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

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

TMS ICME MANUFACTURING DESIGN MATERIALS

Our Collaborators

NIST Information Technology Laboratory

Information Systems Group
Y.-S. Li-Baboud, B. Long

NIST Material Measurement Laboratory

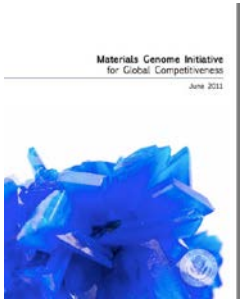
Technical Program Director for Materials Genomics
J.A. Warren

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%

Designed New Material

Material Optimizer

Materials Properties Prediction Tools

Microstructure Prediction Tools

Processing Modeling Tools

Data Informatics & Tools

Error propagation/Uncertainty Analysis

Material Performance Criteria

Data Informatics and Tools

DATA (Experimental and Calculated)

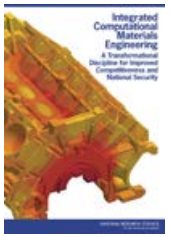
CALPHAD
nist.matdl.org

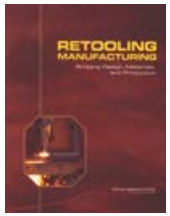
Reference Data

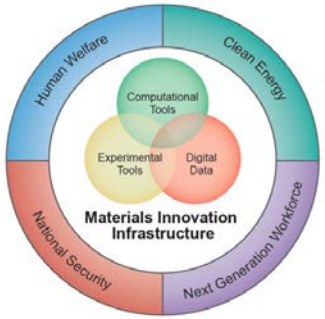
Tools: e.g. CMS

First Principles
nist.matdlorg

Atomistic Simulations
http://www.ctcms.nist.gov/potentials/

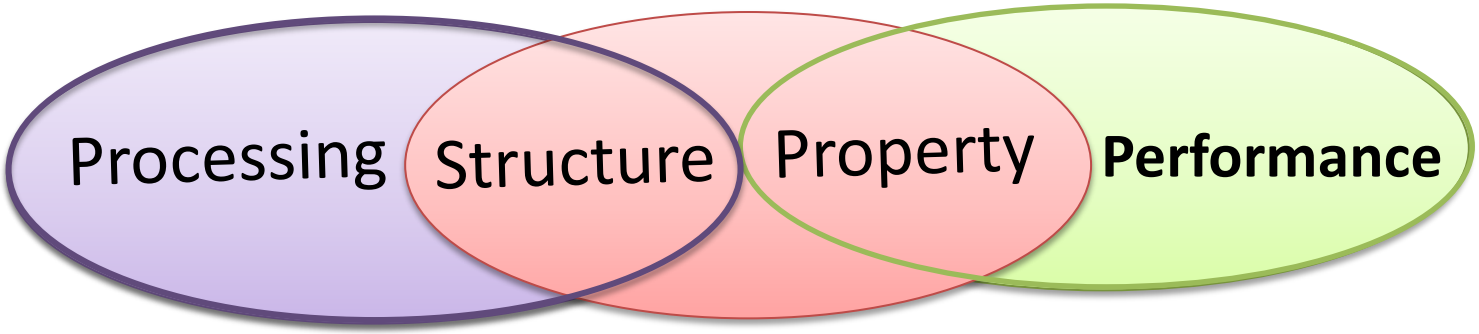




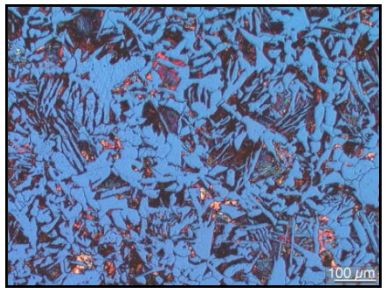


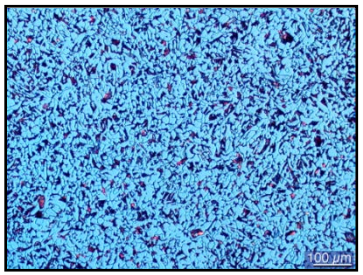
Materials Are Complicated Hierarchical Systems

- 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

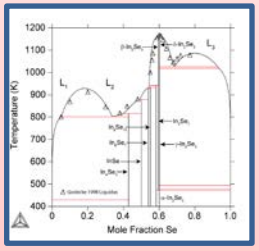
Composition & phase dependent property data (CALPHAD)

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.

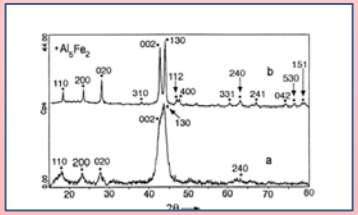
1-D (Points)

Melting Temperatures



Critical Temperatures (Phase Changes)

Lattice Parameters



Heat of Formations

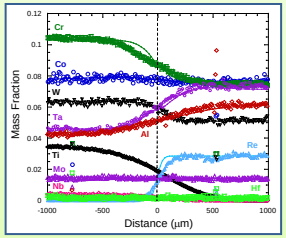
Phase fractions and compositions

Tracer Diffusivities

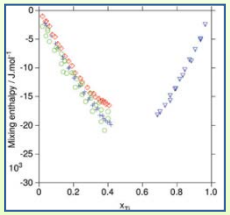
Activation energies

2-D (Lines)

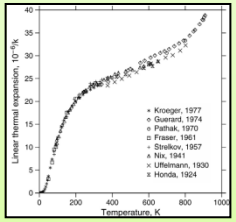
Composition Profiles



Heat Capacities

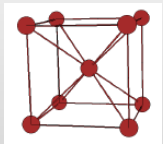


Enthalpies of mixing



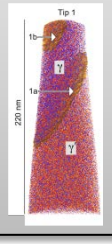
3-D

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

Materials Universe

GRANTA MATERIAL INTELLIGENCE

MATERIALS PROJECT

MatWeb

ASM International The Materials Information Society

Everything Material.

