Data Informatics and Tools

C. E. Campbell, U.R. Kattner, Material Measurement Laboratory, NIST

A. Dima, Information Technology Laboratory, NIST

Laura Bartolo, Kent State University

NIST Diffusion Workshop

May 9, 2013

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

KENT STATE UNIVERSITY

Our Collaborators

NIST Information Technology Laboratory

Information Systems Group

Y.-S. Li-Baboud, P. Dessauw, P. Savonitto, D. Foxvog

NIST Material Measurement Laboratory

Technical Program Director for Materials Genomics

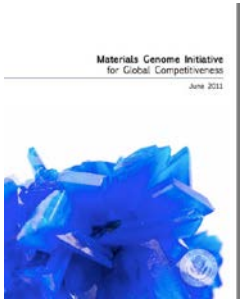
J.A. Warren

Thermodynamics and Kinetics Group

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

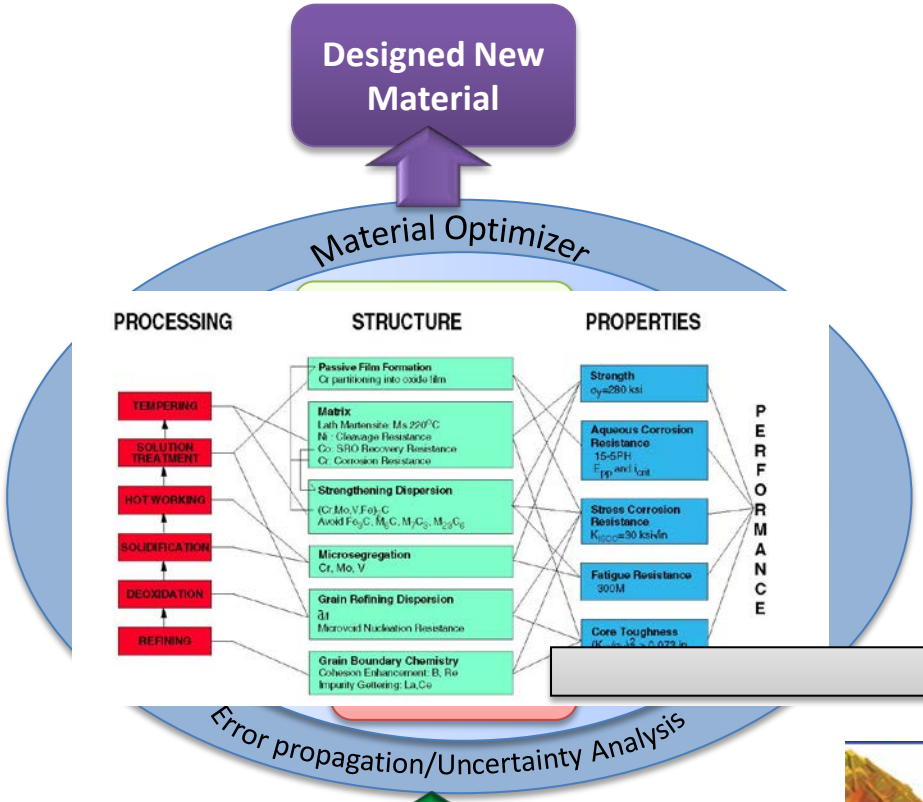
Cell Systems Science Group

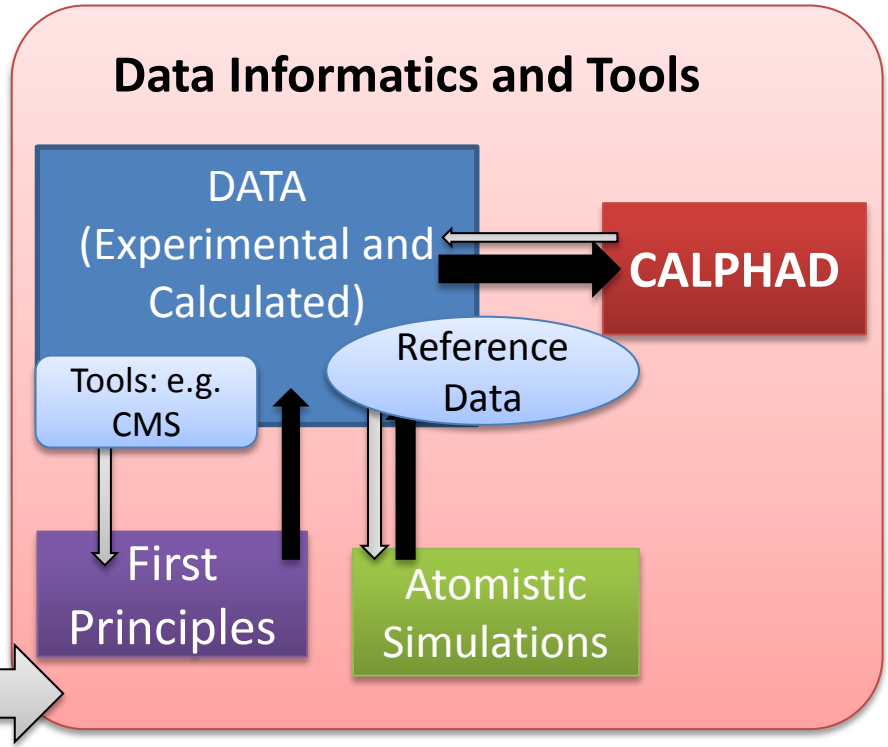
T.N. Bhat, J.T. Elliott

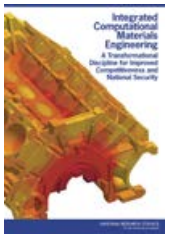


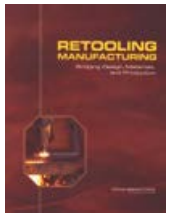
Materials Genome Initiative:
The Need for Data and Informatics

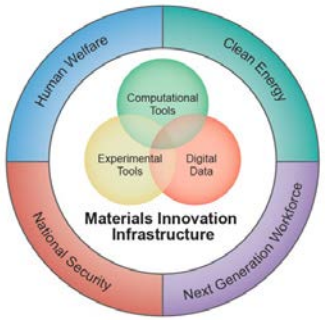
Goal: decrease the cost and time-to-market by 50%





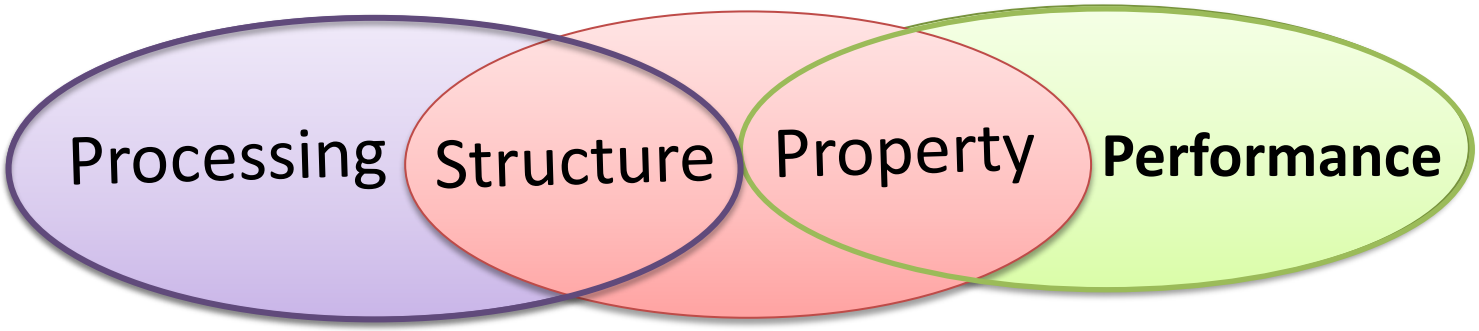




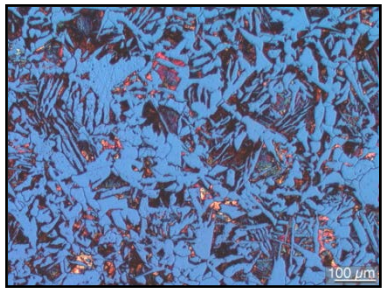


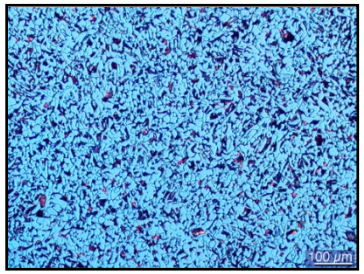
Materials Are Complicated Hierarchical Systems

- 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:

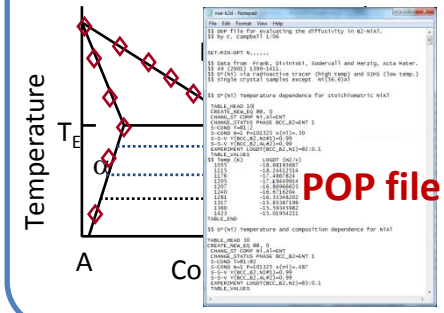
- What phases are present
- Composition and morphology of the phases present

CALPHAD

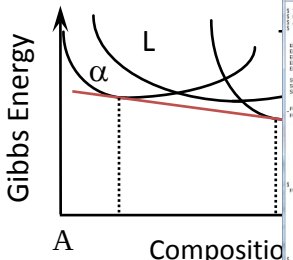
CALPHAD Approach

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

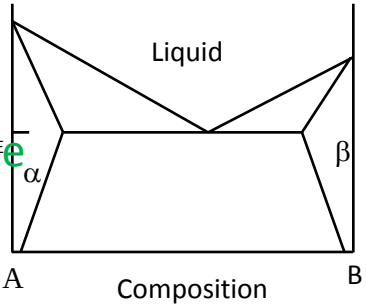
Experimental phase diagram and thermochemical data → Determine Gibbs energy functions for each phase: $G = f(x,T,P)$ → Calculated phase diagram → Combine binaries and ternaries to predict multi-component systems

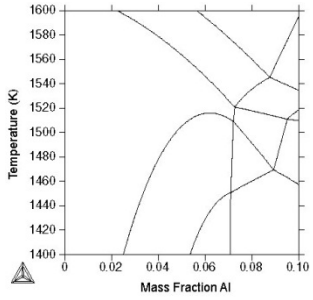


POP file



TDB file





Al isopleth for 8 component system

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

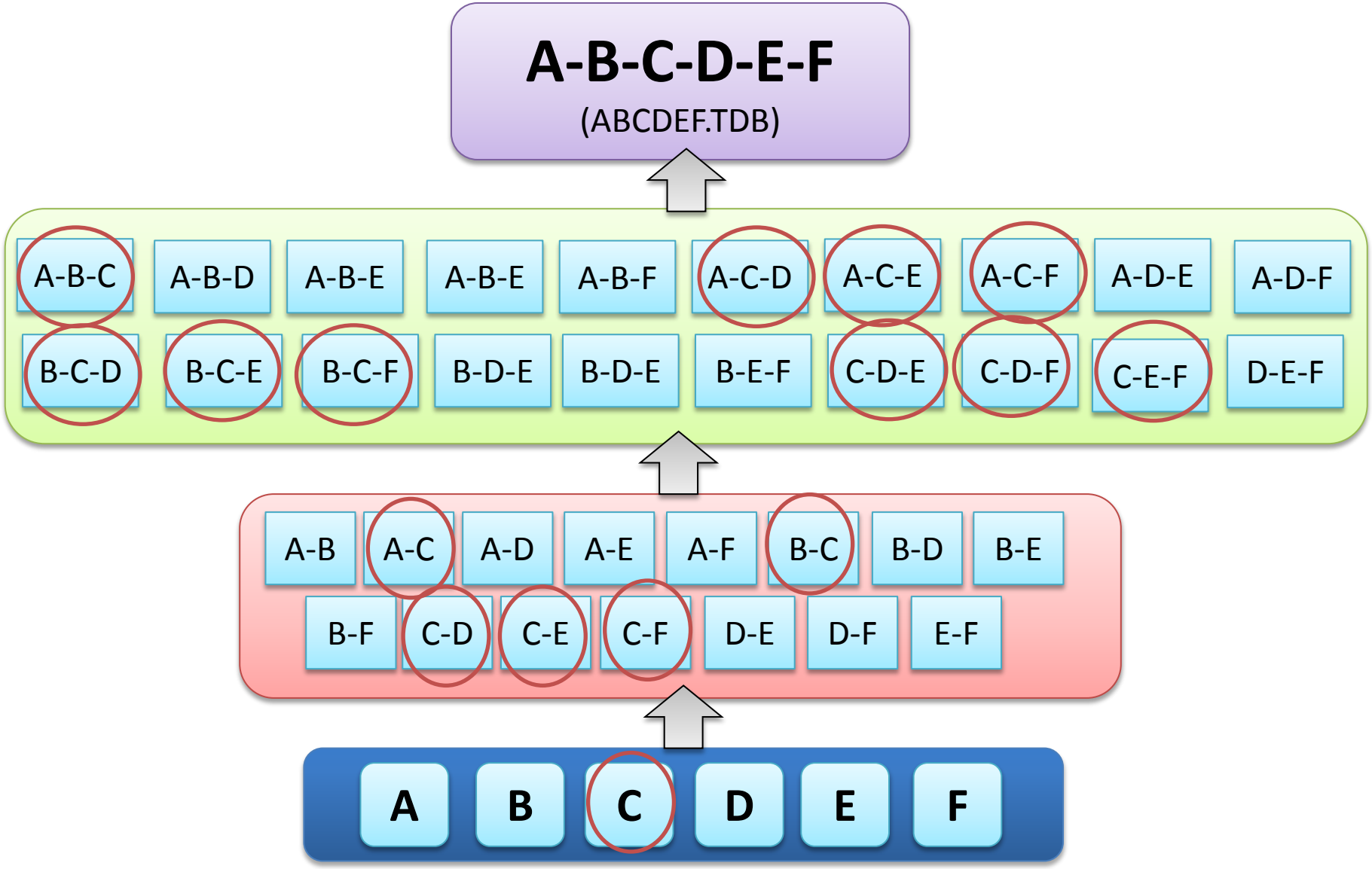
Binaries → Ternaries → Quaternaries → n^{th} Order Systems

