

Data Informatics and Tools for Phase-Based Property Data

C. E. Campbell and U. R. Kattner

Material Measurement Laboratory, Materials Science and Engineering Division

A. Dima, D. Foxvog, P. J. Dessauw, and P. Savoritto

Information Technology Laboratory, Software and Systems Division

2nd World Congress on Integrated Computational Materials Engineering

Salt Lake City, Utah

July 9, 2013

Our Collaborators

NIST Information Technology Laboratory

Information Systems Group

Y.-S. Li-Baboud, B. Long

NIST Material Measurement Laboratory

Technical Program Director for Materials Genomics

J.A. Warren

Thermodynamics and Kinetics Group

B.P. Burton, R.H. Taylor, C.A. Becker, F. Tavazza, Z. Trautt

Cell Systems Science Group

T.N. Bhat, J.T. Elliott

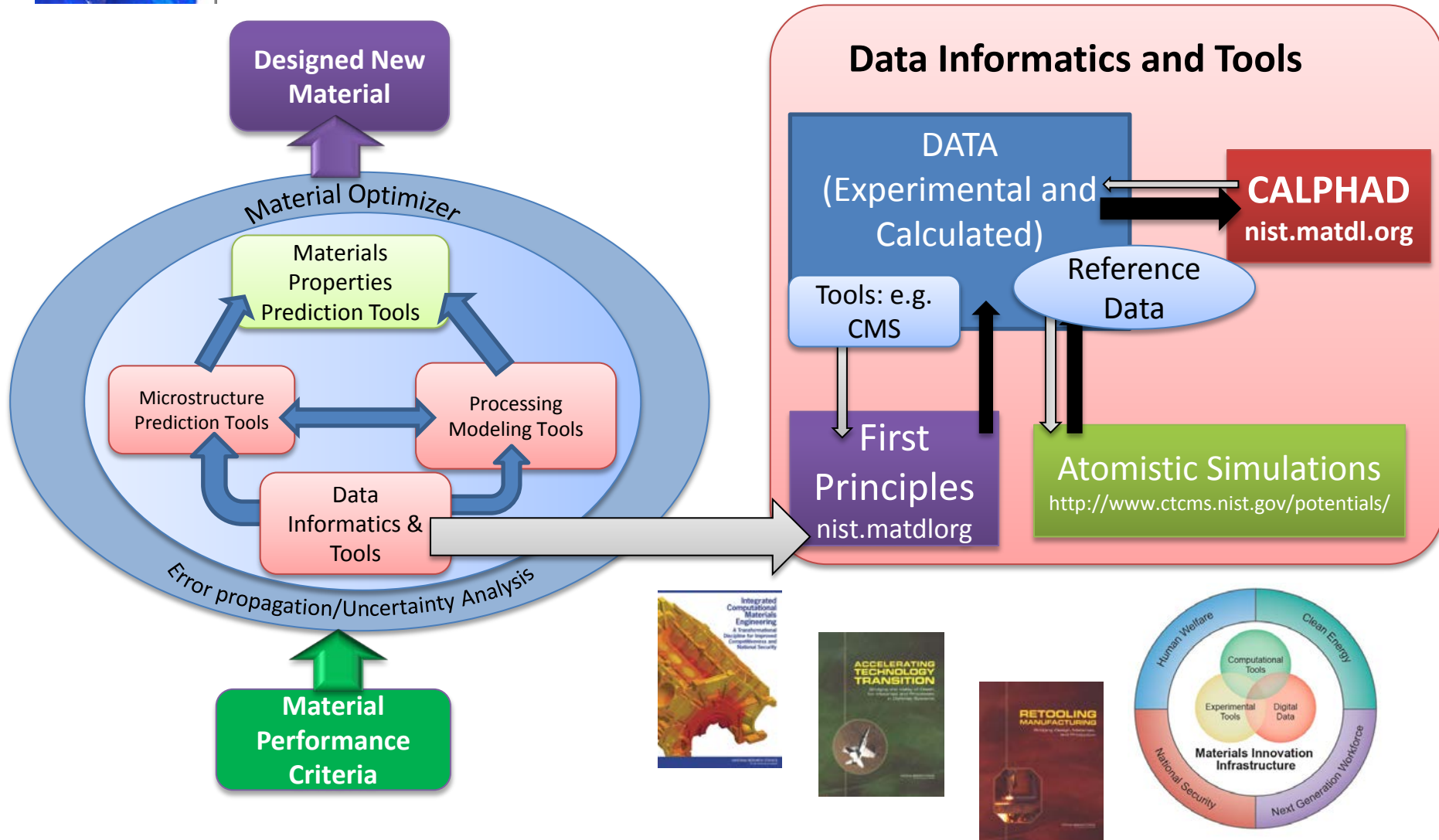
Kent State University, Center for Materials Informatics

Laura Bartolo



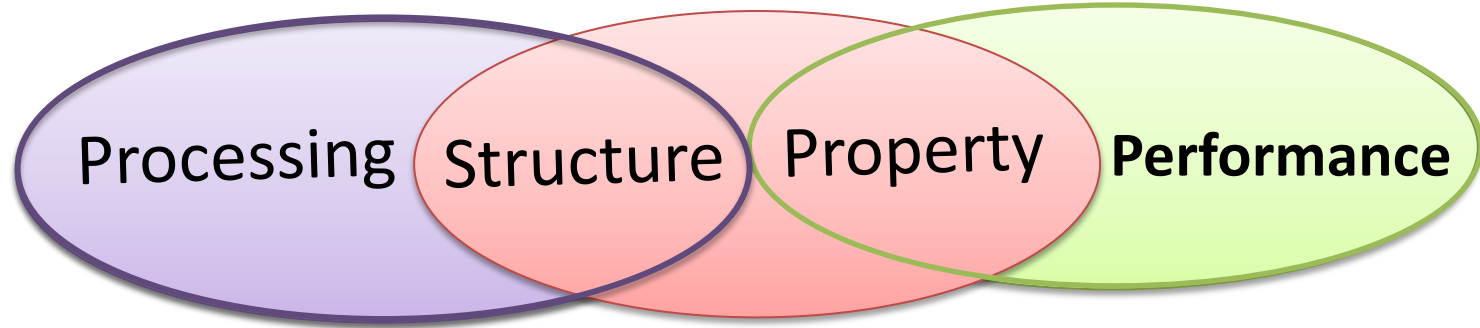
Materials Genome Initiative: The Need for Data and Informatics

