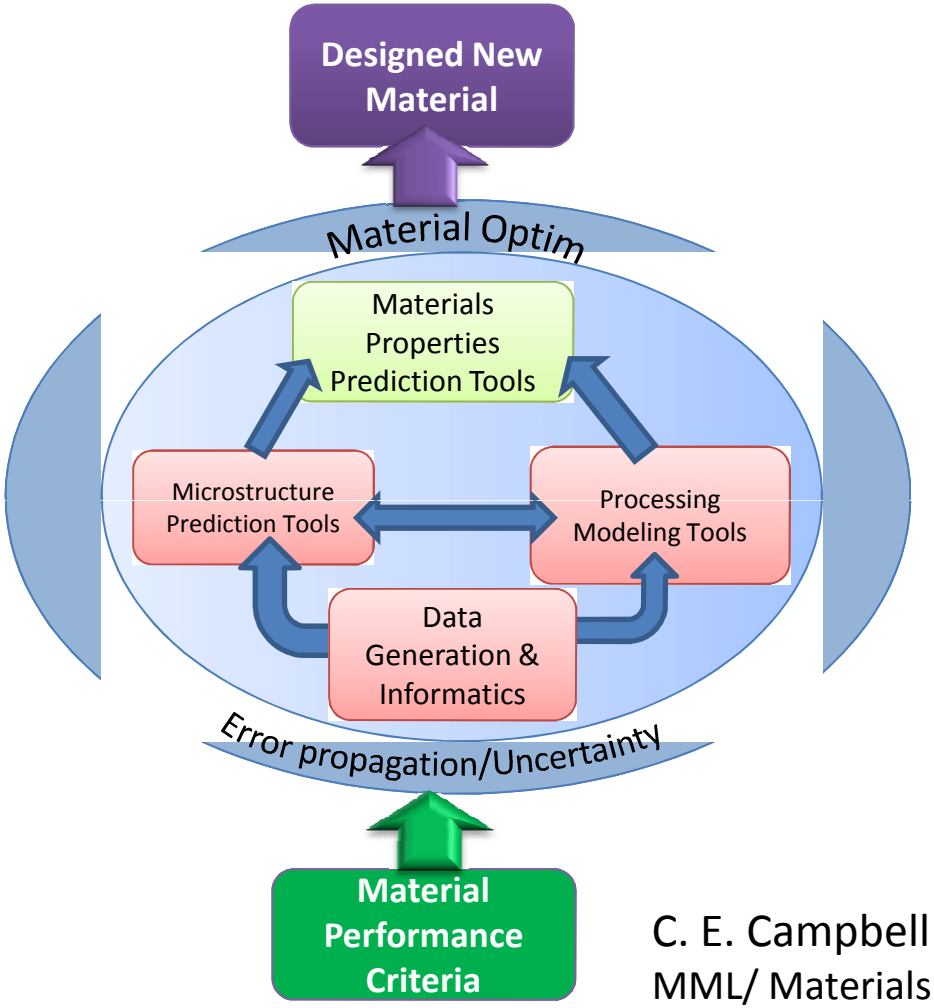
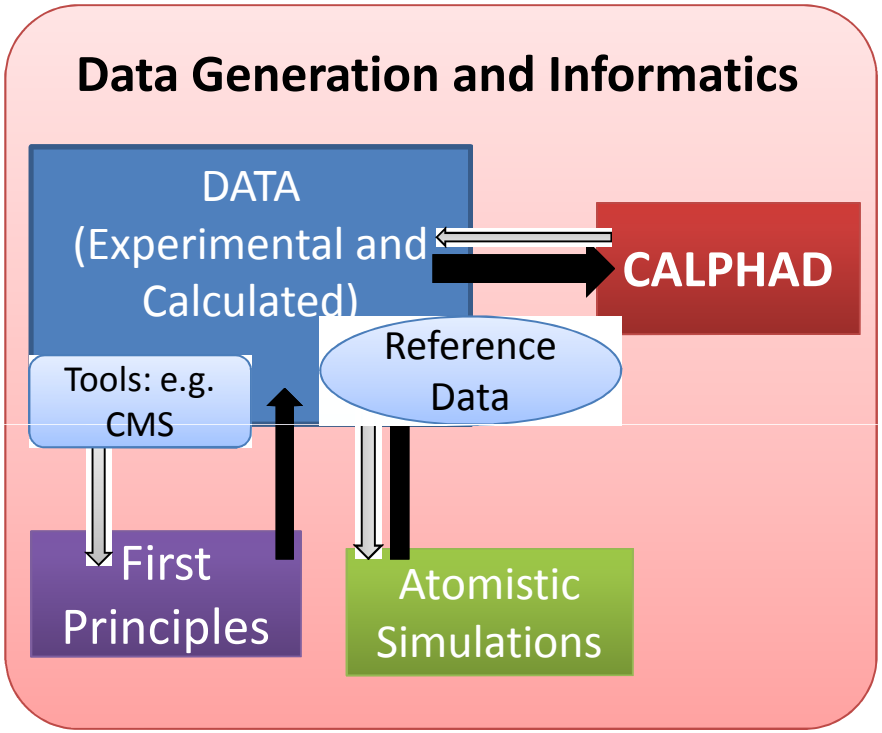
Materials Data and Informatics





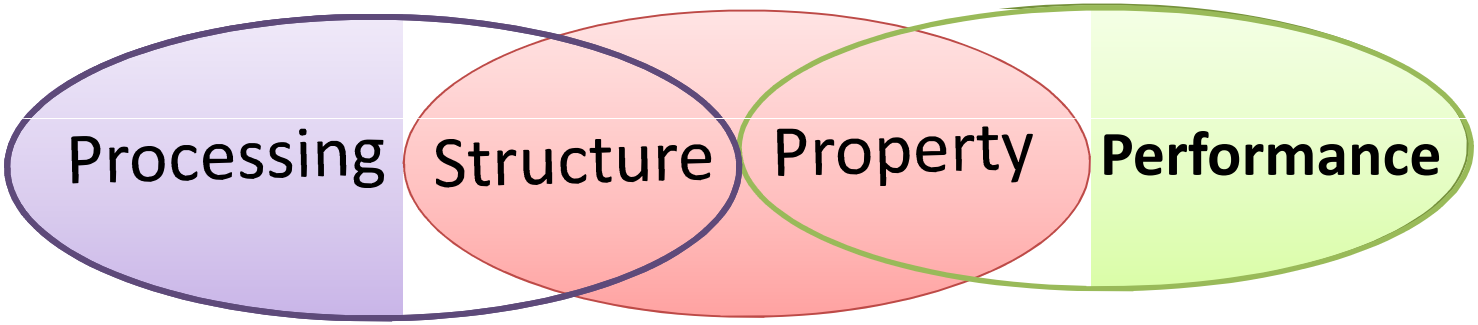
C. E. Campbell and A. Dima
MML/ Materials Science and Engineering Division and ITL/Software and Systems Division

