Short guide to understand XML

Bo Sundman February 27, 2023

1 Introduction

The XML format is so flexible there is no simple introduction how to use it, a lot is up to the user. I have with the help of Richard Otis tried to summarize the basic features, based on his work on converting TDB files to XML files. This is not exhaustive and I may have made errors and mistakes. The terminology is also important in order to avoid misunderstandings.

The letters < and > are used to identify XML directives in the XML file. An XML directive starts with "<XMLelement" and can end with "/>" or "</XMLelement>" as shown below. A nested XMLelement1 can have a partial ending using a simple ">" followed by other XMLelements, data or text. In such a case it is ended by </XMLelement1>.

Note there should be no space between "/" and ">" but in this PDF document such a space is sometimes generated by LaTeX.

One or more single spaces are used to separate parts. The XMLelements are case insensitive. See also the example in section 3.

- An simple XMLelement is written <XMLelement attribute="value" $\cdots$ />. The attribute is followed by an equal sign and its value are enclosed by double quotations such as "value".
- A nested XMLelement has other XML elements as "childs" and is written:
  <XMLelement1 attribute="value" $\cdots$ >
  <XMLsimplelement2 attribute="another value" $\cdots$/>
  any other text (or XMLelements nested or simple)
  </XMLelement1>

  The nesting can be any levels deep. The fact that a simple XMLelement is terminated by a simple "/>" and a nested one by </XMLelement>, confused me initially.
- A "keyword" in a TDB file corresponds roughly to a simple XMLelement. In the XML file the TDB keyword may be divided into several nested XMLelements to simplify consistency checks.
- The XMLelements and its attributes should, at least partially, be defined by a schema file with extention ".rng". Anyting outside an XMLelement and its attributes are either ignored by XML or considered as an error. But it can be inportant content for the TDB database.

- The XMLelements and attributes used by XML are case insensitive.
- The "value" of an attribute of an XMLelement can be used in other XMLelements, for example names of functions. This is useful for consistency tests and finding spelling errors.

2 More things to think about

There are many issues to be considered for implementation in the XMLTDB format, for example:

- Defining the "model parameter identifiers" such as "G", "TC", "BMAGN", "LNTH" etc. but also less used volume parameters and non-thermodynamic data such as "MQ" must be considered carefully. Some of them can be constituent dependent.
- How to define and handle models like magnetism, low T vibrational entropy etc. There are several variants and a model may have specific "model parameter identifiers" such as "TC" or "BMAGN".
- There are also complicated model issues such a "disordered part" of a phase or "permutations" of model parameters for FCC ordering. Such things may or may not be implemented in some software and if so, probably very differently.
- There must also be flexibility for the database manager to add comments whereever he/she thinks necessary. The bibliographic reference for each parameter is not sufficient.
- Maybe duplicate datasets for a phase using different models could be accomodated in the same XML file? For example an FCC phase modeled with or without ordering.

As written in the project application I do not think that the any of the current thermodynamic software will use the XML database directly or provide XMLTDB database files to the users. This format is intended for the database managers to develop, document, extend and verify their multicomponent databases. Considerating in particular that the next database manager will be able to continue his/her work.

In addition to design the format this project will provide a free software to APPEND/MERGE two XMLTDB files with an internal check and another software to UPLOAD/DOWNLOAD a "standard" TDB file to/from an XMLTDB file.

The hope is that each software developer, who also maintains databases, will provide a routine to can convert their TDB dialect to an XMLTDB database. It will also be an appeal to journals which publish assessments that the paper should provide an XMLTDB file of the model parameters as supplementary material.