Goal: decrease the cost and time-to-market by 50%

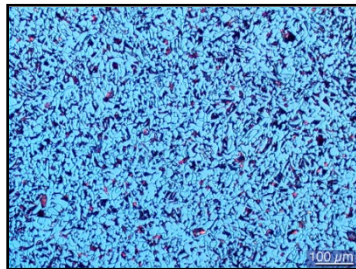
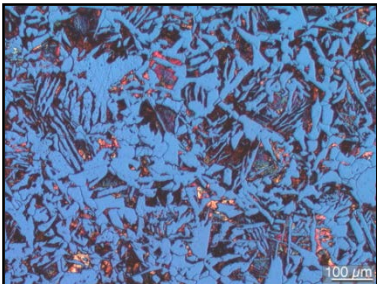


Materials Are Complicated Hierarchical Systems

- 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



Composition & phase dependent property data (CALPHAD)

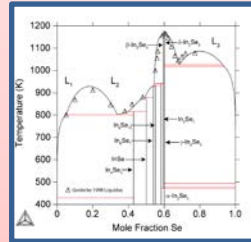
Examples of Types Data

- Phase-based property data: Thermodynamic quantities, diffusion mobilities, molar volume, elastic properties, thermal conductivity, etc.
- Emphasis on binary and ternary data to predict multicomponent properties
- Data can be experimental or computational.

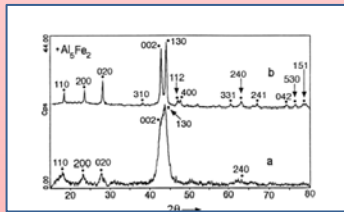
1-D (Points)

Melting Temperatures

Critical Temperatures
(Phase Changes)



Lattice Parameters



Heat of Formations

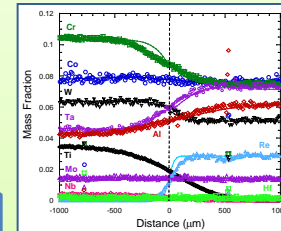
Phase fractions and
compositions

Tracer
Diffusivities

Activation energies

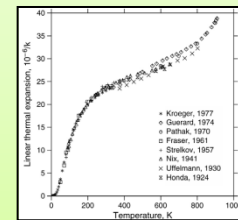
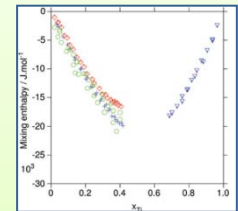
2-D (Lines)

Composition
Profiles



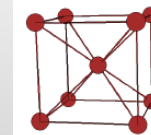
Heat Capacities

Enthalpies of mixing

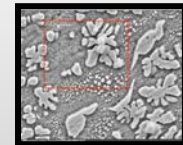


3-D

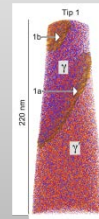
Crystal structures



Micrographs/Morphologies



3-D Atom probe Tomography



Example Information Needed to Describe General Data Entry

Data

- Elements present
- Type of value (e.g. enthalpy, heat of formation, phase boundary, diffusivity, lattice parameter, bulk modulus)
 - Experimental or computational method
 - Type of measurement (direct or indirect)
- Number of phases present
- Datum value and error
 - Type (single value or series)
 - Units
 - Actual value(s) and error(s)
- For each phase present
 - Phase name
 - Composition and fraction and errors
 - Crystal structure (this input will follow the format prescribed by the CCN) or amorphous
 - Lattice parameter
- Temperature and error
- Pressure and error
- Type of Material
 - Bulk composition
 - Material purity
 - Sample preparation
 - Microstructure information
 - Single crystal
 - Polycrystalline (grain size, dislocation density)
 - Non-crystalline

**Need extensible formats
that can evolve with
changing data needs!**

Metadata

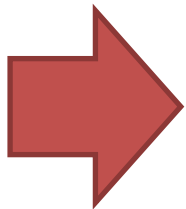
- Data manipulation details (if any, e.g. reference state corrections, analysis method to determine interdiffusion coefficient)
- Reporting format (raw data, digitized data, other)
- Reference (DOI or text ; one must be present)
- Additional information