May 2nd, 2013

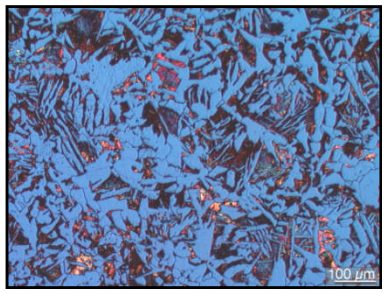
NIST
National Institute of Standards and Technology
Technology Administration
U.S. Department of Commerce

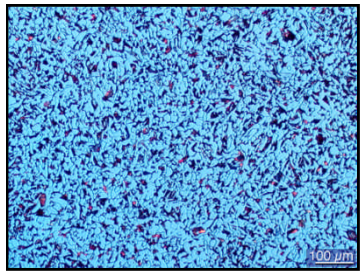
Materials Are Complicated Hierarchical Systems

- Advanced materials often consist of **several components** (generally, $n > 5$) and **multiple phases**.
- The material **properties are dependent on the microstructure.**
- **The microstructures changes as a function of processing and service conditions.**



Material A at Temp 1





Material A at Temp. 2

Key to material design:

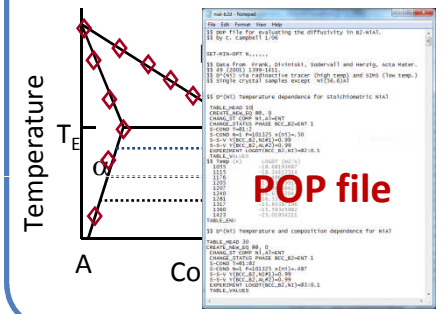
- What phases are present
- Composition and morphology of the phases present

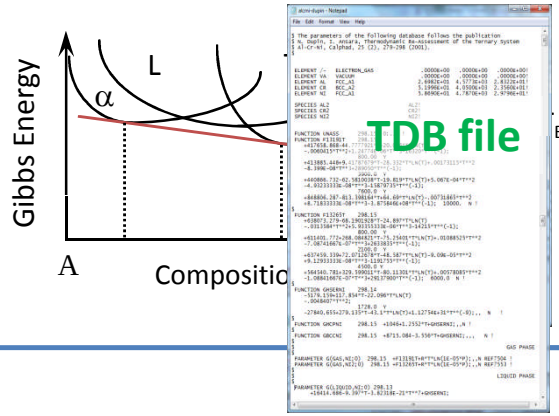
CALPHAD

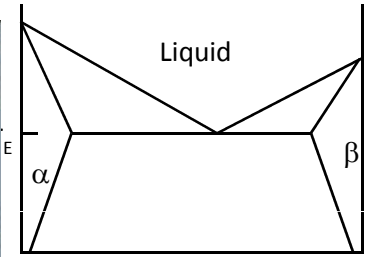
CALPHAD Approach

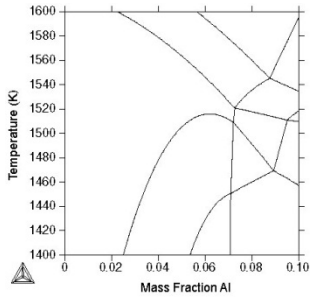
- Collected experimental and computational data are used to fit functions.
- Functions are used to calculate phase equilibria, including phase diagrams.

Experimental phase diagram and thermochemical data → Determine Gibbs energy functions for each phase: $G = f(x,T,P)$ → Calculated phase diagram → Combine binaries and ternaries to predict multi-component systems









Al isopleth for 8 component system

$$G^{\phi} = G^{0} + G^{ideal} + G^{excess}$$

Binaries → Ternaries → Quaternaries → n^{th} Order Systems

True quaternary compounds are rare in metallic systems

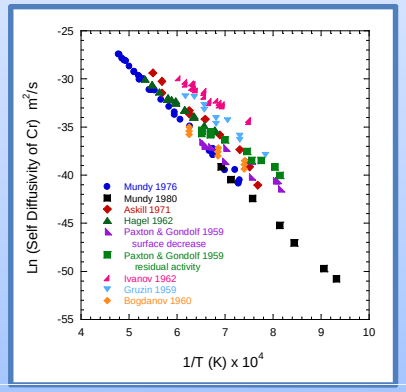
Assessment of ternary systems is usually sufficient for the description of a multicomponent system

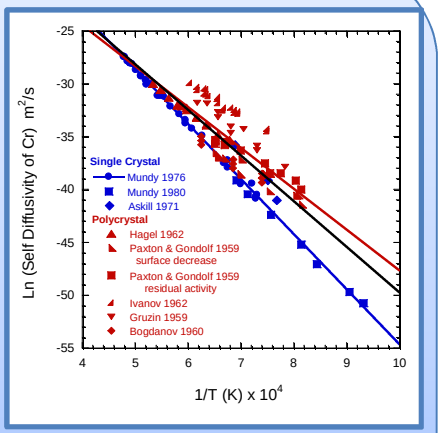
Same methodology can be applied to the description of other property data

Informatics Challenges Associated with CALPHAD data

Need to classify and differentiate data

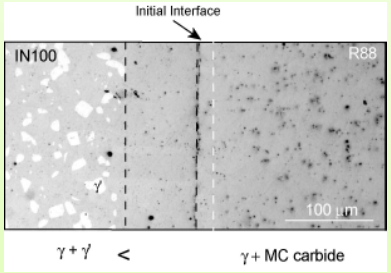
- Need to determine which data to evaluate
- Example: Cr Self-Diffusion Data
  - Separate single crystal from poly crystal data.
  - Only use single crystal data to determine reference values.
- Other examples of sorts include material purity, measurement technique, etc.,





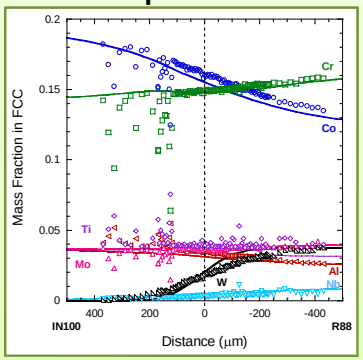
Does experiment = calculation?