3 Example of FeNi in the XML format used by PyCalphad

This is an example, the final XMLTDB format will most likely be different. It will also depend on how much testing of the internal consistency of the file that will be implemented.

```
<?xml version="1.0"?>
<?xml-model href="database.rng" schematypens="http://relaxng.org/ns/structure/1.0"
   type="application/xml"?>
<Database version="0">
  <metadata>
    <writer>pycalphad 0.10.2.dev0+gebcfbdb4.d20220530</writer>
  </metadata>
  <ChemicalElement id="/-" mass="0.0" reference_phase="ELECTRON_GAS" H298="0.0" S298="0.0"/>
  <ChemicalElement id="FE" mass="55.847" reference_phase="BCC_A2" H298="4489.0" S298="27.28"/>
  <ChemicalElement id="NI" mass="58.69" reference_phase="FCC_A1" H298="4787.0" S298="29.796"/>
  <ChemicalElement id="VA" mass="0.0" reference_phase="VACUUM" H298="0.0" S298="0.0"/>
  <Expr id="G0FCCFE">
    <Interval in="T" lower="298.15" upper="6000.0">-1513.82 - 0.0034916*T**2.0
 - 2.86342032e-11*T**4.0</Interval>
  </Expr>
  <Expr id="G0SERNI">
    <Interval in="T" lower="298.15" upper="6000.0">-8333.63278 - 0.00311343009*T**2.0
    - 1.7331937e-07*T**3.0</Interval>
  </Expr>
  <Expr id="GEFCCFE">
    <Interval in="T" lower="0.1" upper="6000.0">GEFCCFE3 + 12.47175*TEFCCFE</Interval>
  </Expr>
  <Expr id="GEFCCFE1">
    <Interval in="T" lower="0.1" upper="6000.0">1.0
   - 1.0*2.71828182845905**(-1.0*T**(-1.0)*TEFCCFE)</Interval>
  </Expr>
  <Expr id="GEFCCFE2">
    <Interval in="T" lower="0.1" upper="6000.0">ln(GEFCCFE1)</Interval>
  </Expr>
  <Expr id="GEFCCFE3">
    <Interval in="T" lower="0.1" upper="6000.0">24.9435*T*GEFCCFE2</Interval>
  </Expr>
  <Expr id="GEINNI">
    <Interval in="T" lower="0.1" upper="6000.0">GEINNI3 + 12.47175*TEINNI</Interval>
  </Expr>
  <Expr id="GEINNI1">
    <Interval in="T" lower="0.1" upper="6000.0">1.0 -
```

```
 1.0*2.71828182845905**(-1.0*T**(-1.0)*TEINNI)</Interval>
</Expr>
<Expr id="GEINNI2">
  <Interval in="T" lower="0.1" upper="6000.0">ln(GEINNI1)</Interval>
</Expr>
<Expr id="GEINNI3">
  <Interval in="T" lower="0.1" upper="6000.0">24.9435*T*GEINNI2</Interval>
</Expr>
<Expr id="GHFCCFE">
  <Interval in="T" lower="298.15" upper="6000.0">G0FCCFE + GEFCCFE + MRFCCFE</Interval>
</Expr>
<Expr id="GHSERNI">
  <Interval in="T" lower="298.15" upper="6000.0">G0SERNI + GEINNI + MRSERNI</Interval>
</Expr>
<Expr id="MRFCCFE">
  <Interval in="T" lower="298.15" upper="6000.0">0.0</Interval>
</Expr>
<Expr id="MRSERNI">
  <Interval in="T" lower="298.15" upper="6000.0">0.0</Interval>
</Expr>
<Expr id="RTEMP">
  <Interval in="T" lower="298.15" upper="6000.0">0.120271814300319*TEMP</Interval>
</Expr>
<Expr id="TEFCCFE">
  <Interval in="T" lower="298.15" upper="6000.0">302.0</Interval>
</Expr>
<Expr id="TEINNI">
  <Interval in="T" lower="298.15" upper="6000.0">284.0</Interval>
</Expr>
<Expr id="TEMP">
  <Interval in="T" lower="298.15" upper="6000.0">T**(-1.0)</Interval>
</Expr>
<Expr id="UN_ASS">
  <Interval in="T" lower="298.15" upper="300.0">0.0</Interval>