➡ **True quaternary compounds are rare in metallic systems**

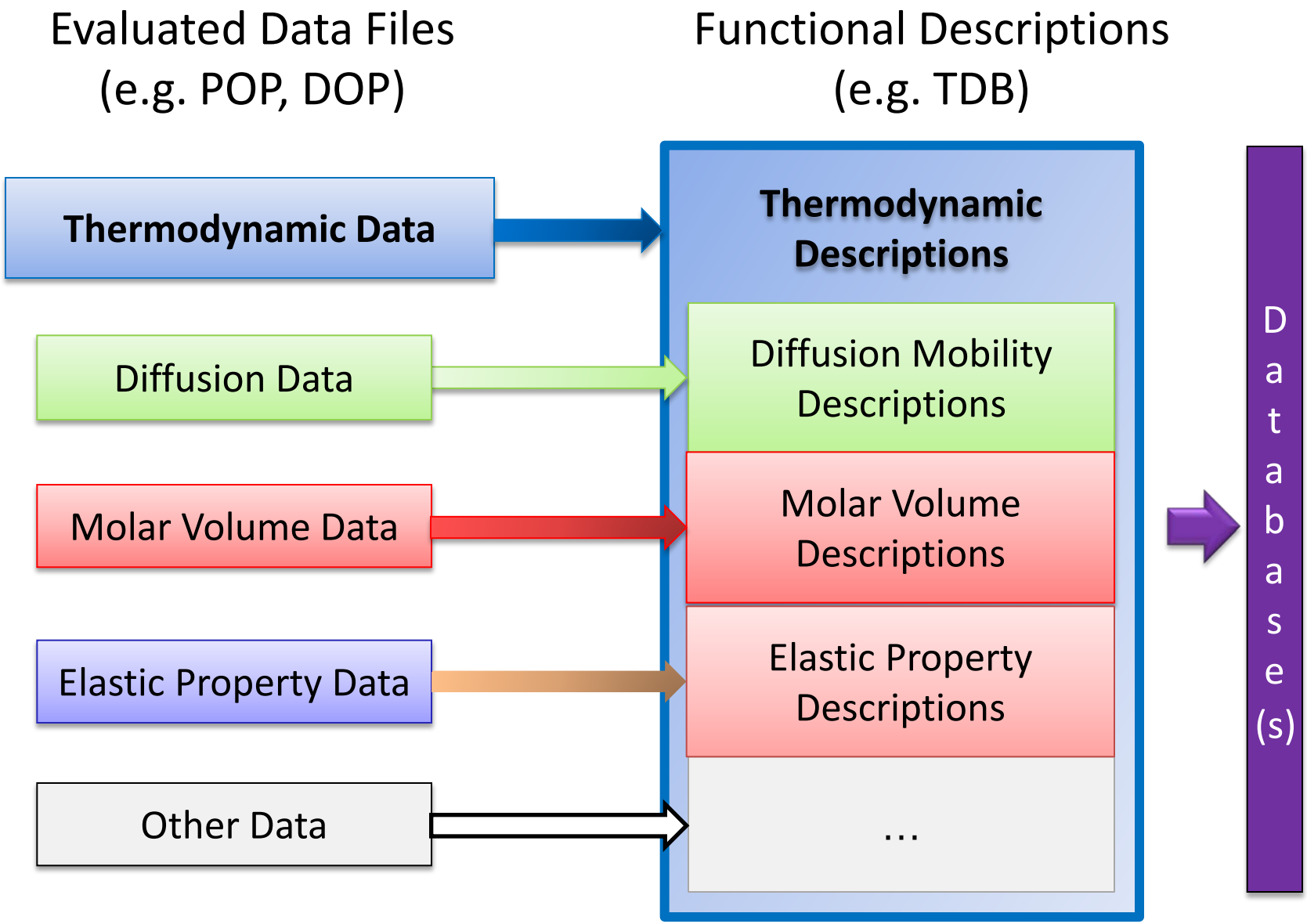
➡ **Assessment of ternary systems is usually sufficient for the description of a multicomponent system**

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

Data Dependencies



Data Flow of CALPHAD "Database" Files



Examples of Files for a CALPHAD Thermodynamic Assessment

Functional Description File Thermodynamics

- Elements: e.g. A,B
- Phases: e.g. FCC_A1 and BCC_A2 (crystal structure)
- Reference states used
- Software/version used
- Reference info
  - Bibliographic info
  - DOI link
- Contributor info
  - Name
  - E-mail
- User Comments
- Links to other files

Evaluated Data Files Thermodynamics (POP)

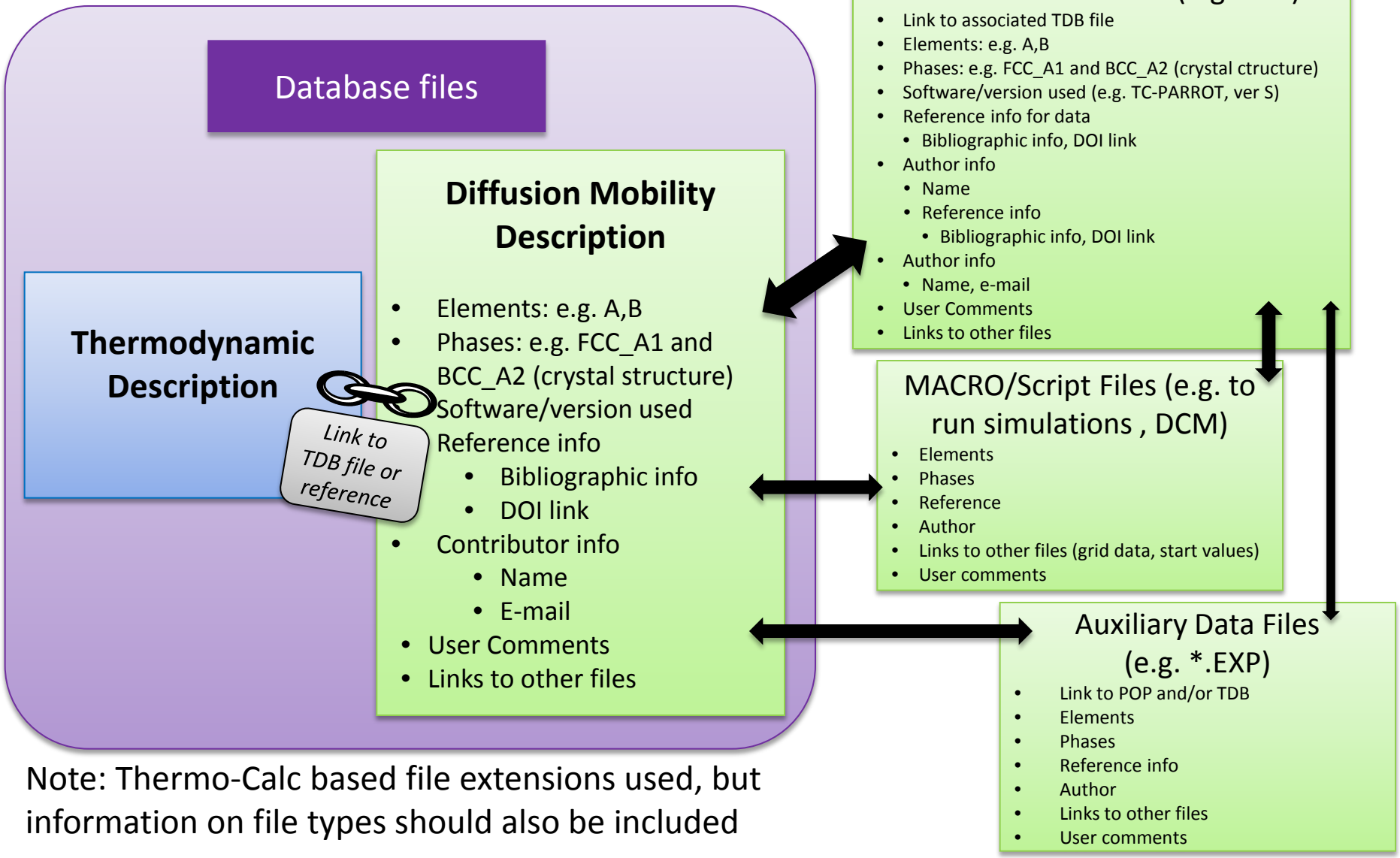
- Link to associated TDB file
- Elements: e.g. A,B
- Phases: e.g. FCC_A1 and BCC_A2 (crystal structure)
- Software/version used (e.g. TC-PARROT, ver S)
- Reference info for data
  - Bibliographic info
  - DOI link
- Author info
  - Name
  - Reference info
    - Bibliographic info
    - DOI link
- Contributor info
  - Name
  - E-mail
- User Comments
- Links to other files

Auxiliary Files (e.g. EXP, TCM)

- Link to POP and/or TDB
- Elements
- Phases
- Reference info
- Author
- User
- Links to other files

Note: Thermo-Calc based file extensions used, but information on file types should also be included

EXAMPLES OF FILES FOR A CALPHAD: DIFFUSION MOBILITY ASSESSMENT



Note: Thermo-Calc based file extensions used, but information on file types should also be included

Selecting a Repository System

- **Must provide**
  - the ability to link files
  - tags for properties, phases and models,
  - customize metadata
  - search capability
  - customization with own tools
  - the ability to link with other repository systems
  - **persistent identifiers for data**
  - **License for data use**

- Must be flexible enough to adapt to future needs
- Must be easy to use and install

- **Systems explored**

Free Software

Populating the Repository

Tags are needed to facilitate searches and provide information about the system and the file contents!

The effort required to populate repository must be minimal to be attractive for contributions from the community!

Essential Tags:

- Elements
- Reference
- Property (thermodynamics, diffusion, …)
- Phases

Suggested Common Tags for CALPHAD; First Principles and Atomstics

- **Methods**
  - First-Principles
  - Atomistics
  - CALPHAD
  - Other
- **Property Type**
  - Thermodynamic
  - Kinetic
  - Molar Volume
  - Mechanical
  - Electrical
  - Optical
  - Magnetic
  - Interface/Surface
  - *Other*
- **Systems**
  - **A-G**
    - Al Binaries
    - Al Ternaries
    - Al High Order Systems
  - **H-L**
  - **……**

- **State**
  - Gas
  - Liquid
  - Other
  - Solid
    - Bulk
    - Nano
    - Phases
- **Software**
  - **First Principles**
    - VASP
    - DMOL
    - ABINIT
    - SIESTA
    - WIEN2k
    - Materials Studio
    - *other*

Different tags for different types of data :
Experimental data

-- CALPHAD
- Thermo-Calc
- Pandat
- FactSage
- OC
- MTData
- *other*

-- ATOMISTICS
- LAMMPS
- IMD
- *other*

Suggested Common Tags for Experiments

- **Alloy Class**
  - **….**
  - **Mg Alloys**
    - Cast Alloys
      - Mg-Al
      - Mg-Al-Mn
      - Mg-Al-Zn
      - Mg-Al-RE (RE-Rare Earth)
    - Wrought Alloys
      - Mg-Al
      - Mg-Li
      - Mg-Zn
- **Systems**
  - **A-G**
    - Al Binaries
    - Al Ternaries
    - Al High Order Systems
  - **H-L**
  - **……**
- **State**
  - Gas
  - Liquid
  - Other
  - Solid (Bulk; Nano; Phase)

- **Property Type**
  - **Thermodynamic**
    - Enthalpy
    - Activity/chemical potential
    - Heat Capacity
    - Phase Equilibria
      - » Phase boundaries
      - » Transition temperature
      - » Single phase defect
  - **Kinetic**
    - Diffusivities
      - Tracer
      - Intrinsic
      - Interdiffusion
    - Nucleation
  - **Molar Volume**
    - Lattice parameters
    - Thermal Expansion
  - **Mechanical**
  - **Electrical**
  - **Thermal Conductivity**
  - **Optical**
  - **Magnetic**

- Interface/Surface
- Corrosion
  - Potential-pH (Pourbaix) diagrams
  - Polarization measurements
  - Electrochemical measurements
  - Corrosion rates
- *Other*

Suggested Common Tags for Experiments

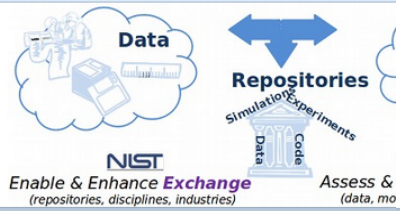
- **Experimental Methods**
  - Image Analysis
    - Optical Metallography
    - SEM
    - TEM
    - Atom Probe
  - X-Ray Diffraction
  - Dilatometry
  - Calorimetry
    - Differential Scanning Calorimetry
    - Solution
    - Drop
    - Other
  - Differential Thermal Analysis
  - Electromotive Force
  - Vapor pressure
    - Knudsen
    - Other
  - Composition Analysis
    - Electron Probe Microanalysis
    - SIMS

- **Processing**
  - Casting
    - Centifugal
    - Die
    - Investment
    - Sand/mold
  - Deformation
    - Drawing
    - Extrusion
    - Forging
    - Rolling
    - Sheet
  - Machining
    - Abrasive
    - Electrical/chemical
    - Machining (i.e. Drilling , Milling)
    - Sheet/film (punching, perforating,)
  - Thermal
    - Electron beam Machining
    - Laser cutting
    - Plasma arc cutting

  - Molding
    - Blow
    - Calendering
    - Compression
    - Injection
  - Powder Processing
    - Cold
    - Hot
  - Thermal Processing
    - Homogenization
    - Aging

File Repository/DSpace

http://www.camm.ohi... Inside NIST Add to Wish List Zip Code 20892



NIST File Repositories → NIST File Repositories

NIST File Repositories

Browse by

- By Issue Date
- Authors
- Titles
- Subjects

Search within this community and its collections: Go

The National Institute of Standards and Technology is establishing essential data ex protocols and mechanisms to ensure the quality of materials data and models need widespread adoption and data sharing.