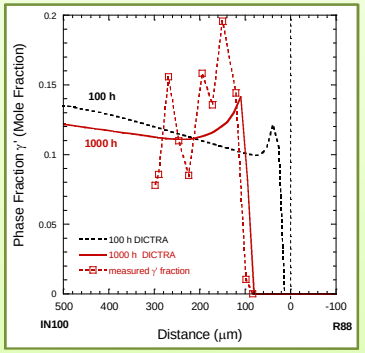
Diffusion example: Does the simulation include all the phases in experiment?



IN100/René-88 at 1150 ºC for 1000 h

If γ' and MC carbide not included, cannot predict the position of the γ+γ'/γ +MC interface correctly





- Do the boundary conditions match the experimental geometry?
- Are the heating and cooling rates used in the experiment included in the calculation?

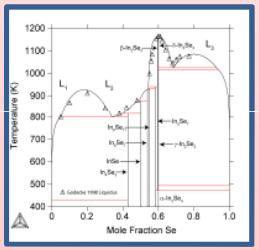
Examples of CALPHAD Data Types

For each assessment: Evaluated data file (e.g., POP, DOP)
Functional descriptions for phase quantity (e.g., TDB)

- Emphasis on binary and ternary data to predict multicomponent properties
- Data can be experimental or computational.

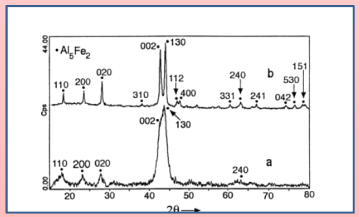
1-D (Points)

Melting Temperatures



Critical Temperatures (Phase Changes)

Lattice Parameters



Heat of Formations

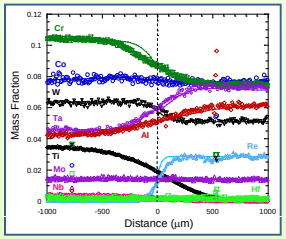
Phase Fractions and Compositions

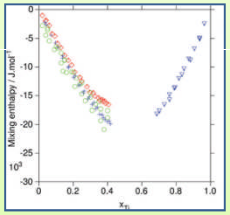
Tracer Diffusivities

Activation Energies

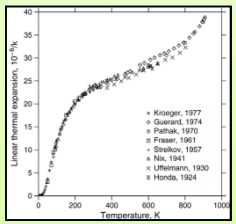
2-D (Lines)

Composition Profiles





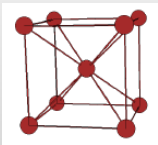
Heat Capacities

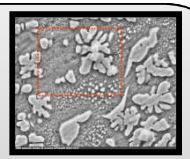


Enthalpies of Mixing

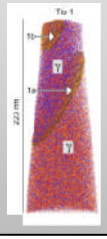
3-D

Crystal Structures



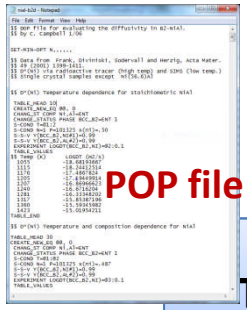


Micrographs/Morphologies



3-D Atom Probe Tomography

Data Flow of CALPHAD "Database" Files



POP file

Evaluated Data Files
(e.g. POP, DOP)

Functional Descriptions
(e.g. TDB files)

Thermodynamic Data

Diffusion Data

Molar Volume Data

Other Data

Thermodynamic Descriptions

Diffusion Mobility Descriptions

Molar Volume Descriptions

...

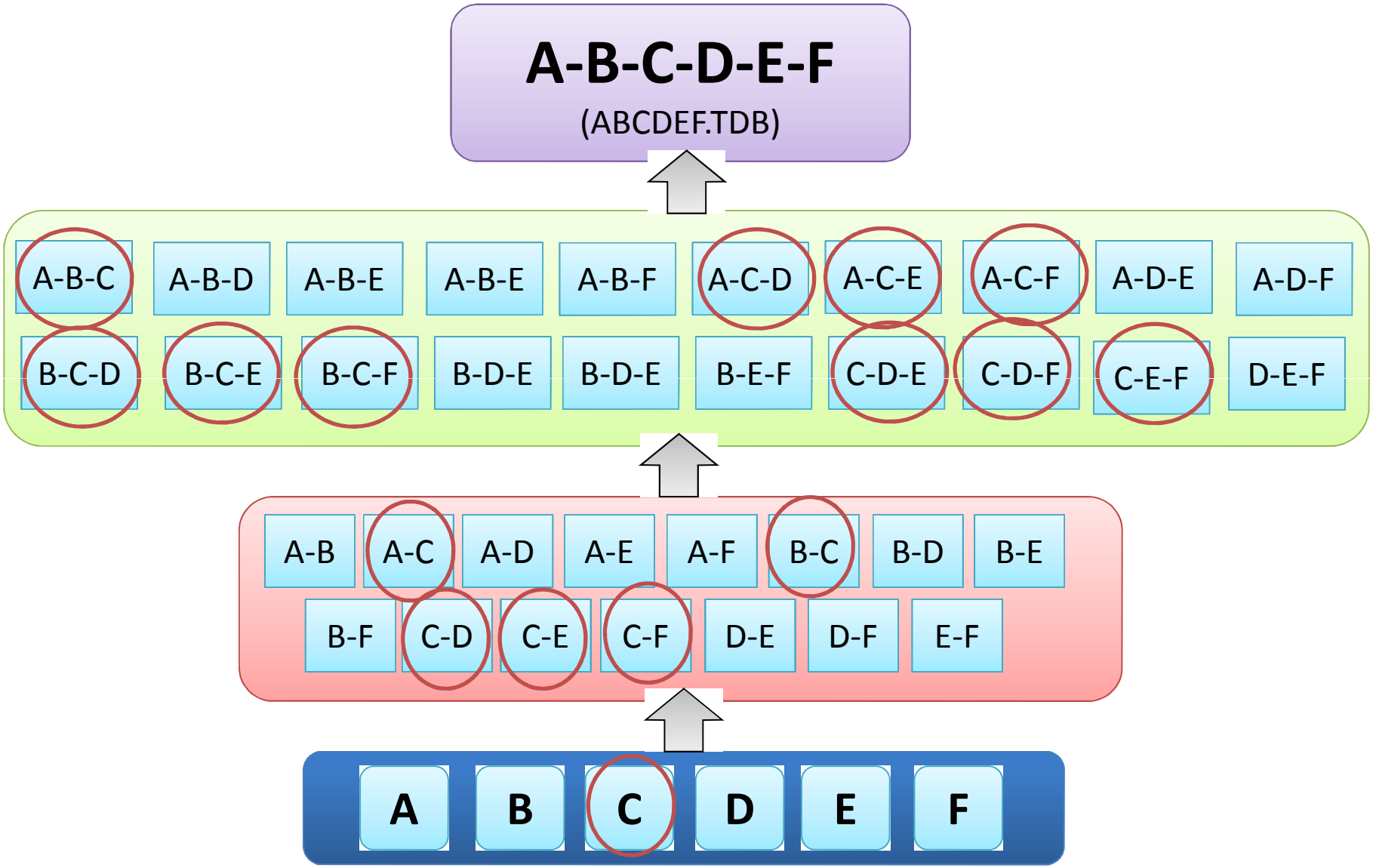
Database(s)