</Expr>
<Phase id="FCC_A1">
  <Model type="CEF">
    <ConstituentArray>
      <Site id="0" ratio="1.0">
        <Constituent refid="FE"/>
        <Constituent refid="NI"/>
      </Site>
      <Site id="1" ratio="1.0">
        <Constituent refid="VA"/>
```

```
    </Site>
  </ConstituentArray>
  <MagneticOrdering type="IHJ" structure_factor="0.25" afm_factor="0.0"/>
</Model>
<Parameter type="G">
  <Order>0</Order>
  <ConstituentArray>
    <Site refid="0">
      <Constituent refid="FE"/>
    </Site>
    <Site refid="1">
      <Constituent refid="VA"/>
    </Site>
  </ConstituentArray>
  <Interval in="T" lower="298.15" upper="6000.0">GHFCCFE</Interval>
</Parameter>
<Parameter type="TC">
  <Order>0</Order>
  <ConstituentArray>
    <Site refid="0">
      <Constituent refid="FE"/>
    </Site>
    <Site refid="1">
      <Constituent refid="VA"/>
    </Site>
  </ConstituentArray>
  <Interval in="T" lower="298.15" upper="6000.0">-192.0</Interval>
</Parameter>
<Parameter type="NT">
  <Order>0</Order>
  <ConstituentArray>
    <Site refid="0">
      <Constituent refid="FE"/>
    </Site>
    <Site refid="1">
      <Constituent refid="VA"/>
    </Site>
  </ConstituentArray>
  <Interval in="T" lower="298.15" upper="6000.0">192.0</Interval>
</Parameter>
<Parameter type="BMAGN">
  <Order>0</Order>
  <ConstituentArray>
    <Site refid="0">
```

```
      <Constituent refid="FE"/>
    </Site>
    <Site refid="1">
      <Constituent refid="VA"/>
    </Site>
  </ConstituentArray>
  <Interval in="T" lower="298.15" upper="6000.0">1.77</Interval>
</Parameter>
<Parameter type="G">
  <Order>0</Order>
  <ConstituentArray>
    <Site refid="0">
      <Constituent refid="NI"/>
    </Site>
    <Site refid="1">
      <Constituent refid="VA"/>
    </Site>
  </ConstituentArray>
  <Interval in="T" lower="298.15" upper="6000.0">GHSERNI</Interval>
</Parameter>
<Parameter type="TC">
  <Order>0</Order>
  <ConstituentArray>
    <Site refid="0">
      <Constituent refid="NI"/>
    </Site>
    <Site refid="1">
      <Constituent refid="VA"/>
    </Site>
  </ConstituentArray>
  <Interval in="T" lower="298.15" upper="6000.0">633.0</Interval>
</Parameter>
<Parameter type="NT">
  <Order>0</Order>
  <ConstituentArray>
    <Site refid="0">
      <Constituent refid="NI"/>
    </Site>
    <Site refid="1">
      <Constituent refid="VA"/>
    </Site>
  </ConstituentArray>
  <Interval in="T" lower="298.15" upper="6000.0">-633.0</Interval>
</Parameter>
```

```
<Parameter type="BMAGN">
  <Order>0</Order>
  <ConstituentArray>
    <Site refid="0">
      <Constituent refid="NI"/>
    </Site>
    <Site refid="1">
      <Constituent refid="VA"/>
    </Site>
  </ConstituentArray>
  <Interval in="T" lower="298.15" upper="6000.0">0.52</Interval>
</Parameter>
<Parameter type="TC">
  <Order>0</Order>
  <ConstituentArray>
    <Site refid="0">
      <Constituent refid="FE"/>
      <Constituent refid="NI"/>
    </Site>
    <Site refid="1">
      <Constituent refid="VA"/>