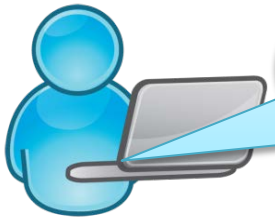
MatNavi

AFLOWLIB.ORG

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



Industry user just wants to the know the diffusion coefficient at given temperature for a given material.

Calphad user who wants to complete a new assessment

Advanced expert wants to understand the diffusion mechanism

D A T A

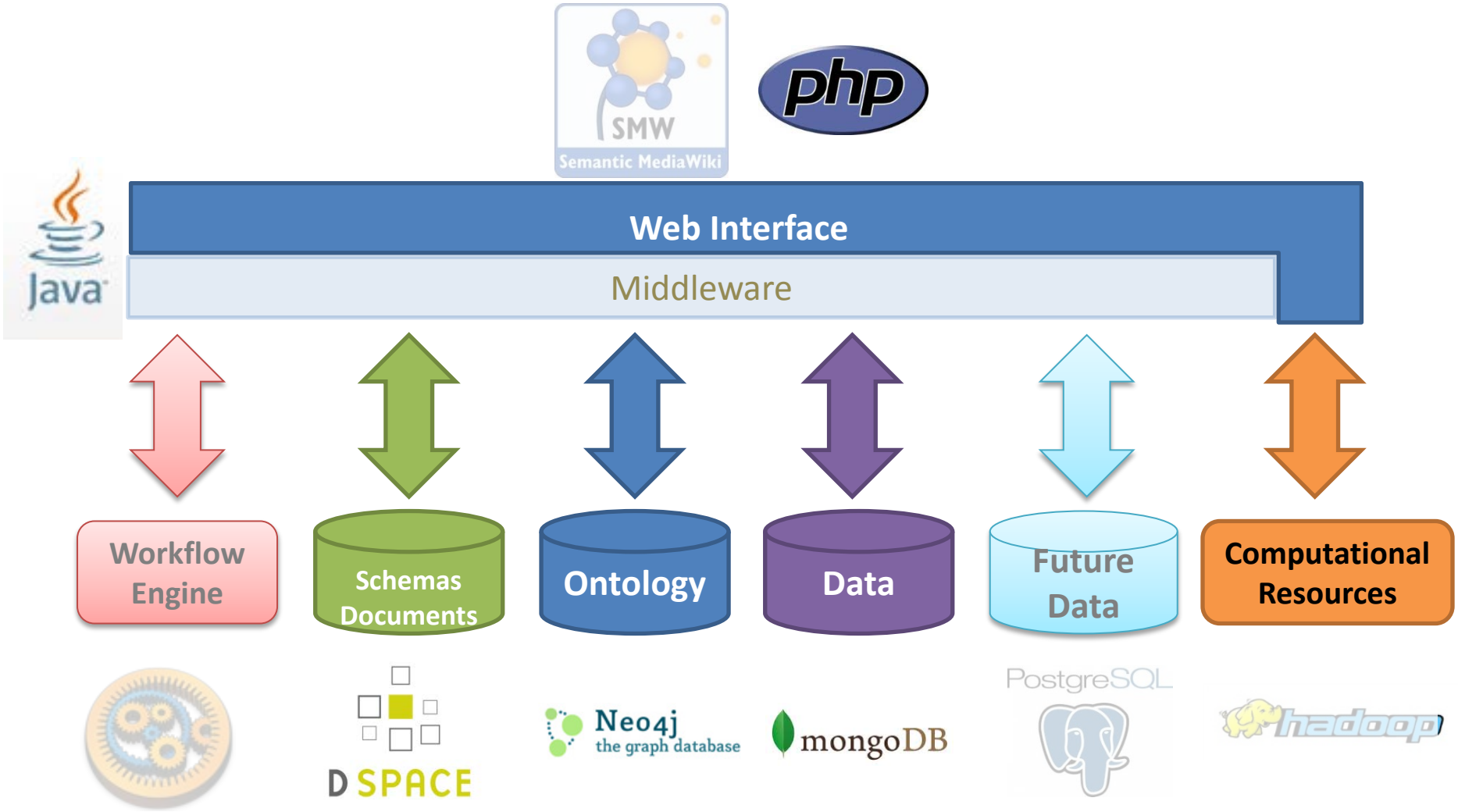
D = 2.16e-15 m2/s
$D = D_0 \exp(-Q/RT)$

Diffusion Coefficients
Diffusion Couple
Composition Profiles

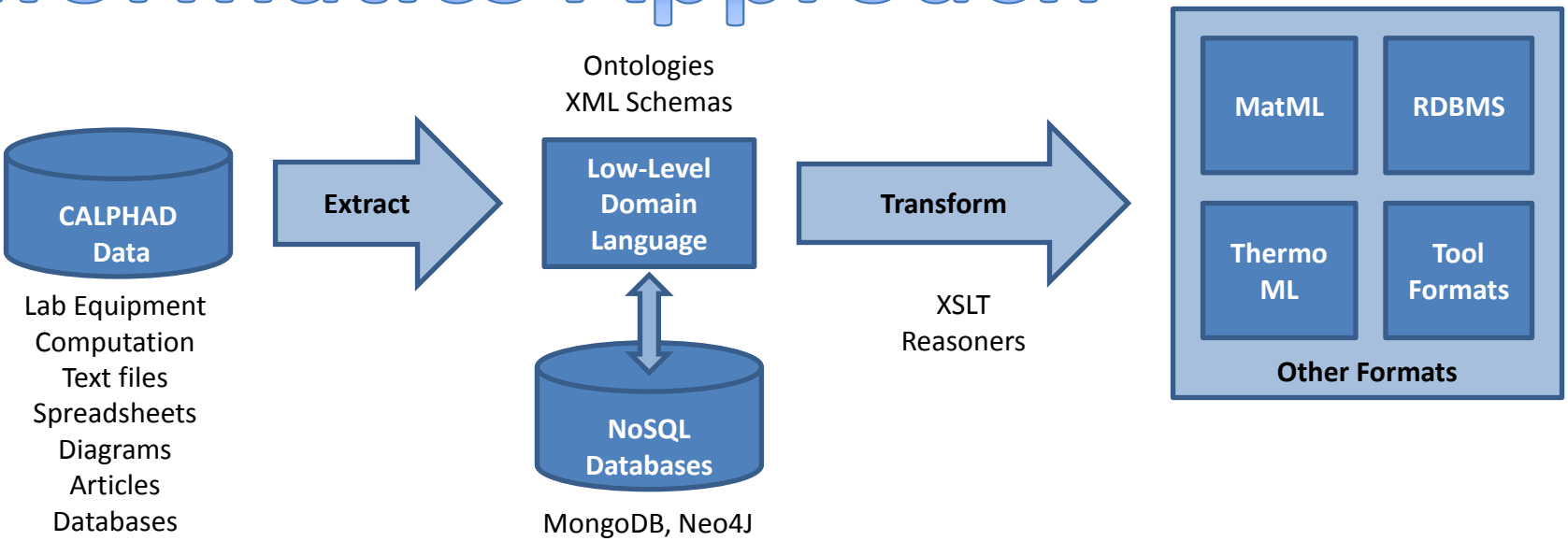
Composition Profiles
Raw Data (i.e Intensity vs Distance)