Phase-Based Property Database

➤ Material Property Database Exist

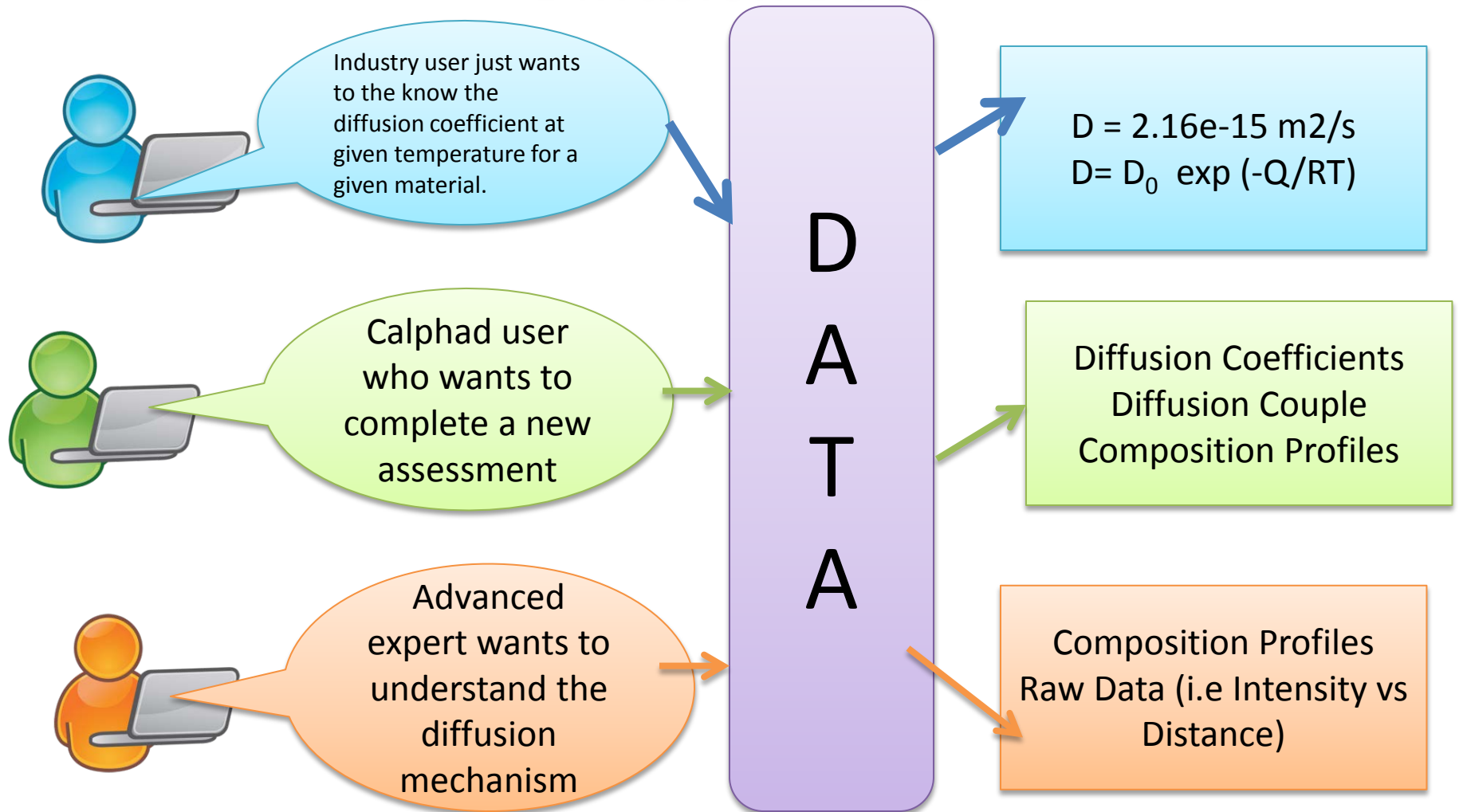


Generally, focused on engineering/design specs or first-principle calculations results.



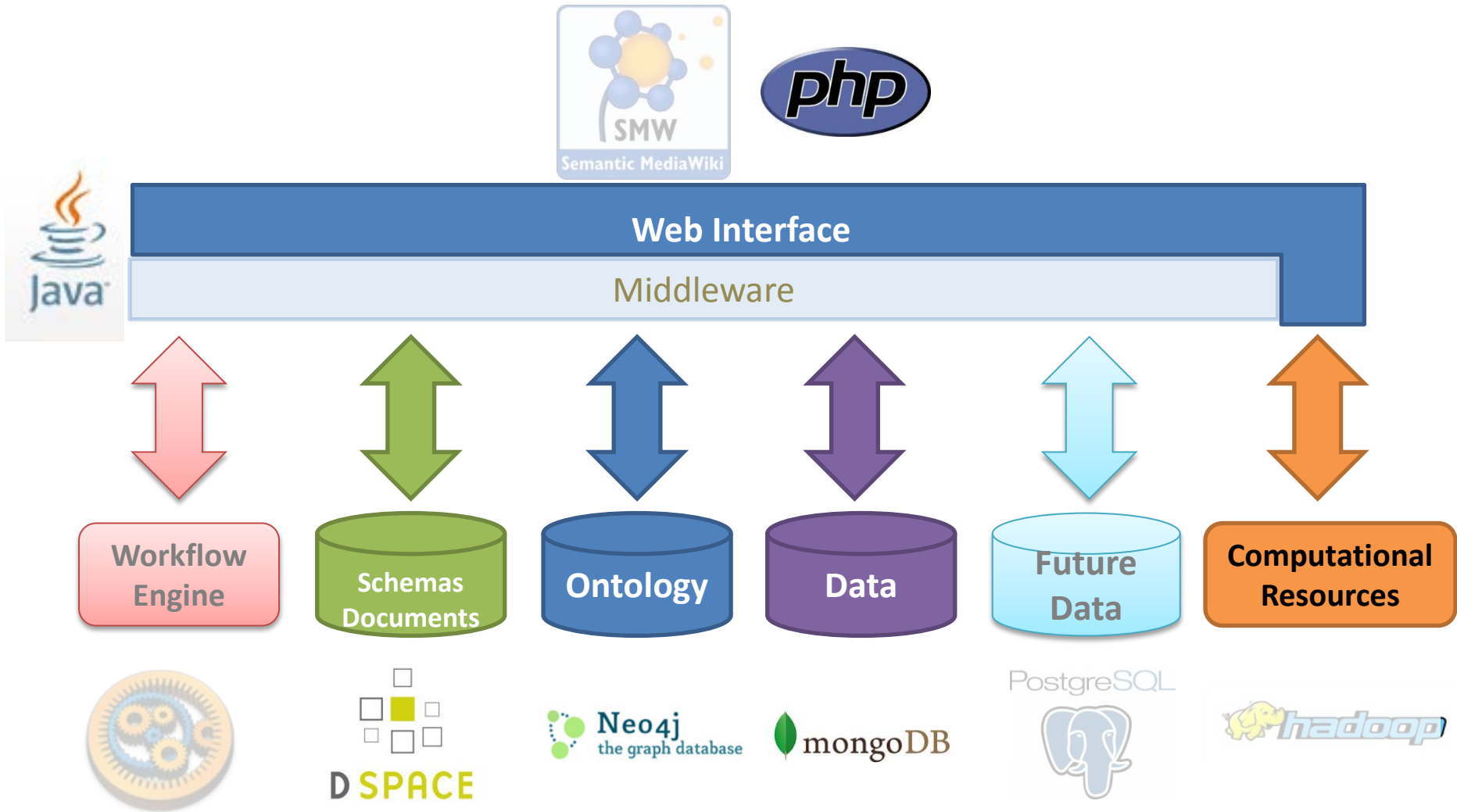
- **Focus on phase-based properties that are needed to describe the composition, temperature, and pressure functions of a phase.**
- **Unary, binary and ternary data are primary focus.**
- **Multicomponent data are needed for validation**