- Unary
- Binary
- Ternary
- ...
- Multi-component

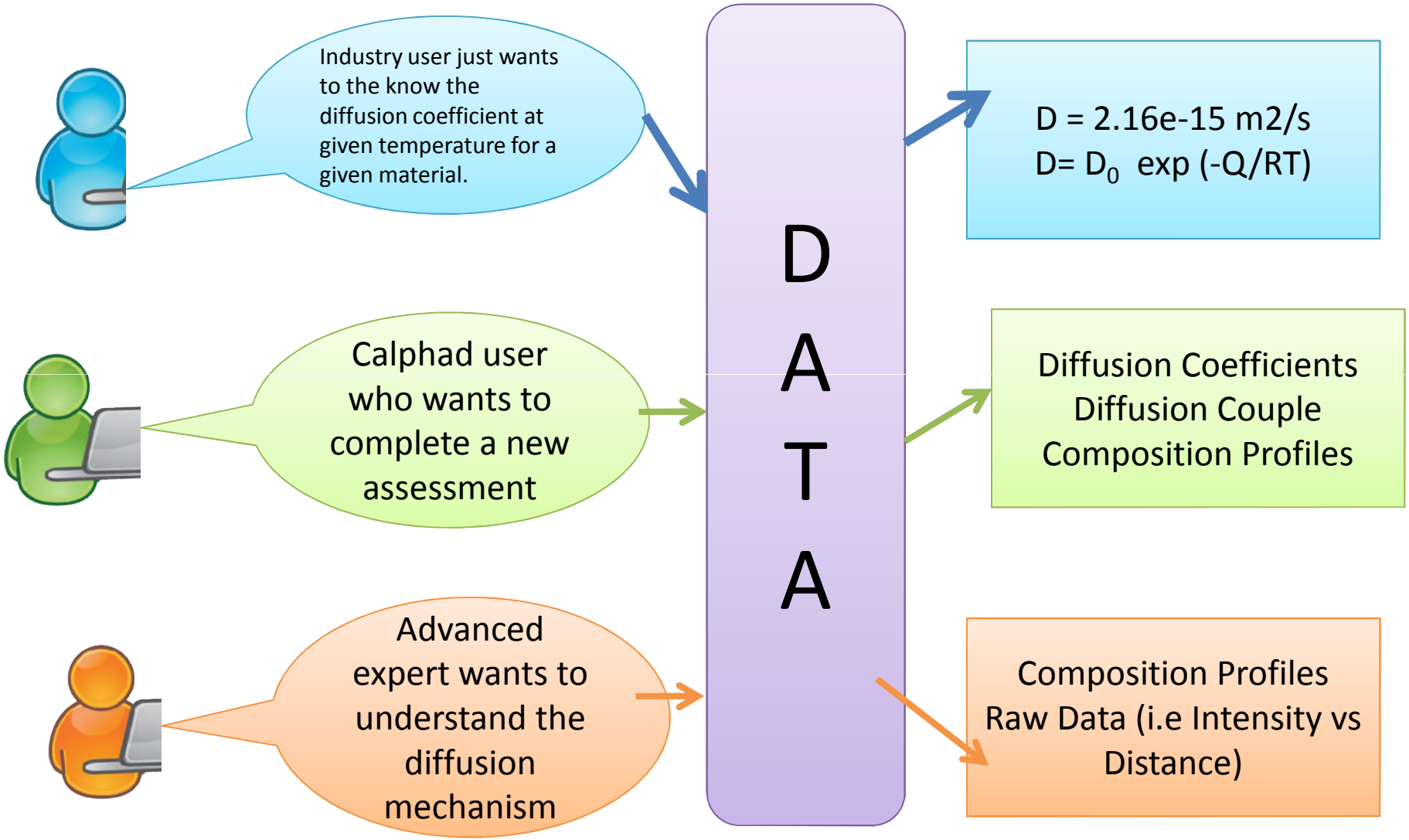
TDB file

A file repository is needed to improve reproducibility and efficiency.