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

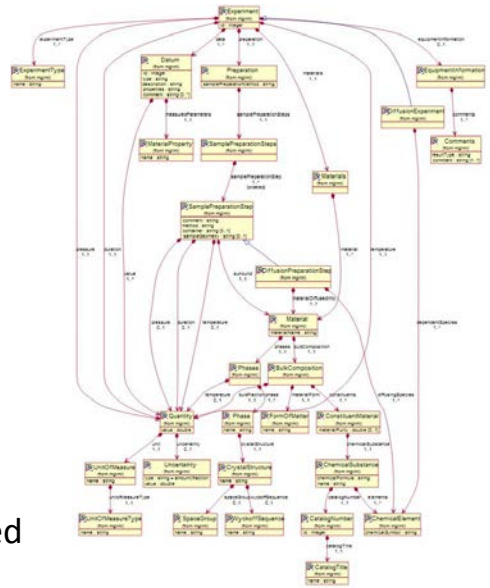
Architectural Strategy



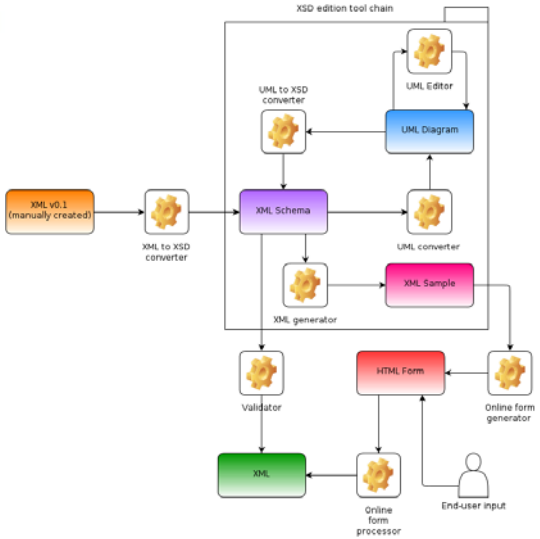
Informatics Approach



Section of XML Schema



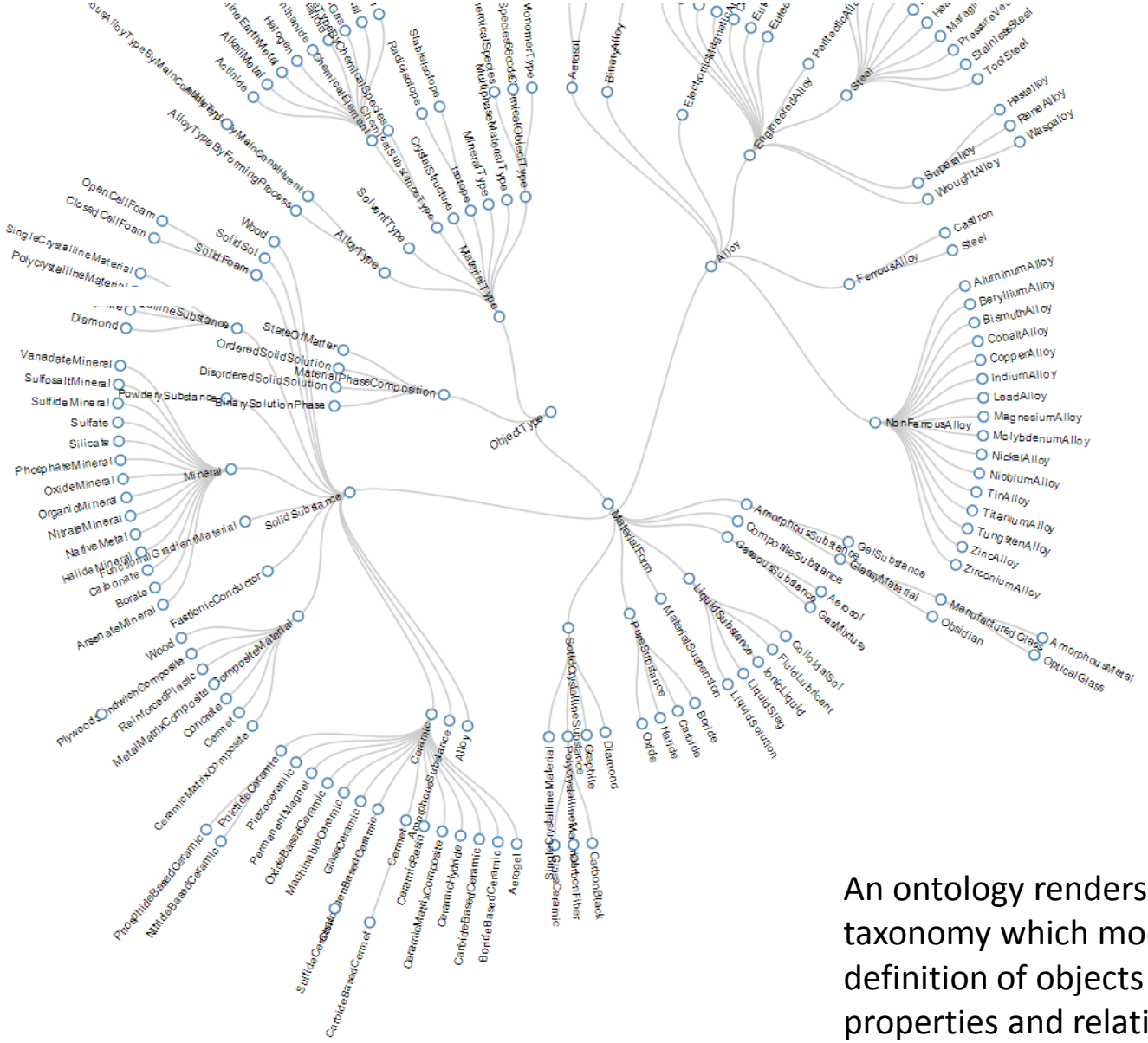
