



Reading XML files with SAS



Reading XML files with SAS

- Basic's of XML Files
- XML Map
- XML Libname statement
- XML mapper

Understanding XML Documents

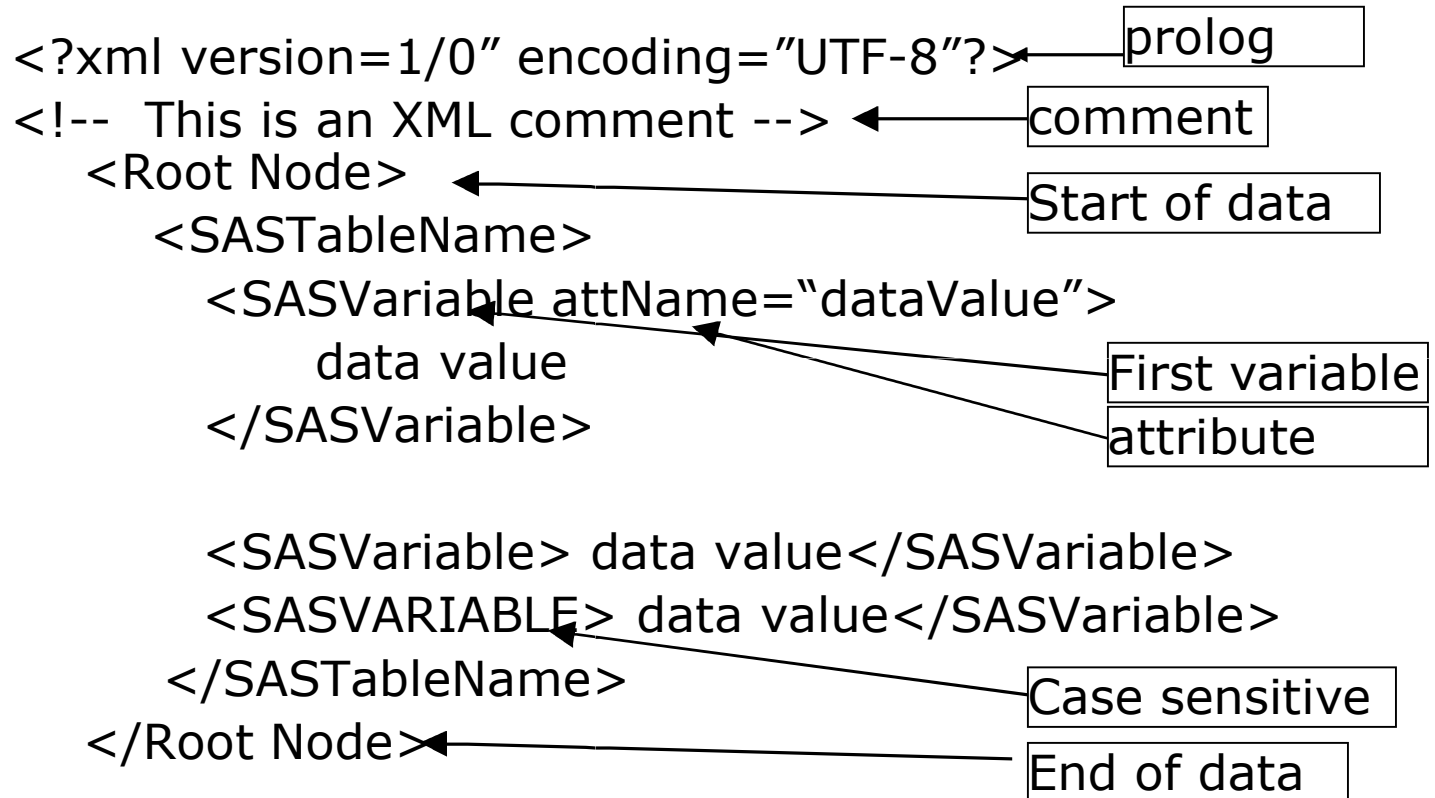
- 5 Rules of a well formed Document
 - One root element
 - Attribute values must be in quotes
 - Tags are case sensitive
 - Starting tags must have ending tags
 - Tags must be properly nested



Pros and Cons of XML Files

- Pros
 - Open Standard
 - Interoperability
- Cons
 - Very verbose

Understanding XML Files



XML Restrictions

- Names can contain letters or numbers
- Names cannot start with XML (in any case format)
- Names must begin with a letter or underscore “_”

XML Data Restrictions

Entity	Character
<	<
>	>
&	&
"	"
'	'