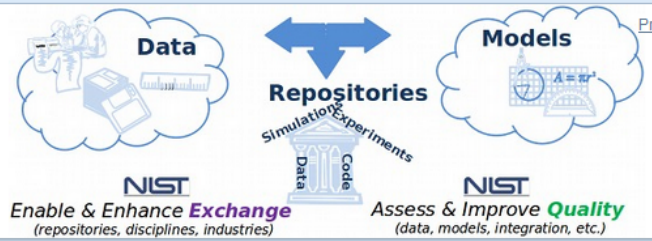
Collections in this community

- CALPHAD Assessments
- First Principles Phase Stability (FPPS) Files
- Interatomic Potentials

Recent Submissions

Al_Cr_Ni Diffusion Mobilities in Gamma Prime and B2
Campbell, C.E. (2013-02-11)
This work presents the assessment of the diffusion mobilities in both the γ' (Ni3Al-L12) and B2 in the Ni–Al–Cr system utilizing the phenomenological model developed by Helander and Åg experimental ...

AlN–GaN; GaN–InN; AlN–InN
Burton, Benjamin; van de Walle, Anton; Kattner, Ursula (2013-01-31)
First principles phase diagram calculations were performed for the wurtzite-structure quasib AlN–GaN, GaN–InN, and AlN–InN.



Profile: Carelyn Campbell | Logout

NIST File Repositories → NIST Data File Repositories → CALPHAD Assessments

CALPHAD Assessments

Browse by

- By Issue Date
- Authors
- Titles
- Subjects

Search within this collection: Go

Submit a new item to this collection

Recent Submissions

Al_Cr_Ni Diffusion Mobilities in Gamma Prime and B2
Campbell, C.E. (2013-02-11)
This work presents the assessment of the diffusion mobilities in both the γ' (Ni3Al-L12) and B2 (NiAl) phases in the Ni–Al–Cr system utilizing the phenomenological model developed by Helander and Ågren. Available experimental ...

Ni-Al-Cr system Thermodynamic Re-Assessment of the Ternary System Al-Cr-Ni
Dupin, N.; Ansara, I.; Sundman, B. (2013-01-31)
A re-assessment of the ternary system A1-Cr-Ni following Dupin's thesis work using a single Gibbs energy function for the gamma and gamma prime phases is presented taking into account new experimental liquidus temperatures. ...

Ag-Al Functional Description
Du, Zeting; Jing, Zhan-Peng; Li, Changrong; Niu, Chunji (2013-01-31)
The energy expressions for GP zones in the Al–Ag binary system, including the ε-state and the η-state ones, are established by combining the essential Gibbs energy for the matrix alloy with the interfacial energy and the ...

Search DSpace

Go

Search DSpace
This Collection
Advanced Search

Browse

All of DSpace
- Communities & Collections
- By Issue Date
- Authors
- Titles
- Subjects

This Collection
- By Issue Date
- Authors
- Titles
- Subjects

My Account

Logout
Profile
Submissions

Context

Edit Collection
Item Mapper
Export Collection
Export Metadata

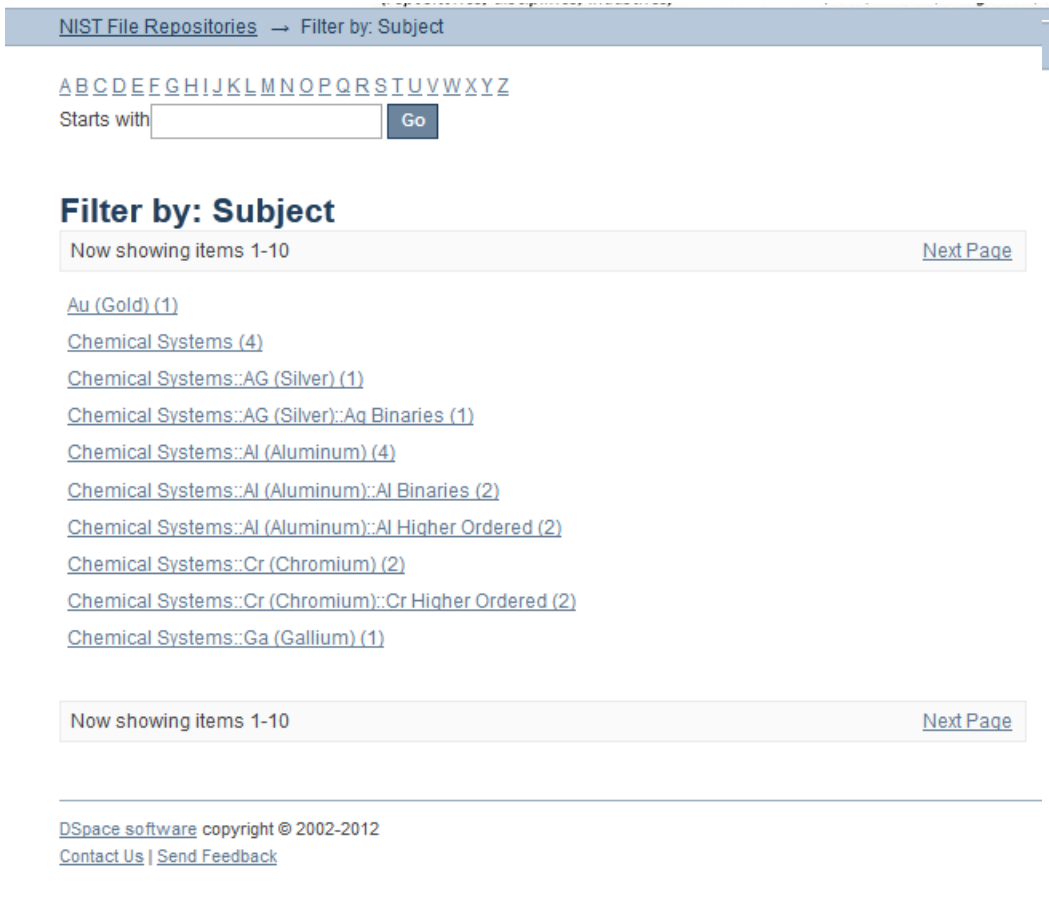
Statistics

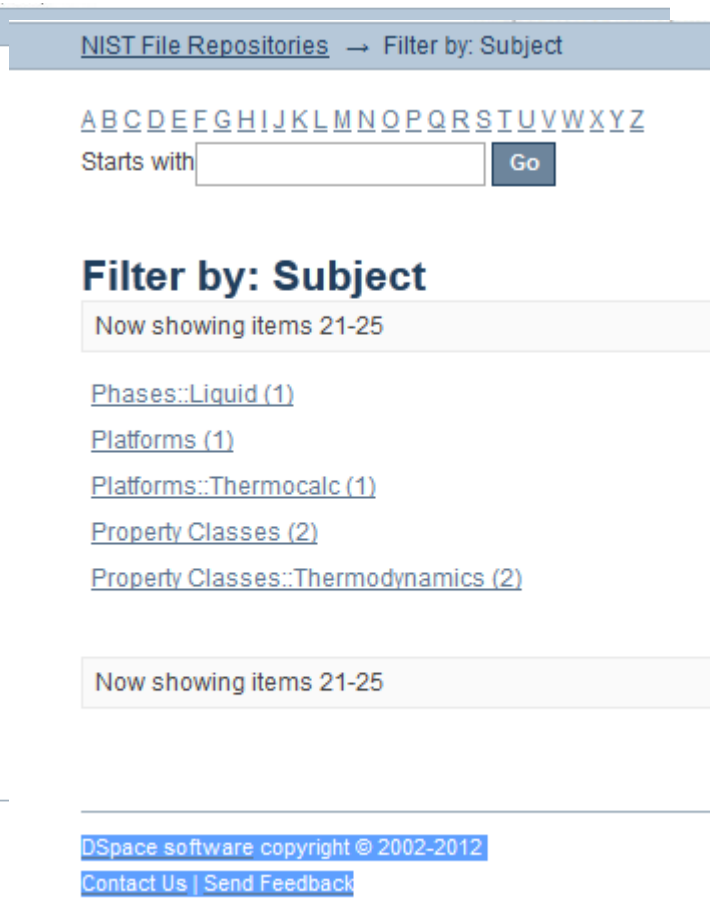
View Usage Statistics
View Search Statistics
View Workflow Statistics

Discover

Sample Entry

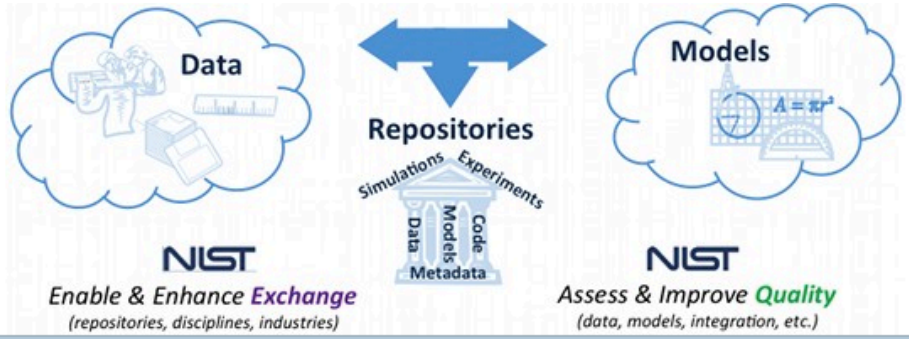
Current Advance Search Options





Tag/Search Terms will eventually will be tie to ontology

Upload & Embargo Unpublished Data



Data submission

Describe → Describe → Access → Upload → Review → CC License → License → Complete

Access Settings

Private Item:

If selected, the item won't be searchable

☑

Embargo Access until Specific Date:

Accepted format: yyyy, yyyy-mm, yyyy-mm-dd

2014-05-12

Reason:

Keep private until article is published

Search NIST

Search NIS
This Collec

Advanced Sea

Browse

All of NIST Re
Communiti
Subjects
Titles
Authors
This Collectior
Subjects
Titles
Authors

My Account

My Exports
Logout
Profile
Submissions

Creative Commons 3.0 with 4 versions

http://creativecommons.org/licenses/

Attribution:
You must attribute the work in the manner specified by the author or licensor. (Not to be used for endorsements)

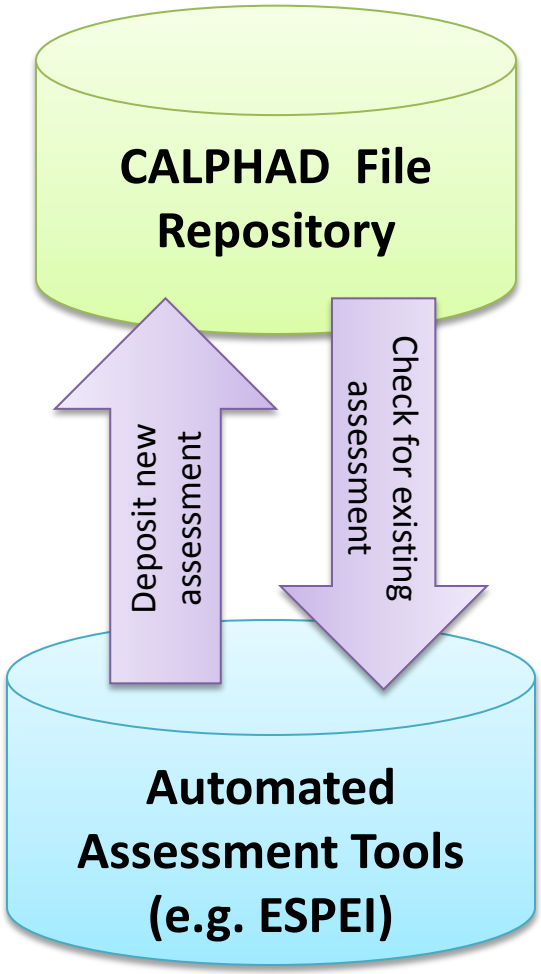
4 versions of 3.0 License allows one to choose:

- To share, adapt, & make commercial use of work.
- To share, adapt & no commercial use of work
- To share, adapt, make commercial use of work & share alike.
- To share, adapt, no commercial use of work & share alike.

