XMLTDB in FactSage Calphad Optimizer and ChemApp: Lessons learned and opportunities generated

F. Tang[1], B. Sundman[2], M. to Baben[1]

[1] GTT-Technologies, Herzogenrath, Germany
[2] OpenCalphad, Gif-sur-Yvette, France

Vote here!

Like
neutral
don‘t care

Current XMLTDB proposal (0.0.6)

```
<Database version="0.0.6">
    <metadata>
        <writer Software="OpenCalphad 6.068" Date="2023-10-26" />
    </metadata>
    <Defaults LowT="10" HighT="6000" Bibref="U.N. Known" Elements="VA /-" />
    <Element Id="AL" Refstate="FCC_A1" Mass="26.982" H298="4577.3" S298="28.322" />
    <Element Id="C" Refstate="GRAPHITE" Mass="12.011" H298="1054" S298="5.7423E+00" />
    <Species Id="VA" Stoichiometry="VA" />
    <Species Id="AL" Stoichiometry="AL" />
    <Species Id="C" Stoichiometry="C" />
    <TPfun Id="R" Expr="8.31451;" />
    <TPfun Id="RTLNP" Expr="R*T*LN(1.0E-5*P);" />
    <TPfun Id="G0AL4C3" Expr=" -277339-.005423368*T**2;" />
    <TPfun Id="GTSERAL" Expr=" -.001478307*T**2-7.83339395E-07*T**3;" />
    <TPfun Id="GTSERCC" Expr=" -.00029531332*T**2-3.3998492E-16*T**5;" />
    <TPfun Id="G0BCCAL" Expr=" +GHSERAL+10083;" />
    <TPfun Id="G0HCPAL" Expr=" +GHSERAL+5481;" />
    <TPfun Id="GHSERAL" Expr=" -8160+GTSERAL;" />
    <TPfun Id="GHSERCC" Expr=" -17752.213+GEGRACC+GTSERCC;" />
    <TPfun Id="G0DIACC" Expr=" -16275.202-9.1299452E-05*T**2-2.1653414E-16*T**5;" />
    <TPfun Id="GEDIACC" Expr=" +0.2318*GEIN(+813.6)+.01148*GEIN(+345.3)-0.236743*GEIN(1601);" />
    <TPfun Id="G0LIQAL" Expr=" -209-3.777*T-.00045*T**2;" />
    <TPfun Id="G0LIQCC" Expr=" +63887-8.2*T-.0004185*T**2;" />
    <TPfun Id="GEGRACC"
        Expr="-0.5159523*GEIN(+1953)+0.121519*GEIN(+448)+0.3496843*GEIN(+947)+
              0.0388463*GEIN(+192.65)+.005840323*GEIN(+64.46);" />
    <Phase Id="LIQUID" Configuration="CEF" State="L">
        <Sites NumberOf="1" Multiplicities="1">
            <Constituents Sublattice="1" List="AL C" />
        </Sites>
        <AmendPhase Models="LIQ2STATE" />
    </Phase>
    <Phase Id="AL4C3" Configuration="CEF" State="S">
        <Sites NumberOf="2" Multiplicities="4 3">
            <Constituents Sublattice="1" List="AL" />
            <Constituents Sublattice="2" List="C" />
        </Sites>
        <AmendPhase Models="GEIN" />
    </Phase>
    <Phase Id="BCC_A2" Configuration="CEF" State="S">
        <Sites NumberOf="2" Multiplicities="1 3">
            <Constituents Sublattice="1" List="AL" />
            <Constituents Sublattice="2" List="C VA" />
        </Sites>
        <AmendPhase Models="GEIN" />
    </Phase>
    <Phase Id="DIAMOND" Configuration="CEF" State="S">
        <Sites NumberOf="1" Multiplicities="1">
            <Constituents Sublattice="1" List="C" />
        </Sites>
        <AmendPhase Models="GEIN" />
    </Phase>
    <Phase Id="FCC_A1" Configuration="CEF" State="S">
        <Sites NumberOf="2" Multiplicities="1 1">