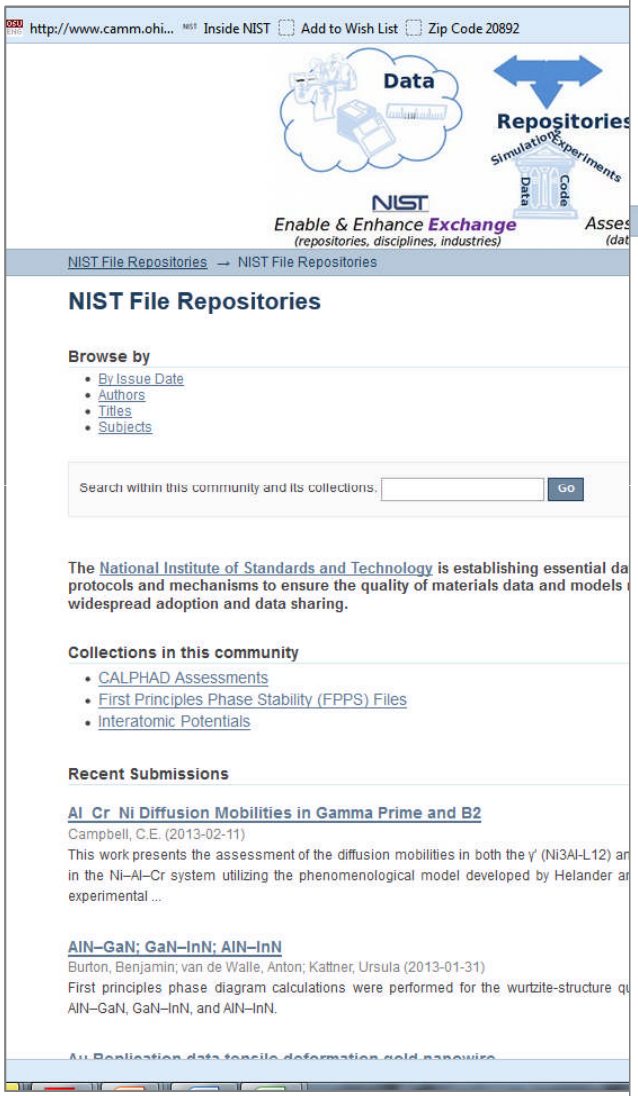
Data Dependencies



Different Data Users: Diffusion Data

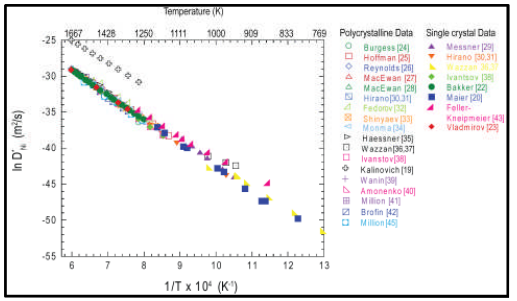


File Repository/DSpace

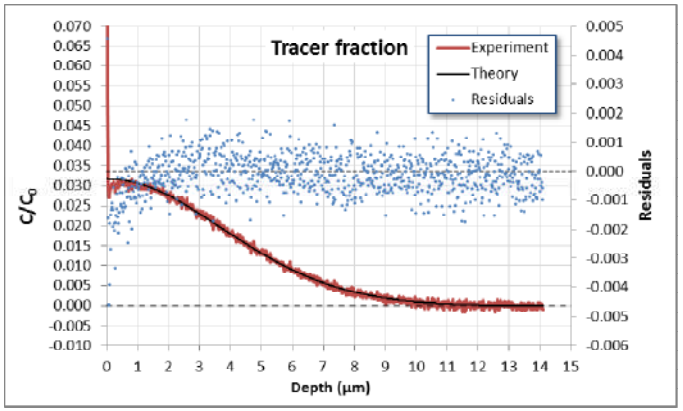


Data Repositories

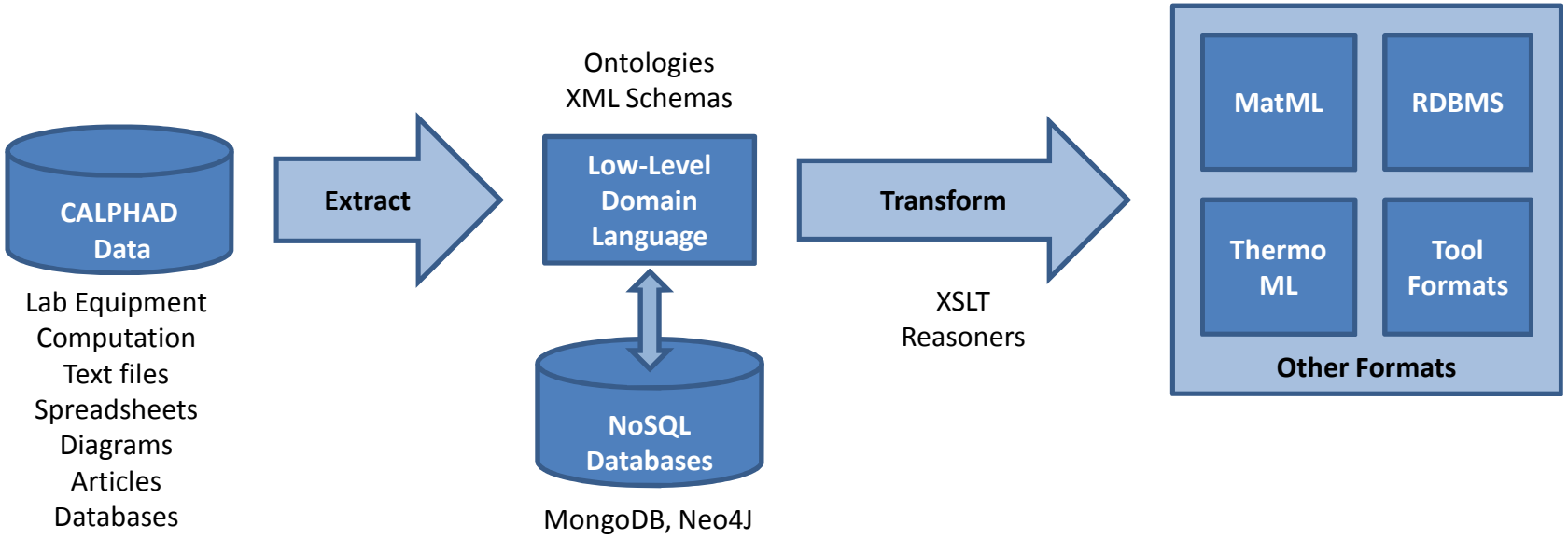
Different communities have different needs



- **Evaluated data**
  are needed by most data users as input computational and simulation tools
- **Original measured data**
  are needed to develop analytical functions
  - Ideally one would have a number of sources with original measured data
  - These data are in general not available in published papers or at best in graphical form
    - Digitization from graphs creates additional errors!
  - Repositories provide the space to accurately store these data for future use