Share — Copy, distribute, and transmit work
Share Alike — re-distribution & re-use of licensed work under the same, similar or a compatible license.
Waiver — Any of the above conditions can be waived if you get permission from the copyright holder.
Public Domain — Where the work or any of its elements is in the public domain under applicable law, that status is in no way affected by the license.
Other Rights — In no way are any of the following rights affected by the license:
Your fair dealing or fair use rights, or other applicable copyright exceptions and limitations;
The author's moral rights;
Rights other persons may have either in the work itself or in how the work is used, such as publicity or privacy rights.
Notice — For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.

Improving the CALPHAD Assessment Process

What if the evaluated data file (e.g. POP) is missing?
Are there new experimental or computational data?



Requires data search

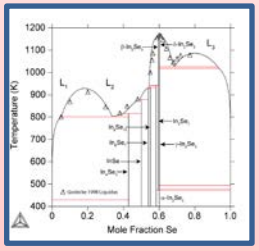
Examples of CALPHAD Data Types

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.

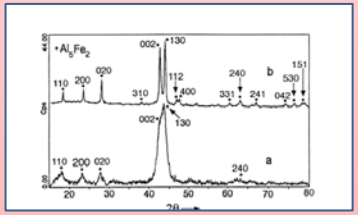
1-D (Points)

Melting Temperatures



Critical Temperatures (Phase Changes)

Lattice Parameters



Heat of Formations

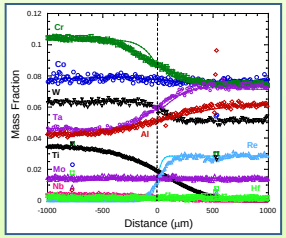
Phase fractions and compositions

Tracer Diffusivities

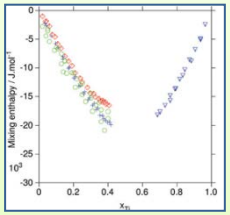
Activation energies

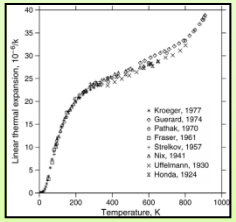
2-D (Lines)

Composition Profiles



Heat Capacities

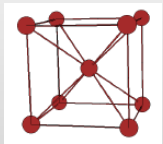




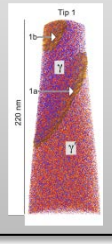
Enthalpies of mixing

3-D

Crystal structures



Micrographs/Morphologies



3-D Atom probe Tomography

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

GRANTA MATERIAL INTELLIGENCE

MATERIALS PROJECT

MatWeb

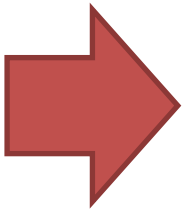
ASM INTERNATIONAL The Materials Information Society

Everything Material.

MatNavi

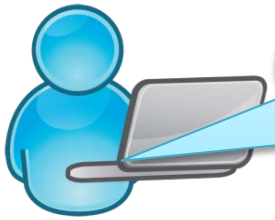
AFLOWLIB.ORG

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

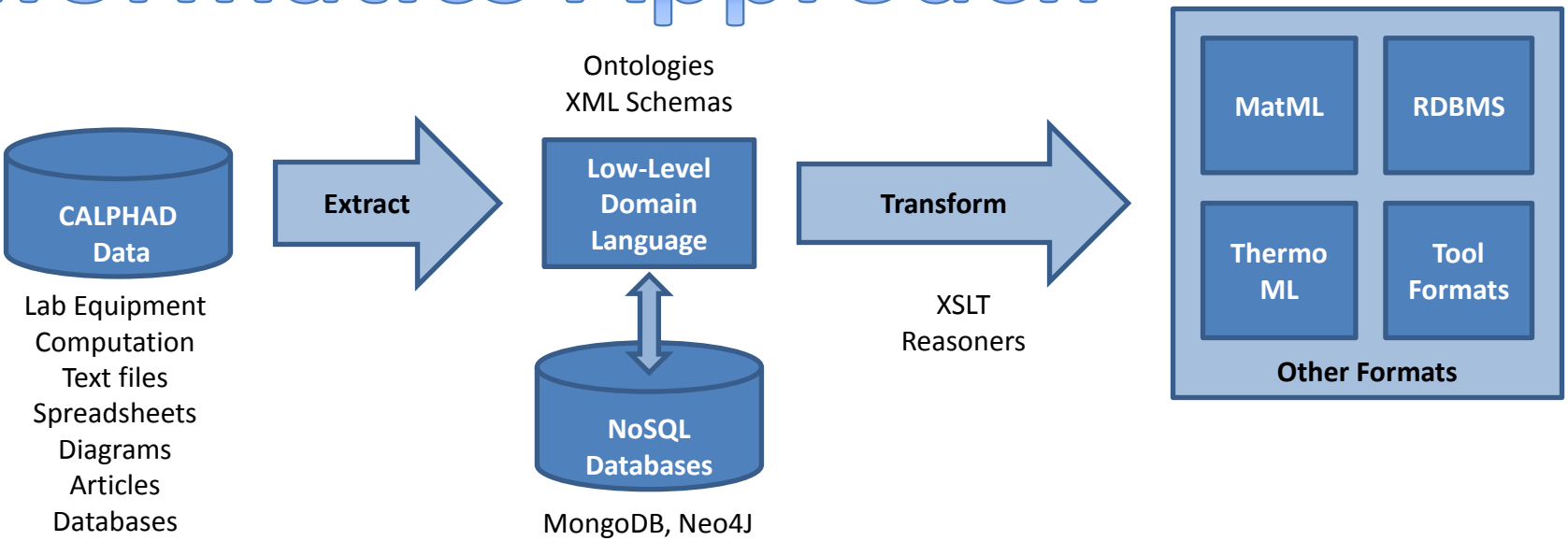
D = 2.16e-15 m2/s
$D = D_0 \exp(-Q/RT)$

Diffusion Coefficients
Diffusion Couple
Composition Profiles

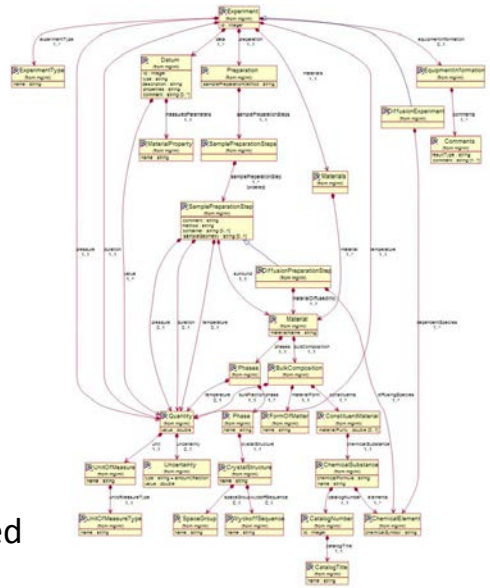
Composition Profiles
Raw Data (i.e Intensity vs Distance)

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

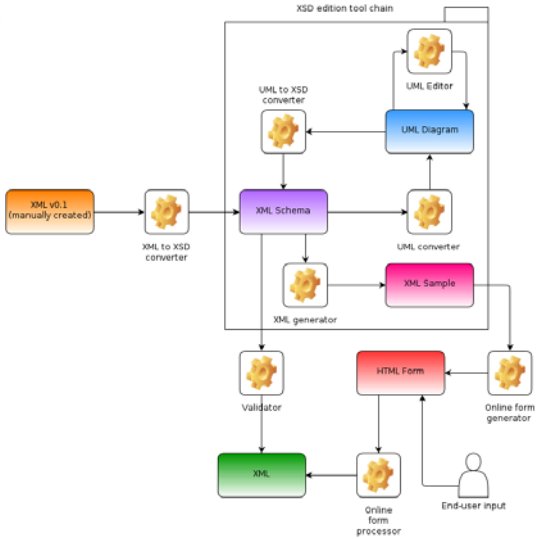
Informatics Approach



Section of XML Schema



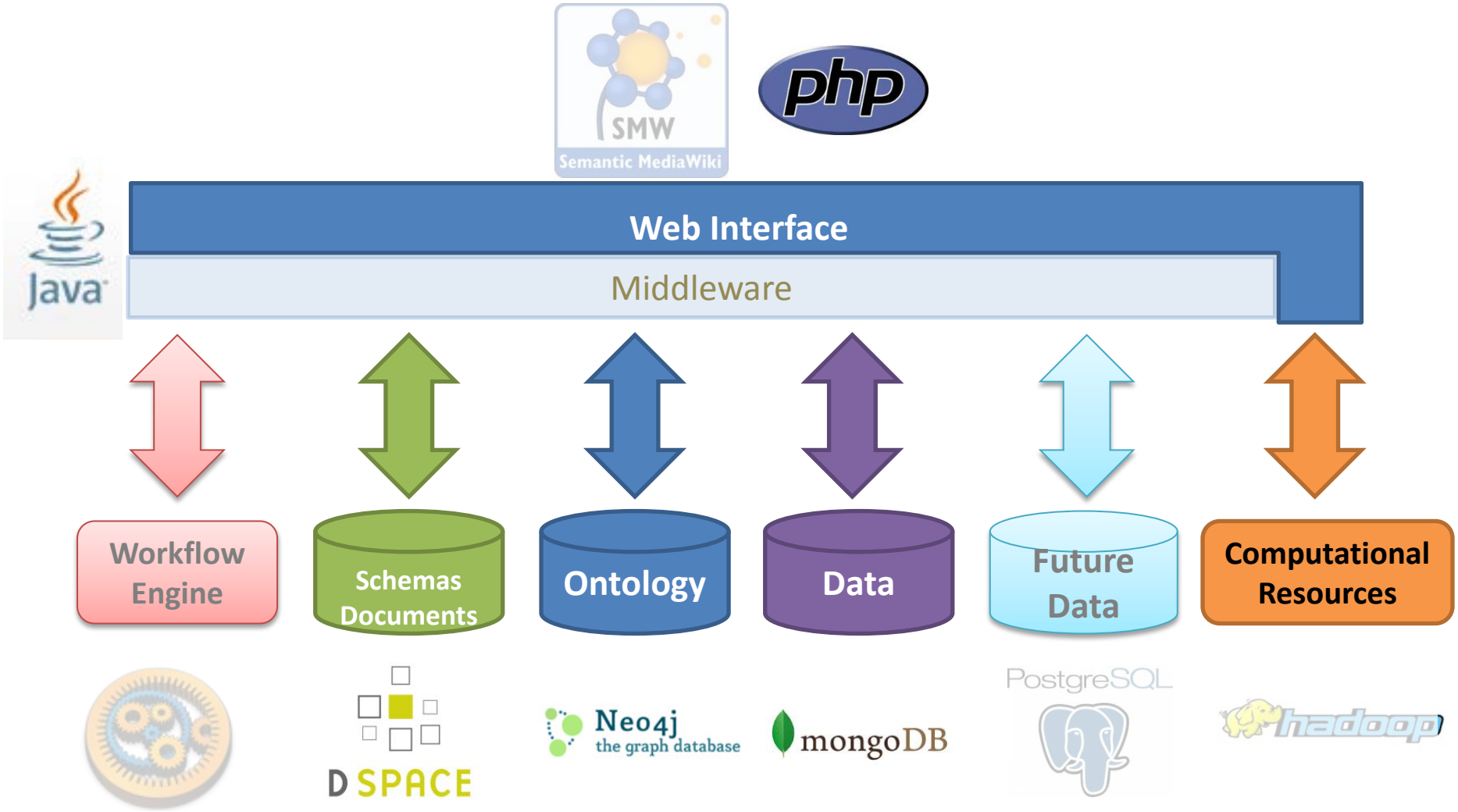
Tool Chain



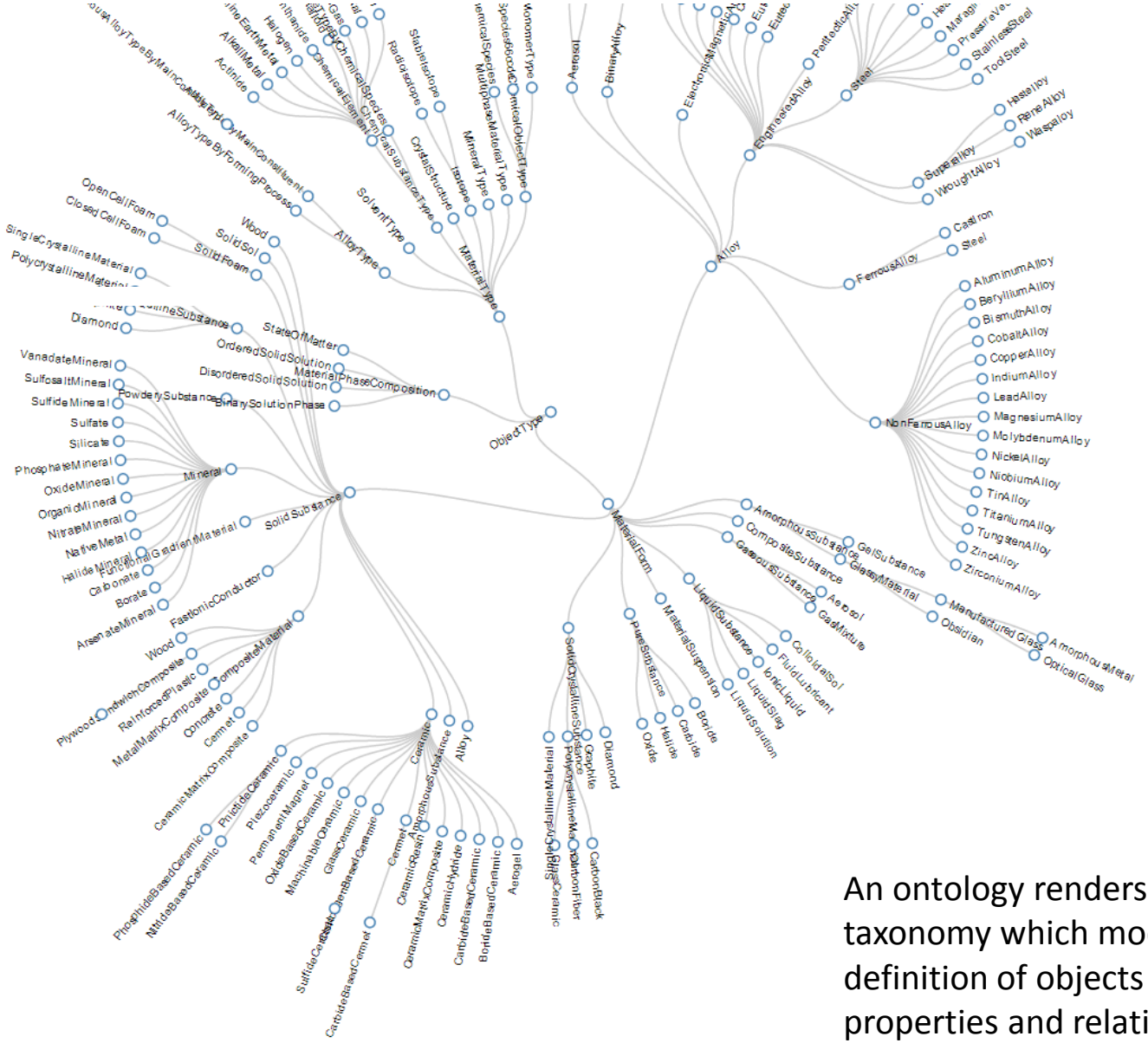
Materials Ontology currently being developed