            <Constituents Sublattice="1" List="
            <Constituents Sublattice="2" List="
        </Sites>
        <AmendPhase Models="GEIN" />
    </Phase>
    <Phase Id="GRAPHITE" Configuration="CEF" S
        <Sites NumberOf="1" Multiplicities="1">
            <Constituents Sublattice="1" List="C" />
        </Sites>
        <AmendPhase Models="GEIN" />
    </Phase>
    <Phase Id="HCP_A3" Configuration="CEF" State="S">
        <Sites NumberOf="2" Multiplicities="1 0.5">
            <Constituents Sublattice="1" List="AL" />
            <Constituents Sublattice="2" List="C VA" />
        </Sites>
        <AmendPhase Models="GEIN" />
    </Phase>
    <Parameter Id="G(LIQUID,AL;0)" Expr=" +G0LIQAL;" Bibref="20HE" />
    <Parameter Id="LNTH(LIQUID,AL;0)" Expr=" +LN(254);" Bibref="20HE" />
    <Parameter Id="G2(LIQUID,AL;0)" Expr="+13398-R*T-0.165
    <Parameter Id="G(LIQUID,C;0)" Expr=" +G0LIQCC;" Bibref="20HE" />
    <Parameter Id="LNTH(LIQUID,C;0)" Expr=" +LN(1400);" Bibref="20HE" />
    <Parameter Id="G2(LIQUID,C;0)" Expr="+59147-49.61*T+2.9806*T*LN(+T);" Bibref="20HE" />
    <Parameter Id="G(LIQUID,AL,C;0)" Expr=" +20994-22*T;" Bibref="20HE" />
    <Parameter Id="G(AL4C3,AL:C;0)" Expr=" +G0AL4C3-3.08*GEIN(+401)+3.08*GEIN(+1077);" Bibref="20HE" />
    <Parameter Id="LNTH(AL4C3,AL:C;0)" Expr=" +LN(401);" Bibref="20HE" />
    <Parameter Id="G(BCC_A2,AL:C;0)" Expr=" +GTSERAL+3*GTSERCC+1006844;" Bibref="20HE" />
    <Parameter Id="LNTH(BCC_A2,AL:C;0)" Expr=" +LN(863);" Bibref="20HE" />
    <Parameter Id="G(BCC_A2,AL:VA;0)" Expr=" +G0BCCAL;" Bibref="20HE" />
    <Parameter Id="LNTH(BCC_A2,AL:VA;0)" Expr=" +LN(233);" Bibref="20HE" />
    <Parameter Id="G(BCC_A2,AL:C,VA;0)" Expr=" -819896+14*T;" Bibref="20HE" />
    <Parameter Id="G(DIAMOND,C;0)" Expr=" +G0DIACC+GEDIACC;" Bibref="20HE" />
    <Parameter Id="LNTH(DIAMOND,C;0)" Expr=" +LN(1601);" Bibref="20HE" />
    <Parameter Id="G(FCC_A1,AL:C;0)" Expr=" +GTSERAL+GTSER
    <Parameter Id="LNTH(FCC_A1,AL:C;0)" Expr=" +LN(549);"
    <Parameter Id="G(FCC_A1,AL:VA;0)" Expr=" +GHSERAL;" Bi
    <Parameter Id="LNTH(FCC_A1,AL:VA;0)" Expr=" +LN(283);"
    <Parameter Id="G(FCC_A1,AL:C,VA;0)" Expr=" -70345;" Bibref="
    <Parameter Id="G(GRAPHITE,C;0)" Expr=" +GHSERCC;" Bibref="
    <Parameter Id="LNTH(GRAPHITE,C;0)" Expr=" +LN(1953);" Bibref="20HE" />
    <Parameter Id="G(HCP_A3,AL:C;0)" Expr=" +GTSERAL+0.5*GTSERCC+2176775;" Bibref="20HE" />
    <Parameter Id="LNTH(HCP_A3,AL:C;0)" Expr=" +LN(452);" Bibref="20HE" />
    <Parameter Id="G(HCP_A3,AL:VA;0)" Expr=" +G0HCPAL;" Bibref="20HE" />
    <Parameter Id="LNTH(HCP_A3,AL:VA;0)" Expr=" +LN(263);" Bibref="20HE" />
    <Parameter Id="G(HCP_A3,AL:C,VA;0)" Expr=" 0;" Bibref="20HE" />
    <Bibliography>
        <Bibitem Id="20HE"
            Text="Zhanglong He,
                  Calphad
    </Bibliography>
</Database>
```