    </Site>
  </ConstituentArray>
  <Interval in="T" lower="298.15" upper="6000.0">2270.64276</Interval>
</Parameter>
<Parameter type="TC">
  <Order>1</Order>
  <ConstituentArray>
    <Site refid="0">
      <Constituent refid="FE"/>
      <Constituent refid="NI"/>
    </Site>
    <Site refid="1">
      <Constituent refid="VA"/>
    </Site>
  </ConstituentArray>
  <Interval in="T" lower="298.15" upper="6000.0">-430.18647</Interval>
</Parameter>
<Parameter type="TC">
  <Order>2</Order>
  <ConstituentArray>
    <Site refid="0">
      <Constituent refid="FE"/>
      <Constituent refid="NI"/>
```

```
    </Site>
    <Site refid="1">
      <Constituent refid="VA"/>
    </Site>
  </ConstituentArray>
  <Interval in="T" lower="298.15" upper="6000.0">-1671.74725</Interval>
</Parameter>
<Parameter type="TC">
  <Order>3</Order>
  <ConstituentArray>
    <Site refid="0">
      <Constituent refid="FE"/>
      <Constituent refid="NI"/>
    </Site>
    <Site refid="1">
      <Constituent refid="VA"/>
    </Site>
  </ConstituentArray>
  <Interval in="T" lower="298.15" upper="6000.0">-6143.57337</Interval>
</Parameter>
<Parameter type="TC">
  <Order>4</Order>
  <ConstituentArray>
    <Site refid="0">
      <Constituent refid="FE"/>
      <Constituent refid="NI"/>
    </Site>
    <Site refid="1">
      <Constituent refid="VA"/>
    </Site>
  </ConstituentArray>
  <Interval in="T" lower="298.15" upper="6000.0">-10552.4704</Interval>
</Parameter>
<Parameter type="TC">
  <Order>5</Order>
  <ConstituentArray>
    <Site refid="0">
      <Constituent refid="FE"/>
      <Constituent refid="NI"/>
    </Site>
    <Site refid="1">
      <Constituent refid="VA"/>
    </Site>
  </ConstituentArray>
```

```
      <Interval in="T" lower="298.15" upper="6000.0">-6018.86703</Interval>
    </Parameter>
    <Parameter type="NT">
      <Order>0</Order>
      <ConstituentArray>
        <Site refid="0">
          <Constituent refid="FE"/>
          <Constituent refid="NI"/>
        </Site>
        <Site refid="1">
          <Constituent refid="VA"/>
        </Site>
      </ConstituentArray>
      <Interval in="T" lower="298.15" upper="6000.0">59.0347415</Interval>
    </Parameter>
    <Parameter type="BMAGN">
      <Order>0</Order>
      <ConstituentArray>
        <Site refid="0">
          <Constituent refid="FE"/>
          <Constituent refid="NI"/>
        </Site>
        <Site refid="1">
          <Constituent refid="VA"/>
        </Site>
      </ConstituentArray>
      <Interval in="T" lower="298.15" upper="6000.0">0.0</Interval>
    </Parameter>
    <Parameter type="G">
      <Order>0</Order>
      <ConstituentArray>
        <Site refid="0">
          <Constituent refid="FE"/>
          <Constituent refid="NI"/>
        </Site>
        <Site refid="1">
          <Constituent refid="VA"/>
        </Site>
      </ConstituentArray>
      <Interval in="T" lower="298.15" upper="6000.0">0.0</Interval>
    </Parameter>
  </Phase>
</Database>
```