Note this is a work in progress

Architectural Strategy



Prototype MGI Ontology



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

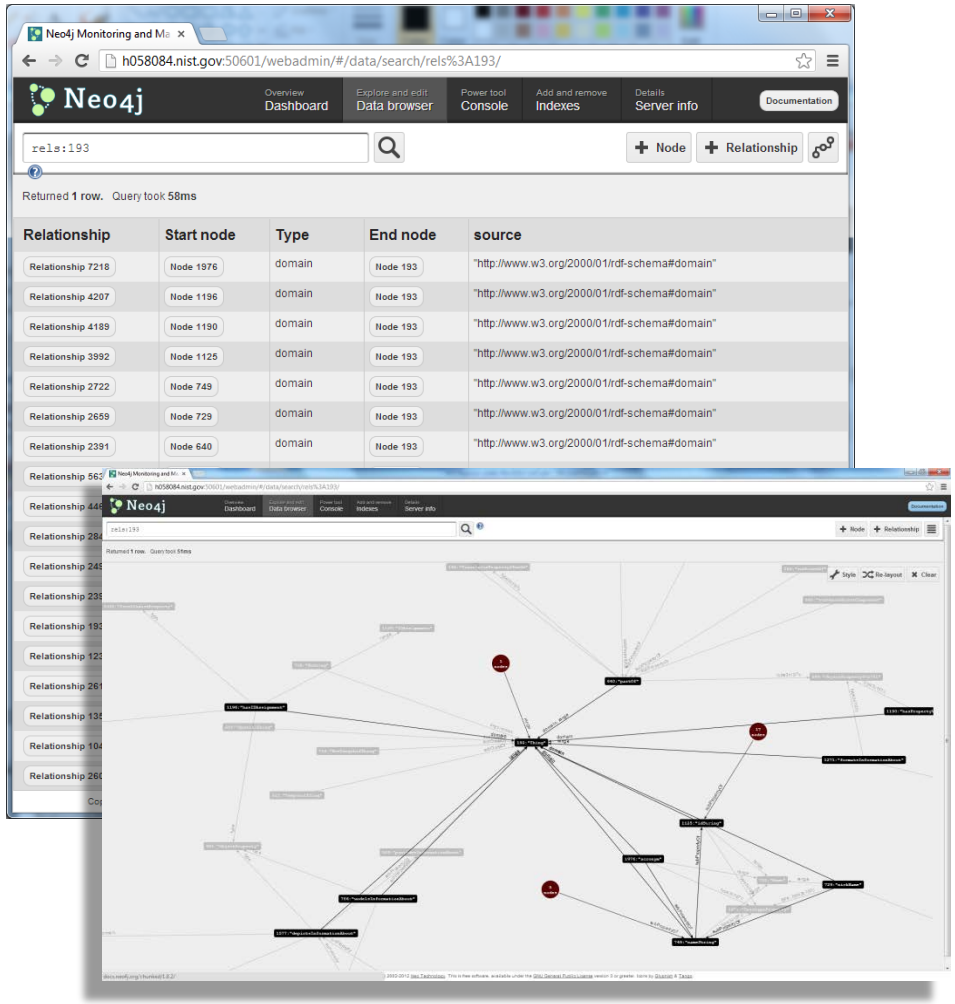
- Objects, Materials, and Events
- Physical Properties
- Documents
- Data Objects & Types
- People & Organizations
- Software
- Relations among these

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

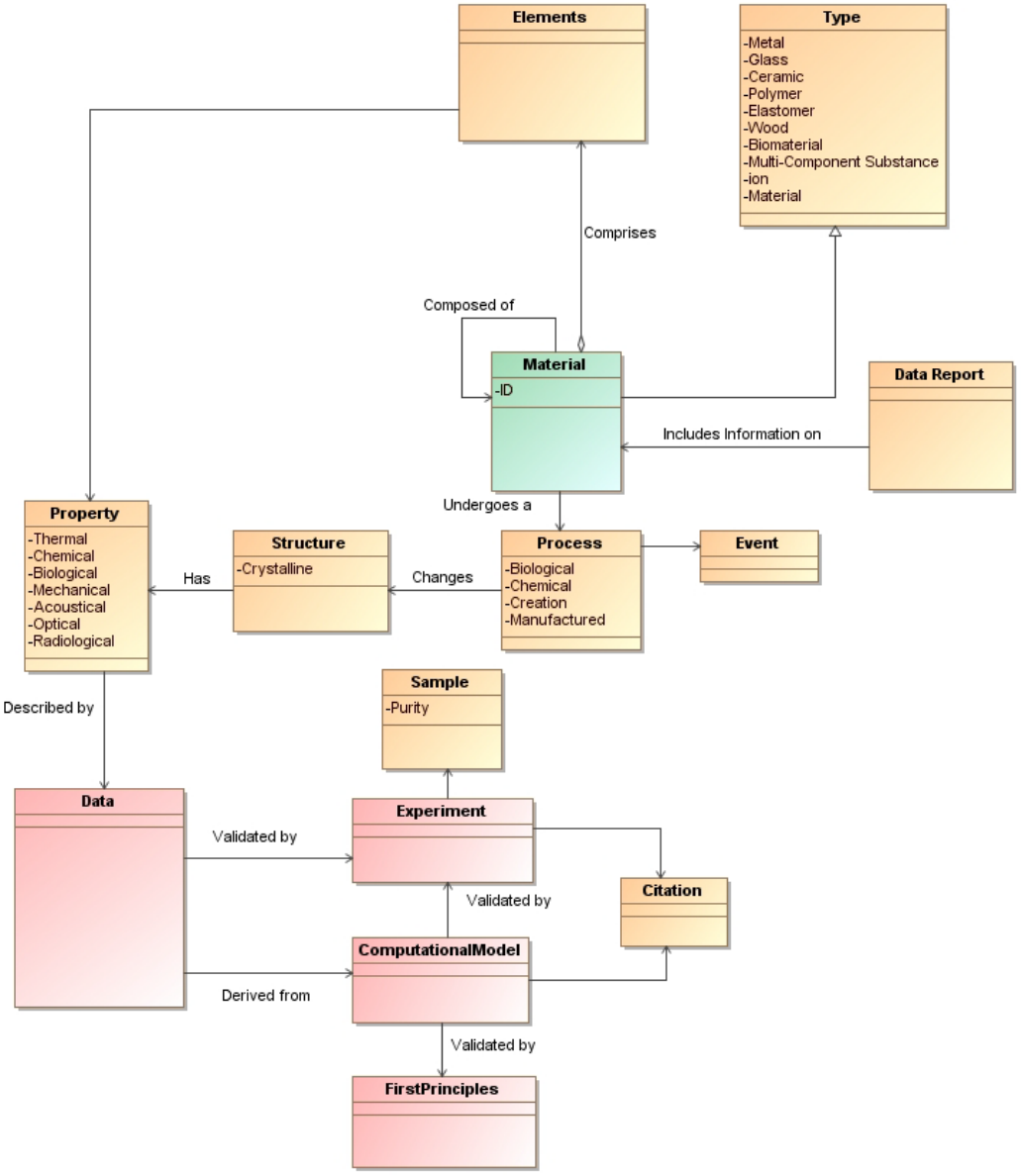
Sources

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

Tools

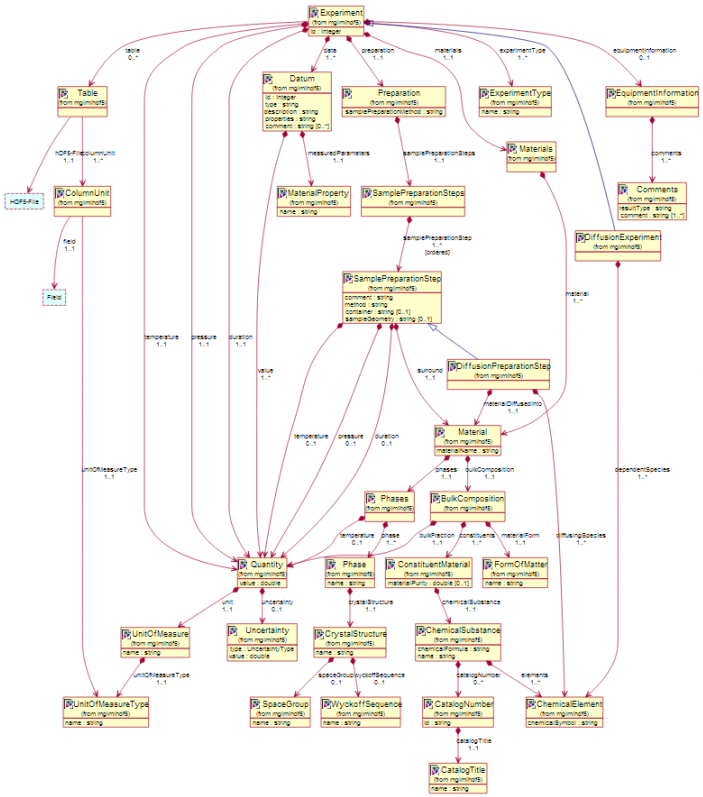
- UML (Unified Modeling Language)
- Semantic Web (RDF, OWL)