Example of Different Types of Data Users: Diffusion Data

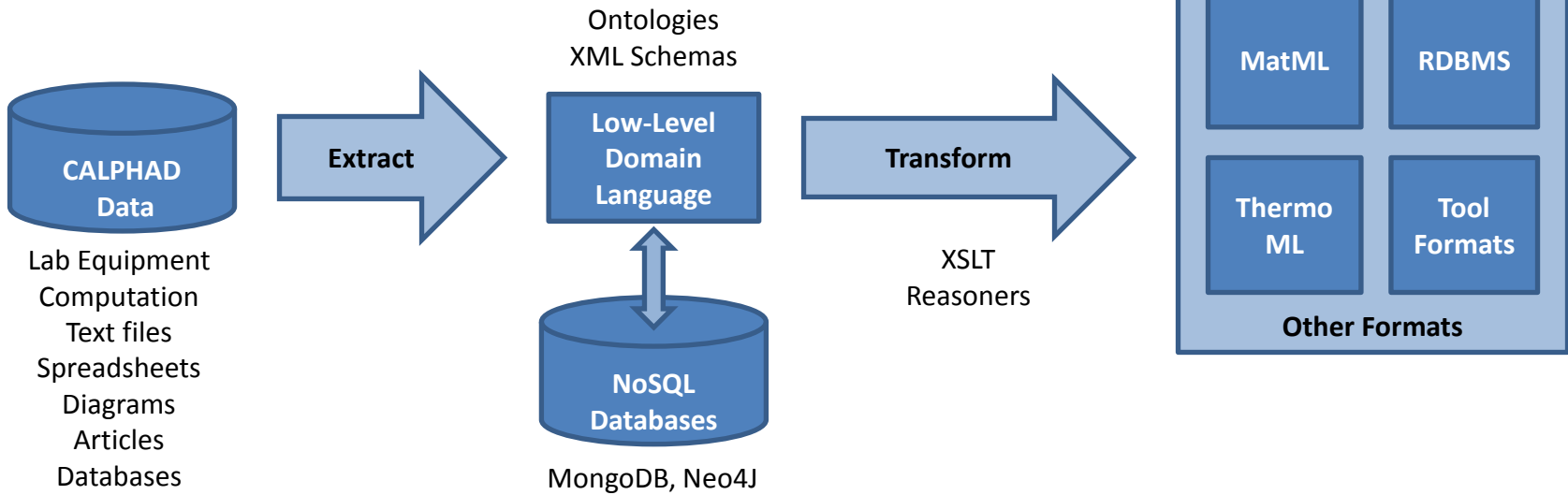


Data are diverse **Data are semi-structured** **Need complete data sets**

Architectural Strategy



Informatix Approach



Section of XML Schema

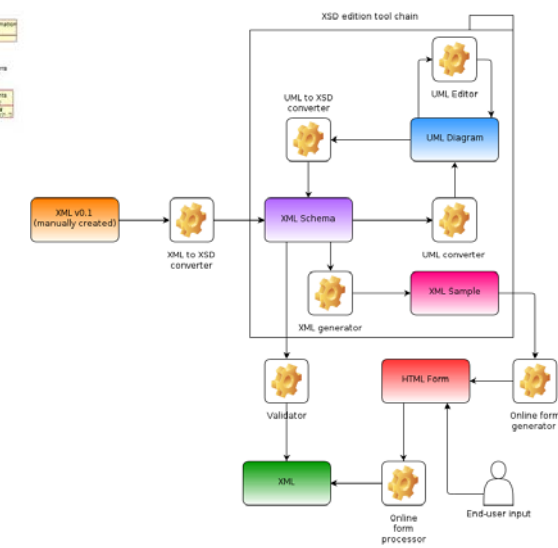
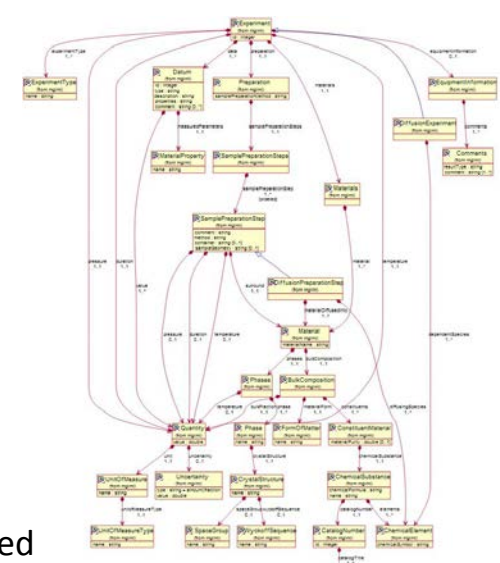
Tool Chain

- Material Substance
 - Chemical Substance
 - Crystalline Substance
 - Engineered Material
 - Fluid Substance
 - Metal
 - NanoMaterial
 - Physical Mixture
 - Polymer Substance
 - Elastomer
 - Thermoplastic
 - Thermoset
 - Radioactive Material
 - Semiconductor
- Alloy
 - Ferrous Alloy
 - Steel
 - Alloy Steel
 - Carbon Steel
 - High Strength Steel
 - Stainless Steel
 - NonFerrous Alloy
 - Aluminum Alloy
 - Cobalt Alloy
 - Copper Alloy
 - Lead Alloy
 - Magnesium Alloy
 - Molybdenum Alloy
 - Nickel Alloy
 - Titanium Alloy
 - Zinc Alloy
 - Superalloy
 - Ceramic
 - Boride Based Ceramic
 - Cermet
 - Nitride Based Ceramic
 - Oxide Based Ceramic
 - Permanent Magnet
 - Composite Material
 - Ceramic Matrix Composite
 - Cermet
 - Concrete
 - Metal Matrix Composite
 - Reinforced Plastic
 - Sandwich Composite
 - Functional Gradient Material
 - Thin Film
 - Vacuum
 - Organic Material