- Helpful hint any Unicode character can be referenced

Reading an XML File

```
<?xml version="1.0" encoding="utf-8" ?>
<FileDescription>
  <simple_table>
    <text_ex> This is obviously text </text_ex>
    <date_ex> 02-MAR-2010 </date_ex>
    <num> 42 </num>
    <date_EX> 2010MAR03 </date_EX>
  </simple_table>
</FileDescription>
```

- To read use xml Libname statement
libname xmlLIB xml 'simple.xml';

Reviewing output

- `proc contents data=xmlLIB.simple_table order=varnum; run;`
- `proc print data=xmlLIB.simple_table; run;`

- **Partial Proc Contents**

#	Variable	Type	Len	Format	Informat	Label
1	DATE_EX1	Num	8	IS8601DA.	ANYDTDTE.	DATE_EX1
2	NUM	Num	8	F8.	F8.	NUM
3	DATE_EX0	Num	8	IS8601DA.	DATE.	DATE_EX0
4	TEXT_EX	Char	22	\$22.	\$22.	TEXT_EX

DATE_EX1	NUM	DATE_EX0	TEXT_EX
2010-03-03	42	2010-03-02	This is obviously text

Complex XML Files

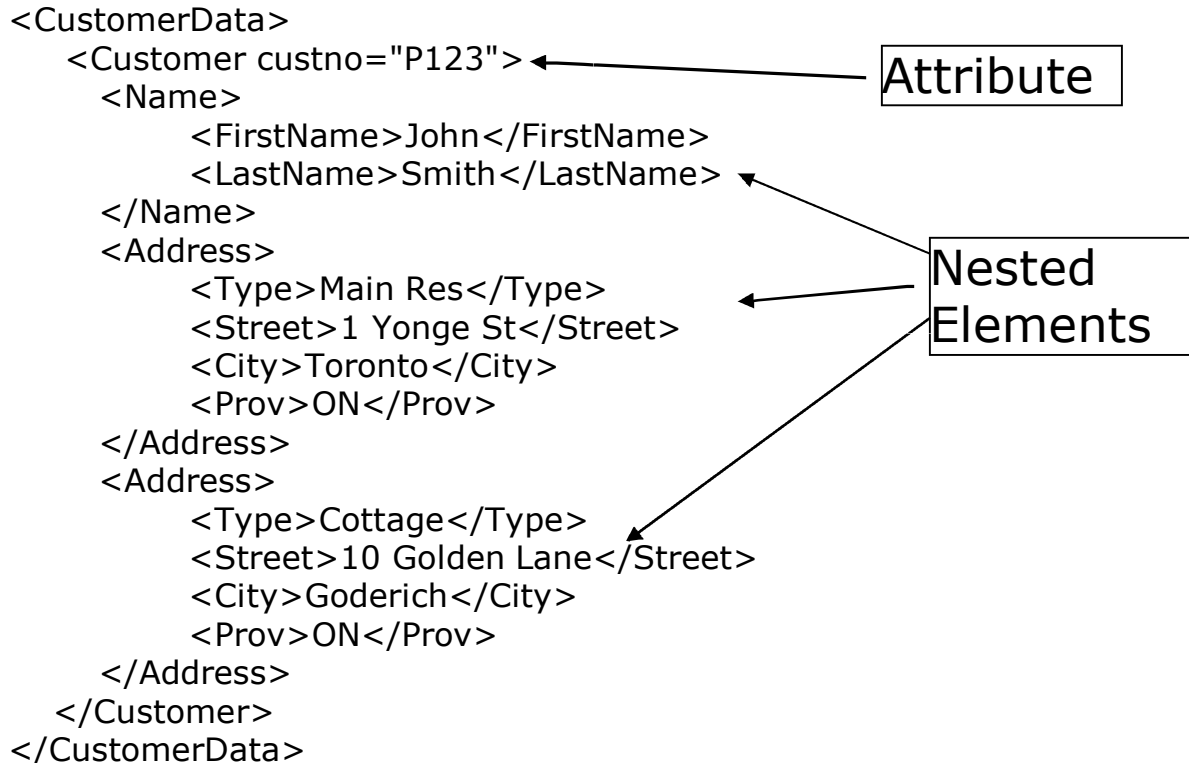
- Defination

- Attributes
- Nested elements

- How to Handle

- A map of instructions to read the file or a .MAP file

Complex XML



XML MAPS

- **Partial XML MAP**

```
<SXLEMAP name="CustomerData" version="1.2">
```

```
<TABLE name="Customer">
```

```
<TABLE-PATH syntax="XPath">/CustomerData/Customer</TABLE-PATH>
```

```
<COLUMN name="custno">
```

```
<PATH syntax="XPath">/Customer/@custno</PATH>
```

```
<TYPE>character</TYPE>
```

```
<DATATYPE>string</DATATYPE>
```

```
<LENGTH>6</LENGTH>
```

```
</COLUMN>
```

```
<COLUMN name="FirstName">
```

```
<PATH syntax="XPath">/CustomerData/Customer/Name/FirstName</PATH>
```

```
<TYPE>character</TYPE>
```

```
<DATATYPE>string</DATATYPE>
```

```
<LENGTH>15</LENGTH>
```

```
</COLUMN>
```