Note: This is a generalized model depicting overall structure

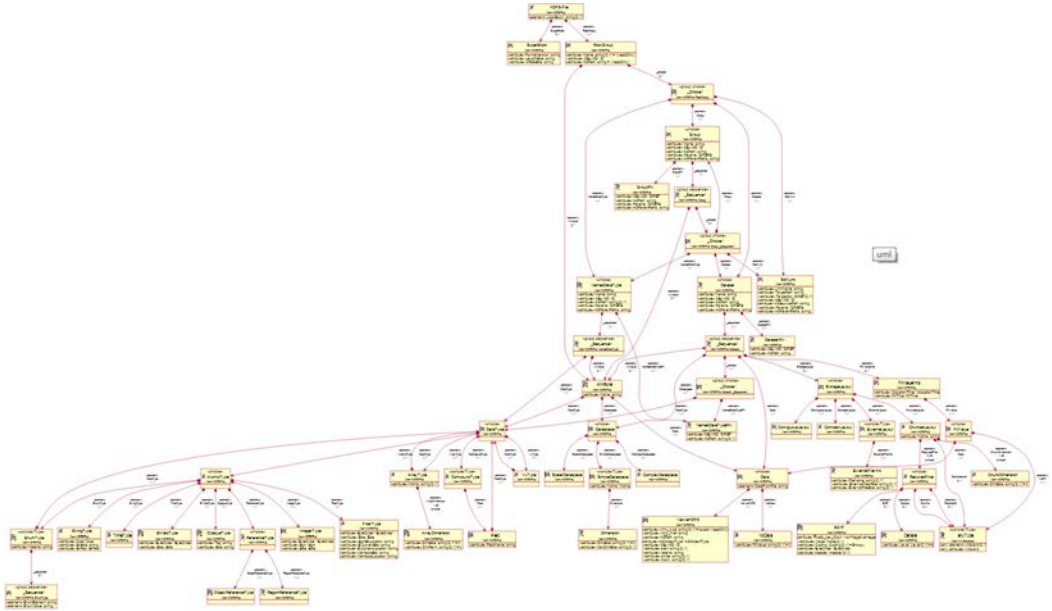


Encoding CALPHAD Data

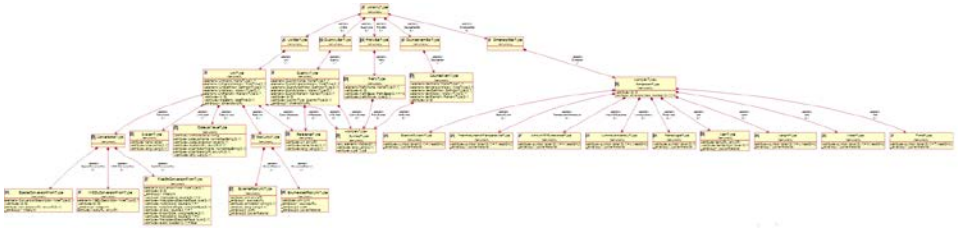
Core CALPHAD



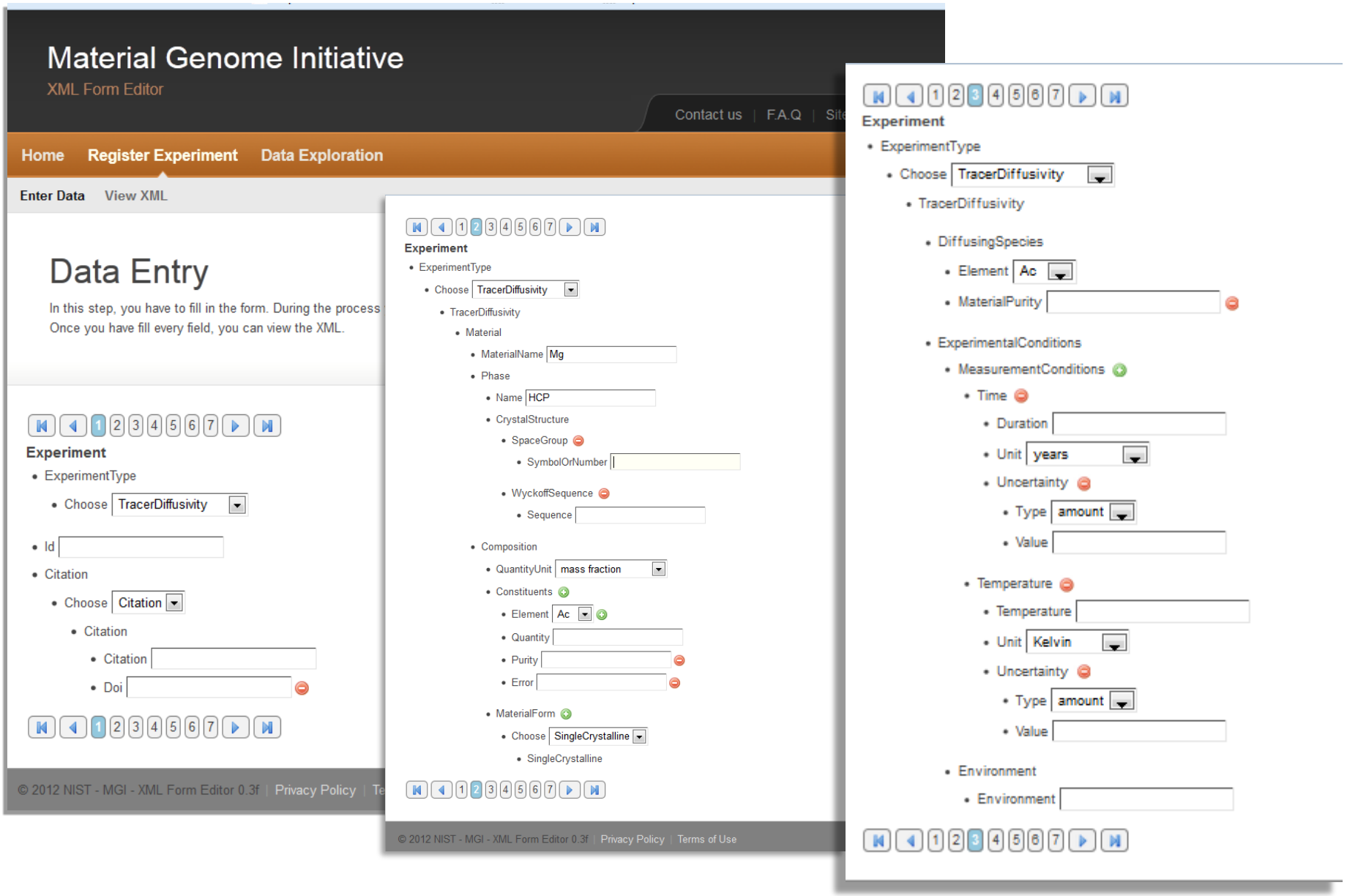
Tabular Data (HDF5)



Measurement Units (Units ML)



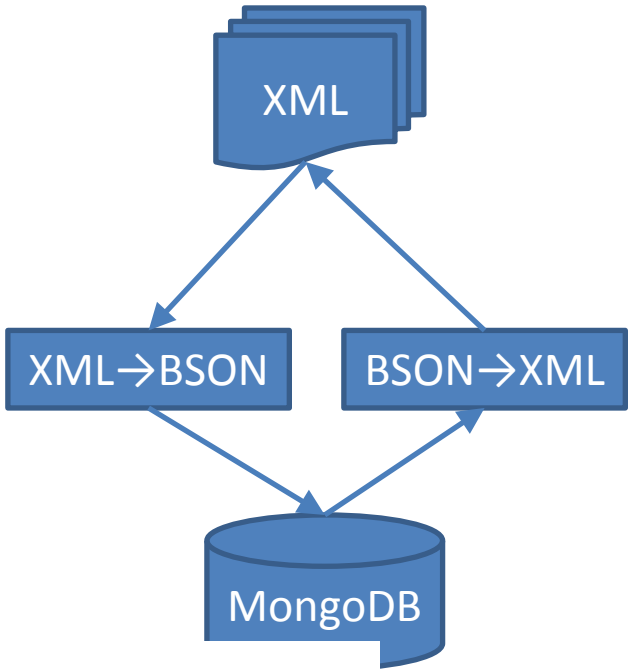
Data Collection: Tracer Diffusivity Test Schema



Snap-Shot of XML Format

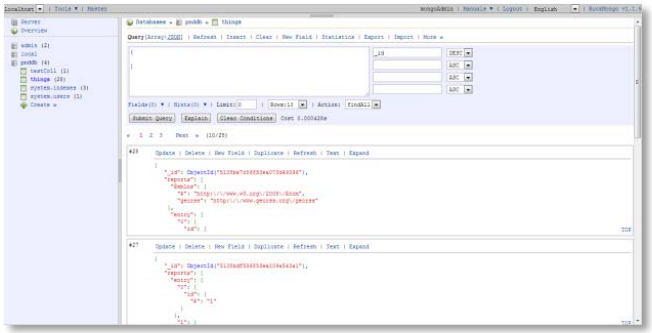
- Why XML?
  - Internationalization
  - Flexibility
  - Transformability
  - Interoperability
  - Longevity
  - Web-Enabled
  - Available Resources