Data Curation

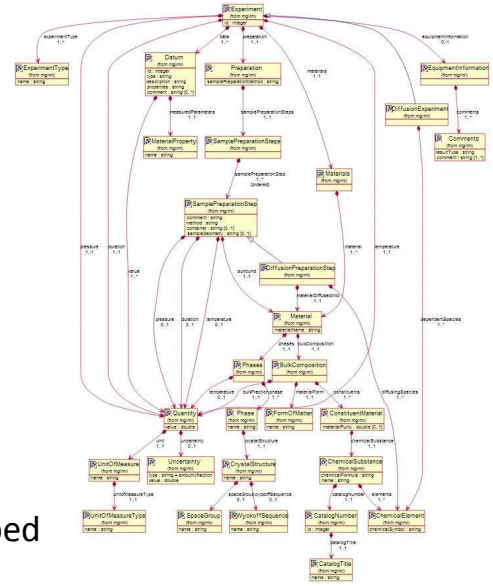


Lab Equipment
Computation
Text files
Spreadsheets
Diagrams
Articles
Databases

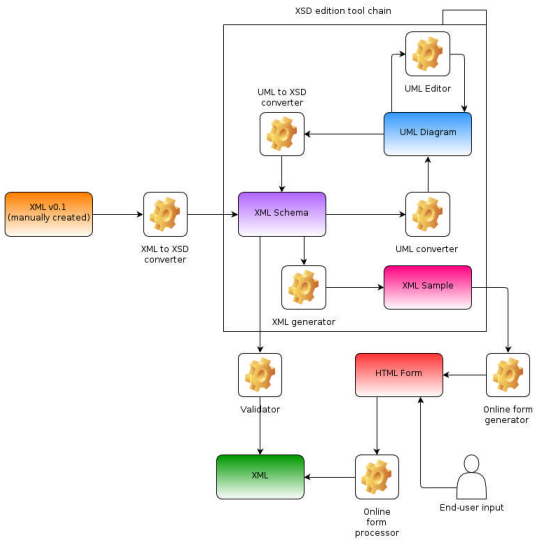
MongoDB, Neo4J

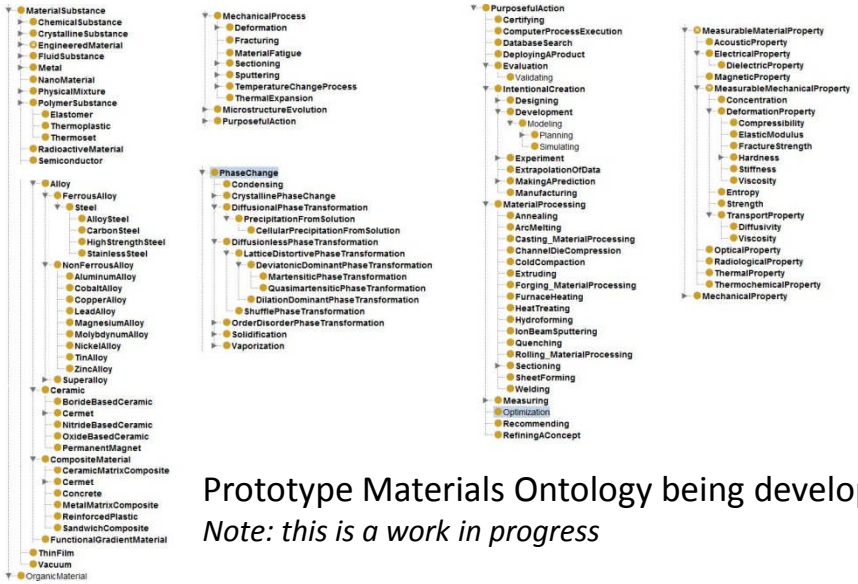
XSLT
Reasoners

Section of XML Schema



Tool Chain

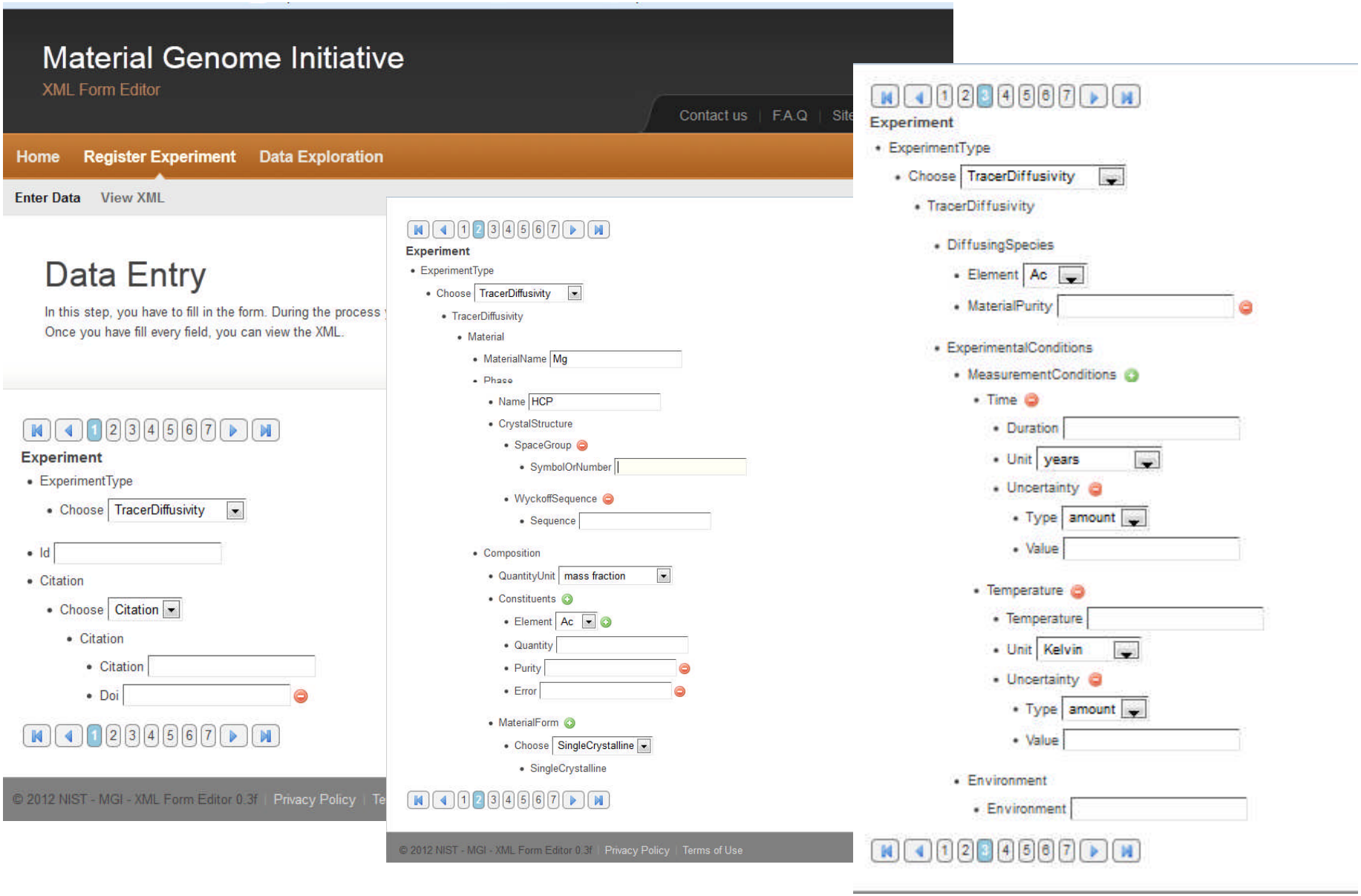




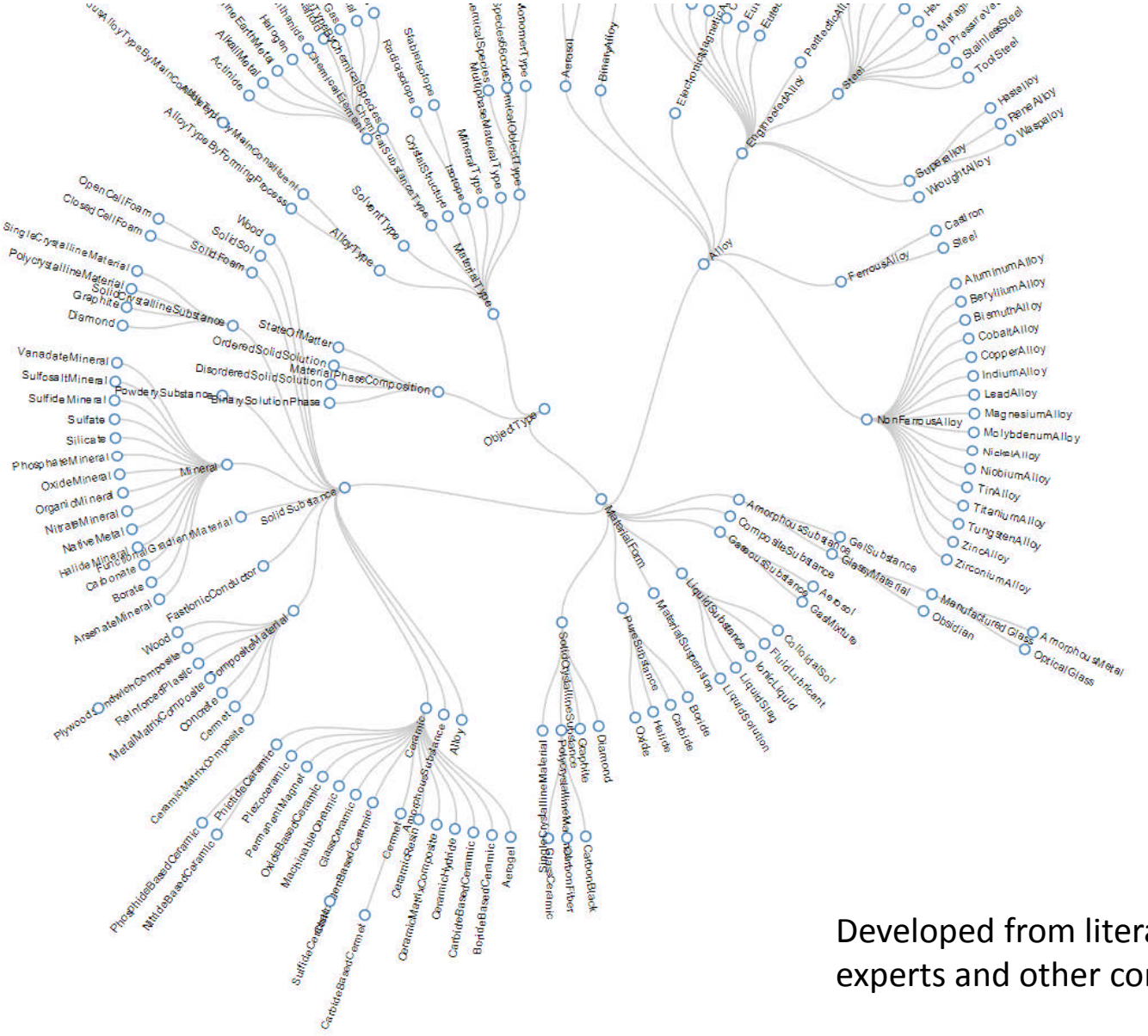
