓ Decimal

Representa un subconjunto de los números reales, que pueden ser representados por los números decimales. El espacio de valores de decimal es el conjunto de números que se pueden obtener al dividir un número entero por una potencia no negativa de diez, es decir, expresar como  $i/10^n$  donde  $i$  y  $n$  son números enteros y  $n \geq 0$ . La precisión no es reflejada en este espacio de valores; el número 2.0 no es distinto del número 2.00. La relación de orden en decimal es la relación de orden de los números reales, restringida a este subconjunto.

- ✓ Float

El tipo de datos float sigue el modelo de la precisión simple de 32 bits tipo de datos de punto flotante de la IEEE [IEEE 754-2008]. Su espacio de valores es un subconjunto de los números racionales. Números de punto flotante a menudo se utilizan para aproximar números reales arbitrarios.

El espacio de valor para los flotantes contiene los números no nulos  $m \times 2^e$ , donde  $m$  es un número entero cuyo valor absoluto es menor que 224, y  $e$  es un número entero entre -149 y 104, inclusive.

- ✓ DateTime

Representa instantes de tiempo, opcionalmente marcados con un desfase de zona horaria en particular. Valores que representan el mismo instante, pero que tienen diferentes desfases de zona de tiempo son iguales, pero no idénticos.

- ✓ Date

Representa intervalos de exactamente un día de duración de las líneas de tiempo del DateTime comenzando al momento inicial de cada día, hasta, pero no incluyendo el momento de inicio del día siguiente



### 3.2.5 Componentes

- ✓ Ordenado (ordered)

Para algunos tipos de datos, este documento especifica una relación de orden por sus espacios de valor; la faceta ordenada refleja esto. Toma los valores *total*, *parcial* y *false*, con el significado que se describen a continuación. Para los tipos primitivos de datos, el valor de la faceta ordenado se especifica en facetas fundamentales. Para ordinarios tipos de datos, el valor se hereda sin cambio con respecto al tipo de base. Para obtener una lista, el valor siempre es falsa; para una unión, el valor se calcula como se describe a continuación.

Un valor *false* significa que no se prescribe un orden; un valor *total* asegura que el orden prescrito es un orden total; un valor *parcial* significa que el orden prescrito es un orden parcial, pero no un orden total.

- ✓ Bounded (limitado)

Algunos tipos de datos ordenados tienen la propiedad de que hay un valor mayor que o igual a cualquier otro valor, y otro que es menor que o igual a cualquier otro valor. En el caso de tipos de datos ordinarios, estos dos valores no están necesariamente en el espacio de valores del tipo de datos derivados, pero siempre van a *estar* en el espacio de valores del tipo de datos primitivo del que se han derivado.) La faceta bounded es booleana y su valor es *true* en general para tales tipos de datos limitados. Sin embargo, permanecerá *false* cuando el mecanismo para la imposición de un límite de este tipo sea difícil de detectar, como, por ejemplo, cuando la acotación se produce debido a la derivación utilizando un componente de patrón.

- ✓ Cardinality (cardinalidad)

Cada espacio de valores tiene un número específico de miembros. Este número puede ser caracterizado como finito o infinito. (En la actualidad no existen tipos de datos con espacios más grandes que un valor finito llamado *countable*). La faceta cardinalidad se define como infinito numerable, aunque en general es finita para tipos de datos con espacios valor finito.