- Mechanical Process
 - Deformation
 - Fracturing
 - Material Fatigue
 - Sectioning
 - Sputtering
 - Temperature Change Process
 - Thermal Expansion
 - Microstructure Evolution
 - Purposeful Action
- Phase Change
 - Condensing
 - Crystalline Phase Change
 - Diffusional Phase Transformation
 - Precipitation From Solution
 - Cellular Precipitation From Solution
 - Diffusional Phase Transformation
 - Diffusionless Phase Transformation
 - Lattice Distortive Phase Transformation
 - Deviatonic Dominant Phase Transformation
 - Martensitic Phase Transformation
 - Quasi Martensitic Phase Transformation
 - Dilution Dominant Phase Transformation
 - Order Disorder Phase Transformation
 - Vaporization

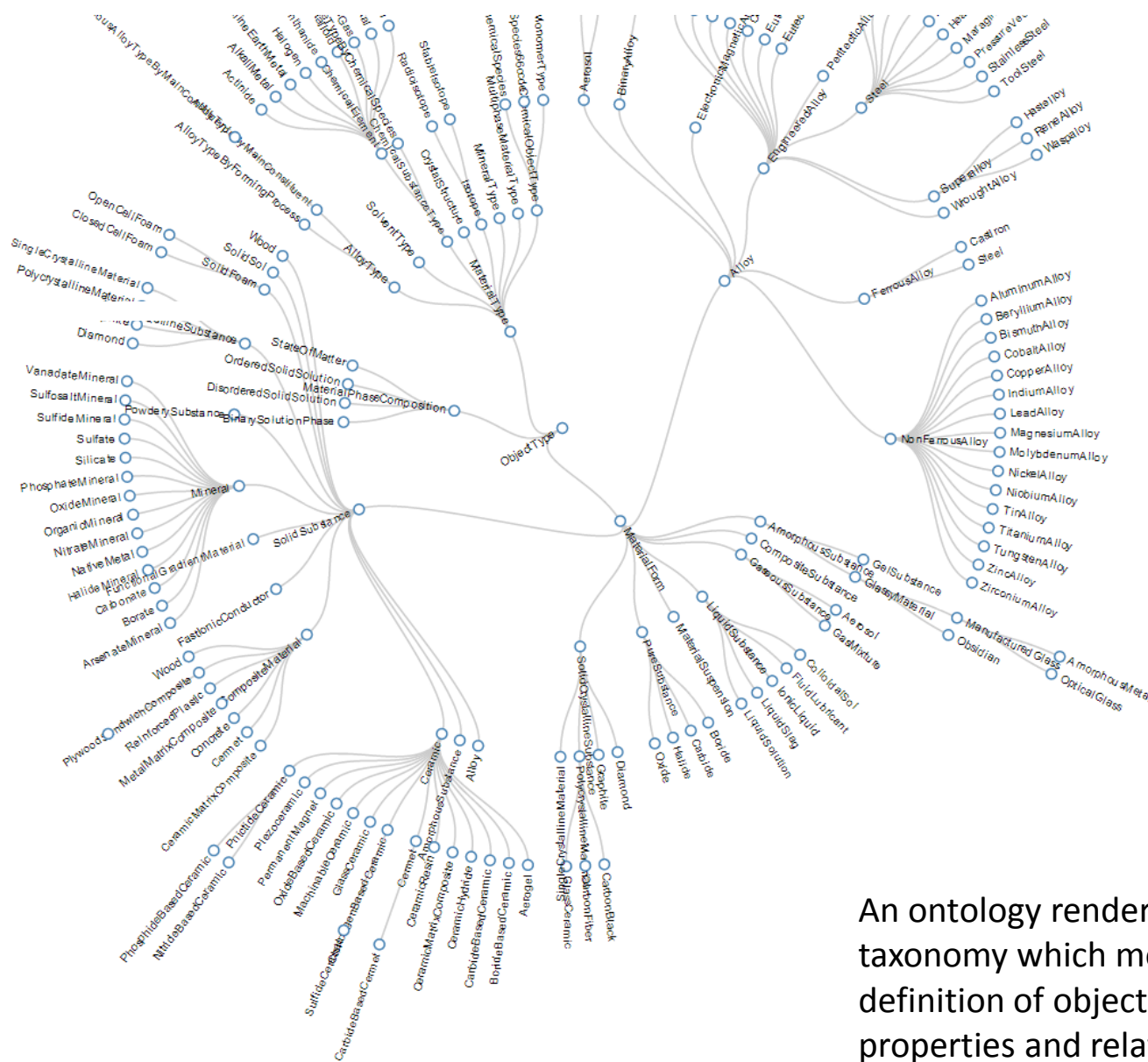
- Purposeful Action
 - Certifying
 - Computer Process Execution
 - Database Search
 - Deploying Product
 - Evaluation
 - Validating
 - Intentional Creation
 - Designing
 - Modeling
 - Planning
 - Simulating
 - Development
 - Experiment
 - Extrapolation Of Data
 - Making Prediction
 - Manufacturing
 - Material Processing
 - Annealing
 - Arcl Melting
 - Casting_Material Processing
 - Channel Die Compression
 - Cold Compaction
 - Extruding
 - Forging_Material Processing
 - Furnace Heating
 - Heat Treating
 - Hydroforming
 - Ion Beam Sputtering
 - Quenching
 - Rolling_Material Processing
 - Sectioning
 - Sheet Forming
 - Welding
 - Measuring
 - Optimization
 - Recommending
 - Retraining Concept

- Measurable Material Property
 - Acoustic Property
 - Electrical Property
 - Dielectric Property
 - Magnetic Property
 - Measurable Mechanical Property
 - Compressibility
 - Elastic Modulus
 - Fracture Strength
 - Hardness
 - Stiffness
 - Viscosity
 - Entropy
 - Strength
 - Transport Property
 - Diffusivity
 - Viscosity
 - Optical Property
 - Radiological Property
 - Thermal Property
 - Thermochemical Property
 - Mechanical Property



Materials Ontology currently being developed
Note this is a work in progress

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

An ontology renders shared vocabulary and taxonomy which models a domain with the definition of objects and/or concepts and their properties and relations.