Prototype Materials Ontology being developed

Note: this is a work in progress

Data Collection: Tracer Diffusivity Test Schema



Prototype MGI Ontology



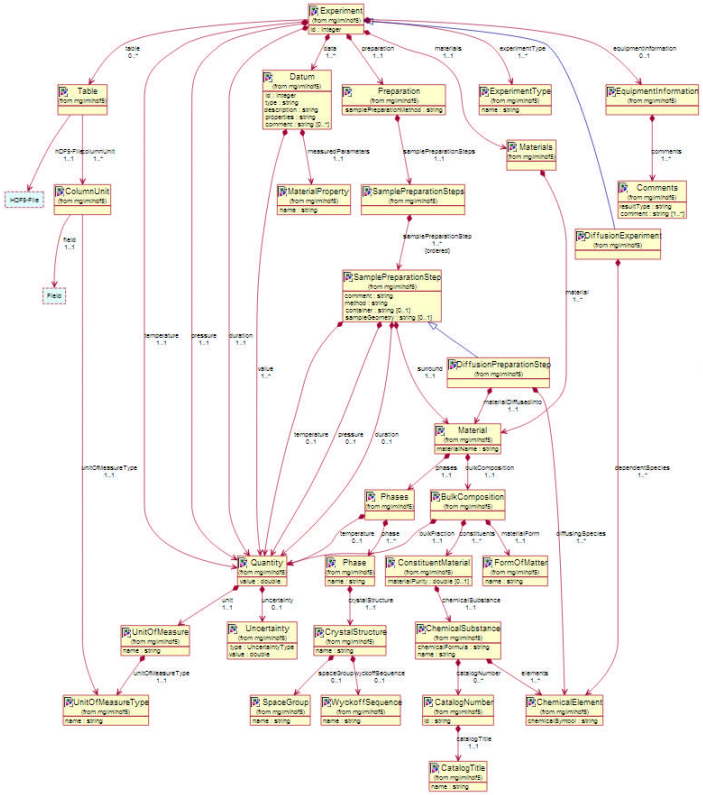
Broad concepts covered in materials data files (data have many types)

- Objects, Materials, and Events
- Physical Properties
- Documents
- Data Objects & Types
- People & Organizations
- Software
- Relations among these