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")
  ]
}
```

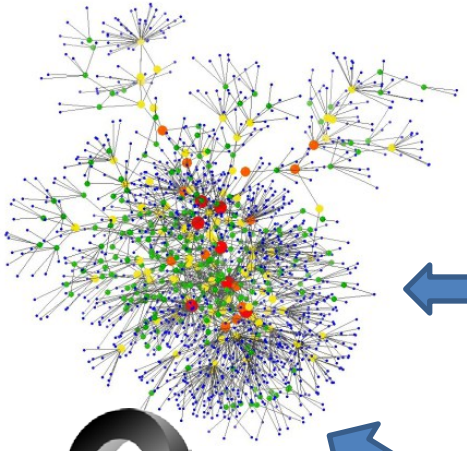


Future Data Informatics

User interface

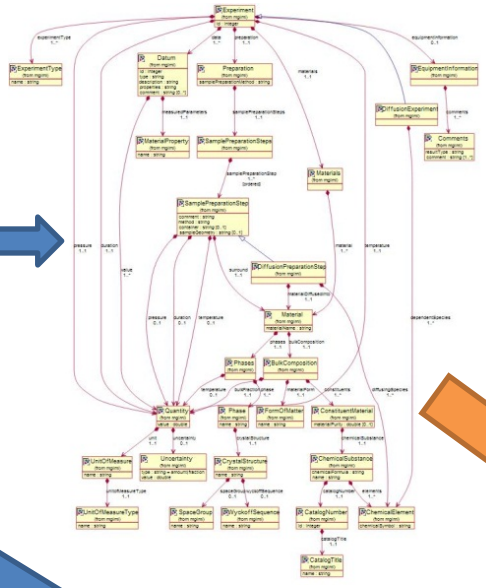
Data Capture

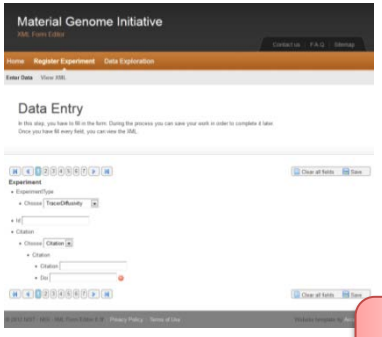
Ontology



JSON

XML Schema





Data Tools:
Statistics; Machine Learning

Various Database Platforms

mongoDB

D SPACE

Neo4j
the graph database

UNIFIED MODELING LANGUAGE

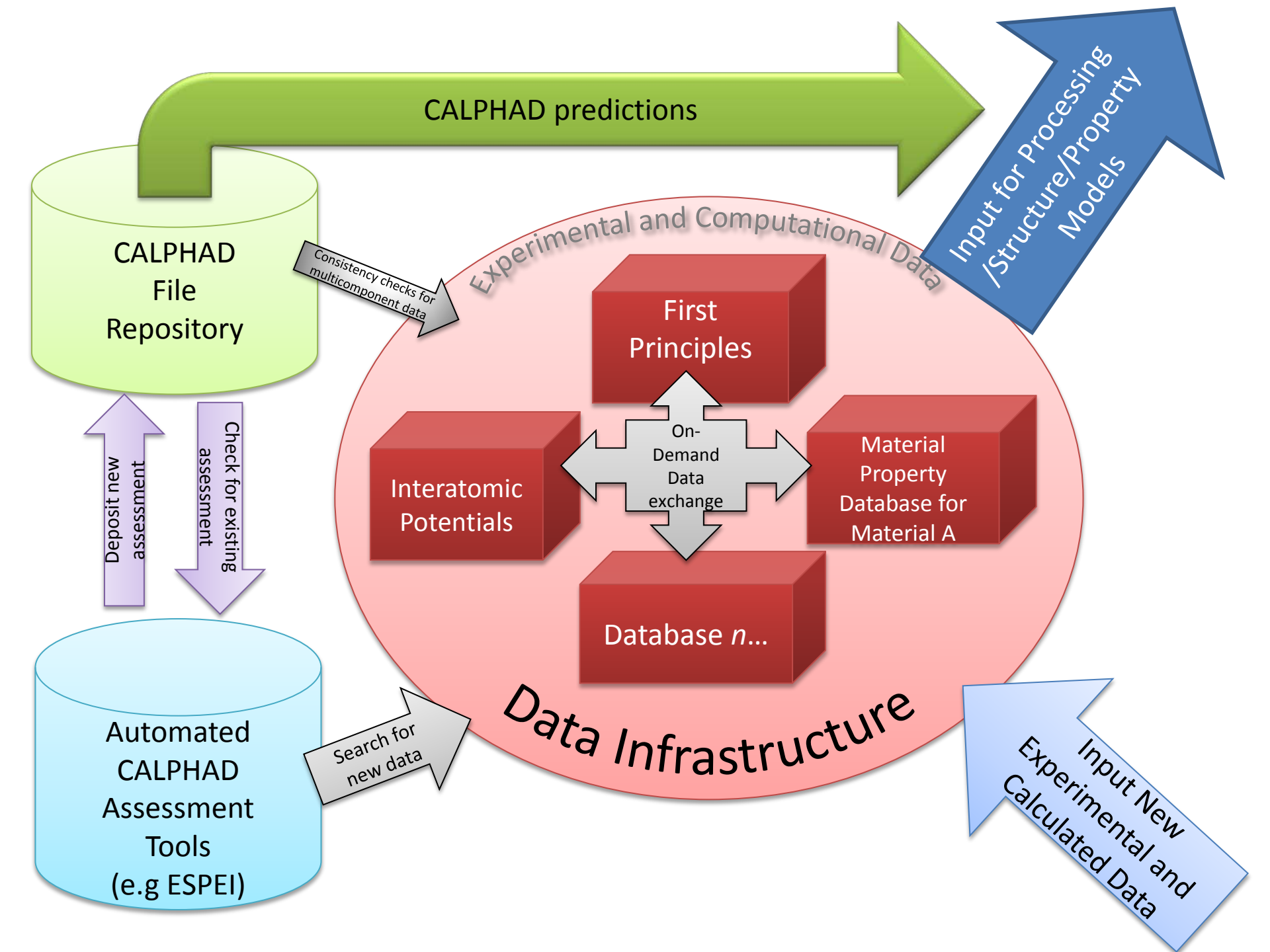
Java

RDF

XML

php

Semantic Web


CALPHAD predictions
Input for Processing /Structure/Property Models
CALPHAD File Repository
Consistency checks for multicomponent data
Experimental and Computational Data
First Principles
Interatomic Potentials
On-Demand Data exchange
Material Property Database for Material A
Database n…
Data Infrastructure
Deposit new assessment
Check for existing assessment
Automated CALPHAD Assessment Tools (e.g ESPEI)
Search for new data
Input New Experimental and Calculated Data

Future: Needs Community Input

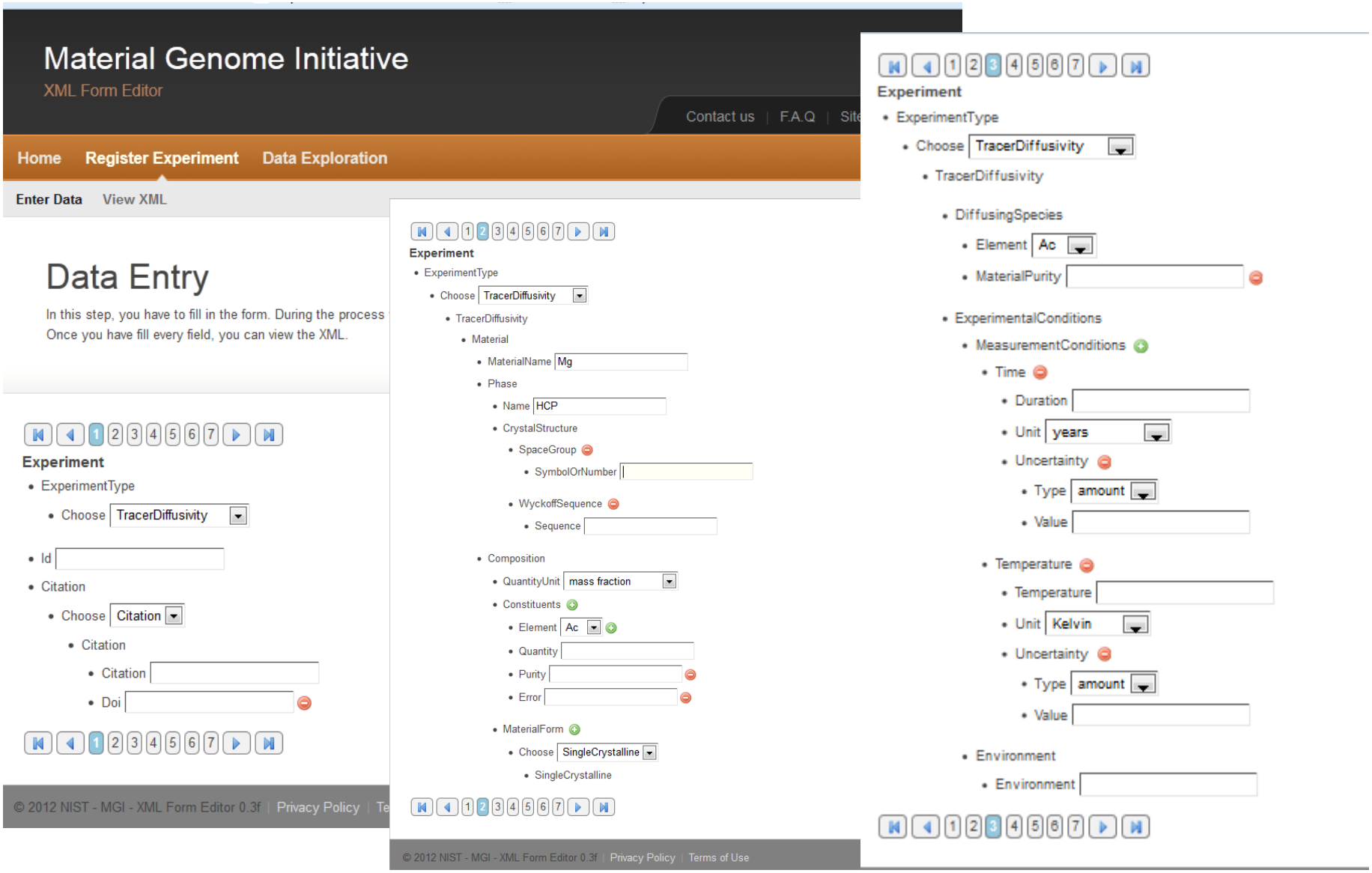
- Need feedback on developing infrastructure
  - Community needs to help define data standards (XML Schemas) – Long term benefit better data for everyone
  - Willingness to share pre-competitive data and respect for data shared by others (need to reference contributed data)

Want to know more: Attend the NIST Diffusion Workshop, May 9-10 at NIST

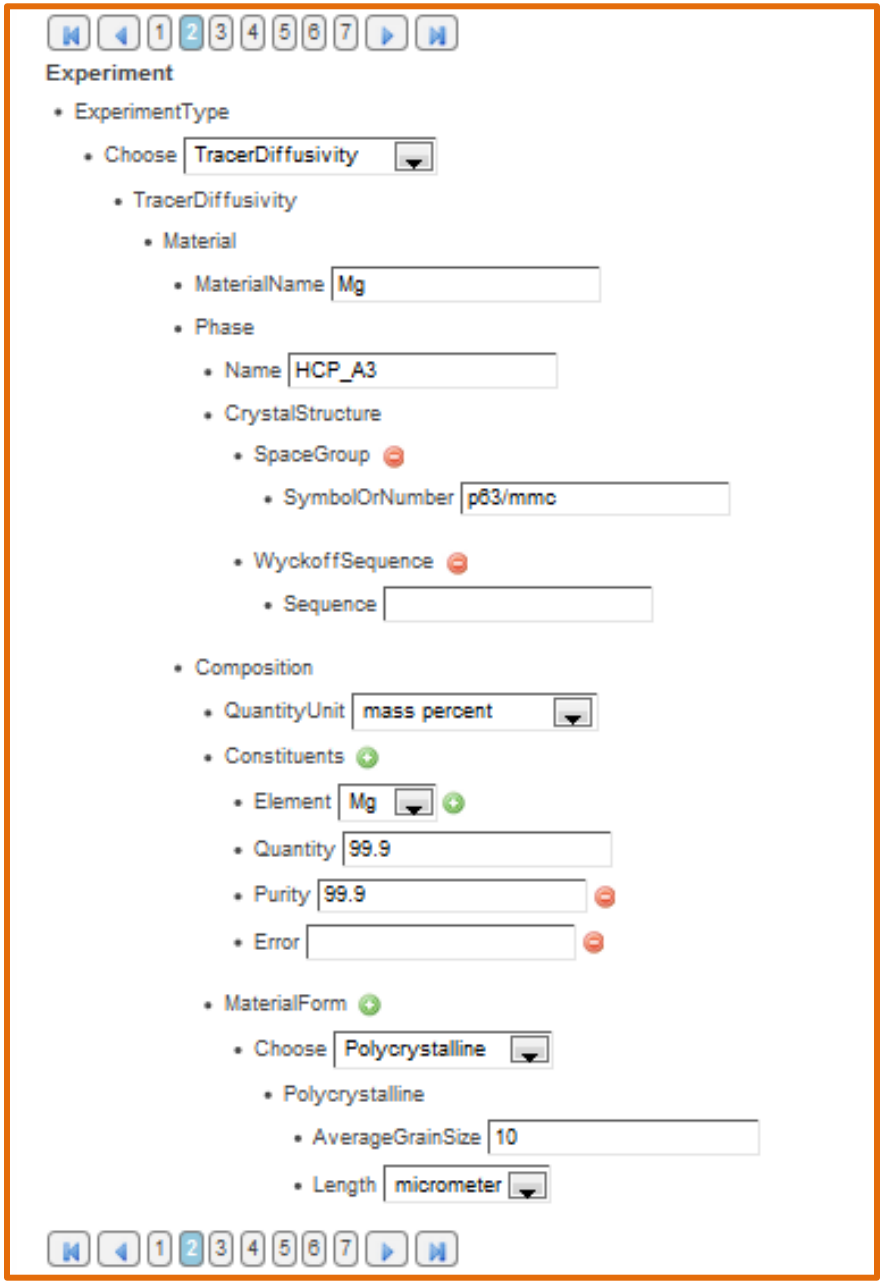
http://www.nist.gov/mml/msed/thermodynamics_kinetics/Diffusion-Workshop-Group.cfm

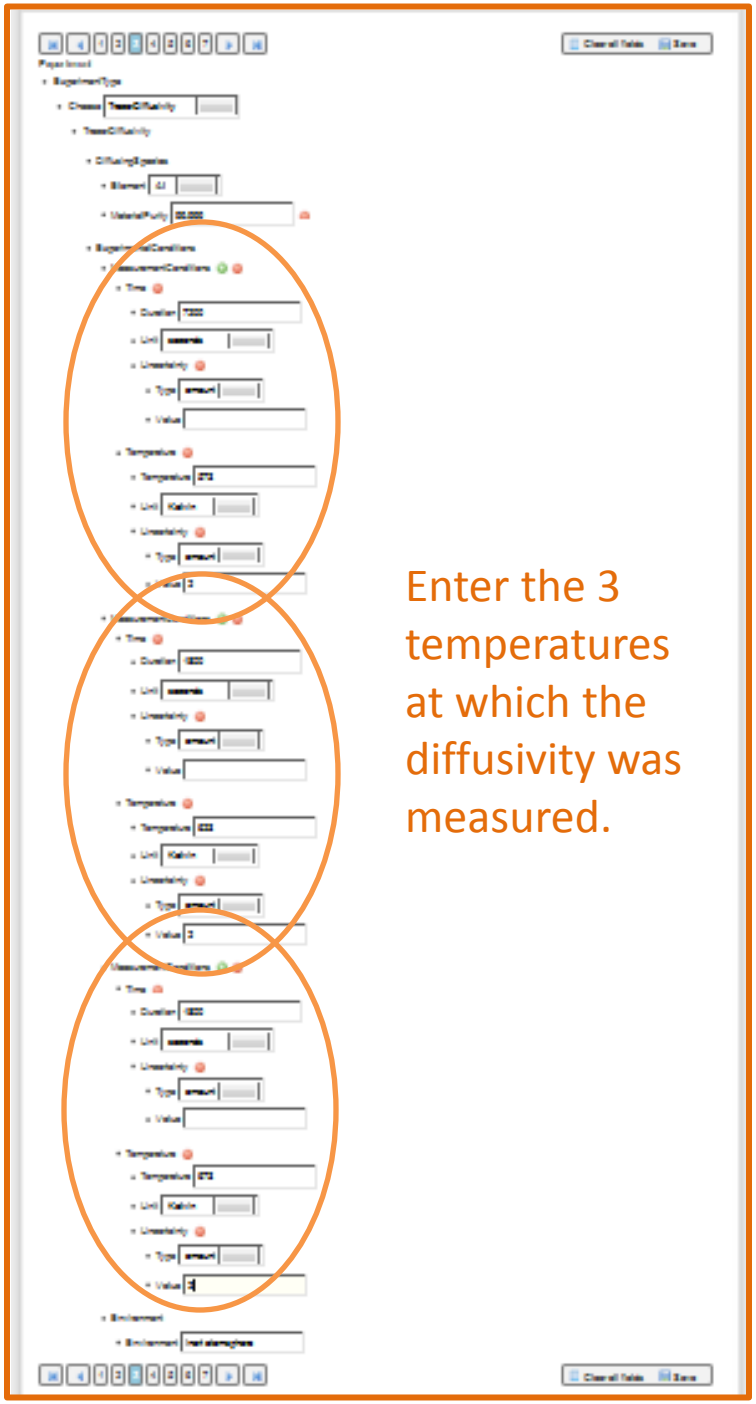
Extra slides on a test data schema for tracer diffusivity

Data Collection: Tracer Diffusivity Test Schema

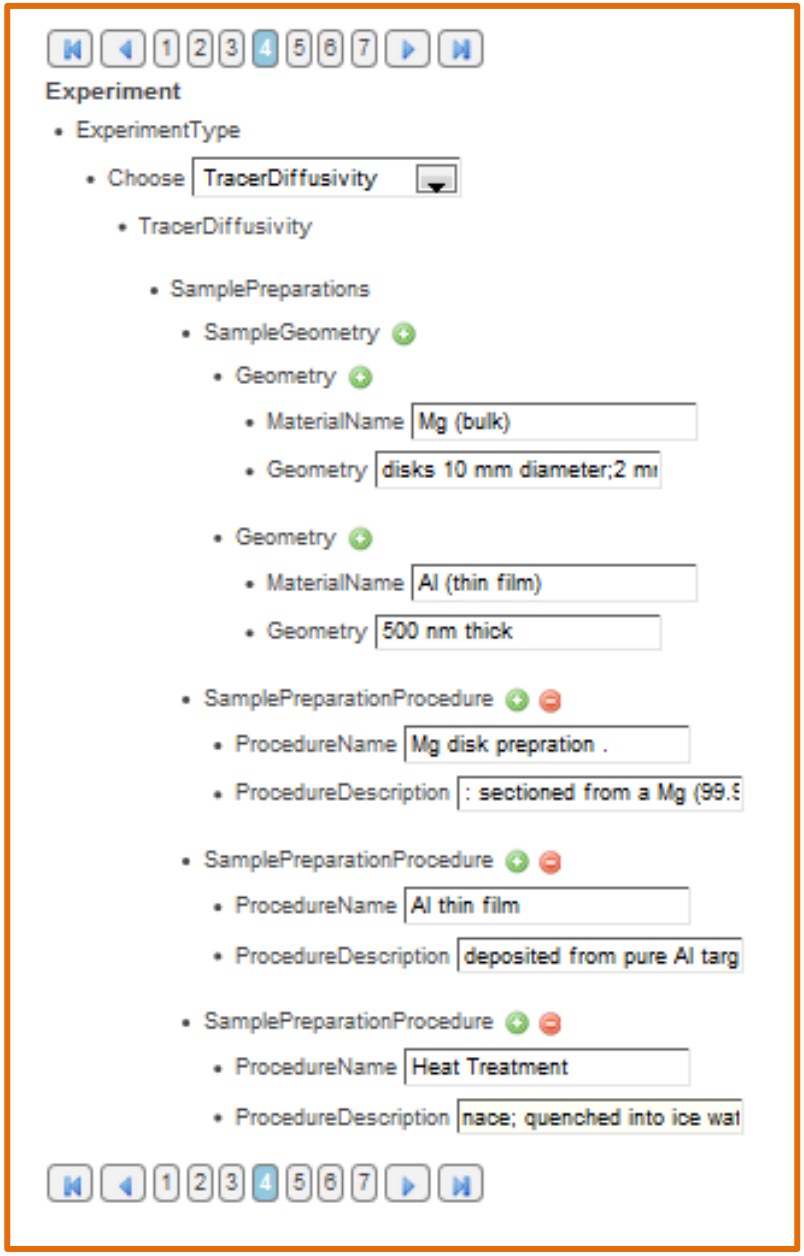


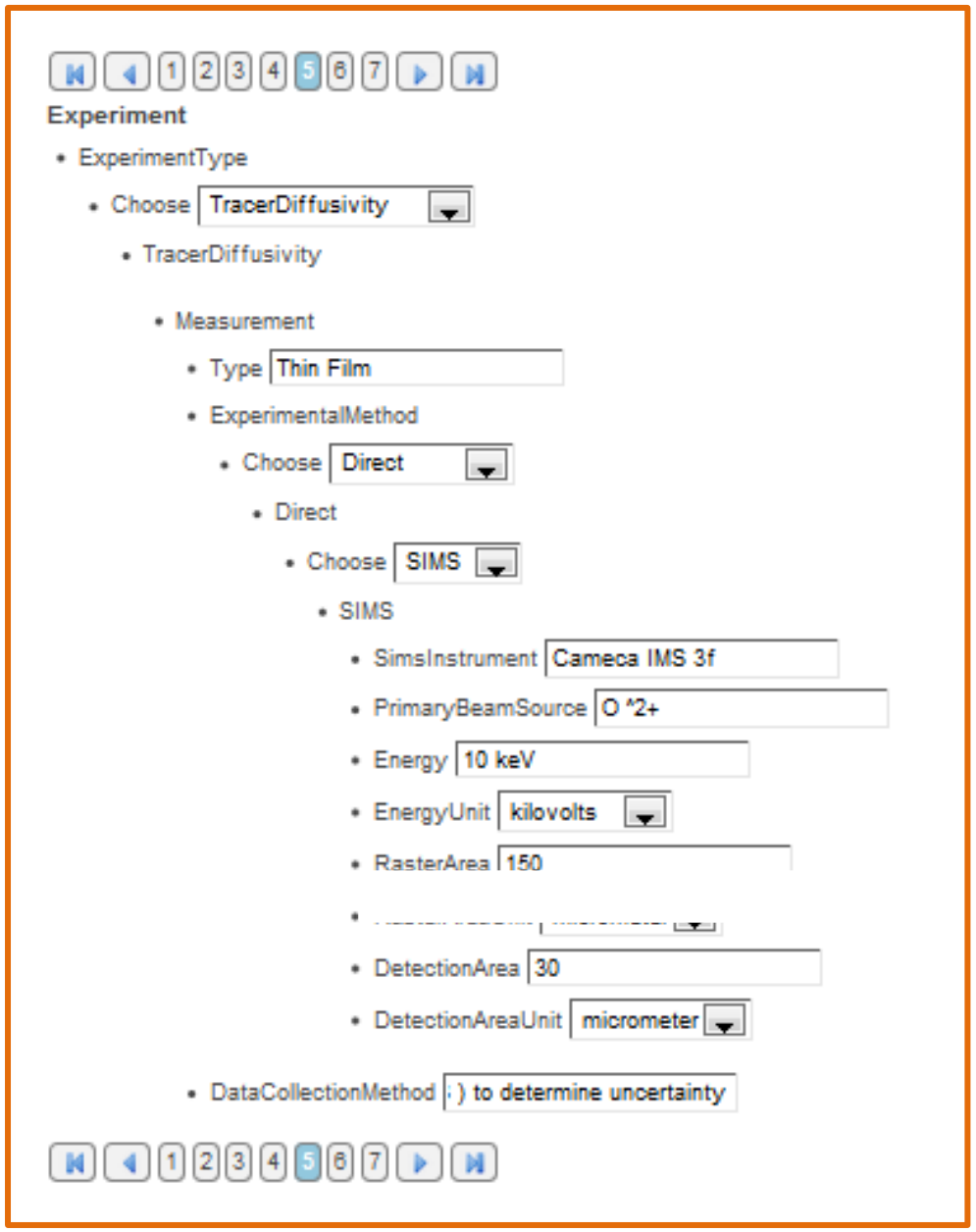
Example Entry