Tool Chain



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)

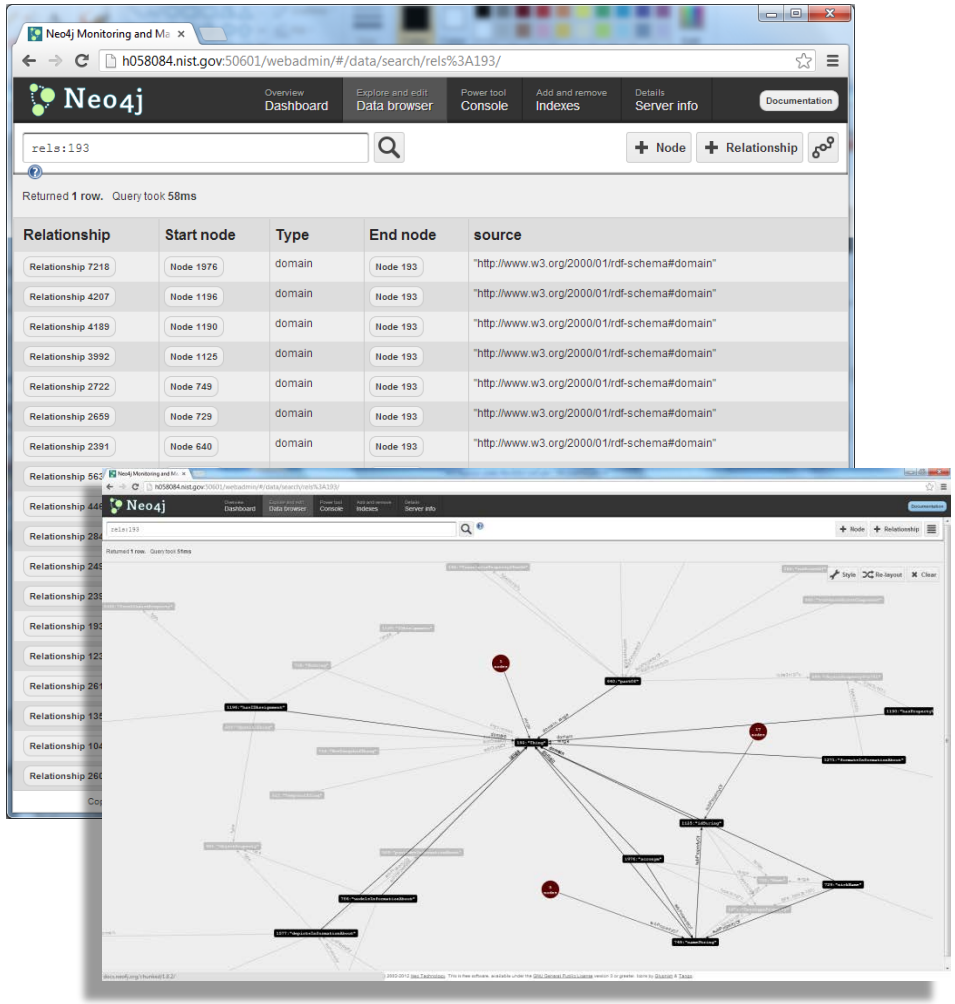
- Objects, Materials, and Events
- Physical Properties
- Documents
- Data Objects & Types
- People & Organizations
- Software
- Relations among these