Benefits from an Ontological Approach

- Semantic Unification
 - The unification of lexically different representations that have the same semantics
 - Example: fcc phase in steels can be referred to as fcc, austenite or γ .
- Ontology-based Data Integration
 - Using ontologies to unify data that share some common semantics but originate from unrelated sources
 - Example: Are property data from two experiments consistent enough to be combined?
- Ontologies are not static and can grow with needs

Ontologies & Graph Databases

The image shows two overlapping screenshots of the Neo4j web interface. The top screenshot displays a search results table for the query 'rels:193'. The table lists relationships with their IDs, start nodes, types, end nodes, and sources. The bottom screenshot shows a graph visualization of the same data, with nodes and relationships represented as a network.

Relationship	Start node	Type	End node	source
Relationship 7218	Node 1976	domain	Node 193	"http://www.w3.org/2000/01/rdf-schema#domain"
Relationship 4207	Node 1196	domain	Node 193	"http://www.w3.org/2000/01/rdf-schema#domain"
Relationship 4189	Node 1190	domain	Node 193	"http://www.w3.org/2000/01/rdf-schema#domain"
Relationship 3992	Node 1125	domain	Node 193	"http://www.w3.org/2000/01/rdf-schema#domain"
Relationship 2722	Node 749	domain	Node 193	"http://www.w3.org/2000/01/rdf-schema#domain"
Relationship 2659	Node 729	domain	Node 193	"http://www.w3.org/2000/01/rdf-schema#domain"
Relationship 2391	Node 640	domain	Node 193	"http://www.w3.org/2000/01/rdf-schema#domain"

- Why use a graph database?
 - True networked database with queries, ACID, and REST interface
 - All apps can share the same representation
 - Overcomes some of the limitations of RDF
 - Flexible visualization...