Defaults are comfortable, but can be dangerous, e.g. when combining data sets

Very closely modeled after `PHASE` and `CONSTITUENTS` commands in TDBs

Model description is nested in phase tag – where are required model parameters?

Other models: MQMA, I2SL, UNIQUAC, …
What is a configuration model?
E.g. Redlich-Kister, Toop (more?)

A lot of information ‘coded’ into a small data field that is computationally ‘hard’ to decode

Definition and location of `GTSERAL` and `GTSERCC` unclear

All parameter lines can be put anywhere in the file. Convention is: ‘by system’

Suggestions and potential future features

Take advantage of XML properties

- All parameters collected under phase
- Parameters don’t ‘float’ freely in file
- Schema validation can verify structural correctness

```
<Phase Id="BCC_A2" Configuration="CEF" State="S">
    <Sites NumberOf="2" Multiplicities="1 3">
        <Constituents Sublattice="1" List="AL" />
        <Constituents Sublattice="2" List="C VA" />
    </Sites>
    <AmendPhase Models="GEIN" />
    <!- Al unary -->
    <Parameter MPID="G" ConstArray="AL:VA" Expr=" +G0BCCAL;" Bibref="20HE" />
    <Parameter MPID="LNTH" ConstArray="AL:VA" Expr=" +LN(233);" Bibref="20HE" />
    <!-- Al-C binary -->
    <Parameter MPID="G" ConstArray="AL:C" Expr=" +GTSERAL+3*GTSERCC+1006844;" Bibref="20HE" />
    <Parameter MPID="LNTH" ConstArray="AL:C" Expr=" +LN(863);" Bibref="20HE" />
    <Parameter MPID="L" ConstArray="AL:C,VA" Degree="0" Expr=" -819896+14*T;" Bibref="20HE" />
</Phase>
```

As it is nested in the *BCC_A2* phase tag, the `MPID` can omit the phase id

separating `MPID` makes this more explicit

Unification of ‘primitive’ technical requirements

- Phase name lengths: 25 characters? 28 characters? Unlimited?
- (dis)allowed characters
- Can end members have ‘common names’?
- Number precision defaults, esp. for output?
- Prevent circular function references

Modern world requirements: FAIR data support

- Verification: checksums, cryptographic signing
- Change management:
  - ‘beautiful diff’ for versioning
  - tracking changes and maintain evolution
- Ontology adaption (integrate into larger data concepts)
- Comments and other meta-data on data/values

Connect atomistic model to phase model

- Allows description of first principle models
- Synchronization with e.g. VASP input / output

```
<Phase Id="HypotheticalMu" Configuration="CEF" State="S">
  <AmendPhase Models="KeuterCryst Vegard" />
  <Sublattices NumberOf=" 3" Multiplicities=" 7 2 4">
    <Constituents Sublattice=" 1" List=" CO CR FE NI TA HF" />
    <Constituents Sublattice=" 2" List=" MO NB TA W HF" />
    <Constituents Sublattice=" 3" List=" CO CR FE MO NB NI TA W HF" />
  </Sublattices>
  <Crystallography>
    <SpaceGroup number=" 166">
      <!-- General configuration -->
      <Configuration const="*:*:*">
        <Wyckoff Symbol="3b" Sublattice=" 1" />
        <Wyckoff Symbol="18h" x="0.5000" z="0.0900" Sublattice="1" />
        <Wyckoff Symbol="6c1" z="0.0520" Sublattice="2" />
        <Wyckoff Symbol="6c2" z="0.1540" Sublattice="3" />
        <Wyckoff Symbol="6c3" z="0.3333" Sublattice="3" />
        <Parameter MPID="LP" a="4.764" b="4.764" c="25.850" Bibref="" />
        <Parameter MPID="LA" a="90" b="90" c="120" Bibref="" />
      </Configuration>
      <!-- Specific to one endmember -->
      <Configuration const="CO:W:FE" Name="Hypothetical Special Case">
        <Wyckoff Symbol="3b" Sublattice=" 1" />
        <Wyckoff Symbol=" 18h" x=" 0.5001" z=" 0.0900" Sublattice=" 1" />
        <Wyckoff Symbol=" 6c1" z=" 0.0521 " Sublattice="2" />
        <Wyckoff Symbol=" 6c2" z=" 0.1541 " Sublattice="3" />
        <Wyckoff Symbol=" 6c3" z=" 0.3334 " Sublattice="3" />
        <Parameter MPID="LP" a=" 4.765" b=" 4.765" c=" 25.851" Bibref="" />
        <Parameter MPID="LA" a=" 90" b=" 90" c=" 120" Bibref="" />
      </Configuration>
    </SpaceGroup>
  </Crystallography>
</Phase>
```

Key crystallographic information to conceptualize the phase model

Relation of Wyckoff sites and sublattice modeling explicitly

Current status (for proposal v0.0.6):
- Reading/Writing from OpenCalphad
- Reading/Writing from ChemApp
- Schema validation XSL file available

Roadmap:
- Version 1.0.0 by the end of the year
- **Getting Feedback!**

Florian Tang
Contact info

https://github.com/flotang-gtt/XMLTDB

https://github.com/sundmanbo/XMLTDB

Biggest questions

- *Uncertainty quantification*
  If a dataset has been researched and evaluated, how can this information be attached to the model?
  How will this data be supported in the software tools?
- *Stewardship for XMLTDB*
  Is it necessary to have authority over the format?
- *Beyond current database data*
  - Safety margins, assessed modeling ranges
  - More meta data – e.g. warnings, dependencies
- _____!!

GTT-Technologies
Kaiserstraße 103
52134 Herzogenrath, Germany
Phone: +49-(0)2407-59533
E-mail: info@gtt-technologies.de