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



- 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

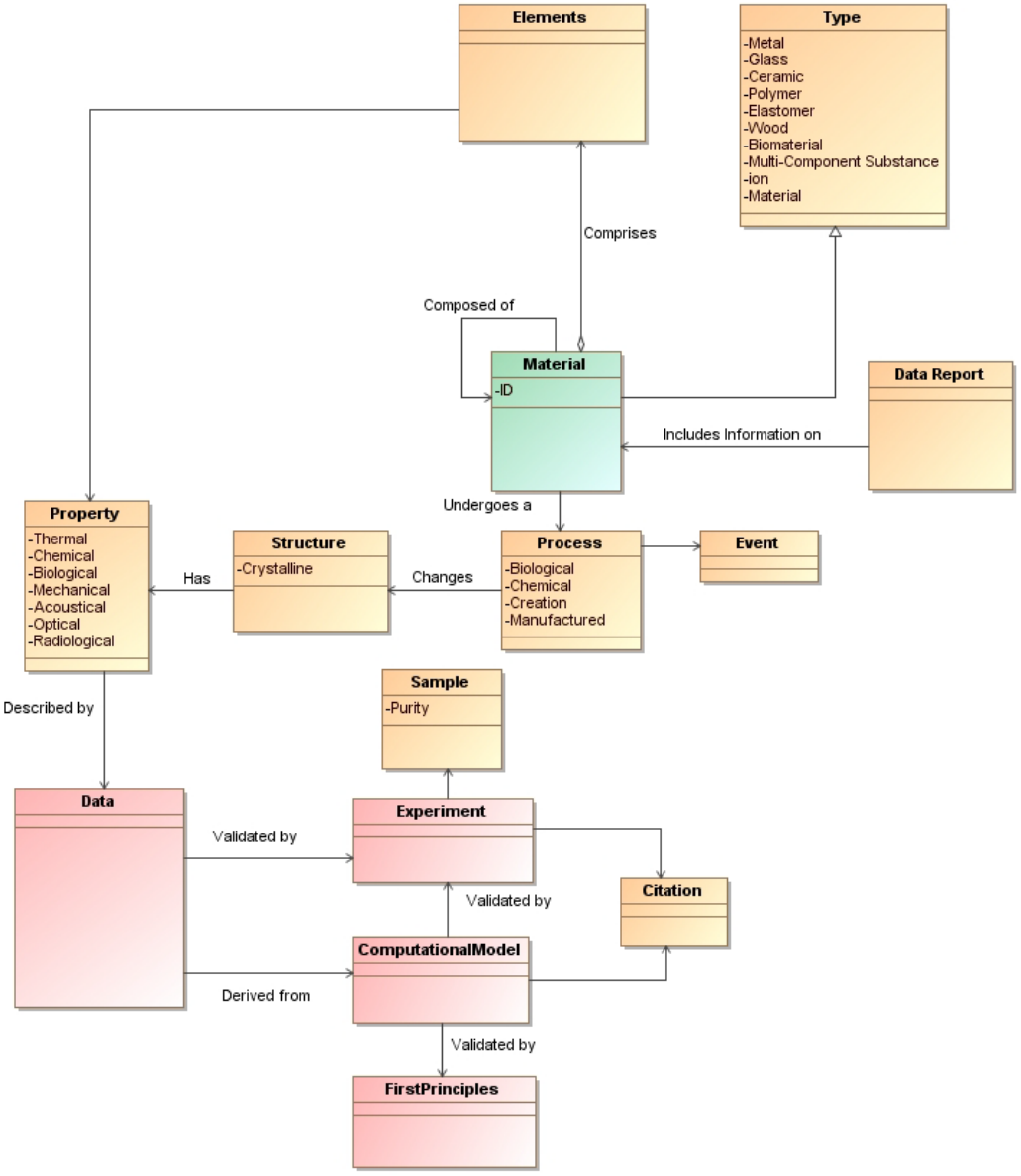
Sources

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

Tools

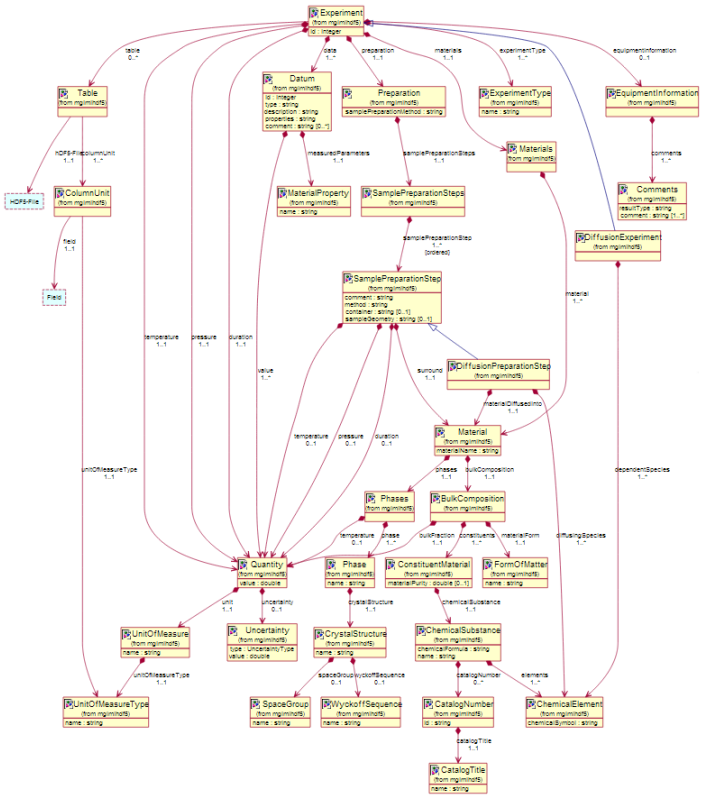
- UML (Unified Modeling Language)
- Semantic Web (RDF, OWL)