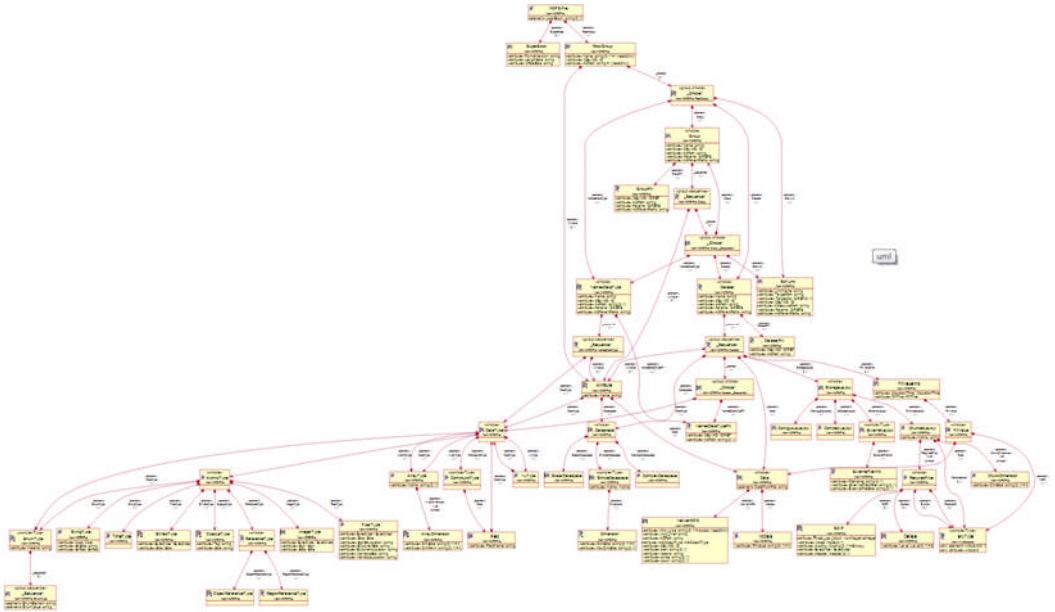
Developed from literature supplied by domain experts and other community resources

Encoding CALPHAD Data

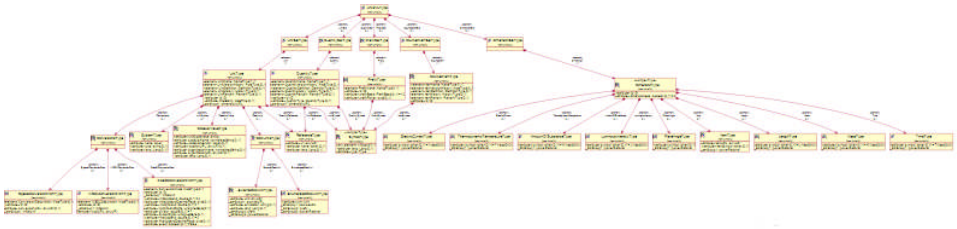
Core CALPHAD



Tabular Data (HDF5)



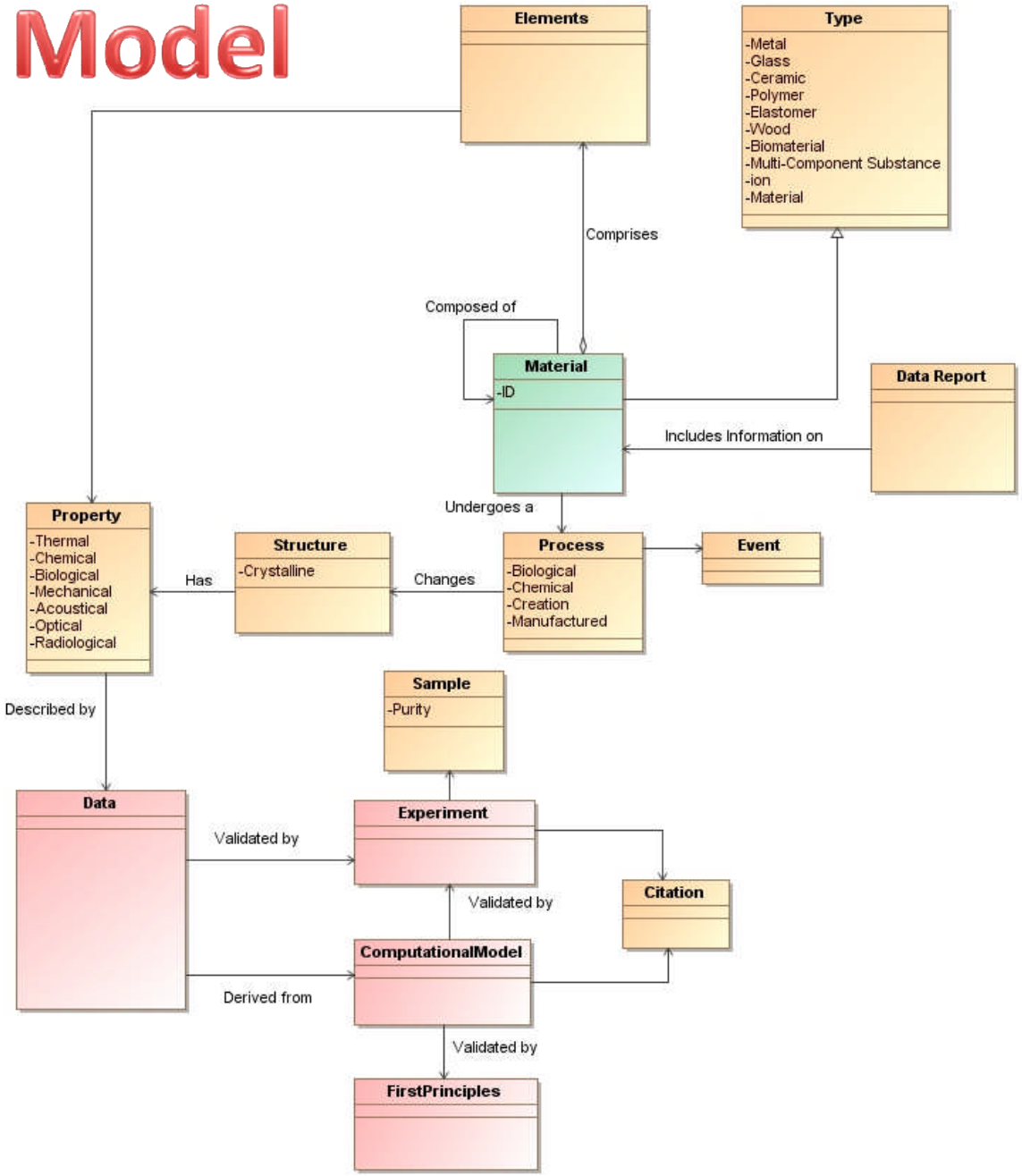
Measurement Units (Units ML)



UML Domain Model

Sources

- Prototype MGI Ontology
- ThermoML
- MatML
- MatSeek
- UnitsML
- ChemML



Questions for the Panel

- Prototype ontology
  - How do we evaluate it?
  - How do we best proceed with its development?
    - What tools are available to support community-based efforts?
    - What is the best way for us to encode our ontology?
- Graph databases
  - Any advice for using property graphs to represent ontologies?
- Software development and ontology
  - How do we best integrate ontology with UML?
  - What is the relationship between ontology and UML domain models?
- Community Experience
  - Are there examples of similar ontology-based systems for us to compare?