UML Domain Model

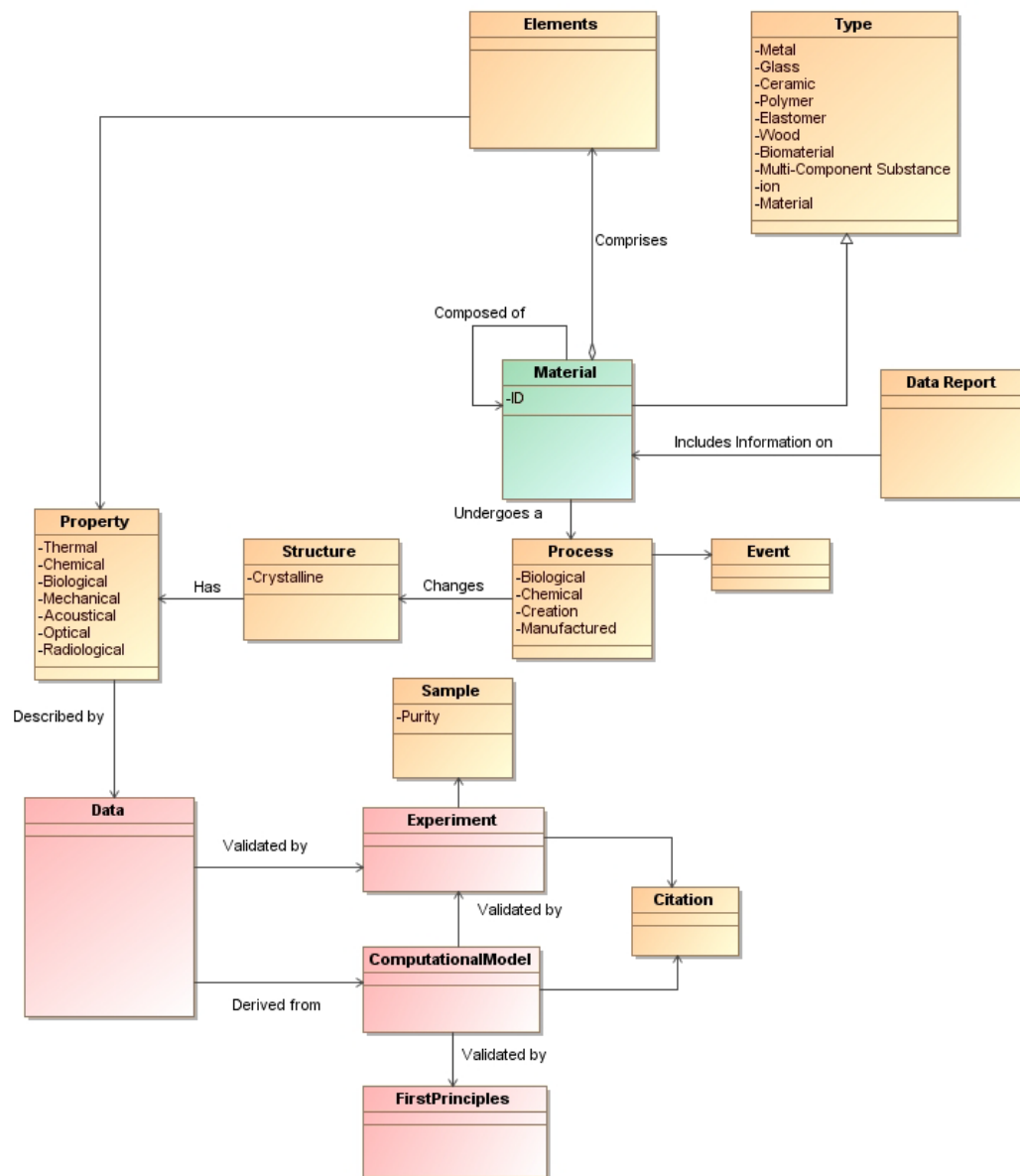
Sources

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

Tools

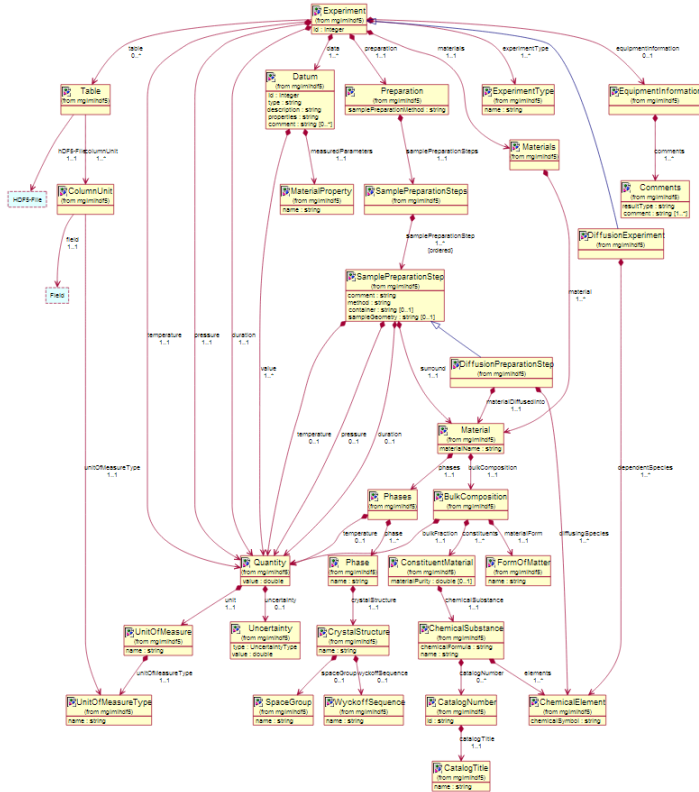
- UML (Unified Modeling Language)
- Semantic Web (RDF, OWL)

Note: This is a generalized model depicting overall structure

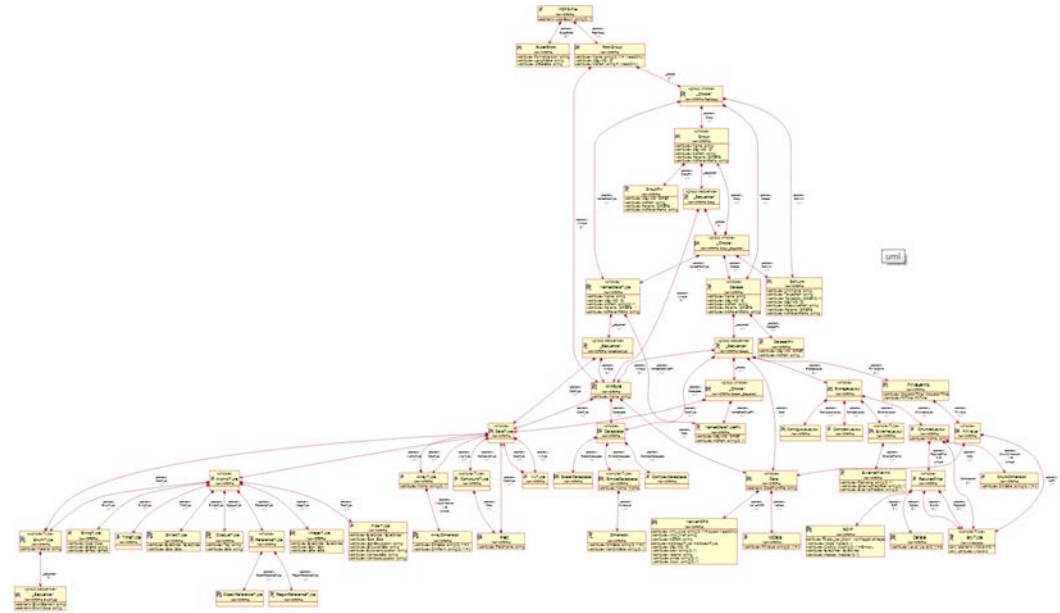


Encoding CALPHAD Data

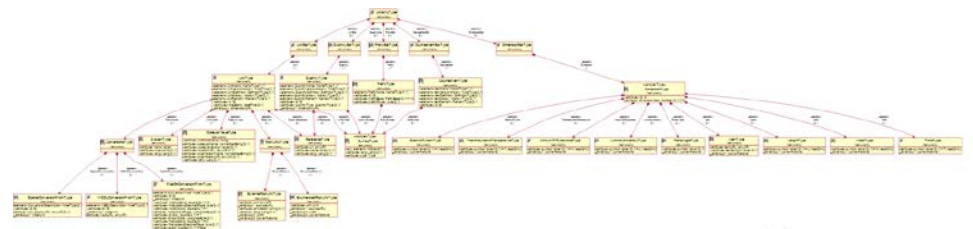
Core CALPHAD



Tabular Data (HDF5)



Measurement Units (Units ML)



Data Collection: Tracer Diffusivity Test Schema

Material Genome Initiative

XML Form Editor

Contact us | F.A.Q | Site

Home Register Experiment Data Exploration

Enter Data View XML

Data Entry

In this step, you have to fill in the form. During the process
Once you have fill every field, you can view the XML.



Experiment

- ExperimentType
 - Choose
- Id
- Citation
 - Choose
 - Citation
 - Doi



Experiment

- ExperimentType
 - Choose
- TracerDiffusivity
 - Material
 - MaterialName
 - Phase
 - Name
 - CrystalStructure
 - SpaceGroup
 - SymbolOrNumber
 - WyckoffSequence
 - Sequence
- Composition
 - QuantityUnit
 - Constituents
 - Element
 - Quantity
 - Purity
 - Error
 - MaterialForm
 - Choose
 - SingleCrystalline



Experiment

- ExperimentType
 - Choose
- TracerDiffusivity
 - DiffusingSpecies
 - Element
 - MaterialPurity
- ExperimentalConditions
 - MeasurementConditions
 - Time
 - Duration
 - Unit
 - Uncertainty
 - Type
 - Value
 - Temperature
 - Temperature
 - Unit
 - Uncertainty
 - Type
 - Value
 - Environment
 - Environment