Note: This is a generalized model depicting overall structure

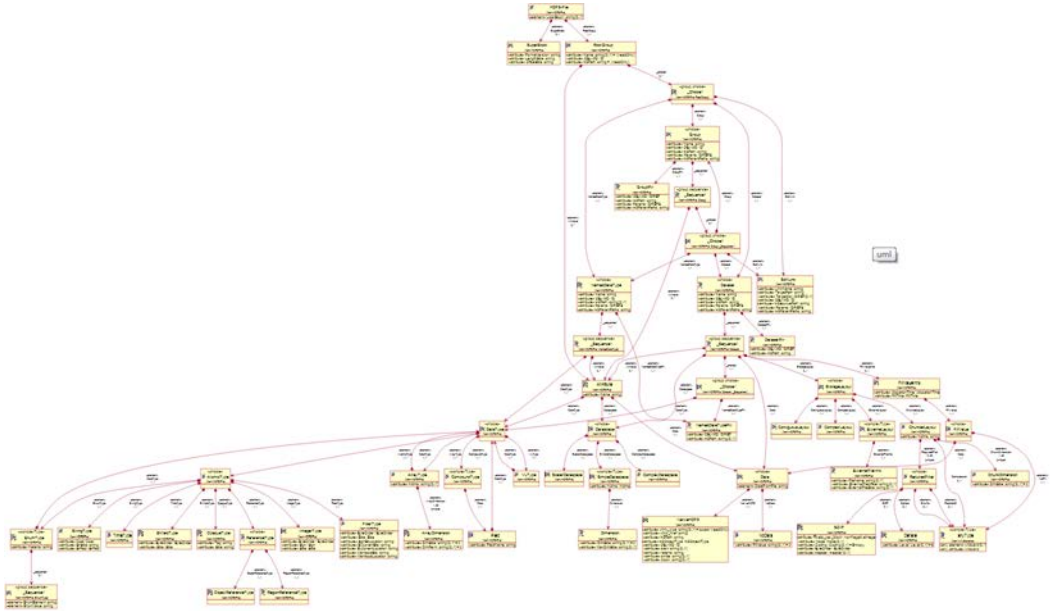


Encoding CALPHAD Data

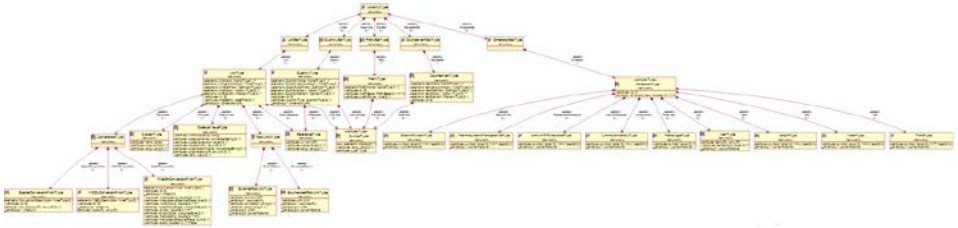
Core CALPHAD



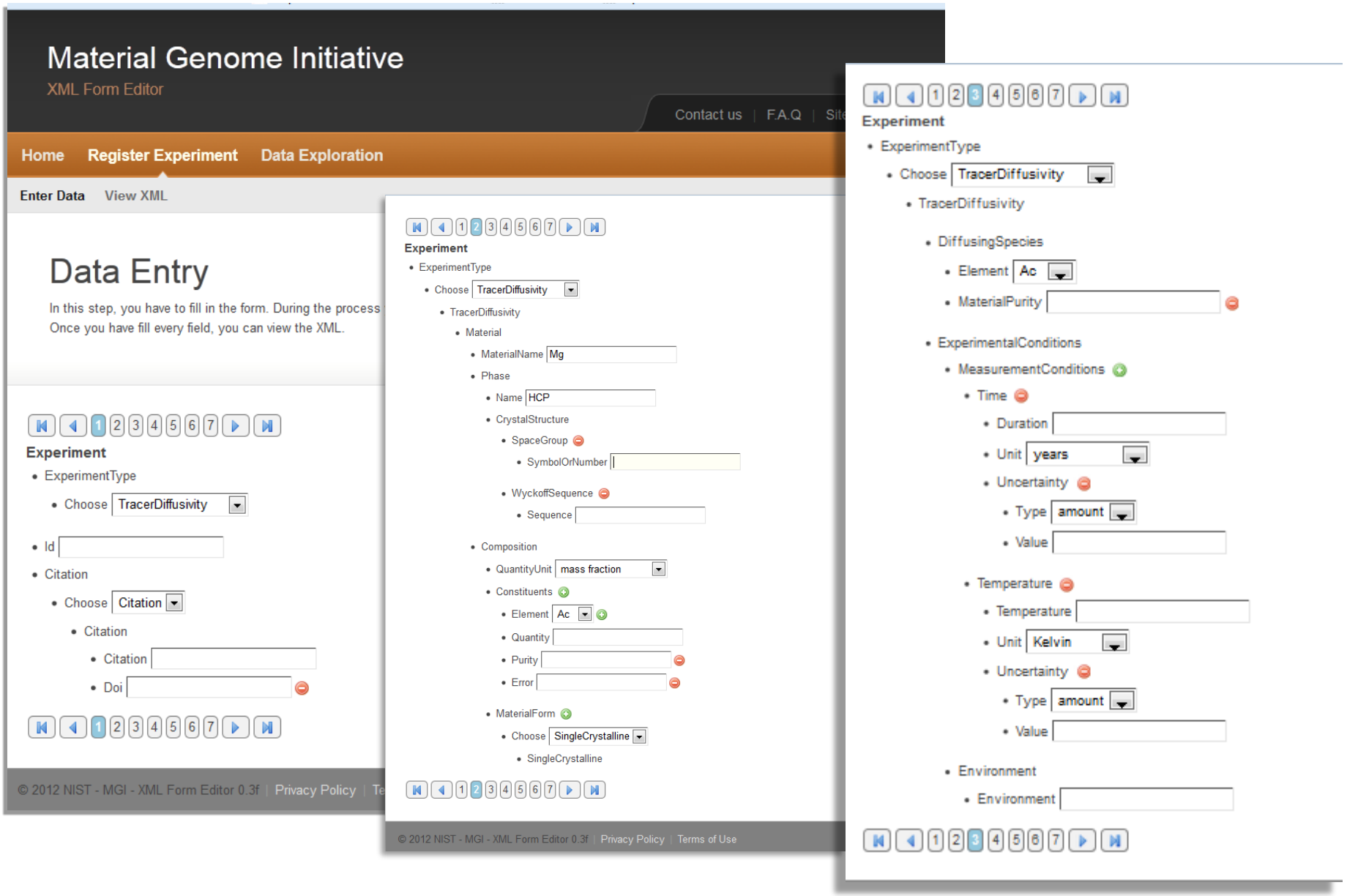
Tabular Data (HDF5)



Measurement Units (Units ML)