SAS Table
Name

SAS
Column
Names

Attribute

Element

XML MAPs

```
<COLUMN name="LastName">
  <PATH syntax="XPath">/CustomerData/Customer/Name/LastName</PATH>
  <TYPE>character</TYPE>
  <DATATYPE>string</DATATYPE>
  <LENGTH>5</LENGTH>
</COLUMN>

<COLUMN name="Cust_ordinal" ordinal="YES" retain="YES">
  <INCREMENT-PATH beginend="BEGIN"
    syntax="XPath">/CustomerData/Customer</INCREMENT-PATH>
  <TYPE>numeric</TYPE>
  <DATATYPE>integer</DATATYPE>
</COLUMN>

</TABLE>
```

XML MAPs Continued

```
<!-- ##### -->
<TABLE name="Address">
  <TABLE-PATH syntax="XPath">/CustomerData/Customer/Address</TABLE-PATH>

  <COLUMN name="Type">
    <PATH syntax="XPath">
      /CustomerData/Customer/Address/Type</PATH>
    <TYPE>character</TYPE>
    <DATATYPE>string</DATATYPE>
    <LENGTH>8</LENGTH>
  </COLUMN>

  <COLUMN name="Street">
    <PATH syntax="XPath">
      /CustomerData/Customer/Address/Street</PATH>
    <TYPE>character</TYPE>
    <DATATYPE>string</DATATYPE>
    <LENGTH>14</LENGTH>
  </COLUMN>
```

XML MAPs Continued

```
<COLUMN name="City">
  <PATH syntax="XPath">/CustomerData/Customer/Address/City</PATH>
  <TYPE>character</TYPE>
  <DATATYPE>string</DATATYPE>
  <LENGTH>8</LENGTH>
</COLUMN>

<COLUMN name="Prov">
  <PATH syntax="XPath">/CustomerData/Customer/Address/Prov</PATH>
  <TYPE>character</TYPE>
  <DATATYPE>string</DATATYPE>
  <LENGTH>2</LENGTH>
</COLUMN>

<COLUMN name="Cust_ordinal" ordinal="YES">
  <INCREMENT-PATH beginend="BEGIN"
    syntax="XPath">/CustomerData/Customer</INCREMENT-PATH>
  <TYPE>numeric</TYPE>
  <DATATYPE>integer</DATATYPE>
</COLUMN>

</TABLE>

</SXLEMAP>
```

Reading Complex XML Files

- filename customer `customer.xml`;
- filename SXLEMAP 'customer.map';
- libname customer xml xmlmap=SXLEMAP
access=READONLY;

Proc print of Customer

Customer Table

Obs	custno	FirstName	LastName	Cust_ordinal
1	P123	John	Smith	1

Address Table

Obs	Type	Street	City	Prov	Cust_ordinal
1	Main Res	1 Yonge St	Toronto	ON	1
2	Cottage	10 Golden Lane	Goderich	ON	1

What's the Ordinal

- SAS can only handle simple xml so in order to handle nested data ordinals are created to be able to join the data in the tables.
- Ordinals are simple counters that increment the data member count

XML Path

Customer

```
<TABLE-PATH  
  syntax="XPath">/CustomerData/Customer</TABLE-  
  PATH>  
<PATH syntax="XPath">  
  /CustomerData/Customer/@custno</PATH>
```

Common

```
<INCREMENT-PATH beginend="BEGIN"  
  syntax="XPath">/CustomerData/Customer</INCRE  
  MENT-PATH>
```

Address

```
<TABLE-PATH syntax="XPath">  
  /CustomerData/Customer/Address</TABLE-PATH>  
<PATH syntax="XPath">  
  /CustomerData/Customer/Address/Type</PATH>
```

Reading Multiple XML Files

Start of allCustomers.xml

```
<Customer> .....data.....</Customer>
```

```
<Customer> .....data.....</Customer>
```

```
<Customer> .....data.....</Customer>
```

```
<Customer> .....data.....</Customer>
```

```
<Customer> .....data.....</Customer>
```

```
<Customer> .....data.....</Customer>
```