Snap-Shot of XML Format

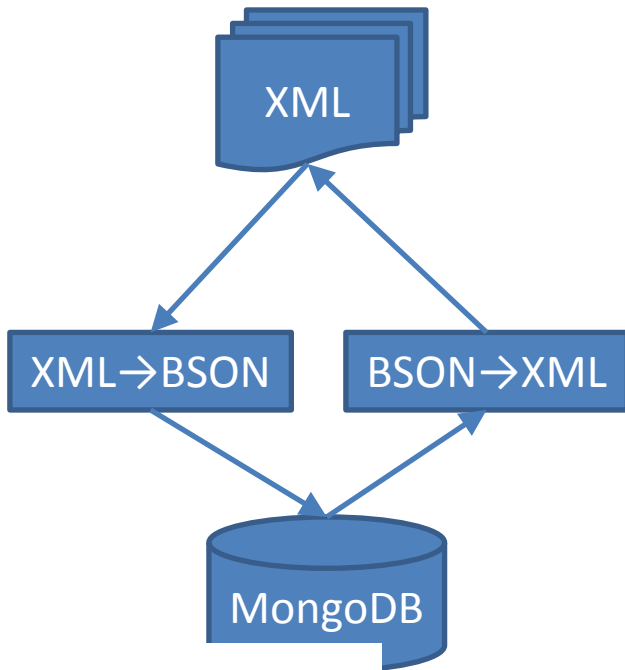
XML View

This is a preview of the XML which will be generated. Each modification you make in the form will be written in this preview.

```
<experiment xmlns:hdf5="http://hdfgroup.org/HDF5/XML/schema/HDF5-File">
  <experimentType>
    <tracerDiffusivity>
      <material>
        <materialName>
          Mg
        </materialName>
        <phase>
          <name>
            HCP
          </name>
          <crystalStructure>
            <spaceGroup>
              <symbolOrNumber />
            </spaceGroup>
            <wyckoffSequence>
              <sequence />
            </wyckoffSequence>
          </crystalStructure>
        </phase>
        <Composition>
          <quantityUnit>
            mass fraction
          </quantityUnit>
          <constituents>
```

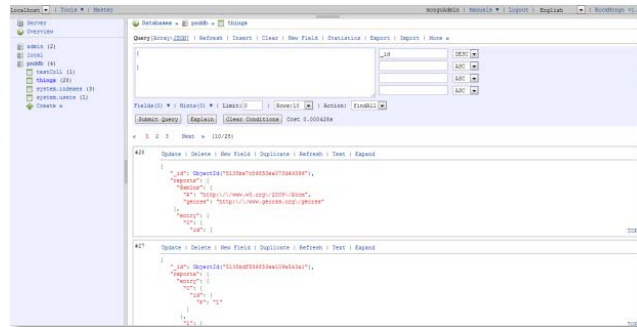
- Why XML?
 - Internationalization
 - Flexibility
 - Transformability
 - Interoperability
 - Longevity
 - Web-Enabled
 - Available Resources

XML Document Storage



- MongoDB
 - Schema-less, cloud-friendly
 - High Performance, scalable
 - Used by CERN enable information discovery on Compact Muon Solenoid data
 - Reason: “dynamic queries, full indexes, including inner objects and embedded arrays, as well as auto-sharing”

```
{
  "_id"      : ObjectId("4be97eabcd1b30e86000003"),
  "title"   : "Ordered List",
  "creator_id" : ObjectId("4be97eabcd1b30e86000001"),
  "memberships" : [
    ObjectId("4be97eabcd1b30e86000001"),
    ObjectId("4be97eabcd1b30e86000002")
  ]
}
```



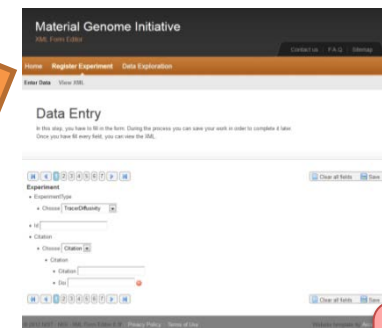
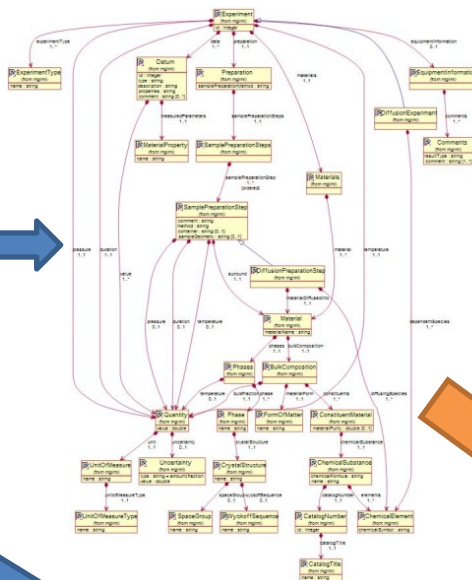
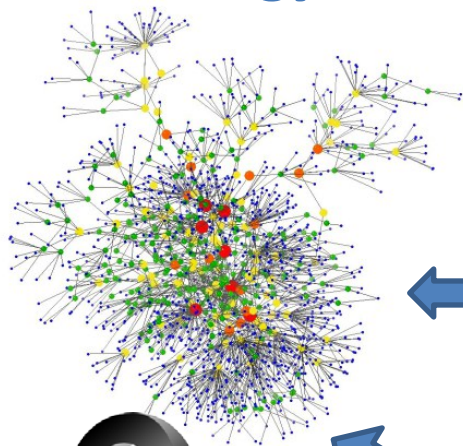
Future Data Informatics

User interface

Data Capture

Ontology

XML Schema

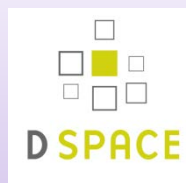


JSON

UNIFIED
MODELING
LANGUAGE



Various Database Platforms



Data Tools:
Statistics; Machine
Learning

Needs for the Future Data Infrastructure



Building a data infrastructure that enables complex searches and allows for data mining and machine learning

- **Need methods to facilitate data capture**
 - User defined templates
 - Automated capture from instruments
 - Electronic laboratory notebooks?
- **Community participation**
 - Incentives from journals and funding sources