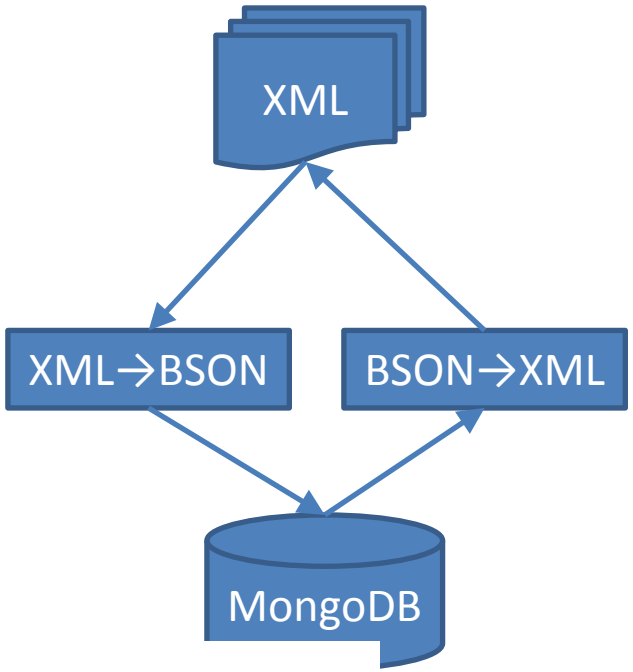
Data Collection: Tracer Diffusivity Test Schema



Snap-Shot of XML Format

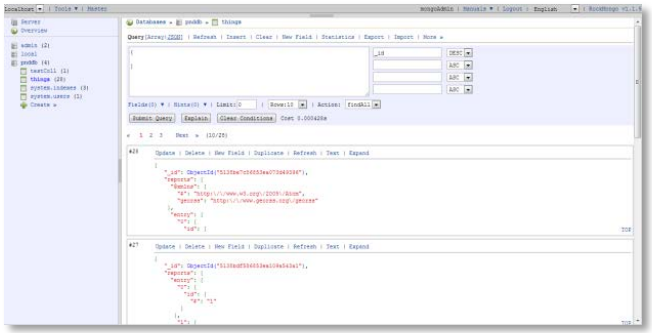
- Why XML?
  - Internationalization
  - Flexibility
  - Transformability
  - Interoperability
  - Longevity
  - Web-Enabled
  - Available Resources