The same content in the normal TDB formmat

```
$ Database file written 2021-3-12
$ For checking the negative S of ordered compounds
 ELEMENT /-   ELECTRON_GAS              0.0000E+00  0.0000E+00  0.0000E+00!
 ELEMENT VA   VACUUM                    0.0000E+00  0.0000E+00  0.0000E+00!
 ELEMENT FE   BCC_A2                    5.5847E+01  4.4890E+03  2.7280E+01!
 ELEMENT NI   FCC_A1                    5.8690E+01  4.7870E+03  2.9796E+01!

 FUNCTION RTEMP     298.15 +R#**(-1)*TEMP#; 6000 N !
 FUNCTION TEMP      298.15 +T**(-1); 6000 N !
 FUNCTION UN_ASS    298.15 +0.0; 300 N !

$ This is Einstein contribution to FCC Fe
 FUNCTION TEFCCFE              298.15 +302;  6000 N !
 FUNCTION GEFCCFE1  0.10 +1-1*EXP(-TEFCCFE#*T**(-1)); 6000 N !
 FUNCTION GEFCCFE2  0.10 +1*LN(GEFCCFE1#); 6000 N !
 FUNCTION GEFCCFE3  0.10 +3*R*T*GEFCCFE2#; 6000 N !
 FUNCTION GEFCCFE   0.10 +1.5*R*TEFCCFE#+GEFCCFE3#; 6000 N !
$============================
$ This is with magnetic entropy zero for FCC Fe at 0 K
 FUNCTION MRFCCFE              298.15 0; 6000 N !
$ This is with magnetic entropy non-zero for FCC Fe at infinite high T
$ FUNCTION MRFCCFE              298.15 +1370.87261-8.47122*T;    6000 N !
$============================
 FUNCTION G0FCCFE              298.15 -1513.82-3.4916E-03*T**2
                                      -2.86342032E-11*T**4; 6000 N !
 FUNCTION GHFCCFE   298.15 +GEFCCFE#+G0FCCFE#+MRFCCFE; 6000 N !

$ This is Einstein contribution to FCC Ni
 FUNCTION TEINNI              298.15 +284;   6000 N !
 FUNCTION GEINNI1  0.10 +1-1*EXP(-TEINNI#*T**(-1)); 6000 N !
 FUNCTION GEINNI2  0.10 +1*LN(GEINNI1#); 6000 N !
 FUNCTION GEINNI3  0.10 +3*R*T*GEINNI2#; 6000 N !
 FUNCTION GEINNI   0.10 +1.5*R*TEINNI#+GEINNI3#; 6000 N !
$================================
$ This is with magnetic entropy zero for FCC Ni at infinite high T
 FUNCTION MRSERNI              298.15 0; 6000 N !
$ This is with magnetic entropy non-zero for FCC Ni at 0 K
$ FUNCTION MRSERNI              298.15 +1857.39449-3.48137127*T; 6000 N !
$================================
 FUNCTION G0SERNI              298.15 -8333.63278-0.00311343009*T**2
                              -1.7331937E-7*T**3;    6000 N !
 FUNCTION GHSERNI   298.15 +GEINNI#+G0SERNI#+MRSERNI#; 6000 N !
```

```
TYPE_DEFINITION % SEQ *!
DEFINE_SYSTEM_DEFAULT ELEMENT 2 !
DEFAULT_COMMAND DEF_SYS_ELEMENT VA /- !

TYPE_DEFINITION & GES A_P_D FCC_A1 MAGNETIC   0.0    0.25000E+00 F !
PHASE FCC_A1  %&  2 1   1 !
   CONSTITUENT FCC_A1  :FE,NI : VA :  !

  PARAMETER G(FCC_A1,FE:VA;0)           298.15 +GHFCCFE#; 6000 N REF0 !
  PARAMETER TC(FCC_A1,FE:VA;0)         298.15 -192; 6000 N REF0 !
  PARAMETER NT(FCC_A1,FE:VA;0)         298.15 +192; 6000 N REF0 !
  PARAMETER BMAG(FCC_A1,FE:VA;0)        298.15 +1.77; 6000 N REF0 !
  PARAMETER G(FCC_A1,NI:VA;0)           298.15 +GHSERNI#; 6000 N REF0 !
  PARAMETER TC(FCC_A1,NI:VA;0)         298.15 +633; 6000 N REF0 !
  PARAMETER NT(FCC_A1,NI:VA;0)         298.15 -633; 6000 N REF0 !
  PARAMETER BMAG(FCC_A1,NI:VA;0)        298.15 +.52; 6000 N REF0 !
  PARAMETER TC(FCC_A1,FE,NI:VA;0)      298.15 +2.27064276E+03; 6000 N REF0 !
  PARAMETER TC(FCC_A1,FE,NI:VA;1)      298.15 -4.30186470E+02; 6000 N REF0 !
  PARAMETER TC(FCC_A1,FE,NI:VA;2)      298.15 -1.67174725E+03; 6000 N REF0 !
  PARAMETER TC(FCC_A1,FE,NI:VA;3)      298.15 -6.14357337E+03; 6000 N REF0 !
  PARAMETER TC(FCC_A1,FE,NI:VA;4)      298.15 -1.05524704E+04; 6000 N REF0 !
  PARAMETER TC(FCC_A1,FE,NI:VA;5)      298.15 -6.01886703E+03; 6000 N REF0 !
  PARAMETER NT(FCC_A1,FE,NI:VA;0)      298.15 +5.90347415E+01; 6000 N REF0 !
  PARAMETER BMAG(FCC_A1,FE,NI:VA;0)     298.15 +0; 6000 N REF0 !
  PARAMETER G(FCC_A1,FE,NI:VA;0)        298.15 +0; 6000 N REF0 !

LIST_OF_REFERENCES
NUMBER  SOURCE
REF0 '1991 Dinsdale, DOI:10.1016/0364-5916(91)90030-N'
!
```