End of allCustomer.xml

```
filename customer 'customer.xml';
```

```
filename SXLEMAP 'customer.map';
```

```
libname customer xml xmlmap=SXLEMAP  
access=READONLY CONCAT=YES;
```

Helpful XML Tools

- XMLMapper – provided by SAS
- Can be downloaded from SAS
 - <http://www.sas.com/apps/demosdownloads/92 SDL sysdep.jsp?packageID=000513>
- Altova XML Spy
 - 30 day trial available for free
- Eclipse
 - <http://www.eclipse.org/downloads/>

XML Mapper

The screenshot displays the SAS XML Mapper application window. The title bar shows 'SAS XML Mapper' and the URL 'T0A08XRDTW60260.ad.cibc.com'. The interface includes a menu bar (File, Tools, Help), a toolbar, and a main workspace divided into several panes. On the left, a tree view shows the XML structure: CustomerData [1] {1} containing Customer [1] {1}, which has Attributes [1] {1} (with custno [1] {1}), Name [1] {1}, and Address [2] {2}. A teal box labeled 'DATA Window' is overlaid on this tree. On the right, a 'Properties' pane shows settings for 'Normal', 'Ordinal', 'Filename', and 'Filepath'. Below it, a tree view shows the mapped XML structure: CustomerData containing Customer (with custno, FirstName, LastName, Cust_ordinal) and Address (with Type, Street, City, Prov, Cust_ordinal). A teal box labeled 'XML Mapper Window' is overlaid on this tree. At the bottom, a 'Table View, Validation' pane shows a table with columns Type, Street, City, Prov, and Cust_ordinal. A warning icon and message state: 'SAS formats and informats are not applied to this view.' Below the table, a teal box labeled 'Table View, Validation' is overlaid.

Table: Address Row: 1 / 2 Columns: 1 / 5

Type	Street	City	Prov	Cust_ordinal
1 Main Res	1 Yonge St	Toronto	ON	1
2 Cottage	10 Golden Lane	Goderich	ON	1

Creating a MAP with XML Mapper

The screenshot displays the SAS XML Mapper application window. The title bar reads "SAS XML Mapper" and the menu bar includes "File", "Tools", and "Help". Below the menu bar is a toolbar with several icons. The main workspace is divided into two panes. The left pane shows a hierarchical tree view of the XML structure:

- CustomerData [1] {1}
- Customer [1] {1}
- Attributes [1] {1}
- custno [1] {1}** (highlighted)
- Name [1] {1}
- Address [2] {2}

The right pane is titled "Properties" and contains fields for "Name" (set to "SXLEMAP"), "Description", "Path", and "End Path". Below these fields are checkboxes for "Retain" and "Replace". A context menu is open over the "SXLEMAP" node in the tree, listing the following actions:

- Get
- Put
- Insert** (highlighted)
- Delete
- Create row count ordinal
- Move first
- Move up
- Move down
- Move last
- Expand

Creating a MAP with XML Mapper

The screenshot displays the SAS XML Mapper application window. The main area is divided into two panes. The left pane shows a hierarchical tree view of an XML schema. The root element is 'CustomerData [1] {1}', which contains a 'Customer [1] {1}' element. This 'Customer' element has an 'Attributes [1] {1}' group containing a 'custno [1] {1}' attribute, and two other elements: 'Name [1] {1}' and 'Address [2] {2}'. The right pane shows the 'Properties' tab for the selected 'custno' attribute. The 'Type' is set to 'character' with a 'Length' of 4. The 'Datatype' is 'string'. Below this, there are fields for 'Name', 'Width', and 'NDec', with 'Width' and 'NDec' both set to 0. At the bottom of the window, there is a 'Table view' tab showing a table with one row and one column. The table header is 'custno' and the data cell contains '1 P123'. A status bar at the bottom indicates 'Table: Customer Row: 1 / 1 Columns: 1 / 1' and includes a warning icon with the text 'SAS formats and informats are not applied to this view.'

CustomerData [1] {1}

- Customer [1] {1}

 - Attributes [1] {1}

 - custno [1] {1}**

- Name [1] {1}
- Address [2] {2}

Properties

Type: character Length: 4

Datatype: string

Name: Width: 0 NDec: 0

Format: Informat: 0 0

SXLEMAP

- Customer

 - custno**

Table: Customer Row: 1 / 1 Columns: 1 / 1

Warning: SAS formats and informats are not applied to this view.

custno
1 P123

Many Thanks and References

- SAS 9.1.3 XML Libname Engine
- SAS and especially Chevell Parker
- www.lexjansen.com with his many papers and examples