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("4be97eaebcd1b30e86000003"),
  "title"        : "Ordered List",
  "creator_id"   : ObjectId("4be97eadbcd1b30e86000001"),
  "memberships" : [
    ObjectId("4be97eadbcd1b30e86000001"),
    ObjectId("4be97eaebcd1b30e86000002")
  ]
}
```

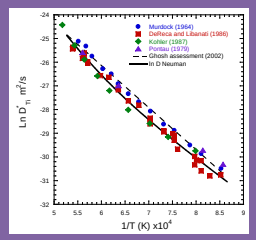


Initial Data Set: Reference Data for Self Diffusion Data for the Pure Elements

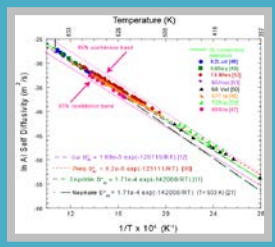
- Need for reference diffusion data for the pure elements
- **Current process:**

Search data

Digitize Data

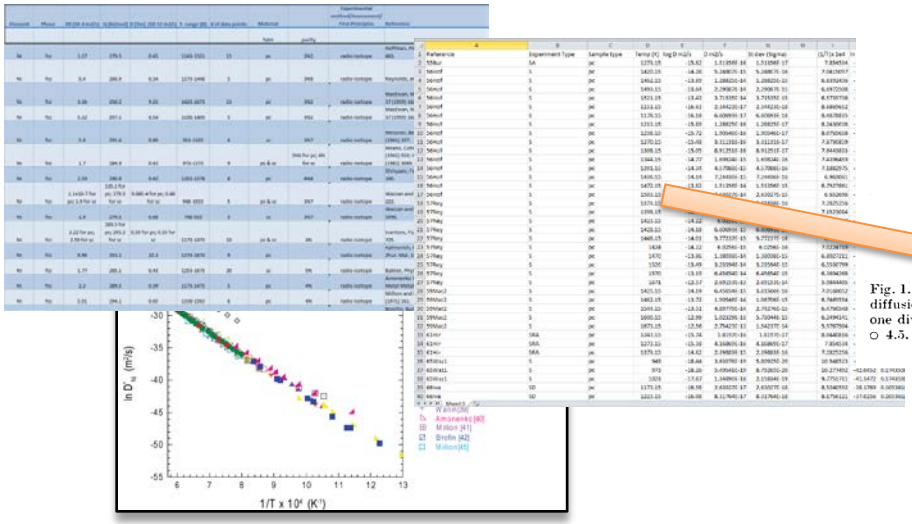


Plot data



Reference values

Future



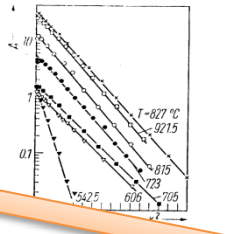
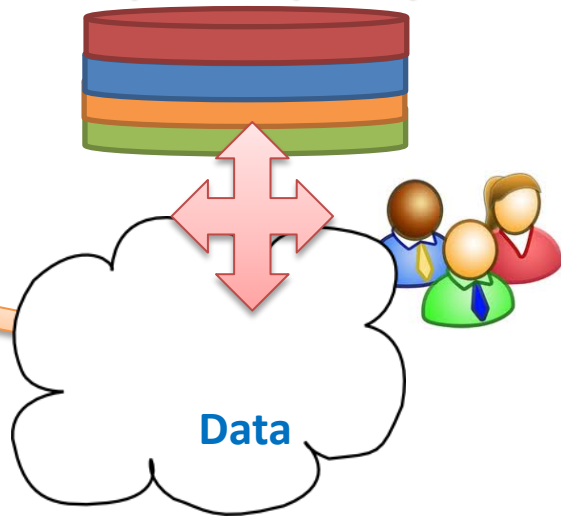


Fig. 1. Gaussian penetration plot... diffusion in nickel single crystals. Value of one division on the abscissa: □ 18, × 5.5, ○ 4.5, ● 5.7, ■ 3.4, ▽ 0.056, ▼ 0.19 × × 10^{-11} cm²

File Repository/DSpace

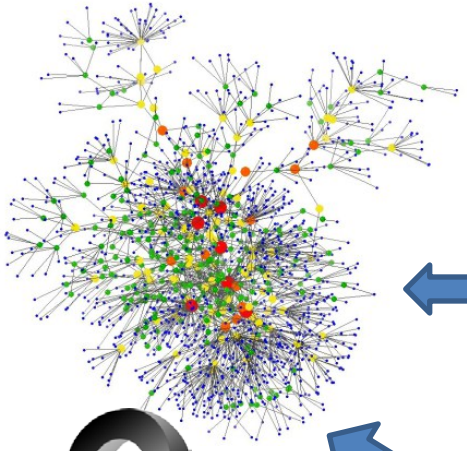


Future Data Informatics

User interface

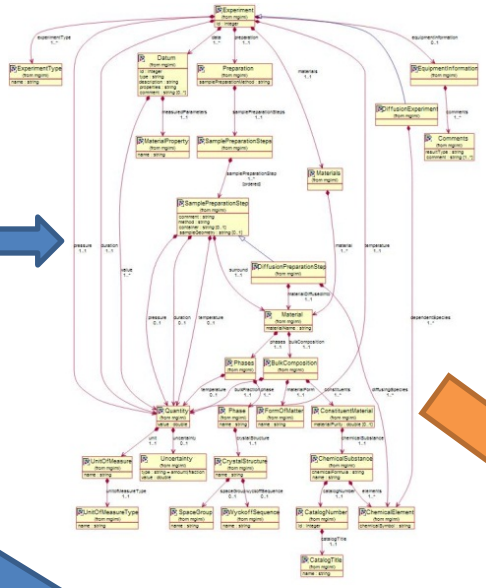
Data Capture

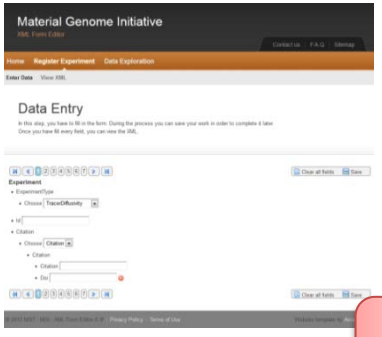
Ontology



JSON

XML Schema





Data Tools:
Statistics; Machine Learning

Various Database Platforms

mongoDB

D SPACE

Neo4j
the graph database

UNIFIED MODELING LANGUAGE

Java

RDF

XML

php

Semantic Web

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**