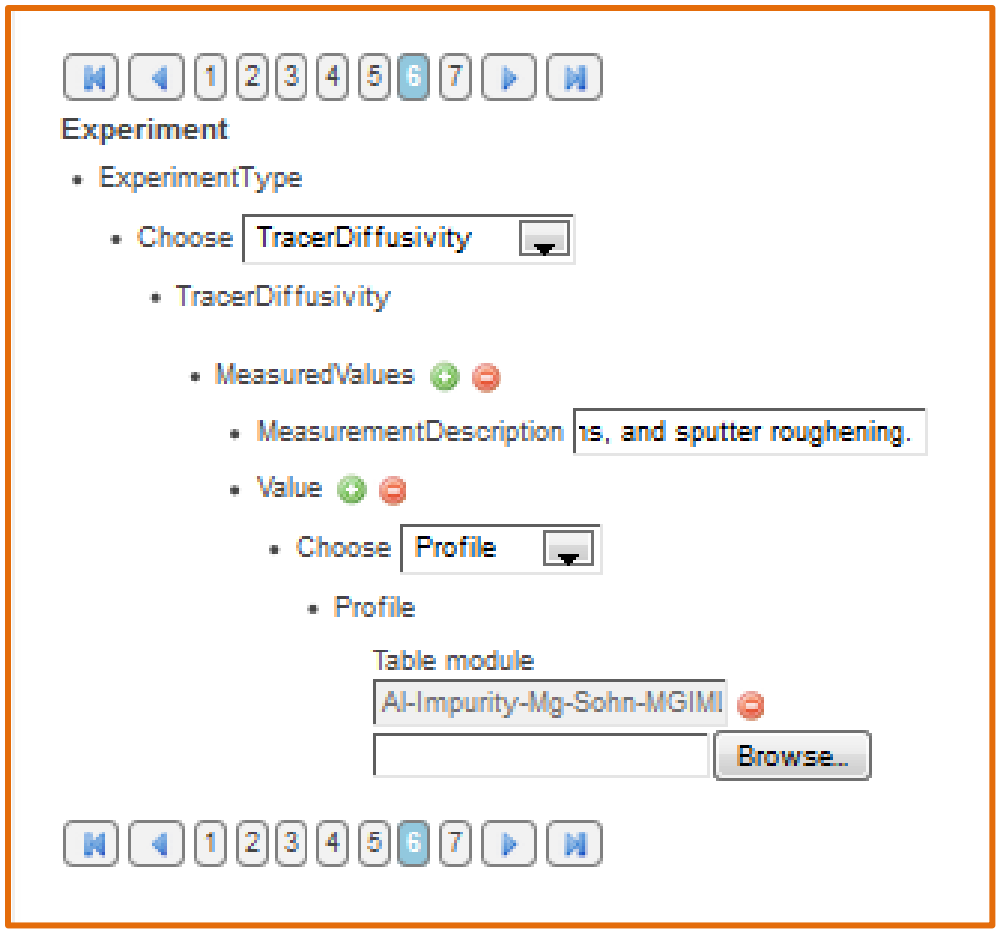


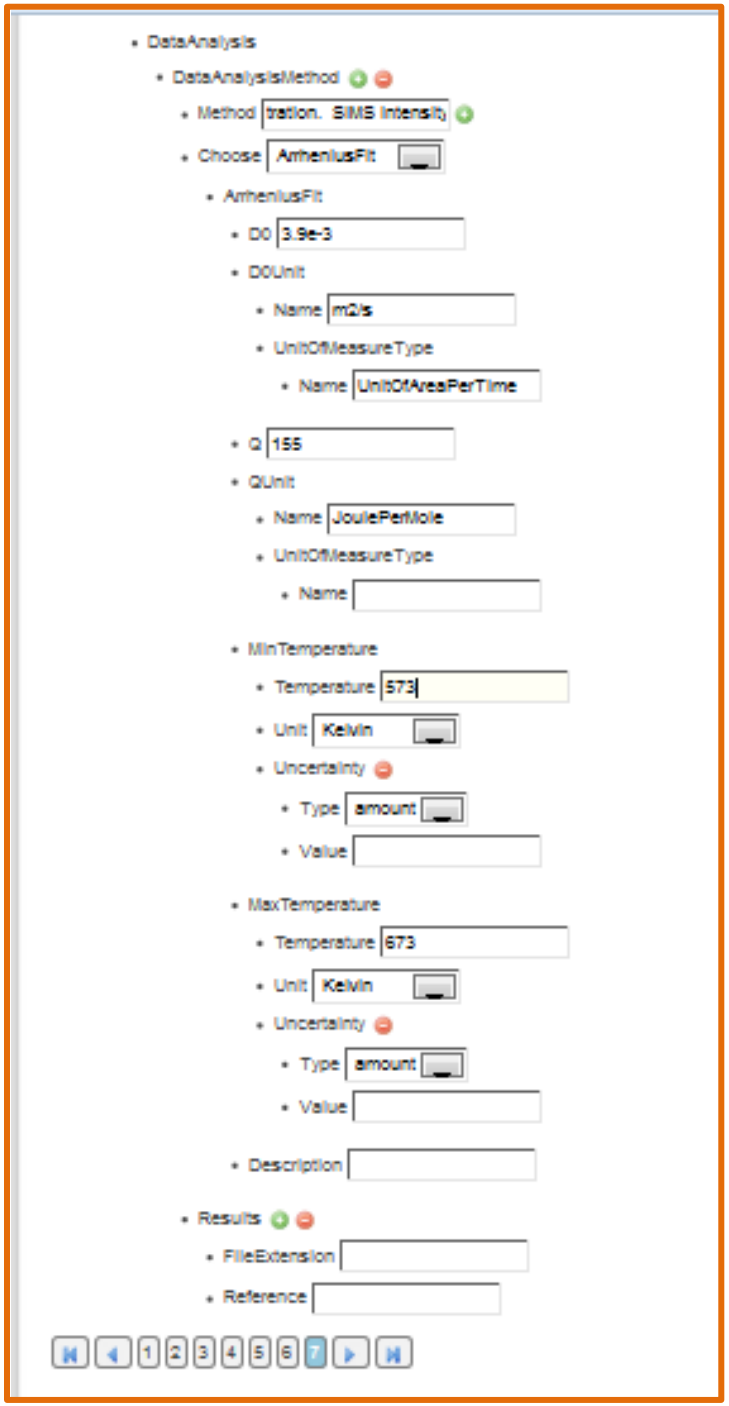


Enter the 3 temperatures at which the diffusivity was measured.


1 2 3 4 5 6 7
Experiment
ExperimentType
Choose TracerDiffusivity
TracerDiffusivity
SamplePreparations
SampleGeometry
Geometry
MaterialName Mg (bulk)
Geometry disks 10 mm diameter;2 mı
Geometry
MaterialName Al (thin film)
Geometry 500 nm thick
SamplePreparationProcedure
ProcedureName Mg disk prepration .
ProcedureDescription : sectioned from a Mg (99.9
SamplePreparationProcedure
ProcedureName Al thin film
ProcedureDescription deposited from pure Al targ
SamplePreparationProcedure
ProcedureName Heat Treatment
ProcedureDescription nace; quenched into ice wat
1 2 3 4 5 6 7


1 2 3 4 5 6 7
Experiment
ExperimentType
Choose TracerDiffusivity
TracerDiffusivity
Measurement
Type Thin Film
ExperimentalMethod
Choose Direct
Direct
Choose SIMS
SIMS
SimsInstrument Cameca IMS 3f
PrimaryBeamSource O ^2+
Energy 10 keV
EnergyUnit kilovolts
RasterArea 150
DetectionArea 30
DetectionAreaUnit micrometer
DataCollectionMethod) to determine uncertainty
1 2 3 4 5 6 7


1 2 3 4 5 6 7
Experiment
ExperimentType
Choose TracerDiffusivity
TracerDiffusivity
MeasuredValues
MeasurementDescription ıs, and sputter roughening.
Value
Choose Profile
Profile
Table module
Al-Impurity-Mg-Sohn-MGIMI
Browse...
1 2 3 4 5 6 7


DataAnalysis
DataAnalysisMethod
Method tration. SIMS Intensity
Choose ArrheniusFit
ArrheniusFit
D0 3.9e-3
D0Unit
Name m2/s
UnitOfMeasureType
Name UnitOfAreaPerTime
Q 155
QUnit
Name JoulePerMole
UnitOfMeasureType
Name
MinTemperature
Temperature 573
Unit Kelvin
Uncertainty
Type amount
Value
MaxTemperature
Temperature 673
Unit Kelvin
Uncertainty
Type amount
Value
Description
Results
FileExtension
Reference
1 2 3 4 5 6 7